

THE SLAYING OF LADY MONDEGREEN,
BEING A STUDY OF
FRENCH TONAL ASSOCIATION AND ALIGNMENT
AND THEIR ROLE IN SPEECH SEGMENTATION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Pauline Susan Welby, B.A., M.A.

* * * * *

The Ohio State University

2003

Dissertation Committee:

Professor Mary E. Beckman, Adviser

Professor Christopher H. Brew

Dr. Hélène Lœvenbruck

Professor Shari R. Speer

Approved by

Adviser

Department of Linguistics

© *Copyright by*
Pauline Susan Welby
2003

ABSTRACT

This dissertation investigates French intonational structure and the role of language-specific intonational cues to speech segmentation, the listener’s task of determining where words begin and end. The Lady Mondegreen of the title never existed: she was created in the imagination of a young child who missegmented a line from one of her favorite poems.

French intonation is characterized by an obligatory fundamental frequency (f_0) rise on the last syllable of a phrase (1a) and an optional early rise near the beginning (1b).

- (1) a. Les gamins SAGES jardinaient.
‘The good kids were gardening.’
- b. Les GAMINS SAGES jardinaient.

A production experiment examined the text-to-tune alignment of these rises, the timing of f_0 peaks and valleys with respect to segmental landmarks. The low starting point of the early rise was consistently located at the function word-content word boundary (e.g., the determiner-noun boundary in *les* | *gamins sages*). There was also sometimes an f_0 inflection or elbow at that boundary, even where there was no early rise. The evidence supports a model in which the early rise is a bitonal phrase accent and the late rise a bitonal pitch accent.

Three perception experiments examined whether the early rise and f_0 elbow serve as cues to content word beginnings. Natural speech was recorded and f_0 manipulated

through resynthesis. In Experiment 1, participants listened to noise-masked minimal and near-minimal pairs of phrases differing only in segmentation and presence versus absence of an early rise (e.g., *le niveau de mécénat* ‘the level of patronage’/ *le niveau de mes sénats* ‘the level of my senates’) and indicated what they heard. Listeners interpreted an early rise as a marker of a content word beginning. In Experiment 2, the timing of the early rise was manipulated in minimal pairs of nonwords. Sequences like [me.la.mõ.din] were perceived as two words (*mes lamondines* ‘my lamondines’) when the early rise started at the second syllable ([la]) and as a single word (*mélamondine*) when it started at the first syllable ([me]). Using a similar paradigm, Experiment 3 showed that not only early rises, but also simple f_0 elbows, cued content word beginnings.

to my niece

Charlotte Jane Ward

my sister

Lenore Christine Welby

my mother

Evelyn Louise Major Welby

and to the memory of my grandmother

Mary Anne Sweeney Welby

ACKNOWLEDGMENTS



The moment I first saw the red-brick gables and tower of Oxley Hall, home to Ohio State’s linguists, I thought to myself, “I am going to love coming to work here every day.” And I have.

I start by thanking my committee members, Mary Beckman, Chris Brew, Hélène Loevenbruck, and Shari Speer, each of whom has contributed to the dissertation and to my growth as a linguist.

I am very grateful to my adviser, Mary Beckman, for sparking my interest in phonetics and intonation and for her constant support and confidence in me and my work throughout my graduate career. Mary’s patience and guidance helped me find a dissertation topic that truly excited me, and her advice along the way was essential in

completing the dissertation. I have enjoyed and immensely benefited from our years of discussions, and I appreciate that her door is always open for her students.

My work on French intonation began as part of a Prosody and Processing seminar co-taught by Mary Beckman and Shari Speer in the summer of 2001. That seminar and Shari's own work on prosody and ambiguity in English helped me understand the importance of examining production and perception together. Shari also provided valuable input on the design of the perception experiments and gave many helpful comments on the manuscript.

Chris Brew also gave valuable feedback on the manuscript, as well as guidance on the analysis of the production data and on the corpus search in Chapter 5. Throughout the dissertation, I have used tools I learned auditing Chris's winter 2000 computational linguistics seminar.

Hélène Loevenbruck, of the Institut de la Communication Parlée in Grenoble, gave invaluable advice and assistance at every stage of the dissertation, from encouraging my nascent interest in French intonation, to helping me hammer out experiment details, to inviting me to ICP to run my experiments, to providing thoughtful, detailed comments on the final manuscript. I am grateful for her innumerable contributions to the dissertation, her generosity with her time, and her enthusiasm for my work.

I have had the good fortune to be part of the Ohio State Department of Linguistics. I thank the chair, Peter Culicover, and the former chair, Brian Joseph, for fostering a department culture that is both intellectually challenging and supportive.

I would like to thank a number of my professors at Ohio State for their contributions to my dissertation research: Keith Johnson, who always has time for his students and who gave me a great deal of advice on phonetics, statistics, and scripting; Julie

Boland, now at the University of Michigan, who introduced me to psycholinguistics; Beth Hume, who gave me valuable comments and a phonologist's perspective on my first paper on French intonation; and Christiane Laeufer, who provided helpful input as a member of my generals exam committee.

Outside of the classroom, I have benefited from the department's very active Speakers Series and discussion groups. I have learned a lot from my work as a research associate on various projects, and I thank the professors who supported me on their grants: Mary Beckman and Chris Brew, and Jan Edwards, Marios Fourakis, and Rob Fox of the Department of Speech and Hearing Science. I have also greatly benefited from the opportunity to teach undergraduate linguistics.

My years at Ohio State have been shared with many fun, smart, and kind people. I have fond memories of: hikes, late-night study sessions, trips to the movies and the symphony and the ballpark, practice talks, rotation ping pong, airport runs, conversations about linguistics and everything else, potlucks, World Cup 2002 matches at 3am on the big screens at Crew Stadium, lunches around the lab table, OSU football games, summer Ultimate, tea breaks, juggling, dinners at the Blue Nile....

I thank my friends for the fun times and for their support in many ways big and small that helped me to stay the course.

My thanks go especially to: Mike Calcagno, Nick Cipollone, Kordula De Kuthy, Robin Dautricourt, Robin Dodsworth, Anna Feldman, Janice Fon, Craig Hilts, Erhard Hinrichs and family, Tsan Huang, Kiwa Ito, Stefanie Jannedy, Sun-Hee Lee, Soyoung Kang, Matt Makashay, Julie McGory, Grant McGuire, Detmar Meurers, Amanda

Miller-Ockhuizen, Youn Noh, Bill Raymond, Anton Rytting, Liz Strand, Tom Stewart, Giorgos Tserdanelis, Shravan Vasishth, Jennifer Venditti, David White, Neal Whitman, Steve Winters, Peggy Wong, and Kiyoko Yoneyama.

And to Allison Blodgett, who has been a devoted friend since the early years of endless problem sets and labs and who is always there to talk through a data analysis problem or to convince me to pack up work for the day and go to the movies, as appropriate. To Mariapaola D’Imperio, Svetlana Godjevac, and Rebecca Herman, who from the very beginnings of my graduate career have inspired me by their example as meticulous and energetic researchers who always make time for important things like good food and friends. To Vanessa Metcalf, who helped ease the pressures of my final year in Columbus with many tasty salads and bad TV shows. I miss sitting on our front porch in Clintonville, sipping ice tea and watching the tomatoes and basil grow. To Laurie Maynell, who is always ready to listen and offer a word of encouragement or a joke. To Jen Muller, who makes me laugh and pumpkin pie. To Ben Munson, who I can always count on to be there for me, personally and professionally. I thank him for his comments on an early version of Chapter 4 and his input on the title of the dissertation. The working title was something deadly boring about “French rises” and speech segmentation. But in an e-mailed comment typical of his trademark wit, Ben observed that “‘French rises’ could mean anything . . . French people getting up in the morning, political revolts in France. . . .” So, thanks to him, the title is now less boring (though still deadly).

I would also like to thank my running buddy, Beth Hume. Our early morning runs and our races with the other members of the Binary Feet, the department running

club, were a great source of stress relief for me. And training to run a half-marathon gave me a lot of confidence in what I could do if I worked hard.

I have had a number of rewarding opportunities to spend time outside of my home department. I thank Julie Boland for inviting me to be a visiting student in her lab at the Rutgers University Center for Cognitive Science in the spring of 1999. I thank Monique and Pierre Philit for generously giving me a place to stay while I was taking classes in Lyon in the summer of 2000.

Much of my dissertation work was done at the Institut de la Communication Parlée in Grenoble, whose members hosted me in their lab first for short data collection trips in July and November 2002 and more recently for a longer stay as a visiting student. I thank the director of the lab, Jean-Luc Schwartz, the assistant director, Pierre Badin, and the former director, Pierre Escudier for their generosity in providing me with space and resources to pursue my research and for making me feel welcome. My dissertation has benefited immeasurably from ready access to experiment participants and from my interactions with ICP researchers. For many discussions about my work and theirs, I thank Christian Abry, Véronique Aubergé, Marion Dohen, Bleicke Holm, and Hélène Loevenbruck. I am very happy to be returning to ICP as a post-doc.

My new friends at ICP have made my move to Grenoble an enjoyable one not only professionally but also personally. I thank Virginie Attina, Marion Dohen, Frédéric Elisei, Matthias Odisio, Jihène Serkhane, and Annemie van Hirtum for quickly making me feel at home there. I am especially grateful to Fred for his support and many thoughtful gestures as I neared the *dernière ligne droite*.

Hélène Loevenbruck has been a wonderful friend to me, and I'm happy that we've been reunited in Grenoble. Hélène, Philippe Pichon, and their son Nicolas have been

uncommonly generous in opening their home to me. I thank them for many delicious dinners, lively conversations, weekend excursions, and for much laughter. Nicolas is only three, but he knows that weekends are for playing, not for working. And now that my dissertation is done, I'm looking forward to lots more time playing with him and his family.

I am grateful to have had abundant assistance, technical support, and resources throughout my graduate school career.

I thank my French consultants, colleagues, and friends who shared their native speaker intuitions with me and assisted me in preparing my experiments: Virginie Attina, Hélène Balladur, Marion Dohen, Frédéric Elisei, Christiane Lauefer, and Hélène Loevenbruck. Their time and their patience are greatly appreciated. I also thank the friends and colleagues who volunteered their time to participate in pilot experiments and gave helpful feedback: Mariette Bessac, Saveria Colonna, Robin Dautricourt, Ludovic Gourgas, and Amandine Penel.

Finishing the dissertation away from my home department posed certain difficulties and stresses. It has been a great source of relief to me to know I had people whose trans-Atlantic assistance I could depend on. For all sorts of help, including, but not limited to, filing paperwork and moving heavy boxes, I thank the following: Allison Blodgett, Jos Bosch, Hope Dawson, Robin Dodsworth, Josja Eggen, Anna Feldman, Kiwa Ito, Martin Jansche, Soyoung Kang, Laurie Maynell, Vanessa Metcalf, and Tom Stewart.

I am grateful to those people who made me feel like Columbus and the department were still my home when I came back to defend. I thank Mary Beckman, Allison Blodgett, and Robin Dodsworth for giving me places to stay, and Allison, Peggy

Wong, and Mike Armstrong for lending me their desks. Their generosity was a great help to me as I was preparing to defend and making revisions.

I thank the members of the administrative staffs at Ohio State and at ICP for their organizational support of my research and teaching: Judy Anderson, Frances Crowell, Jane Harper, Claudia Morettini, and Maxine Vargas at Ohio State, and Nadine Bioud and Yvette Gaudé at ICP.

I am grateful to the systems administrators and technical staff at both Ohio State and ICP for their calm and good-humored assistance in installing software, recovering lost data, and setting up equipment, often under time pressure. I thank Jim Harmon, Jeff McCune, Matt Hyclak, Jignesh Patel, and Mike Katkin at Ohio State, and Nino Medves, Monique Revil-Signorat, and Christophe Savériaux at ICP.

I thank Beau David Case, Miriam Conteh-Morgan, and the staff of the Ohio State University Libraries. The depth of the library's print and electronic holdings and help from the staff greatly facilitated my research. I have also appreciated the document delivery service.

I thank the developers of the software program Praat, Paul Boersma and David Weenink of the Institute of Phonetic Sciences at the University of Amsterdam, for their efforts in maintaining the program as an open-source, free resource. I thank Mary Beckman and Kyuchul Yoon for teaching me to program in Praat.

I thank Svetlana Godjevac for introducing me to the \LaTeX typesetting program and for the sage advice to learn it “before you're writing your dissertation!” I'm glad I listened to her. I also thank the following for their technical advice on the intricacies of the program: Mike Daniels, David Dowty, Frédéric Elisei, Martin Jansche, Matt Makashay, Detmar Meurers, Nathan Vaillette, and Shravan Vasishth.

Fellowship administered by the Ohio State Center for Western European Studies allowed me to study French and to focus on my research. Generous gifts from my sister and brother-in-law allowed me to buy a computer and a car, both of which helped me immensely, particularly on the long, cold nights when I was working on my generals exam.

I am grateful for many wonderful experiences that shaped me as a scholar and led me to linguistics. My French and linguistics classes at Brandeis University and my first linguistics class at Trinity College Dublin laid the groundwork for what was to come. And studying at the CUNY Latin/Greek Institute and teaching for many summers at the Johns Hopkins University Center for Talented Youth, I met many interesting, happy graduate students who encouraged me to apply to graduate school and gave me practical advice about getting in (and getting out).

Finally, I would like to acknowledge the support of my family and friends outside of linguistics. A number of my non-linguist friends have in some way shared my graduate school years with me. I am grateful to Beth Fisher for her steadfast friendship and encouragement through the years and across the miles. I thank her and Amit Shah and also their family for making me feel at home in their homes. I also thank David Pollack and Sue Coghlin, two friends who link my New York and Columbus worlds. I was delighted when David, an old friend from my Bronx Science days, arrived at Ohio State as a post-doc in the math department. It was a joy having him in Columbus and rekindling our friendship. And for the past couple of years, Sue has been cheering me on during her trips to Columbus to visit her family, my trips to New York to visit mine, and in many phone conversations.

I thank my mother, Evelyn Welby, my sister, Lenore Welby, my brother-in-law, Larry Ward, and my cousin, Jeannie Humphries, for their love and support through the years. My niece Charlotte I thank for changing my dream of graduation day—ever since I learned she was on the way, I imagine her in my arms, tugging at my tassel. I am also grateful for having had the love of my grandmother, Mary Welby. Grandma always encouraged me to continue my education, and as I finish this dissertation, I think of her.

Many years ago, Beth Hume advised our phonology class to “study a language with a good beach,” to study the language of a place where we wanted to be. I have taken certain liberties with that advice, but what Grenoble lacks in beaches, it makes up for in mountains. And I look forward to seeing them again soon.



VITA

October 12, 1971	Born - Flushing, New York
summer 1990	Basic Ancient Greek Institute City University of New York Graduate School and University Center and Brooklyn College New York, New York
1991–1992	Classics and linguistics coursework Trinity College Dublin
1993	B.A. in classics, magna cum laude Brandeis University Waltham, Massachusetts
1998	M.A. in linguistics The Ohio State University Columbus, Ohio
1996–1997, 1998–1999, 2000–2002	Graduate Research Associate The Ohio State University
autumn 1997, summer 1999, autumn 2002 ..	Graduate Teaching Associate The Ohio State University
spring 1999	Visiting student Rutgers University Center for Cognitive Science Piscataway, New Jersey (Ohio State Linguistics Department Adjacent Technologies Fellowship)

summer 1999	Center for Cognitive Science Summer Fellow The Ohio State University
1999–2000	Title VI Foreign Languages and Area Studies Fellow (French) The Ohio State University
2003–present	Presidential Fellow The Ohio State University
2003–present	Visiting student Institut de la Communication Parlée Grenoble

PUBLICATIONS

Welby, Pauline. 2003. Effects of pitch accent type and status on focus projection. *Language and Speech* 46(1): 53–81.

Welby, Pauline. 2003. French Intonational Rises and their Role in Speech Segmentation. In *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology*, pp. 2125–2128. Geneva.

Welby, Pauline. 2002. The realization of early and late rises in French: A production study. In *Proceedings of the Speech Prosody 2002 Conference*, pp. 695–698. B. Bel and I. Marlien (eds.) Aix-en-Provence.

Cassidy, Steve, Pauline Welby, Julia McGory, and Mary Beckman. 2000. Testing the adequacy of query languages against annotated spoken dialog. In *Proceedings of the 8th Australian International Conference on Speech Science and Technology*, pp. 428–433.

Welby, Pauline and Neal Whitman. (1999). Ringe revisited: Comments on Ringe's probabilistic comparison method. *The Ohio State University Working Papers in Linguistics*: 52, 63–75.

FIELDS OF STUDY

Major Field: Linguistics

Specializations: Phonetics, psycholinguistics

TABLE OF CONTENTS

	Page
Abstract	ii
Dedication	iv
Acknowledgments	v
Vita	xv
List of Tables	xxi
List of Figures	xxii
Chapters:	
1. Introduction	1
1.1 The problem	1
2. Models of French Intonation	9
2.1 Basics of French Prosody	9
2.1.1 Primary accent and the late rise	10
2.1.2 Secondary accent and the early rise	11
2.1.3 Articulatory correlates	15
2.1.4 Stress in French: Property of the phrase or the word?	17
2.2 Background concepts in intonation	20
2.2.1 Superpositional approaches	20
2.2.2 Functional approaches	22
2.2.3 Autosegmental-Metrical approaches	23
2.2.4 Tonal Association and Alignment	24
2.3 Models of French Intonation	26

2.3.1	Delattre	26
2.3.2	Rossi	28
2.3.3	Martin	31
2.3.4	Di Cristo and Hirst	35
2.3.5	Mertens	42
2.3.6	Jun and Fougeron	45
2.3.7	Post	53
2.3.8	Intonation and the other aspects of French prosody	59
3.	Production Experiment 1: Tonal association and alignment	61
3.1	Y a pas de quoi en faire une thèse ?	61
3.2	Introduction	61
3.3	Methods	65
3.3.1	Materials	65
3.3.2	Subjects	67
3.3.3	Procedures	67
3.4	Results and Discussion	70
3.4.1	Rate manipulation	70
3.4.2	Accent patterns found	72
3.4.3	Occurrence of the LHLH pattern	93
3.4.4	Accent patterns not found	99
3.4.5	Tonal alignment of the early and late rises	99
3.4.6	Distinguishing the early rise and the late rise	122
3.4.7	French Intonational Structure	125
4.	Speech Segmentation	130
4.1	The speech segmentation problem revisited	130
4.2	Research in speech segmentation	132
4.2.1	Phonological pattern frequencies	133
4.2.2	Prosodic cues	136
4.2.3	Generalizations on the nature of speech segmentation	153
4.2.4	The Possible Word Constraint	154
4.2.5	Weighting of cues to speech segmentation	158
4.3	Prosodic cues to speech segmentation in French	160
4.3.1	The early intonational rise as a cue to speech segmentation	167
5.	The Early Rise	
	as a Cue to Speech Segmentation	170
5.1	An investigation of the role of the early rise in speech segmentation	170

5.2	Perception Experiment 1: Presence of the early rise as a cue to speech segmentation (Speech masked by noise)	170
5.2.1	Methods	172
5.2.2	Results	178
5.2.3	Discussion	182
5.3	Perception Experiment 2: Alignment of the early rise as a cue to speech segmentation (segmentation of nonwords)	184
5.3.1	Methods	184
5.3.2	Results	191
5.3.3	Discussion	197
5.4	Perception Experiment 3: Early elbow as a cue to speech segmentation (segmentation of nonwords)	200
5.4.1	Methods	202
5.4.2	Results	210
5.4.3	Discussion	214
5.5	General Discussion	216
6.	Summary and Conclusion	224
6.1	Goals of the dissertation	224
6.2	Summary of the findings	226
6.2.1	Production Experiment: Tonal association and alignment	226
6.2.2	Perception Experiment 1: Presence of the early rise as a cue to speech segmentation	228
6.2.3	Perception Experiment 2: Alignment of the early rise as a cue to speech segmentation	229
6.2.4	Perception Experiment 3: Early elbow as a cue to speech segmentation	230
6.3	Limitations of the study	232
6.4	Directions for future research	233
6.4.1	Association and alignment of the early rise in consonant- versus vowel-initial phrases	233
6.4.2	Intonational cues in online speech segmentation	234
6.4.3	Production of the early rise in difficult listening conditions	236
6.4.4	Articulatory/visual correlates of the early rise: Articulatory tracking	239
6.4.5	Applications to speech technology	240
6.5	The variety of cues to speech segmentation	241
6.6	Models of intonation	244
6.7	What counts as phonological?	245
6.8	Complementarity of production and perception studies	247

Appendices:

A.	Production Experiment Instructions	269
B.	Production Experiment Materials	271
C.	Perception Experiment 1 Instructions	276
D.	Perception Experiment 1 Materials	279
E.	Perception Experiment 2 Instructions	281
F.	Perception Experiment 2 Sample Answer Sheet	284
G.	Perception Experiment 2 Materials	285
H.	Perception Experiment 3 Instructions	287
I.	Perception Experiment 3 Materials	289

LIST OF TABLES

Table		Page
3.1	<i>Logistic regression results for independent variable number of content word syllables.</i>	96
3.2	<i>Logistic regression results for independent variable duration of AP.</i> . . .	97
3.3	<i>Coefficients for the early L elbow regression model.</i>	104
3.4	<i>Coefficients for the three early H regression models.</i>	106
3.5	<i>Coefficients for the late H regression model.</i>	116
3.6	<i>R² values for each speaker for regression analyses using as dependent variables (1) latency from the left edge of the last syllable of the AP to the late L elbow (l-el) tag and (2) latency from the right edge. The independent variable in both cases was duration of the last syllable, excluding any non-sonorant rhyme. For each speaker, the higher R² value is underlined.</i>	119
5.1	Points used by the Praat script in definition of f_0 curves for Perception Experiment 3.	205
5.2	Fundamental frequency values (in Hz) assigned to each of the points by the Praat script for each of the four conditions in Perception Experiment 3.	205

LIST OF FIGURES

Figure		Page
1.1	<i>Sample speech signal. The utterance is Vu le signal acoustique, la segmentation de la parole devrait pas être aussi évidente qu'elle l'est.... 'Looking at the acoustic signal, speech segmentation shouldn't be as easy as it is....'</i>	2
1.2	<i>An example utterance with a fundamental frequency curve superimposed on the spectrogram. It is the continuation of the example in Figure 1.1: ... mais la mélodie facilite la tâche! '... but the melody makes the task easier!'</i>	4
2.1	<i>Illustration of a prosodic phrase (et les gamins sages 'and the good kids') realized with (a) an obligatory late rise and (b) an optional early rise and an obligatory late rise. A gloss for the entire sentence is 'And the good kids were gardening'.</i>	11
2.2	<i>Association of (a) a pitch accent to a syllable, and (b) a boundary tone to the edge of a prosodic unit. T stands for 'tone'.</i>	23
2.3	<i>Schematicization of same f_0 curve with different text-to-tune alignments. Dots represent syllable boundaries. (a) Alignment consistent with a $L+H^*$ pitch accent associated with the syllable WX, (b) Alignment consistent with a L^*+H pitch accent associated with the syllable WX.</i>	25
2.4	<i>Delattre's "dix intonations de base" 'ten basic intonation contours'.</i>	27
2.5	<i>Example of Rossi's intonational morphemes applied to a short sentence. Glosses for the two pronunciations are (a) 'It's (that's) the castle that I bought.' (b) 'It's (that's) the castle (the one that I bought).' The example is adapted from Rossi (1985).</i>	29

2.6	<i>Examples of Martin's (1987) "prosodic markers" and their feature specifications. (a) Marie adore les fraises. 'Marie loves strawberries.' (b) La sœur de Paul s'en va. 'Paul's sister is leaving.'</i>	32
2.7	<i>Five different f_0 shapes (redrawing of Martin (2003)). Martin argues that a transcription system should capture the differences in shape. . .</i>	34
2.8	<i>Tonal templates (a) for the Tonal Unit (UT) and (b) for the Intonation Unit (IU) (from Di Cristo (2000)).</i>	36
2.9	<i>A detailed example of Di Cristo and Hirst's view of prosodic structure. The figure is chiefly a synthesis of Figures 1 and 4 of Di Cristo and Hirst (1993). The tonal conversion level, however, reflects additions to the model discussed in Di Cristo and Hirst (1996). The stylized f_0 curve reflects my understanding of Di Cristo and Hirst's view of tonal scaling, as discussed in Di Cristo and Hirst (1996). A gloss is 'My son and his neighbor argued with each other.'</i>	37
2.10	<i>Example of a prosodic phrase (et les gamins sages 'and the good kids') where (a) the peaks of the early rise and late rise have the same height, (b) the peak of the late rise is higher than that of the early rise, (c) the peak of the late rise is lower than that of the early rise.</i>	42
2.11	<i>Six predicted surface realizations of Accentual Phrase (AP) Jun and Fougeron (2002)).</i>	47
2.12	<i>Tonal associations, Jun and Fougeron (2002). Jun and Fougeron note that the L of the LHi phrase accent can spread over the function words preceding the first content word. The figure is adapted from Jun and Fougeron (2002).</i>	48
2.13	<i>Tonal associations, Jun and Fougeron (2000).</i>	49
2.14	<i>Revision of the AP of Jun and Fougeron (2002) proposed by Welby (2002).</i>	50
2.15	<i>Tonal specifications of the Intonation Phrase, Post (2000). Curly braces indicate a set from which a tone can be selected and parentheses indicate optionality. Tones marked with '*' are pitch accents and tones marked with '%' are boundary tones.</i>	55

2.16	<i>Two patterns with and without a L preceding the late H*: (a) %LH*H*, (b) %LH*LH*.</i>	58
3.1	<i>Possible tonal patterns of (a) the Early H Associated Tone Hypothesis (dashed lines) and (b) the Early H Trailing Tone Hypothesis (solid lines) for syllables of different lengths. The upper tier represents segments and the lower tier represents function word and content word syllables. Cw stands for content word and fw stands for function word.</i>	63
3.2	<i>Schematization of the tonal tagging of a critical phrase. (a) Tags added by hand. (b) Tags whose position was automatically calculated (bold). The reference points (arbitrarily labeled i, j, p, q) are indicated in italics. The dotted lines represent the regression lines calculated by the script that determined elbow position. The elbow tag was placed at the intersection of the two regression lines.</i>	69
3.3	<i>Speaking rate by subject.</i>	72
3.4	<i>Six predicted surface realizations of Accentual Phrase (AP) Jun and Fougeron (2002)).</i>	73
3.5	<i>AP types found as a percentage of the total. The L2H AP type, not predicted by Jun and Fougeron (2002), is discussed on p. 87.</i>	73
3.6	<i>(a) LHLH pattern in the AP et la mélanine in utterance-medial position. The full utterance is Le mélanome et la mélanine étaient étudiés à la fac. ‘Melanomas and melanin were studied at college.’ There is no leading plateau, but a fall from the preceding H to a L right at the function word-content word boundary. The initial AP of the utterance also has an LHLH pattern, with a slight fall to an early elbow. (b) LHLH pattern with a leading plateau in the utterance-initial AP le méli-mélo. The full utterance is Le méli-mélo va déconcentrer Mélanie ‘The mess is going to distract Mélanie.’ (c) LHLH pattern in the utterance-medial AP le linoléum. There is a fall from the preceding H to a low leading plateau. The full utterance is L’aluminium et le linoléum seront installés mercredi. ‘The aluminum and the linoleum will be installed on Wednesday.’</i>	75

3.7	<i>LHLH pattern with a leading low plateau and a low plateau between H1 and H2 in the AP Et le mélomane. The full utterance is Et le mélomane s'était disputé avec eux. 'And the music lover had argued with them.'</i>	76
3.8	<i>Fundamental frequency differences between two tones of the LHLH pattern. Difference is H2 – H1 for each LHLH token.</i>	76
3.9	<i>(b) LHLH with a high plateau extending from the early H to the late L elbow in the AP la monomanie. The full utterance is La monomanie doit être évitée par les empereurs. 'Monomania should be avoided by emperors.'</i>	77
3.10	<i>Schematization of the range of excursion sizes for the dip to the f_0 inflection point preceding the H of the late rise. In (d) there is a clear late elbow and a large "dip" from the early H to the late elbow. In (a) the early H and late L have the same f_0 value, and the dip size is 0.</i>	78
3.11	<i>Fundamental frequency of the late L elbow as a function of fundamental frequency of the early H for APs with the LHLH pattern.</i>	78
3.12	<i>Correlation between size of fall from H of early rise to L of late rise in APs with the LHLH with the duration of the target content words.</i>	79
3.13	<i>Examples of plateau realizations of the early H (H1). Target AP is underlined. (a) Plateau. Difference between end and beginning: 0.5 Hz. The utterance is <u>Et la limonade</u>a été versée par Anna. 'And the lemonade was poured by Anna.' (b) Rising plateau. Difference between end and beginning: 8.7 Hz. The utterance is <u>Et le linoléum</u>sera installé mercredi. 'And the linoleum will be installed on Wednesday.' (c) Plateau with perturbation (caused by the [n] of linoléum). Difference between end and beginning: 6.4 Hz. The utterance is L'aluminium <u>et le linoléum</u>seront installés mercredi. 'The aluminum and the linoleum will be installed on Wednesday.'</i>	81
3.14	<i>Example of plateau realization of the late H (H2). Target AP is underlined. The utterance is Et le magnanime peut être trouvé dans le café. 'And the magnanimous one can be found in the café.'</i>	83

3.15	(a) <i>LH pattern with a slight fall to an early elbow in the AP</i> le moulin. <i>The full utterance is</i> Le moulin va être apprécié par Hélène. ‘The mill is going to be appreciated by Hélène.’ (b) <i>LH pattern with a fall from the late H of the preceding AP to an elbow at the function word-content word boundary in the utterance-medial AP</i> la mignonne. <i>The full utterance is</i> La courageuse, la mignonne et le malveillant étaient observés par l’espion. ‘The courageous woman, the cute woman, and the spiteful man were being watched by the spy.’	85
3.16	(a) <i>LLH in the utterance-initial AP</i> le linoléum. <i>The full utterance is</i> Le linoléum sera installé mercredi. ‘The linoleum will be installed Wednesday.’ (b) <i>LLH in the utterance-medial AP</i> le minimum. <i>The full utterance is</i> Le maximum, le minimum et les écarts-types seront calculés par Manon. ‘The maximum, the minimum, and the standard deviations will be calculated by Manon.’	86
3.17	<i>L2H in the utterance-medial AP</i> le minime. <i>The full utterance is</i> Le randonneur, le mélomane et la forestière s’étaient disputés avec eux. ‘The hiker, the music lover, and the forester had argued with them.’ .	87
3.18	<i>LHH in the utterance-medial AP</i> la limonade. <i>The full utterance is</i> La grenadine, la limonade et l’Orangina ont été versés par Anna. ‘The grenadine, the lemonade, and the Orangina were poured by Anna.’ . .	89
3.19	<i>LHi in the utterance-medial AP</i> la mélanine. <i>The full utterance is</i> Le mélanome, la mélanine et le collagène étaient étudiés à la fac. ‘Melanoma, melanin, and collagen were studied at college.’	90
3.20	<i>The sentence</i> Romain ranima la jolie maman. ‘Romain revived the pretty mom.’ (a) <i>with broad sentence focus and (b) with narrow focus on the verb</i> ranima. <i>In the narrow focus condition, there is a low plateau following the focused verb. Figure adapted from Dohen et al. (2003a) and used with permission.</i>	91
3.21	<i>AP types found by number of function words in AP as percentages of the total.</i>	95
3.22	<i>Location of the early L elbow as a function of duration of the last function word syllable (hypothesized associated syllable). The dashed line is the regression line.</i>	101

3.23	<i>Distribution of the early L elbow. Note the peak at 0 ms. A difference of 0 ms indicates that the elbow was realized precisely at the function word-content word boundary.</i>	102
3.24	<i>LHLH pattern with two early elbows, one at the boundary between the conjunction et and the le (e2-el), the other at the boundary between the determiner and the first content word syllable (e-el). The utterance is Le maximum et le minimum seront calculés par Manon. ‘The maximum and the minimum will be calculated by Manon.’</i>	103
3.25	<i>Location of the early H relative to the Accentual Phrase beginning as a function of the duration of the first content word syllable.</i>	105
3.26	<i>Distribution of the early H.</i>	105
3.27	<i>Location of the early H relative to the Accentual Phrase beginning as a function of AP duration and number of preceding function words. The solid line is the regression line for the one function word condition and the dashed line is the regression line for the two function word condition.</i>	108
3.28	<i>Location of the early H relative to the Accentual Phrase beginning as a function of speech rate.</i>	110
3.29	<i>Slope of the early rise by accent type.</i>	111
3.30	<i>Location of the late H as a function of duration of the last syllable of the AP (for AP types LHLH, LH, LLH, and L2H).</i>	112
3.31	<i>Distribution of the late H.</i>	113
3.32	<i>Location of the late H as a function of duration of the last syllable of the AP by AP type. (a) LHLH (b) LH (c) LLH (d) L2H.</i>	115
3.33	<i>Distribution of the late L elbow.</i>	117
3.34	<i>Location of the late elbow (l-el) as a function of duration of the penultimate syllable of the AP by speaker. (a) Speaker 1. (b) Speaker 2. (c) Speaker 3. (d) Speaker 4. (e) Speaker 5. (f) Speaker 6. (g) Speaker 7.</i>	118
3.35	<i>Rise time of the late rise as a function of speaking rate.</i>	121

3.36	<i>Slope of the late rise.</i>	122
3.37	<i>Durations of (a) first content word syllables (mean = 134 ms, std. dev. = 29.04), (b) second content word syllables (for 3- and 4-syllable content words, where second syllable is non-final) (mean = 131 ms, std. dev. = 28.59), (c) final content word syllables (mean = 230 ms, std. dev. = 71.54), (d) final content word syllables, excluding any coda consonants (mean = 172 ms, std. dev. = 50.21).</i>	124
3.38	<i>Schematic example of an unattested alignment of the L of the late rise. The first row shows syllable boundaries. The star indicates that the pattern is not well-formed.</i>	129
4.1	<i>A 211-year old speech segmentation error. The intended message is l'abondance 'abundance'. Photo by Pauline Welby, 2002. Used by kind permission of the Musée de la Révolution Française, Vizille, France. .</i>	131
5.1	<i>Sample materials for Perception Experiment 1. (a) Target produced as the single three-syllable content word mécénat (natural production). (b) Target produced as a function word followed by a two-syllable content word mes sénats. (natural production). (c) Target produced as a single three-syllable content word mécénat. Fundamental frequency resynthesized (from the original in panel a) to delete the early rise. The new pattern is hypothesized to be consistent with a segmentation with a content word beginning at [se] (e.g., mes sénats). (d) Target produced as a function word followed by a two-syllable content word mes sénats. Fundamental frequency resynthesized (from the original in panel c) to add an early rise. This pattern is hypothesized to favor a segmentation with a content word beginning at [me] (e.g., mécénat). Note the similarity in f_0 between panels a and d and between panels b and c. .</i>	175
5.2	<i>Percentage of two-word responses in Perception Experiment 1.</i>	179
5.3	<i>Identical segment strings with different alignment patterns. The pattern in (a), with a L elbow at the boundary between [me] and [la], is consistent with the interpretation of [me] as the function word mes 'my' (et mes lamondines 'and my lamondines'). The pattern in (b), with an elbow at the boundary between [e] and [me], is consistent with the interpretation of [me] as the first syllable of a content word (et mélamondine 'and mélamondine').</i>	186

5.4	<i>Sample materials for Perception Experiment 2. (a) Target produced as the single 4-syllable content word mélamondine (natural production). (b) Target produced as a function word followed by a 3-syllable content word mes lamondines (natural production). (c) Target produced as the single 4-syllable content word mélamondine. Fundamental frequency resynthesized (from the original in panel a) to correspond to a segmentation with a content word beginning at [la] (e.g., mes lamondines). (d) Target produced as a function word followed by the 3-syllable content word mes lamondines. Fundamental frequency resynthesized (from the original in panel b) to correspond to a segmentation with a content word beginning at [me] (e.g., mélamondine). Note the similarity in f_0 between panels a and d and between panels b and c.</i>	187
5.5	<i>Position of the critical boundary (between the first and second syllable) for responses to the [me.la.mõ.din] (mélamondine/ mes lamondines) items in Perception Experiment 2.</i>	190
5.6	<i>Percentage of responses with a word boundary at the critical region in Perception Experiment 2. Standard error of the mean is indicated. . .</i>	192
5.7	<i>Comparison of responses with boundaries at the critical 1/2 syllable region according to alignment of the early rise.</i>	197
5.8	<i>Illustration of the realization of the L tone of an early rise at the left edge of a syllable ([mo]) formed by a resyllabification process (liaison). A gloss for the sentence is ‘My Arianella was going to spread the marmelade for him or her.’</i>	199
5.9	<i>LH pattern with a fall from the late H of the preceding AP to an elbow at the function word-content word boundary in the utterance-medial AP la mignonne. The full utterance is La courageuse, la mignonne et le malveillant étaient observés par l’espion. ‘The courageous woman, the cute woman, and the spiteful man were being watched by the spy.’ . .</i>	200
5.10	<i>LLH pattern with a fall from the late H of the preceding AP to an elbow at the function word-content word boundary in the utterance-medial AP le mélomane. The full utterance is Le randonneur, le mélomane et la forestière s’étaient disputés avec eux. ‘The hiker, the music lover, and the forester had argued with them.’</i>	201

5.11	<i>Target four-syllable sequence [me.la.mɔ̃.din] (mélamondine/ mes lamondines) embedded in carrier sentence for Perception Experiment 3. .</i>	204
5.12	<i>Sample materials for Perception Experiment 3. All targets were produced as a function word followed by a 3-syllable content word. (a) Target has a LHLH pattern with the early elbow at the starting at the beginning of the first target syllable, a pattern hypothesized to be compatible with a 4-syllable content word interpretation (mélamondine). (b) Target has a LLH pattern with the elbow in the same position and is also hypothesized to be compatible with a 4-syllable content word interpretation (mélamondine). (c) Target has a LHLH pattern with an early rise starting at the boundary between the first and second target syllables, a pattern was hypothesized to be compatible with a function word followed by a three-syllable content word (mes lamondines). (d) Target has a LLH pattern with the same early elbow position and is also hypothesized to be compatible with a function word followed by a three-syllable content word (mes lamondines).</i>	206
5.13	<i>Percentage of two-word responses (responses with a word boundary at the critical region) in Perception Experiment 3. Standard error of the mean is indicated.</i>	211
5.14	<i>Comparison of responses with boundaries at the critical 1/2 syllable region according to alignment of the early f_0 elbow, in condition with no early rise.</i>	213
5.15	<i>LHLH pattern with two early elbows, one at the boundary between the conjunction et and the determiner le (e2-el), the other at the boundary between the determiner and the first content word syllable (e-el). The utterance is Le maximum et le minimum seront calculés par Manon. ‘The maximum and the minimum will be calculated by Manon.’ . . .</i>	215
5.16	<i>LHLH pattern in an utterance-medial, critical AP. There is no low leading plateau across the determiner, but a fall from the preceding H to a L right at the function word-content word boundary.</i>	217
6.1	<i>Illustration of a prosodic phrase (et les gamins sages ‘and the good kids’) realized with (a) an obligatory late rise and (b) an optional early rise and an obligatory late rise. A gloss for the entire sentence is ‘And the good kids were gardening’.</i>	225

6.2	<i>Example item set for a word-spotting experiment. Solid lines represent natural f_0 curves, while dotted lines represent resynthesized curves.</i>	235
6.3	<i>Simplified example map for an experiment testing the influence of discourse context on the realization of the early rise.</i>	238

CHAPTER 1

INTRODUCTION

Ye Highlands and ye Lawlands
Oh, where hae ye been?
They hae slain the Earl of Murray
And laid him on the green.
Percy (1765)

1.1 The problem

How do people engaged in a conversation know where one word ends and the next word begins? If we consider the physical signal, speech is clearly continuous—it does not have the convenient spaces that separate words in written language. The waveform and spectrogram in Figure 1.1 have no empty intervals marking the boundaries between words. In fact, even places where we see apparent breaks in the signal are not pauses, but the silences corresponding to the closure in stop consonants like the [p]s in *parole* and *pas* and the [t]s in *acoustique* and *segmentation*.

This lack of clear markers of word boundaries sometimes leads to misperceptions. Several years ago, R.H., a fellow graduate student confessed to me that she had once thought that there was a well-known phonologist called “Princess Smolensky.” Of course, she soon discovered that the name she had been hearing was in fact “Prince and Smolensky.” When asked whether she hadn’t found it a little odd that members

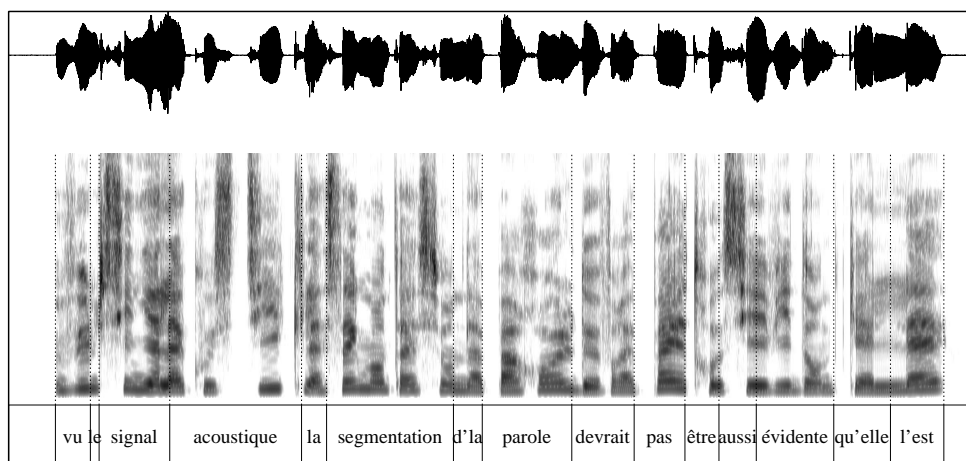


Figure 1.1: *Sample speech signal. The utterance is* Vu le signal acoustique, la segmentation de la parole devrait pas être aussi évidente qu'elle l'est. . . . *'Looking at the acoustic signal, speech segmentation shouldn't be as easy as it is. . . .'*

of royalty would study linguistics, she replied “What did I know?—Trubetskoy was a count!” In solidarity, I confessed that I too had had my difficulties—I was surprised to learn that there was a phonologist named Donca Steriade and that the “Don Casteriade” I had been hearing about existed only in my missegmenting imagination! R.H and I are not alone. A half century ago, the writer Sylvia Wright (Wright (1954)) coined the word *mondegreens* to describe these types of missegmentations:

When I was a child, my mother used to read aloud to me from Percy's *Reliques*, and one of my favorite poems [The Bonny Earl of Murray] began, as I remember:

Ye Highlands and ye Lowlands,
Oh, where hae ye been?
They hae slain the Earl Amurray,
And Lady Mondegreen.

Only years later did Wright realize that there was no Lady Mondegreen, that she had misparsed “laid him on the green,” the last words of the stanza.

Yet such missegmentations are the exception rather than the rule: we somehow experience speech not as an undifferentiated rush of sounds, but as a sequence of words. Listeners manage to segment speech virtually flawlessly. How?

Part of the answer to that question comes from thinking about the one case where the above assertion is obviously false: while we usually have no problem finding the words in our native language, when confronted with a language we don't know, we are often unable to tell where one word ends and another begins. Clearly something about knowing a language helps us segment speech.

Other clues come from examining situations where listeners have problems segmenting their native language. Listeners' errors in perception follow certain patterns, which may give us insight into their segmentation strategies. We can imagine, for example, a couple of reasons why young Sylvia Wright heard "Lady Mondegreen" in the poem. First, *lady* is a real word of English that young children know, and although *Mondegreen* happens not to be a real family name, there's no reason it couldn't be. That is, if the Mondegreen family moved in next door, we'd have no trouble pronouncing their name, which follows all the rules for possible English words. Second, the interpretation makes sense in the context—what a miserable day for the nobles—the earl has been slain, and so has the lady! So, we might suspect (correctly, as it turns out) that factors like discourse context and word knowledge help listeners find word boundaries.

This dissertation examines another factor that helps listeners, in particular French listeners, in speech segmentation—the intonation or distinctive pitch patterns of their language. Related to intonation is prosody, which includes not only intonational

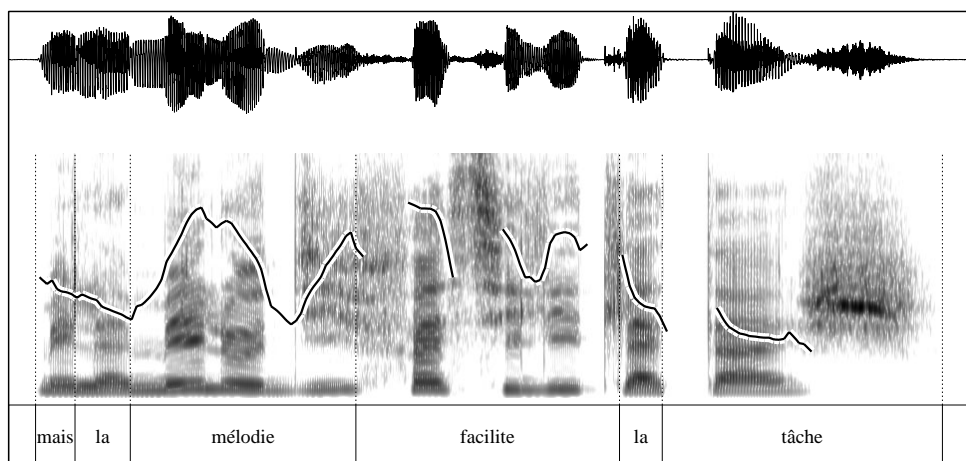


Figure 1.2: *An example utterance with a fundamental frequency curve superimposed on the spectrogram. It is the continuation of the example in Figure 1.1: ...mais la mélodie facilite la tâche! ‘...but the melody makes the task easier!’*

structure, but also the structure of variations in loudness, duration, and rate, and the interactions between these factors.

To examine the role intonation might play, consider the continuation of the earlier example, this time with a fundamental frequency curve superimposed on the spectrogram (Figure 1.2). Fundamental frequency (f_0), the rate of vocal fold vibration, is the primary acoustic correlate of pitch. We can make a few observations about the f_0 curve in the example, which turn out to hold true of French in general. First, there are a number of rises. Second, the position of these rises is not random. Rather the peaks of many rises seem to correspond roughly to the ends of words or phrases. For example, in Figure 1.2, there is an f_0 rise at the end of *mélodie* and one at the end of *facilite*. Other peaks do not occur at the end of words. For example, there is an f_0 rise near the beginning of *mélodie*. The low starting point of this early rise falls right at the boundary between *la* and *mélodie* (there is a vertical line at this boundary).

These early and late rises are intonational characteristics specific to French. A number of researchers have suggested that the early rise helps French listeners in speech segmentation, but this role has rarely been studied experimentally. Establishing the role of the early rise in speech segmentation in French is one of the two major goals of this dissertation.

The other main goal is investigating the structure of the intonational rises and their text-to-tune alignment, that is, the timing of f_0 peaks and valleys with respect to segmental landmarks like consonants, vowels, and syllables. Such an investigation is important for at least two reasons. First, in order to adequately address the role of the early rise in speech segmentation, we must precisely define the phonetic characteristics of the rise. Designing and analyzing a corpus allows us to establish those characteristics and to observe other features potentially relevant to speech segmentation. I take descriptions in the literature as a starting point, but those descriptions are often phonetically vague. A close study of a large corpus of utterances allows us to develop precise, empirically testable hypotheses. It allows us to use f_0 resynthesis to create stimuli for perception experiments that faithfully model the intonation patterns of natural speech.

The study of intonational structure is valuable in its own right, however, beyond its contribution to the design of perception experiments. French intonation has been widely studied for decades, and so, unsurprisingly, there are a number of burning controversies in the field (for example, whether the early rise and the late rise have the same structure). The current production study is designed to address some of these issues. The results will provide evidence for and against competing models, thus contributing to a more accurate, empirically-based description of French intonation.

In addition, although there is a long tradition of research on French intonation, there are no comprehensive studies of text-to-tune alignment (tonal alignment) in French. Although the tonal alignment of other languages has been studied, the subject has been virtually ignored for French. Yet an understanding of the tonal alignment facts is important for a number of reasons. From a theoretical standpoint, evidence from tonal alignment can support one model over another. And in line with my earlier observation, an understanding of tonal alignment is needed to design materials for experiments that test aspects of intonation. Similarly, a more refined account of tonal alignment might improve speech synthesis systems, since inaccurate alignment of rises and falls in the intonation contour makes synthesized speech difficult to understand. And in turn, improvements in speech technology provide scientists with better tools for future speech research.

The topic of this dissertation falls at the intersection of two seemingly very different areas of research—intonational structure and speech segmentation. These subjects are rarely studied by the same researchers—intonation research like my production study is typically conducted by phoneticians and laboratory phonologists, while speech segmentation is typically studied by psycholinguists. And it is certainly not common to find intonational structure and speech segmentation sharing a dissertation. I hope that I have convinced the reader of the intrinsic interest of each subject and of the benefits of considering the two subjects together. Nevertheless, in recognition of the fact that not all readers will be equally interested in both topics, the chapters of the dissertation are designed to be more or less understandable on their own. For example, the reader who is particularly interested in intonational cues

to speech segmentation (but not particularly interested in the structure of French intonation) is invited to skip right to Chapter 5.

The dissertation is organized as follows. Chapter 2 outlines influential models of French intonation and discusses a number of questions in the field. Chapter 3 describes a production experiment designed to answer some of these questions and to inform the construction of materials for the three perception experiments. Participants in the production experiment were recorded reading sentences with embedded target phrases. The critical items differed along a number of dimensions, including speaking rate, position of the target phrase within the utterance, and number of syllables in the target phrase. The range of intonation patterns found, the influence of various factors on the intonation patterns, and the tonal alignment facts are reported. The implications for the structure of French intonation are discussed. Chapter 4 reviews our current knowledge about cues to speech segmentation in general and more specifically to speech segmentation in French. Chapter 5 reports the results of two perception experiments. In one experiment, participants listened to noise-masked minimal pairs or near-minimal pairs of stimuli such as *le niveau de mes sénats* ‘the level of my senates’ and *le niveau de mécénat* ‘the level of patronage’ and indicated what they heard. The experiment examined whether the presence of an early rise on the target syllable (underlined) acted as a cue to the presence of a content word. In a second experiment, the timing of the early rise in pairs of invented nonwords was manipulated. These pairs contained identical sound sequences with two plausible segmentations. Participants listened to the stimuli and indicated what they heard. A sequence like [me.la.mɔ̃.din] was predicted to be perceived as two words (*mes lamondines* ‘my lamondines’) when there was an early rise beginning at the second

syllable [la] and as a single word (*mélamondine*) when there is an early rise beginning at the first syllable [me]. Chapter 6, the conclusion chapter, summarizes the results of each of the experiments, discusses larger implications of the results, and proposes future avenues of research.

CHAPTER 2

MODELS OF FRENCH INTONATION

Of course I am French, why do you think I have this outrrrrageouss accent, you silly king?!

Monty Python and the Holy Grail (Gilliam and Jones, 1975)

2.1 Basics of French Prosody

Perhaps the most striking characteristic of French prosody to speakers of English is that French words seem to be stressed on their last syllables, a perception that is routinely caricatured by English speakers. In the French soldier's famous reply to Galahad in *Monty Python and the Holy Grail*, the word *accent* is (mis)pronounced with stress on the second syllable, *à la française*.

The variety and unpredictability in word stress that is characteristic of many other languages is not found in French. In the other Romance languages and in Germanic languages like English, part of a speaker/listener's tacit knowledge of her native language is knowing which syllable or syllables of a word are stressed. For example, in the word '*syllable*', the first syllable is stressed, while in the word *Ger'manic*, the second syllable is stressed. Melodic prominences or pitch accents can only be associated to these metrically strong syllables (Bolinger (1958)). In languages with stress of this

type, it is possible to find minimal pairs of words that differ only in stress, although these pairs may be rare.¹ Italian *'ancora* ‘anchor’ contrasts with *an'cora* ‘yet’ and English *'differ* contrasts with *de'fer*. No such minimal pairs are possible for French, which does not have contrastive word stress.

2.1.1 Primary accent and the late rise

All models of French intonation include an obligatory primary accent assigned to the final full (non-schwa) syllable of a prosodic phrase. This accenting is marked by syllabic lengthening and increased intensity, as noted by Pasdeloup (1990), Fletcher (1991), and Jun and Fougeron (2000), among others.² If the phrase is not utterance-final, there is also a rise in fundamental frequency (f_0), as schematized in Figure 2.1a, where there is a rise across the last syllable of the phrase *les gamins* SAGES [le.ga.mẽ.saʒ] ‘the good kids’. When the last syllable contains a reduced vowel nucleus, as in some pronunciations of *sage* ([saʒə]), this schwa syllable is not accented.³ The primary accent marks the right edge of a prosodic phrase (Di Cristo (2000), Post (2000), *inter alia*).

¹According to Cutler and van Donselaar (2001), “In all lexical-stress languages, minimal stress pairs are very rare” (p. 175). This may be true for Germanic languages, but in Greek, minimal stress pairs are quite common.

²Wenk and Wioland (1982) give evidence for lengthening of the last full syllable in the phrase, but observe a *decrease* rather than an increase in intensity. This contrary result likely stems from their concentration on utterance-final syllables, rather than syllables that are simply phrase final.

³This word-final *e*, called *e-muet* ‘silent *e*’, represents a historically pronounced vowel. It is still regularly pronounced in some varieties of French, particularly in the South of France. In some words or phrases, the *e-muet* is typically pronounced (e.g., *carte bleue* [kaʁ.tə.blø] ‘credit card’, *calmement* [kal.mə.mã] ‘calmly’), even by speakers who rarely produce it in other contexts. (The French of the Rhône-Alpes region is an exception to this generalization.) These pronunciations avoid difficult-to-pronounce sequences of consonants—note that *carte* is pronounced [kart] and *calme* is pronounced [kalm] (both without word-final schwas).



Figure 2.1: *Illustration of a prosodic phrase (et les gamins sages ‘and the good kids’) realized with (a) an obligatory late rise and (b) an optional early rise and an obligatory late rise. A gloss for the entire sentence is ‘And the good kids were gardening’.*

2.1.2 Secondary accent and the early rise

In addition to this phrase-final late rise, researchers observe that French is characterized by an optional secondary accent, an early rise that occurs near the beginning of the phrase, as in Figure 2.1b. Unlike the late rise, the early rise is not consistently accompanied by syllabic lengthening or increased intensity, although onset consonants may be lengthened or strengthened (see Pasdeloup (1990) and Mertens et al. (2001), *inter alia*).⁴ The role (or roles) of the early rise and its structure are not yet well understood. While it is generally agreed that the presence versus absence of the early rise does not convey a pragmatic contrast comparable to the presence or absence of accent in languages like English, Fonagy (1979) argues that the early and late rise together form an *arc accentuel* that serves to highlight the semantic unity of a phrase. For example, he notes that a fixed collocation like *procès verbal* ‘parking ticket’ is more likely to be produced with an *arc accentuel* (with both an early and a late rise) than a phrase like *procès genant* ‘annoying process’. Recent research has shown that

⁴One of the five speakers in Pasdeloup (1990) did show evidence of slight (but significant) lengthening of syllables identified as having a secondary accent. Pasdeloup suggests that “cet allongement pourrait être volontaire et caractéristique du dialecte méridional parlé par cette locutrice” ‘this lengthening could be voluntary and characteristic of the southern dialect spoken by this speaker (pp. 134, 135).’ In any case, the early rise is not routinely accompanied by lengthening, in contrast to the late rise.

the early rise can help to resolve adjective scope ambiguities in both production and perception (Astésano et al. (2002) and Astésano and Bard (2003)). For example, in *les bouées et les balises vermeilles* ‘the vermillion buoys and beacons’, an early rise on the second noun *balises* is more likely in the narrow scope reading (in which only the beacons (*balises*) are vermillion (*vermeilles*)). Note that together with a late rise, the early rise would form an *arc accentuel* on *balises vermeilles*. Astésano and Bard (2003) argue that the early rise marks the beginning of a minor prosodic phrase. Researchers have also claimed that the early rise helps establish a well-formed rhythm by avoiding long stretches of unaccented syllables (e.g., Pasdeloup (1990), Astésano and Bard (2003)). Some researchers have also suggested that the early rise may aid the listener in speech segmentation by identifying the beginning of a content word (e.g., Vaissière (1997), Di Cristo (2000)) or a morpheme (Delais-Roussarie (1995)). The current evidence suggests that the early rise fulfills more than one role. Its role in speech segmentation is the topic of chapter 5.

Many researchers assume more than one type of early rise. For example, Di Cristo (1999b) argues that there are four types: two *non-emphatiques* ‘non-emphatic’ types, *rythmique* ‘rhythmic’ and *lexicale* ‘lexical’ and two *emphatiques* types *d’intensification* ‘intensifying’ and *de contraste* ‘contrastive’ (he also sometimes refers to emphatic accents as *accents d’insistance* ‘accents of insistence’, following other researchers) (pp. 27, 28). This classification is not clearcut, though, since Di Cristo (1999b) also claims that “l’accent emphatique représente la forme surface marquée (ou Hyper) de l’accent initial ...” ‘the emphatic accent represents the marked (or Hyper[-articulated, according to the Adaptive Dispersion Theory (Lindblom (1990))]) surface form of the initial accent (p. 39). Rossi distinguishes between the *ictus mélodique*

‘melodic accent’ and the *accent énonciatif* or *accent externe* (The *externe* is in opposition to the *interne* of the *accent interne*, Rossi’s term for the primary accent. The primary accent is ‘internal’ in that it is assumed to be a property of the word.) (Rossi (1985), Rossi (1999)). While the number of types of early rise remains an open question, I note that in my experiments, there was no pragmatic context favoring emphasis or contrast.

It is worth noting that the early rise appears to be a relatively recent development in the language—relative to the late rise, that is, since we find evidence of the early rise in the beginning of the 19th century (see discussion in Fonagy (1979)). As Fonagy (1979) and Di Cristo (1999b) both note, earlier scholars considered the early rise a threat to the French language. Di Cristo (1999b) writes:

L’accent emphatique dit d’insistance, qui frappe l’initiale du mot, est référencé depuis longtemps dans les travaux sur le français (Grammont (1914))... Il a été cependant mis à l’écart de la norme prosodique du français, sous prétexte qu’il était l’émanation du français populaire (Fouché (1933–1934)) et que la “prononciation soignée de la langue” se devait par conséquence de l’éviter. Pour d’autres, l’accent d’insistance est incompatible avec le “génie de la langue française” ou nuit à sa beauté (Delattre (1940)). A cet égard on ne peut s’empêcher de citer ici les propos de Grammont (1963), p. 119, suivant lesquels le français, “langue souple et toute en nuances” n’a nullement besoin d’avoir recours à l’accent d’insistance, alors qu’en anglais “on en fait un très grand usage, parce que la banalité des tournures employées d’ordinaire dans la conversation oblige le sujet parlant à se servir de ce procédé lorsqu’il veut que ses paroles ne passent point inaperçues” (sic!)... [En 1956, Fouché] ne pourra que constater, à l’instar d’autres phonéticiens, la prolifération de l’accent d’insistance en français et exprimer son inquiétude de le voir gagner progressivement tous les usages du français (Fouché, 1956: LXII).

p. 163

‘The emphatic accent, called [the *accent*] *d’insistance*, that strikes the beginning of the word, has long been cited in work on French (Grammont (1914))... It was nevertheless set apart from the French prosodic norm, with the claim that it was specific to the French of the masses (Fouché (1933–1934)) and the “refined pronunciation of the language” should therefore avoid it. For

others, the *accent d'insistance* is incompatible with the “genius of the French language” or destroys the beauty of the language (Delattre (1940)). On this subject, we cannot help but cite here the words of Grammont (1963), p. 119, according to which French, a “supple language full of nuances” has absolutely no need to resort to the *accent d'insistance*, while in English “one makes great use of [such an accent], since the banality of the expressions used in ordinary conversation requires the speaker to use this device when he does not want his words to go by undetected” (sic!) [*sic!* (of disbelief) in the original] [In 1956, Fouché] could only note, following other phoneticians, the proliferation of the *accent d'insistance* in French and express his concern in seeing it progressively gain ground in all styles of French (Fouché, 1956: LXII).’

Di Cristo also notes that other scholars have since noted that the *accent d'insistance* has become fully integrated into the language. In connection with this apparent change in acceptability, it is relevant to note that the early rise in current Hexagonal French⁵ is common in even the most formal speaking styles. For example, when Jacques Chirac, the President of the French Republic, gives a speech, his French is characterized by all the markers of formal French: the use of the negative particle *ne*, of the pronoun *nous* ‘we’ in subject position (in everyday French, *on* is used), of the past subjunctive, of inverted questions, and the pronunciation of liaison consonants in optional contexts. But he does not avoid early rises. If anything, formal speaking styles are more dense in early rises than more casual styles.

In addition, we note that naive French listeners and even trained researchers find it difficult to even identify the presence or absence of primary and secondary accents in the language. As Vaissière (1997) remarks, “maints phonéticiens ont noté la difficulté particulière qu’ont les français à percevoir de façon cohérente des proéminences dans leur langue maternelle (a fortiori, dans les autres langues)...” ‘many a phonetician has noted the particular difficulty that the French have in perceiving in a

⁵The French spoken in France, nicknamed the *Hexagone* for its shape.

coherent way the prominences in their mother tongue (much less in other languages)” (p. 56). Fonagy (1979), for example, calculated that the French subjects in his study performed ten times more poorly than Italian and Hungarian subjects in determining the presence of primary and secondary accents in their native language. The naive subjects in Pasdeloup (1990) identified secondary accents only 56% of the time. Pasdeloup herself reports difficulty in identifying the secondary accents in her corpus: she was able to unambiguously code only 73% of the syllables in her corpus.

2.1.3 Articulatory correlates

A few studies have examined the articulatory correlates of prosodic phrasing or of the primary and secondary accent in French. These studies have all found some evidence for such correlates along various articulatory measures, but have also typically reported inter-speaker variability. For example, Fletcher and Vatikiotis-Bateson (1991) found longer opening gestures and bigger lip apertures for syllables with the primary (“tonic”) accent for each of their two speakers, although these effects are not found for the same syllable for each speaker. They also report a potential articulatory effect for the secondary accent in the speech of one speaker, who consistently produced the secondary (“initial”) accent. Fougeron (2001a) reports that the articulatory characteristics of prosodic-phrase initial segments reflect the level of the prosodic boundary—segments are strengthened (more linguo-palatal contact, glottalization of vowels) at the beginning of higher level phrases. This study explicitly excluded phrases with early rises.

Løevenbruck (2000) compared the articulation of syllables realized with an early rise and a late rise in both a “natural” condition and a condition with contrastive

emphasis (contrastive focus), measuring tongue and jaw movements. She found preliminary evidence that both the early rise and the late rise may be accompanied by hyperarticulation. In the natural condition, the late rise was marked by a hyperarticulation stronger than that of the early rise. This pattern was reversed in a contrastive condition, where the hyperarticulation of the early rise became as strong as or even stronger than the late rise. (See also Løevenbruck (1999). Comparable results were found in Tabain (2003).) We do not yet know, however, how the observed articulatory effects interact with the presence or absence of an early rise, since no study has yet systematically studied stimuli with and without the early rise.

Dohen et al. (2003b) further investigated the articulatory correlates of contrastive focus. Dohen et al. (2003a) present evidence that these articulatory/visual correlates are salient to naive listeners and can be used to disambiguate utterances that differ only in focus structure. They argue that prosody is “multimodal and multigestural” (p 245).

French accents are discussed in the literature using a variety of terms, including “accent primaire,” “continuation mineure,” “tonic accent,” and “primary stress” for the late rise, and “accent secondaire,” “accent probabilitaire,” “secondary stress,” and “accent d’insistance” for the early rise. I will use the term “late rise” for the obligatory late peak and “early rise” for the optional earlier peak. Note that even in cases where the obligatory rise is the only rise in the phrase, I will refer to it as the “late rise.” When necessary for the sake of clarity, I will use terms specific to certain researchers in discussing their models. I use SMALL CAPS to indicated the presence of primary and secondary accents.

2.1.4 Stress in French: Property of the phrase or the word?

There is no general agreement about whether accent or stress in French is a property of the phrase or of the word.⁶ The claim that stress in French is fixed at the word level is common in current and historical descriptions of French intonation. For example, Post (2000) writes that “all [French] words have final stress (‘fixed stress’ Garde (1968))” (p. 1). Similarly, Dupoux et al. (2001) claim that “...word stress is final in French” (p. 1617). In some accounts, (e.g., Garde (1968)), certain classes of words (mostly content words) are *accentogènes*, that is, capable of generating or giving rise to accents. These accounts all have some mechanism for ensuring that “stress” is realized on the last syllable of a prosodic phrase. According to Peperkamp and Dupoux (2002), “First, in French, stress falls on the word’s final vowel (Schane 1968). Function words, which are typically unstressed elements, attract stress if they are phrase-final” (p. 7). Hirst and Di Cristo (1998) propose the following set of principles:

Following Garde (1968), we can assume that stress placement in French is based on three elementary principles. A principle of *autogénéité* (stressability) selects lexical items (generally content words) that are stressable; a grouping principle specifies that a stress group is constituted by a stressable word and by adjacent pro/en-clitics governed by it; a right-heading principle assigns stress to the last full syllable of the construction so formed. The last two principles explain why clitics, which are not usually stressable, can occasionally be stressed in French, as in [(2)].

- (2) Dis LE | à ton jeune FILS
Tell it to your young son
‘Tell it to your young son.’
-

⁶I hesitate between the terms *stress* and *accent*. The term *stress* is typically used to describe contrasts between more prominent and less prominent syllables. As we will see, French listeners have difficulty identifying differences in prominence across syllables. The term *accent*, however, might be interpreted to mean *pitch accent* in the Autosegmental Metrical understanding of the term (see §2.2.3), which may not be an appropriate characterization of the “secondary accent.”

In a later paper, Di Cristo (2000) argues for a *bipolarisation prosodique* ‘prosodic bipolarization’ that follows from a principle of *promotion des extrémités* ‘promotion of the extremities’. According to this view:

...le mot, qui constitue l’unité accentuelle virtuelle minimale en français, est doté dans la représentation sous-jacente d’une proéminence initiale et d’une proéminence finale.

‘...the word, which constitutes the virtually minimal accentual unit in French, is endowed in the underlying representation with an initial prominence and a final prominence.’

Di Cristo (2000), p. 36

Another principle, the *principe de dominance finale* ‘principle of final domination’ accounts for the fact that the late accent rather than the early accent is the primary accent. Accent is clearly treated as a property of the word, as illustrated in (3), taken from Di Cristo (2000). The original caption is preserved, and an English translation is added.

				x
(3)	x			x
	x	x	x	x
	Fé	li	ci	té

Représentation de la forme prosodique sous-jacente du mot “Félicité” par une grille métrique simple. ‘Simple metrical grid representation of the underlying prosodic structure of the word Félicité.’

Other accounts treat stress as a property of the prosodic phrase (e.g., Grammont (1963), Martin (1979), Jun and Fougeron (2002)). Jun and Fougeron (2002) note that this sets French apart from its sister languages: “...in contrast with most other Romance languages, the domain of stress in French has changed over the course of its evolution from Latin from a lexical to a phrasal domain” (p. 1). Martin (1979) explicitly rejects the idea that certain word classes are *accentogènes*, that is, capable

of generating or giving rise to accents. Martin notes the contrast in pairs like those in (4), where the function word *le* ‘it’ is either accented or unaccented. As others have noted, the realization of the late rise on phrase-final enclitic pronouns is common and unexceptional (see, for example, Mertens (1993) and Delais-Roussarie (1999)).⁷ For Martin, that the pronoun in (4b) is accented is not exceptional; it simply follows from a set of grouping rules.

- (4) a. Tu le PRENDs
 you it take
 ‘Take it.’
 b. Prends LE
 take it
 ‘Take it.’

In the case of phrase-final function words (as in (4b)), schwa syllables can be accented—they are lengthened and the peak of the late rise is realized on them. However, the same is not true for a phrase-final content word—in that case, the last *full* syllable will be accented, not a final schwa syllable. Post (2002) and other researchers argue that this difference is evidence of the existence of lexical stress in French:

The coincidence of stressed syllables and word-group boundaries has led some authors to deny the existence of word stress in French (e.g. Coustenable and Armstrong (1934):4; Pulgram (1965):132-133; see Di Cristo (1999) for a discussion). Nevertheless, metrical prominence appears to be relevant to the assignment of pitch movements, as is argued by Dell (1984) on the basis of the different alignment of intonation contours with utterances that end in a full syllable and those in which the final syllable was a schwa.

p. 8

⁷Hirst and Di Cristo (1998) writes that clitics “can occasionally be stressed in French” (p. 196), but imperative constructions of that type found in (2) and (4b) are encountered regularly.

Another possible explanation for this difference, though, is a requirement that the primary accent be realized on the last word of the phrase. If a full syllable is available (as in the case of a schwa-final content word like *sage* in Figure 2.1), the accent will be realized on that syllable. If not (as in the case of monosyllabic function words like *le* in (4b)), it must be realized on a schwa syllable.

It is not clear, then, that there is any advantage to accounts that treat stress or primary accent as a word-level property, since these accounts must then separately provide for the phrase-final realization of accent and the apparently exceptional accentuation of phrase-final function words. To be sure, a French word uttered in isolation will be accented on its last full syllable, but this follows from the fact that any word uttered in isolation is in itself a complete prosodic phrase. With no clear advantage to accounts that treat stress or primary accent as a property of the word, I prefer a simpler account that treats primary accent as a property of the prosodic phrase and thus provides a unified treatment of the accenting of phrase-final syllables.

2.2 Background concepts in intonation

Before discussing specific models of French intonation, it is useful to first outline some basic concepts and different types of approaches that come into play in the various accounts.

2.2.1 Superpositional approaches

In a superpositional or overlay model of intonation, component parts of the fundamental frequency (f_0) curve are superimposed on each other. The essence of this approach is succinctly captured by Bolinger (1972), who likens the f_0 variations of

speech to the movements of the ocean, distinguishing between ripples, waves, swells, and tides:

It would be more accurate to say ripples *on* waves *on* swells *on* tides, because each larger movement carries the smaller ones on its back... In speech... the ripples are the accidental changes in pitch, the irrelevant quavers. The waves are the peaks and the valleys that we call *accent*. The swells are the separations of our discourse into its larger segments. The tides are the tides of emotion.

[Bolinger (1972), p. 19, as quoted in Ladd (1996), pp. 23, 24]

Ladd (1996) recognizes the aptness of the metaphor with respect to microprosodic perturbations (the “ripples”) and non-linguistic effects on f_0 variation (the “tides”), but notes that most work in the superpositional framework has emphasized the “wave”/“swell” distinction. That is, researchers working in this class of models assume a distinction between “accentuation,” which concerns pitch movements on certain syllables, and “intonation,” which concerns global pitch movements across larger domains. In these types of approaches, accentuation phenomena include lexically specified tones in languages such as Mandarin and Thai, as well as accents associated with metrically prominent syllables in languages like English and Dutch. Intonation phenomena in these approaches include boundary pitch shapes associated with certain sentence modalities as well as long range trends such as declination. The intonation modules of a number of speech synthesis systems are based on superpositional approaches (see for example, Fujisaki and Hirose (1984) for Japanese; and Aubergé (1991), Aubergé (1992), Morlec (1997), Morlec et al. (2001), and Holm (2003), *inter alia* for French).

2.2.2 Functional approaches

Some models of intonation (e.g., the classic model of Delattre (1966)) take a functional approach in describing French intonation. The melody of an utterance has a clear influence on its linguistic function and researchers have attempted to map given f_0 shapes onto different linguistic functions. For example, a fall may indicate a statement, a rise to a mid-level may indicate a continuation, and a higher rise may indicate a question. The attraction of a unified treatment of form and function in intonation probably has much to do with the observations of cross-linguistic similarities in intonation, discussed for example, by Lieberman (1980), Ohala (1984), and Vaissière (1995).

In a recent review of the advantages and disadvantages of different types of transcription systems, Vaissière (2002) questions systems that do not encode functional meanings:

A phonological transcription should avoid using one and the same symbol (for example, H%), as these types of rises [rises used to convey questions, continuations, or to “hold the floor”], which may sometimes correspond to the same f_0 contours, are perceptually distinguished (Fonagy and Bérard (1973)).

The underlying assumption here, that there is a one-to-one correspondance between form and function, is questionable. In other areas of linguistics, it is widely accepted that form and function are *distinct*. For example, there is nothing about the strings of segments in *cat* [kæt] that corresponds to cat-ness. Shuffling the phonemes to produce *tack* [tæk], produces something that is not at all soft and fluffy like a cat. Moreover, it is widely agreed that ambiguity (lexical, syntactic, semantic) is the rule rather than the exception in natural languages and that almost all utterances

are multiply ambiguous. That ambiguity in intonation should exist seems therefore unremarkable, unproblematic, and expected.

2.2.3 Autosegmental-Metrical approaches

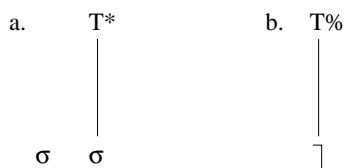


Figure 2.2: Association of (a) a pitch accent to a syllable, and (b) a boundary tone to the edge of a prosodic unit. T stands for ‘tone’.

In the Autosegmental-Metrical (AM) framework (see, for example, Pierrehumbert (1980), Beckman and Pierrehumbert (1986), Pierrehumbert and Beckman (1988)), there are three categories of tones. Pitch accents (marked with a ‘*’) are associated to metrically strong syllables or moras. Boundary tones (marked with a ‘%’) are realized at the boundaries of prosodic units. Phrase accents (marked with a ‘-’) are similar to boundary tones, but tend often have a secondary association to designated syllables (cf. Grice et al. (2000)). The descriptions of many languages use two tones, high (H) and low (L), while some require a mid (M) tone (e.g., Cantonese (see Chao (1947), Hashimoto-Yue (1972), Yip (1980), Yip (1993), Yip (1995), and Wong et al. (1999)) and Yoruba (see Connell and Ladd (1990) and Laniran and Clements (2003))). The association of a pitch accent to a syllable and the association of a boundary tone to a prosodic boundary are schematized in Figure 2.2. In the AM framework, only

certain landmarks are specified for tone, and the pitch of non-specified stretches is determined by interpolation.

2.2.4 Tonal Association and Alignment

Two concepts central to Autosegmental-Metrical approaches are *association* and *alignment*. Text-to-tune *alignment* is the timing of important parts of the f_0 curve (the tune) of an utterance relative to important parts of the segmental string (the text). Alignment is related to, though distinct from, the *association* of tones to prosodic units or edges. In Autosegmental Metrical phonology, tones may be associated to syllables or moras (pitch accents) or to prosodic edges (boundary tones, phrase accents).

One of the fundamental goals of much work on alignment has been to characterize the intonational structure of particular languages and of spoken languages in general. As Pierrehumbert and Beckman (1988) put it:

The task of phonetic realization rules is to describe how speakers pronounce phonological entities... The realization process does not make arbitrary use of the information in the phonological representation; instead, it uses the information in particular ways, which give rise to generalizations supporting one sort of representation over others.

p. 160

One of the goals of the production experiment described in Chapter 3 is to use a detailed investigation of tonal alignment to evaluate intonational structures proposed by different models of French intonation. Differences in text-to-tune alignment can signal lexical differences (in Swedish, for example) or pragmatic differences (in Italian and English, for example). In Chapter 5, we investigate whether cues offered by differences in tonal alignment in French can aid listeners in speech segmentation.

A schematized example of a minimal pair contrasting in alignment is given in Figure 2.3. The capital letters represent segments (consonants and vowels). Dots represent syllable boundaries. The lines are stylized f_0 curves. In each of the panels, the segmental material and the shape of the f_0 curve is identical. What differs is the relationship of the f_0 curve to the segmental material. In (a) panel, there is an f_0 inflection point or elbow located near the end of the “syllable” *U* and an f_0 peak toward the end of the following “syllable” *WX*. In (b), the low elbow is reached within the *WX* syllable, while the f_0 peak is realized after the offset of the syllable.

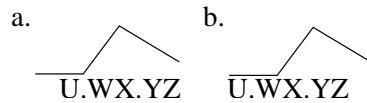


Figure 2.3: *Schematicization of same f_0 curve with different text-to-tune alignments. Dots represent syllable boundaries. (a) Alignment consistent with a $L+H^*$ pitch accent associated with the syllable *WX*, (b) Alignment consistent with a L^*+H pitch accent associated with the syllable *WX*.*

This difference in alignment might reflect a categorical distinction in the phonology of the language. This contrast would be described in many Autosegmental-Metrical accounts as a contrast between a L^*+H accent (Figure 2.3b), where the L tone is associated to the accented syllable (in this case *WX*), and a $L+H^*$ accent (Figure 2.3a), where it is the H tone is associated to the accented syllable. In many languages, a tone is realized within the syllable to which it is associated, but this is not always the case. Such variation is not unexpected if we assume that basic intonational units can vary in the same way basic units of language vary. For example, both French

and English have a series of phonemically voiced stops /b,d,g/ that are distinguished from their phonemically voiceless counterparts /p,t,k/, but there are systematic voice onset time (VOT) differences between the two languages in the implementation of these contrasts (see Caramazza and Yeni-Komshian (1974)).

2.3 Models of French Intonation

Having reviewed some relevant concepts, we will now outline several models of the intonation of Hexagonal French. There is a vast body of research on the topic, and we cannot give an exhaustive account of all models of French intonation. For a comprehensive review, the reader is referred to Lacheret-Dujour and Beaugendre (1999), which includes discussions of various proposals, including those discussed below and those of Verluyten, Dell, Padeloup, and Delais-Roussarie (see also Verluyten (1982), Dell (1984), Padeloup (1990), and Delais-Roussarie (1995)). In this section, we will focus on models that make precise and therefore empirically testable predictions about the two rises, particularly with respect to text-to-tune alignment. This focus is motivated by the two primary goals of the dissertation, first, to examine French intonational structure through the window of text-to-tune alignment and second, to examine whether the presence of an early rise and its alignment act as cues to speech segmentation.

2.3.1 Delattre

According to the classic description of Delattre (1966), French intonation patterns can be decomposed into ten basic intonation contours (*dix intonations de base*). The shape of these contours is defined with respect to four distinct pitch levels, with four rising contours, four falling contours, and two level contours. These contours are

illustrated in Figure 2.4. Each contour is associated with a pragmatic function, for example, question or exclamation for utterance-final contours, and major and minor continuation for utterance-medial ones.

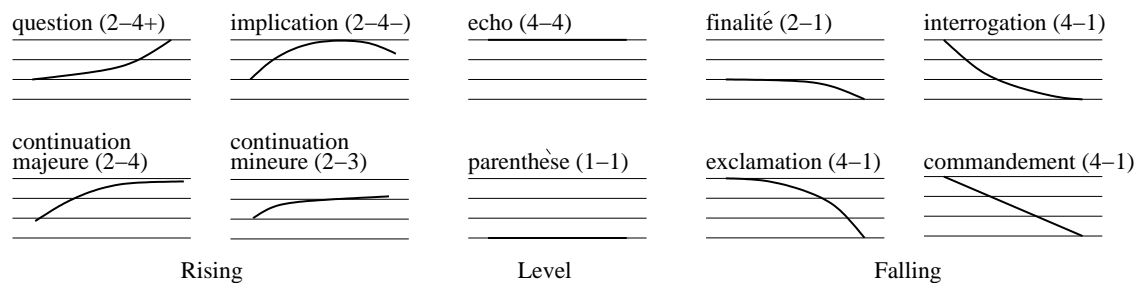


Figure 2.4: *Delattre's "dix intonations de base" 'ten basic intonation contours'.*

The Delattre model fails to capture important aspects of intonational structure. The model, like models that rely on holistic contours in general, makes no prediction about text-to-tune alignment. The model is therefore unable to capture distinctions based on tonal alignment (see §2.2.4). In addition, the model does not account for pitch movements that are not phrase final, including the early rise that is a major focus of this dissertation. In Bolinger's terms, the account describes the phrase-final swells of the late rise, but not the waves of the early rise (see §2.2.1).

In addition, some of the claimed distinctions have been called into question. As shown in the figure, both questions and implications are defined by rising movements from level 2 to level 4. They are distinguished only by the shape of the curve—convex for questions and concave for implications. The results of Di Cristo (1976) and Rossi (1978), however, show that this difference is not reliable.

Of course, Delattre worked at a time before technological advances made examining phenomena such as text-to-tune alignment and contour shape fairly straightforward. Early observations like his, although ultimately not completely accurate, have fed the work of more recent researchers.

2.3.2 Rossi

Like Delattre, Rossi argues for the existence of a certain number of intonational morphemes or *intonèmes* (Rossi (1981), Rossi (1985), Rossi (1999)). These intonemes are divided into three main classes, which are distinguished by their function and their form. The three continuatives (*continuatif majeur* (/CT/ or /CM/), *continuatif mineur* (/ct/ or /cm/), and *continuatif majeur appellatif* (/CT+/ or /CA/)) are characterized by f_0 rises. The two conclusives, *conclusif majeur* (/CC/) and *conclusif mineur* (/cc/), are characterized by f_0 falls. Parentheticals (/PAR/) are characterized by flat f_0 . This intonational morpheme accounts for f_0 plateaux in dislocated constituents. The model includes six intonational levels: super-sharp (*suraigu*), sharp (*aigu*), sub-sharp (*infra-aigu*), médium (*medium*), grave (*grave*), and infra-grave (*sub-grave*). Figure 2.5 gives an example of the relationship between a subset of the intonational morphemes and the intonational levels.

In the model, intonation is tightly linked to syntax and pragmatics. For example, intonation signals syntactic hierarchization and certain combinations of syntactic structures and intonational patterns are ruled out. A continuative morpheme (CT) is ruled out after *le frère* in (5) by a rule that prevents the constituents of the subject noun phrase *le frère de Brutus* from being separated by a boundary of equal or greater strength than the boundary at the end of the noun phrase.

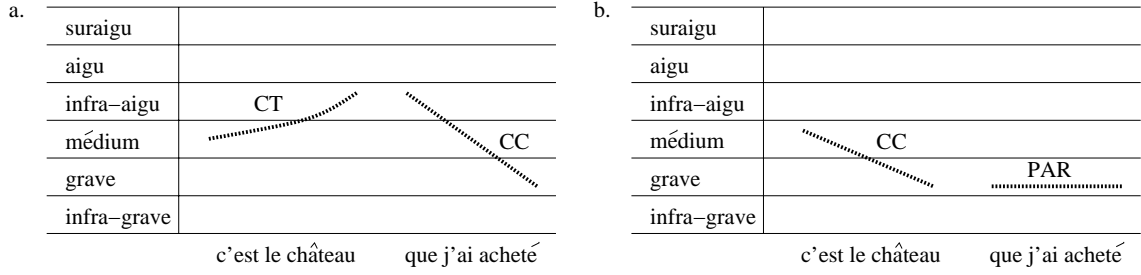


Figure 2.5: *Example of Rossi's intonational morphemes applied to a short sentence. Glosses for the two pronunciations are (a) 'It's (that's) the castle that I bought.' (b) 'It's (that's) the castle (the one that I bought).'* The example is adapted from Rossi (1985).

- (5) * Le frère /CT/ de Brutus /ct/ n'a pas tué César /CC/
 The brother of Brutus not has not killed Caesar
 'Brutus's brother did not kill Caesar.'

In many cases, the continuative intonemes (or more precisely, their various actualizations) correspond to the late rise of the primary accent described by other researchers.

As discussed in §2.1.2, Rossi assumes two types of accents corresponding to the early rise, the *ictus mélodique* and the *accent énonciatif* or *accent externe* (AE). The AE is distinct from the *ictus mélodique* in that the “AE est réalisé par un relief de la mélodie et de l'intensité qui dépasse celui qui caractérise l'ictus mélodique” ‘AE is realized by an excursion in melody and intensity that surpasses that of the melodic accent’ (Rossi (1985), p. 147). According to Rossi (1985), the AE can be used as a type of focus accent (Rossi (1999) discusses the AF or *accent de focalisation*), but is also the early rise found in the speech of newsreaders (footnote, pp. 147, 148). The *ictus mélodique* (IC) is realized “par des contraintes purement rythmiques” ‘by

purely rhythmic constraints'. These constraints avoid long sequences of syllables with no intonational morphemes.

The realization of the IC proceeds according to an algorithm detailed in Rossi (1999). The algorithm operates on long sequences of unaccented syllables, assigning H and L tones (*(H)aut* 'high' and *B(as)* 'low' in the original) to syllables. A series of grouping rules then operates to simplify the string of Hs and Ls. Finally the number of Hs for each syllable is calculated, and the IC is realized on the syllable with the highest score.

It is not clear why the mechanism of tone assignment and simplification is used in calculating the position and strength of the *ictus mélodique*. In any case, it is an open question whether the early rise or secondary accent is in fact associated with a single syllable. Nevertheless, a strength of Rossi's approach is that his model explicitly addresses the secondary accent, and raises interesting questions about how many categories of secondary accent/early rise the language contains.

Like the Delattre model, the Rossi model does not make predictions how intonational morphemes are aligned with the segment string.

In addition, as Lacheret-Dujour and Beaugendre (1999) note, the generalizability of the model to more complex sentences has not been demonstrated. In this regard, it is worthwhile to note that Jun (1998) has shown for Korean that while the syntactic approach to deriving prosodic structure makes many correct predictions for citation form utterances, it does not correctly predict the range of prosodic phrasing found in spontaneous utterances. She also notes that Croft (1995) has made similar observations about the claimed syntax-intonation mapping in English. For French, Post (2000) found that the predictions of Prosodic Phonology theory (Nespor and

Vogel (1986)) with respect to the restructuring of syntactically derived Phonological Phrases did not match her corpus data.

2.3.3 Martin

Martin (1979) proposes that the *accent tonique* ‘tonic accent’ or primary accent is a property of the minimal prosodic unit of an utterance, the prosodic word, and serves to demarcate it. Prosodic words are then organized in a hierarchical prosodic structure, which is in most cases isomorphic to syntactic structure. The tonic accent appears on the “last pronounced [i.e., non-schwa] syllable” of the prosodic word. This model holds that syntactic and semantic relationships among words determine the accentability of a phrase and rejects the traditional notion (see, e.g., Garde (1968)) that certain classes of words are *accentogènes*, that is, capable of generating or giving rise to accents.

Four quasi-semantic/syntactic principles then serve to group together words, and predict differences in accentuation. For example, *un baTEAU* ‘a boat’ is predicted to differ in accentuation from *UN baTEAU* ‘one boat’.

In later work, Martin (1987, 1998, 1999) describes the accents realized on prosodic words as prosodic markers. These markers are holistic contours defined by three binary features: low (low/high), rising (rising/falling), and ample (large/small). Figure 2.6 illustrates some of the contours defined by these features.

Noting that it has been demonstrated that rhythmic factors play a role in the distribution of accents (e.g., Wenk and Wioland (1982)), Martin (1987) incorporates the principle of *eurhythmicity* into his model. This principle states that “structures that balance the number of syllables of the prosodic words AT EVERY LEVEL OF THE

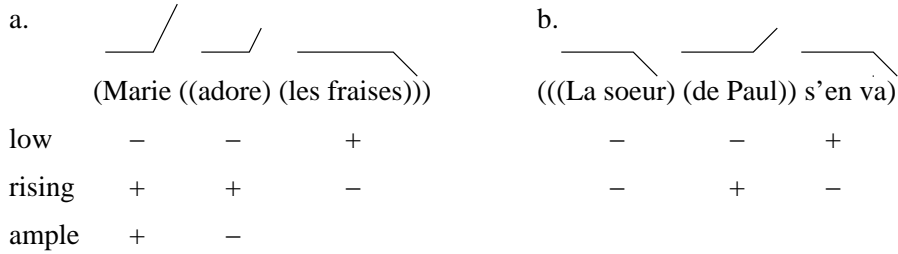


Figure 2.6: *Examples of Martin's (1987) "prosodic markers" and their feature specifications. (a) Marie adore les fraises. 'Marie loves strawberries.' (b) La sœur de Paul s'en va. 'Paul's sister is leaving.'*

STRUCTURE are favored" (p. 934, emphasis in the original). The principle predicts that an utterance like the one in Figure 2.6b would have different prosodic phrasing if the two-syllable subject *la sœur* 'the sister' were replaced by the four-syllable *l'arrière grand-père* 'the great-grandfather'.

One strength of Martin's models is that they emphasize that the domain of the accent is a unit larger than the word. Martin, however, does not discuss the early rise ("l'accent d'insistance ou d'emphase" in his terminology), though, except to note that a tonic accent can be displaced to the left in cases of accent clash. The lack of a discussion about the role of the early rise is particularly striking in the 1987 paper since that paper focuses on the apparent pressure to produce balanced prosodic units, and that is one of the claimed functions of the early rise.

Like many models that assume holistic contours as their basic units, Martin's models do not specify what the text-to-tune alignment should be. The only description we are given is that "l'accent tonique se plac[e] sur la dernière syllabe prononcée" 'the tonic accent is placed on the last pronounced syllable' (Martin (1987), p. 5). It is probably reasonable to assume that the end of the contour is assumed to coincide with

the last syllable, but beyond that, the model makes no predictions about text-to-tune alignment. Yet we know the details of text-to-tune alignment signal important distinctions in many languages (e.g., Swedish, Italian, English). We would expect that an adequate description of the intonational structure of those languages and that of French would make reference to the same *kinds* of structures (although, of course, not necessarily to the same structures). Pierrehumbert (2000) points out that the prosodic and intonational systems of the world’s spoken languages exploit the same inventory of parameters (relative duration, pitch range, tonal alignment, etc.), just as the vowel systems of different languages do (backness, rounding, etc.). She argues that we should expect to find cross-linguistic similarities in intonational structure:

For example, although the French high front unrounded vowel is broadly analogous to the English one, its exact degree of height, fronting, and spreading is different. . . Similarly, we do not expect to see tonal elements literally equated across languages. Instead, the expectation is that broad patterns of contrast and tonal realization might be echoed from one system to the next.

p. 26

We might expect, for example, that text-to-tune alignment is an important dimension in all spoken languages, although it plays different roles in different languages. For example, text-to-tune alignment signals semantic and pragmatic distinctions in some languages and word boundaries in others. Some languages may not “use” text-to-tune alignment to signal contrasts, but we still expect that there will be language-specific alignment patterns that are part of the competence of native speakers. To continue the comparison with vowel systems, consider the articulation of vowels in a language with an uncrowded vowel space, as in the five-vowel system of Spanish. The high front unrounded vowel of Spanish has a distinct articulation, even though

it is not articulatorily or acoustically close to any other vowel. A Spanish [i] and an English [i], for example, are recognizably different.

In a recent paper, Martin (2003) outlines a number of perceived problems with the ToBI (Tones and Break Indices) intonational transcription system (Beckman and Ayers (1993), Beckman and Hirschberg (1994)) in particular and Autosegmental Metrical framework in general. His discussion reveals a number of fundamental misunderstandings that undoubtably influence his choice of holistic contours and his rejection of the tonal targets approach. For example, he notes that all the f_0 rises in Figure 2.7 are likely to be transcribed as L H* sequences, regardless of the slope, excursion size or shape (convex vs. concave) of the rise:

The reductionist character of the notation by high and low targets and their variants (downstepping) arises from there. Consider, for example, the melodic variants [in Figure 2.7]. Since the time parameter has been abandoned in the ToBI notation, each of the [5] f_0 movements is *a priori* subject to the same notation, L H*

pp. 110, 111

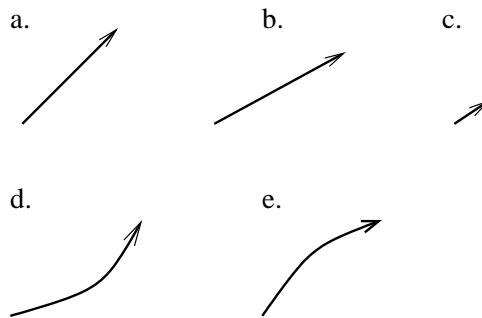


Figure 2.7: *Five different f_0 shapes (redrawing of Martin (2003)). Martin argues that a transcription system should capture the differences in shape.*

Yet one of the strengths of the Autosegmental Metrical framework, and indeed, any adequate system of phonological transcription, is that it does not distinguish between variants that are not contrastive along some dimension. For example, if the curves in Figure 2.7a and Figure 2.7c, which have different excursion sizes, do not contrast within a given language, they should not have different transcriptions, just as a phonemic transcription of a language should not use separate symbols for two instances of /b/ with quantitatively different VOTs. Moreover, Autosegmental-Metrical approaches hardly abandon the time parameter. In fact, AM approaches, unlike holistic approaches, provide an explicit mechanism for capturing phonological distinctions signalled by differences in temporal alignment.

2.3.4 Di Cristo and Hirst

Di Cristo and Hirst’s description of French intonation assumes three basic units: the Tonal Unit (TU), the Rhythmic Unit (RU), and the Intonation Unit (IU) (Hirst and Di Cristo (1996), Di Cristo and Hirst (1996), Di Cristo (1998), Di Cristo (1999b), Di Cristo (2000), Di Cristo et al. (2000)).⁸ The boundaries of the Tonal Units are determined by the three post-lexical grouping rules or principles given in (6).

- (6) a. INITIAL PROMINENCE (optional): Form a stress group up to and including the initial syllable of each lexical item.
 - b. FINAL PROMINENCE (obligatory): Form a stress group up to and including the final syllable of each lexical item.
 - c. READJUSTMENT When a monosyllabic stress group is both preceded and followed by a stressed syllable in the same intonation unit [IU], combine the last two groups into one.
- from Hirst and Di Cristo (1998)

⁸ The model of French intonation promoted by DiCristo and Hirst and colleagues has been under development for two decades. Different aspects of the model are discussed in different papers and the authors often refer the reader to more complete descriptions in earlier papers. I therefore cite papers ranging from their earliest to their latest work. I have taken care not to include aspects of earlier models that have since been superseded.

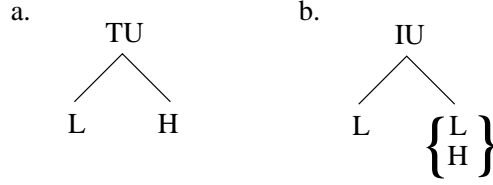


Figure 2.8: *Tonal templates (a) for the Tonal Unit (TU) and (b) for the Intonation Unit (IU) (from Di Cristo (2000)).*

Each of these stress groupings is coextensive with a melodic unit, the Tonal Unit (TU). A Tonal Unit is delimited by a L tone to the left and an H tone to the right (the tonal structure is therefore [LH]). This tonal template is shown in Figure 2.8a. In this model, both the early rise and the late rise are considered to be Tonal Units. A detailed example with Tonal Unit, Rhythmic Unit, and Intonation Unit boundaries marked are shown in Figure 2.9.

The model includes a unit called the Rhythmic Unit (RU) to account for the observation that the primary accent, but not the secondary accent, is accompanied by syllable lengthening. The Rhythmic Unit is not tonally defined, but specifies differences in metrical strength. As Figure 2.9 shows, an RU boundary occurs at the end of a word or group of words. Syllables that are RU-final (like the last syllables of *voisin* and *disputés* in Figure 2.9) are lengthened. The non-melodic character of the Rhythmic Unit is explicitly stated:

Par la suite (Di Cristo & Hirst, 1992)⁹, afin de rendre compte de la base rythmique du français et de la distinction entre l’accent primaire (final) et l’accent secondaire (initial), nous avons proposé d’introduire dans le modèle un constituant prosodique de rang intermédiaire: l’Unité Rythmique (UR) dont il ne sera pas question ici dans la mesure [où] nous traitons uniquement de l’organisation intonative (ou mélodique) du français.

⁹The intended reference here is apparently Di Cristo and Hirst (1993) .

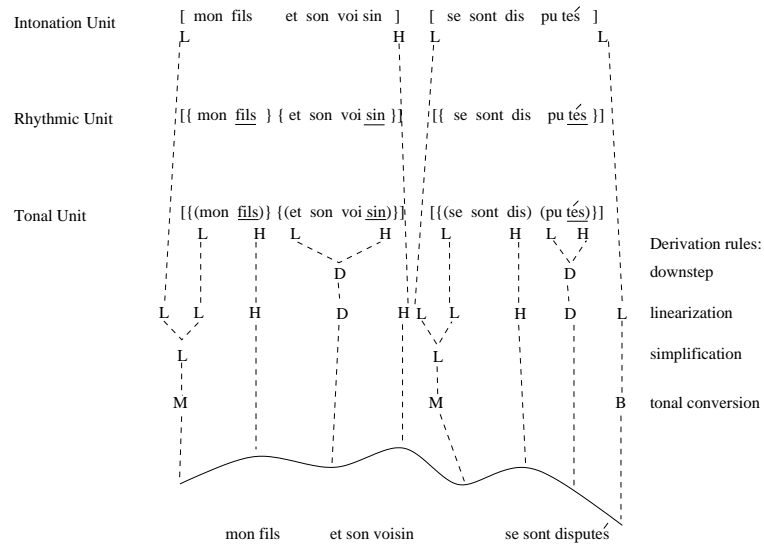


Figure 2.9: A detailed example of Di Cristo and Hirst's view of prosodic structure. The figure is chiefly a synthesis of Figures 1 and 4 of Di Cristo and Hirst (1993). The tonal conversion level, however, reflects additions to the model discussed in Di Cristo and Hirst (1996). The stylized f_0 curve reflects my understanding of Di Cristo and Hirst's view of tonal scaling, as discussed in Di Cristo and Hirst (1996). A gloss is 'My son and his neighbor argued with each other.'

‘Eventually (Di Cristo & Hirst, 1992), in order to account for the rhythmic basis of French and the distinction between the primary accent (final) and the secondary accent, we proposed introducing into the model an intermediate prosodic constituent, the Rhythmic Unit (RU), which will not be discussed here since we are treating only the intonational (or melodic) organization of French.’

Di Cristo and Hirst (1996) pp. 219, 220

The third and highest unit in Di Cristo and Hirst’s prosodic hierarchy is the Intonation Unit (IU). As Hirst and Di Cristo (1998) note, the grouping rules that define the stress groups or tonal units “must be applied once the focal (rhematic) part of the utterance has been selected and boundaries of Intonation Units affected” (p. 199). The mechanism by which the boundaries of the Intonation Unit are defined is not entirely clear, but according to Di Cristo (2000):

... la métricit  des Unit s Intonatives (UI) est motiv e en premier lieu par des contraintes s mantico-pragmatiques, ce qui n’exclut pas pour autant l’influence de la syntaxe qui est   m me d’exercer ses propres contraintes   un niveau moins profond.

... the *m tricit * of the Intonation Units (IU) is motivated first by semantico-pragmatic constraints, which does not as such exclude the influence of syntax, which also exercises its own constraints at a shallower level.”

pp. 36, 37

For example, Di Cristo and Hirst (1993) (citing Di Cristo (1978) and Hirst and Di Cristo (1984)) maintain that there is generally an Intonation Unit boundary after the noun phrase subject of a sentence. An Intonation Unit has two possible tonal patterns: [LH] and [LL]. The choice of pattern is pragmatically defined: the [LL] pattern corresponds to an assertion, while the [LH] pattern corresponds to a continuation (which can be utterance medial or utterance final) or a question. The tonal template for the IU is shown in Figure 2.8b.

Figure 2.9 illustrates grouping at the three proposed levels, tonal assignment at the two tonally defined levels (Intonation Unit, Tonal Unit), distinguishing of primary from secondary accents at the Rhythmic Unit level (primary accents are indicated by underlining). It also shows the operation of ordered derivation rules that act upon the tones at the Tonal Unit and Intonational Unit levels. The initial L of an Intonation Unit and the L of the first Tonal Unit are simplified to a single L. The initial L is converted to M, a level in the middle of the f_0 range, and the final L is converted to B, a level at the bottom of the f_0 range.¹⁰

Di Cristo and Hirst explicitly reject the association of tones to individual syllables:

A la différence de la théorie autosegmentale standard (Beckman and Pierrehumbert (1986)) les segments tonals B(as) et H(aut) qui constituent les *primitives* de la description ne sont pas assignés dans le modèle à des syllables particulières ou à des accents, mais aux [Unités Intonatives et aux Unités Tonales].

‘Contrary to standard autosegmental theory (Beckman and Pierrehumbert (1986)), the L(ow) and H(igh) tonal segments that constitute the *primitives* of the description are not assigned in the model to particular syllables or to accents, but to [Intonation Units and Tonal Units].’

Di Cristo (2000), p. 34

The assumption of association of tones to syllables is also rejected in the earlier versions of the model (e.g., Hirst and Di Cristo (1984), Di Cristo and Hirst (1993)).

This rejection seems incompatible, however, with various of Di Cristo and Hirst’s claims about the nature of accent in French. For example, they argue for the primacy of pitch cues in French prosody (contrary to Wenk and Wioland (1982), for whom length is the most important cue (see §2.3.8): “[s]ince the main cue for rhythmic stresses is pitch prominence (R and D), [in Di Cristo and Hirst (1993)] we used the

¹⁰These initials are standard in the INTSINT (International Transcription System for Intonation) system developed by Hirst and Di Cristo (see Hirst and Di Cristo (1998)).

term Tonal Unit” They clearly treat syllables as accented, for example, employing notations like small capital letters to signal accented syllables and metric grids that assign varying levels of stress to different syllables. Di Cristo (1999a) discusses the phrase *sur les théories de la littérature* taken from the larger example in (7).

- (7) Il avait rédigé un ouvrage sur les théories de la littérature qui avait
 he had written a work on the theories of the literature that had
 rencontré un certain succès
 met a certain success
 ‘He had written a work on theories of literature that had had a certain amount of
 success.’

He describes:

un arc accentuel délimité, à sa gauche, par un accent mélodique sur la syllabe
 “ thé ” et à sa droite, par un accent de même nature sur la syllabe “ tur(e) ”.
 Six syllabes séparent ces deux accents. . . entraîne, par voie de conséquence, la
 réalisation d’un accent intermédiaire sur la syllabe “ ries ”.

‘an *arc accentuel* delimited on its left by a melodic accent on the syllable *thé*
 and on its right by an accent of the same type on the syllable *tur(e)*. Six
 syllables separate these two accents. . . as a result, an intermediary accent is
 realized on the syllable *ries* [emphasis added].’

p. 197

In any case, although the account makes reference to individual tones and to syllables, the association of the tones with the segment string is not clear. It is therefore not possible to predict, on the basis of the Di Cristo and Hirst model, what text-to-tune alignments should be possible. However, differences in alignment between the early rise and the late rise are unexpected in an account in which they are both of the same type (i.e., Tonal Unit), even considering the effect of their position with respect to the Rhythmic Unit. As we will see in the following chapter, these alignment differences are in fact considerable.

A number of questions have been raised about the model proposed by Di Cristo and Hirst and colleagues. Jun and Fougeron (2000) observe that the model violates the Strict Layer Hypothesis above the level of the foot. This hypothesis (Selkirk (1984), Nespor and Vogel (1986)) constrains prosodic structure according to a number of principles. It assumes that there are multiple layers of hierarchically organized prosodic structure. A prosodic unit at a given level is composed of one or more units of the next lowest level, and each prosodic unit is exhaustively contained in the next higher level. It is well known, however, that French prosodic structure below the word level violates the Strict Layer Hypothesis in two common post-lexical processes, liaison and enchainment (see, for example, Selkirk (1986)). In Figure 2.9, however, the word *disputés* is split between two Tonal Units. Jun and Fougeron (2000) question whether “placing a prosodic boundary in the middle of a word agrees with native speaker intuitions” (p. 11).

Hirst and Di Cristo (1996) claim that a major indicator of the presence of an Intonation Unit, the highest unit in their prosodic hierarchy, is a break in “melodic cohesion” (*cohésion mélodique*). Specifically, the occurrence of a H tone whose f_0 is higher than that of the preceding H tone is claimed to signal the start of a new Intonation Unit. There is, however, no evidence that suggests that listeners perceive this f_0 difference as a signal to a boundary. As a concrete example, it is not known whether there would be a difference in strength of the boundary at the end of the first prosodic phrase (*et les gamins sages*) among the three patterns in Figure 2.10—in the (a) version, the late rise and the early rise reach the same peak, in the (b) version, the peak of the late rise is higher than that of the early rise, and in the (c) version, the peak of the late rise is lower than that of the early rise. The issue of number of

levels of phrasing and what counts as a certain level of phrasing can only be resolved by systematic studies which tap native listener judgments. Such studies might use stimuli like the examples shown in patterns in Figure 2.10 and methodology like that of Rolland and Løevenbruck (2002), who tested listeners' perception of prosodic phrasing.

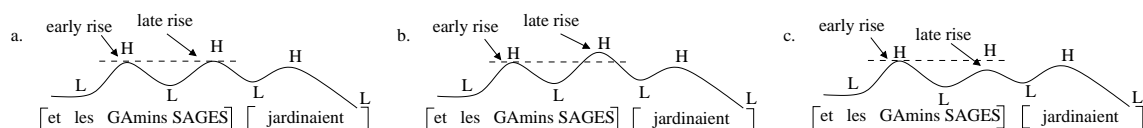


Figure 2.10: *Example of a prosodic phrase (et les gamins sages ‘and the good kids’) where (a) the peaks of the early rise and late rise have the same height, (b) the peak of the late rise is higher than that of the early rise, (c) the peak of the late rise is lower than that of the early rise.*

2.3.5 Mertens

In the Mertens model (Mertens (1987), Mertens et al. (2001), Mertens (2002)), intonation contours are decomposed into sequences of tones associated with individual syllables. Unlike standard Autosegmental Metrical models, all syllables are specified for tone. A tone is defined as the set of pitch levels associated with a syllable. The model includes four pitch levels: *L*- bottom of pitch range, *H*+ top of pitch range, *L* low, *H* high. For unaccented syllables, these pitch levels are written with lower case letters. In addition, there are minor pitch intervals that raise or lower the base pitch level, noted by adding a preceding / or \ (/L, \L, /H, \H).

Mertens distinguishes three tone classes, tones associated with a sequence of unaccented syllables (NA, *syllabes non accentuées*), “final accents” (AF, *accent final*), and “initial accents” (AI, *accent initial*), the last two corresponding to the primary and the secondary accent, respectively. The syllable bearing the primary accent is lengthened and may be followed by a pause. The syllable bearing the initial accent is not lengthened, but may be preceded by a pause. The three-way distinction is justified by the observed distributional properties of the three classes—they are not interchangeable. For example, one cannot substitute an initial accent for a final accent or add a final accent to a sequence of unaccented syllables.

The model contains three levels of grouping: the intonation group (IG), the stress group (SG), and the intonation package (IP). The structure of the intonation group (IG) is given in (8), where square brackets indicate optionality.

$$(8) \text{ IG} \rightarrow [[\text{NA}]\text{AI}] [\text{NA}] \text{AF} [\text{NA}]$$

The three tone classes are grouped together into intonation groups according to the following formula: an IG “is defined by the presence of a stressed syllable of type AF (accent final)” and can be preceded by one or more unstressed syllables (NA, non-accented) or by “a stressed syllable of type AI” (accent initial) (Mertens (2002), p. 499). In addition, an optional *appendix* of unaccented syllables follows the final accent (the [NA] to the right of the AF in Figure 8). This accounts for the intonation of sentence-level adverbs and postposed expressions.

An AF (final accent) tone is realized on the last full syllable of the IG and always contains two pitch levels. A two-level notation is maintained even for “tons statiques” ‘static tones’ in order to distinguish AI tones from AF tones (H vs. HH, for example) (Mertens et al. (2001), p. 7). An AI (initial accent) is typically realized on the first

syllable of non-clitic word. Syllables that bear the initial accent are not lengthened, although the articulation of their onset consonant is emphasized. The optional NA (non-accented) stretch to the left of the AI corresponds to preceding clitics (certain types of function words—Mertens et al. (2001), p. 11 gives an exhaustive list).

An intonation group (IG) in turn consists of one or more stress groups (SG). With the stress group, the model allows for more than one intonation pattern for sentences with the same syntactic structure. Stress groups are determined syntactically according to the following formula:

[a stress group] consists of a word *w* carrying lexical stress (word stress) and of all the adjacent clitic words (without lexical stress) that are governed by *w*.... When a SG is indeed stressed, the stress is located on [the] last full syllable ... of the SG which coincides with the AF of the resulting IG. When it is unstressed, it will be integrated in the IG on its right.

Mertens (2002) p. 499.

This merging of two SGs is illustrated in (9). As Mertens notes, the stress group corresponds to what others have called the prosodic word.

- (9) a. [nous avons visité] [son nouvel appartement]
 we have visited his/her new apartment
 ‘We visited his/her new apartment.’
- b. [nous avons] [visité] [son nouvel] [appartement]
 we have visited his/her new apartment
 ‘We visited his/her new apartment.’
- example from Mertens et al. (2001)

The intonation package (IP), which seems to specify a boundary tone for the entire utterance, is described as follows:

A boundary level is associated with each tone of type AF. A sequence of intonation groups therefore implies a sequence of prosodic boundaries resulting in a recursive and hierarchical organisation of the intonation groups into *intonation packages*. The largest package is that which carries a maximal boundary.

Mertens (2002), p. 500

From its earliest stages, the Mertens model of French prosody has been used as a model for synthesis. Examples from the implementation of the model, called *Mingus* (Modular Intonation Generation Using Syntax) can be found at:

<http://bach.arts.kuleuven.ac.be/pmertens/prosody/mingus.html>. According to native speakers, the system produces good quality results.

But airplanes do not flap their wings. And producing good quality synthesis and modeling a speaker’s tacit knowledge of intonation structure do not necessarily go hand in hand. One aspect of the Mertens model that is unlikely to be part of the phonology of the language, for example, is the tonal specification of unaccented syllables. Mertens et al. (2001) describes the tonal specification of these “atonal” stretches: “Une suite atone (NA) est représentée par deux localisations, indiquant le niveau de hauteur au début et à la fin de la séquence.” ‘An atonal sequence (NA) is represented by two locations, indicating the [pitch] level at the beginning and the end of the sequence’ (p. 7). As Post (2000) notes, unaccented syllables can be assigned pitch values by interpolation between the targets of accented syllables. As Post also notes, the high number of categories in the Mertens model seems unlikely to reflect phonological contrasts in the language.

2.3.6 Jun and Fougeron

Jun and Fougeron (2002) argue that the early rise and the late rise together form the Accentual Phrase (AP), which is claimed to be the basic unit of French intonation. An AP contains at least one content word (W_c), optionally preceded by additional content words and function words (W_f). In their Autosegmental Metrical account,

each rise is structurally a sequence of a L tone followed by a H tone (a LH unit), and the AP has the default structure /LHLH/. In this respect, it is similar to the Korean Accentual Phrase described in Jun (1998). In the French Accentual Phrase, the first rise is described as a phrase accent and the second as a pitch accent. This difference in treatment is due largely to the different behavior of the two H tones. The H of the early rise is not always realized on a full-vowel syllable and syllables upon which the H of the early rise is realized do not have a significantly longer duration than unaccented syllables. The H of the late rise, however, is systematically realized on the final full (non-schwa) syllable of an AP. In order to distinguish the two H tones notationally, Jun and Fougeron use *Hi* (initial H) for the high of the phrase accent (the early high) and *H** for the high of the pitch accent (the late rise). The star (*) conventionally indicates association of a pitch accent to a particular syllable. Jun and Fougeron (2002) use the notation /LHiLH*/ for the default AP pattern.

The model predicts five variant surface shapes formed by the absence of one or more tones of the default /LHiLH*/. The factors influencing which variants are realized are not fully described, but these tonally simpler variants often arise through undershoot when there is not enough time to realize all four tones of /LHiLH*/. A chart illustrating the six patterns is given in Figure 2.11. The structure of the last variant, in which there is a pitch fall, is slightly more complicated than the others. According to Jun and Fougeron (2002), the LHiL* is the least common of the five variants and occurs when “the pitch accent is immediately preceded by Hi and followed by Hi” (p. 11). The L part of the LHiL* pitch accent is realized as the starred tone on the accented syllable. Jun and Fougeron suggest that this variant avoids an impermissible series of three H tones.

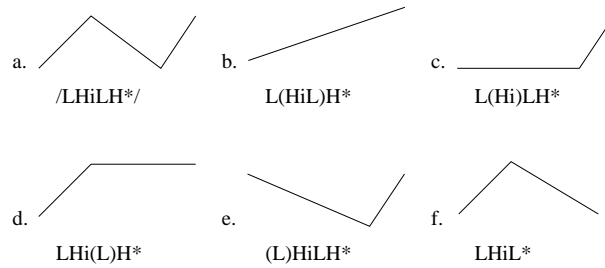


Figure 2.11: *Six predicted surface realizations of Accentual Phrase (AP) Jun and Fougeron (2002)).*

The Jun and Fougeron (2002) model abandons the claim found in an earlier model (Jun and Fougeron (2000)) that each tone is associated to a single syllable. Although Jun and Fougeron (2002) do not report new data, this change is probably motivated by a re-analysis of their data from the earlier paper, which were not consistent with the claimed tone/syllable correspondence. For example, the discussion of the patterns of tonal realization patterns in the Jun and Fougeron (2000) production study data makes it clear that two tones are sometimes realized within a single syllable.

In the current model, the tones of the early LHi sequence are not associated to syllables. Rather the LHi is treated as a phrase accent, with an association to the left edge of the Accentual Phrase. Jun and Fougeron observe that the L part of the LHi “can spread over all of the clitic syllables preceding the AP initial content word” (p. 5). The late rise is considered to be a LH* pitch accent. As in the earlier model, the H* tone is associated to the last full (non-schwa) syllable of the AP. The preceding L, however, is not associated to a particular syllable. This accounts for the observed

fact that the L of the LH* is not regularly realized in the syllable preceding the H*-toned syllable, but often occurs in the same syllable as the H* (Jun and Fougeron (2000), Welby (2002)).

The model also includes the Intonation Phrase (IP), a prosodic unit higher than the Accentual Phrase (AP). An IP contains one or more APs and is marked at the right edge by a L% or H% boundary tone, which is accompanied by “significant lengthening of the phrase final syllable” (p. 7) and an optional pause. As in the earlier model, this model assumes that an AP-final H* is “pre-empted” by the L% boundary tone, yielding a /LHiLL%/ pattern for some IP-final APs (in declaratives and *wh*-questions, for example). As Jun and Fougeron (2002) note, the details of the boundary tones and their interaction with the other tones remain to be worked out. The structure of the Jun and Fougeron (2002) Intonation Phrase is shown in Figure 2.12. As a point of comparison, the earlier Jun and Fougeron (2000) model, in which every tone is associated to a syllable is illustrated in Figure 2.13.

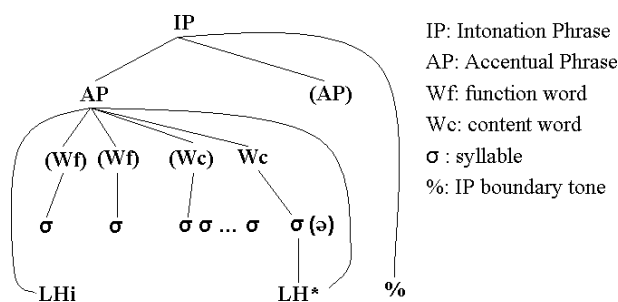


Figure 2.12: *Tonal associations, Jun and Fougeron (2002).* Jun and Fougeron note that the L of the LHi phrase accent can spread over the function words preceding the first content word. The figure is adapted from Jun and Fougeron (2002).

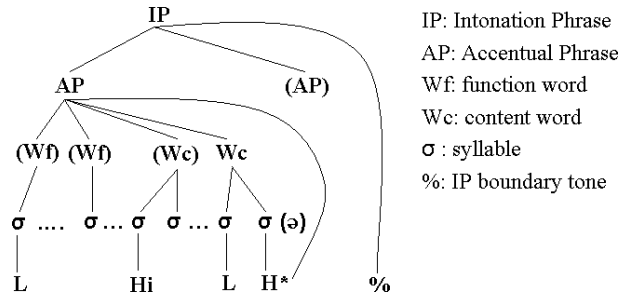


Figure 2.13: *Tonal associations, Jun and Fougeron (2000).*

Jun and Fougeron offer evidence from a number of sources for the claimed difference in status of the early and late rises. For the early rise, they note that the L and the Hi both seem to have a close association with the left edge of the AP. Even in cases where there are long stretches of unaccented syllables (due to a string of function words) the fall to the early L is fairly steep and is reached very near the left edge of the AP. (Figure 4c of Jun and Fougeron (2002) is a good illustration of this). According to Jun and Fougeron, an AP usually starts with a L tone, but it can also start with a Hi, surfacing as $/(L)HiLH^*/$. The location of the Hi tone of the early rise is much more variable than that of the H^* of the late rise. The early Hi is usually realized either on the first or second content word syllable of the AP (and even occasionally on the third), while the late H^* is consistently realized within the last full syllable of the AP. In addition, Jun and Fougeron note that studies have consistently found that the syllable in which the late H^* is realized is lengthened, while the syllable in which the early Hi is realized is not. Jun and Fougeron's observations make clear that the two rises have different timing characteristics, beyond the realization of the two H tones. They note that the L of the LHi phrase accent can

stretch over any preceding non-content words, while no such “spreading” is observed for the L of the late rise.

Although Jun and Fougeron note that the L of the early rise can extend from the left edge of the AP across any preceding function words, the tonal associations in the model do not predict this. The data in Welby (2002) show that the low inflection point (elbow) in tokens with preceding function words is consistently realized either very late in the last function word syllable or very early in the first content word syllable, a pattern consistent with a target at the function word/content word boundary. Welby (2002) accounted for the timing of the low plateau stretching from the left edge of the phrase to the L elbow with a double association of the L tone of the LH- phrase accent (early rise) to the left edge of the AP and to the left edge of the first content word of the AP (illustrated in Figure 2.14). The proposal follows the double associations described in Pierrehumbert and Beckman (1988) and Grice et al. (2000), but in the French case, there is a double association to two boundaries.

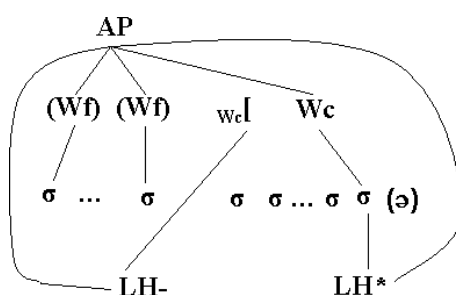


Figure 2.14: *Revision of the AP of Jun and Fougeron (2002) proposed by Welby (2002).*

There are a number of problems with considering the Accentual Phrase as a single unit with a four-tone structure. One question is raised by the observation that more than one early rise is possible in cases where there are long, polymorphemic content words or long stretches of clitics (see also Padeloup (1990) and Delais-Roussarie (1995) for discussion of cases with more than one early rise). Jun and Fougeron (2002) examined the realization of early rises in these exceptional or “over-accented” cases. They observe that the location of additional early H tones is variable, that additional early Hs are not accompanied by syllable lengthening, and that native speakers do not perceive a juncture at these Hs. They also note that others have observed that an early rise is more likely to occur on certain kinds of function words than others (e.g., polysyllabic words, negative adverbs, etc.; see Delais-Roussarie (1995) and Mertens et al. (2001)). They conclude that these accents should be considered phrase accents (LHi) rather than pitch accents (LH*), and that their presence may be conditioned by rhythmic and meaning constraints. But even if these “exceptional” early rises are more like unexceptional early rises than late rises, it is not clear how they are claimed to fit into the Jun and Fougeron Accentual Phrase. To which prosodic unit would an exceptional early rise be associated in the schematization in Figure 2.12?

Accounting for the utterance-final falls as the “pre-empting” of the H* pitch accent by the higher-level L% is also problematic. Jun and Fougeron describe the situation as follows: “Since the AP final syllable is also the Intonation Phrase final syllable, the AP final H* and Intonation Phrase final L% are supposed to be realized on the same syllable. In this case, the AP final H* is preempted by the higher level (Intonation Phrase) boundary tone” (p. 9). The suggestion seems to be that the two tones are not able to be realized in the same syllable. It seems unlikely that the failure of a

H*L% sequence to surface could be motivated by time pressure, since we know that a single syllable is commonly lengthened to accommodate two tone targets in the case of the LH early and late rise sequences. It is therefore unclear why both the H* pitch accent and the L% Intonation Phrase boundary tone could not be realized within the same syllable.

Finally, the assertion that the Accentual Phrase with a /LHiLH*/ pattern is the basic unit of French intonation is questionable. Jun and Fougeron (1995) found that the four-tone pattern occurred utterance initially in only 36% of their read speech corpus. The pattern is likely to be even less common in spontaneous speech, where evidence suggests that the early rise (corresponding to Jun and Fougeron's LHi sequence) occurs less frequently than in read speech. Finally, it is not known what the distribution of intonation patterns is in child-directed speech. Fonagy and Fonagy (1976) report that read fairy tales are characterized by "undulating" *arcs accentuels*, but the extent to which this pattern is present in spontaneous, child-directed speech needs to be investigated. Two scenarios seem possible. The first is that caretakers use the LHLH pattern (or *arc accentuel*) frequently in their speech to young children. This would be in line with studies that have shown distinctive prosodic characteristics of infant- and child-directed speech such as higher overall pitch and larger pitch excursions (see Fernald and Simon (1984), Jacobson et al. (1983), also Vihman (1996) for other references). Another possibility is that since caretakers probably use shorter words in speaking to children, their phrases will also be shorter and less likely to be long enough to accommodate a four tone pattern. Jun and Fougeron (2000) report that an average Accentual Phrase in their read corpus contained about 2.3-2.6 words (1.2 content words) and 3.5-3.9 syllables, in

line with the averages observed by other researchers for comparable phrasal units. If APs in child-directed speech contained fewer syllables on average (and so were physically shorter), the LHLH pattern would probably occur less frequently. This raises questions about how a child would learn that the LHLH pattern is the default pattern. In a study of the acquisition of intonation by French-acquiring and Japanese-acquiring infants, Hallé et al. (1991) found that French-acquiring infants produced rising patterns on 73% of their disyllabic productions, while the Japanese-acquiring infants produced falling or level contours on 74% of their productions. This matches the patterns found in the adult languages. Presumably, if the basic structure of French Accentual Phrases is LHLH, we could find evidence for this in research on infants or child learners. For example, we might expect that as children's utterances grow long enough, they show a preference for the two-rise pattern.

2.3.7 Post

Post (2000) proposes an account of the phonology of French intonation in which patterns of pitch accent distribution are accounted for by a set of ranked Optimality Theoretic well-formedness constraints.¹¹ Optimality Theory (OT) holds that there is a universal, language-independent set of violable constraints on well-formedness (Prince and Smolensky, 1993). Differences among languages, differences among speakers, and even differences within a single speaker, are claimed to arise from different rankings of these constraints. The candidate output form that incurs the fewest violations of the most highly ranked constraint or constraints is selected as the optimal candidate. Post's constraints make reference to the syntactic, morphological, and

¹¹She follows Delais-Roussarie (1995) in offering an Optimality Theoretic account of French intonation phenomena.

rhythmic structure of an utterance. A number of constraints are proposed to divide an utterance into Phonological Phrases of appropriate length and to require the presence of a pitch accent on the last full syllable of the Phonological Phrase. A larger unit, the Intonation Phrase, is marked by a boundary tone on either side and can contain one or more Phonological Phrase.

The tonal specifications of the Intonation Phrase in this model are shown in Figure 2.15. An Intonation Phrase can begin with either a %L or %H boundary tone and end with either a L% or H% boundary tone or zero boundary (0%). The initial boundary tone is followed by an optional early H* pitch accent and by an obligatory late H* or H+H* pitch accent. If there is a sequence of two H tones, a L tone may be inserted between the two. This account differs from the Jun and Fougeron models in a number of respects. One important difference is that the H tones of both the early and the late rise are treated as pitch accents. The first L tone is treated as an initial boundary tone, while the phonological status of the second L tone and its alignment are less clear—it is described as a tone that is “inserted” between two H tones. Post claims that the “low tone is usually aligned with the pre-accentual syllable...” (p. 153), although she later states that “[a]lthough the low target is often realised in the accented syllable in IP-final rising and IP-internal rising-falling movements, ... it can also occur in the penultimate syllable...” (p. 168).

Post follows Verluyten (1982), Christophe (1993), and Delais (1994) in assuming that the lowest level of prosodic grouping corresponds to the syntactically derived Phonological Phrase of Nespor and Vogel (1982). The Phonological Phrase is the domain of accent assignment. The Phonological Phrase formation rule is given in

$$\left\{ \begin{array}{c} \%L \\ \%H \end{array} \right\} (H * (L))_0 \left\{ \begin{array}{c} H* \\ H + H* \end{array} \right\} \left\{ \begin{array}{c} L\% \\ H\% \\ 0\% \end{array} \right\}$$

Figure 2.15: *Tonal specifications of the Intonation Phrase, Post (2000)*. Curly braces indicate a set from which a tone can be selected and parentheses indicate optionality. Tones marked with ‘*’ are pitch accents and tones marked with ‘%’ are boundary tones.

(10), and an illustration of the rule is given in (11). Phonological phrases are subject to restructuring under certain conditions.

(10) **Phonological Phrase formation rule**

A Phonological Phrase groups together a lexical head (X) with all the items on its non-recursive side (i.e., the left [for French]) within the maximal projection and with any other non-lexical item on the same side. (French pre-nominal adjectives cannot function as heads of PPs.)

- (11) [Ces petits enfants]_{PP} [apprennent]_{PP} [à parler]_{PP} [le français]_{PP}
these little children learn to speak the French
‘These little children are learning to speak French.’

I will not review in depth all the proposed OT constraints. Rather I will sketch some of the relevant details.

The account includes constraints that require the edges of lexical words and prosodic words to coincide and prosodic words to have pitch accents at their right edge. It also includes a constraint that avoids stress clash (“accent clash”).

To partially account for the finding in her data and elsewhere that Phonological Phrases tend to contain a maximum of six or seven syllables, Post adds a constraint to limit the length of the PP. Her model correctly excludes long PPs like (12b) but allows a PP-final function word to be accented in cases like *Prends-LE* ‘take it’.

- (12) a. [des hivers]_{PP} [autres qu'en Afrique]_{PP}
 some winters other than in Africa
 'winters other than those in Africa'
- b. * [des hivers autres qu'en Afrique]_{PP}
 some winters other than in Africa

To account for cases in which Phonological Phrases have two rises, a late rise and an early rise, Post introduces two constraints, one requiring every Prosodic Word to have a pitch accent on its left side, and a more highly-ranked constraint requiring a Prosodic Word to have a pitch accent on its right side. Of course, all prosodic words do not have early rises, but the relative ranking of the two constraints and a constraint against pitch accents prevents the over-generation of secondary pitch accents.

Post also attempts to account for the prosodic variation due (presumably) to post-lexical factors such as speaking rate. The early rise does not always appear in phrases where it is predicted to be possible, and it sometimes appears on a syllable not predicted by the current constraint rankings. For example, *de jOLIS enFANTS* is predicted, but *de JOlis enFANTS* is also possible. Post models this through the variable ranking of certain constraints.

An advantage of the Post account is that it accounts for both categorical and gradient aspects of intonation patterns attested in the language, although it is not always clear that the line between categorical and gradient has been correctly drawn.

One weakness of the account is that the OT analysis concentrates on the division of an utterance into Phonological Phrases and the distribution of H* and H+H* accents within the PP, but has little to say about the behavior or timing of preceding L tones. Post claims that:

...the speaker chooses to realise a low tone, and thereby modifies the interpretation of the pitch accent. L-insertion leads to a more explicit separation

of the items marked by the starred tone. The low tone is usually aligned with the pre-accentual syllable, which results in a contour that is very similar to the one described by Gussenhoven (1984) for Dutch and British English as partial linking. When the low tone [the L before the late H] is not inserted, the items form a conceptual unity (cf. Gussenhoven (1984): 246-248).

p. 154

In other words, there is predicted to be a pragmatic difference between the %LH*H* and the %LH*LH* patterns in Figure 2.16. There is, however, no evidence provided for this claim. It is difficult to reconcile the claim with the observation that patterns with and without the L preceding the late H* are routinely observed for utterances that have the same text (e.g., Welby (2002)). The claimed difference is also surprising since Fonagy (1979) found that the *arc accentuel* (Post's %LH*LH* pattern) was common in cases where there was *more* semantic unity within the phrase (see §2.1.2). Vaissière (2002) makes a similar observation, noting that a pronunciation of *le petit gamin* 'the little kid' with f_0 peaks at the beginning of *petit* and at the end of *gamin* (a pronunciation corresponding to an early rise followed by a late rise, Fonagy's *arc accentuel*) will convey more "semantic dependency" than a pattern with f_0 peaks at the end of *petit* and at the end of *gamin* (corresponding to two late rises). Fonagy contrasted the presence of the *arc accentuel* with the absence of any early rise, not with the absence of an intermediate low point. Following Fonagy, we might expect both patterns in Figure 2.16 to indicate conceptual unity relative to a phrase without the early rise.

In addition, Post treats both early H and late H as pitch accents, although their temporal alignment and that of the L tones that precede them are very different.

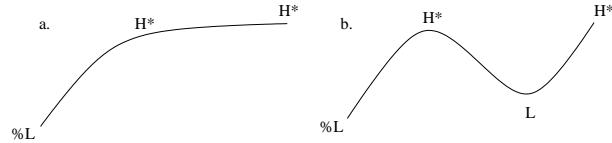


Figure 2.16: *Two patterns with and without a L preceding the late H*:* (a) %LH*H*, (b) %LH*LH*.

The differing alignment of the f_0 maxima corresponding to the H tones is well established and uncontroversial, and Welby (2002) provides experimental evidence for the differing alignment of the two L tones. Of the early and late rises Post (2000) writes:

In the view adopted in this thesis, functional differences between intonation contours are not necessarily mirrored by structural differences, and therefore do not in themselves provide a sufficient basis for positing distinctions at the phonological level. This means that word-initial and word-final pitch movements would need to show a consistent formal difference for them to be phonologically distinct.

p. 140

Based on the results of the production study reported in chapter 3 and of Welby (2002), we will argue that formal differences do exist.

Another problem for the Post model is that it is unable to account for all accent patterns reported in the literature. For example, the fact that a non-initial L tone can only be inserted between two H tones, excludes a LLH* pattern. This pattern, described by Jun and Fougeron (2000), and illustrated in Figure 2.11c, is probably the same as the Post (2000) one-accent IP pattern with a “level onset” (illustrated in her Figure 7 (p. 122)). The “level onset” description, however, does not adequately account for the first L of the LLH* pattern—that this first L is a distinct target is particularly clear in utterance-medial phrases, where f_0 falls to reach a target at a content word beginning.

Finally, the claimed phonological distinction between the H^* and $H+H^*$ pitch accent types is suspect. According to Post (2000), “the leading high tone [of the $H+H^*$] associates with the penultimate syllable, and the H^* with the final accented syllable of the Intonation Phrase” (p. 158). The critical difference, however, seems to be the timing of the preceding L—the L preceding the $H+H^*$ is realized earlier than the L preceding the H^* pitch accent (see Post (2000), p. 159, Figure 10). Since the leading H tone of the $H+H^*$ is an associated tone, it follows that a preceding L tone must be realized earlier.¹² Post herself notes that her data and evidence from the literature show a great deal of variability in the position of this L tone. She argues on the basis of her production and perception studies that this variability is gradient, not categorical, and should therefore not be captured in the phonology (Post (2000), Post (1999)). This argues against the proposed H^* vs. $H+H^*$ contrast.

2.3.8 Intonation and the other aspects of French prosody

In this dissertation we focus on the structure of French intonation and its role in speech segmentation. In so constraining the research topic, we do not deny or minimize the importance of other prosodic factors or their contribution to speech segmentation. We agree the observation of Di Cristo (1999b), that the problem is multiparametric:

... la pluralité des stratégies d’encodage de la proéminence, qui varient, notamment, selon les particularités du système prosodique de la langue (Berinstein (1979)) et pour une langue donnée, selon le contexte linguistique et la catégorie dont relève l’accent (Beckman and Edwards (1992)).

p. 147

¹²Note that association of the leading or trailing tone of a bitonal accent is not standard in Autosegmental Metrical frameworks, but has been suggested by Arvaniti et al. (2000).

‘...the numerous strategies for encoding prominence, which vary, notably among the specifics of the language (Berinstein (1979)) and for a given language, according to the linguistic context and the category of the accent (Beckman and Edwards (1992)).’

In fact, as we will see, the results of the perception experiments reported in chapter 5 show clear contributions of factors other than intonation.

Yet some researchers have emphasized the importance of certain parameters to the exclusion of others. For example, some researchers claim that duration is the primary correlate of accent in French. Wenk and Wioland (1982) write of “the length accents of French” (p. 202) and call syllable lengthening “the single most effective cue to French accent” (p. 201) and “the most dependable manifestation of French accent” (p. 214). Wenk and Wioland (1982) describe Delattre’s position:

Delattre (1966), pp. 190,193 states unequivocally that French ‘stress is perceived *exclusively* by an excess of duration.’ For him (Delattre (1951), p. 45), frequency variation is no more than a possible (“non-indispensable”) carrier of accent in French. [emphasis added]

pp. 195, 196

Wenk and Wioland themselves explicitly allow a role for intonation as a cue to accent in French, but argue that it is not the primary cue. Fletcher and Vatikiotis-Bateson (1991), on the other hand write, “final lengthening in French involves a specific lengthening at the phrase-edge. Accentuation, by contrast, is a change in linguistic prominence and not essentially a duration contrast”(p. 18).

In the domain of perception of French accent, most previous research on the role of prosodic cues to speech segmentation in French have focused on the contribution of duration. This dissertation aims thus to partially fill this gap in the literature.

CHAPTER 3

PRODUCTION EXPERIMENT 1: TONAL ASSOCIATION AND ALIGNMENT

[D]ans aucune langue il n'est aussi facile de déterminer la place de l'accent que dans le français. . . .

'In no other language is it as easy to determine the position of the accent as in French. . . .'

Friederich Diez (1836), quoted in Fonagy (1979)

3.1 Y a pas de quoi en faire une thèse ?

If Diez was right, then *y a pas de quoi en faire une thèse*!¹³ Happily, though, the characterization of French intonational structure turns out to be a problem complex enough to occupy the energy of many researchers and to fill many dissertations, including this one.

3.2 Introduction

The production experiment presented here addresses a number of hypotheses about the association of the tones of the early and late rises in French intonation. The

¹³ *Y a pas de quoi en faire une thèse* is a popular expression which literally means 'There's nothing to write a dissertation about' and corresponds roughly to the English 'there's not much to write home about'.

first two hypotheses deal with the placement of the H tone of the early rise, which has thus far not been adequately described. The first is given in (13).

- (13) EARLY H ASSOCIATED TONE HYPOTHESIS: The H tone of the early rise is associated to a particular syllable.

This association has been proposed by Jun and Fougeron (2000) and Post (2000). If the early H is an associated tone, the f_0 peak should consistently appear either within the target syllable or in some position systematically related to a syllable landmark (e.g., just after the syllable offset), regardless of the length of the syllable. A regression analysis should show that the distance from the syllable onset to the early H tag varies across syllable lengths. The second hypothesis is given in (14).

- (14) EARLY H TRAILING TONE HYPOTHESIS: The H is a trailing tone of a bitonal LH accent.

Bitonal pitch accents are standardly assumed to consist of a starred tone (L^* or H^*) and a non-starred tone that is not associated to a particular syllable but leads or trails the starred tone at a “fairly invariant” interval (Pierrehumbert and Beckman (1988), p. 123 and Pierrehumbert (1980)). Factors such as segmental composition are not predicted to have a systematic effect on the timing of the non-starred tone. If the early H is a trailing tone of the initial L, we would expect it to follow at a fairly fixed distance, regardless of syllable length. In this case, the H might consistently appear within or just beyond the target syllable with phonetically long vowels, but often appear well after the syllable offset for syllables with short vowels. For example, the f_0 peak might appear in the first syllable of the target *magnanime* ‘magnanimous one’, if that syllable is phonetically long, but in the second syllable of the target word *minimum* ‘minimum’, if the first syllable is short. Figure 3.1 shows possible patterns for both hypotheses for syllables of different lengths. The dashed lines represent the

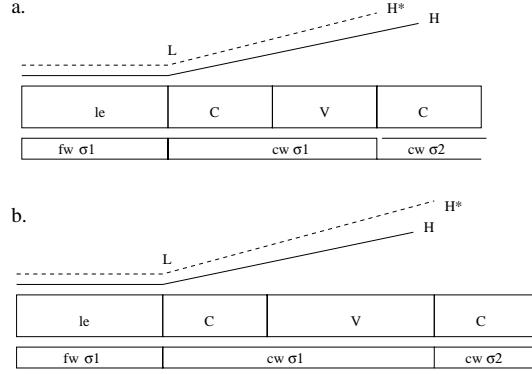


Figure 3.1: Possible tonal patterns of (a) the Early H Associated Tone Hypothesis (dashed lines) and (b) the Early H Trailing Tone Hypothesis (solid lines) for syllables of different lengths. The upper tier represents segments and the lower tier represents function word and content word syllables. Cw stands for content word and fw stands for function word.

patterns expected for the Early H Associated Syllable Hypothesis. In each case, the f_0 peak (marked H*) is reached at the same position relative to the hypothesized associated syllable (late in the syllable) regardless of vowel length. The solid lines represent the patterns expected for the Early H Trailing Tone Hypothesis. In this case, the f_0 peak (marked H) is reached at the same latency from the L for all three syllable lengths.

The next hypothesis, given in (15), deals with the association of the L of the early rise, whose alignment has been reported to be much more stable than the H of the early rise. We test the account of the double association of the early rise given in Welby (2002).

- (15) EARLY L ASSOCIATED TONE HYPOTHESIS: The L tone of the early rise is doubly associated to the beginning edge of the first content word syllable of the prosodic phrase and (optionally) to the beginning edge of the phrase.

The double association of a tone was first proposed by Pierrehumbert and Beckman (1988) to account for the alignment of the phrasal H of Japanese. Double association has recently been used to account for alignment phenomena in German, Greek, Hungarian, and Romanian (Grice et al. (2000)).

Two hypotheses are tested for the L of the late rise. The first is given in (16).

- (16) LATE L ASSOCIATED TONE HYPOTHESIS: The L tone of the late rise is associated to a particular syllable.

Jun and Fougeron (2000) argue that the L of the late rise is associated to the penultimate syllable of the prosodic phrase, although the Jun and Fougeron (2002) version of the model does not include this claim.

According to reports in the literature, the L elbow preceding the f_0 peak of the late rise can appear in the same syllable as the H of the late rise or in the preceding syllable (e.g., Post (2000), Jun and Fougeron (2000), Welby (2002)). This variation may be compatible with an account of the L of the late rise as a leading tone, as described in the hypothesis in (17).

- (17) LATE L LEADING TONE HYPOTHESIS: The L tone of the late rise is a leading tone of the H of the late rise.

If this hypothesis is correct, the L of the late rise should appear at some fixed interval before the H of the late rise.

A last hypothesis, given in (18), addresses the association of the H of the late rise. It is motivated by the widely reported observation that the peak of the late rise is typically realized in the last full syllable of the phrase. If the hypothesis is correct, we expect the late H to be realized in some position systematically related to that syllable.

- (18) LATE H ASSOCIATED TONE HYPOTHESIS: The H of the late rise is associated to the last full syllable of the phrase.

The experiment also tests factors that may influence tonal alignment. These factors include position of the phrase in the utterance and speaking rate. For example, the L of the early rise may be more consistently aligned when the phrase is utterance initial than when it is utterance final, since the speaker's f_0 does not need to fall from a preceding H. The early L may also be more consistently aligned when the speaking rate is slower, since the speaker has more time to reach the L target.

3.3 Methods

3.3.1 Materials

A set of 108 sentences was designed. The following four factors were manipulated: position of the target word within the sentence (initial, medial), and number of function word syllables preceding the target word (1, 2). Length of the hypothesized associated syllables was not directly manipulated, but was expected to vary across the different target words.

There were 27 target words ($\times 4$ factors = 108). All target words were directly preceded by a definite article (*le*, *la*, *les*). Nine target words were masculine, singular, and consonant-initial; 9 were feminine, singular and consonant-initial; and 9 were plural and vowel-initial. The targets varied in length, with 3 2-syllable words, 3 3-syllable words, and 3 4-syllable words for each type.

In order to maximize the chances of obtaining a smooth f_0 curve, the target words contained almost all sonorant or voiced segments. An example set of four sentences is given in (19). The target words are italicized. The preceding function words (*et* and *le*) are underlined and the hypothesized associated syllable of the early H (the first syllable) is double-underlined. Examples (19a) and ((19b)) show the target words

in utterance-initial position and examples (19c) and (19d) show the target words in utterance-medial position. The entire 108 sentence set is given in Appendix B.

- (19) a. Le minimum sera calculé par Manon.
 ‘The minimum will be calculated by Manon.’
 b. Et le minimum sera calculé par Manon.
 ‘And the minimum will be calculated by Manon.’
 c. Le maximum, le minimum et les écart-types seront calculés par Manon.
 ‘The maximum, the minimum and the standard deviations will be calculated by Manon.’
 d. Le maximum et le minimum seront calculés par Manon.
 ‘The maximum and the minimum will be calculated by Manon.’

The variation in materials was intended to reduce the monotony of the reading task and to allow us to examine other claims made in the literature. Various predicates (each 8 syllables in length) were used for the different target word sets. If a carrier phrase had been used and only the subject varied, the target words might have been read with an intonation consistent with contrastive narrow focus (emphasis on *minimum* in (19), for example). The syllable length variation allowed us to examine predictions about when the 4-tone LHLH pattern should be possible. Since Jun and Fougeron (2000) claim that each tone is associated to a separate syllable, that model predicts that the 4-tone LHLH pattern should never appear on 3-syllable phrases like *le moulin* ‘the mill’, with a 2-syllable content word and a preceding monosyllabic function word. Vowel-initial targets were included to allow us to investigate one of the findings of Welby (2002). Underlyingly vowel-initial words can become consonant-initial through the process of liaison (e.g., *les anomalies* [le.zanomali] ‘the anomalies’). For cases involving liaison consonants, Welby (2002) found that the L of the early rise was aligned not to the morphological boundary, but to the syllable boundary. For example, we would expect to find the early L aligned to the boundary between the

syllables [le] and [za] for *les anomalies*, not to the morphological boundary between the determiner *les* and the following content word.

3.3.2 Subjects

Seven native speakers of French participated in the experiment. They were all women from Paris or the surrounding region. Female speakers were chosen to minimize tracking errors: compared with male voices, female voices typically have higher fundamental frequency and wider pitch ranges with respect to microprosodic effects. The speakers ranged in age from 21 to 35, with an average of 28.6. Five of the speakers had doctoral degrees; the other two had the equivalent of a bachelor's degree (*baccalauréat* + 4 years of university education).

3.3.3 Procedures

The 108 sentences were randomized in order to avoid contrastive focus effects. Participants in the experiment read the corpus aloud two times, first in a self-selected normal rate and second in a fast rate, and were recorded onto digital audio tape (DAT). The instructions given to participants are provided in Appendix A.

Data analysis

Recordings were transferred from DAT to computer using a digital-to-digital cable and downsampled to 22.05 kHz using CoolEdit software. The soundfiles were segmented and each utterance saved as a separate file. F_0 curves and spectrograms were created using Praat speech analysis software (Boersma and Weenink (1992–2001)).

Word and syllable boundaries were tagged for each target word/function word set (which corresponded to the critical phrase), using waveforms and spectrograms to guide the segmentation. The beginning and end of each utterance were also tagged.

Unfortunately, the voiced, non-sonorant segments included in the materials introduced segmental perturbations which made it difficult to take critical f_0 measurements and sometimes to accurately categorize tone pattern. These perturbations were not apparent in the pilot recordings. All of the plural critical items, which included the [z] liaison consonant, were therefore excluded from the analysis, as were three sets of items that included target words with voiced, non-sonorant segments. Sixty usable sentences remained for each subject. Since each list was read at two speaking rates, a total of 120 tokens remained for each of the seven speakers. Of these 840 (120×7) tokens, 13 contained disfluencies in the critical region and were discarded. Another 37 that began with the conjunction *et* were set aside for future analysis, since they were produced with f_0 rises on the conjunction. A total of 790 tokens remained for the analysis.

Intonational features of the early and late rises were tagged to allow the hypotheses discussed in §3.2 and the prevalence of different accent patterns to be examined. The classification of accent patterns is discussed in §3.4.2. Figure 3.2 shows a schematization of a labeled f_0 curve. The following features were tagged by hand: f_0 at the beginning of the utterance, early L tone (for initial phrases) or late H of preceding phrases (for medial phrases), early H tone, and late H tone. Since each of the H tones was typically preceded and followed by a L tone, the location of a H tone could be defined as the time of the maximum f_0 at the end of the rise from the preceding L and the beginning of the fall to the following L target. When there was a plateau, a

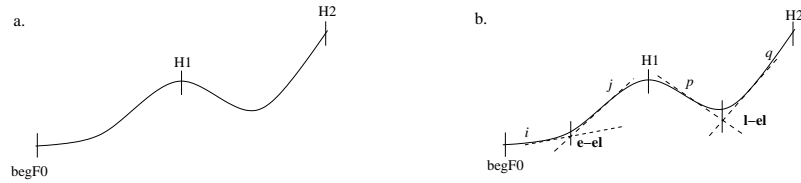


Figure 3.2: *Schematization of the tonal tagging of a critical phrase. (a) Tags added by hand. (b) Tags whose position was automatically calculated (bold). The reference points (arbitrarily labeled *i*, *j*, *p*, *q*) are indicated in italics. The dotted lines represent the regression lines calculated by the script that determined elbow position. The elbow tag was placed at the intersection of the two regression lines.*

series of points having fairly level f_0 values, the first point of this series was selected as the peak location. High plateaux were labeled as described in §3.4.2.

Low tones other than the L of the initial f_0 could not be tagged in a consistent manner by hand since they did not always correspond to f_0 minima (see Figure 3.2). The following procedure was therefore adopted. Reference points around the region of the “elbow” (inflection point) flanking the early and late H were tagged. The elbows were calculated and inserted automatically using a series of scripts. Fundamental frequency and time information was first extracted for each file using a Praat script. The position of each elbow was then calculated using an R script encoding a procedure described by D’Imperio (2000):

[A]n automatic procedure was employed by which two straight lines were fitted to the f_0 segment going from [the reference points (e.g., *i* and *j* in the example in Figure 3.2)]. The parameters of the two linear models were estimated by means of conventional linear least-squares methods. To estimate the elbow position, i.e., the intersection of the two fitted lines, two linear regressions were computed for each possible elbow location. . . . The location eventually selected as the “elbow” was the one leading to the smallest total modeling error.

p. 92

A final Praat script extracted the elbow f_0 and time information for the elbows from the files outputted by the R script and inserted labels into the Praat TextGrid files. The automatic elbow labeling was visually inspected, and obvious errors in elbow placement were hand-corrected for cases where the elbow formed a sharp angle and the algorithm had clearly chosen an elbow position on the basis of anomalous f_0 points. These hand corrections were rare: of the 1,332 elbow tags, 69 (or 5.2%) were hand-corrected. The tag *e-el* refers to the elbow immediately preceding the *early* rise, usually appearing near the function word-content word boundary. The tag *e2-el* marks an optional early elbow that precedes *e-el*. This elbow can occur in medial position, when the fall from the preceding H reaches a L target before the function word-content word boundary and there is a low leading plateau before the early rise. The tag *l-el* marks the elbow immediately preceding the late rise. The tag *l2-el* marks an optional late elbow which can precede *l-el*.

Time values for all tags, durations, and f_0 values were automatically extracted from the label files by a Praat script.

3.4 Results and Discussion

3.4.1 Rate manipulation

Speaking rate was calculated according to the following procedure. First, each of the 108 sentences was assigned a phonetic transcription corresponding to the assumed most likely pronunciation of that sentence and a phone count was assigned to each sentence. With the aid of a Praat script, I systematically listened to each utterance and noted any deviation from the base pronunciation. The phone count for each

utterance was adjusted by adding or subtracting the appropriate number of phones from the base count. An example is give in (20).

- (20) a. Le magnanime peut être trouvé dans le café.
 [lə maɲanim pø t-ɛtrə truve dɑ̃ lə kafe] (base phone count: 29)
 ‘The magnanimous one can be found in the cafe.’
- b. Le magnanime peut être trouvé dans le café.
 [lə maɲanim pø ɛt truve dɑ̃ lə kafe] (corrected phone count: 29 – 3 phones)

The base pronunciation, given in (20a), has 29 phones, including a liaison [t] that avoids vowel hiatus between the main verb *peut* and the infinitive *être*. In the fast speaking rate, Speaker 2 produced this sentence without the liaison [t] and without the final syllable ([rə]) of *être*. Three phones are therefore subtracted from the base count of 29 to yield a phone count of 26 for this utterance.¹⁴ A separate Praat script extracted utterance duration from each label file. The duration file and the phone count file were combined and rates were calculated in phones per second.

Speaking rates for all speakers are summarized in Figure 3.3. Speaking rates in the normal rate ranged from 11.5 phones/s for Speaker 4 to 14.2 phones/s for both Speaker 6 and Speaker 7. Rates in the fast rate ranged from 13.6 phones/s for Speaker 4 to 17.7 phones/s for both Speaker 1 and Speaker 6. All seven subjects succeeded in increasing their speaking rate in the fast rate. Speaker 2 had the smallest percentage change, raising her rate from 13.1 to 15.2 phones/s for a 15.66% change.

¹⁴Phone deletions are likely to be part of a speaker’s strategy for increasing speaking rate. Considering this, my calculation is a conservative measure of the fast speaking rate. I also recognize that speaking rate is not adequately characterized by a simple quantitative measure like phones per second. Studies on French have shown that changes in speaking rate lead to modifications in pitch range, intonation patterns, and prosodic phrasing (Fougeron and Jun (1998)) and that phenomena such as enchainment, pause insertion, and syllable lengthening demonstrate qualitative differences linked to speaking rate (Zellner (1998), Zellner-Keller (in press); see also references therein). The important thing in the current study, however, is to show that participants speeded up from the normal to the fast rate.

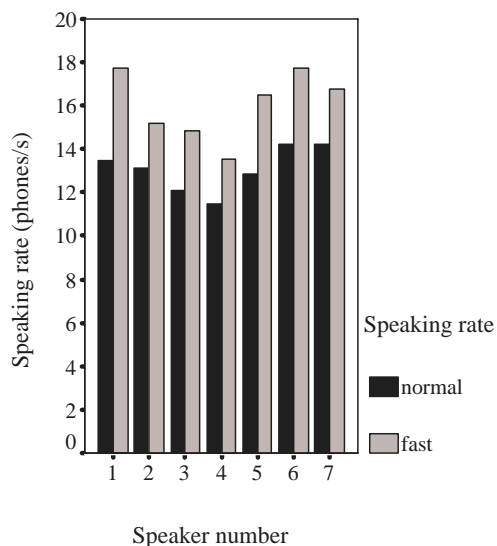


Figure 3.3: *Speaking rate by subject.*

Speaker 1 had the largest percentage change, raising her rate from 13.5 phones/s to 17.7 phones/s for a 31.5% change.

3.4.2 Accent patterns found

In order to describe the accent patterns found, the inventory proposed by Jun and Fougeron (2000) and Jun and Fougeron (2002) is adopted as a starting point. I chose to work with the Jun and Fougeron model because, of all the models discussed, it makes the clearest predictions about the alignment details to be examined. I will, however, need to make some modifications to the proposed inventory. The patterns of the Jun and Fougeron model, which are discussed in Chapter 2, are presented again as Figure 3.4. The Accentual Phrase patterns found in the data are summarized in Figure 3.5.

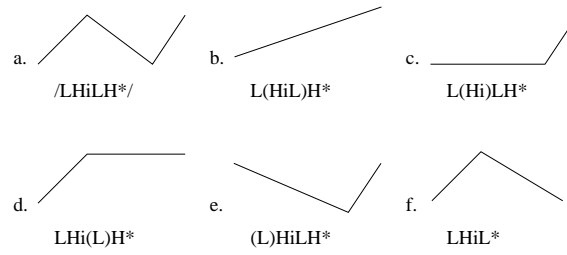


Figure 3.4: *Six predicted surface realizations of Accentual Phrase (AP) Jun and Fougeron (2002).*

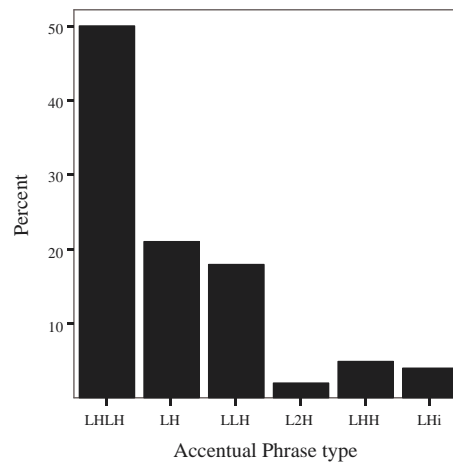


Figure 3.5: *AP types found as a percentage of the total. The L2H AP type, not predicted by Jun and Fougeron (2002), is discussed on p. 87.*

LHLH

By far the most prevalent accent pattern was LHLH, which corresponds to Jun and Fougeron's LHiLH*. The pattern was defined by the presence of both an early and a late rise, with a clear elbow between the two H tones. The LHLH pattern was produced on 50% of the critical APs.¹⁵ The LHLH pattern was often preceded by a low plateau or a fall to the initial L. Examples are given in Figures 3.6a, 3.6b, and 3.6c. For all examples, the content word of the critical AP is labeled *target*. Preceding function words are labeled *et* for the conjunction and *det* for determiners. The three tiers shown are the word tier, the segment tier, and the tone tier. To minimize the clutter on the tone tier, the reference points for the placement of the elbow tags have been removed.

The medial L was sometimes realized as a plateau, with two clear elbows, as illustrated in Figure 3.7.

Scaling of H1 and H2 of LHLH In the data, the H tone of the late rise (H2) generally had a higher f_0 value than the H tone of the early rise (H1). The peak of the late rise was higher than that of the early rise in 83% of cases. The results of a Wilcoxon Signed Ranks Test showed that this difference was significant ($Z = -14.571$, $p < .001$). The histogram in Figure 3.8 charts the difference. Note that the values are

¹⁵The percentage of critical APs with the LHLH pattern would probably have been even greater, had it not been necessary to exclude three four-syllable target words (*malavisé*, *misogynie*, and *mammalogie*) from the analysis because of segmental perturbations. However, this still left three other four-syllable target words, read in eight conditions by each of the seven speakers ($3 \times 8 \times 7 = 168$). Of these, a total of 155 tokens with four-syllable target words were included in the analysis. The contribution of syllable length and duration of the targets in clock time is discussed later in the chapter.

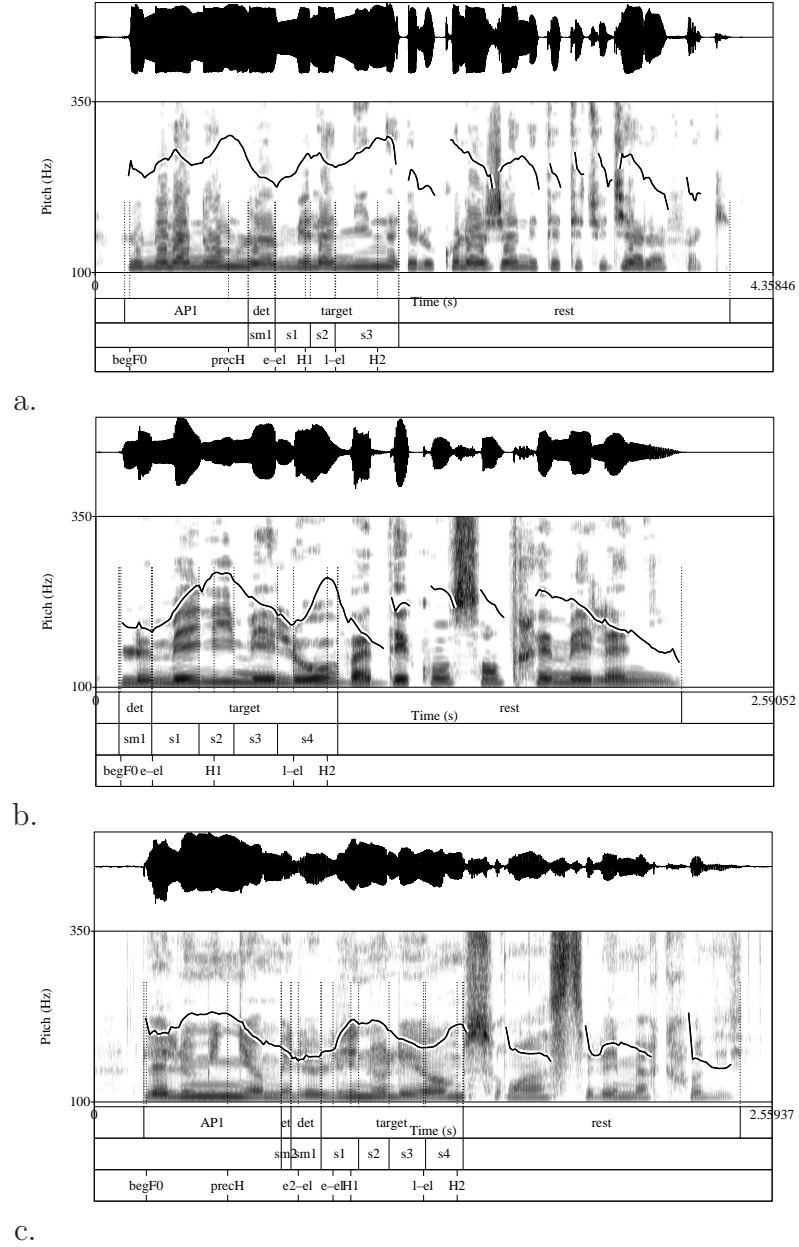


Figure 3.6: (a) *LHLH pattern in the AP et la mélanine in utterance-medial position. The full utterance is Le mélanome et la mélanine étaient étudiés à la fac. ‘Melanomas and melanin were studied at college.’ There is no leading plateau, but a fall from the preceding H to a L right at the function word-content word boundary. The initial AP of the utterance also has an LHLH pattern, with a slight fall to an early elbow.* (b) *LHLH pattern with a leading plateau in the utterance-initial AP le méli-mélo. The full utterance is Le méli-mélo va déconcentrer Mélanie ‘The mess is going to distract Mélanie.’* (c) *LHLH pattern in the utterance-medial AP le linoléum. There is a fall from the preceding H to a low leading plateau. The full utterance is L’aluminium et le linoléum seront installés mercredi. ‘The aluminum and the linoleum will be installed on Wednesday.’*

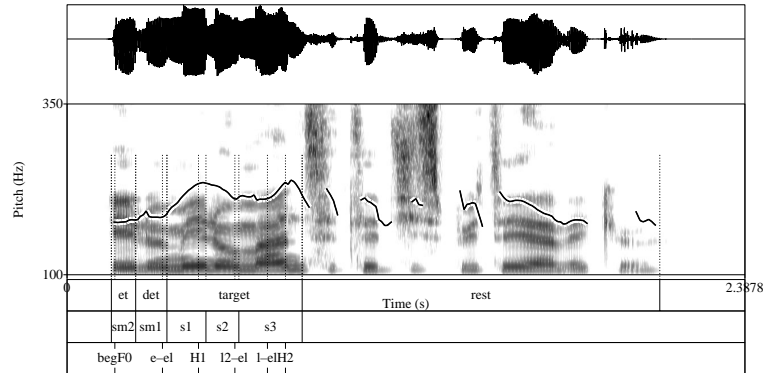


Figure 3.7: *LHLH* pattern with a leading low plateau and a low plateau between *H1* and *H2* in the *AP* Et le mélomane. The full utterance is Et le mélomane s'était disputé avec eux. 'And the music lover had argued with them.'

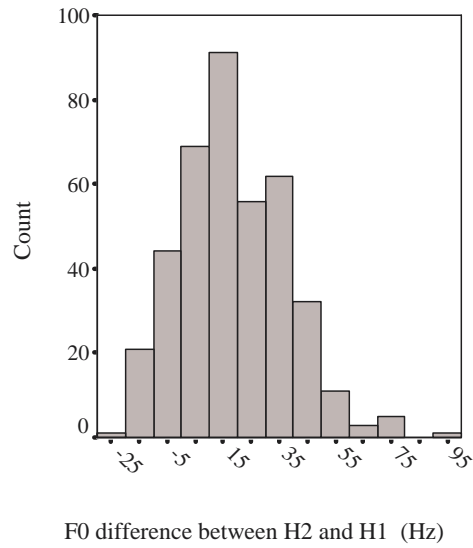


Figure 3.8: *Fundamental frequency differences between two tones of the LHLH pattern. Difference is $H2 - H1$ for each LHLH token.*

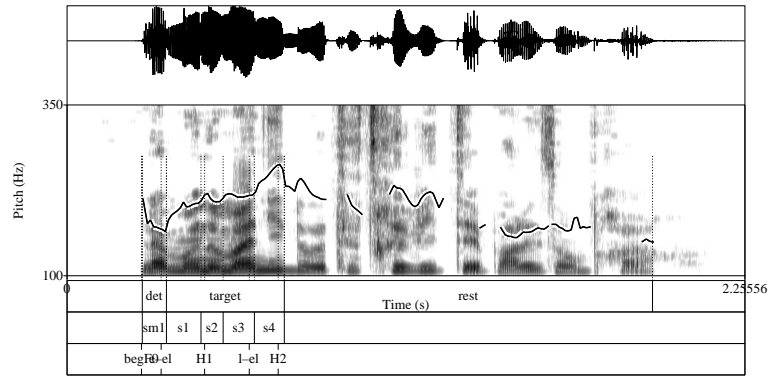


Figure 3.9: (b) *LHLH with a high plateau extending from the early H to the late L elbow in the AP la monomanie. The full utterance is La monomanie doit être évitée par les empereurs. ‘Monomania should be avoided by emperors.’*

skewed to the right of a 0 Hz difference. A similar scaling difference was also found by Rolland and Lœvenbruck (2002).

Variation in the scaling of the late L of the LHLH pattern One variant of the LHLH pattern found was a rise to the early H, followed by a high plateau that stayed level until a rise to higher late H. An example is given in Figure 3.9.

While Post (2000) treats this as %L H* H* H%, the evidence suggests that it is better treated as a variation of LHLH. The excursion size of the dip varies tremendously from a clear L to a very reduced dip, so it is likely that the elbow here is in fact an undershot L. This argument is supported by the fact that if we examine the difference in f_0 between the early H and late L elbow, we find a continuum of values rather than a dichotomy between the values for two AP types. This continuum is schematized in Figure 3.10.

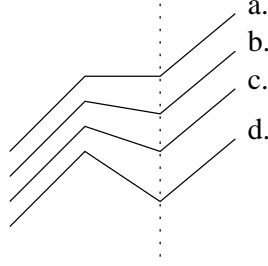


Figure 3.10: *Schematization of the range of excursion sizes for the dip to the f_0 inflection point preceding the H of the late rise. In (d) there is a clear late elbow and a large “dip” from the early H to the late elbow. In (a) the early H and late L have the same f_0 value, and the dip size is 0.*

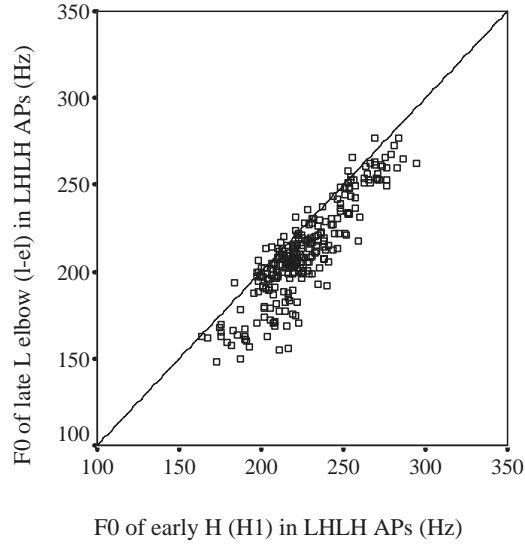


Figure 3.11: *Fundamental frequency of the late L elbow as a function of fundamental frequency of the early H for APs with the LHLH pattern.*

Figure 3.11 charts fundamental frequency of the late L elbow against fundamental frequency of the early H for APs with the LHLH pattern. Note that there is a single cloud of data points, some lying on top of the $x = y$ line, indicating that the

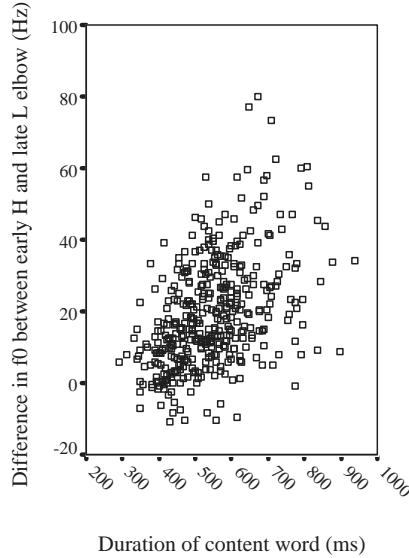


Figure 3.12: *Correlation between size of fall from H of early rise to L of late rise in APs with the LHLH with the duration of the target content words.*

early H and late L elbow have roughly the same f_0 and form a high plateau. For LHLH tokens, the “dip size,” the difference in fundamental frequency between the late L elbow and the early H is correlated with the number of syllables in the AP ($r = .418$), with the duration of the AP ($r = .405$), and with the duration of the target content word ($r = .449$) (all three of which are correlated with one another). The relationship between dip size and duration is plotted in Figure 3.12. As the target word (and the AP) get longer, the dip from the early H to the late elbow gets larger. This pattern is consistent with undershoot in shorter phrases (more precisely, shorter content word parts of APs), where there is time pressure to realize all tonal targets. Jun (1998) makes a similar observation about the scaling of the late L target in non-final Accentual Phrases in Seoul Korean, which, as in French, have a characteristic LHLH pattern. She notes that the L is inversely correlated to the number of syllables

in the AP—as the number of syllables in the AP increases, the L gets lower. Although that specific correlation does not exist for French, the basic observation of undershoot of tones when there is not enough time or when there are not enough syllables is the same.

Further evidence for the analysis of patterns like the one in Figure 3.9 as LHLH comes from the fact that the f_0 of the late L elbow of the LLH AP type is lower (average 193 Hz across subjects) than that of the LHLH (average 209 Hz), which suggests a perseveratory co-articulation in LHLH APs where the late L (elbow) is pulled up by the preceding early H. The results of a Mann-Whitney Test for Two Independent Samples shows that this difference is significant ($U = 19271$, $Z = -5.630$, $p < .001$)).

Plateau realizations of the early H (H1) and late H (H2) of the LHLH pattern In about 31% of the LHLH patterns, the early H target was realized as a high plateau remaining more or less level for a period of at least 40 milliseconds before starting to fall to reach the late L elbow target. Examples are given in Figure 3.13. There is no apparent meaning difference between a plateau realization and a simple peak realization of the early H. The beginning point of a plateau was defined as the H1 tag and the end point was defined as the point immediately before the relatively steep drop to the L of the late H.¹⁶ The average of the absolute values of the differences between the beginning point and the end point of a plateau was 3.1 Hz. In 75% of cases, the beginning and end of the plateau differed by less than

¹⁶I had originally written Praat scripts to automatically calculate plateau end points. This approach turned out to be impractical for two reasons. First, there were many cases with segmentation perturbations of the f_0 in the area of the plateau. Second, I wanted to include “rising plateaux” and “falling plateaux” (described later in this section).

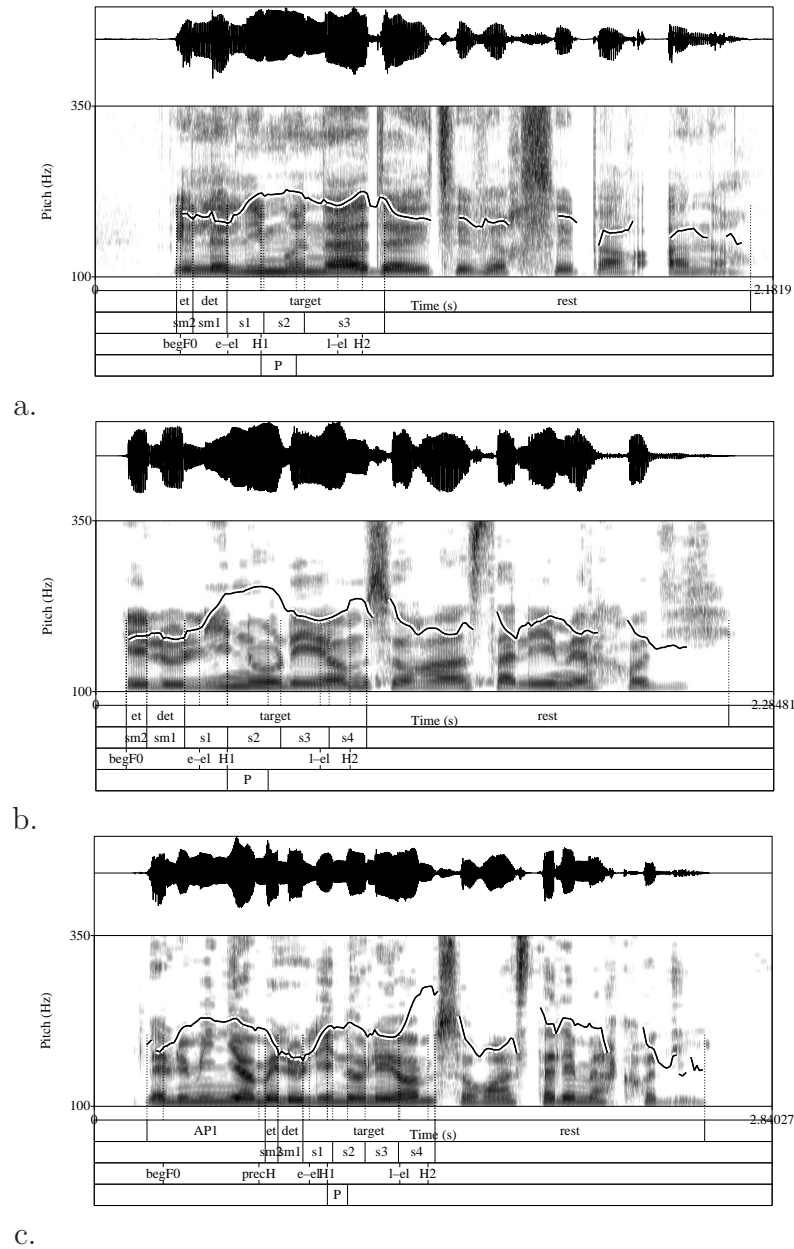


Figure 3.13: *Examples of plateau realizations of the early H (H1). Target AP is underlined. (a) Plateau. Difference between end and beginning: 0.5 Hz. The utterance is Et la limonade a été versée par Anna. ‘And the lemonade was poured by Anna.’ (b) Rising plateau. Difference between end and beginning: 8.7 Hz. The utterance is Et le linoléum sera installé mercredi. ‘And the linoleum will be installed on Wednesday.’ (c) Plateau with perturbation (caused by the [n] of linoléum). Difference between end and beginning: 6.4 Hz. The utterance is L’aluminium et le linoléum seront installés mercredi. ‘The aluminum and the linoleum will be installed on Wednesday.’*

4 Hz. For the remaining 25%, there was a difference of up to 9.8 Hz. In about two thirds of these cases, there were perturbations in the f_0 curve that could account for the f_0 differences. In the other cases there were “rising plateaux” where f_0 reached a preliminary peak, then rose again at a fairly shallow slope before falling to the L of the late rise. “Falling” plateaux were much rarer: in only three cases was the plateau end value more than 4 Hz lower than the plateau beginning. The average length of the early H (H1) plateaux was 92.8 ms, with a maximum of 231.4 ms.

The choice in whether to produce an early H (H1) as a plateau may be connected to the duration of the content word (or words) or the duration of the AP. In the corpus, the average duration of LHLH Accentual Phrases with early H plateaux was longer (duration of content word, 568.2 ms; duration of AP, 716.5 ms) than those without early H plateaux (duration of content word, 534.0 ms; duration of AP, 674.8 ms). It seems plausible that a plateau realization may make an early H tone more salient and thus more effective in highlighting the preceding L tone marking the left edge of a content word. It is also possible that the realization of a target as a simple peak or a plateau may be a matter of speaker choice.¹⁷

Plateau realizations of the H of the late rise (H2) were less common, occurring in only 42 cases, about 11% of LHLH APs. An example is given in Figure 3.14. Unlike the plateau realization of the H of the early rise, at least some plateau realizations of the late H may signal a pragmatic difference. According to informally collected native speaker intuitions, the late H plateaux in many of the items seem to convey continuation or demarcation. This potential difference between the peak and plateau realizations of the late rise merits further study.

¹⁷Actually, the choice is not a binary one—there is not a neat separation between simple peak realizations and plateau realizations.

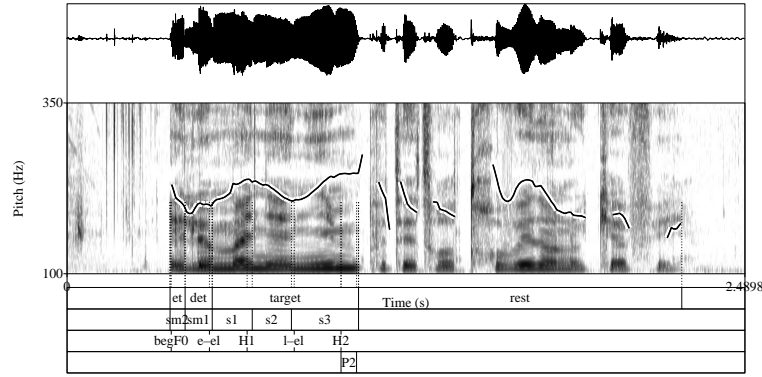


Figure 3.14: *Example of plateau realization of the late H (H2). Target AP is underlined. The utterance is Et le magnanime peut être trouvé dans le café. ‘And the magnanimous one can be found in the café.’*

The duration and tonal scaling of plateau realizations of the late H were examined. The average of the absolute values of the differences between the beginning point and the end point of late H plateaux was 3.1 Hz, with no difference over 6.5 Hz. Late H plateaux rarely extended beyond the end of the Accentual Phrase: only 3 (less than 1%) of the plateaux ended more than 10 ms from the end of the AP. Late H plateaux tended to be shorter than early H plateaux, with an average duration of 73.3 ms and a maximum of 156.4 ms. The timing and duration facts are unsurprising given that the late H marks the right edge of an AP and that the late H is typically realized toward the end of the last syllable of the AP (see §3.4.5). We would not expect the late H plateau to extend into the following AP. The duration of the late H plateau has relatively little room to vary—it begins late in the last (full) syllable of the AP and cannot extend very far past that syllable. In contrast, the position of the early H (the starting point of an early H plateau) is much less fixed. Since the position of

the following L target (the L of the late rise) also can vary a great deal, there is little time pressure to end the plateau quickly in order to reach the low target.

The average duration of LHLH Accentual Phrases with late H plateaux was longer (duration of content word, 559.7 ms; duration of AP, 698.8 ms) than those without late H plateaux (duration of content word, 542.9 ms; duration of AP, 686.1 ms). Given the possible pragmatic difference between the peak realization and the plateau realization of the late H, it may be the case that the AP lengthening follows from the need to realize a plateau.

To summarize, the H of the early rise (H1) is more likely than the H of the late rise (H2) to be realized as plateau. The plateau realization of the H of the late rise (H2) may convey a contrast in meaning. An early H realized as a plateau is likely to be longer in duration than a late H realized as a plateau. Plateau realizations of both the early H and the late H are more likely in longer APs, although this observation must be interpreted with caution with respect to the late H, given potential pragmatic differences between peak and plateau realizations.

LH, LLH

The pattern LH, corresponding to the L(HiL)H* of Jun and Fougeron (2002) is defined as a rise from a L at the beginning of the AP or the content word to a H in the last full syllable of the AP (or very rarely, just beyond the end of this syllable). This pattern was the second most common accent pattern at 21%. This pattern was sometimes accompanied by an early leading low plateau or a slight fall to an early elbow. Examples are given in Figures 3.15a and b.

Accent pattern LLH, corresponding to the L(Hi)LH* of Jun and Fougeron (2002), accounted for an additional 18%. This pattern is characterized by a H on the last

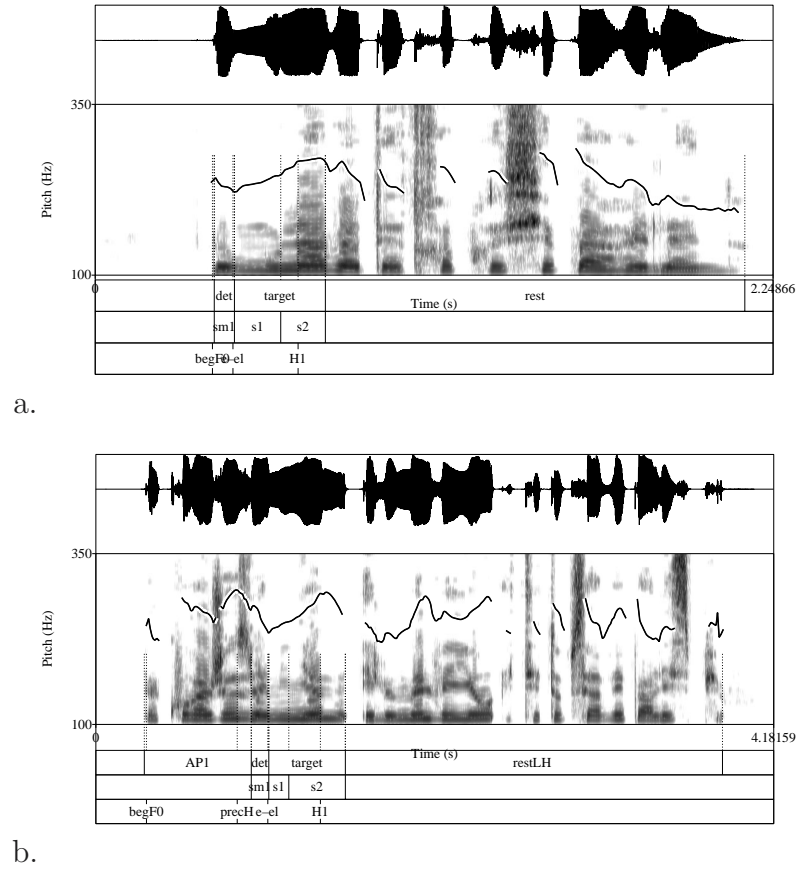


Figure 3.15: (a) *LH pattern with a slight fall to an early elbow in the AP le moulin. The full utterance is Le moulin va être apprécié par Hélène. ‘The mill is going to be appreciated by Hélène.’* (b) *LH pattern with a fall from the late H of the preceding AP to an elbow at the function word-content word boundary in the utterance-medial AP la mignonne. The full utterance is La courageuse, la mignonne et le malveillant étaient observés par l’espion. ‘The courageous woman, the cute woman, and the spiteful man were being watched by the spy.’*

full syllable of the AP (or rarely just beyond it), with a leading low plateau ending at an f_0 elbow in a non-initial content word syllable (i.e., an elbow late in the AP). Examples are given in Figures 3.16a and b.

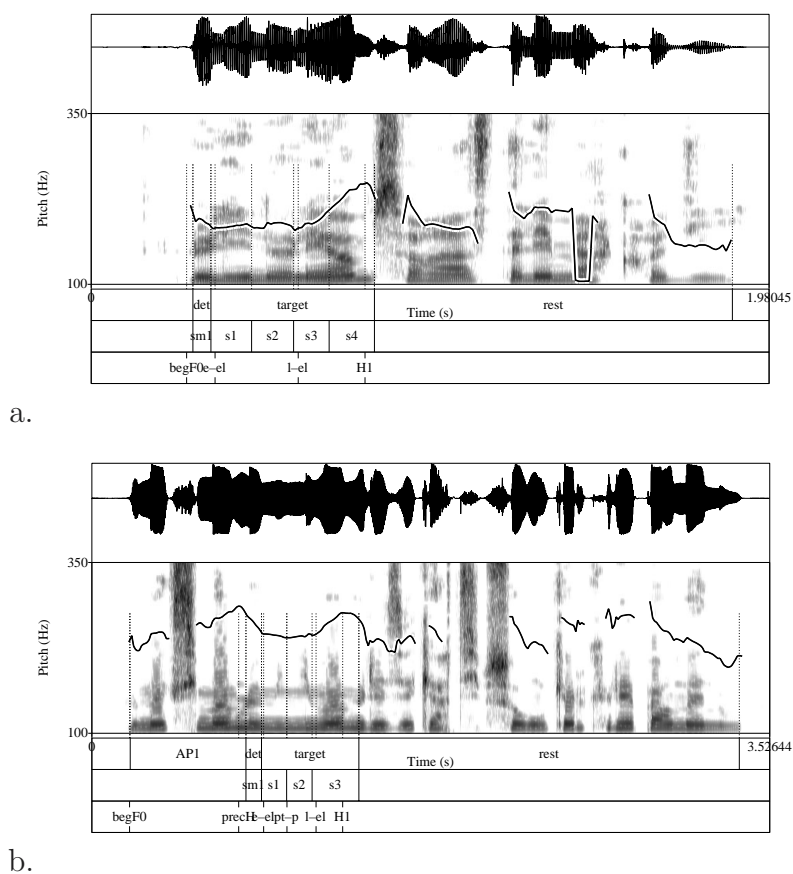


Figure 3.16: (a) *LLH in the utterance-initial AP* le linoléum. The full utterance is Le linoléum sera installé mercredi. ‘The linoleum will be installed Wednesday.’ (b) *LLH in the utterance-medial AP* le minimum. The full utterance is Le maximum, le minimum et les écarts-types seront calculés par Manon. ‘The maximum, the minimum, and the standard deviations will be calculated by Manon.’

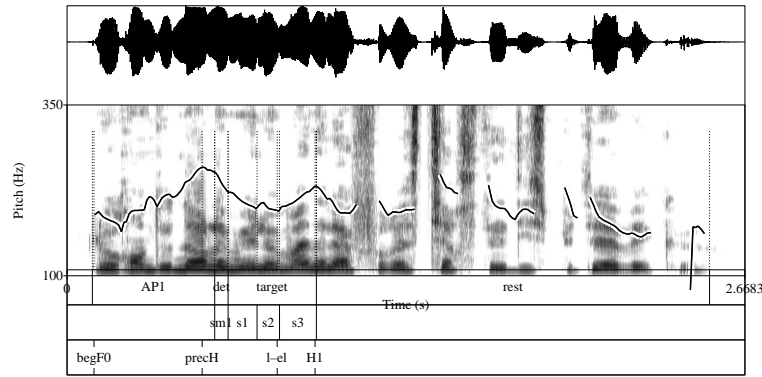


Figure 3.17: *L2H in the utterance-medial AP* le minime. *The full utterance is* Le randonneur, le mélomane et la forestière s'étaient disputés avec eux. 'The hiker, the music lover, and the forester had argued with them.'

Note that a LH pattern preceded by a low plateau and a LLH pattern are distinguished by the timing of their elbows. The bitonal LH pattern with a preceding leading low plateau has an elbow early in the AP, typically near the beginning edge of the content word. In the tri-tonal LLH, there is an obligatory elbow either in the last syllable or the penultimate syllable of the AP. There may also be an early elbow near the beginning edge of the content word, particularly in medial position. Jun and Fougeron analyze the LH pattern as $L(HiL)H^*$, with the Hi of the initial LHi phrase accent and the following L undershot. In LLH, only the Hi is undershot ($L(Hi)LH^*$).

L2H

Jun and Fougeron (2002) predict that the pattern $(LHi)LH^*$, where there is undershoot of both tones of the initial LHi, should not be found: "... though the realization of L or Hi is optional, one of these must be realized to make the AP initial boundary" (p. 151). Though rare, this pattern is attested in the data, accounting for 14 of 790 cases, or 2%. Figure 3.17 shows one such case. The critical AP is in medial position,

and there is a fall from the late H of the preceding AP to a L late in the critical AP, followed by a rise to the late H. This accent pattern is labeled L2H to distinguish it from the LH pattern. In the L2H pattern, there is a rise from the late L or L2. The LH (the L(LHi)H* of (Jun and Fougeron (2002))) and the L2H ((LHi)LH*) patterns differ in that in the more common LH pattern has a rise from a L near the beginning of the AP, while the rare L2H pattern has a rise from a L late in the AP. Note that the L of the L2H in Figure 3.17 is 193 ms from the beginning of the content word.

Of the 14 L2H APs found in the data, 12 are found in fast rate utterances. In addition, eight of the 14 L2H APs were produced by one speaker, Speaker 1, who had the fastest fast speaking rate (along with Speaker 6) and the highest percentage change from the normal to the fast speaking rate. These facts suggest that speakers produce this rare pattern under time pressure.

LHH

The pattern LHH, corresponding to Jun and Fougeron's LHi(L)H* was also fairly uncommon, occurring in only 4.5% of cases. The pattern is defined by a high plateau that stretches from one of the beginning two syllables to the final full syllable, with no intervening late L elbow. An example is given in Figure 3.18. It was sometimes difficult to determine whether the accent pattern of an AP should be categorized as a LHH pattern or as a LHLH with an undershot dip between the two H tones. In difficult cases, I assumed that there was no L tone and labeled the AP LHH. Even with this generous labeling heuristic, the pattern is very rare.

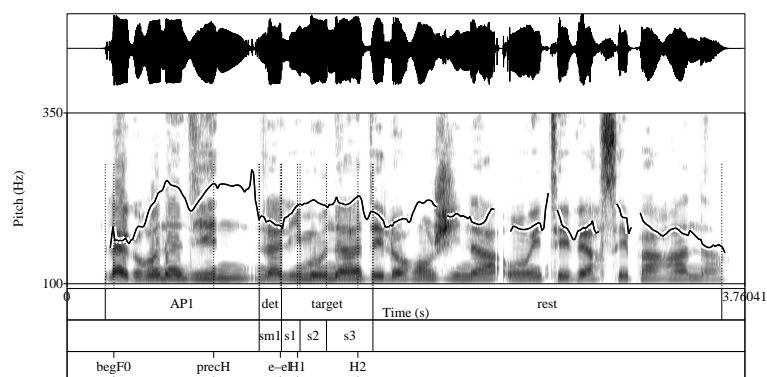


Figure 3.18: *LHH in the utterance-medial AP la limonade. The full utterance is La grenadine, la limonade et l’Orangina ont été versés par Anna. ‘The grenadine, the lemonade, and the Orangina were poured by Anna.’*

LHi

Another rare pattern was the LHi pattern, in which there is only an early rise, but no late rise. This pattern occurred in only 4% of critical APs. It is defined by a rise from a L at the beginning of the content word to a H on one of the first two content word syllables. The timing facts make it clear that this is a bitonal pattern, not a LHL pattern. While there is a fall from the early (only) H of the AP, the fall ends at a L in the following AP. This L is often right at the function word-content word boundary, forming a canonical early elbow. An example of the LHi AP type is given in Figure 3.19.

In one respect, this pattern resembles the focal accent described by Jun and Fougeron (2000)—there is rise near the beginning of the phrase, but none at the end. Focus intonation in French, however, is characterized not only by this particular

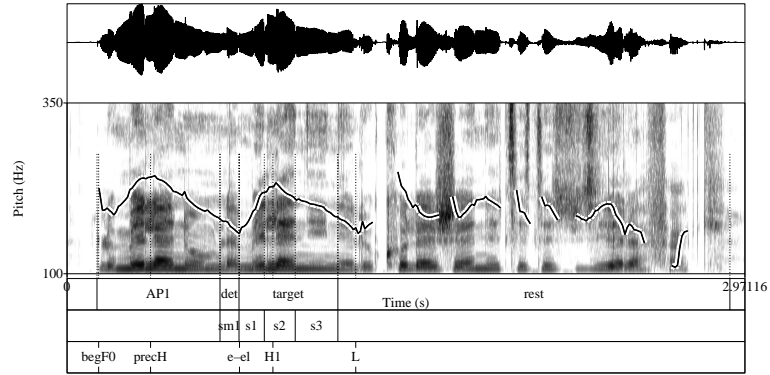


Figure 3.19: *LHi in the utterance-medial AP la mélanine. The full utterance is Le mélanome, la mélanine et le collagène étaient étudiés à la fac. ‘Melanoma, melanin, and collagen were studied at college.’*

intonation pattern on the focused element, but also by a low plateau on following material (see Jun and Fougeron (2000) for a description). This pattern is illustrated in Figure 3.20b, which is adapted from Dohen et al. (2003a). In the current experiment, there is no following low plateau for any of the items labeled LHi.

It is also clear that these LHi are not the first part of a longer LHLH AP. The last syllable of these APs has an average duration of 213.1 ms, while the first syllable and the second (non-final) syllable are much shorter (124.9 ms and 123.6 ms, respectively). This is consistent with lengthening of AP-final syllables. As illustrated in Figure 3.19, the tonal alignment patterns of a sequence of two LHi APs are incompatible with those expected for a LHLH AP. The L of the second LHi is consistently aligned with the function word-content word boundary, and the Hi is realized on a non-AP-final syllable. Finally, the tonal scaling facts support the LHi account. In LHLH patterns, the L of the late rise typically has a higher f_0 value than that of the L of the early rise

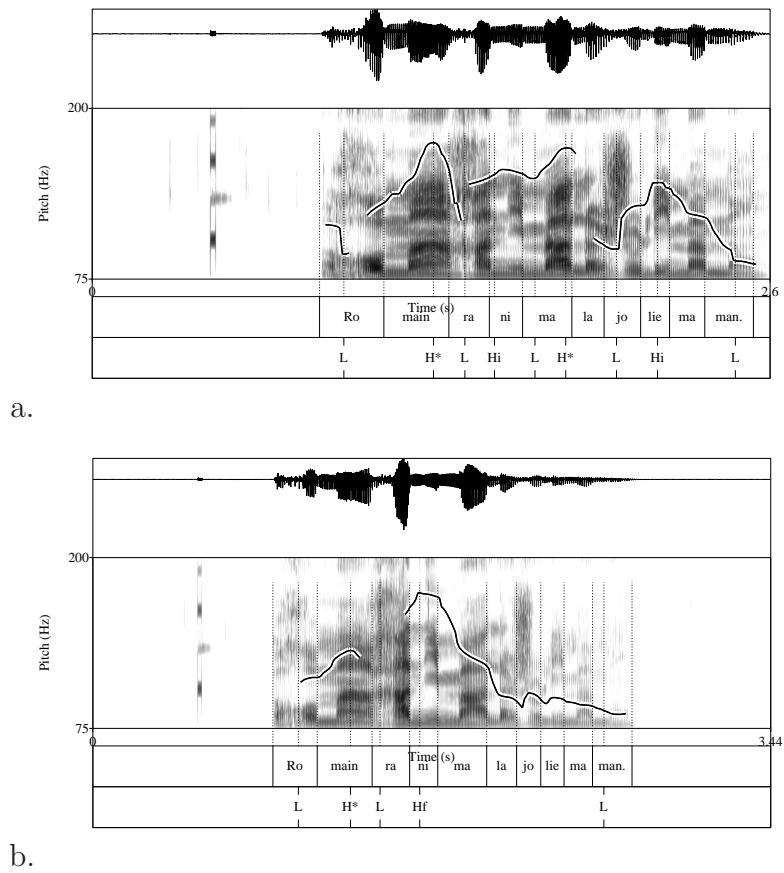


Figure 3.20: *The sentence Romain ranima la jolie maman. ‘Romain revived the pretty mom.’ (a) with broad sentence focus and (b) with narrow focus on the verb ranima. In the narrow focus condition, there is a low plateau following the focused verb. Figure adapted from Dohen et al. (2003a) and used with permission.*

(see Figures 3.6, 3.7, and 3.13 and the discussion of perseveratory coarticulation of the early H and late L in §3.4.2). In a sequence of two LHi APs, however, the L of the second LHi has a lower f_0 value than the L of the preceding LHi (as in Figure 3.19).

Two speakers accounted for all 28 of the productions of the LHi AP type in the critical APs (Speaker 5 produced five and Speaker 7 produced 23). Further examination of all APs (critical and non-critical APs) in all 790 utterances showed that the other five speakers never produced an AP with a LHi pattern.

The LHi AP type also had an interesting distribution. Thus far in the discussion we have referred to utterance-initial position and utterance-medial position. Utterance-medial position includes APs that are not the first AP in the utterance and not the last AP in the utterance. This distinction, however, is not adequate for describing the distribution of the LHi AP type. Again considering both critical and non-critical APs, the LHi AP type was only found in APs that did not immediately precede the verb phrase. Positions where the LHi pattern was found are indicated in (21) by underlining. APs that immediately preceded the verb phrase always contained a late rise.

- (21) a. [La mélanine] [était étudiée à la fac.]
 the melanin was studied at the college
 ‘The melanin was studied at college.’
- b. [Le mélanome, la mélanine et la collagène] [étaient étudiés à la
 the melanoma the melanin and the collagen were studied at the
 fac.]
 college
 ‘The melanoma, the melanin, and the collagen were studied at college.’
- c. [Le mélanome et la mélanine] [étaient étudiés à la fac.]
 the melanoma and the melanin were studied at the college
 ‘The melanoma and the melanin were studied at college.’

There are at least two possible explanations for this distribution. One is that there is a prosodic boundary higher than an Accentual Phrase boundary that coincides with the syntactic boundary between the noun phrase subject and the verb phrase. Di Cristo suggests that such a boundary (an Intonation Phrase boundary) is common in this position. The H of the late H in APs at this boundary (APs immediately preceding the verb) may be a required boundary tone marking the right edge of this unit. Another is that a LHi pattern is not pragmatically appropriate in this position, on the last element of a compound subject. Use of the LHi pattern creates the list intonation (“intonation énumérative”) that Fonagy (1979) describes finding in a pilot study: “[t]ous les mots avaient été accentués d’une façon uniforme, soit sur la dernière, soit sur la première syllabe” ‘all the words had been accented in the same way, either on the last or on the first syllable’ (p. 140). This list impression is particularly strong when the LHi pattern is repeated across more than one AP. In fact, the two speakers tended to produce two LHi patterns in a row on the first and second APs of a three-conjunct compound subject (as in Figure 3.19). It was rare to find cases where there was a LHi pattern on the second AP and a different pattern on the preceding AP. It seems likely that the LHi pattern suggests that an element is non-final, constraining the distribution of the pattern.

3.4.3 Occurrence of the LHLH pattern

A model in which each tone is associated to a separate syllable predicts that the 4-tone pattern will be found only in an Accentual Phrase of four or more syllables. All of the critical APs consisted of a noun preceded by one or more function words. The Jun & Fougeron (2000) model restricts the realization of the early H to content word

syllables (except in exceptional circumstances). For our data, the model therefore predicts that only critical APs with content words of three or more syllables should be realized with the LHLH pattern. However, of the 314 two-syllable words in the corpus, 41 were produced with the four-tone LHLH pattern (13%).

Models which require every tone to be associated to a syllable also cannot account for other patterns found in our data. The four-tone LHLH pattern was found for 83% (129 of 155) of APs with content words of 4 syllables in length, while the pattern was found for only 70% (226 of 321) of APs with content words of 3 syllables in length. Recall that all target content words in the corpus are preceded by at least one function word. This difference is not predicted since in both cases there are four syllables to which the four tones of the pattern can associate. We also find that the number of LHLH patterns produced varies with number of function words in the AP: APs with two function words are more likely to have the four tone pattern than APs with only one function word. This is true even though the corpus contains more APs with one preceding function word (413) than APs with two preceding function words (377). (Recall that 37 utterances with two preceding function words in the critical APs were excluded from the analysis because they had f_0 rises on the conjunction *et*.) The chart in Figure 3.21 shows this difference, which is also not predicted by a syllabic association model.

These distribution facts suggest that it is perhaps not number of syllables that explains the realization of a four-tone LHLH pattern, but rather the duration of the AP. That is, a four-tone pattern can be realized only when the speaker has enough time to do so, and APs with two function words are longer than APs with only one function word, *ceteris paribus* (in this corpus, average durations are 558 ms for APs

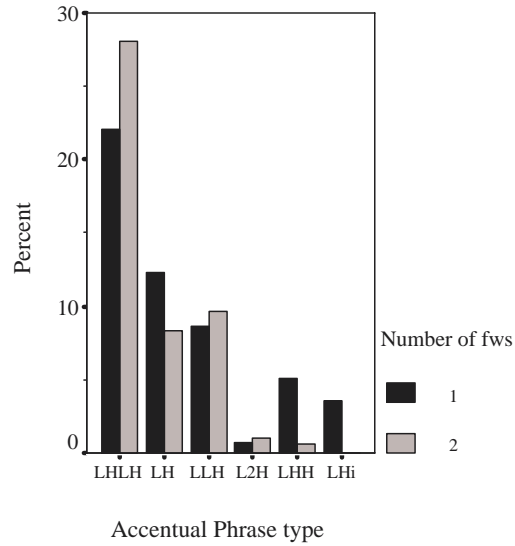


Figure 3.21: *AP types found by number of function words in AP as percentages of the total.*

with one function word and 648 ms for APs with two function words). A number of binary logistic regression analyses were performed to examine the factors influencing the realization of the LHLH pattern in the critical APs.

Logistic regression is a type of regression analysis used to examine the influence of independent or predictor variables on a categorical dependent variable (a dependent variable with different levels or categories). The predictor variables can be categorical, continuous, or a combination of categorical and continuous. In the usual case, the dependent variable is binary and is coded as a dummy variable with two levels (e.g., 1 and 2, or YES and NO). The goal of logistic regression is to provide a model that predicts the outcome category for individual cases. For example, we want to build a model that can predict whether or not a given Accentual Phrase will have the LHLH pattern.

	B	S.E.	B/S.E.
number of content word syllables	2.019	.145	13.92
constant	-5.584	.407	-13.72

Model chi-square: 292.0

Percentage correct: 79.5

$p < .001$

Table 3.1: *Logistic regression results for independent variable* number of content word syllables.

Analyses were performed separately for each of the seven subjects and by pooling all data for all subjects. The dependent variable was *LHLH status of AP pattern*, with values of 1 (is LHLH AP) and 0 (is any other AP type). Many of the AP types were relatively uncommon and are therefore represented by relatively few tokens. The decision was therefore made to proceed with a binary logistic regression by collapsing all non-LHLH AP types into a single category, rather than to perform a multinomial regression analysis with all AP types as levels of the dependent variable. The first regression analysis used number of syllables in the content word as the independent variable. The results are summarized in Table 3.1.

Number of syllables is an excellent predictor of whether or not an Accentual Phrase will be produced with the 4-tone LHLH pattern. The model is significant at the .001 level according to the model chi-square statistic. The model predicts 79.5% of the LHLH accent patterns correctly. The results of the second analysis, which used duration of the AP as the independent variable, are summarized in Table 3.2. Duration of the AP was also an excellent predictor. The model is significant at the .001 level and correctly predicts 76.2% of the LHLH patterns.

	B	S.E.	B/S.E.
duration of AP	.012	.012	12.00
constant	-7.125	.531	-13.42

Model chi-square: 323.013

Percentage correct: 76.2

$p < .001$

Table 3.2: *Logistic regression results for independent variable duration of AP.*

Although the percentage correct number for the model with number of syllables as a predictor of occurrence of the LHLH pattern is slightly higher than that of the model which uses AP duration as the independent variable, this is not evidence for models which require every tone to be associated to a separate syllable. First, as noted above, those models do not predict that there should be a steady increase in realization of LHLH patterns as number of syllables increase, simply that LHLH should only appear in APs that are at least 4 syllables long and contain at least 3 content word syllables. Secondly, logistic regressions were done for each of the seven subjects individually, and for some subjects, the percentage correct number was slightly higher for the model with AP duration as the independent variable. I therefore conclude that duration of the AP is as good a predictor of whether a LHLH patterns will be realized as is the number of syllables in the AP.¹⁸ It is common in the literature to note that LHLH patterns (or phrases with both primary and secondary accents) are more prevalent in longer APs, but length is typically expressed in number of syllables (e.g., Jun and Fougeron (2000), Pasdeloup (1990)). To my knowledge, this result is the first evidence

¹⁸Note that it is not possible to enter AP duration and number of content word syllables as independent variables in the same analysis, since they are highly correlated. Inclusion of both would engender a multicollinearity problem.

that duration (expressed in clock time) is a predictor of the realization of the LHLH pattern.

Stepwise logistic regression analyses were conducted to determine whether adding more independent variables would significantly improve either model. In this technique, a number of variables are entered into the model, and only the variable or set of variables that provides the model with the best fit is retained. It is common to set aside a portion of the data in order to test the model chosen by the training data. This is accomplished by running a second stepwise regression analysis on the test data and verifying that variables chosen by the original regression analysis are retained.

For our examination of the factors predicting the realization of the LHLH pattern, the variables added were: *number of function words (1, 2)*, *position (initial, medial)*, and *rate (normal, fast)*. It has been observed that Accentual Phrases in utterance-initial position rarely lack a secondary accent or early rise.¹⁹ It has also been noted that secondary accents are less common in fast speech and more common in slower, deliberate speech.

The addition of these additional variables has only a small impact on the goodness of the model. A model with the independent variables *number of content word syllables*, *number of function words (1, 2)*, *position (initial, medial)*, and *rate (normal, fast)* actually performs slightly worse than the model with only *number of content word syllables* (percentage correct is 78.6). A model with the independent variables *AP duration*, *number of function words (1, 2)*, *position (initial, medial)*, and *rate (normal, fast)* performs only marginally better than a model with only *AP duration*,

¹⁹An exception is the speaker from Tours in Welby (2002), who often produced LLH patterns utterance-initially.

with a percent correct of 78.7, with the contribution of *number of function words* being insignificant.

3.4.4 Accent patterns not found

Having discussed the accent patterns found in the data, it is worthwhile to discuss briefly the accent patterns *not* found. I do not, for example, find examples of Jun and Fougeron's accent patterns (L)HiLH* and LHiL* (illustrated in Figure 3.4 e and f). This gap is likely to be an artefact of the restricted nature of the critical APs in the corpus.

The lack of APs with the (L)HiLH* pattern may be due to the composition of the critical APs, all of which started with one or more function words, followed by a content word. It may be more common for an AP starting with a content word to have a (L)HiLH* pattern. The LHiL* pattern is the least common pattern found by Jun and Fougeron, who observe it only when the pitch accent is immediately preceded and immediately followed by an early Hi. Jun and Fougeron (2002) argue that the realization of a L* rather than a H* avoids a sequence of three consecutive H tones (*[HiH*Hi]). Given this distributional constraint, a LHiL* AP is expected only immediately before a (L)HiLH* AP. The absence of LHiL* APs in my data is therefore expected, since there are no (L)HiLH* APs. So the accent patterns found in the current data are not an exhaustive list of possible accent patterns in the language.

3.4.5 Tonal alignment of the early and late rises

Early L elbow

The timing of the early L elbow was examined for all APs which contained the elbow. Early L elbows were possible for the LHLH, LH, LLH, LHH, and LHi patterns.

Note that the early L was not always realized with an initial low leading plateau and elbow. The elbow is possible in cases in which there is a leading low plateau or a fall from the late H of a preceding AP. In medial position there is a fall from the H of the preceding AP to a L; in initial position, there can be a slight fall from the beginning f_0 . The 660 tokens produced with an early L elbow were included in the analysis.

Welby (2002) found that the early elbow was consistently realized at the function word-content word boundary. I argued that the L tone of the early elbow is not associated to the last function word syllable, as claimed by Vaissière (1997, 2002). I also argued against the Jun and Fougeron (2000) account of the low leading plateaux as a spreading of the early L to the last function word syllable. Rather the pattern of results supported a model in which the early L is doubly associated to the left edge of the AP and to the left edge of the first content word (see Chapter 2, §2.3.6). This double association must be optional since a leading low plateau and resulting low elbow are not always found. The association to the content word boundary (rather than that of the function word) provides a unified account of phrases with and without preceding function words. Materials in the Welby (2002) study included both prosodic phrases which began with content words and those that began with function words.

The results of the current experiment lend support to the conclusions of Welby (2002) and to the Early L Associated Tone Hypothesis. As in the earlier experiment, the early elbow straddles the function word-content word boundary, as illustrated in Figure 3.22. In the graph, data points below the $x = y$ line represent cases in which the early L elbow was realized within the last function word syllable (always the definite article), and data points above the line represent cases in which the elbow

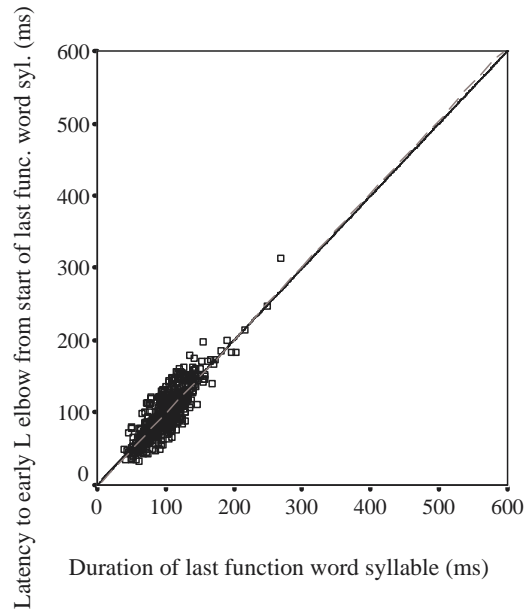
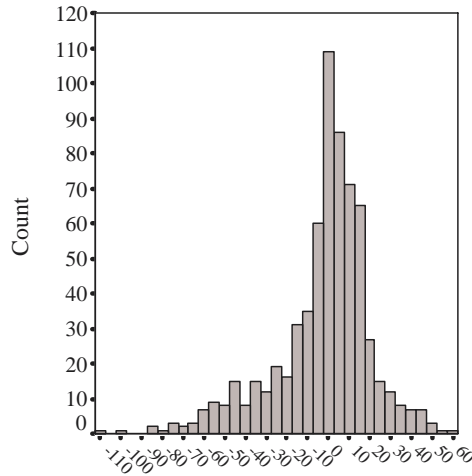


Figure 3.22: *Location of the early L elbow as a function of duration of the last function word syllable (hypothesized associated syllable). The dashed line is the regression line.*

was realized in the first content word syllable. Note that the regression line lies almost directly on top of the $x = y$ line. The fact that many data points fall on top of one another (since they fall very close to the $x = y$ line) obscures the distribution of the early elbow data points. The histogram in Figure 3.23 sheds light on this distribution. Twenty-nine percent of the early elbows are realized within 5 ms of the function word-content word boundary; 47% within 10 ms; 63% within 15 ms; and 73% within 20 ms.

Another pattern in the data leads me to revise the formulation of the optional double association of the early L to the left edge of the AP and the left edge of the first content word proposed in Welby (2002) and in the Early L Associated Tone Hypothesis. In some cases of medial APs, the early L plateau does not stretch all the way to the left boundary of the AP, but stops at another syllable boundary. In



Latency to early L elbow from function word-content word boundary (ms)

Figure 3.23: *Distribution of the early L elbow. Note the peak at 0 ms. A difference of 0 ms indicates that the elbow was realized precisely at the function word-content word boundary.*

some of these cases, there may not be enough time to fall from the preceding H to a L target at the left edge of the AP. In others, it is not clear why the plateau stops before reaching the edge of the AP. In Figure 3.24, for example, there is an early elbow (*e-el*) at the boundary between the determiner *le* and the first syllable of the content word *minimum* and a second elbow (*e2-el*) at the boundary between the conjunction *et* and the determiner *le*. To explain such cases, I revise my account of the association of the L of the early rise:

DOUBLE ASSOCIATION OF L OF EARLY RISE: The L of the early rise is associated to the beginning edge of the first content word of the Accentual Phrase and optionally to the beginning edge of the Accentual Phrase. When this second association is unachievable due to physical constraints on the realization of sequences of H and L tones or for other reasons, it defaults to the syllable boundary nearest the beginning edge of the Accentual Phrase.

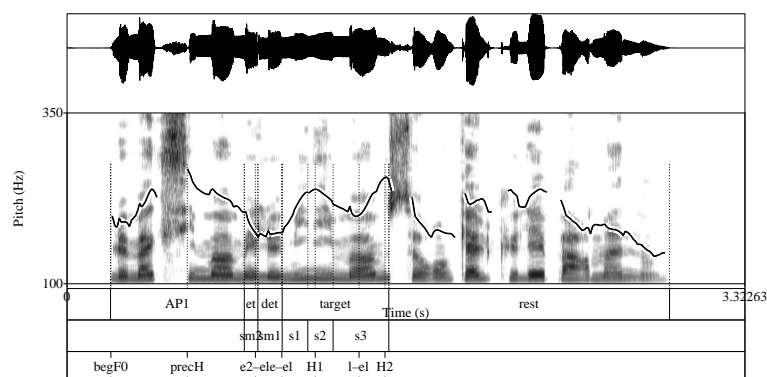


Figure 3.24: *LHLH pattern with two early elbows, one at the boundary between the conjunction et and the le (e2-el), the other at the boundary between the determiner and the first content word syllable (e-el). The utterance is Le maximum et le minimum seront calculés par Manon. ‘The maximum and the minimum will be calculated by Manon.’*

I propose that the L of the early rise is *edge-seeking*—it seeks to associate to the beginning edge of the first content word syllable of the Accentual Phrase and also to the beginning edge of another unit, preferably that of the Accentual Phrase.

A stepwise multiple regression analysis was conducted to see what variables contributed to the alignment of the early L elbow. The dependent variable was latency from the left edge of the last function word syllable. The independent variables were *duration of the first content word syllable*, *number of preceding function words (1 or 2)*, *position (initial or medial)*, and *rate (normal or fast)*. The model that accounted for the greatest amount of variability in the training data included the following independent variables: *duration of last function word syllable*, *position*, and *number of function words*. This model accounted for 60% of the variation. The vast majority of that variation (58%), however, is accounted for by a model with *duration of last*

constant	coeff 1	coeff 2	adjusted R^2
-3.192	1.016 (duration last fw)	9.279 (position)	59%

Table 3.3: *Coefficients for the early L elbow regression model.*

function word syllable as the only independent variable. Testing the model with the test data fails to confirm that the variable *number of function words* makes a significant contribution in predicting the variation ($p = .651$), but the variable *position* is marginally significant ($p = .056$). Excluding *number of function words* from the model leaves us with a model that accounts for 59% of the variance. The coefficients for the model are given in Table 3.3.

The coefficient for the binary dummy variable *position* should be interpreted to mean that early L elbows for APs in initial position appeared on average 9.279 ms later than those for APs in medial position (the reference category).

It is somewhat surprising that the position of the early L elbow was not affected by speaking rate. I had suspected that elbows in the fast rate might be less precisely placed than elbows in the normal rate. In fact, the distribution of data points for the two rates is nearly identical. The two sets of data points regress to very close lines ($y = 1.056x - 5.4751$, $R^2 = 0.59$ for the normal rate, and $y = 1.0564x - 1.2291$, $R^2 = 0.56$ for the fast rate).²⁰

H of Early Rise

The H of the early rise was produced in the first content word syllable 46% of the time and in the second content word syllable 54% of the time. This variety is seen in [Figure 3.25](#), which plots the peak delay from the left edge of the first content word

²⁰ R^2 values are reported in tables as percentages.

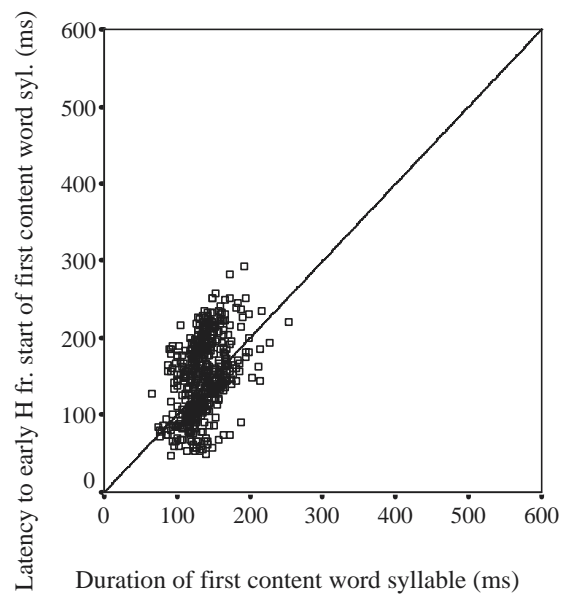


Figure 3.25: *Location of the early H relative to the Accentual Phrase beginning as a function of the duration of the first content word syllable.*

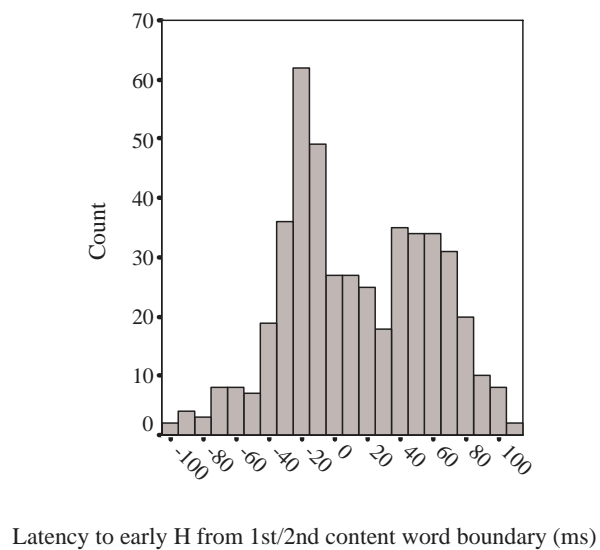


Figure 3.26: *Distribution of the early H.*

model	constant	coeff 1	coeff 2	adjusted R^2
<i>A</i>	40.658	.744 (s1 duration)	17.078 (no. fws)	19%
<i>B</i>	98.436	.306 (AP duration)	-29.766 (no. fws)	53%
<i>C</i>	50.073	.174 (CW duration)	18.147 (no. fws)	20%

Table 3.4: *Coefficients for the three early H regression models.*

syllable (the hypothesized associated syllable) against the duration of that syllable. Data points that fall below the $x = y$ line represent early H tones that were realized within the first content word syllable, and points above the line represent early H tones realized in the second content word syllable. The histogram in Figure 3.26 shows the distribution of the early H on either side of the boundary between the first and second content word syllable. Note that there is no single peak (as in the histogram for the early elbow in Figure 3.23). Unlike the early elbow, the position of the early H cannot be specified with respect to a syllable boundary.

This variety in the realization of the early H is in line with reports in the literature. The position of the peak of the early rise has been reported to be less predictable than that of the late rise, typically appearing in the first, second or sometimes even third syllable of the AP. The seven speakers in the current experiment realized the early H in the first content word syllable of the AP more often than the two speakers in Welby (2002), who realized the early H in the first syllable only about 25% of the time. This difference may be due to the fact that the critical APs in the current experiment all contain one or more function word, while the earlier experiment included critical APs with 0, 1 or 2 function words.

Stepwise multiple regression analyses were performed to attempt to find which variables most accurately predict the text-to-tune alignment of the early H. For all

analyses, a cross-validation technique was employed: most of the data was used as training data to find a model which best fit the data, and a certain percentage of the data was held back to be used as test data, to verify the independent variables chosen by the model. The timing of the early H was examined for all AP types with an early H: LHLH, LHH, and LHi (a total of 469 tokens). The placement of the early H was plotted for each subject, and since no timing differences were apparent, data for all subjects was pooled for the regression analyses.

Three sets of analyses were carried out, each with a different dependent variable. Table 3.4 shows the regression coefficients for each of the models. The first used latency from the left edge of the first content word syllable to the early H as the dependent variable and the following independent variables: *duration of the first content word syllable*, *number of preceding function words (1 or 2)*, *position (initial or medial)*, and *rate (normal or fast)*. The model chosen using these parameters and the training data included the independent variables *duration of first content word syllable* and *number of function words*. The R^2 for this model (*Model A*) was only .193 or 19%, indicating that the model accounted for a scant 19% of the variance in peak placement.

The poor performance of this model is evidence against the Early H Associated Tone Hypothesis, which predicts that the early H will appear at a fixed distance from a syllable landmark of the hypothesized associated syllable (here, the first content word syllable). It is apparent from looking at the graph in Figure 3.26 that using latency from the right edge of the first content word syllable as the dependent variable would produce even poorer results.

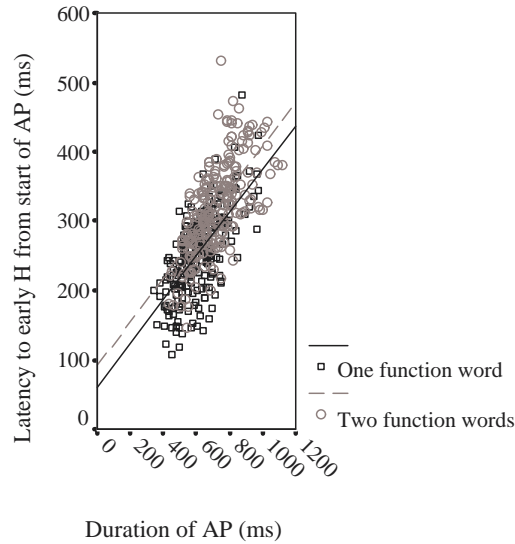


Figure 3.27: *Location of the early H relative to the Accentual Phrase beginning as a function of AP duration and number of preceding function words. The solid line is the regression line for the one function word condition and the dashed line is the regression line for the two function word condition.*

Since Jun and Fougeron (2002) have hypothesized that the early H is part of an AP-initial phrase tone, it is possible that the position of this tone depends on the length of the AP. The correlation between these two factors is clearly illustrated in Figure 3.27.

The second regression analysis therefore used latency from the left edge of the Accentual Phrase to the early H as the dependent variable and the following independent variables: *duration of the Accentual Phrase*, *number of function words*, *position*, and *rate*. The model chosen retained both *duration of AP* and *number of function words*. The R^2 for this model was 53%. Most of this variation is accounted for by *duration of the AP*—addition of *number of function words* improves the model by only 4%. Since *number of function words* is a dichotomous variable for this data, its coefficient must

be interpreted with respect to the reference category, which is *two function words*. Given the coefficient -29.766 of Model B, this means peak placement is on average 29.766 ms earlier for APs with only one function word. This earlier placement can be seen on the chart in Figure 3.27, where the solid regression line for the one function word condition is below the dashed regression line for the two function word condition. The model is validated by the test data, where the two independent variables are significant and account for 60% of the variance.

A third regression analysis examined the possibility that duration of the content word was a better predictor of early H placement than duration of the Accentual Phrase. The regression used latency from the left edge of the content word to the early H as the dependent variable and the following independent variables: *duration of the content word*, *number of function words*, *position*, and *rate*. A model with the independent variables *content word duration* and *number of function words* was selected, but this model (Model C) accounted for only 20% of the variance.

Timing of early H and preceding L with respect to each other

If the Early H Trailing Tone Hypothesis is correct, and the early H is a trailing tone of the preceding L, it should follow the L at a fairly fixed distance. To test this hypothesis, the distance between an early H and the preceding L (early L elbow or initial L target in cases where there was no elbow) was extracted for all AP types with early rises (LHLH, LHH, and LHi). The results are shown in Figure 3.28. It is clear from the chart that there is no invariant rise time at either rate—rise time varies from under 50 ms to nearly 300 ms. The Early H Trailing Tone hypothesis is therefore rejected.

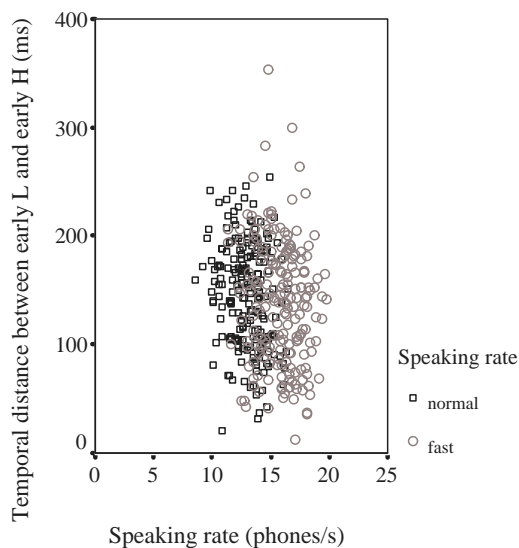


Figure 3.28: *Location of the early H relative to the Accentual Phrase beginning as a function of speech rate.*

Slope of the early rise

We have so far rejected the Early H Associated Tone Hypothesis (the timing of the early H cannot be predicted by reference to a syllable landmark) and the Early H Trailing Tone Hypothesis. It is possible that the variation in the data can be accounted for outside an Autosegmental Metrical model. It may be that the tones are not timed and scaled separately, that the speaker rather attempts to produce a rise with a fairly consistent slope. To examine this hypothesis, the slope of the early rise was examined for each token for each AP type with an early rise. Slopes were calculated by the formula $m = (y_2 - y_1) / (x_2 - x_1)$, where y_1 is the f_0 value of the L of the early rise (the low starting point), y_2 is the f_0 value of the H of the early rise (the peak), x_1 is the time of the L, and x_2 is the time of the H. The results, which are plotted in Figure 3.29, showed a great deal of scatter in the data. A regression

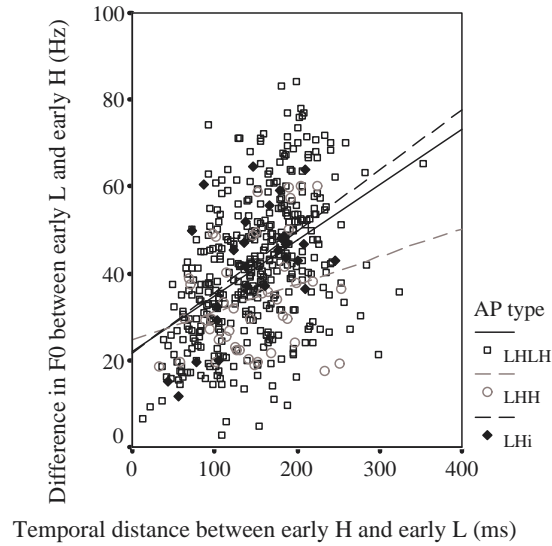


Figure 3.29: *Slope of the early rise by accent type.*

analysis showed that only 18% of the variation in the f_0 difference from the early L to the early H was explained by the distance between the two tones. The hypothesis that speakers attempt to produce a rise with a specific slope is therefore rejected.

H of late rise

Researchers have observed that the f_0 peak of the late rise typically appears in the last full syllable of the Accentual Phrase or, rarely, just beyond it. The timing of the late H was examined for the following AP types: LHLH, LH, LLH, and L2H. In the labeling, the late H tags and the early H tags were systematically placed at either the f_0 peak or at the beginning of a plateau, if one was present. Plateaux were more rare for the late H, but they did occur. The other AP type with a late H, LHH, was excluded from this analysis. In the LHH pattern, there is a plateau extending from the early H to the late H. The late H tag for the LHH is therefore

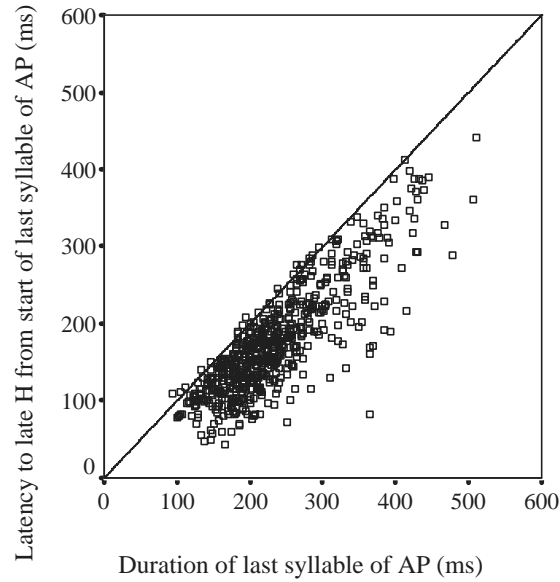


Figure 3.30: *Location of the late H as a function of duration of the last syllable of the AP (for AP types LHLH, LH, LLH, and L2H).*

placed at the end of the plateau and its timing is not comparable to that of the late H for the other AP types. A total of 45 LHH utterances were present in the data and were excluded. Two utterances with the LLH were excluded from the timing analysis because of perturbations in the critical region of the elbow. These exclusions left a total of 715 cases for the analysis. Of these 715 cases, only 22 or 3% were realized past the offset of the last syllable of the Accentual Phrase.

Figure 3.30 shows the location of the late H as a function of duration of the last syllable of the AP, which is hypothesized to be the associated syllable in the Late H Associated Tone Hypothesis and in accounts such as Jun and Fougeron (2000), Jun and Fougeron (2002), and Post (2000), which treat this tone as a pitch accent. Note that most of the data points fall close to the end of the hypothesized associated

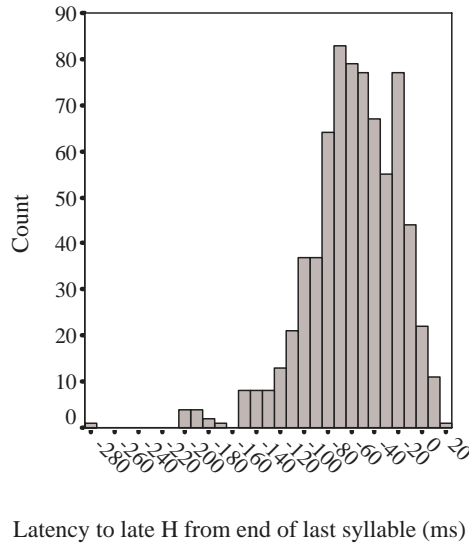


Figure 3.31: *Distribution of the late H.*

syllable (indicated by the $x = y$ line), and that the 3% of the late H data points that do not fall within the last syllable fall just beyond it. The small number of late H tokens falling beyond the last syllable is also evident in the histogram in Figure 3.31. This distribution provides support for the Late H Associated Tone Hypothesis.

Note that the alignment of the H of the late rise, which is associated to a syllable, is very different from that of the L of the early rise, which is associated to an edge. The L of the early rise is realized either right at the function word/content word boundary, just before the boundary or just after the boundary. This distribution is shown in Figure 3.22, where many data points fall right on the $x = y$ line and roughly equal numbers fall on either side of the line, and in Figure 3.23, where there is a peak at 0 ms. In contrast, the H of the late rise is realized toward the end of the last full syllable of the Accentual Phrase, but rarely just beyond the syllable boundary. This

is shown in Figure 3.30, where most data points fall below the $x = y$ line, and in Figure 3.31, where the peaks are located at negative latencies from the end of the associated syllable. In addition, researchers routinely note that the H of the late rise is not realized on AP-final schwa syllables. This restriction provides further support for an account in which this H is associated to a syllable rather than a phrase edge.

An examination of the location of the late H revealed a potential difference across different AP types. Late H location is graphed separately for each AP type in Figure 3.32. The late H data points in the LHLH AP type ((a) in the figure) seem to cluster closer to the syllable end than do the data points for the other AP types.

Stepwise multiple regression analyses were performed to attempt to find which variables most accurately predict the text-to-tune alignment of the late H. As with the regression analyses for the tones of the early rise, a portion of the data was set aside to be used as test data for cross-validation. The location of the late H was plotted for each subject, and since no timing differences were apparent, data for all subjects was pooled for the regression analyses. *Latency from the beginning edge of the last syllable of the AP* was the dependent variable, and the following were entered as independent variables: *duration of the last syllable of the AP*, *rate (normal, fast)*, *position (initial, medial)*, *IS-LHLH (yes, no)*. The last three are all binary-coded dummy variables. The model with the best fit kept the variables *duration of the last syllable of the AP*, *IS-LHLH*, and *rate*. Together these variables accounted for 73.8% of the variance in late H placement. Cross-validation using the test data, however, failed to confirm the contribution of the *rate* variable. This is not surprising, since a model without this variable still accounts for 73.6% of the variability in the training data. The details of the best model are given in Table 3.5. The coefficient for the

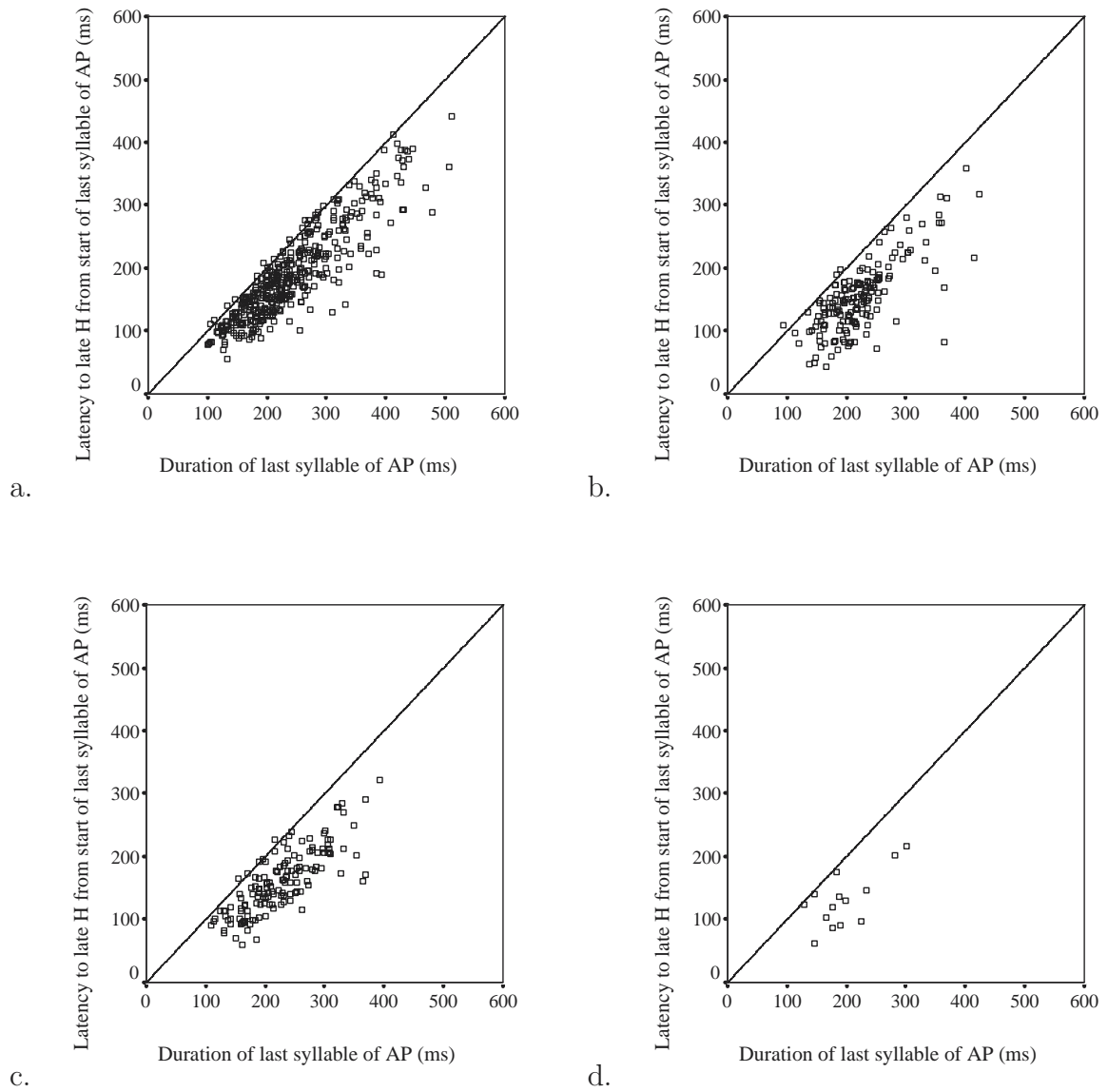


Figure 3.32: *Location of the late H as a function of duration of the last syllable of the AP by AP type. (a) LHLH (b) LH (c) LLH (d) L2H.*

constant	coeff 1	coeff 2	adjusted R^2
-16.996	.762 (duration last AP syl)	22.030 (IS-LHLH)	73.6%

Table 3.5: *Coefficients for the late H regression model.*

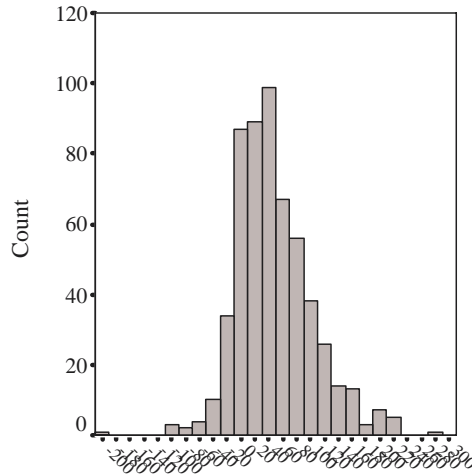
dummy variable *IS-LHLH* should be interpreted to mean that the late H for APs of the type LHLH will be realized on average 22 ms later than the late H for other AP types. This later placement of the H tone is perhaps attributable to tonal crowding—the late H may be pushed farther to the periphery to allow realization of all four tones.

Late L elbow

The position of the late L elbow is reported in the literature to be less stable than that of the late H, appearing sometimes in the same syllable as the late H and sometimes in the penultimate syllable. The timing of the late L elbow was examined for all AP types which contained the elbow. A total of 552 APs of types LHLH, LLH, and L2H were examined.²¹ The late L elbow (l-el) was realized in the last syllable 82% of the time and in the penultimate syllable the remaining 18% of the time. This variation is shown in the histogram in Figure 3.33.

As Figure 3.34 demonstrates, there were differences across speakers in the realization of the late elbow. Some speakers almost never realized the late elbow in the penultimate syllable: Speakers 3 and 4 produced late elbows in the penultimate syllable in less than 5% of cases. Speaker 1, on the other hand, produced 49% of her late elbows in the penultimate syllable. (No speaker produced the majority of her late L

²¹Note that AP types LH and LHH are not included in this analysis because although these accent types contain a late H, any elbow present is an early elbow, not a late elbow.



Latency to late L elbow from penultimate syl.-last syl. boundary (ms)

Figure 3.33: *Distribution of the late L elbow.*

elbows in the penultimate syllable.) Because of this difference, no attempt was made to pool the data for all speakers.

Each subject's data were examined to see whether the location of the late elbow could be explained in terms of some syllable landmark of the penultimate syllable of the AP, as predicted by the Late L Associated Tone Hypothesis and Jun and Fougeron (2000). There was, however, no correlation for any subject between duration of the penultimate syllable of the AP and latency from either the left edge of the penultimate syllable or the right edge of the penultimate syllable.

Since the late L elbow appeared most often in the last syllable of the AP, separate regression analyses were performed for each speaker to see whether there was any evidence that the late L tone was associated to that syllable. One set of analyses used latency from the left edge of the last syllable of the AP as the dependent measure

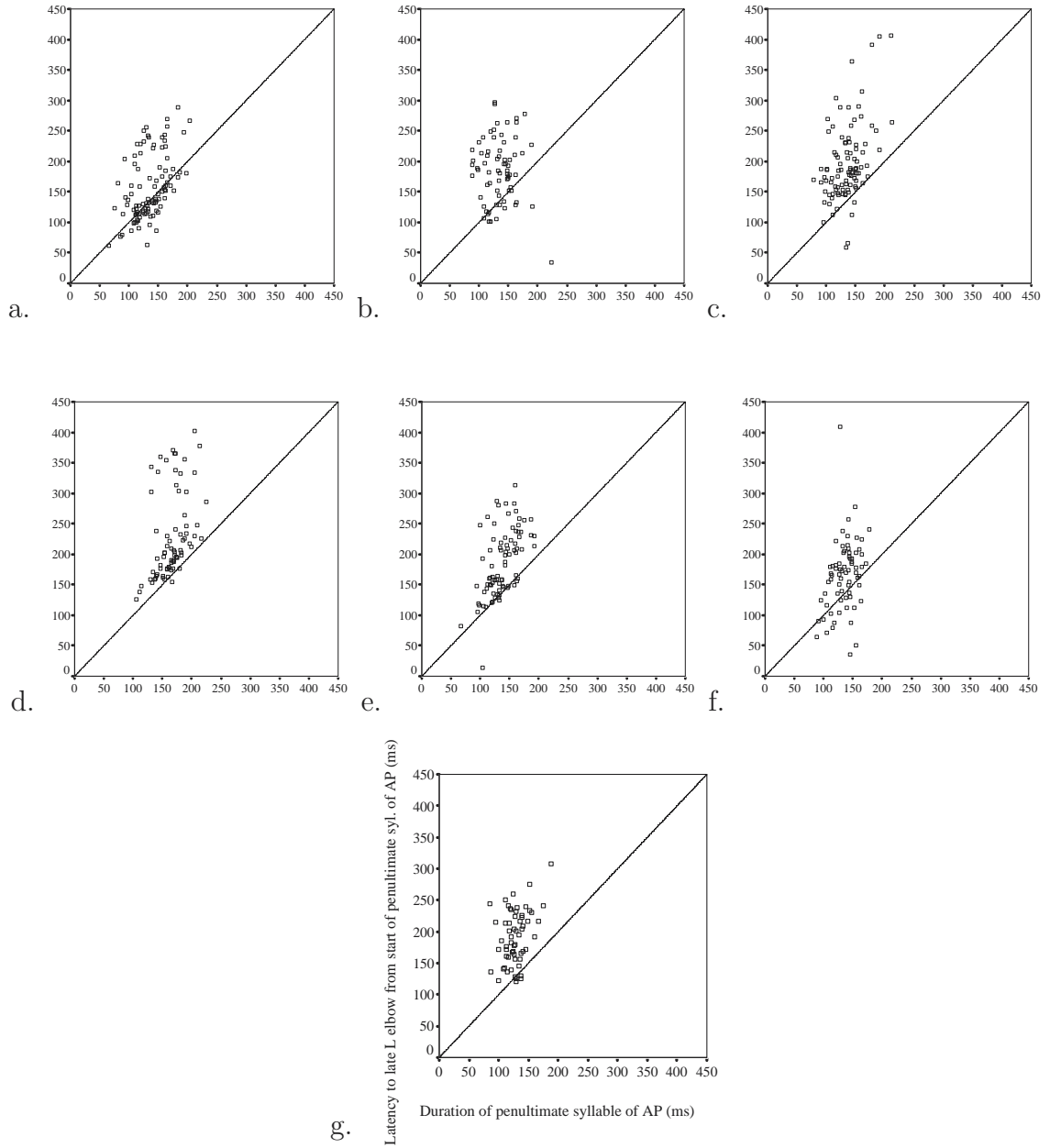


Figure 3.34: *Location of the late elbow (l-el) as a function of duration of the penultimate syllable of the AP by speaker. (a) Speaker 1. (b) Speaker 2. (c) Speaker 3. (d) Speaker 4. (e) Speaker 5. (f) Speaker 6. (g) Speaker 7.*

speaker	R^2 , lat. from left edge of last syl	R^2 , lat. from right edge of last syl
1	.224	<u>.459</u>
2	<u>.289</u>	.205
3	<u>.521</u>	.455
4	.423	<u>.582</u>
5	.416	<u>.460</u>
6	.213	<u>.362</u>
7	.116	<u>.524</u>

Table 3.6: R^2 values for each speaker for regression analyses using as dependent variables (1) latency from the left edge of the last syllable of the AP to the late L elbow (l-el) tag and (2) latency from the right edge. The independent variable in both cases was duration of the last syllable, excluding any non-sonorant rhyme. For each speaker, the higher R^2 value is underlined.

and duration of the last syllable of the AP, excluding any non-sonorant rhyme, as the independent measure.²² Another set used latency from the right edge of the last syllable of the AP as the dependent measure. The results from these analyses are hard to interpret. For some subjects, latency from the left edge of the syllable was a better measure; for others, latency from the right edge yielded higher R^2 scores. The R^2 values for these analyses are given in Table 3.6.

If we were to take this as evidence about the association of the L of the late rise (the late L elbow) to the last syllable of the AP, we would have to argue that speakers of French have induced very different models of intonation. Some speakers have a model in which the association to the last syllable of the AP is implemented phonetically with reference to the distance from the left edge of the syllable and some speakers have a model in which that association is implemented with reference to the distance from the right edge of the syllable. This seems unlikely.

²²The duration of the entire last syllable of the AP and the duration of the sonorant portion of the AP differed for only two target words: *limonade* and *malaise*.

It is possible that the earlier and late timings of the L of the late rise may convey pragmatic differences, with later elbows conveying affirmativeness or indignation. Post (2000) discusses this possibility (p. 126).

Another possibility is that since the early L and the late H mark edges (the left edge of the AP and the function word-content word boundary for the early L and the right edge of the AP for the late H), their timing is more constrained and less free to vary than that of the early H and the late L elbow, which are AP internal.

A third possibility is that these tones are phased with an articulatory dimension. Recent work suggests such a phasing relationship between certain tonal and articulatory events in the Neapolitan variety of Italian. In Neapolitan Italian, the H of yes/no question rises is systematically later than the H of narrow focus statement rises (D’Imperio (2000), D’Imperio (2001); D’Imperio and House (1997)). D’Imperio et al. (2003) report preliminary evidence that these H targets are more closely phased with the articulatory dimension of between-lip distance than with two common segmental landmarks (onset and offset of the stressed vowel). Statement H tones are phased with maximum between-lip distance within the stressed syllable. The authors note that this location does not correspond to any identifiable segmental boundary, acoustic event or phonological unit, nor does it overlap with RMS peak amplitude. A pilot study examining the potential phasing of tonal and articulatory events in French is underway.

Timing of the late H and preceding L with respect to each other

In order to test the Late L Leading Tone Hypothesis, times for the late H and late elbow were extracted and rise time (temporal distance between these two tones) was calculated for all tokens with a late rise. The AP types examined were LHLH,

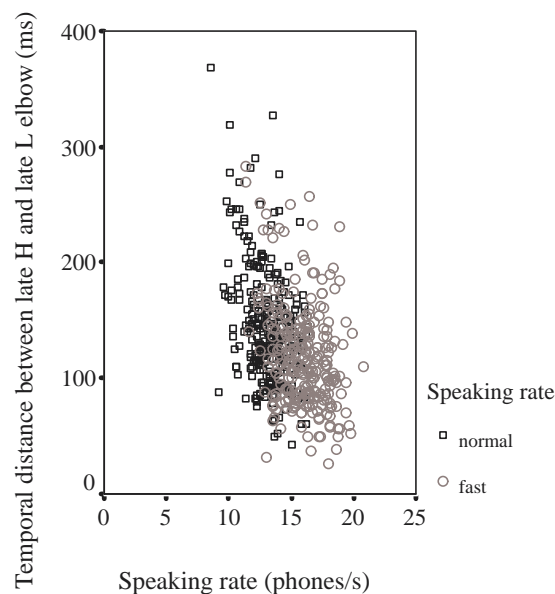


Figure 3.35: *Rise time of the late rise as a function of speaking rate.*

LLH, and L2H. Note that while AP types LH and LHH contain late Hs, they are not included in this analysis: LH contains a rise from an early L to a late H and the rise in LHH is an early rise. If the hypothesis is correct, the distance between a late H and its preceding elbow will be fairly invariant. The results are shown in Figure 3.35.

As with the early rise, it is clear from the chart that there is no invariant rise time for the late rise at either rate—rise time varies from under 50 ms to over 300 ms. The Leading Tone hypothesis is therefore rejected.

Slope of the late rise

We also examined the hypothesis that the slope of the late rise was invariant, using the same procedure for calculating slope described in §3.4.5. As the graph in Figure 3.36 makes clear, that hypothesis can be quickly rejected. There is a great

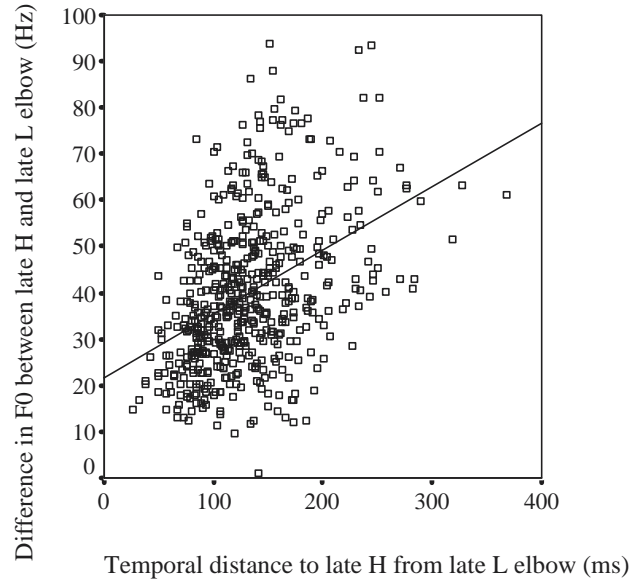


Figure 3.36: *Slope of the late rise.*

deal of scatter in the data points and a regression line fit to the data has an R^2 of only 18%.

3.4.6 Distinguishing the early rise and the late rise

Thus far, we have concentrated on the characteristics of the “primary accent” and “secondary accent” related to fundamental frequency and the late rise and early rise. These f_0 differences are considerable: the L of the early rise is aligned to the boundary between a function word and a content word and the H of the late rise is located close to the end of the last syllable of the phrase; the early H is often realized as a plateau, while such realizations are rare for the late H; and the late H is often higher than the early H. There are also other types of cues available to the listener. For example, there is a marked lengthening of the last content word syllable of a phrase. The histograms

in Figure 3.37a, b, and c illustrate this lengthening, contrasting the durations of content word-initial syllables, content word-final (and AP-final) syllables, and second syllables (not including second syllables of two-syllable content words, since those syllables are also final). Final syllables are almost 100 ms longer on average than non-final syllables. Wilcoxon Signed Ranks Test showed that this difference is significant (for the final syllable/first content word syllable comparison $Z = -23.836$, $p < .001$; for the final syllable/second (non-final) syllable comparison $Z = -18.764$, $p < .001$). There is no significant difference in duration between first and (non-final) second syllables ($Z = -1.275$, $p = .202$). These results are in line with those of other studies that have shown that the late rise, but not the early rise, is accompanied by lengthening.

We must interpret these comparisons with caution, however, given that there are syllable structure differences between the AP-final syllables and non-AP-final syllables. While almost all non-final syllables had a CV structure, most AP-final syllables had a CVC structure (e.g., *le mélomane* [me.lo.man] CV.CV.CVC). The materials were designed to maximize sonorant segments, and this imbalance in syllabic structure was unfortunately not recognized as potentially problematic until after the data had been collected. However, as Figure 3.37d shows, even if we subtract the duration of any coda consonant (e.g., the [n] of *mélomane* [me.lo.man]) from the duration of AP-final syllables, these stretches of the AP-final syllable are still significantly longer on average than both content word-initial syllables ($Z = 18.709$, $p < .001$) and non-final, second syllables ($Z = 12.835$, $p < .001$).

We note also that Rolland and Loevenbruck (2002) found very similar results in a production study whose targets contained only CV syllables. In that study, the

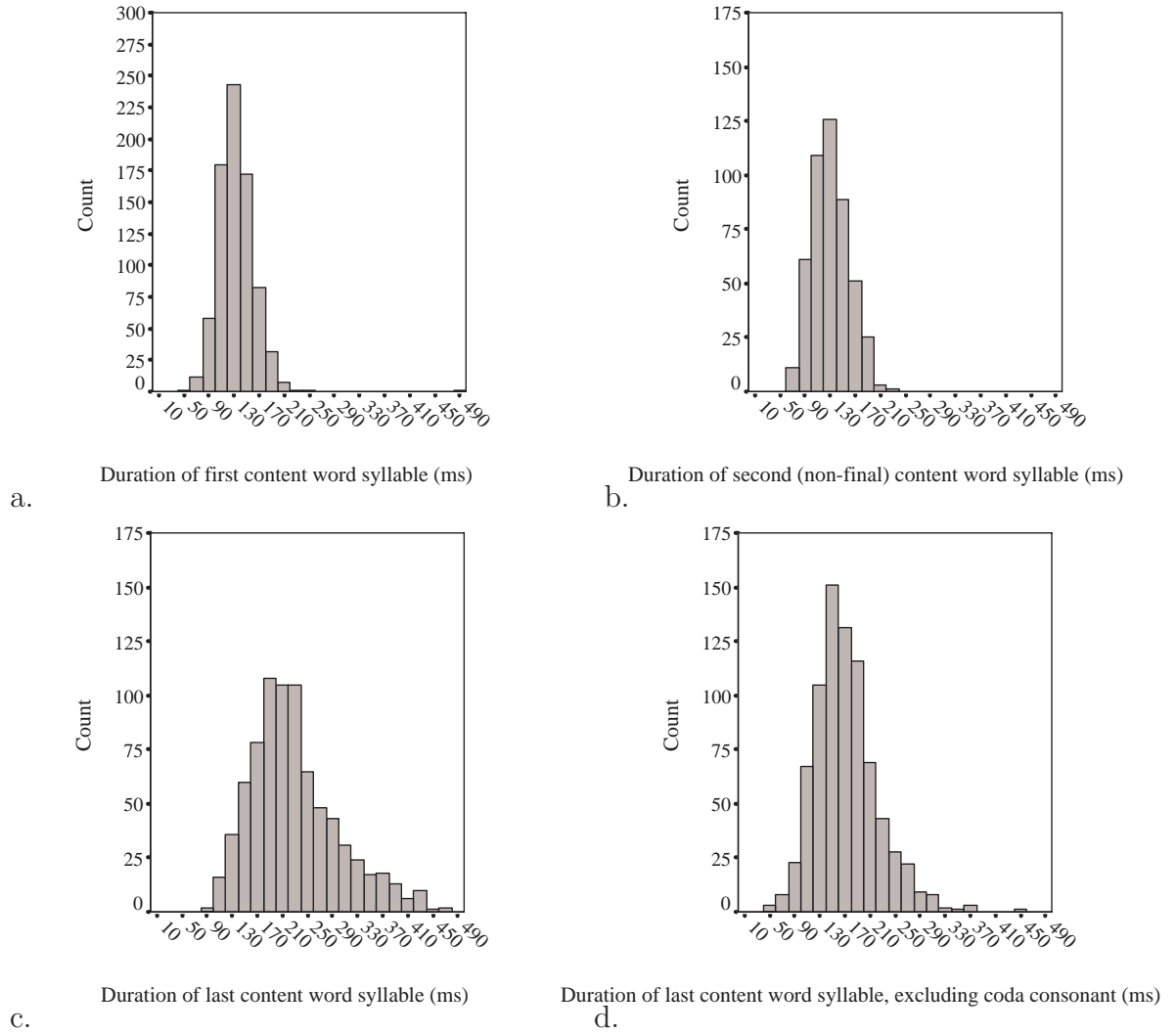


Figure 3.37: Durations of (a) first content word syllables (mean = 134 ms, std. dev. = 29.04), (b) second content word syllables (for 3- and 4-syllable content words, where second syllable is non-final) (mean = 131 ms, std. dev. = 28.59), (c) final content word syllables (mean = 230 ms, std. dev. = 71.54), (d) final content word syllables, excluding any coda consonants (mean = 172 ms, std. dev. = 50.21).

peak-bearing syllable of the late rise was significantly longer than that of the early rise. These researchers conclude that "...H* is probably perceived as the primary accent because it reaches a higher f_0 value, and because the f_0 rise is also more dynamic than that of Hi (larger in amplitude and shorter in time) and [borne] by a single syllable" (p. 613). We find a similar difference in "dynamics" in our results. The temporal distance between the L and H of the early rise (the rise time) was greater than that between the L and the H of the late rise, whether measured in clock time (148.2 ms vs. 131.5 ms) or average number of syllables spanned ²³ (1.03 vs. 0.17). Although the f_0 peaks of late rises in our corpus typically reached f_0 values higher than those of early rises, we note that the size of the pitch excursion (40.4 Hz for early rise vs. 39.7 Hz for the late rise) is not significantly different. Given the same excursion size, but a longer rise time, the slope of the early rise (298.6 Hz/s) is shallower than that of the late rise (324.3 Hz/s). The results of a Mann-Whitney Test for Two Independent Samples shows that this difference is significant ($U = 116372$, $Z = -2.984$, $p < .01$)).

3.4.7 French Intonational Structure

Our results allow us to make a number of conclusions about the structure of the early and late rises.

First, it is clear that the early rise and the late rise are structurally different. Each can be described as a L H tone sequence, but the structural similarity ends there. The L tone of the early rise is aligned to the boundary between a function word and a content word and can stretch leftward to mark the left edge of the phrase. The H

²³As an example, if the L of a rise is realized in the last function word syllable (coded *syllable 0*) and the H is realized in the second content word syllable (coded *syllable 2*), the syllable span is 2 ($2 - 0 = 2$). If the two tones are realized in the same syllable, the syllable span is 0).

of the early rise can be realized on the first or second content word syllable (or later, according to reports in the literature), and it can be realized as a plateau with no change in meaning. In contrast, the H of the late rise is consistently realized at the end of the last full syllable of the AP. Plateau realizations are less common and at least sometimes may signal a difference in meaning. The alignment of the L of the late rise varies greatly within and across speakers. We also find evidence for a difference in scaling between the two rises. These alignment and scaling differences argue against the Di Cristo and Hirst claim that the early rise and the late rise have the same structure, that they are both Tonal Units. The evidence is all the more compelling because the Di Cristo and Hirst account explicitly predicts only a duration difference between the early rise and the late rise, accounted for by the Rhythmic Unit.

The evidence also argues against Post's claim that the H of the early rise is a pitch accent. Post (2000) treats both the H of the early rise and the H of the late rise as pitch accents (noted H*). We can make a good argument that the H of the late rise is a pitch accent (or part of a bitonal pitch accent). The peak of the late rise is uncontroversially associated with the last full syllable of the prosodic phrase and is consistently realized late in last full syllable of the phrase. In addition, the late H can only be realized on full (non-schwa) syllables, in line with many other languages in which pitch accents are associated to metrically strong syllables, including all of the other Romance languages. The association of a tone to a metrically strong syllable is an essential structural characteristic of a pitch accent. Yet, as we have seen, the early H shows no evidence of an association to a particular syllable. In our data, the peak of the early rise was sometimes realized in the first syllable, sometimes well into the second syllable (see Figure 3.25). Based on this evidence we reject the

classification of the early H as a pitch accent. We note that our observations agree with the Padeloup (1990) claim that the utterance-initial secondary accent is not linked to a specific syllable, that “...ce n’est pas tant la démarcation précise du début d’un mot ou d’un groupe de mots qui importe, mais plutôt la réalisation rapide d’un accent indépendamment des limites syllabiques” ‘...it is not so much the precise demarcation of the beginning of a word or a group of words that is important, but rather the quick realization of an accent independent of syllabic limits’ (p. 142).

The results of the current experiment strongly support the hypothesis that the L of the early rise is an edge tone (or part of a compound edge tone) with a double association to the left edge of the prosodic phrase and the left edge of the first content word, as proposed in an earlier study (Welby (2002)). In cases where the L cannot stretch to the left edge of a prosodic phrase, it can stretch to the left edge of another syllable boundary. Claims of the affiliation of the L of the early rise to the last function word (Vaissière (1997) and Jun and Fougeron (2000)) are unsupported. Post’s (2000) distinction between a L boundary tone (%L in her notation) anchoring the early H and another type of L tone preceding the late H is supported.

The results also support Jun and Fougeron’s (2002) revision of the Jun and Fougeron (2000) model in which all tones were associated to separate syllables. We found patterns in the realization of the rise not predicted by the earlier model. For example, the four-tone LHLH pattern occurred in prosodic phrases with fewer than four syllables (see §3.4.3). In addition, the tonal alignment facts run counter to a number of predictions of the model. For example, both the L of the early rise and the L of the late rise are associated to syllables (and only to syllables) (see Figure 2.13,

p. 38). We expect this similarity in association to be reflected by a similarity in alignment. Yet the early L is systematically aligned to a boundary, while the alignment of the late L varies considerably.

Our results support an Autosegmental Metrical model along the lines of the Jun and Fougeron (2002) model, in which the early rise is treated as a bitonal phrase accent (LH-) and the late rise is treated as a bitonal pitch accent (LH*). The L of the early rise is clearly an edge tone, with a double association. And the H* of the late rise is clearly associated to a syllable and so is structurally a pitch accent.

The behavior of the two interior tones of the LH-LH*, however, leads us to revise a certain number of assumptions about the nature of our structural units. As reported in §3.4.5, there was no evidence that the H of the early rise was associated to a syllable or that it trailed its preceding L at an invariant distance. Similarly, the L of the late rise was not associated to a particular syllable and did not lead the late H at an invariant distance. Standard AM models assume that the tones of a bitonal pitch accent lead or trail their starred tone at a fixed distance.

At the same time, we have reason to believe that the interior tones are in fact part of compound tones. The results of Perception Experiments 1 and 2, for example, suggest that both tones of the LH- sequence of the early rise play a role in marking the left edge of the content word (see Chapter 5). Similarly, although the position of the L of the late rise varies considerably by comparison to that of the L of the early rise, the late L is always realized near the right edge of the AP. We never find, for example, cases like the one illustrated in Figure 3.38, in which the late L is realized just after the early H, in the middle of the AP rather than at the end. This timing

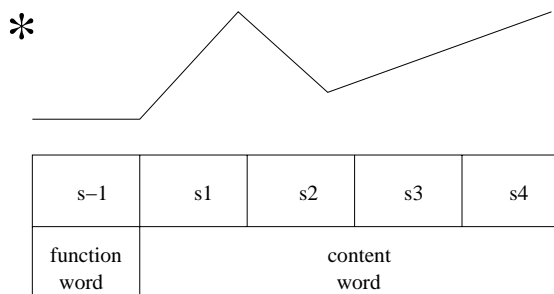


Figure 3.38: *Schematic example of an unattested alignment of the L of the late rise. The first row shows syllable boundaries. The star indicates that the pattern is not well-formed.*

is physically possible—we routinely find steep falls from late H targets to the early L targets of following APs. The L of the late rise may act to highlight the starred tone.

We also note that the distinction between phrase accent (LH- of the early rise) and pitch accent (LH* of the late rise) is purely a structural one. The LH* accent has the structural hallmarks of a Germanic-style pitch accent (association to a metrically strong syllable and lengthening of that syllable relative to other unreduced syllables), yet it clearly marks the right boundary of an Accentual Phrase. Thus it acts as an edge tone much like that of the early LH-.

CHAPTER 4

SPEECH SEGMENTATION

The scene: Nicolas (age 2;11) and Pauline reading an *Astérix* comic together

Nicolas: Ils font quoi? ‘What are they doing?’

Pauline: Ils se donnent des coups d’épées. ‘They’re fighting with swords (*épées*).’

Nicolas: C’est quoi, les dépees? ‘What are *dépees*?’

4.1 The speech segmentation problem revisited

The average two-year old learns about 10–15 words a day (Bloom, 1993). In order to accomplish this daunting task, a child must have strategies to find the beginnings and ends of words. The effectiveness of these strategies is shown by the apparent rarity of missegmentations like the one quoted above. Indeed, part of what makes these errors funny is that they are unexpected and surprising.

As we have seen in the Introduction, adults too make occasional errors, often with words that are new or unfamiliar to them. The photograph in Figure 4.1, taken at the Musée de la Révolution Française in Vizille, France, gives an example of such an error from the end of the 18th century. The maker of the plate clearly intends to proclaim *l’abondance* ‘abundance’, but missegments the presumably unfamiliar word as *la bondance*. The *la* on the first line is one form of the definite article, but *bondance* is not a word in Standard French.



Figure 4.1: A 211-year old speech segmentation error. The intended message is *l'abondance* 'abundance'. Photo by Pauline Welby, 2002. Used by kind permission of the Musée de la Révolution Française, Vizille, France.

Occasionally such errors become permanent. In the *patois* of the town of Abondance in Haute-Savoie, *la bondance* means 'mangel-wurzel', a type of beet used as fodder for cattle (Christian Abry, p.c.). This word most likely stems from the same missegmentation of *l'abondance*.²⁴ An example from English of the preservation of a missegmentation is the Modern English word *apron*. Hundreds of years ago, the English phrase *a naperon* was misheard as *an aperon* often enough that the word *naperon* lost its initial 'n' and became the familiar *apron*.

²⁴Note that the direction of the missegmentation is not in doubt. *Abondance* ultimately comes from Latin *abundāre* 'to overflow', which in turn is a compound of *ab* 'away' + *undāre* 'to flow'.

In this chapter I examine what is known about the cues that listeners use to effectively segment speech, making examples like these rare and curious exceptions rather than ever-present roadblocks to communication.

4.2 Research in speech segmentation

A large body of research shows that listeners use their tacit knowledge of a wide range of patterns in their native language to help them segment speech, making the process automatic and seemingly effortless. Listeners are sensitive to patterns at many levels of linguistics organization. They exploit, for example, their knowledge of regularities in allophonic variation, phonological pattern frequencies and constraints, and prosodic information such as duration, stress, and intonation patterns.

A growing body of evidence shows that listeners are sensitive even to very fine-grained patterns in their native language. Models of phonology have been proposed in which a speaker's phonological knowledge includes not only abstract categories for things like syllable, consonant, and individual phonemes, but also fine phonetic detail (see Pierrehumbert (2002, 2003)). Such models of phonology are able to explain not only patterns that make reference to these abstract categories (a language-specific constraint against /mr/ onsets, for example), but also those that point to a listener/speaker's more detailed knowledge about phonetic encoding (such as sensitivity to what voice onset time differences distinguish voiced from voiceless consonants in a given language). According to exemplar model theory, this detailed knowledge is built up through the encoding of many different tokens for each label or category (see, for example, Johnson (1997) and Pierrehumbert (2001a)).

Different types of cues are likely to be more or less relevant for different languages. Even within a single language, some cues may be more reliable and hence more used in a given linguistic or real-word context. For example, people talking at a busy subway station may rely on prosodic cues like syllable durations and intonation patterns, since those cues are preserved in noise, while other phonetic details may be overwhelmed (see §4.2.5).

4.2.1 Phonological pattern frequencies

Listeners are sensitive to sound sequence restrictions in their native language. Evidence that listeners are sensitive to phonotactic restrictions comes from the results of word-spotting experiments in which listeners are asked to detect words in longer nonsense word carriers. Listeners find it easier to detect words when their onsets are aligned with phonotactic boundaries than when they are misaligned. In Dutch, voiced obstruents are not found word finally, and the sequence [mr] is not a possible onset or coda cluster. Such a sequence can only occur across a syllable boundary and therefore a possible word boundary. Examples are given in (22).²⁵

- (22) Een vallende boomraakte de tramrails waardoor de trams moesten
 a falling tree touched the trolley rails so the trolleys had to
 omrijden
 make a detour
 ‘A tree fell and touched the trolley rails, so the trolleys had to make a detour.’

For Dutch listeners, the word *rok* ‘skirt’ is easier to detect in [fim.rɒk] where the [mr] sequence forces a syllable boundary, than in [fi.drɒk] where the sequence [dr] must be syllabified as a syllable onset (McQueen (1998)).

²⁵I thank Peter van der Heijden for providing the example.

Finnish listeners show similar sensitivity to word segmentation cues provided by the vowel harmony constraints in their language that restrict which vowels can cooccur within the same word (Suomi et al. (1997), Vroomen et al. (1998)). Finnish has a neutral vowel class /i,e/ and two harmony classes, one containing the three front vowels /y,ø,æ/ and the other containing the three back vowels /u,o,a/. If a word begins with a syllable containing a member of one of the harmony classes, all following vowels must be members of the same harmony class or of the neutral class. Suomi et al. (1997) found that Finnish listeners were able to use this phonotactic constraint as a cue to word segmentation: they were quicker to detect real words embedded after disharmonious contexts (e.g., the real word *hymy* ‘smile’ in *puhymy*, where the first syllable contained a member of the back vowel harmony class and the following syllables members of the front vowel harmony class), than those embedded after harmonious contexts (e.g., *pyhymy*).

Listeners use not only phonotactic constraints (i.e., zero or near-zero probability of occurrence) in their native language, but also weaker statistical tendencies as cues to speech segmentation. Saffran et al. (1996) found that adults learning an artificial language used the relative strength of transitional probabilities between phonemes to hypothesize word boundaries in the language. Van der Lugt (1999) showed that in combination with a syntactic boundary, a common word-initial CV sequence facilitated the detection of a target word in a word-spotting task with Dutch listeners. Mattys et al. (1999) used head turn preference experiments to show that infants as young as nine months old are sensitive to transitional probabilities. Infants in the study reacted differently to high and low probability cross-syllabic consonant sequences in CVC.CVC stimuli.

These findings are in line with a host of studies that have shown that listeners are exquisitely sensitive to sublexical phonological pattern frequencies in their native language. Naive listeners judge nonwords with more common sequences to be more wordlike than those with less common sequences (Coleman and Pierrehumbert (1997), Vitevitch and Luce (1999), Frisch et al. (2000), Munson (2000), Hay et al. (in press), *inter alia*). In addition, nonwords with common sequences are retained longer in recognition memory (Frisch et al. (2000)). Pattern frequency effects have also been found in phoneme monitoring (McQueen and Pitt (1996)) and phoneme categorization (Pitt and McQueen (1998)) tasks. Hay (2000) showed that language users use low probability phoneme sequences across morphological boundaries in morphological decomposition.

Phonological pattern frequency has also been found to have effects on production. In a nonword repetition task, latency from the nonword prompt to the participant's repetition of the prompt was longer and the repetitions were less accurate for items containing low frequency sound sequences (Vitevitch et al. (1997)). Participants in a language game task were more likely to preserve high-frequency VC rimes than low frequency ones (Treiman et al. (2000)). Munson (2000) also found an effect of phonological pattern frequency with child participants: children's repetitions were less accurate, longer in duration, and more variable for nonwords with less common sound sequences.

4.2.2 Prosodic cues

Stress cues

Listeners also use language-specific metrical cues in speech segmentation. English listeners know, for example, that content words (nouns, adjectives, etc.) in their language tend to begin with stressed syllables (Cutler and Carter (1987)). As a concrete example, in the preceding sentence, all but two of the content words follow this generalization (the exceptions are *e'xample* and *be'gin*). Cutler and Norris (1988) showed that listeners were able to use this pattern in speech segmentation. They detected the *mint* [mɪnt] embedded in ['mɪntəf] more easily than the [mɪnt] embedded in [mɪn'tef]. The results of this word-spotting task were interpreted to mean that listeners hypothesize a boundary to the left of the strong second syllable in [mɪntef], hindering detection of [mɪnt]. Observational studies and controlled laboratory experiments have shown that when English listeners missegment speech, their errors reflect the stress pattern of the language (e.g., Smith et al. (1989) and Cutler and Butterfield (1992)). For example, errors in which a word boundary is placed before a stressed syllable, as in the *i'llegal* of (23), are much more common than those in which a word boundary is placed before an unstressed syllable. This use of stress information was dubbed the Metrical Segmentation Strategy (Cutler and Norris (1988)).

(23) *The speaker said:* The parade was illegal.

The listener heard: The parade was an eagle.

[from Garnes and Bond (1980), an early study of perception errors in spontaneous speech]

Even nine-month old infants acquiring English use the strong syllable/weak syllable pattern predominant in bisyllabic words in the language as a cue to a word boundary (Mattys et al. (1999)). Although infants this age are also sensitive to phonotactic

cues, prosodic information overrides phonotactic information. The results of a head turn preference task showed that infants listened longer to strong/weak stimuli (the common pattern) with infrequent phoneme sequences than to weak/strong stimuli with common sequences.

English listeners appear to be sensitive to a specific correlate of stress in their language: stressed syllables always contain full vowels, while unstressed syllables often (though not always) contain reduced vowels. Although there are other correlates to stress, including increased intensity and duration and frequent association with pitch accents, the evidence suggests that English listeners do not use these so-called “suprasegmental” cues in lexical access and that segmental vowel quality information provides a much stronger cue. This primacy of segmental cues is supported by the finding that each member of minimal pairs like *'forbear/for'bear*, in which the unstressed syllable contains a non-reduced vowel, primes semantic associates for both members of the pair (Cutler (1986)). Bond and Small (1983) found that word misstressing did not have an effect on shadowing performance unless it was accompanied by a change in vowel quality (see also Slowiack (1990) and Cutler and Clifton (1984) for similar results). In a cross-splicing experiment, Fear et al. (1995) found that listeners were more sensitive to the difference between full and reduced vowels than to stress differences between full vowels.

Dutch, a Germanic sister language to English, has similar stress patterns, and Dutch listeners show the same pattern of missegmentations, placing word boundaries to the left of stressed syllables (Vroomen and de Gelder (1995), Vroomen et al. (1996)). Yet there are differences between the two languages. Unlike their English-speaking counterparts, Dutch listeners appear to be sensitive to correlates of stress

other than vowel reduction (Cutler and van Donselaar (2001)). Stress in Dutch has many correlates, but vowel reduction does not play a major role in signalling the stressed/unstressed distinction: Dutch unstressed syllables are very likely to contain strong, unreduced vowels. In a fragment identification task, Dutch listeners reliably identified the source of a syllable taken from a stress minimal pair with no difference in vowel quality (e.g., the fragment *lo* from '*doorlopen* 'run on' or *door'**lopen* 'proceed through'). And in a lexical decision task, there was no evidence that one member of a stress minimal pair primed the other member, suggesting that information from increased intensity, duration, and/or pitch excursion effectively signalled the difference between the two items.

Another study directly examined the use of nonsegmental word stress information on lexical access by native speakers of English and native speakers of Dutch (Cooper et al. (2003)). In an offline forced-choice identification task, Australian and Dutch participants listened to English word fragments like [myuz] and judged which of a pair of words ('*music* or *mu'**seum*) was the source of the fragment. The members of all pairs differed in stress placement, but not in vowel quality. Although the stimuli were English, Dutch listeners actually outperformed the native speaker listeners, correctly identifying 72% of the fragments, while the Australian listeners correctly identified only 59%. The Dutch listeners were able to exploit their native language sensitivity to correlates to stress other than vowel quality in making judgments about the English fragments. Of course, this information was also available to the Australian listeners. The relatively poor performance of the native listeners of English further strengthens the primacy of segmental information in the use of stress cues to speech segmentation and lexical access in English. The observed differences between English and Dutch

might be explained by a model that includes a more detailed level of phonological representation in addition to the traditional level that includes abstract categories and features like *syllable* and *stress*. In an exemplar theory account, native listeners of English would encounter enough exemplars of stressed and unstressed syllables to support the generalization that unstressed syllables in English contain reduced vowels. Dutch listeners would gain similarly fine-grained knowledge about the characteristics of stressed and unstressed syllables in their language.

Syllabification cues

Speaker/listeners of languages that do not have the type of stress that is important in word segmentation in languages like English and Dutch may use other language-specific prosodic cues to speech segmentation. It has been argued for French, a language without contrastive word stress, that listeners use the syllable as a speech segmentation unit. In a landmark study, (Mehler et al. (1981)) showed that French listeners are able to detect auditorily presented target sequences like [ba] more quickly in carrier words like *balance* ‘scales’ than words like *balcon* ‘balcony’. In the first case, the target sequence [ba] matches the syllabic structure of *ba.lance*, while in the second, there is a mismatch (*bal.con*). The reverse pattern is found if the target sequence is [bal]. This interaction between target and carrier was interpreted as evidence that French listeners use syllable boundaries to hypothesize potential word boundaries. The authors argued that the syllable was a language-universal “processing unit in speech perception.”

A later study, however, showed evidence of language specificity: the response pattern characteristic of the so-called syllable effect was found with French listeners presented with English stimuli, but not with English listeners presented with English

or French stimuli (Cutler et al. (1986)). These researchers explicitly linked the absence of syllable-based segmentation strategies among English listeners to the greater diversity of syllable types in English and a lack of clear syllable boundaries in the language, factors which were postulated to make a syllable-based segmentation strategy inefficient for that language. As an illustration of the claimed cross-linguistic difference in syllabification, the authors compare the syllabification of the French word *palace* ‘luxury hotel’ and the English word *palace*:

In French the syllabification is clear—there is a syllable boundary between *pa-* and *lace*. In English, however, the syllable boundary falls neither before nor after the /l/... Phonologists represent the syllable structure of French *palace* as [pa][lace], but of English *palace* as [pa[l]ace]; that is, the /l/ properly belongs to both first and second syllables (Anderson and Jones (1974), Kahn (1976)). Segments which belong to two syllables at once are said to be ambisyllabic. In stress languages, intervocalic consonants preceding an unstressed vowel are frequently ambisyllabic.

p. 387

The authors also clearly assume a link between ease of syllabification and inherent physical properties of the experimental items:

Experiment 2, in which listeners heard French materials, added the further interesting finding that English listeners do not use syllabification even when the words they are listening to can be easily syllabified. Experiment 4, on the other hand, presented French listeners with English materials and showed that French listeners do employ syllabification even when some of the words they are listening to [i.e., words like English *balance*, where the medial /l/ is assumed to be ambisyllabic] are hard to syllabify.

p. 397

To use a concrete example, the claim is that English *palace* is inherently “hard to syllabify” and French *palace* inherently “easy to syllabify,” regardless of the native language of the listener. The claim that cues to syllabification are present in the

signal, *independent of the native language of the listener*, goes against what we know about the language-specificity of syllabification. It is well known that listeners use native language knowledge in their segmentation decisions. A well-known example is the case of the phoneme sequence [ski] (corresponding to the English word *ski*). English speakers hear one syllable, while Japanese speakers, whose phonology does not allow consonant clusters, hear two (see Beckman (1996)). In addition, we might expect that predictable allophonic variation in native language stimuli would give native speakers an advantage in syllabification. For example, English /l/ has two well-known positional variants: clear-*l* [l] and dark-*l* [ɫ]. English-speaking listeners, who expect /l/ in coda position to be realized as [ɫ] in their native language, may encounter difficulty in syllabifying the French word *balcon*, where the coda /l/ is pronounced [l].

Cutler et al. (1986) predict that a syllable-based segmentation strategy will be used by listeners of any language with a clear syllable structure:

Furthermore, we assume that the effects are not specific to French and English, but that speakers of any language with clearly bounded regular syllables should show syllabification effects, while speakers of any language with irregular, hard-to-segment syllables should not.

p. 397

Recently a language-specific role for the syllable in speech segmentation has been called into question. The strong claim linking clear syllable structure and the use of a syllable-based segmentation strategy has motivated a number of studies on the possible syllable effect in other Romance languages, whose syllable structures are widely assumed to be straightforward. A description from Sebastián-Gallés (1996) exemplifies this assumption: “As all Romance languages, [Catalan and Spanish] have

clear syllabic boundaries (no ambisyllabicity as is found in Germanic languages)” (p. 172). However, these studies have had mixed results. A study of Italian found no evidence for the syllable effect (Tabossi et al. (2000)). One study of both Catalan and Spanish failed to find strong support for the effect (Sebastián-Gallés et al. (1992)). For Catalan, the characteristic target/carrier interaction was found only for items with a certain word stress pattern. For Spanish, the effect was only present when overall reaction times (RTs) were particularly slow. While another study of Spanish found the interaction pattern, RTs in that study were also long (Bradley et al. (1993)). As Dupoux (1993) and Content et al. (2001b) note, the fact that the effect is reliably found only when performance is slow suggests that the apparent syllable effect may not reflect earliest stages of processing, but rather conscious segmentation strategies.

The case for the “syllable effect” is particularly vulnerable since all of the evidence put forth for the effect comes from a single paradigm, the target-detection paradigm. This stands in contrast to the converging evidence for the use of stress information in segmentation decisions in English, where the evidence comes from a variety of paradigms including observational data from naturally occurring and laboratory-induced speech errors and the perception of noise-masked and low amplitude speech (although see the caveat in §4.2.5). Recent studies have cast doubt on the conclusions of the original target-detection study on French (Mehler et al. (1981)). Noting that the study used only five pairs of words, all of which contained the vowel [a] in their first syllable, followed by a liquid consonant (the “pivotal consonant,” either [l] or [r], e.g., *balance* / *balcon*, *palace* / *palmier*), Content et al. (2001b) set out to replicate the original result with nonword materials with more varied segmental composition. The new study used pairs of nonwords that varied in their segmental composition,

such as [bupis]/[buptal] and [gife]/ [giffɔ̃ʁ], where the pivotal consonants are a stop [p] and a fricative [f], respectively. Subjects in the study were native speakers of Swiss French. The results showed a three-way target \times carrier \times pivotal consonant class interaction. Separate analyses of each pivotal consonant class showed that the interaction between target and carrier characteristic of the syllable effect was reliably present only for items with liquid pivotal consonants (e.g., [firu]/ [firnɛʁ]), the type of items used in the original study. The size of the effect was strongest for the slowest participants, in line with other studies on the “syllable effect.” For items with liquid pivotal consonants, the effect was reliable only in conditions which blocked items by vowel (of the first syllable) or pivotal consonant class. The demonstrated sensitivity to segmental composition is strong evidence against an account which uses syllable classification as a mechanism for speech segmentation. As Content et al. (2001b) put it, “...the fact that the syllabic effect is limited to one particular class of pivotal consonants seems hard to reconcile with the notion of a ‘bank of syllabic analysers’ (Mehler et al. (1990)) ” (p. 632).

Despite these results, some might argue that syllable classification plays an important role in French speech segmentation, but only for syllables involving a limited class of segments, namely liquids. *A priori*, the idea that a segmentation cue could be class specific is entirely plausible. This is clearly the case for some segmentation cues, such as those from allophonic variation. English listeners are sensitive to the fact that voiceless stops have release bursts syllable initially, but tend to be unreleased syllable finally (Lehiste (1960)). Information about the distribution of bursts is irrelevant for segment classes like liquids and fricatives, though, since they lack bursts. Yet on closer examination, the idea that syllable classification is an effective segmentation

strategy for syllables with liquids seems implausible. Content et al. (2001b) question even the claim that syllable boundaries in French are straightforward, noting that French-speaking subjects show a great deal of variability in explicit syllabification (see Content et al. (2000) and Goslin and Frauenfelder (2000)). This variability turns out to be greatest in the case of syllables with liquid consonants. In a study of syllabification by French-speakers (Swiss and Belgians) using a variety of tasks (syllable reversal, repetition of word “parts,” repetition of word “parts” in a sentence frame), Content et al. 2001 found that subjects often assigned the medial consonant to the “first part” of CVCV words or to both the first and second parts. This type of assignment was significantly more common for items with liquid medial consonants (e.g., *palais* [palɛ] ‘palace’) than for items with obstruent medial consonants (e.g., *souci* [susi] ‘worry’). While the CVCV structure of these items mirrors that of the CVCV items in Content et al. (2001b) and other studies on the syllable effect, listeners’ explicit syllabification decisions disagreed with the presumed CV.CV syllabification most often for items with *liquid pivotal consonants*. This is precisely the case for which the target \times carrier crossover interaction characteristic of the syllable effect is found in some conditions. Something is clearly different about the items with liquid pivotal consonants, but it would be difficult to maintain that their special behavior in the target-detection task is evidence for a limited syllable classification effect.

Content et al. (2001b) argue that listeners typically use subsyllabic phonetic information in their segmentation decisions: “the perceptual process is based on smaller-sized units, and it delivers fine-grained phonetic information so that participants can continuously evaluate the phonetic and sub-phonetic match between the incoming

signal and target” (p. 632). According to this view, listeners build a syllabic representation only when their performance in the experimental task is slow. Content et al. (2001b) note that studies of the syllable effect have consistently shown that response times for CV targets are faster than RTs for CVC targets. They examined the time course of phonetic information to test the hypothesis that listeners detect a target as soon as enough phonetic information is available to make a match. Their results showed that most (85%) of the variance in RTs in responses to CVC targets could be accounted for by a regression equation incorporating latencies to segmental landmarks, although the results were less clear for CV targets.

Are all parts of the syllable created equal?

Content et al. (2001b) do not discount a role for syllable structure, but argue for the “Syllable Onset Segmentation Heuristic” (SOSH, called elsewhere the “Syllable Onset Segmentation Hypothesis”) in which syllable onsets are “privileged alignment points” (p. 634). Their view of the special status of onsets follows from their rejection of standard assumptions about syllabification. In their view (described in Content et al. (2001a)), “distinct processes are involved in locating onsets and offsets of syllables. Onset determination is more reliable and dominant” (p. 177). They contrast this view with the traditional “boundary” conception of syllabification in which locating a syllable onset entails locating a syllable offset. The original “syllable classification” model of French speech segmentation and subsequent versions of that model (Mehler et al. (1981), Mehler et al. (1990)) do not predict processing differences whether a syllable onset or a syllable offset is misaligned with a word boundary. If speech is classified into syllables, misalignment with an offset should be just as disadvantageous to processing as misalignment with an onset. To assess the potentially differing role

of syllable onsets and offsets in speech segmentation, Dumay et al. (2002) conducted a series of word-spotting tasks with French listeners in which misalignments occurred either at syllable onset or coda. For example, the real word *lac* ‘lake’ is aligned to a syllable onset in the nonsense carrier [zun.lak], but misaligned at its onset in [zu.glak]. The real word *jupe* ‘skirt’ is aligned in [ʒup.tɛʃ], but misaligned at its offset in [ʒu.plun]. In line with the SOSH, the results showed a consistently significant effect of RT and error rates for misalignment at the onset. Effects of misalignment at the offset were found only for certain measures (e.g., for one experiment, a significant effect was found only for RTs, by participants) and these effects were much smaller than the effects for onset misalignment.

As Dumay et al. (2002) note, the results of other studies are consistent with the view of the primacy of onsets and other word-initial information. For example, results from Finnish show that listeners use disharmonious initial contexts as cues to segmentation (*puhmy*), but not disharmonious final contexts (*hymyp*) (Suomi et al. (1997)). McQueen (1998) found a stronger effect with Dutch listeners for misalignment at syllable onset than misalignment at syllable offset. Van der Lugt (1999) also found a similar asymmetry; he concludes that “the likelihood of a word’s onset seems to be more important than the likelihood of a word’s offset.” He notes that these findings are in line with results that show an important role for word beginnings in word recognition:

Many studies have confirmed the intuitive notion that the beginnings of words are particularly important for the fast recognition of words in continuous speech. Gating studies (Grosjean (1980); Tyler and Wessels (1983, 1985)) have shown that words can be recognised reliably before their offset, at the point where they become unique from all other words in the language (the

“uniqueness point”). Other evidence from phoneme monitoring (Marslen-Wilson, 1984), auditory lexical decision (Taft, 1986), mispronunciation detection (Cole (1973); Marslen-Wilson and Welsh (1978)) and cross-modal priming (Marslen-Wilson and Zwitserlood (1989), Zwitserlood (1989)) also suggests that word-initial information is crucial for efficient word recognition.

p. 46

Dumay et al. (2002) propose that the Syllable Onset Segmental Heuristic may be a (spoken) language universal, noting that it is not incompatible with the Metrical Segmentation Strategy proposed for languages like English and Dutch: “segmenting the speech stream at strong syllables ... necessarily entails some syllable onset segmentation” (p. 157). But as Norris et al. (2001) caution, “[p]sycholinguistics has a long and embarrassing tradition of claims to language-universality based on data from English alone” (p. 652). Claims of the universality of the SOSH are therefore premature. Dumay et al.’s (2002) study on French would need to be complemented by studies on many other languages before such a proposal can be seriously entertained.

Syllable and word boundary mismatches in French

An obvious complication for any account of French speech segmentation that makes reference to the syllable is the well-known fact that syllable boundaries and word boundaries in the language frequently do not coincide. A number of processes lead to this mismatch: *liaison*, in which a latent (normally unpronounced) word-final consonant surfaces and is resyllabified as the onset of the vowel- or glide-initial syllable of the following word (e.g., *petit appartement*) [pə.ti.ʔa.paʁ.tə.mɑ̃]; *enchainment*, a process similar to *liaison*, except that it involves the resyllabification of a final consonant that is always pronounced (e.g., *petite orange* [pə.ti.ʔo.Rɑ̃ʒ]); and the addition of an epenthetic consonant, a process in which a consonant (usually [t])

is added to avoid vowel hiatus (e.g., *Comment va-t-elle?* [ko.mã#va.tɛl] ‘How is she?’). Other mismatches between syllable and content word boundaries commonly arise when function words are elided before vowel-initial content words as in *l’incendie* (from *le incendie*) ‘the fire’, *je t’écoute* (from *te écoute*) ‘I’m listening to you’, *elle s’ennuie* (from *se ennueie*) ‘she’s bored’. If the goal of speech segmentation is to help the listener identify words in the speech stream, then what good is it to find syllable onsets or offsets, if they don’t coincide with word boundaries?

A priori, this syllable/word boundary mismatch might be seen as an argument against reference to the syllable in French speech segmentation. But such a rejection of a role for the syllable on those grounds would be too hasty. First, it is generally agreed that most cues to speech segmentation do not determine word boundaries, but identify *possible* word boundaries. Identification of a syllable boundary (or onset) clearly must identify only a possible word boundary: most words have more than one syllable, so many syllable boundaries are word-internal. Second, there is evidence that there are acoustic cues that allow listeners to resolve the mismatches between syllable and word boundaries that arise from post-lexical resyllabification processes. Fougeron et al. (2002) found spectral and durational differences differentiating the critical consonants in all three conditions in minimal triplets like *cale égale* ‘equal wedge’ (enchainment consonant), *cas légal* ‘legal case’ (underlying word-initial consonant), and the nonword *qualégale* (word-internal syllable onset consonant) (the pronunciation of all three is [ka.le.gal]). This result is in line with the spectral and durational differences found by Laeuffer (1987), who concludes that resyllabification in enchainment is incomplete. Other studies have also examined acoustic differences between liaison versus underlying consonants, generally finding that liaison consonants

are shorter than underlyingly word-initial consonants in minimal pairs like *petit ami* ‘boyfriend’ vs. *petit tamis* ‘small sieve’ ([pə.ti.ta.mi]) (e.g., Dejean de la Bâtie (1993), Wauquier-Gravelines (1996), Yersin-Besson and Grosjean (1996), and Spinelli et al. (2003)).

A number of studies have shown that recognition of vowel-initial words is not slowed by the presence of a liaison or enchainment consonant. These findings suggest that listeners are sensitive to differences between underlyingly word-initial consonants and consonants arising through resyllabification processes, and exploit these differences in speech segmentation (Gaskell et al. (2002), Spinelli et al. (2002), Spinelli et al. (2003)). For example, in a cross-modal priming experiment, listener response times to a visual target like *italien* ‘Italian’ were equally fast for liaison (*un généreux italien* [ẽ ʒe.ne.RØ ʒi.ta.ljẽ] ‘a generous Italian’), enchainment (*un virtuose italien* [ẽ vir.tuø ʒi.ta.ljẽ] ‘an Italian virtuoso’), and non-resyllabified conditions (*un chapeau italien* [ẽ ʃa.po i.ta.ljẽ] ‘an Italian hat’). The results of a word-detection task actually showed a facilitatory effect of liaison and enchainment—targets in these conditions were detected faster than in the condition where there was no syllable misalignment. Interestingly Dejean de la Bâtie and Bradley (1995) found that *potential* liaison slowed participants’ performance in a phoneme-monitoring task. Responses were slow to items like *petit tableau* ‘small painting’ where the syllable-initial [t] could be either an underlyingly word-initial consonant or a liaison consonant (as in *excellent acteur* ‘excellent actor’). This effect disappeared, however, in a second experiment that replaced the original neutral carrier sentence *C’est un _____* ‘it is a _____’ with a biasing carrier phrase like *Dans ce film, j’ai découvert _____* ‘in this film, I discovered _____’, showing that listeners could use context to facilitate segmentation.

The suggestion that resyllabified input should pose little processing difficulty for the French listener seems eminently reasonable, even obvious. Liaison and the other resyllabification processes are pervasive, systematic phenomena in the language, and all available evidence shows that listeners are able to find patterns in their native language and use them as cues to speech segmentation. We should therefore expect that listeners would extract and exploit cues from resyllabification in the speech segmentation processes. For example, listeners surely have tacit knowledge of the fact that only five consonants are possible liaison consonants /p,t,n,r,z/²⁶ and accordingly do not treat other onset consonants as potential liaison consonants. In addition, while some liaisons are optional or a matter of style, others are obligatory, such as those in the determiner+noun pairs in (24) and the plural adjective+noun pairs in (25).

- (24) a. les élèves
 ‘the students’
 b. mon enfant
 ‘my child’
 c. huit Américains
 ‘eight Americans’
- (25) a. de bons amis
 ‘good friends’
 b. de vieux hôtels
 ‘old hotels’
 c. de nouveaux avions
 ‘new airplanes’

One might even expect that failure to produce the liaison consonant (by a non-native speaker, for example) in obligatory contexts like these would hinder speech segmentation and word recognition. Furthermore, listeners are likely to have a fine-grained knowledge of the relative probability of liaison in different contexts.

²⁶Liaison with /p/ is possible for only two words, *beaucoup* ‘a lot, very much’ and *trop* ‘too (much)’. Liaison with these adverbs is not obligatory: one hears, for example, both [bo.ku.pɛ.me] and [bo.ku.ɛ.me] for *beaucoup aimé*.

In short, the existence of resyllabification processes in French does not force us to abandon reference to the syllable in French speech segmentation. Doing so would underestimate the ensemble of tacit information that listeners have at their disposal in solving the speech segmentation problem.

Moraic cues

The syllable may be an important speech segmentation unit, but some researchers have suggested that the mora, rather than the syllable, is the relevant unit in Japanese speech segmentation (Otake et al. (1993), Cutler and Otake (1994), Otake et al. (1996), McQueen et al. (2001)). The mora is a subsyllabic unit with five different forms: V (e.g., /a/); CV (e.g., /ta/); CCV (e.g., /gja/)²⁷; N, a nasal coda consonant (e.g., *Honda* /ho- N.- da/); and Q, the moraic first part of a geminate consonant (e.g., *Nippon* /ni- Q.- po-N/, where the first part of the geminate is the coda of the first syllable and the second part is the onset of the second syllable) (McQueen et al. (2001)).²⁸ McQueen et al. (2001) used the word-spotting paradigm to investigate whether Japanese listeners used mora boundaries as cues to speech segmentation. In one experiment, they found that listeners were significantly better at spotting words embedded in vowel contexts (e.g., the *agura* ‘to sit cross-legged’ in [oagura]) than those embedded in consonant contexts (e.g., [tagura]). In another experiment, listeners were no better at spotting words in vowel contexts (e.g., the *saru* ‘monkey’ in [sarua]) than in moraic nasal contexts (e.g., [saruN]). A third experiment included all three conditions (vowel context (e.g., the *uni* ‘sea urchin’ in [gjaouni]), moraic nasal

²⁷Note that this syllable structure is possible if the onsets of these syllables are analyzed as clusters rather than as palatalized consonants.

²⁸Following McQueen et al. (2001), periods indicate syllable boundaries and hyphens mora boundaries.

context (e.g., [gjaNuni]), and consonant context (e.g., [gjabuni])). Listeners had no difficulty in spotting words in the vowel and moraic nasal contexts (response times were actually faster in the moraic nasal contexts, although error rates were higher in that condition. Both these differences were significant only by subjects). Listeners, however, found it nearly impossible to spot words in the consonant context condition. McQueen et al. (2001) note that listeners spot words quickly in cases where a word boundary and a mora boundary are aligned, i.e., the vowel and moraic nasal contexts (e.g., [o-a-gu-ra], [sa-ru-a], [gja-o-u-ni] and [sa-ru-N], [gja-N-u-ni-ŋ]). Listeners had difficulty in just those cases where the boundaries were misaligned (e.g. [ta-gu-ra], [gja-bu-ni]). They interpret these results as evidence that “segmentation in Japanese is based on the mora” (pp. 123).

An alternate account for this pattern of results is that Japanese listeners are sensitive to the distribution of segment classes in their language, they have tacit knowledge of possible syllable onsets and codas, and they make use of this knowledge in their parsing decisions. They know, for example, that moraic nasals cannot appear syllable-initially or word-initially,²⁹ while non-geminate stops can only appear syllable-initially. This tacit knowledge of the distribution of moraic nasals excludes a word boundary after the first vowel in *gja#N.u.ni, just as knowledge of the distribution of non-geminate stops excludes the word boundary in *gja.b#u.ni. This explanation accounts for the pattern of results in McQueen et al. (2001) and other studies examining the use of the mora in Japanese speech segmentation, but does not require that the listener explicitly parse the speech input into moras.

²⁹The only potential exceptions to this general rule are the words [N] ‘uh-huh, yes’ and [NN.N] ‘uh-uh, no’.

In addition, the results of a developmental study suggest that evidence for mora-based segmentation strategies may be influenced by adult listeners' knowledge of the *kana* writing system, which transparently represents moraic structure (Yoneyama et al., in preparation). Pre-literate children in the study did not show the same “mora effect” as (literate) adults in other studies.

4.2.3 Generalizations on the nature of speech segmentation

While the exact nature of cues to speech segmentation is far from resolved, it is certain that such cues do exist and that listeners exploit them. As discussed in §4.2.2, there is a consensus among researchers that these cues do not *determine* word boundaries, but rather identify *possible* word boundaries. For example, English-speaking listeners are able to recognize words beginning with weak syllables (*e'xample*, for example), even though the beginnings of such words are not identified by the Metrical Segmentation Strategy. And obviously, not all syllable boundaries coincide with word boundaries, so any segmentation strategy based on syllables or syllable onsets cannot be deterministic.

In addition, the evidence also suggests that listeners cannot “turn off” their native segmentation strategies, that they use these strategies in the processing of a non-native language. As noted above, Dutch listeners used their native language sensitivity to correlates of stress other than vowel reduction in a task with English language materials (Cooper et al. (2003)). And whether or not their segmentation strategies are appropriately labelled “syllable-based” and “mora-based,” French and Japanese listeners show native-language processing effects when listening to foreign language stimuli (e.g., Cutler et al. (1986), Otake et al. (1993), Cutler and Otake (1994), Otake

et al. (1996)). The results of Weber (2000) show that listeners employ their native language strategies in listening to a foreign language, even when these strategies are ineffective and lead the listener to hypothesize inappropriate word boundaries. For example, in a word-spotting task with German participants and English materials, German phonotactics provided a stronger cue to word segmentation than did English phonotactics. Words were spotted faster when their onsets were aligned to syllable boundaries forced by German phonotactics (e.g., the *luck* in [mɔɪs.lʌk], where the [sl] sequence requires a syllable boundary) rather than by English phonotactics (e.g., [gʌʃ.lʌk], where the [ʃl] sequence requires a syllable boundary, at least in some dialects). The results of this study also show, however, that the phonotactics of a non-native language may influence speech segmentation. Although their native language phonotactic cues provided the strongest segmentation cues, native German listeners proficient in English were able to spot English words embedded in nonsense words faster when those words were aligned with English phonotactic boundaries than with no phonotactically required boundary in either language (Weber (2000)).

4.2.4 The Possible Word Constraint

Of course, the problem of word recognition does not end with the identification of possible word boundaries. The Possible Word Constraint (PWC) model offers an integrated account of the role that cues to word boundary locations might play in a complete model of word recognition (Norris et al. (1997)). In line with earlier research (Cutler et al. (1992)), the model assumes that listeners in all (spoken) languages use the same *types* of cues to hypothesize word boundaries and that cues to speech segmentation are redundant. In addition to the phonotactic and prosodic cues

discussed, these cues include presence of leading or following silence (Norris et al. (1997)), knowledge of the distribution of allophones such aspirated stops in English (Lehiste (1960), Nakatani and Dukes (1977)), and durational cues (Lehiste (1972), Beckman and Edwards (1990), Banel and Bacri (1994)). The PWC model assumes that the specifics of these cues differ from language to language. For example, since English syllables cannot begin or end with the consonant cluster [pn], listeners will use this phonotactic constraint to hypothesize a syllable boundary when a [pn] sequence is encountered ([p.n]). French speakers, however, have no such constraint, since [pn] is a possible onset in their language (e.g., *pneu* ‘tire’). According to the PWC model, the extent to which listeners rely on each type of cue may vary across languages. In addition, not all types of cues will necessarily be present in the signal, but if they are, listeners will use them.

The PWC model follows the standard assumption that word recognition proceeds through the activation of multiple candidates that compete with one another during word recognition, with a winning candidate finally emerging (see Norris et al. (1995) and references therein for a discussion of competition models). Language-specific cues to speech segmentation operate to identify potential word boundaries. Candidate parses that are misaligned with hypothesized word boundaries are penalized. The Possible Word Constraint eliminates parses of the input string that leave residues that are not possible words. Evidence for the operation of this constraint comes from word-spotting experiments that have found that listeners have more difficulty in spotting, for example, the word *apple* in *fapple*, where the residue is a single consonant ([f]) and therefore not a possible word than in *vuffapple*, where the residue is a possible nonword (*vuff*).

Misalignment of syllable and word boundaries in French clearly poses a challenge for the Possible Word Constraint, as the formulators of the hypothesis recognize: “[i]ronically perhaps, notorious examples of resyllabification occur in French, which is a syllable-timed language with clear syllabification, so that all syllabification should be relatively clearly marked in the signal” (p. 237). They note that for an input like *petit éléphant* [pə.ti.te.le.fã], *éléphant* should be penalized because of the misaligned syllable/word boundary. They suggest that the PWC is “turned off” in these cases, arguing:

[T]he liaison process is highly systematic in a way that should allow the PWC penalty to be disabled in contexts where liaison is likely... [T]he plausibility of disabling the PWC in liaison contexts is supported by the fact that the set of environments in which liaison can occur is very restricted in French. This is a further example of the necessity for language-specificity in implementing the input features to which the PWC is sensitive.

p. 237

Some researchers have interpreted research on the PWC as supporting a claim that the constraint is not sensitive to what counts as a well-formed word *in a particular language*. Rather, it is claimed to be a (spoken) language-universal constraint that forbids parses that strand single consonants that could not be free-standing words in any language.³⁰ This proposed universal is not inconsistent with the evidence that listeners use language-specific patterns in the identification of possible word boundaries, as the authors of the PWC themselves explicitly state. Potential evidence for the generality of the PWC comes from experiments on English, Dutch, and the Bantu language Sesotho. For example, English words cannot end with certain lax

³⁰As the proposers of the Possible Word Constraint note, this may not be true of Imzadn Tashlhit Berber, an Afro-Asiatic language spoken in Morocco, in which single consonants have been claimed to count as syllable nuclei. See Dell and Elmedlaoui (1996), Coleman (1996), and Louali and Puech (1999).

vowels. While *deck* [dɛk] is a word in the language, the sequence [dɛ], with no coda consonant, could not be a well-formed English word. Despite this well-formedness violation, English listeners do not treat residue like [dɛ] in the same way that they do residue consisting of a single consonant. They are no slower to spot the real word *perturb* in *dɛperturb*, where the residue [dɛ] is not a possible word, than they are in *daperturb*, where the residue [da] is a possible word. They are much slower to spot the real word in *sperturb*, where the residue [s] is a single consonant (Norris et al. (2001)). Similarly, although words in the Bantu language Sesotho must contain at least two syllables, monosyllabic residue is not treated in the same way as single-consonant residue by Sesotho listeners in speech segmentation. The real word *alafa* ‘to prescribe’ is as easy to spot in [roalafa], where the monosyllabic residue [ro] is not a possible word, as in [pafoalafa], where the disyllabic residue [pafo] is a possible word. The real word is harder to spot in [halafa], where the residue is a single consonant (Cutler et al. (2002)).

While these results are interesting, they can be explained without the Possible Word Constraint. For example, it is unsurprising that English listeners are slow to spot *perturb* in *sperturb*, since the /p/ in the embedded word is unaspirated in the fricative+stop cluster and listeners expect an aspirated stop syllable- (and word-) initially. We have long known that listeners are sensitive to precisely this type of allophonic variation (Lehiste (1960), Nakatani and Dukes (1977)).

In addition, many future studies are needed before we can realistically propose a language universal. Further evidence should, of course, come from many different languages, but should also include careful examination of the treatment of many different patterns within a single language. It is also important not to underestimate the

potential influence of language-specific phonology on the Possible Word Constraint. For example, in Japanese, the nonword [flaku] would be parsed as /huraku/, with the single consonant residue [f] parsed as /hu/ (or possibly / ϕ u/) and potentially posing no problem to listeners in spotting the real word *raku* ‘easy’.³¹

4.2.5 Weighting of cues to speech segmentation

While many researchers accept that there are multiple cues to word boundaries, very little research has examined the interaction of these different types of cues. A recent study on English does just that. In a series of three cross-modal fragment priming experiments, Mattys (2003) examines the weighting of four different cues to speech segmentation (stress patterns, phonotactics, coarticulation, and lexical information) in clear speech and in the presence of noise. The auditory test primes were the first two syllables of three-syllable stress-initial or non-stress-initial words (e.g., [ˈkʌstə] from *customer* and [kəˈθi] from *cathedral*). These primes were embedded at the end of sequences of nonsense syllables. For example, in the experiment that tested the influence of phonotactic cues versus stress cues, the test utterances included [ˈbesɪwɒmkʌstə], [ˈbesɪwɒnkʌstə], [besɪwɒmkəˈθi], and [besɪwɒnkəˈθi]. The visual targets were written words (e.g., *customer* and *cathedral*). Participants listened to each context-prime sequence, saw the written target, and decided whether the target was a real word.

The Metrical Segmentation Strategy account (see §4.2.2) predicts that response times to *customer*, which begins with a stressed syllable ([ˈkʌstə]), will be faster than those to *cathedral*, which does not [kəˈθi]. Listeners are sensitive to the fact that most English content words begin with stressed syllables. The stress information,

³¹The example is due to Mary Beckman and Kiwako Ito.

however, is in conflict with phonotactic information in the case of ['besɪwomkʌstə]. The sequence [mk] is a very uncommon word-internal sequence in the language, as are most non-homorganic nasal-stop sequences, and there is likely to be a word boundary separating the two segments [m#k]. The results of the experiment showed a clear difference between the two listening conditions in the weighting of the two cues. In clear speech, there was a main effect of phonotactics—listeners were faster in the low-frequency biphone sequence condition than in the high frequency condition—but no effect of stress. The pattern was reversed for speech in noise (reported as SNR = 0, but presumably SNR = 1). There was a main effect of stress pattern—as predicted by the MSS, listeners were faster to find words that began with stressed syllables—but no effect of phonotactics. Two other experiments examining the interaction of stress cues with coarticulation cues and with lexical cues (i.e., whether or not the preceding syllable string forms a real word) found similar results: listeners rely on stress only in the presence of noise. As Mattys (2003) puts it:

These results highlight the need to study speech segmentation as an integrated phenomenon, in which segmentation power is not equally distributed between previously-documented cues. Instead, cues are hierarchically organized, and the hierarchy can be altered as a function of the quantity and quality of lexical, segmental, and acoustic information available in the signal.

Although it does not directly follow from the Mattys (2003) results, it also seems clear that the weighting of different types of cues might change over the course of language acquisition. Certain cues might be more important to infants, children or adults. As Mattys himself has shown in earlier work (Mattys et al. (1999)), nine-month old infants are sensitive to both phonotactic and prosodic information, but when the two types of cues are pitted against each other, the influence of prosody is stronger than that of phonotactics. The Monte Carlo simulation in Pierrehumbert

(2001b) predicts this disparity—in that study, the regularity of English word stress was learnable even with the smallest vocabulary in the experiment (400 monomorphemic words), while the phonotactic constraint examined was perfectly learned only with a much larger vocabulary (3200 words). Although Mattys et al. (1999) studied infants, the observation that robust generalizations about phonotactic patterns require a larger vocabulary than those about prosodic patterns is still relevant.

4.3 Prosodic cues to speech segmentation in French

The two perception experiments described in the next chapter focus on how French listeners might use the presence and timing of intonational rises as cues to word boundaries. To date very few studies have examined French listeners' use of prosodic patterns as cues to speech segmentation.

Earlier research on French has established that the last syllable of a phrase or a word spoken in isolation is lengthened and has special prominence (see Chapter 2). Banel and Bacri (1994) investigated the influence of this pattern on segmentation. They constructed stimuli pairs ambiguous in interpretation between a single, disyllabic word and a sequence of two monosyllabic words (e.g., *bagage* 'luggage' vs. *bas* 'low' + *gage* 'pledge'). Two types of metrical patterns, an iambic (short-long) and a trochaic (long-short) pattern were constructed by concatenating repetitions of monosyllabic words. Monosyllabic words were read both in isolation and as the initial or final member of a pair, pronounced in a monotone to control for the possible influence of intonational cues. Banel and Bacri (1994) give the following example: a short *bas* (290 ms) was concatenated with a long *gage* (810 ms) to create an iambic version of [bagaʒ], and a long *bas* (440 ms) was concatenated with a

short *gage* (530 ms) to create a trochaic version. Since the end of a prosodic phrase is marked by lengthening, the iambic pattern (short-long) is the expected pattern within a prosodic phrase. Participants listened to the disyllabic stimuli and indicated whether they heard one word or more than one word. As expected, listeners were more likely to hear one word (e.g., *bagage*) for stimuli in the iambic condition (short-long) and two words (*bas, gage*) in the trochaic condition (long-short). This finding corresponds to the expectation that a phrase-final syllable will be lengthened. The lengthening of the first syllable in the trochaic (long-short) pattern signals a phrase boundary. Since phrase boundaries do not typically occur in the middle of a single word, only the two-word interpretation is consistent with this interpretation. The main effect of metrical pattern was large—listeners gave a “multi-word” response to only 27% of the iambic pattern (short-long) stimuli, versus 46% of the trochaic pattern (long-short) stimuli. But even given a trochaic (long-short) pattern, listeners showed a bias for the “one-word” response, giving “two-word” responses less than half the time (46% of responses). This response bias is almost certainly linked to the fact that the two-word interpretation is usually neither semantically nor syntactically well-formed.

Rietveld (1980) examined the phonetic differences between minimal pairs like those in (26). Rietveld notes that in his data, “... the fundamental frequency showed a rise on the last syllable of the Noun in the NP” (p. 295).

- (26) a. Le comtat saccagé sera en proie à la famine.
 The county devastated will be in prey to the famine
 ‘The devastated county will be beset by famine.’
 b. Le comte a saccagé une grande partie de la ville.
 The count has laid waste a large part of the city
 ‘The count laid waste to a large part of the city.’

The observed differences in f_0 , intensity, and duration correspond to a primary accent and late rise on the final syllable [ta] of *le comtat* ([lə kɔ̃.ta]) in (26a) and a primary accent and late rise on the [kɔ̃] of *le comte a* ([lə kɔ̃.ta]) in (26b), with a fall across the following [ta].³² The [a] of *comtat* in (26a) is significantly longer than the [a] of the auxiliary verb *a* ‘has’ in (26b), which is expected for a prosodic phrase-final syllable.

The results of a perception test showed that French listeners could reliably identify the source sentence of fragments like the underlined phrases in (26) and that durational differences were the most reliable cue.

In addition, Rietveld compared the durations of the noun-final [a] in the phrases *le comtat saccagé* and *le syndicat sommé* ‘the summoned union’, the [a] of the auxiliary in the phrases *le comte a saccagé* and *le syndic a sommé* ‘the president of the apartment building association summoned’, and the two word-medial [a]s of *ambassador* ([ã.ba.sa.dɔ̃ʁ]), which was also in the corpus. Noun-final [a] was significantly longer (at 116.3 ms) than both auxiliary [a] (79.7 ms) and word-medial [a] (85.2 ms).

Rietveld interprets his results as evidence that the “opinion proposed by many linguists and phoneticians—that the breath-group ‘de-stresses’ the words it contains—[is] not confirmed ...” (p. 295). Christophe (1993) follows him in this interpretation:

Ce résultat n’a rien de trivial, dans la mesure où l’opinion est assez répandue parmi les linguistes et les phonéticiens que les mots français perdent leur identité (et leur accent) au profit d’une unité prosodique plus importante, appelée “groupe rythmique” (par exemple, “muscat perdu” formerait un seul groupe rythmique, et par conséquent la syllabe “ca” ne serait pas accentuée).

‘There is nothing trivial about this result, to the extent to which the opinion is fairly wide-spread among linguists and phoneticians that French words lose

³²I assume that there is enchainment of the coda [t] of *comte* in *le comte a*, that the [t] is resyllabified as the onset of the following syllable. A pronunciation without enchainment is also possible, with a glottal stop before the [a].

their identity (and their accent) for the benefit of a larger prosodic unit, called the “rhythmic group” (for example, *muscat perdu* would form a single rhythmic group, and as a result, the syllable ‘ca’ [ka] would not be accented).’

p. 41, footnote 23

Yet Rietveld’s own description of an f_0 rise on the noun-final syllables of his experimental items suggests that *le muscat* and the other article+noun pairs in his corpus form prosodic phrases (Jun and Fougeron’s APs) to the exclusion of their following adjectives. Given that, the lengthening of the noun-final and, crucially, phrase-final syllables is part of the marking of the primary accent.

Christophe (1993) tested whether French adults and French infants were sensitive to prosodic differences in their native (or to-be-native) language (see also Christophe et al. (1994)). One experiment used repetitions of the bisyllabic sequence [ma.ti] excised from intra-word contexts like *mathématicien* and inter-word contexts like *panama tisane* ‘panama (hat) herbal tea’. Tokens in the intra-word and inter-word condition were significantly different along several dimensions: the [a] was longer and had higher intensity and higher fundamental frequency in the inter-word condition, and the [t] closure was longer in the inter-word condition. Adults listeners were able to correctly categorize the French [ma.ti] tokens, assigning tokens to a non-linguistic category (*noir* ‘black’ or *blanc* ‘white’) at a rate better than chance. This result suggests that the adults were able to exploit the phonetic cues available in the signal.

In addition, results of an experiment using the high-amplitude sucking paradigm showed that three-day old French-acquiring infants were able to distinguish [ma.ti] tokens taken from intra-word contexts from those taken from inter-word contexts. The infants first listened to a stimulus series (either intra- or inter-word [ma.ti]) until they were habituated to that series. The stimuli then switched to a second series,

either to a series of the other condition (an experimental condition, e.g., INTRA \rightarrow INTER) or to a series of physically different tokens from the same condition (a control condition, e.g., INTRA1 \rightarrow INTRA2). The rate of sucking was significantly higher in the experimental conditions, in the fourth minute after the change in stimulus series.

Other experiments extended the results to categorization by adults of phonemically varied stimuli and of non-native language stimuli. English adults, for example, were able to correctly categorize the French [ma.ti] stimuli. French adults were able to categorize Spanish intra- and inter-word [la.ti] tokens, even though the phonetic differences between categories did not exactly mirror those found for French.

At least two interpretations of these results are available. It could be, as Christophe suggests, that the newborns, whatever their language environment, exploit word boundary cues universal to spoken languages to discover the difference between the two [ma.ti] conditions. Christophe (1993) concludes that “it seems that certain prosodic cues are universal (lengthening of the first consonant of a word), while others are language-specific” (p. 112) and suggests that French newborns are likely to be able to discriminate intra-word and inter-word minimal pairs in languages other than French:

La similitude de situation entre le cas du français et celui de l’espagnol (observation de variations prosodiques systématiques, capacité d’adultes [à] catégoriser les stimuli) nous rend optimistes sur la capacité de nouveau-nés français à discriminer les stimuli espagnols.

‘The similarity between the French case and the Spanish case (observation of systematic prosodic variations, ability of adults to categorize the stimuli), makes us optimistic about the ability of French newborns to discriminate the Spanish [intra- and inter-word [la.ti]] stimuli.’³³

p. 112

³³Due to organizational changes at the hospital where the infant experiments were run, Christophe was unable to run the crucial experiment with French infants and Spanish stimuli.

It seems likely, however, that three-day old infants are sensitive to a pattern in their to-be-native language. Even at this tender age, they may have already learned enough about the prosodic patterns of the ambient language to help them distinguish between the two types of [ma.ti]. It is well established that newborns can distinguish their to-be-native language from a foreign language. Infants as young as two days have been shown to prefer listening to their to-be-native language (Moon et al. (1993)). Four-day old French-acquiring infants can distinguish between French and Russian (Mehler et al. (1988)). In particular, newborns seem to be sensitive specifically to the prosody of their to-be-native language—French-acquiring infants no older than five days were able to distinguish low-pass filtered French stimuli from English stimuli and from Japanese stimuli (Nazzi et al. (1998)). Low-pass filtering preserves prosodic cues such as intonation and duration, but removes most segmental information. In the Nazzi et al. (1998) study, French-acquiring newborns were unable to distinguish low-pass filtered stimuli from two prosodically similar languages, Dutch and English. They could, however, distinguish Dutch and English from Spanish and Italian, prosodically similar languages.

Under either interpretation, the Christophe (1993) results show that newborn infants can use prosodic cues in speech segmentation. Exactly which of the available cues they use (duration, intensity, f_0) remains unclear. In particular, we can not make any conclusions about the potential contribution of f_0 , since intonation patterns likely varied within conditions. All [ma.ti] tokens were read in the carrier sentence in (27) and excised.

- (27) A la radio, on parle de _____, du moins je le crois.
 ‘On the radio, they talk about _____, at least I think they do.’

It seems likely that a four-syllable target like (28a) could be uttered as one prosodic phrase, with a primary accent and late rise on the last syllable [pik], while a six-syllable target like (28b) would typically be uttered as two prosodic phrases, with primary accents and late rises on both [ma] and [ɛsk].

- (28) a. lama typique ‘typical lama (Buddhist monk)’
 b. cinéma titanésque ‘massive cinema’

Furthermore, even if the infants did use intonational cues in the experiments, these cues are not available in their full, intact form, since the stimuli are excised from their carrier sentences.

Fougeron (in preparation) similarly exploited the potential ambiguity between a series of monosyllabic words and a single polysyllabic word. She conducted a series of experiments with Swiss French listeners testing the hypothesis that the presence of a rise on the first syllable of ambiguous sequences like *main-te-nue* ‘maintained’ signals a word boundary to the right of the rise (*mains* | *tenues*) ‘hands held’, and biases listeners to a two-word interpretation. Participants listened to ambiguous three-syllable items like *main-te-nue* and indicated whether each item had been taken from a sentence corresponding to a two-word interpretation (*Le pantin a les mains tenues par les fils* ‘The hands of the jumping-jack toy are held by the strings’) or one corresponding to a one-word interpretation (*La proposition est maintenue par l’assemblée* ‘The proposal is upheld by the assembly’). Contrary to the hypothesis, listeners perceived patterns with a rise across the first syllable *main* and a rise across the last syllable *nue* as a single word, *maintenue*. One interpretation of this finding is that listeners did in fact use the presence of a rise on *main(s)* in their segmentation decisions. They may have interpreted the rise as the early rise in a LHLH pattern. As Fougeron points out (p.c.), the dispreference for the two-word parse may be influenced

by the rarity of very short accentual phrases (there is likely a phrase boundary after *mains* for the two-word segmentation *mains tenues*).

Another possibility is that fundamental frequency is not an essential cue to speech segmentation in French. As Wenk and Wioland (1982) note, for Delattre “frequency variation is no more than a possible (‘indispensable’) carrier of accent in French” (p. 196, referring to Delattre (1951), p. 45). While they themselves do not “deny[] the influence of pitch variation in determining group limits,” Wenk and Wioland (1982) argue that increased duration of phrase-final syllables is “the most dependable manifestation of French accent” (p. 215) and refer to the “length accents of French” (pp. 196, 202).

4.3.1 The early intonational rise as a cue to speech segmentation

Other researchers have suggested that listeners may exploit the early intonational rise as a cue to speech segmentation. Observing that the early rise is typically realized at the beginning of a content word, Di Cristo (2000) claims that one of the roles of the early rise is to serve as a cue to speech segmentation:

Dans cette perspective, l’actualisation de l’accent initial assume une double fonction, car il participe au rythme et il signale le début de l’unité lexicale à laquelle il est affecté. L’accent finale participe également à la construction du rythme et démarque la fin de l’unité lexicale qui le reçoit.

‘According to this perspective, the realization of the initial accent assumes a double function, because it participates in the rhythm and signals the beginning of the lexical unit to which it is assigned. The final accent also participates in the construction of the rhythm and demarcates the end of the lexical unit that bears it.’

p. 39

Vaissière (1997) also suggests a role for intonational patterns in speech segmentation, emphasizing the importance of tonal alignment and specifically citing the fundamental frequency elbow of the early rise as a boundary marker. She proposes that the alternation between low and high fundamental frequency in French may help the listener to identify word boundaries and draws a parallel between the function of this alternation in French and that of the alternation in English between strong and weak syllables:

C'est l'alternance régulière de hauteur mélodique entre mots lexicaux et mots grammaticaux qui nous semble une caractéristique essentielle du français (au même titre que l'alternance entre syllabes accentuées et syllabes réduites en anglais). Alors que les mots lexicaux sont en général marqués par un passage (au moins) dans le registre haut du locuteur, les mots grammaticaux ... sont caractérisés par une valeur cible basse de f_0 [...] A la présence de mots grammaticaux correspond donc une échancrure du fondamental..., qui est un indice de rupture: la valeur basse, marque de frontière, s'associe avec le mot grammatical.

'It is the regular alternation in melodic height between content words and function words that seems to us to be an essential characteristic of French (along the same lines as the alternation between strong syllables and weak syllables. While content words are in general marked by a passage (at least) in the high register of the speaker, function words ... are characterized by a low target f_0 value[...]. Corresponding to the presence of function words, there is an elbow ..., which is a cue to a break: the low value, a boundary marker, is associated with the function word.'

pp. 66, 67

This proposed cue differs in two important respects from the cues discussed above. First, it would identify actual word boundaries rather than simply potential word boundaries. Second, it would mark the boundary between a content word and a preceding function word. A number of studies have suggested that the successful identification of function words plays an important role in both adult and infant speech segmentation (e.g., Christophe et al. (1997)). The alignment of the early

rise is therefore potentially a very powerful cue, but one that has so far not been empirically tested. The perception experiments described in the next chapter aim to partially fill this void.

CHAPTER 5

THE EARLY RISE AS A CUE TO SPEECH SEGMENTATION

Early to bed and early to rise
Makes a content word easy to excise.

Benjamin Franklin (sort of)

5.1 An investigation of the role of the early rise in speech segmentation

Three perception experiments were conducted to explore the possible use of intonational cues in French speech segmentation. Reports in the literature and the production experiment findings allowed us to develop hypotheses about the potential role of intonation cues.

5.2 Perception Experiment 1: Presence of the early rise as a cue to speech segmentation (Speech masked by noise)

This experiment tested the hypothesis that listeners use the presence of an early rise as a cue to the beginning of a content word. Previous studies have established that early rises are rarely realized across proclitic function word syllables. Proclitics differ from enclitics in the realization of intonational rises. The late rise is frequently

realized on enclitic pronouns (e.g., *regarde-LA* ‘look at her’ (see, for example, Mertens (1993), Delais-Roussarie (1995), and Delais-Roussarie (1999)).

There are some exceptions to the generalization that early rises are not realized on proclitics. Jun and Fougeron (2002) suggest that an early rise is likely to be realized in order to avoid unaccented clitic sequences of over four or five syllables, although long sequences of function words are rare in the language. Function words may take an early rise in metalinguistic negation, as Fonagy (1979) notes (e.g., *On ne dit pas LE table, on dit LA table!* ‘You don’t say LE table, you say LA table!’). We can think of this as a kind of focus accenting in which the function word is promoted to the status of a pseudo-content word. Early rises are more common across disyllabic function words and certain classes of monosyllabic proclitics (pronouns like *moi* and *lui* and negative adverbs like *pas*) (Delais-Roussarie (1995)). In addition, the early rise can sometimes appear on determiners in enumerations or lists.³⁴ In addition, it is very common in the speaking style of television newscasters, who often realize it on determiners (Fonagy and Fonagy (1976), Fonagy (1979), Vaissière (1983)). Despite these exceptions, an early rise is much more likely to occur on a content word than a function word. Furthermore, the results of the production experiment showed that the low starting point of the early rise began consistently at the boundary between a function word and a content word. We suspected that listeners might exploit this regularity in speech segmentation. A perception experiment tested the hypothesis in (29).

(29) PRESENCE OF EARLY RISE AS CUE TO CONTENT WORD HYPOTHESIS: Listeners interpret the presence of an early rise as a cue to a content word beginning.

³⁴An example is a promo for a television show heard at the end of the summer of 2003: “Après LA rentrée, LE boulot, LES impôts, vous n’avez qu’une seule envie — vous marrer!” ‘After the end of summer vacation, work, taxes, you just want one thing — to laugh!’

A noise-masking paradigm previously used to study cues to word segmentation in English was adapted. Smith et al. (1989) presented English-speaking listeners with six-syllable phrases embedded in noise. Even in the absence of most segmental information, listeners correctly identified the rhythmic patterns of the stimuli. In a forced choice task, they reliably selected items that matched the pattern of strong and weak syllables in the stimuli. For example, given noise-masked *rust prevents a nuisance*, with a SWSWSW rhythmic pattern, listeners preferred the rhythmic and word boundary match *crushed defence was useless* over mismatched *the crust seldom improves* (WSSWWS). When a choice matching in both rhythm and word boundary location was available, listeners reliably chose the exact match over an item that matched only in stress (for example, *tough expense misused them* matches *rust prevents a nuisance* in stress pattern (SWSWSW), but not in word boundary location). These results showed that listeners were sensitive to speech rhythm and that they were able to use prosodic cues (duration, fundamental frequency, intensity or some combination) to locate the positions of word boundaries.

5.2.1 Methods

Materials

Corpus design and recording

Stimulus materials were twenty-four pairs of phrases like the ones in (30).

- (30) a. le niveau de mes sénats
 lə nivo də me se.na
 ‘the level of my senates’
- b. le niveau de mécénat
 lə nivo də me.se.na
 ‘the level of patronage’

Each phrase contained seven syllables, with the last four syllables constituting the target sequence. One member of the pair contained a target sequence with two monosyllabic function word syllables followed by a two-syllable content word (*de mes sénats* ‘of my senates’ [də.me.se.na]). The second member of the pair contained a target sequence with one monosyllabic function word syllable followed by a three-syllable content word (*de mécénat* ‘of patronage’ [də.me.se.na]). These two segmentations will be referred to as the two-word segmentation (*mes sénats*) and the one-word segmentation (*mécénat*). An early rise beginning at the second syllable of the phrase *de mécénat* is possible, since that syllable is the initial syllable of a content word. By contrast, an early rise is unlikely in the case of *de mes sénats*, since *mes* is a function word, and early rises are rarely realized on function words. Although the phrases contained real words, they were designed to be nonsensical. This minimized the chance that listener segmentation preferences would be influenced by context.

The requirements of the experimental materials made it impossible to include only pairs of items that were exact minimal pairs. The construction of exact minimal pairs requires finding a three-syllable content word whose first syllable is interpretable as a determiner (*la, les, ma, mes, mon*, etc.) and whose last two syllables form either a noun or an adjective. Of the 24 pairs, therefore, only eight were exact minimal pairs (e.g., *le niveau de mes sénats* / *le niveau de mécénat* and *la montagne a nos minets* / *la montagne a nominé*). Eight pairs differed in only one phoneme (e.g., *positif et mes dictées* / *positif et médité* and *original et ma carie* / *original et macareux*), and eight pairs differed by more than one phoneme (e.g., *mélomane et ma lagune* / *mélomane et malaga* and *la baleine et mes lamées* / *la baleine et Mélanie*). It was hoped that despite the inevitable imperfections in the segmental structure of the stimuli, the

presence of an early rise would be a strong enough cue to the beginning of a content word to provide support for the hypothesis. This seemed reasonable since the stimuli were presented masked in noise, which eliminates many cues to segmental identity.

These phrases were inserted into carrier sentences like the one given in (31), read by a female native speaker of Parisian French in a sound attenuated chamber and recorded onto digital audio tape (DAT), sampling at 48 kHz with 16-bit resolution.

(31) Ils écrivent X dans les marges.
'They write X in the margin.'

Carrier phrases were used in order to elicit target sequences that ended with a late rise, since the primary (obligatory) accent is realized as a fall utterance-finally. In the initial recording sessions, the speaker consistently produced the pattern in Figure 5.1b for the two-word segmentation (*mes sénats*), with a late rise on the last syllable of the target sequence and no early rise. For the one-word segmentation (*mécénat*), she sometimes produced an early rise and sometimes produced only the late rise. This was not unexpected, since the early rise is optional. Since the hypothesis to be tested examines the role of the early rise, the speaker was prompted in subsequent recording sessions to produce an early rise in the one-word segmentation condition, as in Figure 5.1a.

Recordings were transferred from DAT to computer using a digital-to-digital cable and CoolEdit software. Each item was saved as a separate file and the individual files were downsampled to 22.05 kHz using Praat software (Boersma and Weenink (1992–2001)). All further phonetic analyses and manipulations were performed using Praat. Fundamental frequency curves were created and inspected to ensure that each phrase had been produced as intended.

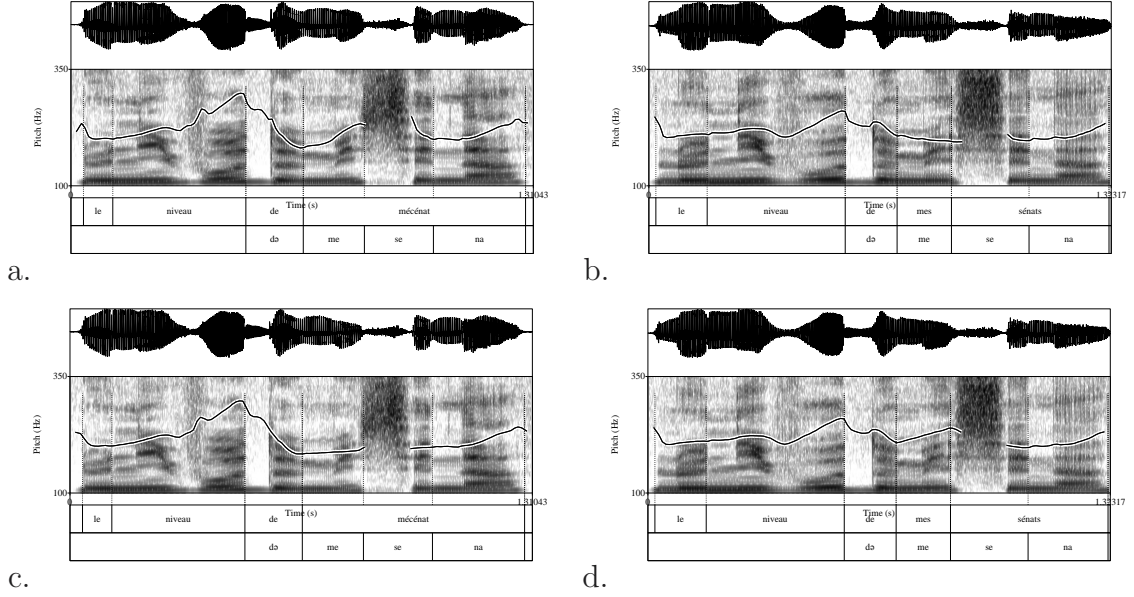


Figure 5.1: *Sample materials for Perception Experiment 1. (a) Target produced as the single three-syllable content word mécnat (natural production). (b) Target produced as a function word followed by a two-syllable content word mes sénats. (natural production). (c) Target produced as a single three-syllable content word mécnat. Fundamental frequency resynthesized (from the original in panel a) to delete the early rise. The new pattern is hypothesized to be consistent with a segmentation with a content word beginning at [se] (e.g., mes sénats). (d) Target produced as a function word followed by a two-syllable content word mes sénats. Fundamental frequency resynthesized (from the original in panel c) to add an early rise. This pattern is hypothesized to favor a segmentation with a content word beginning at [me] (e.g., mécnat). Note the similarity in f_0 between panels a and d and between panels b and c.*

Fundamental frequency resynthesis

Fundamental frequency resynthesis using the PSOLA (pitch-synchronous overlap and add) technique was performed to create two additional conditions for each pair (for a description of the technique, see Moulines and Charpentier (1990)). For the one-word segmentation items, the early rise was removed and for the two-word segmentation items, an early rise was added. There were therefore four files for each pair: two natural files and two files with resynthesized f_0 , corresponding to four conditions as illustrated in Figure 5.1.

RMS amplitude levelling and addition of noise

The root mean squared (RMS) amplitude of all 96 items was normalized to a maximum value of 65 dB in each file. The amplitude levelling script is a C program that calculates RMS amplitude in overlapping frames (of 1024 samples, with 512 samples between the beginnings of successive frames), finds the maximum RMS amplitude in a sound file, then scales the samples so that the maximum amplitude will be 65 dB. Noise was added to the normalized stimuli at a signal-to-noise ratio (SNR) of -10, with 500ms of padding at the beginning and end of each sound file. The padding allowed the listener to acclimate to the noise at the presentation of each stimulus. At a -10 dB SNR, listeners can hear prosodic information such as duration and intonation, but cannot identify segments (see Smith et al. (1989) and references therein).

Procedures

The experimental instructions and stimuli were presented to participants using a presentation script developed with E-Prime software. Participants were tested individually in a sound attenuated chamber. They listened to stimuli played over headphones and indicated what they heard by circling one of two choices on an answer

sheet (e.g., (a) *le niveau de mes sénats*, (b) *le niveau de mécénat*).³⁵ Each participant heard six items in each of the four conditions, for a total of 24 items. There were four lists and the lists were counterbalanced so that each participant heard only one version of each item. Since the presence of so many exact minimal pairs on the answer sheets was impossible to disguise, no fillers were included. Each participant heard the items in the same pseudo-randomized order. The order was such that two items in the same condition never followed each other. There were two versions of the answer sheet: to control for order effects, in version A, if choice (a) contained two words in the ambiguous critical region (*le niveau de mes sénats*) and choice (b) contained one word in the ambiguous critical region (*le niveau de mécénat*) (or vice versa), the ordering was reversed in version B of the answer sheet. Full instructions and materials are given in Appendix C and Appendix D. There were two short practice sessions, each with three items, to familiarize participants with the task. Noise was added to the stimuli in the second practice session, but not the first. The experimenter was present during the practice sessions to answer procedural questions, but no feedback was given on the answers chosen.

Participants

Forty native speakers of French (34 women, 6 men) participated in the experiment, immediately after participating in the nonword experiment described in §5.3. None of the participants reported a speech or hearing disorder. Thirty-two of the participants were undergraduate students pursuing a course of study in linguistics (*sciences du*

³⁵I decided use an answer sheet rather than to collect input from the keyboard since the E-Prime program was not compatible with the French AZERTY keyboard layout. The standard American QWERTY keyboard has a slightly different layout, and lacks accented letters.

langage) at the Université Stendhal de Grenoble.³⁶ The other eight participants were undergraduate or graduate students at the Ohio State University. These participants had been raised in France and were native speakers of French. They had a variety of majors and were either pursuing regular degrees at the university or participating in year-long exchange programs. One participant was a recent graduate of the Ohio State Master's of Business Administration program. The Ohio State participants had lived in the United States for between a few months and 5 years. The mean age of all participants in the experiments was 21.0 years. Participants in France were paid €10 and participants in the United States were paid \$10 for completion of Perception Experiments 1 and 2. The two experiments combined took a total of about an hour to complete.

5.2.2 Results

Responses were coded according to whether a function word/content word boundary was perceived at the critical boundary, that is whether a one-word or two-word response was given. For example, a *le niveau de mes sénats* response is a two-word response—there is a function word/content word boundary between the syllables [me] and [se] (*mes* | *sénats*) and was coded 1. The one-word response *le niveau de mécénat* has no word boundary at the critical position and was coded 0. The percentage of two-word responses for each condition was calculated.

The results are shown in the graph in Figure 5.2. The bars represent the percent of two-word responses to each of the four conditions, defined by the speaker's original segmentation and by presence versus absence of an early rise. The first and the last

³⁶Most *sciences du langage* students are women, which explains the lack of gender balance in the participants. We have no reason, however, to believe that there are gender-related differences in speech segmentation.

bars represent responses to original productions and the middle bars responses to the resynthesized versions.

The category axis shows the original segmentation (as read by the speaker) of the four syllables of the target (one word (e.g., *mécénat*) or two word (*mes sénats*)). The scale axis shows the percentage of two-word responses, that is responses that included a word boundary at the critical boundary (e.g., *le niveau de mes | sénats*). Dark bars represent items with an early rise in the target, and light bars represent items with no early rise.

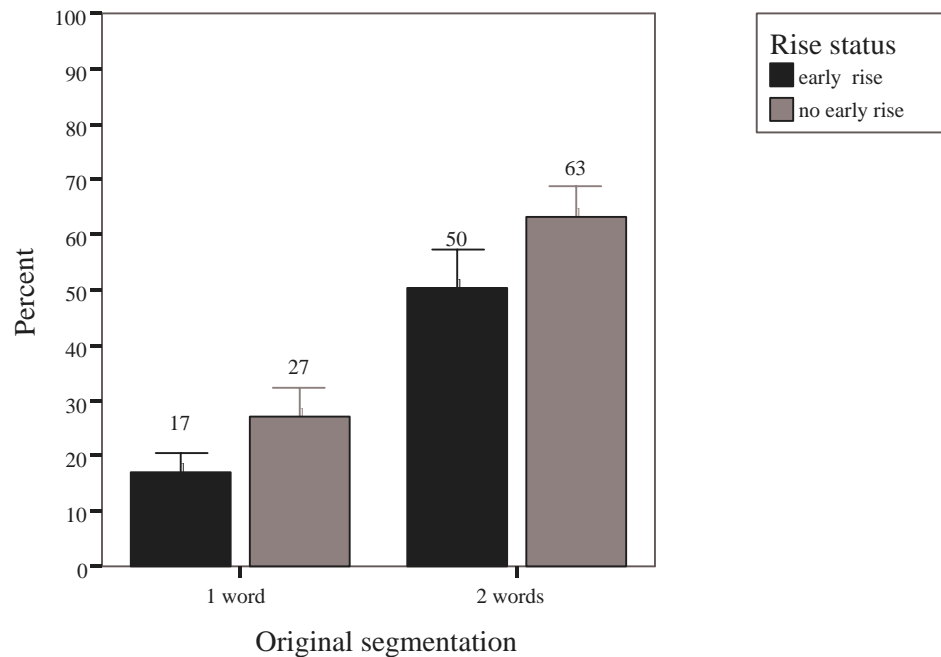


Figure 5.2: *Percentage of two-word responses in Perception Experiment 1.*

A series of Wilcoxon Signed Ranks Tests was performed.³⁷ The dependent measure was percentage of two-word responses (*mes sénats*). The results showed an effect of *original segmentation*, whether the item had been produced with one word or two words in the target (for the subjects analysis, $Z = -5.416$, $p < .001$; for the items analysis, $Z = -3.273$, $p < .001$). Listeners were more likely to perceive two words when the speaker had in fact produced two words: a mean of 56.7% of responses were two-word responses in the two two-word segmentation conditions. Listeners were more likely to perceive one word when the speaker had produced one word: only a mean of 22.1% of responses in the two one-word segmentation conditions were two-word responses. This finding is expected for two reasons. First, not all items were exact minimal pairs. Second, for exact minimal pairs, cues to the original segmentation are preserved, even in cases with conflicting intonational information (e.g., the items in Figures 5.1c and d). For example, a comparison of syllable durations shows systematic differences between the one-word and two-word conditions. The second syllable of the target sequence (e.g., the [me] in [də.me.se.na]) is significantly longer when it is content-word initial (as in *de mécénat*) than when it is a function word (as in *mes sénats*) (183.8 ms vs. 154.6 ms, $p < .001$).

Listeners were also more likely to perceive a long content word (to give a one-word response) when the early rise was present than when it was absent (for the subjects analysis, $Z = -3.977$, $p < .001$; for the items analysis, $Z = -2.578$, $p < .05$). In the one-word segmentation conditions, which contained a long content word

³⁷A more common statistical choice would have been to perform ANOVAs. Yet none of the hypotheses in the perception experiments depends on an interaction between main effects. In addition, the nonparametric Wilcoxon test is suited for examining percentage data, which do not meet the assumptions of ANOVA. Nevertheless, ANOVAs were performed on the arcsines of the percentages. Those results are available by request.

(e.g., *mécénat*), only 17.1% of responses incorrectly identified the stimuli as two-word productions when an early rise was present, versus 27.1% of incorrect responses when no early rise was present. This difference is marginally significant in both the subjects analysis ($Z = -1.926$, $p = .054$) and the items analysis ($Z = -1.840$, $p = .066$). In the two-word segmentation conditions, which contained a function word followed by a shorter content word (e.g., *mes sénats*), 63.0% of responses correctly identified the segmentation of the stimuli when no early rise was present. Correct identification dropped to 50.4% of responses when an early rise was present. This difference is significant (for the subjects analysis, $Z = -2.785$, $p < .01$; for the items analysis, $Z = -2.217$, $p < .05$).

The overall percentage of two-word responses across all four conditions is quite low (39.4%). Even in the two conditions for which the speaker had produced two words, the percentage of two-word responses is only 56.7%. This suggests that there was a bias toward one-word responses, a possibility that was investigated through a post-hoc norming test. Participants in the norming were 14 native speakers of French drawn from the same population as most of the participants in the noise-masking experiment, undergraduate students in *sciences du langage* at the Université Stendhal de Grenoble. Twelve participants were women, two were men, and the average age of participants was 18.7. They were paid €5 for their participation in the experiment, which lasted about 15 minutes. They completed the norming test immediately after their participation in Perception Experiment 3 (§5.4).

Participants were presented with a response sheet which contained the minimal and near-minimal pairs of the noise-masking experiment, presented in the same order as in the original experiment. As with the noise-masking experiment, there were

two counterbalanced lists to control for order effects. Participants were instructed to indicate which of the two phrases they found “more plausible” (*plus plausible*) by checking a box to the left of their choice (as shown in (32)).

(32) ☐ le niveau de mes sénats ☐ le niveau de mécénat

The results of the norming test showed that there was indeed a bias to the one-word response. Participants chose two-word responses only 33.93% of the time ($t(23) = -2.816, p < .01$).

5.2.3 Discussion

The results of the experiment provide evidence for the Early Rise as Cue to Content Word Hypothesis. Despite the bias to one-word responses and despite the cues to original segmentation that were present in the signal (duration, segmental info for the pairs that were only near-minimal pairs), the presence or absence of the early rise had a significant influence on segmentation. In an offline task, listeners used the presence of an early rise as evidence for a content word beginning. Listeners may use their knowledge of the distribution of early rises as a cue in natural conversations. However, given the Mattys (2003) observation that stress cues to segmentation in English are useful only in noisy conditions (see §4.2.5), we must be cautious about making claims about the use of the cue in clear speech. Perception Experiment 2 examines the role of the rise in segmentation in a paradigm that uses clear speech.

It is also worthwhile to consider that the early rise is not consistently present in every Accentual Phrase. About half of the critical Accentual Phrases in the production experiment were produced without an early rise, and the speaker of the experimental items in Perception Experiment 1 did not always spontaneously produce an early

rise on Accentual Phrases like *de mécénat*. All earlier studies of French intonation agree that the rise is optional. Many accounts argue that speaking style is one factor that influences the appearance of the rise. Early rises are less commonly produced in conversation than in read speech (see, for example, (Fonagy (1979), Vaissière (1983), Post (2000)). In addition, many Accentual Phrases contain more than one content word (*les gamins sages*)—clearly one of these content words (in this case *sages*) is not marked by an early rise. The presence of an early rise therefore cannot be a necessary cue to the presence of a content word in natural conversations. Speech segmentation in French proceeds effortlessly even when no early rise is present.

In fact no cue to speech segmentation in any language is reliably present in all utterances. For example, as discussed earlier, although English listeners are sensitive to the rhythmic pattern of English and can use their tacit knowledge of this pattern in speech segmentation, they have no difficulty in understanding utterances with words that violate the dominant pattern. Similarly, phonotactic constraints may contribute to speech segmentation, but they clearly cannot be the only cue. Harrington et al. (1989), for example, showed that phonotactic constraints could mark only 37% of the word boundaries in their corpus if the analysis relied on phonemic transcription. Similarly, intonational cues to speech segmentation will not always be present in the speech signal, but when they are there, listeners will likely make use of them. A listener's reliance on certain cues may depend not only on the segmental and prosodic characteristics of the utterance, but also on listening conditions. For example, in a noisy environment (whether a naturally occurring noisy subway station or that created by an experimental manipulation), segmental information may be less

robustly conveyed than prosodic information like duration and intonation. Listeners may adjust their use of cues to speech segmentation accordingly.

5.3 Perception Experiment 2: Alignment of the early rise as a cue to speech segmentation (segmentation of non-words)

Perception Experiment 1 showed that listeners used the presence of an early rise as a cue to a content word beginning in noise-masked speech. Perception Experiment 2 further examines the nature of intonational cues to speech segmentation and tests the hypothesis in (33).

- (33) ALIGNMENT OF EARLY RISE AS CUE TO CONTENT WORD HYPOTHESIS:
Listeners interpret the low starting point (L tone) of an early rise as a cue to a content word beginning.

The experiment used phonotactically permissible nonwords, that is, sequences of syllables that are not words in the language, but could be. Listeners' strategies in the segmentation of nonwords were assumed to reflect their strategies in the segmentation of unknown (but real) words. This use of nonwords is common in studies of sensitivity to pattern frequency (see, for example, Vitevitch et al. (1997)).

5.3.1 Methods

Materials

Corpus design and recording

Thirty-six minimal pairs with identical segment sequences were designed, as in (34). In (34a), the possessive adjective *mes* 'my' ([me]) is followed by the nonsense word *lamondines*, while in (34b), [me] is the first syllable of the nonsense word *mélamondine*.

- (34) a. mes lamondines
 me la.mõ.din
 ‘*my lamondines*’
- b. mélamondine
 me.la.mõ.din
 ‘*mélamondine*’

Each critical string was four-syllables long, with the first syllable interpretable as one of the following determiners: *mon* [mõ] ‘my’ (masc. sing.), *ma* [ma] ‘my’ (fem. sing.), *mes* [me] ‘my’ (pl.), *tes* [te] ‘your’ (pl.), *ses* [se] ‘his, her, its’ (pl.), *la* [la] ‘the’ (fem. sing.), *les* [le] ‘the’ (pl.), *nos* [no] ‘our’ (pl.), *des* [de] (plural indefinite article or contraction of the preposition *de* and the definite article *les*). Twenty nonword filler items were also included to distract participants from the goals of the experiment.³⁸ The fillers varied in number of syllables and segmentation: there were four one-syllable fillers (e.g., *noche*), four two-syllable fillers (e.g., *nucard*), three three-syllable fillers (e.g., *pailliment*), three fillers with a determiner followed by a one-syllable nonword (e.g., *ses frûles*), five fillers with a determiner followed by a two-syllable nonword (*son racon*), and one filler with a determiner followed by a three-syllable nonword (*ses malliformes*). A full list of the critical items is given in Appendix G.

The target sequences were embedded in the carrier sentence given in (35) and recorded by a female native speaker of Parisian French.

- (35) Et _____ pourrait être une expression utilisée par les Français.
 ‘And _____ could be an expression used by the French.’
-

³⁸This was also the reason participants completed the nonword experiment before the noise-masking experiment—in the noise-masking experiment, it was clear that the experimental items had ambiguous segmentations. The noise-masking experiment is discussed first because the hypothesis tested there in some sense logically precedes the hypothesis tested in the nonword experiment.

Stimuli were recorded and transferred to computer by the same method described in §5.2.1. As predicted by the results of the production experiment, the L tone of the early rise was consistently realized at the boundary between the function word and the content word. For example, the L tone is realized between *et* and *mé-* in the case of *et mélamondine* and between *mes* and *la* in the case of *et mes lamondines*. These alignment differences are schematized in Figure 5.3.



Figure 5.3: *Identical segment strings with different alignment patterns. The pattern in (a), with a L elbow at the boundary between [me] and [la], is consistent with the interpretation of [me] as the function word mes ‘my’ (et mes lamondines ‘and my lamondines’). The pattern in (b), with an elbow at the boundary between [e] and [me], is consistent with the interpretation of [me] as the first syllable of a content word (et mélamondine ‘and mélamondine’).*

Actual productions used in the experiment are shown in Figures 5.4a and c. The twenty fillers were also read embedded in the carrier sentence in (35).

Fundamental frequency resynthesis

The alignment of the early rise was manipulated using Praat software, producing two additional versions of each item. For items produced with a function word followed by a content word (e.g., *mes lamondines*), the alignment of the early rise was made compatible with a single four-syllable content word (e.g., *mélamondine*): the low starting point of the early rise was moved to the beginning of the first syllable

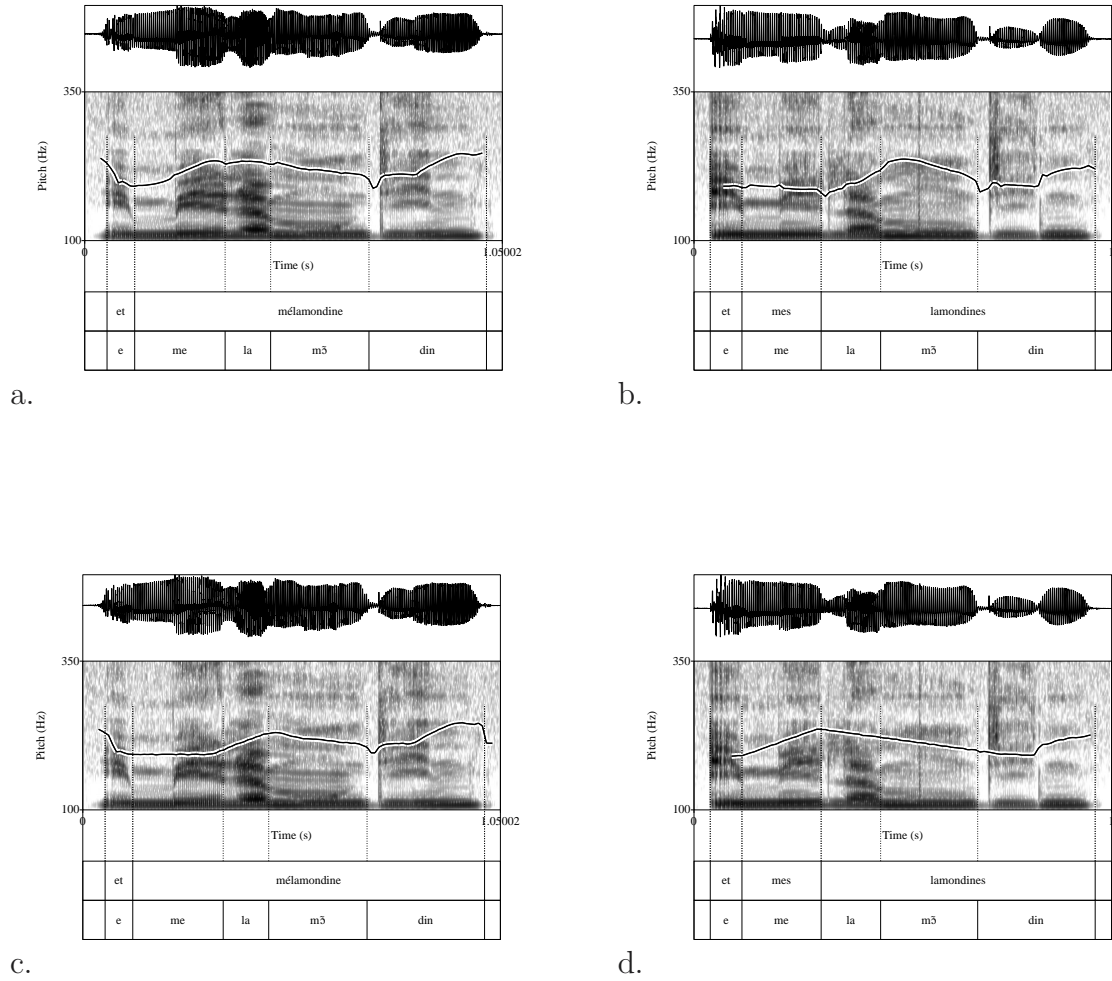


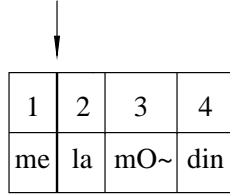
Figure 5.4: *Sample materials for Perception Experiment 2. (a) Target produced as the single 4-syllable content word mélamondine (natural production). (b) Target produced as a function word followed by a 3-syllable content word mes lamondines (natural production). (c) Target produced as the single 4-syllable content word mélamondine. Fundamental frequency resynthesized (from the original in panel a) to correspond to a segmentation with a content word beginning at [la] (e.g., mes lamondines). (d) Target produced as a function word followed by the 3-syllable content word mes lamondines. Fundamental frequency resynthesized (from the original in panel b) to correspond to a segmentation with a content word beginning at [me] (e.g., mélamondine). Note the similarity in f_0 between panels a and d and between panels b and c.*

([me] for *mes lamondines*) and the peak of the early rise was moved to the end of the first syllable (see Figure 5.4d). For items produced with a single four-syllable content word (*mélamondine*), alignment of the early rise was changed to correspond to a function word followed by a three-syllable content word: the low starting point was moved forward to the beginning of the second syllable ([la] for *mé | lamondine*) and the peak was moved to the end of the second syllable (Figure 5.4c). In the post-experiment debriefing sessions, participants routinely reported that they had been unaware that the stimuli had been manipulated. The patterns for all four versions of the critical items are shown in Figure 5.4. The early elbow is not always located exactly at the function word/content word boundary—the elbow is often realized 10 ms or more from the boundary, as noted in Chapter 3. In order for natural items to differ from their resynthesized counterparts only in presence or absence of an early rise, the elbow position was maintained.

Participants

Twenty-eight native speakers of French participated in the experiment, immediately before completing the experiment described in §5.2. Of these participants, 26 were students at the Université Stendhal de Grenoble and two were students at the Ohio State University. Twenty-four were women and four were men. The mean age of all participants in the experiment was 20.1. One Ohio State participant had lived in the United States for five years, the other for less than a year. As noted in §5.2.1, participants in France were paid €10 and participants in the United States were paid \$10 total for Perception Experiments 1 and 2.

Data for fourteen participants were excluded from the analysis for failure to meet language background criteria or performance thresholds. For two participants, examination of language background questionnaires revealed that they were not speakers of Hexagonal French (they were from Guadeloupe and Morocco). Data for an additional 12 participants was excluded from the analysis. Seven of these participants gave 5 or fewer two-word (*mes lamondines*-type) responses (with an average of 2.29) to the 36 critical items. Recall that the 36 critical items include 18 items produced as two words, nine of which were presented to listeners in the no-resynthesis version. In addition these participants gave three or fewer two-word responses (with an average of 1.29) to the filler items, even though 9 fillers were pronounced as two words. It cannot simply be that these participants are not sensitive to intonational cues—the participants correctly segmented only 19% of non-resynthesized items that had been produced as two words. These items retained all cues to segmentation, including fundamental frequency, duration, and intensity, but the listeners did not use these cues. Although the experimental instructions included an example of a determiner+nonword segmentation, it is possible that these participants tried to provide as many single-nonword answers as possible—the instructions refer to invented words (see Appendix E), and determiners are real words. Data from five other participants whose responses fell on the other extreme were also excluded from the analysis. These participants produced fewer than five one-word responses (with an average of 3) to the critical items. They did give an average of 11 one-word responses to the fillers. These participants may have adopted a strategy of assigning the longer (four-syllable) critical items and the shorter fillers to different categories.



1	2	3	4
me	la	mO~	din

Figure 5.5: *Position of the critical boundary (between the first and second syllable) for responses to the [me.la.mõ.din] (mélamondine/ mes lamondines) items in Perception Experiment 2.*

Procedures

Participants were tested individually in a sound attenuated chamber. They listened to a total of 36 critical items and 20 filler items and were instructed to “fill in the blank” on a response sheet with the word or words they heard between *Et* and *pourrait*. That is, there was no constraint on responses—participants were free to write down whatever they heard. There were four lists, each containing 9 items in each of the four conditions and the same 20 fillers. The lists were counterbalanced so that each participant heard only one version of the critical items. Items were pseudo-randomized so that two critical items in the same condition never followed each other and there were never more than three critical items in a row. Items were presented in the same pseudo-randomized order to all participants. The instructions are given in Appendix E. An example answer sheet is given in Appendix F.

There was a short practice session consisting of six items. Some participants gave one-word or two-word responses for all six items. Before going on to the critical trials, all participants were reminded that they would sometimes hear one word and sometimes more than one word between the *Et* and the *pourrait* of the carrier sentence.

This instruction was particularly important for those participants who had given all one-word or all two-word responses for the six practice items, but was given to all participants for the sake of uniformity.

5.3.2 Results

Responses were coded for the presence of a word boundary between the first and second syllables of the four-syllable critical sequence. This boundary is indicated by an arrow in Figure 5.5. A response containing a boundary (e.g., *mes lamondines* or *mé lamondine*) was coded 1 and a response with no boundary (e.g., *mélamondine* or *méla mondine*) was coded 0. Percentage of responses with a boundary was calculated for each of the four conditions. Results are shown in Figure 5.6.

A series of Wilcoxon Signed Ranks Tests was performed. The dependent measure was percentage of responses with a word boundary between the first syllable and the second syllable.³⁹

Results showed an effect of *original segmentation* (for the subjects analysis, $Z = -2.410$, $p < .05$; for the items analysis, $Z = -3.176$, $p < .01$). Listeners were more likely to perceive a word boundary at the critical first/second syllable boundary when the speaker had actually produced a boundary (e.g., *mes | lamondines*). They detected a boundary 58.5% of the time in those cases versus only 48.5% of the time when she had produced a four-syllable content (non)word (e.g., *et mélamondine*). This finding is expected, since cues to the original segmentation are present even

³⁹Note that it was not possible to use *percentage of two-word responses* as the dependent measure, as in the noise-masking experiment. In the current experiment, responses are freeform rather than forced choice. A three-word response like *mé la mondine* is coded 1 for the presence of a word boundary between the first and second syllables ([me] and [la]), while a two-word response like *méla mondine* is coded 0, since it does not have a word boundary between the first and second syllables.

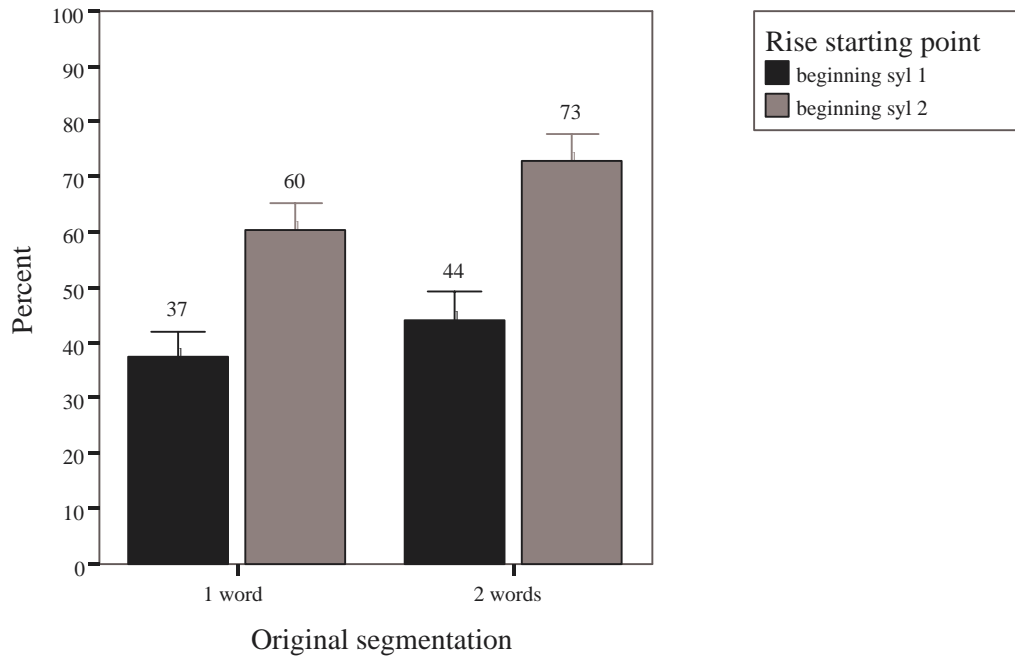


Figure 5.6: *Percentage of responses with a word boundary at the critical region in Perception Experiment 2. Standard error of the mean is indicated.*

in cases with conflicting intonational cues (e.g., the items in Figures 5.4c and d). For example, a number of durational cues are available. The second syllable of the ambiguous four-syllable sequence (e.g., the [la] of [me.la.mõ.din]) was longer when it is content word initial (*mes lamondines*) than when it is word-medial *méllamondine* (165.1 ms vs. 152.4 ms, $p < .001$). The onset consonant of the second syllable (e.g., the [l] of [la]) was also longer when the syllable was word-initial (186.0 ms vs. 166.5 ms, $p < .001$). Lengthening of prosodic-phrase initial consonants has been reported for French (see, for example, Fougeron (2001a)). Somewhat surprisingly, there was no significant difference in the duration of the first syllable (e.g., the [me]

of [me.la.mõ.din]): that syllable was on average 184.4 ms long at the beginning of content words like *mélamondine* and 187.0 ms long for function words like the *mes* of *mes lamondines*. The onset consonant of this syllable (e.g., the [m] of [me]), however, was longer when it began a content word than when it began a function word (89.0 ms vs. 81.6 ms, $p < .05$). There was also an unexpected significant difference in the duration of the last syllable (e.g., the [din] of [me.la.mõ.din]) across the two conditions: it was longer in *et mes lamondines* cases than in *mélamondine* cases (337.4 ms vs. 314.0 ms, $p < .001$).

These responses also reflect a clear bias toward the two-word responses (*mes lamondines*)—more than one word was perceived in almost half of the cases where there was a 1-word original segmentation. This bias is undoubtedly a reflection of the listener’s tacit awareness of frequency patterns in the language. Instances of 4-syllable nouns or adjectives beginning with the syllables [me], [no], etc. (like the experimental items in (36a)) are rare compared to instances of the determiners *mes*, *nos*, etc. followed by three-syllable nouns or adjectives (like the experimental items in (36b)).

- (36) a. i. *mélamondine*
 ‘*mélamondine*’
 ii. *nobrérateur*
 ‘*nobrérateur*’
 b. i. *mes lamondines*
 ‘*my lamondines*’
 ii. *nos brérateurs*
 ‘*our brérateurs*’

To calculate a rough estimate of how much more common items like those in (36b) are than items like those in (36a), I performed several searches using LEXIQUE, an online French database drawn from texts and containing 15 million word tokens (New et al. (2001)). Since the nonwords in the experimental materials have morphology

consistent only with nouns and adjectives, I restricted the searches to those parts of speech. Of the 456,651 noun and adjective tokens per million words of the database, those with 3 syllables are almost 3 times more common than those with 4 syllables: there are 55,104 3-syllable noun and adjective tokens per one million words vs. only 19,434 tokens per million for 4-syllable nouns and adjectives. In other words, 12.1% of nouns and adjectives have 3 syllables, while only 4.3% have 4 syllables.

Of all 4-syllable noun and adjective tokens, a scant 1713 per million begin with the 9 first syllables found in the experimental items ([me], [no], etc. For a complete list of these syllables see the discussion of the corpus design in §5.3.). This count corresponds to an average of 190 noun and adjective tokens per million beginning with each of the initial syllables of the experimental materials ($1713/9 = 190$) (items like those in (36a)).

As function words, however, the initial syllables of the experimental items (as in (36b)) are quite common: there are 57,794 tokens per million of the 9 function words used in the critical items, or an average of 6422 per function word ($57,794/9 = 6422$).⁴⁰ To obtain an estimate of the frequency of items like those in (36b) (determiner + 3-syllable noun or adjective), I made two assumptions. First I make the simplifying assumption that a determiner or possessive will always be followed by a noun or an adjective. The default word order in the language is DET N (ADJ), although some adjectives precede the nouns they modify (DET ADJ N). A small number of adverbs (degree adverbs or intensifiers) can come between the noun and the adjective (e.g., *un enfant assez sage* ‘a pretty well-behaved child’ or *un très bon exemple* ‘a very

⁴⁰The average is pulled down by the low token count for the possessive *tes* ‘you (informal)’ (94/million), which is rare in the written texts from which the database is drawn. This count would surely be much higher in speech. As a point of comparison, the count for *les* ‘the (pl.)’ is 16,011/million.

good example'). Second, I assume that determiners will be followed by a 3-syllable noun or adjective 12.07% of the time, since 3-syllable nouns and adjectives comprise this percentage of all nouns and adjectives.⁴¹ I therefore multiply the average number of tokens per critical function word by this percentage ($6422 \times .1207 = 774.88$), arriving at the estimate of an average of 775 tokens per million for the determiner + 3-syllable noun or adjective critical items.

Tokens like those in (36b) therefore occur at a rate of 775/per million, while tokens like those in (36a) occur at a rate of only 190 tokens per million. Put another way, items like those in (36b) are four times more common than items like (36a). It is therefore unsurprising that participants in the experiment should show a strong bias to two-word responses like those in (36b).

In addition to the effect of *original segmentation*, there was an effect of *starting position of early rise* (for the subjects analysis, $Z = -4.425$, $p < .001$; for the items analysis, $Z = -4.847$, $p < .001$). Listeners perceived a word boundary at the critical region (giving responses like *mes lamondines*) 66.5% of the time when the starting point of the early rise (the elbow) was located in the later position, at the beginning of the second syllable. Only 40.5% perceived a word boundary when the elbow was located at the earlier position, at the beginning of the first syllable. This difference holds within both original segmentations. Within the one-word original segmentation condition, listeners perceived a boundary at the critical region 37.3% of the time in the naturally produced condition, where the starting point of the early

⁴¹ Note that the count of 3-syllable nouns and adjectives beginning with the syllables in the critical items (55,104/million) includes both 3-syllable nouns and adjectives immediately following the determiner (e.g., the *mécano* of *le mécano mécontent* 'the dissatisfied mechanic') and those in other positions (e.g., *mécontent*). This is unproblematic, since it is also true of the count for all nouns and adjectives, not just those with 3 syllables (e.g., for the 2-syllable noun and the 2-syllable adjective in *le méchant message* 'the mean message').

rise was located at the beginning of the first syllable. Responses with a boundary in the critical region jumped to 60.3% when the start of the early rise was moved by f_0 resynthesis to the later position at the beginning of the second syllable. This difference is significant in both the subjects analysis ($Z = -3.946$, $p = .001$) and the items analysis ($Z = -4.229$, $p = .001$). Similarly, in the two-word original segmentation condition, listeners perceived a boundary at the critical region 73.0% of the time in the naturally produced condition, where the early rise started at the beginning of the second syllable. When the starting point of the rise was moved back to the beginning of the first syllable by f_0 resynthesis, responses with a boundary in the critical region dropped to 44.1%. This difference is significant in both the subjects analysis ($Z = -3.840$, $p = .001$) and the items analysis ($Z = -4.462$, $p = .001$).

Examination of the data also revealed some inter-subject variability in use of the early rise cue. This is illustrated in Figure 5.7. A data point above the $x = y$ line indicates that a participant was more likely to give a response with a boundary at the critical 1/2 syllable boundary (e.g., *mes* | *lamondines*) when there was an early rise starting at the 1/2 syllable boundary. A few participants did not use the early rise in their segmentation decisions: their data points lie on or near the $x = y$ line. The vast majority of data points for the 28 participants, however, fall above the $x = y$ line line.⁴² For six participants, the difference between the responses with a boundary at the critical region in the two conditions (early rise at syllable 2 versus early rise at syllable 1) was significant according to the results of a chi-square test (calculated

⁴²Only 26 data points (rather than 28, corresponding to the number of participants) are visible on the scatterplot. This is due to overlapping data points: two participants had boundaries at the critical region 39% of the time when the early rise started at the beginning of syllable 1 and 67% of the time when it started at the beginning of syllable 2 boundary. For an additional two participants, those numbers were 50% and 83%, respectively.

on the raw numbers, with Yates correction). For an additional five participants, the difference was marginally significant (p 's ranging from .063 to .087). Data points for these 11 participants fall above the line in the chart in Figure 5.7 (only ten data points are visible because of two points overlapping at (50.0, 83.3)).

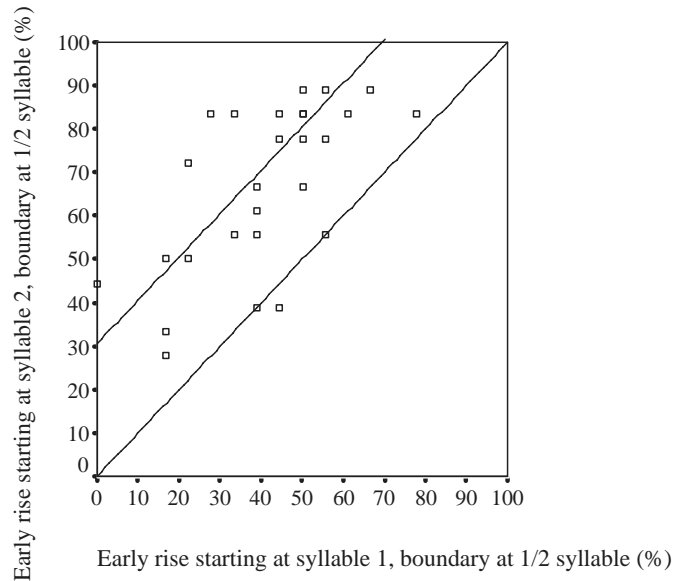


Figure 5.7: *Comparison of responses with boundaries at the critical 1/2 syllable region according to alignment of the early rise.*

5.3.3 Discussion

The results provide strong support for the Alignment of Early Rise as Cue to Content Word Hypothesis. Given ambiguous sequences of syllables, most listeners use the alignment of the low starting point to a syllable boundary as a cue to the beginning of a content word. Note that the peak of the early rise is often realized in the syllable *after* the perceived content-word-initial syllable (e.g., the [la] of *mélamondine*

or the [m̃] of *lamondines*), either as a simple peak or as a high plateau. Listeners are clearly sensitive not merely to the presence of the rise, but to the location of its low starting point. As the results of the production experiment show, the L tone of the early rise is the most stable tone of the two-rise /LHLH/ pattern. Although the early rise is not consistently present in all or even most Accentual Phrases, when it is present, the L tone of the early rise is consistently aligned to the beginning of a content word, and listeners exploit this consistency in speech segmentation.

I also find support for listeners' use of their tacit knowledge of a second pattern in the language, namely the relative frequency of *mes lamondines* and *mélamondine*-type phrases. This use of multiple cues is expected.

My conclusion about the contribution of the alignment of the early rise needs to be somewhat revised, however. Recall that we know that word boundaries and syllable boundaries do not always coincide in French due to common processes of resyllabification (see discussion in §4.2.2). As Welby (2002) notes, in liaison environments, the L tone of the early rise is realized at the beginning of the syllable formed by resyllabification, rather than at the beginning of the morphological word. For example, in the prosodic phrase *mon Arianella* 'my Arianella' in Figure 5.8, the latent /n/ of *mon* [m̃] surfaces before the vowel-initial *Arianella*.⁴³ The example is taken from the Welby (2002) data. The low elbow corresponding to the L tone of the early rise appears at the left edge of the resulting syllable [na], not at the beginning of the lexical word *Arianella*.⁴⁴

⁴³Note that the vowel preceding the latent consonant typically loses its nasality in some dialects, particularly in the South of France. The speaker is from the southern city of Marseille. *Mon* loses its nasal vowel, and the first syllable of the utterance is [mo].

⁴⁴The exact alignment details deserve to be systematically studied. Vaissière (1997), for example, claims that there is greater f_0 variability in items like *l'incompatibilité* than in items like *la compatibilité*. In the former case, there is competition between two tones:

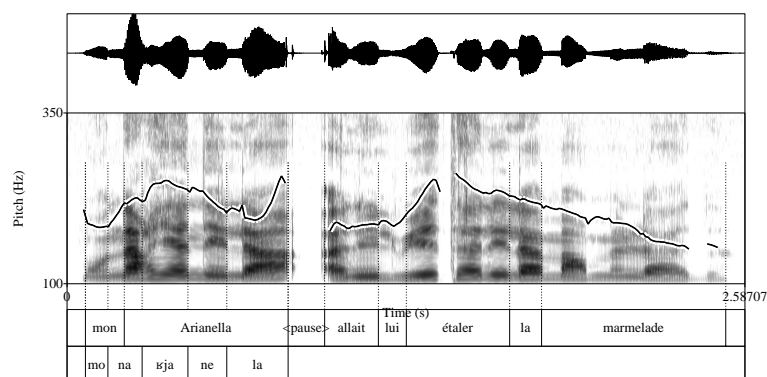


Figure 5.8: *Illustration of the realization of the L tone of an early rise at the left edge of a syllable ([mo]) formed by a resyllabification process (liaison). A gloss for the sentence is ‘My Arianella was going to spread the marmelade for him or her.’*

As I have argued earlier (§4.2.2), the existence of processes of resyllabification is not likely to pose a problem for French listeners. They have tacit knowledge both of these processes and of intonational cues to speech segmentation. It seems highly likely that listeners also have some mechanism for dealing with exceptions to the general rule that early rises are aligned to the content word beginnings.

The design of Perception Experiment 2, however does not explore whether an f_0 elbow at the function word-content word boundary acts as a cue to a content

Si le mot lexical est précédé d’un article contracté (comme dans “l’incompatibilité”), la première syllabe du mot est associée à la fois avec un ton Bas sousjacent [] du fait de la présence du morphème grammatical, et un ton Haut (H lexical, du fait de la présence d’un début de mot long[]).

‘If the lexical word is preceded by a contracted article (as in “l’incompatibilité”), the first syllable of the word is associated at the same time with a low tone [] following from the presence of the grammatical morpheme, and a high tone (lexical H, following from the presence of the beginning of a long word[]).’ (p. 68)

Although Welby (2002) did not systematically study variability of this type, no differences in alignment of the early rise were immediately apparent between the liaison items (like *mon Arianella*...) and non-liaison items (like *ma Marianella* ‘my Marianella’).

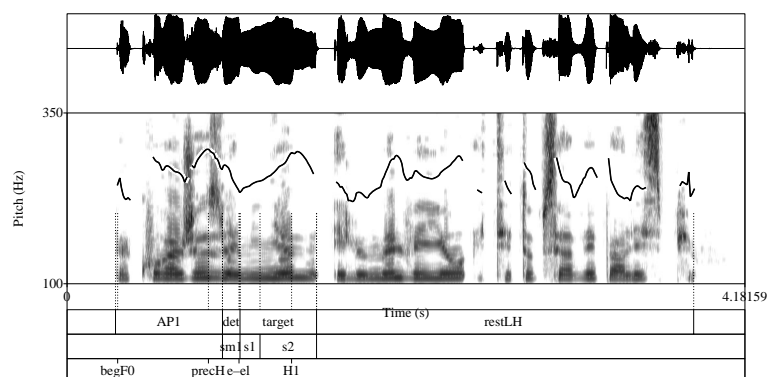


Figure 5.9: *LH pattern with a fall from the late H of the preceding AP to an elbow at the function word-content word boundary in the utterance-medial AP la mignonne. The full utterance is La courageuse, la mignonne et le malveillant étaient observés par l’espion. ‘The courageous woman, the cute woman, and the spiteful man were being watched by the spy.’*

word beginning, whether or not there is an accompanying early rise. For example, there is an elbow at the determiner-noun boundary in Figure 5.9 (also included in Chapter 1), but there is no early rise—the H is a late H. In Figure 5.10, the medial phrase has a LLH pattern with an early elbow at the determiner-noun boundary. The potential role of an early elbow with no following early rise was investigated in the next perception experiment.

5.4 Perception Experiment 3: Early elbow as a cue to speech segmentation (segmentation of nonwords)

I interpreted the results of Perception Experiment 2 as evidence that listeners used the alignment of the early rise as a cue to a content word beginning. I noted, however, that there is sometimes an early elbow at a function word-content word boundary,

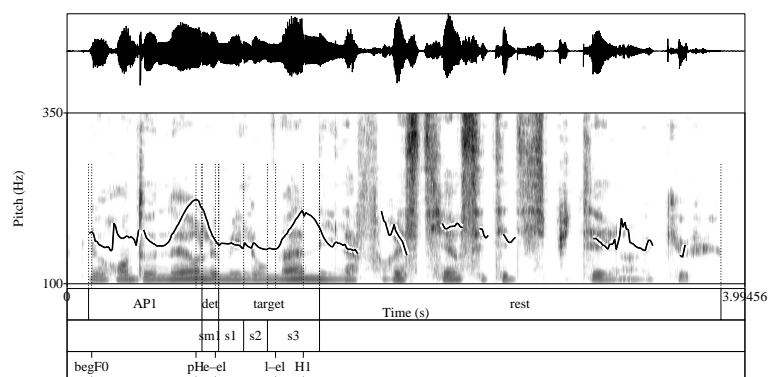


Figure 5.10: *LLH pattern with a fall from the late H of the preceding AP to an elbow at the function word-content word boundary in the utterance-medial AP le mélomane. The full utterance is Le randonneur, le mélomane et la forestière s'étaient disputés avec eux. 'The hiker, the music lover, and the forester had argued with them.'*

even in the absence of an early rise. In Perception Experiment 3, I examine whether this elbow plays a role in speech segmentation, independent of that of the early rise.

(37) **EARLY ELBOW AS CUE TO CONTENT WORD HYPOTHESIS:** Listeners interpret an early fundamental frequency inflection (elbow) as a cue to a content word beginning, even in the absence of a following early rise.

In addition, the materials used allow us to re-test the Alignment of Early Rise as Cue to Content Word Hypothesis using different materials.

The experiment used a design very similar to that of Perception Experiment 2, although participants performed a forced choice task (one-word *mélamondine* vs. two-word *mes lamondines* segmentation).

5.4.1 Methods

Materials

Corpus design and recording

Thirty-six minimal pairs with identical segment sequences were designed. Most of these sequences were different from the ones used in Perception Experiment 2. The *mes lamondines* and *mélamondine* minimal pair from Perception Experiment 2, however, was included.

As with the items in Perception Experiment 2, the critical items in Perception Experiment 3 were four-syllables long, with the first syllable interpretable as a determiner. The syllables used were: *ma* [ma] ‘my’ (fem. sing.), *mes* [me] ‘my’ (pl.), *la* [la] ‘the’ (fem. sing.), *les* [le] ‘the’ (pl.), *nos* [no] ‘our’ (pl.).

Thirteen nonword filler items were also included to distract participants’ attention from the experimental manipulations. A full list of the critical items is given in Appendix I.

The target sequences were embedded in the carrier sentence given in (38) and recorded by a female native speaker of French from Lille, a city in northern France near the Belgian border.

- (38) On a lu _____ mollement.
 ʒ̃.na.ly _____ mɔl.mã
 ‘We read _____mollement.’

The thirteen fillers were also read embedded in the carrier sentence. The speaker read each critical item in its function word + 3-syllable content word segmentation (e.g., *mes lamondines*). This experiment therefore does not test the effect of the original segmentation (one word vs. two word), as did Experiment 2. Not including original segmentation as a factor in our design allowed us to keep the number of

lists and therefore number of participants manageable. Of course, we expect that there will be an even stronger bias to two-word responses across all conditions than in Experiment 2.

The stimuli were recorded in a sound attenuated chamber directly to computer, using Europec-EVA software developed by SQlab (<http://www.sqlab.com>), sampling at 25 kHz (the default for the software) with 16-bit resolution. Each item was automatically saved as a separate file. These files were converted to WAV format and silences at the beginning and end of the carrier sentence were trimmed. All further phonetic analyses and manipulations were performed using Praat.

Fundamental frequency resynthesis

As in the previous experiment, there were four versions of each item. There were however, a number of differences in the creation of the stimuli. As noted above, all the critical items were read in their function word + 3-syllable content word segmentation (e.g., *mes lamondines*). All four versions therefore contained durational cues (and possibly other types of cues) to this segmentation. Other changes involved the f_0 resynthesis. The earlier experiment contained stimuli with their naturally produced f_0 and stimuli with resynthesized f_0 . In all four versions of Perception Experiment 2, the fundamental frequency of the entire 4-syllable target sequence (and the region immediately preceding the sequence) was resynthesized. With one exception, the target sequences in Perception Experiment 3 contained only sonorant segments. The exception is the *mes lamondines/mélamondine* pair, which contains the voiced stop [d]. It was important to maximize sonorant segments, since the f_0 of the entire 4-syllable target sequence and of part of the carrier sentence was resynthesized.

Two Praat scripts were written to semi-automate the resynthesis process. For each critical item, the following segmental landmarks were labelled using a Praat script: beginning of utterance, end of utterance, beginning of *lu*, beginning of the vowel of *lu*, end of *lu*, end of first target syllable, end of second target syllable, end of third target syllable, end of fourth target syllable. A second Praat script calculated f_0 for each soundfile (created Pitch objects), erased the f_0 values for the stretch beginning at the beginning of *lu* and ending at the end of the fourth target syllable, then resynthesized the f_0 by adding points (Pitch Points) at specified values. Four versions of each item were created. Two versions contained early rises. In one, the early rise began at the left edge of the first syllable of the target sequence; in the other, the early rise began at the second target syllable. Two other versions had no early rises, but did have f_0 elbows. One version had an elbow at the beginning of the first syllable of the target sequence. The other had an elbow at the second target syllable.

Seven time points were used by the script to define the f_0 patterns for the four versions. The locations of these points, labelled A–G, are described in Table 5.1. The f_0 values assigned to each of the four versions are given in Table 5.2. Values between specified points are determined by linear interpolation. For reference, an example four-syllable sequence embedded in the carrier sentence is shown in Figure 5.11.

	1	2	3	4	
On a lu	me	la	mO~	din	mollement

Figure 5.11: *Target four-syllable sequence* [me.la.mɔ̃.din] (mélamondine/ mes lamondines) *embedded in carrier sentence for Perception Experiment 3.*

Point	Location
A	25 ms after the beginning of the [l] of <i>lu</i>
B	75 ms after the beginning of the [u] of <i>lu</i>
C	beginning of syllable 1 of target sequence
D	beginning of syllable 2 of target sequence
E	beginning of syllable 3 of target sequence
F	60% into syllable 4 of target sequence
G	90% into syllable 4 of target sequence

Table 5.1: Points used by the Praat script in definition of f_0 curves for Perception Experiment 3.

	version a	version b	version c	version d
	early rise	no early rise	early rise	no early rise
	elbow at s1	elbow at s1	elbow at s2/s3	elbow at s2/s3
A	190	190	190	190
B	225	225	225	225
C	190	190	na	na
D	220	na	190	190
E	na	na	220	na
F	205	195	205	195
G	225	225	225	225

Table 5.2: Fundamental frequency values (in Hz) assigned to each of the points by the Praat script for each of the four conditions in Perception Experiment 3.

The alignment of the preceding late rise on *lu* has the same definition for all four conditions, as does the alignment of the late rise at the end of the target sequence. The intonation patterns of the 4-syllable target sequence (e.g., [me.la.mõ.din]) varied across the four versions. Examples of these four versions are given in Figure 5.12. Version a has a LHLH pattern with an early rise starting at the beginning of the first syllable of the target sequence and ending at the end of the syllable (e.g., [me] of [me.la.mõ.din]). This pattern was hypothesized (by the Alignment of Early Rise as

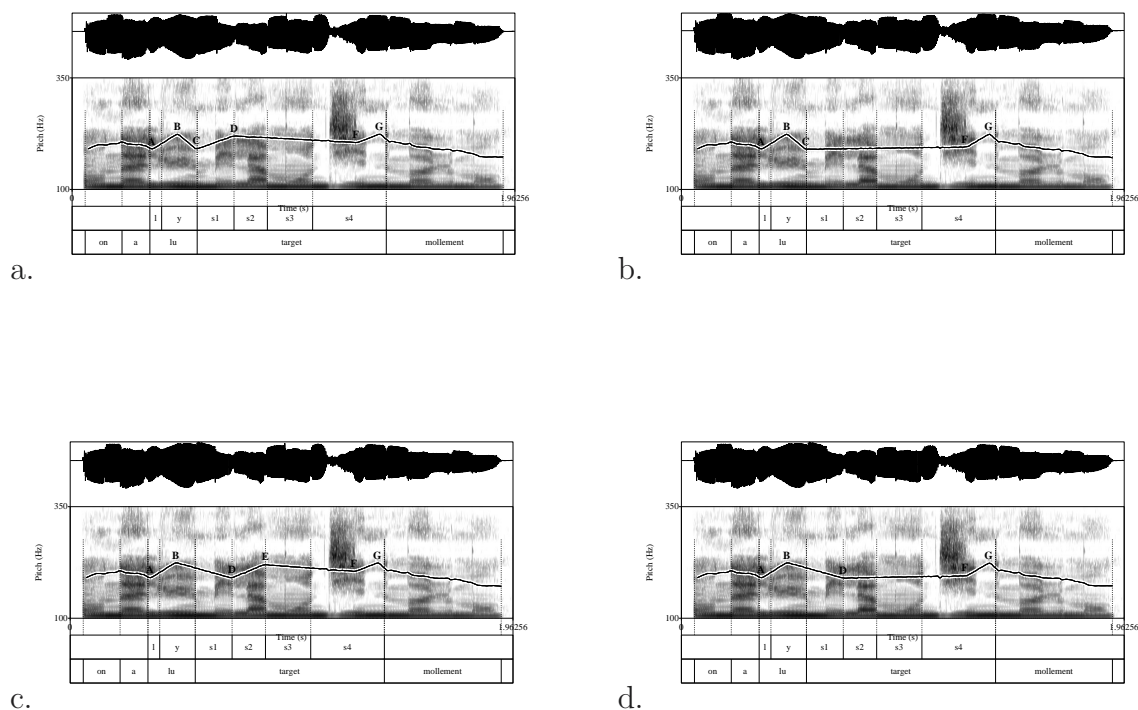


Figure 5.12: Sample materials for Perception Experiment 3. All targets were produced as a function word followed by a 3-syllable content word. (a) Target has a LHLH pattern with the early elbow at the starting at the beginning of the first target syllable, a pattern hypothesized to be compatible with a 4-syllable content word interpretation (mélamondine). (b) Target has a LLH pattern with the elbow in the same position and is also hypothesized to be compatible with a 4-syllable content word interpretation (mélamondine). (c) Target has a LHLH pattern with an early rise starting at the boundary between the first and second target syllables, a pattern was hypothesized to be compatible with a function word followed by a three-syllable content word (mes lamondines). (d) Target has a LLH pattern with the same early elbow position and is also hypothesized to be compatible with a function word followed by a three-syllable content word (mes lamondines).

Cue to Content Word Hypothesis) to be compatible with a four-syllable content word (e.g., *mélamondine*). Version b has a LLH pattern that retains the position of the early elbow in version a. This pattern was also hypothesized (by the Early Elbow as Cue to Content Word Hypothesis) to be compatible with a four-syllable content word (e.g., *mélamondine*). Version c has a LHLH pattern with an early rise starting at the second syllable of the target sequence (e.g., [la] of [me.la.mõ.din]). This pattern was hypothesized (by the Alignment of Early Rise as Cue to Content Word Hypothesis) to be compatible with a function word followed by a three-syllable content word (e.g., *mes lamondine*). Version d has a LLH pattern that retains the position of the early elbow in version c. This pattern was also hypothesized (by the Early Elbow as Cue to Content Word Hypothesis) to be compatible with a function word followed by a three-syllable content word (e.g., *mes lamondine*).

The f_0 values and their alignment with the text were guided by findings from the production study about alignment and scaling, observations from the original productions, and feedback from native speakers about naturalness. For example, the H of late rise of the target sequence is higher than that of the H of the early rise, a scaling that was commonly observed in the production data. In addition, the L elbow of the late rise in LLH patterns (versions b and d) is lower (at 195 Hz) than that of the L elbow in LHLH patterns (versions a and c), reflecting the finding from the production results that late elbows of LLH patterns are typically lower than those of LHLH patterns.

Nevertheless, some compromises were made to ensure that the materials sounded as natural as possible. In particular, in the un-resynthesized productions, the late rise that precedes the target sequence (the rise across *lu*) was typically higher than

the early rise on the target sequence. Preliminary attempts at resynthesizing the f_0 curve, however, produced very unnatural sounding stimuli for versions a and b. In those versions, the f_0 fell too quickly from the H of the preceding late rise to the early elbow of the target sequence. To resolve this problem, I lowered the peak of this preceding late rise and moved the L starting point and the H peak of the rise earlier in the syllable. As a result, the *lu* has an artificial quality in some of the stimuli, and the downdrift of the original utterances is not respected.

In retrospect, a better solution would have been to add function word syllables before the target syllable as in (39), thereby reducing the slope of the f_0 fall from the preceding H.

- (39) On a lu “et _____” mollement.
 ‘We read “and _____” mollement.’

Participants

Thirty-two native speakers of Hexagonal French participated in the experiment. The participants were undergraduate students studying *sciences du langage* at the Université Stendhal de Grenoble or undergraduate engineering students participating an internship at the Institut de la Communication Parlée in Grenoble. One participant was a student in the first week of a doctoral program. The experiment lasted less than half an hour and participants were paid €5.

Data from eight participants who did not meet language background criteria or performance thresholds were excluded from the analyses. Of these, two were excluded because they failed to meet the language background criteria. One participant had not grown up in metropolitan France, but had spent eight years of her childhood

in Tahiti; another spoke both Arabic and French at home. Six participants were excluded because their pattern of responses across the fillers and critical items indicated that they were not performing the task. Two participants gave all two-word responses for all 36 critical items and all one-word responses for all 13 filler items. One of these participants volunteered during the debriefing session that he had misunderstood the task. A third participant had exactly the reverse pattern—all one-word responses for critical items and all two-word responses for all filler items. Note that the fillers, which do not have resynthesized f_0 curves, retain all cues to segmentation. These three participants nevertheless gave the same type of response to every filler (7 of which had two-word segmentations and 6 of which had one-word segmentations). They apparently performed a task where they separated critical items (all 4-syllables long) from filler items (shorter than the critical items, with variable numbers of syllables). An additional participant gave incorrect responses for all 13 of the filler items, giving two-word responses for all but one. Another participant gave two-word responses to all fillers. Another gave two-word responses to all fillers and all critical items. The mean age of the retained participants in the experiment was 19.7. Twenty-nine were women and three were men.

Procedures

Participants were tested individually in a sound attenuated chamber or a quiet room. They listened to a total of 36 critical items and 13 filler items presented over headphones through a Matlab script. As the auditory stimulus was played, two possible segmentations (e.g., *mélamondine*, *mes lamondines*) were presented on the computer screen, one underneath the other. Listeners were instructed to select one of the two segmentations by clicking a box just to the left of their choice. This

forced choice task was adopted since so many participants had difficulty with the free response task used in Perception Experiment 2. The instructions are given in Appendix H.

There were four lists, each containing 9 items in each of the four conditions and the same 13 fillers. The lists were counterbalanced so that each participant heard only one version of the critical items. Items were pseudo-randomized so that two critical items in the same condition never followed each other and there were never more than three critical items in a row. Items were presented in the same pseudo-randomized order to all participants. In order to control for order effects related to the presentation of segmentation choices on the screen, there were two lists of segmentation pairs. For example, in List 1, *mélamondine* preceded *mes lamondines*, while in List 2, *mes lamondines* preceded *mélamondine*. There was a short practice session consisting of four items.

5.4.2 Results

The dependent measure was the percentage of two-word responses (responses with a boundary at the critical region (*mes lamondines*)). Results are shown in Figure 5.13.

A series of Wilcoxon Signed Ranks Tests was performed. The dependent measure was percentage of two-word responses (responses with a word boundary between the first syllable and the second syllable; e.g., *mes lamondines*). The results showed an effect of *position of early elbow* (for the subjects analysis, $Z = -3.721$, $p < .001$; for the items analysis, $Z = -4.366$, $p < .001$). Listeners were more likely to perceive a word boundary at the first/second target syllable boundary and give a two-word response (e.g., *mes | lamondines*) when the elbow was located at the beginning of

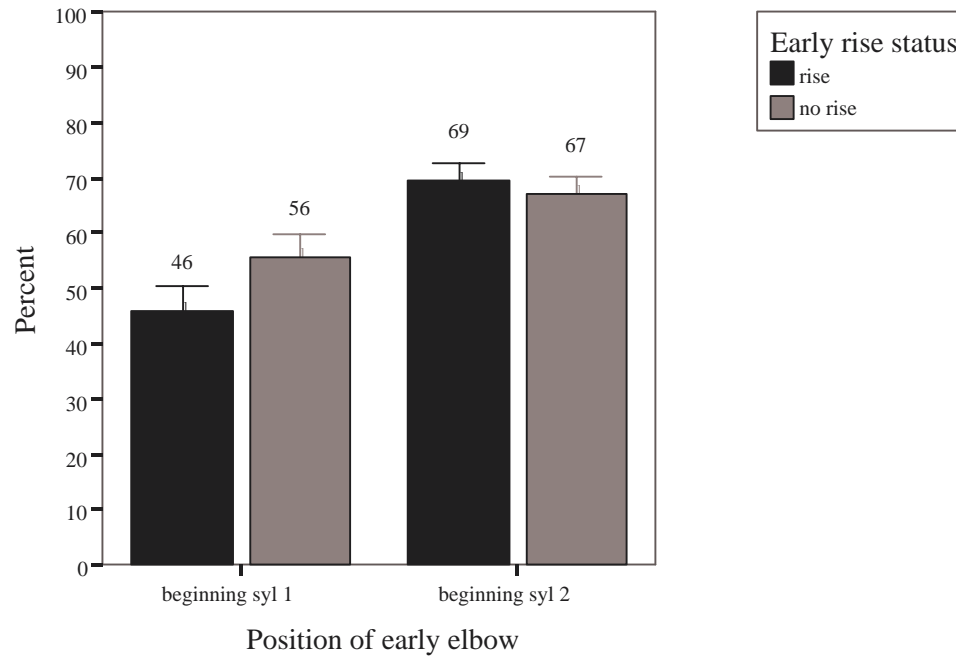


Figure 5.13: *Percentage of two-word responses (responses with a word boundary at the critical region) in Perception Experiment 3. Standard error of the mean is indicated.*

the second target syllable (as in Figures 5.12c and d). They gave two-word responses 68.2% of the time in those cases versus only 50.7% of the time when the elbow was located at the beginning of the first target syllable (as in Figures 5.12a and b).

The comparison crucial to the Early Elbow as Cue to Content Word Hypothesis is that between the two conditions with no early rise but different elbow positions (the two gray bars in Figure 5.13). This difference was significant (for the subjects analysis, $Z = -2.532$, $p < .05$; for the items analysis, $Z = -2.732$, $p < .01$). As predicted by the hypothesis, listeners heard two words (e.g., *mes lamondines*) more often (67.0% of the time) when the elbow fell at beginning of the second target syllable

(as in Figure 5.12d) and less often (only 55.6% of the time) when when the elbow fell earlier, at the beginning of target syllable 1 (as in Figure 5.12b). There was also a significant difference between the two conditions with early rises but different elbow positions (the two black bars in Figure 5.13) (for the subjects analysis, $Z = -3.432$, $p < .01$; for the items analysis, $Z = -4.394$, $p < .001$)—participants gave two-word responses 69.4% of the time when the early rise began at target syllable 2 but only 45.8% of the time when the rise began at target syllable 1. This replicates the finding of an effect of *starting position of early rise* found in Perception Experiment 2. Two further comparisons explored the influence of the presence or absence of the early rise in each of the two elbow position conditions. When the elbow was at beginning of target syllable 2, listeners were no more likely to give two-word responses when there was an early rise than when there was no early rise (the two rightmost bars, 69.4% with rise, vs. 67.0% without rise, (for the subjects analysis, $Z = -.607$, $p = .544$; for the items analysis, $Z = -.759$, $p = .448$). This lack of a difference is unsurprising since all the critical items were produced as two words (a function word followed by a content word) and thus retained other cues to segmentation (such as segment and syllable durations). In contrast, when the elbow was aligned to the beginning of target syllable 1, an alignment inconsistent with a two-word interpretation, listeners gave two-word responses less frequently when there was an early rise than when there was no rise. This difference, though, is significant only by items ($Z = -2.568$, $p = .01$) not by subjects ($Z = -1.567$, $p = .117$).

As with the previous perception experiment, there was inter-subject variability. This variability is illustrated in Figure 5.14, which plots the percentage of two-word responses to items in which the elbow was located at the beginning of syllable 1 against

the percentage of two-word responses to items in which the elbow was located at the beginning of syllable 2.

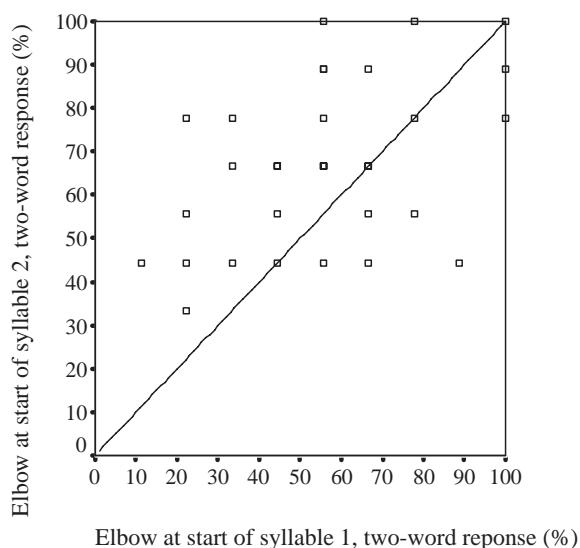


Figure 5.14: *Comparison of responses with boundaries at the critical 1/2 syllable region according to alignment of the early f_0 elbow, in condition with no early rise.*

A data point above the $x = y$ line indicates that a participant was more likely to give a two-word response (e.g., *mes | lamondines*) when there was an early rise starting at the beginning of the second syllable ([la]).⁴⁵ Data points for twenty participants fall above the line, suggesting that these participants used the presence of an f_0 elbow as cue to the beginning of a content word. However, the rest of the data points either

⁴⁵Due to overlapping data points, only 26 data points (rather than 32, corresponding to the number of participants) are visible on the scatterplot. Three participants gave two-word responses 56% of the time when the early rise started at the beginning of syllable 1 and 67% of the time when it started at the beginning of syllable 2 boundary. For two participants, those numbers were 44% and 67%, respectively. For another two, they were 55% and 89%, respectively. For yet another two, they were 67% and 67%.

fall right on the $x = y$ line (5 data points) or below the line (7 data points), indicating that these participants did not use the f_0 elbow as expected in their segmentation decisions. This variability is greater than that in the previous experience. This may have to do with differences in the strength of a simple f_0 elbow versus an early f_0 rise as a cue to speech segmentation, or it may be due in part to the fact that the stimuli in Perception Experiment 2 were more natural sounding than those in Perception Experiment 3. Chi-squared tests examining these differences could not be carried out since the criterion for minimum number of observations per cell is not met (there are 18 observations per participant in a 2×2 contingency table). Variation across listeners in the use of an f_0 elbow as a cue to speech segmentation clearly merits further study.

5.4.3 Discussion

The results of Perception Experiment 3 support the Early Elbow as Cue to Content Word Hypothesis—listeners interpret the presence of an early elbow as evidence of a content word beginning, even when that elbow is not followed by an f_0 rise.

Given the findings of Perception Experiments 1 and 2, how should this intonational cue (these intonational cues) be best characterized? There are at least two possibilities. One is that the cue is not the f_0 rise at all, but the elbow, the f_0 inflection. The presence of an early rise simply makes the elbow more salient (by making the angle smaller, for example). The finding in Perception Experiment 3 that two-word responses were rarer when an early rise were present and the elbow was aligned to the beginning of target syllable 1 could be interpreted to mean that the cue of the early elbow was enhanced by the following rise. Another interpretation is

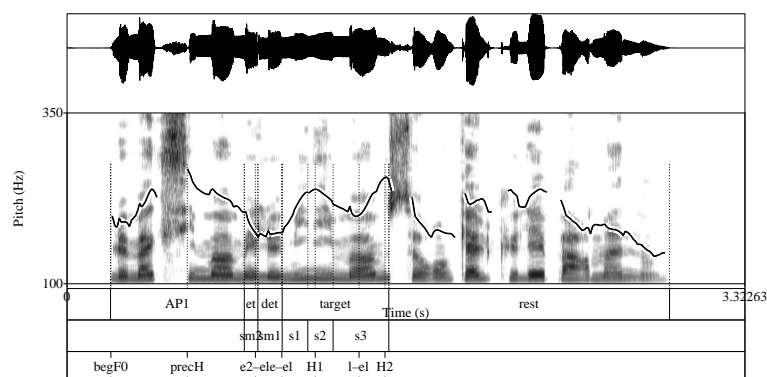


Figure 5.15: *LHLH pattern with two early elbows, one at the boundary between the conjunction et and the determiner le (e2-el), the other at the boundary between the determiner and the first content word syllable (e-el). The utterance is Le maximum et le minimum seront calculés par Manon. ‘The maximum and the minimum will be calculated by Manon.’*

that the early rise is in fact the cue: this cue is composed of two parts, a L target followed by a H target. The cue is most efficient when both components are present (i.e., when there is an f_0 peak), but listeners can make use of the cue even when only one is present. The latter account is supported by the observation that the L target is often, but not always realized as an f_0 elbow. There will be no elbow, for example, in utterance-initial position if the utterance does not begin with a function word. A further argument in favor of this interpretation is the existence of more than one early elbow in a single AP. An example is given in Figure 5.15 (which also appears as Figure 3.24 of Chapter 3). It is possible that the f_0 elbows mark boundaries other than those between function words and content words.

The interpretation of the early rise as a two-part cue to the presence of a content word is compatible with the treatment of the LH- (or LHi, in Jun and Fougeron's notation) of the /LHLH/ pattern as a bitonal phrase accent.

5.5 General Discussion

The results of Perception Experiments 1 and 2 provide the first direct experimental evidence of the use of intonational cues in speech segmentation among French listeners, a role suggested by both Vaissière (1997) and Di Cristo (2000).

In Vaissière's account, it is the alternation between low and high fundamental frequency in French that helps the listener to identify word boundaries. As discussed in §4.3.1, Vaissière draws a parallel between the function of this alternation in French and that of the alternation in English between strong and weak syllables. Vaissière's crucial insight is that French listeners are likely to exploit regularities in intonation as cues to speech segmentation. She also points to the importance of tonal alignment, specifically citing the f_0 elbow as a boundary marker.

My account differs from Vaissière's in several important respects. First, Vaissière's account of the association of the L tone of the early rise (the elbow) is at odds with the alignment results from the production experiment. Vaissière claims that this tone is typically associated with a preceding function word. In Chapter 1, I argued that this L tone is doubly associated to the left edge of the content word and the left edge of the Accentual Phrase. This account allows a uniform treatment of cases in which there is no preceding function word. Vaissière (1997) suggests that in those cases, the L tone may be associated to a schwa syllable (her example is *mécanisME souvent* 'mechanism often'). This remains to be systematically investigated, but a separate

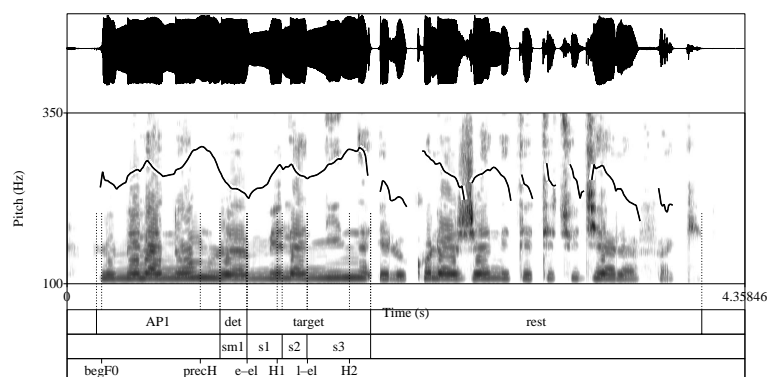


Figure 5.16: *LHLH pattern in an utterance-medial, critical AP. There is no low leading plateau across the determiner, but a fall from the preceding H to a L right at the function word-content word boundary.*

account of the association of the L tone in absolute utterance-initial position would still be needed.

In addition, Vaissière’s account states that function words have a “valeur cible basse de f_0 ” ‘low target f_0 value’ while content words are generally marked by “un passage (au moins) dans le registre haut du locuteur” ‘a passage (at least) in the high register of the speaker’ (Vaissière (1997), p. 66). This generalization is not supported by the data. If the Accentual Phrase is utterance medial, there will not typically be a low plateau across the function word, but a fall from a preceding H across the function word. This is illustrated in Figures 5.9 and 5.16. In addition, listeners in Perception Experiment 3 gave two word interpretations to targets with the pattern in Figure 5.12d, where f_0 falls across the first syllable (interpreted as a function word) then stays low at a much lower level across the next two syllables (interpreted as part of a content (non)word). Given these observations, generalizations about low

and high f_0 values are difficult to support and to test. Our observations from the production experiment, on the other hand, allowed us to make testable predictions about the contribution of the presence of early rises, f_0 elbows, and their text-to-tune alignment as cues to speech segmentation.

The findings in support of an intonational cue to a function word/content word boundary are particularly important in light of evidence that function words play a special role in infant language acquisition and adult speech processing. Christophe et al. (1997) propose the following:

Because phonological phrases are constructed with reference to syntactic structure, closed-class items such as determiners, pronouns and conjunctions tend to occur at their edges (beginning or end in general, depending on the language). Thus it is a good strategy to look for closed-class items at the edges of prosodic units. Babies could exploit this regularity of languages and compile a list of the syllables that occur at the beginning and end of prosodic units, storing the most frequent syllables in a separate list, and subsequently identifying these syllables as closed-class items when encountered at the borders of a prosodic unit.

p. 595

Christophe et al. (1997) note that a number of studies have shown that very young children are sensitive to function words. The results of an ERP study show that 10- to 11-month olds acquiring English notice when the function words of stories are replaced with nonsense monosyllables like [gu] (Shafer et al. (1992)). And Gerken and McIntosh (1993) and Gerken et al. (1994) have shown that two-year olds understand function words that they do not produce. Christophe et al. (1997) propose that function words and morphemes are represented separately in the brain. In their model, this separate representation aids infants in early syntactic and lexical acquisition. In the latter process, babies quickly develop a list of function words—these will be the

“most frequent syllables at prosodic edges” (p. 599)—then strip off these syllables to discover the content words (or at least content word beginnings) of their language.

The results of a phoneme-recognition experiment reported in Christophe et al. (1997) suggest that adult listeners perform “function-word-stripping” in language processing. Participants listened to sentences containing determiner-noun-adjective phrases like those in (40) and performed two tasks.

- (40) a. un fou larmoyant [target follows determiner]
 a crazy man tearful
 ‘a tearful crazy man’
- b. son pas gracieux [target follows noun]
 his/her step gracious
 ‘his/her gracious step’

In the first, the “generalized” task, they pressed a button as quickly as possible to indicate whether they heard the target sound (represented by a visually presented letter). In the second, the “word-initial” task, they responded only if the target occurred word-initially. When the target was preceded by a noun (as in (40b)), listeners were significantly slower to detect the target phoneme in the word-initial condition than in the generalized condition, suggesting an added processing load to identify the target as word-initial. In contrast, when the target was immediately preceded by a determiner (as in (40a)), listeners were just as fast to spot the target phoneme in the word-initial as in the generalized condition. This result suggests that listeners were able to quickly recognize the determiner as a function word, “strip” it, then identify the following syllable as the beginning of a content word. As Christophe et al. (1997) put it, “some word boundaries are more readily available than others” (p. 606).

Christophe et al. (1997) argue that infants likely employ a similar strategy, that there is unlikely to be a complete disconnect between adult and child processing. There will of course be a different weighting of cues and some of the cues available to adult listeners will not be available to infants (lexical cues). But the function-word-stripping account depends on two factors that are within an infant's competence, sensitivity to intonational patterns and knowledge of function words.

The results of Perception Experiments 1, 2 and 3 show that French has a specific intonational cue to the beginning of a content word. It seems likely that this cue could play a role in language acquisition, particularly if the early rise is particularly common in child-directed speech (for discussion of this possibility, see discussion in Chapter 2). In establishing their list of function words, French-acquiring babies may well look not simply for the "most frequent syllables at prosodic edges," but for syllables to the left of the early rise. The presence of the early may also speed lexical access in adult processing, as Vaissière (1997) suggests. In fact, in the Christophe et al. (1997) experiment, the early rise may have been present to facilitate lexical access in determiner-target items like (40a).⁴⁶

Our observations about the presence and availability of a language-specific intonational cue to content word beginnings are thus compatible with the Christophe et al. (1997) account. Some of Christophe et al.'s assumptions, however, should be reexamined in light of the results of the current experiments. They write, for example:

... one could assume that function words are somehow phonetically marked (an assumption, however, that seems rather implausible for French). In this view, subjects would be able to identify real function words just by listening to them and no content word syllable would ever be truly homophonous to a function

⁴⁶The sentences, however, were read "naturally but at a rather fast speech rate" (p. 602). To the extent to which prosodic phrases are shorter in fast speech, this instruction may have limited the production of early rises.

word. . . . Each of these proposals makes different predictions that can be tested experimentally. (footnote: For instance, the prosodic hypothesis predicts that subjects will be misled when encountering a syllable homophonous to a function word at the beginning of a prosodic unit (such as the French verb “décider” where “des” is the plural [in]definite article).)

p. 608

Our results, however, clearly demonstrate that function words are in fact phonetically marked—they are shorter in duration than the corresponding content word-initial syllables and they have distinct intonational properties. These markers are effective, though imperfect, in disambiguating segmentally ambiguous syllables. Furthermore, a number of studies have demonstrated that purportedly homophonous words in fact show systematic differences in spontaneous speech. For example, Lavoie (2002) found that the English preposition [foɪ], but never its phonemic homophone *four*, was often realized as [fə]. So claims of homophony must be examined carefully.

In addition, the intonational cue to content word beginnings could be particularly helpful to listeners in cases where there are sequences of more than one function word syllable. In these cases, the content word may begin two or three (or rarely more) syllables from the beginning of the prosodic phrase, as in the prosodic phrase in Figure 5.8 (p. 199) that begins with three function word syllables (*allait lui* ‘was going to for him/her’). The function words may then be more toward the middle of the prosodic phrase than at its edge.

Finally, the extent to which the early rise, particularly with respect to its alignment details, is used as a cue to speech segmentation deserves further study. All of our experimental stimuli were constructed to maximize the presence of sonorant segments in the critical region at which the low starting point of the early rise might be found. In the production experiment, this control allowed us to examine precise

alignment details. In the perception experiments, it was crucial to our ability to produce natural-sounding stimuli with f_0 resynthesis. In light of the Content et al. (2001b) results, which replicated the classic syllable effect in French only with the liquid pivotal consonants used in the original experiments (see Chapter 4 for details), we must ask how general a cue the early rise and the precise alignment of its starting point to the beginning of a content word. There are at least two possible scenarios—the first is that the cue is mostly useful for sequences of sonorants or sequences of voiced consonants, where there is minimal microprosodic perturbation. In this view, an early rise would be a most effective cue to the beginning of a content word in (41a), whose first syllable is completely composed of sonorants. It would be a less effective cue in (41b), where the content word begins with [d], a voiced stop and in (41c), where the content word begins with [t], a voiceless stop that will cause a break in the f_0 curve corresponding to the stalling of the vocal folds during the stop closure. This proposed restriction is similar to the claim that certain allophonic cues to speech segmentation will be relevant only to certain segment classes (see discussion in Chapter 4).

- (41) a. *et le mégaphone*
 b. *et le démagogue*
 c. *et le téléphone*

The other possibility is that the listener is well-equipped to deal with microprosodic perturbations and that these perturbations do not hinder speech perception. After all, segmentally conditioned f_0 perturbation is pervasive in all spoken languages, yet we do not perceive pitch as discontinuous. In addition, these perturbations may help in speech perception. Reinholt Petersen (1986), for example, showed that segmentally conditioned f_0 perturbations could act as cues to segment identity. It seems

plausible that listeners confronted with input like (41c) can use microprosodic perturbations to help them identify segments, while factoring out these perturbations to “reconstruct” the low elbow of the early rise and its intended alignment. According to this view, which I favor, the difference in fundamental frequency differences before and after the [t] in in (41c) acts as a cue to segmental identity and to the presence of the L beginning of the early rise.

CHAPTER 6

SUMMARY AND CONCLUSION

Le grand Dieu fit les planètes et nous faisons les plats nets!
‘God made the planets and we make our plates clean!’⁴⁷

Rabelais (*ca.* 1534)

6.1 Goals of the dissertation

My goals in this dissertation were twofold. I first sought to clarify the structure of French intonation through a production study. I examined the types of prosodic phrases found in the data and the text-to-tune alignment (tonal alignment) of those phrases, that is, the timing of fundamental frequency peaks and valleys with respect to segmental landmarks like syllable edges. My results allowed me to address a number of controversial issues in the literature. I then drew on the results of the production study to build materials for a series of perception experiments that tested the role of intonational cues in speech segmentation, the listener’s task of determining where words begin and end. In these experiments, I used PSOLA resynthesis to manipulate the fundamental frequency of the experimental utterances, varying the presence and absence of fundamental frequency rises and inflections and their alignment with the text.

⁴⁷The sense is ‘lick the platter clean’.

French intonation is characterized by a “late rise,” an obligatory fundamental frequency rise at the end of a phrase that is not the last phrase of the utterance and an “early rise,” an optional rise near the beginning, as illustrated in Figure 6.1. As indicated in the example, these rises are often described as a series of high (H) and low (L) targets; the notation LHLH, for example, indicates a two-rise pattern.



Figure 6.1: *Illustration of a prosodic phrase (et les gamins sages ‘and the good kids’) realized with (a) an obligatory late rise and (b) an optional early rise and an obligatory late rise. A gloss for the entire sentence is ‘And the good kids were gardening’.*

The peak of the late rise is aligned with the end of the accented syllable and thus marks the boundary of a prosodic phrase. By contrast, the alignment of the early rise and its role are not well understood. The first part of this dissertation examined the structure of the early and late rises. The second part examined whether the early rise aids the listener in speech segmentation by marking boundaries between function words (determiners, conjunctions, etc.) and content words (nouns, adjectives, etc.). Such a role had been proposed by other researchers, but never examined experimentally.

6.2 Summary of the findings

6.2.1 Production Experiment: Tonal association and alignment

The production experiment described in Chapter 3 had two goals. First, I wanted to investigate claims in the literature about the structure of the early and late rises. To do this, I examined text-to-tune alignment. Fundamental frequency (f_0), the rate of vocal fold vibration, is the primary acoustic correlate of pitch. This detailed phonetic study adds to our knowledge of intonation in French, a language for which very few studies of intonational timing exist. Second, I needed to provide a basis for the development of stimuli for subsequent perception studies.

Seven native speakers of Parisian French read sentences with embedded target words at two speaking rates. The following factors were manipulated: number of syllables in the target word, position of the target within the sentence, number of function words preceding the target word, and form of the definite article preceding the target word. Some main results are summarized here.

The two-rise LHLH pattern was the most common pattern. Duration in clock time (milliseconds) and length in syllables were equally good predictors of the realization of the LHLH pattern. Length in syllables (but not duration) has been proposed as a predictor by some models.

In addition to the LHLH pattern, several other patterns were observed: LH, LLH, LHH, LHi, and L2H. The LHi is a pattern with an early rise, but no late rise; its use is appropriate for an enumeration or list. One of the patterns, the L2H, is not reported in the literature. The pattern, which was rare even in my data, is characterized by

a fall from the late H of a preceding phrase to a L in the last syllable of the target phrase, followed by a rise to the late H.

The results show that the two tones at the edges of the two-rise LHLH pattern had the most stable alignment. The L tone (the starting point) of the early rise was consistently aligned with the function word/content word boundary across speakers, speaking rates, and position of the target within the utterance. The H of the late rise was consistently realized near the end of the last syllable of the phrase. There was considerably more variation in the timing of the H end point of the early rise and the L starting point of the late rise. The evidence from tonal alignment is inconsistent with models that give the same account of both the early and the late rise.

Like Jun and Fougeron (2002), I argue that the early rise and the late rise are structurally different. I treat the LH* of the late rise as a bitonal pitch accent in which the H* is associated with the last syllable of the phrase. I treat the LH- of the early rise as a bitonal phrase accent. I argue that the L of this phrase accent is doubly associated to the left edge of the content word and to the left edge of an early boundary, preferably the prosodic phrase boundary, revising somewhat the account of double association given in Welby (2002).

In addition, I observed that L fundamental frequency inflections or “elbows” were sometimes realized at boundaries between function words and content words, even when there was no immediately following early rise (no immediately following H). This observation led to the development of the hypothesis tested in Perception Experiment 3.

6.2.2 Perception Experiment 1: Presence of the early rise as a cue to speech segmentation

I performed a series of three perception experiments, all of which are described in Chapter 5. The first tested the hypothesis that listeners use the presence of an early rise as a cue to the beginning of a content word. Previous studies had shown that early rises are rarely realized on function word syllables. In addition, the results of the production experiment showed that the low starting point of the early rise began consistently at the boundary between a function word and a content word. I suspected that listeners might exploit this regularity in finding word beginnings.

Minimal or near-minimal pairs like the one in (42) were constructed, whose last three syllables could be interpreted either as a function word followed by a two-syllable content word (as in (42a)) or as a three-syllable content word (as in (42b)).

- (42) a. le niveau de mes sénats two-word condition
 lə ni.vo də me se.na
 ‘the level of my senates’
- b. le niveau de mécénat one-word condition
 lə ni.vo də me.se.na
 ‘the level of patronage’

These pairs were read by a native speaker of French. Items in the one-word segmentation condition had early f_0 rises (for example, beginning at [me] in (42b)), while items in the two-word condition had no early rise. Two additional versions of each item were created through resynthesis of the f_0 curve. Early rises were added to items in the two-word condition and removed from those in the one-word condition. Stimuli were masked by noise.

Participants listened to the stimuli and indicated what they heard in a forced choice task (e.g., *le niveau de mécénat* or *le niveau de mes sénats*). The hypothesis was supported. Listeners were more likely to choose the one-word segmentation when there was an early rise, even for items with resynthesized f_0 whose targets had been produced as two words. Similarly, they were less likely to choose the two-word segmentation when an early rise was present.

6.2.3 Perception Experiment 2: Alignment of the early rise as a cue to speech segmentation

Perception Experiment 2 tested whether speech segmentation is influenced not only by the presence or absence of an early rise, but also by the alignment of the rise. It built on results of the production experiment that showed that the low starting point of the rise is aligned to the boundary between a function word and following content word. Unlike the previous experiment, it used clear speech, rather than speech masked by noise.

Minimal pairs with identical segment sequences were designed, as in (43).

- (43) a. mes lamondines
 me la.mõ.din
 ‘*my lamondines*’
 b. mélamondine
 me.la.mõ.din
 ‘*mélamondine*’

Each sequence was ambiguous between an interpretation as a determiner followed by a three-syllable nonword or as a four-syllable nonword. These sequences were read in the carrier sentence in (44) by a native speaker of French.

- (44) Et _____ pourrait être une expression utilisée par les Français.
 ‘And _____ could be an expression used by the French.’

Each sequence was read two times, once with each interpretation. Early rises consistently began at the left edge of the content (non)word (e.g., at [la] in (43a) and [me] in (43b)), as predicted by the results of the production experiment. The f_0 of each naturally produced item was resynthesized to produce versions with an early rise alignment compatible with the other segmentation. For example, the early rise of *mes lamondines* was shifted left to begin at [me], an alignment compatible with a *mélamondine* interpretation, and the early rise of *mélamondine* was shifted to the right, to an alignment compatible with a *mes lamondines* interpretation.

Participants listened to the items and “filled in the blank” on a response sheet to indicate what they heard between the *Et* and *pourrait* of the carrier sentence.

The results support the hypothesis. A shift in the position of the rise caused a shift in listeners’ perception of the location of the content word boundary. Participants were more likely to perceive *mélamondine* when the low starting point of the early rise was located at the beginning of [me] and *mes lamondines* when it was located at the beginning of [la]. This was true even of stimuli with resynthesized f_0 , which retained other cues to their original segmentation.

6.2.4 Perception Experiment 3: Early elbow as a cue to speech segmentation

Perception Experiment 3 tested whether the alignment of an inflection in the fundamental frequency curve (an “elbow”) with a function word/content word boundary could cue the beginning of a content word, even when the elbow was not followed by an early rise. This question was inspired by the observation of utterances with these types of f_0 elbows in the production study data.

As the previous experiment, ambiguous four-syllable sequences like [me.la.mõ.din] were developed. These sequences were read in the carrier sentence in (45).

- (45) On a lu _____ mollement.
 ‘We read _____ listlessly.’

For each item, f_0 resynthesis was used to create four versions varying in the presence or absence of an early rise and in position of the f_0 elbow. All versions had an f_0 elbow. In two versions, the elbow was aligned to the beginning of the first target syllable (e.g., the [me] of [me.la.mõ.din]). One of these versions had an early rise beginning at the low f_0 elbow, while the other did not. In the other two versions, the f_0 elbow was aligned to the beginning of the second target syllable (e.g., the [la] of [me.la.mõ.din]). One version had an early rise, while the other did not.

Participants listened to the items embedded in the carrier sentence and indicated in a forced choice task what they had heard (e.g., *mes lamondines* or *mélamondine*).

The results provide preliminary evidence that some listeners perceive an f_0 elbow as a cue to a content word beginning. Overall, listeners were more likely to give one-word responses (like *mélamondine*) when the elbow was located at the beginning of the first syllable ([me]) and more likely to give two-word responses (like *mes lamondines*) when the elbow was located at the beginning of the second syllable ([la]). The presence of an early rise enhanced the effect in the case of the one-word segmentation responses.

I interpret the results of the three perception experiments to show that the early rise and its tonal alignment are cues to content word beginnings. The early rise is a compound boundary tone (as argued in Jun and Fougeron (2002)) composed of two parts: a L target followed by a H target (LH-). While the boundary is most effectively marked when both the L and the H are realized, a single tone can also be interpreted as a boundary marker.

6.3 Limitations of the study

One major limitation of the production study is that it used read speech. We know that read speech and spontaneous speech differ on many levels, including the level of intonation. In particular, the early rise is known to be less common in spontaneous speech than in read speech.

The design of the materials prevented me from answering certain questions. The target phrases in the production study was constrained to a single content word preceded by one or two function words. Tonal alignment in phrases of different structures (with more than one content word and no preceding function words, for example) remains to be examined. The lack of symmetry in syllabic structure between the first and last content word syllables in the target APs made it difficult to compare the durations of the two types of syllables. First syllables were always CV, but last syllables often had a coda consonant (CVC).

While the results of the perception experiments provide compelling evidence that the early rise can be used as a cue to speech segmentation, more research is needed to establish that this cue is used in more typical listening conditions (i.e., outside the laboratory). This type of research is crucial given recent findings calling into question the generality of the use of stress segmentation cues in English (Mattys (2003)).

In addition, I suspect that the influence of the f_0 elbow as a cue to the beginning of a content word may be stronger than the results suggest. Recall that in the critical stimuli in Perception Experiment 3, the fall to the early L elbow of the target sequence from the preceding H was shortened to produce a more natural-sounding *lu*. In addition, all stimuli were read in their two-word (*mes lamondines*) interpretation. The

experiment therefore lacks the balance of the two previous perception experiments, and I am unable to investigate the factor of original segmentation and its interactions.

6.4 Directions for future research

A number of additional directions for future research present themselves.

6.4.1 Association and alignment of the early rise in consonant- versus vowel-initial phrases

A widespread claim in the literature is that the nature of a word’s first segment affects the realization of the early rise (Padeloup (1990), Vaissière (1997), Mertens et al. (2001), Post (2002)). These accounts have in common that they claim that the early rise is typically realized on the first syllable in consonant-initial words and on the second syllable in vowel-initial words. What “counts” as an onset consonant differs according to the account. According to Padeloup (1990), an onset consonant need not be underlying for a word to be treated as consonant-initial with respect to the placement of the early accent. It can arise through article contraction, enchainment, liaison, or glottal stop insertion (e.g., before the vowel in *après-demain*).⁴⁸ Post (2000), however, claims that “the underlying structure of the syllable must be involved in this phenomenon and not its surface realisation (i.e., the /l/ is syllabified with the following vowel in *l’impossibilité*)” (p. 84). Vaissière distinguishes between the behavior of words with initial consonants arising from liaison and words with initial consonants arising from contraction of the definite article. She includes as consonant-initial underlyingly vowel-initial words that gain an onset consonant through liaison: “[l]a cible initiale est en général retardée si le mot commence par une voyelle (comme

⁴⁸Padeloup’s position on this is confused, though, by her inclusion of the example *l’inTOxication* in her discussion of vowel-initial words (p. 148).

‘inauguré’), s’il n’y a pas de consonne de liaison et avec des sonnantes” ([t]he initial target is generally delayed if the word begins with a vowel (like “inauguré”), if there is no liaison consonant, and with sonorants’) (p. 68).

These claims about the differing behavior of consonant- vs. vowel-initial words have not yet been systematically tested. The evidence we do have, however, suggests that these claims will not stand up to a systematic test. First, even researchers who note that speakers tend to realize the early rise on the second syllable of a long, vowel-initial word observe that segmental composition is not an absolutely reliable predictor of peak location. Padeloup (1990), for example, reports that for vowel-initial words, 76% of early rises were realized on the second syllable, and 24% were realized on the first syllable. Second, as noted in Chapter 5, no differences in alignment of the L of the early rise were apparent between items with underlying consonant and liaison consonant onsets.

6.4.2 Intonational cues in online speech segmentation

While my perception experiments have shown that listeners can use intonational cues in speech segmentation, the experiments used two offline tasks. It remains to be demonstrated that listeners use this information online. Vaissière (1997) specifically suggests that the presence of an early rise might speed lexical access. To test this, one could use the word-spotting task, a paradigm widely used to investigate factors that facilitate lexical access, the retrieval of words from the mental lexicon (see, for example, Norris et al. (1997) and van der Lugt (1999)). In the task, participants listen to a series of non-words. Some of these non-words have real words embedded within them, while most do not. The listener’s task is to press a button as soon as he

or she hears a real word and to repeat the word. Both reaction time in milliseconds and error rate (wrong button presses or incorrect repetitions) are measured.

Using word-spotting to study the influence of the early rise on lexical access would require a number of modifications to the paradigm. Since word length influences the appearance of the early rise, long target words (or multi-word target phrases) embedded in longer syllable strings must be used. Most French syllables, however, are also real words, so it will be difficult to construct long strings of syllables with no embedded words or only a single embedded word. One possible modification may be to instruct participants to monitor for multisyllable (“long”) words. For example, in (46) both the critical and the distractor example items contain other shorter real words (e.g., *jus* ‘juice’ and *arc* ‘arc’ in (a), and *âne* ‘donkey’ in (b)). Listeners would indicate that (a) but not (b) contains the multisyllabic real word *juvénile*.

- (46) a. léjuvénilemaitarque [critical item (*juvénile* embedded)]
 b. mélanamufaleux [distractor item]

A series of items could be created and recorded both without an early rise and with an early rise (as illustrated in Figure 6.2). A second version of each sentence would be created using computer resynthesis and the pitch curve altered to add or remove

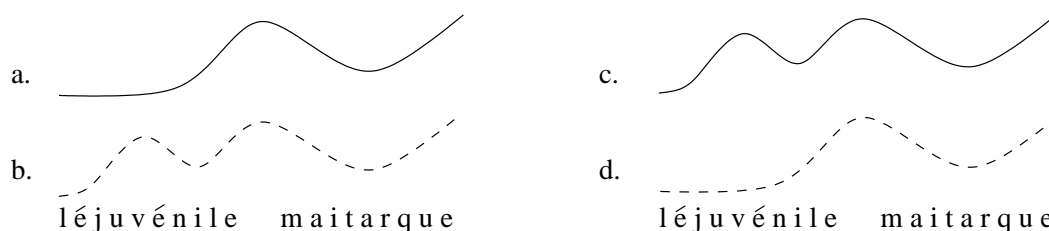


Figure 6.2: *Example item set for a word-spotting experiment. Solid lines represent natural f_0 curves, while dotted lines represent resynthesized curves.*

early rises, as shown in Figures 6.2b and d. If the early rise facilitates lexical access, then listeners should have faster response times in cases like Figures 6.2b and c, where there is an early rise in the critical region, than in cases like Figures 6.2a and d, where there is no early rise. Such a finding would provide evidence that listeners use the presence of an early rise even at the earliest stages of processing.

6.4.3 Production of the early rise in difficult listening conditions

In order to make ecologically valid claims about the use of the early rise in speech segmentation, it must be demonstrated not only that the rise can be used as a cue to speech segmentation, but also that it is available to the listener. If intonational cues are used by listeners in speech segmentation, it may be that speakers adapt their speech to provide or enhance cues in difficult speaking/listening conditions, that is, in those conditions where listeners may have the most difficulty in finding word boundaries. Early rises might be more common, for example, in conversations in noisy environments. They might also be realized more often when a word or phrase is used for the first time and is thus assumed to be less salient.

Production of the early rise in noisy conditions

People are known to adapt their speech in noisy conditions—they speak louder and more slowly and articulate their words more clearly. These adaptations, collectively referred to as the Lombard Effect, have been extensively studied (see, for example, Junqua (1993)). There is also specific evidence that speakers enhance cues to word boundaries when their listeners express difficulty understanding (Cutler and Butterfield (1990)). Since the early rise in French is a cue to a word boundary, one might

suspect that French speakers produce more early rises in difficult speaking/hearing conditions.

To test this hypothesis, one could record participants reading sentences out loud, first under normal conditions, then speaking over noise. If the hypothesis is correct, I expect to find more early rises on targets in the noise condition than in the no-noise condition. Such a finding would demonstrate that speakers adapt their speech to provide cues that are useful to listeners.

Production of the early rise in different discourse contexts

Another factor that is likely to affect the appearance of the early intonational rise is discourse context. It is well-known that a speaker's pronunciation of a word varies according to a variety of factors (Fowler et al. (1997), Bard et al. (2000), *inter alia*). For example, in a story or discourse, a word tends to be longer and more clearly articulated when it is uttered for the first time. Speakers spend less articulatory effort in producing a word for the second or third time, either because repeated words have been rehearsed and are thus easier to produce or because recently heard words are likely to be salient to the listener. In either case, we might expect first-mention words or phrases to be produced with an early rise more often than repeated words or phrases. This question could be examined using an adaptation of the Map Task developed by Anderson et al. (1991). In the task, two participants are given similar maps with a number of landmarks drawn and labelled. A simplified example map is given in Figure 6.3.

In the task, participants sit across from each other, but cannot see each other's map. One participant (the director) sees a route drawn on his or her map and directs the second participant (the follower) to follow the specified route to reach a target.

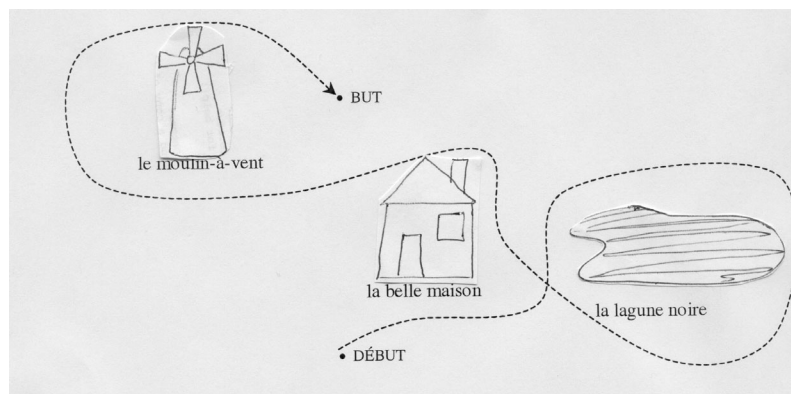


Figure 6.3: *Simplified example map for an experiment testing the influence of discourse context on the realization of the early rise.*

The map task has the advantage of eliciting spontaneous speech with many instances of the target phrases. The presence of slight differences between the director's map and the follower's map encourages the repetition of target phrases. If the hypothesis is correct, participants would produce early rises more often in cases like (47a) where the target (underlined) is a first mention, than in cases like (47b) where the target is repeated.

- (47) a. Director: Tu continues tout droit en passant devant la belle maison sur ta gauche.
 'You keep going straight and you pass in front of the pretty house on your left.'
 b. Director : Juste à côté de la belle maison, tu vois la lagune noire?
 'Right next to the pretty house, do you see the black lagoon?'

6.4.4 Articulatory/visual correlates of the early rise: Articulatory tracking

A number of studies have focused on the articulatory/visual correlates of French intonation (e.g., Fletcher and Vatikiotis-Bateson (1991), Løevenbruck (1999), Løevenbruck (2000), Fougeron (2001a), Fougeron (2001b), Dohen et al. (2003a), Dohen et al. (2003b), and Tabain (2003)).

A promising area of research may be the search for articulatory correlates of the early rise. For example, one might hypothesize that an early rise is accompanied by the hyperarticulation of the segments across which it is realized. One of the characteristics of speech in noise is clearer articulation of segments. Fougeron (2001b) provides evidence that this hyperarticulation is localized, that French speakers hyperarticulate consonants and vowels at the beginning of intonation phrases. The study, however, was limited to items without the optional early rise.

A further study could examine cases both with and without early rises. For example, the audio signal could be recorded simultaneously with articulatory signals (tongue, jaw, and lip movements). An example sentence, with the target word underlined, is given in (48).

- (48) Le mammifère sera dessiné par l'artiste.
'The mammal will be drawn by the artist.'

If the hypothesis is correct, there should be evidence of hyperarticulation at the beginnings of intonation phrases with early rises. For example, if there is an early rise on the target *mammifère* in the sentence above, the following articulatory markers might be found: larger and faster jaw opening for the initial vowel, and longer lip closure for the first [m] relative to the second [m] (see Løevenbruck (1999) and Løevenbruck (2000)). Such findings would add to our understanding of the multi-parametric

nature of cues to word and phrase boundaries. If articulatory/visual correlates were found, a follow-up perception experiment could be designed to investigate whether listeners/viewers are able to make use of these cues, adapting the methods used by Dohen et al. (2003a).

6.4.5 Applications to speech technology

The results have potential applications to the development of speech technologies. Synthesized speech sounds natural only if its intonation closely parallels patterns found in real speech. A basic requirement of any speech synthesis system is therefore a set of rules defining the alignment of fundamental frequency peaks and valleys with respect to syllable landmarks. The findings of the production experiment can be applied to refining such rules. In addition, the findings about the types of accent patterns found (e.g., LHLH with and without a plateau realization of the early H, LLH, etc.) could be used to improve the naturalness of synthesis systems, which often lack the variability found in natural speech.

Similarly, a more complete understanding of speech segmentation strategies may contribute to improving automatic speech recognition systems, which lag far behind humans in segmentation accuracy. The problem is particularly acute at faster speaking rates, in noisy conditions, and for those systems designed to recognize the speech of many users rather than just one (for example, telephone bank account information systems). Currently most speech recognition systems do not use intonational information, but a small body of research suggests that this is a fruitful avenue for further research. A study on Japanese, for example, showed that fundamental frequency information could be used to disambiguate between recognition candidates differing

in accent type (Hirose (1997)). Intonational cues to speech segmentation in French might be particularly valuable, since intonation is known to be robustly conveyed, even in quiet speech or in noisy listening conditions.

6.5 The variety of cues to speech segmentation

Although in this dissertation I have focused on one particular cue to speech segmentation, I have stressed throughout that there are many cues to segmentation. We have seen, for example, that cues such as syllable and segment duration and frequency of certain patterns in the language influence segmentation decisions. And I noted that since the early rise is not always present, it cannot be a necessary cue to speech segmentation.

There are also likely to be cases in which the early rise is more or less useful as a cue to disambiguation. For the Revolutionary plate shown earlier in the dissertation (p. 131), the segmentations *la bondance* and *l'abondance* would have distinctive intonation patterns. In the first case, the early rise, if present, begins at the boundary between the first and second syllables. In the second, it begins at the first syllable. Interestingly, the early rise was almost surely less common at the time in which the plate was created, a fact that may help to explain the missegmentation. In contrast, the ambiguity in the play on words that begins this chapter, *planètes/ plats nets* (both pronounced [pla.net]), is unlikely to be intonationally disambiguated. Considering the first part of the quote, *le grand Dieu fit les planètes*, we would expect any early rise to begin at the boundary between the determiner *les* and the first content word syllable ([pla]), whether or not we are talking about *plats nets* or *planètes*.

Here we would expect durational differences to play an important role in disambiguation. The results of Banel and Bacri (1994) suggest that the word-final [pla] of the noun+adjective phrase *plats nets* would be significantly longer than the word-initial [pla] of *planètes* (see §4.3).

The precise details of the interplay between different cues remains to be studied, as do the mechanisms for resolution when cues are in conflict. The results of Mattys (2003) demonstrate that certain cues to speech segmentation may be more useful given certain listening conditions or attentional demands on the listener. Although that study examined cues to speech segmentation in English, weighting of segmentation cues should be relevant in all human languages. In addition, certain cues may be tied to certain types of segments or phrases. For example, while Finnish listeners can use vowel harmony mismatches to locate word boundaries (Suomi et al. (1997), Vroomen et al. (1998)), they cannot use this cue in utterances in which there are no such mismatches. The sentence in (49), from Sadeniemi (1949), contains only back or neutral vowels. Such sentences are not hard to find in the language.⁴⁹

- (49) Runokieli rakentuu samoista aineksista kuin proosakielikin.
 poetic language built up of same source elements as prose language
 ‘Poetic language is built up of the same source elements as prose language.’

Similarly, since one of the main predictors of the appearance of the early rise in French is length of the prosodic phrase (as discussed in §3.4.3), it will necessarily be a more reliable cue in longer phrases.

Speech as a robust transmitter depends on redundant coding of information.⁵⁰ It therefore seems very unlikely that just one cue will be used in a given situation. It

⁴⁹I thank Ilse Lehisté for the example sentence.

⁵⁰I would also expect that there would be redundant coding in signed languages.

6.6 Models of intonation

My experimental results provide evidence for Autosegmental Metrical models that straightforwardly account for the observed facts of tonal alignment. The results show that the early rise as a cue to a content word beginning depends on the alignment of its low starting point, the L target of the rise. The results also suggest that the cue is composed of two targets, a L followed by a H, since a low f_0 elbow aligned to a function word-content word boundary can signal a content word boundary, even in the absence of an immediately following rise.

Accounts which treat holistic units or configurations as their primitives must stipulate the existence of additional features (e.g., [+delay]) to account for the association of part of the tune with units internal to the prosodic phrase. And these accounts have no way of accounting for the apparent similarity of patterns with a simple f_0 elbow and those with an early rise preceded by an f_0 elbow.

Nevertheless, my results are not entirely compatible with the assumptions of standard Autosegmental Metrical Theory. For example, I have argued that the early rise and the late rise are both composed of a L tone followed by a H tone. Yet, if the early H is a trailing tone and the late L is a leading tone, they are not bitonal tones as traditionally understood in Autosegmental Metrical phonology. There is no invariant distance between the L of the early rise, which is doubly associated to the left edge of the first content word syllable and to the left edge of the prosodic phrase, and the following H. Nor is there an invariant distance between the H* of the late rise, which is associated to the last full syllable of the phrase, and the L that precedes it. The timing of these tones also does not follow that predicted by Grice (1995), who proposes that a leading tone is associated with a syllable preceding the starred syllable, while a

trailing tone follows its starred tone at a fixed interval in “normalized time” (p. 121). A large body of evidence from tonal alignment largely supports core Autosegmental Metrical principles Prieto et al. (1995), Arvaniti et al. (1998), Grice et al. (2000), and (D’Imperio (2000)). Yet other researchers have argued that refinement of the assumptions of standard Autosegmental Metrical theory are necessary to account for observed phenomena. For example, Arvaniti et al. (2000) discuss the case of Greek, where both tones (LH) of a rising pitch accent seem to be associated to the accented syllable. An additional question in my own data arises from the characterization of the H* of the late rise as a pitch accent. While this H* is clearly associated to the last full syllable of the phrase, it marks the right boundary of this phrase. It is not clear that this accenting renders the syllable more prominent, as is predicted for a pitch accent in languages such as English.

In addition, the prevalence of plateau realizations of the H of the early rise poses a challenge to models in which the f_0 of unspecified material is defined by a simple linear interpolation between specified targets. Level plateau realizations might be accounted for by tone spreading of an unstarred tone, but the account of “rising plateaux” is not straightforward.

6.7 What counts as phonological?

As noted in Chapter 1, the presence of an early rise makes no obvious contribution to meaning. That is, a two-rise LHLH pattern does not convey a different meaning from a LLH pattern with no early rise. This is in apparent contrast to the situation in other languages in which differences in tonal specifications are claimed to convey categorical meaning distinctions and to therefore be phonologically distinct (e.g., the

L+H* vs. L*+H of Neapolitan Italian (D’Imperio (2000)) or the H* and L+H* of English (Pierrehumbert and Hirschberg (1990), though see Welby (2003) for further discussion).

In the classic conception, phonology is the system of sound contrasts in a language:

Phonologists find it crucial to be able to represent differences of sound that can be used in a language to distinguish distinct lexical items or distinct grammatical items and categories. It is necessary to say that the differences of sound are used to refer to either distinct lexical *or* distinct grammatical items because not *all* differences need to be distinguished in the formalism. *Some* differences, that is, may perfectly well be part of a person’s linguistics knowledge (using that term in a pretheoretical way), but fail to satisfy the criterion of being relevant to lexical or grammatical information. ...this knowledge does not contribute to distinguishing any lexical items, nor any grammatical items, and thus does not enter into a strictly phonological account (though it is linguistic knowledge and involves knowledge of sound systems). [Emphasis in the original.]

Goldsmith (1996), pp.9–10

Following this definition of phonology, the knowledge of possible tonal patterns in prosodic phrases is not part of the phonology. Indeed, the lack of distinctive word stress or accent led some earlier researchers to conclude that French was a “*langue sans accent*” ‘language without accent’ (see Rossi (1980) for a discussion). In an alternative view, phonology involves the recovery of controls needed for the production of one’s native language (see, for example, Durand (2000)). According to this view, a native speaker of French’s tacit knowledge of the range of possible tonal patterns and of what tonal pattern or patterns are appropriate to a given Accentual Phrase falls squarely under phonological knowledge. The choice of tonal pattern is in part determined by factors such as duration of the prosodic phrase (see §3.4.3), but it may be that the pattern realized is motivated at least partly by speaker choice. For example, a speaker may be more likely to produce an early rise in difficult listening

conditions, where there is a need to enhance the salience of the content word (see §6.4.3).

Pierrehumbert (2003) stresses the importance for phonology both of the notion of contrast and of the building up of detailed knowledge of statistical variation across all levels of representation. For example, speaker-listeners are exquisitely sensitive to the frequency of occurrence of segment sequences in their native language. Although Pierrehumbert does not directly address variation in intonational patterns, her discussion would seem to extend to the detailed knowledge that French speakers must possess to produce tonal patterns. Pierrehumbert (2003) argues that this building up of statistical knowledge is reflected in developmental differences in the speech of children and adults, since children have encountered fewer forms over which to build a probability distribution. For example, 12-year old children have a voice onset time categorization boundary for /k/ vs. /g/ (as in *coat/goat*) that differs from that of adults (Hazan and Barrett (2000)). Similar differences may exist between French children and adults: children may not produce the full range of tonal patterns or may not produce them with adult-like competence.

6.8 Complementarity of production and perception studies

Finally, on a general methodological note, the dissertation illustrates the value of developing perception studies guided by the results of complementary production studies. This methodology allows researchers to first establish that the phenomena under consideration really exist and exist under conditions that do not focus the

speaker's attention on the phenomena. It preemptively responds to the (entirely reasonable) criticism that the phenomena studied in perception experiments are artificial or not representative of natural speech.

An example of this two-pronged approach is Schafer et al. (2000). These authors used a cooperative board game task to elicit spontaneous repetitions of syntactically ambiguous strings like those in (51). Other authors had questioned the generality of disambiguating prosody, suggesting that it might be an artefact of experimental conditions drawing the speaker's attention to ambiguity (Watt and Murray (1996), Straub (1997)).

(51) a. When that moves the square it should land in a good spot.

b. When that moves the square will encounter a cookie.

Having established that speakers in a fairly natural conversational situation produce disambiguating prosody even when they are unaware of the ambiguity, Schafer et al. (2000) then tested the ability of naive listeners to disambiguate these utterances. (See also D'Imperio (2000), Rolland and Løevenbruck (2002), Dohen et al. (2003a), and Dohen et al. (2003b), who combine perception studies with production studies.)

The speech collected in the production experiment was not spontaneous speech, but read speech (for discussion of the limitations of using read speech see §6.3). Crucially, though, the materials in the production experiment did not draw attention to the hypotheses to be tested in the perception study. In particular, the corpus contained no sentences with ambiguous sequences hypothesized to be prosodically disambiguated. I am therefore confident that speakers produce the early rise even in the absence of ambiguity, at least in read speech. I am also confident of the generality

of this finding and of the alignment results, since they hold for all seven of the speakers in the production experiment.

The results of the production study guided the development of my materials for the perception experiments. My review of the literature had given me initial ideas about what to study—there are descriptions in the literature about the early rise, including some observations related to its tonal alignment. But the claims about the role of the early rise in speech segmentation found in the literature were not precise enough to be empirically tested. Also, no sufficiently detailed studies were available to inform the construction of the materials, particularly since I used f_0 resynthesis to create stimuli. The validity of materials constructed in the absence of precise information about alignment could be questioned.

In addition, examination of primary data can lead to the new observations, untested in the literature. For example, the hypothesis in Perception Experiment 3, that a simple f_0 elbow could act as a cue to speech segmentation, was inspired by my finding such cases in the data.

Finally, thinking about production and perception together leads to the formation of new hypotheses of the type described in §6.4—if a given phenomenon performs some role in perception or comprehension, what factors affect its production?

BIBLIOGRAPHY

- Anderson, John and Charles Jones. 1974. Three theses concerning phonological representations. *Journal of Linguistics* 10, 1–26.
- Anne H. Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson and Regina Weinert. 1991. The HCRC map task corpus. *Language and Speech* 34(4), 351–366.
- Arvaniti, Amalia, D. Robert Ladd and Ineke Mennen. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26, 3–25.
- Arvaniti, Amalia, D. Robert Ladd and Ineke Mennen. 2000. What is a starred tone? Evidence from Greek. In Michael B. Broe and Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon*, Cambridge: Cambridge University Press.
- Astésano, Corine and Ellen G. Bard. 2003. Structural and rhythmic influences on the occurrence of the initial accent in French. In *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, pp. 503–506.
- Astésano, Corine, Ellen Gurman Bard and Alice Turk. 2002. Functions of the French initial accent: A preliminary study. In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence, pp. 139–142.
- Aubergé, Véronique. 1991. La synthèse de la parole : des règles aux lexiques. Doctoral thesis, Université de Grenoble.
- Aubergé, Véronique. 1992. Developing a structured lexicon for synthesis of prosody. In Gérard Bailly and C. Benoît (eds.), *Talking Machines: Theories, models and designs*, Amsterdam/New York: North Holland, pp. 307–321.
- Banel, Marie-Hélène and Nicole Bacri. 1994. On metrical patterns and lexical parsing in French. *Speech Communication* 15, 115–126.

- Bard, Ellen Gurman, Anne H. Anderson, Catherine Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon and Alison Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42, 1–22.
- Beckman, Mary E. 1996. When is a syllable not a syllable? In T. Otake and A. Cutler (eds.), *Phonological Structure and Language Processing: Cross-linguistic studies*, Berlin/New York: Mouton de Gruyter, pp. 96–123.
- Beckman, Mary E. and Gayle Ayers. 1993. Guide to ToBI labelling. Manuscript. The Ohio State University.
- Beckman, Mary E. and Jan Edwards. 1990. Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston and M.E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, Cambridge: Cambridge University Press, pp. 152–178.
- Beckman, Mary E. and Jan Edwards. 1992. Intonational categories and the articulatory control of duration. In Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure*, Tokyo: IOS Press, pp. 359–376.
- Beckman, Mary E. and Julia Hirschberg. 1994. The ToBI annotation conventions. Manuscript. The Ohio State University and AT&T Bell Telephone Laboratories.
- Beckman, Mary E. and Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255–310.
- Berinstein, Ava E. 1979. A cross-linguistic study on the perception and production of stress. *UCLA Working Papers in Phonetics* 47.
- Bloom, Lois. 1993. *The Transition from Infancy to Language: Acquiring the power of expression*. Cambridge, England and New York: Cambridge University Press.
- Boersma, Paul and David Weenink. 1992–2001. Praat: A system for doing phonetics by computer. Available from www.praat.org.
- Bolinger, Dwight. 1958. A theory of pitch accent in English. *Word* 14, 109–149.
- Bolinger, Dwight. 1972. *Intonation: Selected readings*. Harmondsworth: Penguin.
- Bond, Zinny S. and Larry H. Small. 1983. Voicing, vowel, and stress mispronunciations in continuous speech. *Perception and Psychophysics* 34, 470–474.
- Bradley, Dianne C., Rosa M. Sánchez-Casas and José E. García-Albea. 1993. The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes* 8, 197–233.

- Caramazza, A. and G. H. Yeni-Komshian. 1974. Voice onset time in two French dialects. *Journal of Phonetics* 2, 239–245.
- Chao, Yuen-Ren. 1947. *Cantonese Primer*. Cambridge: Harvard University Press.
- Christophe, Anne. 1993. Rôle de la prosodie dans la segmentation en mots. Doctoral thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- Christophe, Anne, Emmanuel Dupoux, Jacques Bertoncini and Jacques Mehler. 1994. Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America* 95, 1570–1580.
- Christophe, Anne, Teresa Guasti, Marina Nespor, Emmanuel Dupoux and Brit Van Ooyen. 1997. Reflections on phonological bootstrapping: Its role for lexical and syntactic aquisition. *Language and Cognitive Processes* 12(5/6), 585–612.
- Cole, Ronald A. 1973. Listening for mispronunciations: A measure of what we hear during speech. *Perception and Psychophysics* 13, 153–156.
- Coleman, John. 1996. Declarative syllabification in Tashlhit Berber. In J. Durand and B. Laks (eds.), *Current Trends in Phonology*, CNRS, Paris X and University of Salford Publications, pp. 217–244.
- Coleman, John S. and Janet B. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Third Meeting of the ACL Special Interest Group in Computational Phonology*. Somerset, N.J.: Association for Computational Linguistics, pp. 49–56.
- Connell, Bruce and D. Robert Ladd. 1990. Aspects of pitch realization in Yoruba. *Phonology* 7, 1–29.
- Content, Alain, Nicolas Dumay and Uli H. Frauenfelder. 2000. The role of syllable structure in lexical segmentations: Helping listeners avoid mondegreens. In *Proceedings of the Spoken Word Access Processes (SWAP) Meeting*. Nijmegen: Max Planck Institute for Psycholinguistics, pp. 39–42.
- Content, Alain, Ruth K. Kearns and Uli H. Frauenfelder. 2001a. Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language* 45, 177–199.
- Content, Alain, Christine Meunier, Ruth K. Kearns and Uli H. Frauenfelder. 2001b. Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes* 16(5/6), 609–636.
- Cooper, Nicole, Anne Cutler and Roger Wales. 2003. Constraints of lexical stress on lexical access in English: Evidence from native and nonnative listeners. Manuscript. Max Planck Institute for Psycholinguistics and LaTrobe University.

- Coustenable, Hélène N. and Lillas E. Armstrong. 1934. *Studies in French Intonation*. Cambridge: W. Heffer and Sons.
- Croft, William. 1995. Intonational units and grammatical structure. *Linguistics* 33, 839–882.
- Cutler, Anne. 1986. *Forbear* is homophone: Lexical prosody does not constrain lexical access. *Language and Speech* 29, 201–220.
- Cutler, Anne and Sally Butterfield. 1990. Durational cues to word boundaries in clear speech. *Speech Communication* 9(5), 485–495.
- Cutler, Anne and Sally Butterfield. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31, 218–236.
- Cutler, Anne and David M. Carter. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2, 133–142.
- Cutler, Anne and Charles Clifton, Jr.. 1984. The use of prosodic information in word recognition. In Herman Bouma and Don G. Bouwhuis (eds.), *Attention and Performance X: Control of language processes*, Hillsdale, NJ: Erlbaum, pp. 183–196.
- Cutler, Anne, Katherine Demuth and James M. McQueen. 2002. Universality versus language-specificity in listening to running speech. *Psychological Science* 13(3), 258–262.
- Cutler, Anne, Jacques Mehler, David Norris and Juan Seguí. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25, 385–400.
- Cutler, Anne, Jacques Mehler, Dennis G. Norris and Juan Seguí. 1992. The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology* 24, 381–410.
- Cutler, Anne and Dennis Norris. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance* 14, 113–121.
- Cutler, Anne and Takashi Otake. 1994. Mora or phoneme? Further evidence for language-specific listening. *Journal of Experimental Psychology: Human perception and performance* 14, 113–121.
- Cutler, Anne and Wilma van Donselaar. 2001. *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44(2), 171–195.

- Dejean de la Bâtie, Bernadette. 1993. Word Boundary Ambiguity in Spoken French. Doctoral thesis, Monash University, Victoria, Australia.
- Dejean de la Bâtie, Bernadette and Dianne C. Bradley. 1995. Resolving word boundaries in spoken French: Native and non-native strategies. *Applied Psycholinguistics* 16, 59–81.
- Delais, Elisabeth. 1994. Rythme et structure prosodique en français. In C. Lyche (ed.), *French Generative Phonology: Retrospective and perspectives*, Salford: AFLS/ESRI, pp. 131–150.
- Delais-Roussarie, Elisabeth. 1995. Pour une approche parallèle de la structure prosodique: étude de l'organisation prosodique et rythmique de la phrase française. Doctoral thesis, Université de Toulouse, Le Mirail.
- Delais-Roussarie, Elisabeth. 1999. Accentuation et réalisation des clitiques en français. *Cahiers de Grammaire* 24, 17–37. [=Phonologie: théories et variations].
- Delattre, Pierre. 1940. Le mot est-il une entité phonétique en français? *Le Français Moderne* 8, 47–56.
- Delattre, Pierre. 1951. *Principes de phonétique française: à l'usage des étudiants anglo-américains*. Middlebury, Vermont: Ecole française d'été, 2 ed.
- Delattre, Pierre. 1966. Les dix intonations du français. *The French Review* 40(1), 1–14.
- Dell, François. 1984. L'accentuation dans les phrases en français. In François Dell, Daniel Hirst and J.-R. Vergnaud (eds.), *La forme sonore du langage*, Paris: Hermann, pp. 65–122.
- Dell, François and Mohamed Elmedlaoui. 1996. Nonsyllabic transitional vocoid in Imdalwn Tashlhit Berber. In J. Durand and B. Laks (eds.), *Current Trends in Phonology*, CNRS, Paris X and University of Salford Publications, pp. 217–244.
- Di Cristo, Albert. 1976. Indices prosodiques et structure constituante. *Cahiers de Linguistique, d'Orientalisme et de Slavistique* 7, 27–40.
- Di Cristo, Albert. 1978. De la microprosodie à l'intonosyntaxe. Doctoral thesis, Université de Provence.
- Di Cristo, Albert. 1998. Intonation in French. In Daniel Hirst and Albert Di Cristo (eds.), *Intonation Systems: A survey of twenty languages*, Cambridge: Cambridge University Press, pp. 195–218.

- Di Cristo, Albert. 1999a. Le cadre accentuel du français contemporain : Essai de modélisation (première partie). *Langues* 2(3), 184–205.
- Di Cristo, Albert. 1999b. Vers une modélisation de l’accentuation du français : première partie. *French Language Studies* 9, 143–179.
- Di Cristo, Albert. 2000. Vers une modélisation de l’accentuation du français : seconde partie. *French Language Studies* 10, 27–44.
- Di Cristo, Albert, Philippe Di Cristo, Estelle Campione and Jean Véronis. 2000. A prosodic model for text-to-speech synthesis. In A. Botinis (ed.), *Intonation: Models, analysis and applications*, Dordrecht: Kluwer.
- Di Cristo, Albert and Daniel Hirst. 1993. Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l’Institut de Phontique d’Aix-en-Provence* pp. 9–24.
- Di Cristo, Albert and Daniel Hirst. 1996. Vers une typologie des unités intonatives du français. In *Actes des XXIèmes Journées d’Etudes sur la Parole*. Avignon, pp. 219–222.
- Diez, Friedrich. 1870–1872. *Grammatik der romanischen Sprachen*. Bonn: E. Weber. [First published, 1836.].
- D’Imperio, Mariapaola. 2000. The Role of Perception in Defining Tonal Targets and their Alignment. Doctoral thesis, The Ohio State University.
- D’Imperio, Mariapaola. 2001. Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33(4), 339–356.
- D’Imperio, Mariapaola and David House. 1997. Perception of questions and statements in Neapolitan Italian. In G. Kokkinakis, N. Fakotakis and E. Dermatas (eds.), *Proceedings of Eurospeech*. Rhodes, Greece, vol. 1, pp. 251–254.
- D’Imperio, Mariapaola, Noël Nguyen and Kevin G. Munhall. 2003. An articulatory hypothesis for the alignment of tonal targets in Italian. In *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, pp. 253–256.
- Dohen, Marion, Hélène Lœvenbruck, Marie-Agnès Cathiard and Jean-Luc Schwartz. 2003a. Audiovisual perception of contrastive focus in French. In *Proceedings of the Audiovisual Speech Processing Conference*. St. Jorioz, France, pp. 245–250.
- Dohen, Marion, Hélène Lœvenbruck, Marie-Agnès Cathiard and Jean-Luc Schwartz. 2003b. Potential audiovisual correlates of contrastive focus in French. In *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology*. Geneva, pp. 145–148.

- Dumay, Nicolas, Uli H. Frauenfelder and Alain Content. 2002. The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain and Language* 81, 144–161.
- Dupoux, Emmanuel. 1993. The time course of prelexical processing: The syllable hypothesis revisited. In G.T.M. Altmann and R. Shillcock (eds.), *Cognitive Models of Speech Processing: Second Sperlonga Meeting*, Hillsdale, NJ: Erlbaum, pp. 81–114.
- Dupoux, Emmanuel, Sharon Peperkamp and Núria Sebastián-Gallés. 2001. A robust method to study stress “deafness”. *Journal of the Acoustical Society of America* 110(3), 1606–1618.
- Durand, Jacques. 2000. In Pierre Escudier and Jean-Luc Schwartz (eds.), *La parole : des modèles cognitifs aux machines communicatives*, Paris: Hermès.
- Fear, Beverley D., Anne Cutler and Sally Butterfield. 1995. The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America* 97, 1893–1904.
- Fernald, A. and T. Simon. 1984. Expanded intonation contours in mothers’ speech to newborns. *Developmental Psychology* 20, 104–113.
- Fletcher, Janet. 1991. Rhythm and final lengthening in French. *Journal of Phonetics* 19, 193–212.
- Fletcher, Janet and Eric Vatikiotis-Bateson. 1991. Articulation of prosodic contrasts in French. In *Proceedings of the 12th International Congress of Phonetic Science*. Aix-en-Provence, pp. 18–21.
- Fonagy, Ivan. 1979. L’accent français: accent probabilitaire. *Studia Phonetica* 15, 123–133.
- Fonagy, Ivan and Eva Bérard. 1973. Questions totales, simples et implicatives en français parisien. *Studia Phonetica* 8.
- Fonagy, Ivan and Judith Fonagy. 1976. Prosodie professionnelle et changements prosodiques. *Le Français Moderne* 44, 193–228.
- Fouché, Pierre. 1933–1934. L’évolution phonétique du français du XVI^e siècle à nos jours. *Le Français Moderne* 4, 217–236.
- Fougeron, Cécile. 2001a. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29, 109–135.
- Fougeron, Cécile. 2001b. Domain-initial articulatory variations in French. *Journal of Phonetics* 29, 109–135.

- Fougeron, Cécile, Odile Bagou, Muriel Stefanuto and Uli H. Frauenfelder. 2002. A la recherche d'indices de frontière lexicale dans la resyllabation. In *Proceedings of the 24th Journées d'Étude sur la Parole*. Nancy.
- Fougeron, Cécile and Sun-Ah Jun. 1998. Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics* 26, 45–69.
- Fowler, Carol A., Elena T. Levy and Julie M. Brown. 1997. Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language* 37, 24–40.
- Frisch, Stefan A., Nathan R. Large and David B. Pisoni. 2000. Perception of word-likeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42, 481–496.
- Fujisaki, Hiroya and Keikichi Hirose. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustic Society of Japan* 5(4), 233–241.
- Garde, Paul. 1968. *L'accent*. Paris: Presses Universitaires de France.
- Garnes, Sara and Zinny S. Bond. 1980. A slip of the ear? A snip of the ear? A slip of the year? In Victoria A. Fromkin (ed.), *Errors in Linguistic Performance: Slips of the tongue, ear, pen and hand*, New York: Academic Press.
- Gaskell, M. Gareth, Elsa Spinelli and Fanny Meunier. 2002. Perception of resyllabification French. *Memory & Cognition* 30(5), 798–810.
- Gerken, LouAnn, Peter W. Jusczyk and Denise R. Mandel. 1994. When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition* 51, 237–265.
- Gerken, LouAnn and Bonnie J. McIntosh. 1993. Interplay of function morphemes and prosody in early language. *Developmental Psychology* 29, 448–457.
- Gilliam, Terry and Terry Jones. 1975. *Monty Python and the Holy Grail*. film.
- Goldsmith, John. 1996. Phonological Theory. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, Cambridge, Mass./Oxford, England: Blackwell, pp. 1–23.
- Goslin, Jeremy and Uli H. Frauenfelder. 2000. A Comparison of theoretical and human syllabification. *Language and Speech* 44(4), 409–436.
- Grammont, Maurice. 1914. *Traité pratique de prononciation française*. Paris: Delagrave, 1st ed. [9th edition, 1938].

- Grammont, Maurice. 1963. *Traité de phonétique*. Delagrave.
- Grice, Martine. 1995. Leading tones and downstep in English. *Phonology* 12, 183–233.
- Grice, Martine, D. Robert Ladd and Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17(2), 143–185.
- Grosjean, François. 1980. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28, 267–283.
- Gussenhoven, Carlos. 1984. *On the Grammar and Semantics of Sentence Accent*. Dordrecht/Cinnaminson: Foris.
- Hallé, Pierre, Bénédicte de Boysson-Bardies and Marilyn M. Vihman. 1991. Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by French and Japanese infants. *Language and Speech* 34, 299–318.
- Harrington, J., G. Watson and M. Cooper. 1989. Word boundary detection in broad class and phoneme strings. *Computer Speech and Language* 3, 367–382.
- Hashimoto-Yue, Oi-kan. 1972. *Phonology of Cantonese*. Cambridge: Cambridge University Press.
- Hay, Jen, Janet Pierrehumbert and Mary Beckman. in press. Speech perception, well-formedness and the statistics of the lexicon. In *Papers in Laboratory Phonology VI*, Cambridge: Cambridge University Press.
- Hay, Jennifer Bohun. 2000. Causes and Consequences of Word Structure. Doctoral thesis, Northwestern University.
- Hazan, Valerie and Sarah Barrett. 2000. The development of phonemic categorization in children aged 6–12. *Journal of Phonetics* 28(4), 377–396.
- Hirose, Keikichi. 1997. Disambiguating recognition results by prosodic features. In Yoshinori Sagisaka, Nick Campbell and Norio Higuchi (eds.), *Computing Prosody*, New York: Springer, pp. 327–342.
- Hirst, Daniel and Albert Di Cristo. 1984. French intonation: A parametric approach. *Die Neueren Sprachen* 83(5), 554–569.
- Hirst, Daniel and Albert Di Cristo. 1996. Y-a-t-il des unités tonales en français? In *Actes des XXIèmes Journées d'Etudes sur la Parole*. Avignon, pp. 223–226.
- Hirst, Daniel and Albert Di Cristo. 1998. A survey of intonation systems. In Daniel Hirst and Albert Di Cristo (eds.), *Intonation Systems: A survey of twenty languages*, Cambridge: Cambridge University Press, pp. 1–44.

- Holm, Bleicke. 2003. SFC: un modèle de superposition de contours multi-paramétriques pour la génération automatique de la prosodie – Apprentissage automatique et application à l’énociation de formules mathématiques. Doctoral thesis, Institut National Polytechnique de Grenoble.
- Jacobson, J.L., D.C. Boersma, R.B. Fields and K.L. Olson. 1983. Paralinguistic features of adult speech to infants and small children. *Child Development* 54, 436–442.
- Johnson, Keith. 1997. Speech perception without speaker normalization. In K. Johnson and J. Mullenix (eds.), *Talker variability in speech processing*, San Diego: Academic Press, pp. 146–165.
- Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15, 189–226.
- Jun, Sun-Ah and Cécile Fougeron. 1995. The accentual phrase and the prosodic structure of French. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*. vol. 2, pp. 722–725.
- Jun, Sun-Ah and Cécile Fougeron. 2000. A phonological model of French intonation. In A. Botinis (ed.), *Intonation: Analysis, modelling and technology*, Boston: Kluwer.
- Jun, Sun-Ah and Cécile Fougeron. 2002. Realizations of Accentual Phrase in French. *Probus* 14, 147–172. A special issue on the intonation of Romance languages, edited by J. Hualde.
- Junqua, Jean-Claude. 1993. The Lombard reflex and its role on human listeners and automatic speech recognizers. *Journal of the Acoustical Society of America* 93(1), 510–24.
- Kahn, Daniel. 1976. Syllable-based Generalizations in English Phonology. Ph.D. thesis, Massachusetts Institute of Technology.
- Lacheret-Dujour, Anne and Frédéric Beaugendre. 1999. *La prosodie du français*. Paris: CNRS Editions.
- Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Laeuffer, Christiane. 1987. Some constraints on the domain of phrasal resyllabification in French. *The Ohio State University Working Papers in Linguistics* 36, 75–100.
- Laniran, Yetunde O. and G. Nick Clements. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* 31(2), 203–250.

- Lavoie, Lisa. 2002. Some influences on the realization of *for* and *four* in American English. *Journal of the International Phonetic Association* 32, 175–202.
- Lehiste, Ilse. 1960. An acoustic-phonetic study of internal open juncture. *Phonetica* 5(1–54).
- Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51, 2018–2024.
- Lieberman, Philip. 1980. The innate central aspect of intonation. In Linda R. Waugh and C.H. van Schooneveld (eds.), *The Melody of Language*, Baltimore: University Park Press, pp. 187–199.
- Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H and H Theory. In W. Hardcastle and A. Marchal (eds.), *Speech production and speech modelling*, Dordrecht: Kluwer, pp. 403–440.
- Lœvenbruck, Hélène. 1999. An investigation of articulatory correlates of the Accentual Phrase in French. In *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, vol. 1.
- Lœvenbruck, Hélène. 2000. Effets articulatoires de l’emphase contrastive sur la Phrase Accentuelle en français. *Actes des XXIIIèmes Journées d’Etude sur la Parole* pp. 165–168.
- Louali, Naïma and Gilbert Puech. 1999. La syllabe en berbère tachelhit. In *Journée d’Etudes Linguistiques*. Nantes, pp. 35–40.
- Marslen-Wilson, William D. and Alan Welsh. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10, 29–63.
- Marslen-Wilson, William D. and Pienie Zwitserlood. 1989. Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance* 15, 576–585.
- Martin, Philippe. 1979. Une théorie syntaxique de l’accentuation en français. *Studia Phonetica* 15, 1–12.
- Martin, Philippe. 1987. Prosodic and rhythmic structure in French. *Linguistics* 25, 925–949.
- Martin, Philippe. 1998. Association prosodie-syntaxe: validation par synthèse. In *Proceedings of the 22nd Journées d’Étude sur la Parole*. Martigny, pp. 119–122.

- Martin, Philippe. 1999. Prosodie des langues romanes: analyse phonétique et phonologie, Recherches sur le français parlé. *G.A.R.S.* 15, 233–253. Publications de l’Université de Provence.
- Martin, Philippe. 2003. ToBI: L’illusion scientifique? In Véronique Aubergé, Anee Lacheret-Dujour and Hélène Lœvenbruck (eds.), *Proceedings of the Journées Prosodie*. Grenoble: Institut de la Communication Parlée, pp. 109–113.
- Mattys, Sven L. 2003. Stress-based speech segmentation revisited. In *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology*. Geneva, pp. 121–124.
- Mattys, Sven L., Peter W. Jusczyk, Paul A. Luce and James L. Morgan. 1999. Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology* 38, 465–494.
- McQueen, James and Mark Pitt. 1996. Transitional probability and phoneme monitoring. In *Proceedings of the International Conference on Spoken Language Processing*. Delaware: Alfred I. du Pont Institute, vol. 4, pp. 2502–2505.
- McQueen, James M. 1998. Segmentation of continuous speech using phonotactics. *Journal of Memory and Language* 39, 21–46.
- McQueen, James M., Takashi Otake and Anne Cutler. 2001. Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language* 45, 103–132.
- Mehler, Jacques, Jean Yves Dommergues, Uli H. Frauenfelder and Juan Seguí. 1981. The syllable’s role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20, 298–305.
- Mehler, Jacques, Emmanuel Dupoux and Juan Seguí. 1990. Constraining models of lexical access: the onset of word recognition. In G.T.M. Altmann (ed.), *Cognitive Models of Speech Processing*, MIT Press, pp. 236–261.
- Mehler, Jacques, Peter W. Jusczyk, Ghislaine Lambertz, Nilofar Halsted, Josiane Bertoncini and Claudine Amiel-Tison. 1988. A precursor of language acquisition in young infants. *Cognition* 29, 143–178.
- Mertens, Piet. 1987. L’intonation du français. De la description linguistique à la reconnaissance automatique. Doctoral thesis, Université de Leuven.
- Mertens, Piet. 1993. Accentuation, intonation et morphosyntaxe. *Travaux de Linguistique* 26, 21–69.

- Mertens, Piet. 2002. Synthesizing Elaborate Intonation Contours in Text-to-Speech for French. In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence, pp. 499–502.
- Mertens, Piet, Jean-Philippe Goldman, Éric Wehril and Arnaud Gaudinat. 2001. La synthèse de l’intonation à partir de structures syntaxiques riches. *Traitement Automatique des Langues* 42(1), 145–192.
- Moon, Christine, Robin Panneton Cooper and William P. Fifer. 1993. Two-day-olds prefer their native language. *Infant Behavior and Development* 16, 495–500.
- Morlec, Yann. 1997. Génération multiparamétrique de la prosodie du français par apprentissage automatique. Doctoral thesis, Institut de la Communication Parlée, Grenoble.
- Morlec, Yann, Gérard Bailly and Véronique Aubergé. 2001. Generating prosodic attitudes in French: Data, model and evaluation. *Speech Communication* pp. 357–371.
- Moulines, Eric and Francis Charpentier. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* pp. 453–467.
- Munson, Benjamin. 2000. Phonological Pattern Frequency and Speech Production in Children and Adults. Doctoral thesis, The Ohio State University.
- Nakatani, Lloyd H. and Kathleen D. Dukes. 1977. Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62, 714–719.
- Nazzi, Thierry, Josiane Bertoncini and Jacques Mehler. 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.
- Nespor, Marina and Irene Vogel. 1982. Prosodic domains of external sandhi rules. In H. van der Hulst and N. Smith (eds.), *The Structure of Phonological Representations, part I*, Dordrecht/Cinnaminson: Foris, pp. 225–265.
- Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht/Riverton: Foris.
- New, B., C. Pallier, L. Ferrand and R. Matos. 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L’Année Psychologique* 101(447–462). Available from <http://www.lexique.org>.

- Norris, Dennis, James M. McQueen and Anne Cutler. 1995. Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Human perception and performance* 21, 1209–1228.
- Norris, Dennis, James M. McQueen, Anne Cutler and Sally Butterfield. 1997. The Possible-Word Constraint in the segmentation of continuous speech. *Cognitive Psychology* 34, 191–243.
- Norris, Dennis, James M. McQueen, Anne Cutler, Sally Butterfield and Ruth Kearns. 2001. Language-universal constraints on speech segmentation. *Language and Cognitive Processes* 16(5/6), 637–660.
- Ohala, John J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41, 1–16.
- Otake, Takashi, G. Hatano, Anne Cutler and Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32, 258–278.
- Otake, Takashi, Giyoo Hatano and Kiyoko Yoneyama. 1996. Speech segmentation by Japanese listeners. In T. Otake and A. Cutler (eds.), *Phonological Structure and Language Processing: Cross-linguistic studies*, Berlin and New York: Mouton de Gruyter, Speech Research 12, pp. 183–201.
- Pasdeloup, Valérie. 1990. Modèle de règles rythmiques du français appliqué à la synthèse de la parole. Ph.D. thesis, Institut de Phonétique d’Aix-en-Provence, Université de Provence.
- Peperkamp, Sharon and Emmanuel Dupoux. 2002. A typological study of stress “deafness”. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology VII*, Berlin: Mouton de Gruyter, pp. 203–240.
- Percy, Thomas. 1906. *Reliques of Ancient English Poetry*. London and Bungay: Richard Clay and Sons. [Originally published 1765.].
- Pierrehumbert, Janet and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip Cohen, Jerry Morgan and Martha E. Pollack (eds.), *Intentions in Communication*, Cambridge: MIT Press, pp. 271–311.
- Pierrehumbert, Janet B. 1980. The Phonology and Phonetics of English Intonation. Doctoral thesis, Massachusetts Institute of Technology. Distributed 1987 by Indiana University Linguistics Club.
- Pierrehumbert, Janet B. 2000. Tonal elements and their alignment. In Merle Horne (ed.), *Prosody, Theory and Experiment: Studies presented to Gösta Bruce*, Kluwer.

- Pierrehumbert, Janet B. 2001a. Exemplar dynamics: Word frequency, lenition and contrast. In J.L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam: Benjamins, pp. 137–157.
- Pierrehumbert, Janet B. 2001b. Why phonological constraints are so coarse-grained. *Language and Cognitive Processes* 16, 691–698.
- Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology VII*, Berlin: Mouton de Gruyter, pp. 101–139.
- Pierrehumbert, Janet B. 2003. Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay and S. Jannedy (eds.), *Probabilistic Linguistics*, Cambridge, Mass.: MIT Press, pp. 177–228.
- Pierrehumbert, Janet B. and Mary E. Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: The MIT Press.
- Pitt, Mark and James McQueen. 1998. Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39, 347–370.
- Post, Brechtje. 1999. Solving a controversy in the analysis of French rising pitch movements. In *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco.
- Post, Brechtje. 2000. *Tonal and Phrasal Structures in French Intonation*. The Hague: Thesus.
- Post, Brechtje. 2002. French tonal structures. In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence, pp. 11–13.
- Prieto, Pilar, Jan van Santen and Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23, 429–451.
- Prince, Alan and Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Manuscript. Rutgers University and University of Colorado.
- Pulgram, Ernst. 1965. Prosodic Systems: French. *Lingua: International review of general linguistics* 13, 125–144.
- Rabelais, François. 1997. *Gargantua*. Imprimerie Nationale Editions. Originally published, 1534. Presentation and notes by Nicole Cazauran. Transcription of the text by Marie-Claire Thomine.
- Reinholt Petersen, Niels. 1986. Perceptual compensation for segmentally conditioned fundamental frequency perturbation. *Phonetica* 43, 31–42.

- Rietveld, A.C.M. 1980. French Word Boundaries. *Language and Speech* 23(3), 289–296.
- Rolland, Guillaume and Hélène Lœvenbruck. 2002. Characteristics of the Accental Phrase in French: An acoustic, articulatory and perceptual study. In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence, pp. 611–614.
- Rossi, Mario. 1978. La perception des glissandos descendants dans les contours prosodiques. *Phonetica* 35, 11–40.
- Rossi, Mario. 1980. Le français, langue sans accent ? *Studia Phonetica* 15, 13–51.
- Rossi, Mario. 1981. Vers une théorie de l'intonation. In Mario Rossi, Albert Di Cristo, Daniel Hirst, Philippe Martin and Y. Nishinuma (eds.), *L'Intonation: de l'acoustique à la sémantique*, Paris: Klincksieck.
- Rossi, Mario. 1985. L'intonation et l'organisation de l'énoncé. *Phonetica* 42, 135–153.
- Rossi, Mario. 1999. *L'intonation: Le système du français*. Gap: Orpheus.
- Sadeniemi, Matti. 1949. *Metriikkamme perusteet: ja sovellutusta moderneihin ja antiikin mittoihin*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Saffran, Jenny R., Elissa L. Newport and Richard N. Aslin. 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35, 606–621.
- Schafer, Amy J., Shari R. Speer, Paul Warren and S. David White. 2000. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29(2), 169–182.
- Sebastián-Gallés, Núria. 1996. The role of accent in speech perception. In T. Otake and A. Cutler (eds.), *Phonological Structure and Language Processing: Cross-linguistic studies*, Berlin and New York: Mouton de Gruyter, Speech Research 12, pp. 171–182.
- Sebastián-Gallés, Núria, Emmanuel Dupoux, Juan Seguí and Jacques Mehler. 1992. Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language* 31, 18–32.
- Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3, 371–405.
- Selkirk, Elisabeth O. 1984. *Phonology and Syntax: The relation between sound and structure*. Cambridge, Mass.: MIT Press.

- Shafer, V., L. A. Gerken, J. Shucard and D. Shucard. 1992. "The" and the brain: An electrophysiological study of infants' sensitivity to English function morphemes. Paper presented at the Boston University Conference on Language Development, Boston. Cited in Christophe et al. (1997).
- Slowiacek, L.M. 1990. Effects of lexical stress in auditory word recognition. *Language and Speech* 33, 47–68.
- Smith, Mary R., Anne Cutler, Sally Butterfield and Ian Nimmo-Smith. 1989. The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Science Research* 32(4), 912–920.
- Spinelli, Elsa, Anne Cutler and James M. McQueen. 2002. Resolution of liaison for lexical access in French. *Revue Française de Linguistique Appliquée* 7, 83–96.
- Spinelli, Elsa, James M. McQueen and Anne Cutler. 2003. Processing resyllabified words in French. *Journal of Memory and Language* 48, 233–254.
- Straub, Kathleen Ann. 1997. The Production of Prosodic Cues and their Role in the Comprehension of Syntactically Ambiguous Sentences. Ph.D. thesis, University of Rochester.
- Suomi, Kari, James M. McQueen and Anne Cutler. 1997. Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language* 36, 422–444.
- Tabain, Marija. 2003. Effects of prosodic boundary on /aC/ sequences: Acoustic results. *Journal of the Acoustical Society of America* 113, 516–531.
- Tabossi, Patrizia, Simona Collina, Michela Mazzetti and Marina Zoppelle. 2000. Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human perception and performance* 26, 758–775.
- Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff and Margo Bowman. 2000. English speakers' sensitivity to phonotactic patterns. In *Papers in Laboratory Phonology V: Acquisition and the lexicon*, Cambridge: Cambridge University Press.
- Tyler, Lorraine K. and Jeanine Wessels. 1983. Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics* 34, 409–420.
- Tyler, Lorraine K. and Jeanine Wessels. 1985. Is gating an on-line task? Evidence from naming latency data. *Perception and Psychophysics* 38, 217–222.
- Vaissière, Jacqueline. 1983. Language-independent prosodic features. In Anne Cutler and D. Robert Ladd (eds.), *Prosody: Models and measurements*, Berlin: Springer.

- Vaissière, Jacqueline. 1995. Phonetic explanations for cross-linguistic prosodic similarities. *Phonetica* 52, 123–130.
- Vaissière, Jacqueline. 1997. Langues, prosodies et syntaxe. *Traitement Automatique des Langues* 38(1), 53–82.
- Vaissière, Jacqueline. 2002. Cross-linguistic prosodic transcription: French versus English. In N. B. Volskaya, N. D. Svetozarova and P.A. Skrelin (eds.), *Problems and Methods of Experimental Phonetics: In honor of the 70th anniversary of Prof. L. V. Bondarko*, St. Petersburg: St. Petersburg University Press, pp. 147–164.
- van der Lugt, Arie. 1999. From Speech to Words. Doctoral thesis, Katholieke Universiteit Nijmegen.
- Verluyten, S. Paul. 1982. Recherches sur la prosodie et la métrique du français. Doctoral thesis, Universiteit Antwerpen.
- Vihman, Marilyn May. 1996. *Phonological Development: The origins of language in the child*. Cambridge, Mass.: Blackwell.
- Vitevitch, Michael S. and Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40, 374–408.
- Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce and David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40, 47–62.
- Vroomen, Jean and Beatrice de Gelder. 1995. Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human perception and performance* 21, 98–108.
- Vroomen, Jean, Jyrki Tuomainen and Beatrice de Gelder. 1998. The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language* 38, 133–149.
- Vroomen, Jean, Monique von Zon and Beatrice de Gelder. 1996. Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition* 24, 744–755.
- Watt, Sheila M. and Wayne S. Murray. 1996. Prosodic form and parsing commitments. *Journal of Psycholinguistic Research* 25(2), 291–318.
- Wauquier-Gravelines, Sophie. 1996. Organisation phonologique et traitement de la parole continue, Contraintes prosodiques et phonologiques de l'accès lexical. Ph.D. thesis, Université Paris 7 Denis Diderot.

- Weber, Andrea. 2000. The role of phonotactics in the segmentation of native and non-native continuous speech. In A. Cutler, J. McQueen and R. Zondervan (eds.), *Proceedings of the Workshop on Spoken Access Processes*. Nijmegen: Max Planck Institute for Psycholinguistics, pp. 143–146.
- Welby, Pauline. 2002. The realization of early and late rises in French: A production study. In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*. Aix-en-Provence, pp. 695–698.
- Welby, Pauline. 2003. Effects of pitch accent type and status on focus projection. *Language and Speech* 46(1), 53–81.
- Wenk, Brian J. and François Wioland. 1982. Is French really syllable-timed? *Journal of Phonetics* 10, 193–216.
- Wong, Wai Yi P., Marjorie K.-M. Chan and Mary E. Beckman. 1999. An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In Sun-Ah Jun (ed.), *Prosodic Models and Transcription: Towards Prosodic Typology*. Oxford University Press. [Collection of papers from the 14th International Congress of Phonetic Sciences 1999 Satellite Workshop on Intonation: Models and ToBI Labeling, San Francisco. 1-7 August 1999].
- Wright, Sylvia. 1954. The death of Lady Mondegreen. *Harper's Magazine* 209, 48–51.
- Yersin-Besson, Carole and François Grosjean. 1996. L'effet de l'enchaînement sur la reconnaissance des mots dans la parole continue. *L'Année Psychologique* 96, 9–30.
- Yip, Moira. 1993. Tonal register in East Asian languages. In H. van der Hulst and K. Snider (eds.), *The Phonology of Tone: The Representation of Tonal Register*, Berlin/ New York: Mouton de Gruyter, pp. 245–268.
- Yip, Moira. 1995. Tone in East Asian languages. In J. Goldsmith (ed.), *Handbook of Phonological Theory*, Oxford: Basil Blackwell, pp. 476–494.
- Yip, Moira Jean. 1980. The tonal phonology of Chinese. Doctoral thesis, Massachusetts Institute of Technology. [published in 1991 by Garland (New York)].
- Zellner, Brigitte. 1998. Caractérisation du débit de parole en français. In *Proceedings of the 22nd Journées d'Étude sur la Parole*. Martigny.
- Zellner-Keller, Brigitte. in press. Prediction of temporal structures for various speech rates. In *Progress in Speech Synthesis II*, Berlin/Heidelberg: Springer.
- Zwitserslood, Pienie. 1989. The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32, 25–64.

APPENDIX A

PRODUCTION EXPERIMENT INSTRUCTIONS

The instructions were read aloud to each participant before the recording at each speaking rate. An English translation is given after the instructions.

Instructions

Cette expérience comprend deux parties et dure une heure environ.

Lors de la première partie, vous allez lire trois fois une liste de phrases. Lisez chaque phrase mentalement avant de la prononcer à haute voix. Parlez aussi clairement que possible, d'une voix normale et à une vitesse normale. Essayez de ne pas faire de bruits parasites (en tournant les pages ou en touchant le microphone, par exemple). Si, toutefois, vous hésitez ou toussiez en prononçant une phrase, ce n'est pas grave; répétez simplement la phrase. N'oubliez pas que vous pouvez cesser de participer à tout moment sans pénalité. Avez-vous des questions ?

Vous voulez un verre d'eau? Vous voulez faire une pause?

[after the first reading of the sentences (at a normal speaking rate)]

Maintenant vous allez relire les phrases, mais cette fois-ci, parlez plus rapidement. Si vous faites une erreur, répétez simplement la phrase. Parlez plus vite que la première fois. Mais assurez-vous de rester intelligible. Est-ce que vous êtes prêt(e)? Vous pouvez commencer.

English translation of the instructions

This experiment consists of two parts and lasts one hour.

During the first part, you are going to read a list of sentences three times. Read each sentence silently before saying it out loud. Speak as clearly as possible, in a normal voice at a normal speed. Try not to make any other noise (while turning the pages or by touching the microphone, for example). If, however, you hesitate or cough while saying a sentence, it's okay; just repeat the sentence. Don't forget that

you can discontinue your participation at any point, without penalty. Do you have any questions?

Would you like a glass of water? Would you like to take a break?

[after the first reading of the sentences (at a normal speaking rate)]

Now you are going to reread the sentences, but this time, speak faster. If you make a mistake, just repeat the sentence. Speak faster than the first time, but make sure that your speech stays intelligible. Are you ready? You can begin.

APPENDIX B

PRODUCTION EXPERIMENT MATERIALS

Target phrases are underlined. The target phrases have two-syllable, (items 1–9), three-syllable (items 10–18), or four-syllable content words (items 19–27). In the (a) condition, the target phrase is utterance initial, with one preceding function word. In the (b) condition, it is utterance initial, with two preceding function words. In the (c) condition, it is utterance medial, with one preceding function word. In the (d) condition, it is utterance medial, with two preceding function words.

Within each set, an English translation is given only for the (c) condition (the longest) sentences.

1. a. Le malaise va être très craint par Julien.
b. Et le malaise va être très craint par Julien.
c. L'hésitation, le malaise et la méchanceté vont être très craints par Julien.
d. L'hésitation et le malaise vont être très craints par Julien.
 'Hesitation, malaise, and meanness are going to be very much feared by Julien.'
2. a. Le moulin va être apprécié par Hélène.
b. Et le moulin va être apprécié par Hélène.
c. Le beau paysage, le moulin et la cathédrale vont être appréciés par Hélène.
d. Le beau paysage et le moulin vont être appréciés par Hélène.
 'The beautiful countryside, the mill, and the cathedral are going to be appreciated by Hélène.'
3. a. Le minime sera captivé par le match.
b. Et le minime sera captivé par le match.
c. Les entraîneurs, le minime et les spectateurs seront captivés par le match.
d. Les entraîneurs et le minime seront captivés par le match.
 'The coach, the young player, and the spectators will be captivated by the match.'
4. a. La mignonne était observée par l'espion.

- b. Et la mignonne était observée par l'espion.
- c. La courageuse, la mignonne et le malveillant étaient observés par l'espion.
- d. La courageuse et la mignonne étaient observés par l'espion.
- 'The courageous woman and the cute woman were watched by the spy.'
5. a. La lumière était désirée par Aimée.
- b. Et la lumière était désirée par Aimée.
- c. La solitude, la lumière et la propreté étaient désirées par Aimée.
- d. La solitude et la lumière étaient désirées par Aimée.
- 'Solitude, light, and cleanliness were desired by Aimée.'
6. a. La mouillure est un vrai problème au sous-sol.
- b. Et la mouillure est un vrai problème au sous-sol.
- c. La moisissure, la mouillure et l'humidité sont de vrais problèmes au sous-sol.
- d. La moisissure et la mouillure sont de vrais problèmes au sous-sol.
- 'Mildew, dampness, and humidity are real problems in the basement.'
7. a. Les olives ont été mangées à midi.
- b. Et les olives ont été mangées à midi.
- c. Le camembert, les olives et le reblochon ont été mangés à midi.
- d. Le camembert et les olives ont été mangés à midi.
- 'The camembert, the olives, and the reblochon were eaten at lunchtime.'
8. a. Les anneaux ont été achetés ce matin.
- b. Et les anneaux ont été achetés ce matin.
- c. Les ornements, les anneaux et les tapisseries ont été achetés ce matin.
- d. Les ornements et les anneaux ont été achetés ce matin.
- 'The ornaments, the rings, and the tapestries were bought this morning.'
9. a. Les annonces étaient sûrement lues soigneusement.
- b. Et les annonces étaient sûrement lus soigneusement.
- c. L'hebdomadaire, les annonces et le manuscrit étaient sûrement lus soigneusement.
- d. L'hebdomadaire et les annonces étaient sûrement lus soigneusement.
- 'The weekly magazine, the announcements, and the manuscript were surely carefully read.'
10. a. Le magnanime peut être trouvé dans le café.
- b. Et le magnanime peut être trouvé dans le café.
- c. La courageuse, le magnanime et le valeureux peuvent être trouvés dans le café.

- d. La courageuse et le magnanime peuvent être trouvés dans le café.
 ‘The courageous woman, the magnanimous man, and the brave man can be found in the cafe.’
11. a. Le mélomane s’était disputé avec eux.
 b. Et le mélomane s’était disputé avec eux.
 c. Le randonneur, le mélomane et la forestière s’étaient disputés avec eux.
 d. Le randonneur et le mélomane s’étaient disputés avec eux.
 ‘The hiker and the music lover had argued with them.’
12. a. Le minimum sera calculé par Manon.
 b. Et le minimum sera calculé par Manon.
 c. Le maximum, le minimum et les écart-types seront calculés par Manon.
 d. Le maximum et le minimum seront calculés par Manon.
 ‘The maximum, the minimum, and the standard deviations will be calculated by Manon.’
13. a. La mélanine était étudiée à la fac.
 b. Et la mélanine était étudiée à la fac.
 c. Le mélanome, la mélanine et le collagène étaient étudiés à la fac.
 d. Le mélanome et la mélanine étaient étudiés à la fac.
 ‘Melanomas and melanine were studied at college.’
14. a. La limonade a été versée par Anna.
 b. Et la limonade a été versée par Anna.
 c. La grenadine, la limonade et l’Orangina ont été versés par Anna.
 d. La grenadine et la limonade ont été versées par Anna.
 ‘The grenadine, the lemonade and the Orangina were poured by Anna.’
15. a. La lanoline a été vendue à la ferme.
 b. Et la lanoline a été vendue à la ferme.
 c. Le basilic, la lanoline et les aubergines ont été vendus à la ferme.
 d. Le basilic et la lanoline ont été vendus à la ferme.
 ‘The basil, the lanolin, and the eggplants were sold at the farm.’
16. a. Les animaux sont pris en photo frequemment.
 b. Et les animaux sont pris en photo frequemment.
 c. Le monument, les animaux et les villageois sont pris en photo frequemment.
 d. Le monument et les animaux sont pris en photo frequemment.
 ‘Pictures of the monument, the animals, and the villagers are often taken.’

17. a. Les anémones étaient dessinés rapidement.
 b. Et les anémones étaient dessinés rapidement.
 c. Le géranium, les anémones et le magnolia étaient dessinés rapidement.
 d. Le géranium et les anémones étaient dessinés rapidement.
 ‘The geranium, the anemones, and the magnolia were quickly drawn.’
18. a. Les homonymes sont utilisés par le poète.
 b. Et les homonymes sont utilisés par le poète.
 c. Les enchaînements, les homonymes et les métaphores sont utilisés par le poète.
 d. Les enchaînements et les homonymes sont utilisés par le poète.
 ‘Enchaînements, homonyms and metaphors are used by the poet.’
19. a. Le linoléum sera installé mercredi.
 b. Et le linoléum sera installé mercredi.
 c. L’aluminium, le linoléum et le formica seront installés mercredi.
 d. L’aluminium et le linoléum seront installés mercredi.
 ‘The aluminum, the linoleum, and the formica will be installed Wednesday.’
20. a. Le malavisé va être évoqué par René.
 b. Et le malavisé va être évoqué par René.
 c. Le reporter, le malavisé et le dissolu vont être évoqués par René.
 d. Le reporter et le malavisé vont être évoqués par René.
 ‘The reporter, the unwise one, and the dissolute one are going to be evoked by René.’
21. a. Le méli-mélo va déconcentrer Mélanie.
 b. Et le méli-mélo va déconcentrer Mélanie.
 c. L’incohérence, le méli-mélo et la confusion vont déconcentrer Mélanie.
 d. L’incohérence et le méli-mélo vont déconcentrer Mélanie.
 ‘The inconsistency, the disorder, and the disarray are going to distract Mélanie.’
22. a. La misogynie était très mal vue par Noël.
 b. Et la misogynie était très mal vue par Noël.
 c. L’impolitesse, la misogynie et la grossiereté étaient très mal vues par Noël.
 d. L’impolitesse et la misogynie étaient très mal vues par Noël.
 ‘Noël took a dim view of rudeness, misogyny, and vulgarity.’
23. a. La monomanie doit être évitée par les empereurs.
 b. Et la monomanie doit être évitée par les empereurs.
 c. La malveillance, la monomanie et le fanatisme doivent être évités par les empereurs.

- d. La malveillance et la monomanie doivent être évitées par les empereurs.
 ‘Malevolence, monomania, and fanaticism must be avoided by emperors.’
24. a. La mammalogie va être importante pour Jérôme.
 b. Et la mammalogie va être importante pour Jérôme.
 c. L’agriculture, la mammalogie et la fauconerie vont être importantes pour Jérôme.
 d. L’agriculture et la mammalogie vont être importantes pour Jérôme.
 ‘Agriculture, mammalogy, and falconry are going to be important for Jerome.’
25. a. Les anomalies seront toutes oubliées rapidement.
 b. Et les anomalies seront toutes oubliées rapidement.
 c. Les différences, les anomalies et les désaccords seront tous oubliés rapidement.
 d. Les différences et les anomalies seront tous oubliés rapidement.
 ‘The differences, the anomalies, and the disagreements will all be quickly forgotten.’
26. a. Les Himalayens se sont rencontrés l’autre semaine.
 b. Et les Himalayens se sont rencontrés l’autre semaine.
 c. Les photographes, les Himalayens et les Grenoblois se sont rencontrés l’autre semaine.
 d. Les photographes et les Himalayens se sont rencontrés l’autre semaine.
 ‘The photographers, the Himalayans, and the Grenoblois met one another the other week.’
27. a. Les enlumineurs avaient travaillé très longtemps.
 b. Et les enlumineurs avaient travaillé très longtemps.
 c. La cordonnière, les enlumineurs et le violoniste avaient travaillé très longtemps.
 d. La cordonnière et les enlumineurs avaient travaillé très longtemps.
 ‘The cobbler, the illuminators, and the violinist had worked for a very long time.’

APPENDIX C

PERCEPTION EXPERIMENT 1 INSTRUCTIONS

The experimental instructions were presented to participants on the computer screen immediately prior to the experiment. The program used (E-Prime) did not allow us to type accented letters (é, à, etc.). The instructions were therefore presented to participants in capital letters, for which accents are optional. A series of asterisks indicates the beginning of a new screen. An English translation is given after the instructions.

Instructions

ECRIVEZ “ LISTE 1 ” SUR VOTRE FICHE-REPONSE.

APPUYEZ SUR “ ENTER ” POUR CONTINUER.

INSTRUCTIONS

VOUS ALLEZ ECOUTER DES GROUPES DE MOTS ET VOUS DEVREZ
INDIQUER CE QUE VOUS VENEZ D’ENTENDRE EN VOUS SERVANT DE LA
FICHE-REPONSE DEVANT VOUS.

LA PLUPART DE CES GROUPES DE MOTS N’ONT PAS DE SENS. DE PLUS,
UN BRUIT PARASITE ASSEZ FORT A ETE AJOUTE. IL VOUS SERA DONC
DIFFICILE D’ENTENDRE DISTINCTEMENT LES MOTS. NEANMOINS, IL EST
TRES IMPORTANT DE VOUS CONCENTRER ET D’ECOUTER
ATTENTIVEMENT.

APPUYEZ SUR “ ENTER ” POUR CONTINUER.

VOUS ALLEZ COMMENCER PAR DEUX ESSAIS TRES COURTS QUI VOUS
FAMILIARISERONT AVEC LE PRINCIPE DE L’EXPERIENCE. IL N’Y AURA
DE BRUIT QUE LORS DU DEUXIEME ESSAI.

APPUYEZ SUR “ ENTER ” POUR CONTINUER.

AVEZ-VOUS DES QUESTIONS ?

QUAND VOUS SEREZ PRET(E), VEUILLEZ VERIFIER QUE VOS ECOUTEURS SONT BIEN EN PLACE.

LE CHERCHEUR RESTERA ICI PENDANT LES ESSAIS POUR REPONDRE A VOS QUESTIONS.

APPUYEZ SUR “ ENTER ” POUR COMMENCER.

LE PREMIER ESSAI EST TERMINE.

APPUYEZ SUR “ ENTER ” POUR COMMENCER LE DEUXIEME ESSAI.

LE DEUXIEME ESSAI EST TERMINE. ETES VOUS PRET(E) A COMMENCER L'EXPERIENCE ? SI VOUS AVEZ DES QUESTIONS, N'HESITEZ PAS A LES POSER AU CHERCHEUR.

APPUYEZ SUR “ ENTER ” POUR COMMENCER.

L'EXPERIENCE EST TERMINEE.

MERCI DE VOTRE PARTICIPATION !

VEUILLEZ OUVRIR LA PORTE DE LA CHAMBRE SOURDE POUR REJOINDRE LE CHERCHEUR.

English translation of instructions

Write “List 1” on your answer sheet.

Press “Enter” to continue.

Instructions

You are going to listen to groups of words, and you need to indicate what you just heard using the answer sheet in front of you.

Most of these groups of words do not have any meaning. In addition, we have added a fairly loud noise. It will therefore be difficult for you to distinctly hear the words. Nevertheless, it is very important to concentrate and to listen attentively.

Press “Enter” to continue.

You are going to start with two very short practice sessions that will familiarize you with the experiment. There will only be noise during the second practice session.

Press “Enter” to continue.

Do you have any questions?

When you are ready, please check to make sure your headphones are in place.

The experimenter will stay here during the practice sessions to answer your questions.

Press “Enter” to begin.

The first practice is finished.

Press “Enter” to begin the second practice.

The second practice is finished. Are you ready to start the experiment? If you have any questions, do not hesitate to ask the experimenter.

Press “Enter” to begin.

The experiment is finished.

Thank you for your participation.

Please open the sound booth door to rejoin the experimenter.

APPENDIX D

PERCEPTION EXPERIMENT 1 MATERIALS

A sample answer sheet for Perception Experiment 1 is given on the next page. Participants listened to the noise-masked stimuli and indicated what they heard by circling either (a) or (b). There were two answer sheets, to minimize order effects (see §5.2.1). NB: The word *galops* ‘gallops’ was inadvertently misspelled (with two *l*’s, as in English) in the experimental response sheets. This misspelling likely biased participants to the other segmentation. The misspelling was retained in the norming experiment.

Sample Answer Sheet

- | | | | | |
|-----|----|----------------------------|----|-----------------------------|
| 1) | a. | positif et médité | b. | positif et mes dictées |
| 2) | a. | la montagne a nos minets | b. | la montagne a nommé |
| 3) | a. | a tenu à nos lignées | b. | a tenu un moulinet |
| 4) | a. | Nathalie a ménagé | b. | Nathalie a mes nacelles |
| 5) | a. | le noyau a majoré | b. | le noyau a ma journée |
| 6) | a. | a parlé à mon archet | b. | a parlé aux monarchies |
| 7) | a. | original et ma carie | b. | original et macareux |
| 8) | a. | le manuel et un mandat | b. | le manuel et Amanda |
| 9) | a. | revenu et m'a serré | b. | revenu et macéré |
| 10) | a. | le menu au lémurien | b. | le menu ou les mûriers |
| 11) | a. | englouttis sont méprisants | b. | englouttis sont mes prisons |
| 12) | a. | Amélie va en voyage | b. | Amélie va envoyer |
| 13) | a. | le niveau de mes sénats | b. | le niveau de mécénat |
| 14) | a. | laminaire et déloyaux | b. | laminaire et des loyaux |
| 15) | a. | la baleine et Mélanie | b. | la baleine et mes lamées |
| 16) | a. | le melon est malhabile | b. | le melon et ma labile |
| 17) | a. | terminal ou mon olive | b. | terminal ou monolithe |
| 18) | a. | elle menait à nos ragots | b. | elle menait un ouragan |
| 19) | a. | le ballon de mes manteaux | b. | le ballon de mémentos |
| 20) | a. | le lino est enlacé | b. | le lino est en lacets |
| 21) | a. | le trottoir est monogame | b. | le trottoir et mon nougat |
| 22) | a. | Noémie ou magnolia | b. | Noémie ou ma molaire |
| 23) | a. | prodigieux et mes gallops | b. | prodigieux et mégalo |
| 24) | a. | mélomane et malaga | b. | mélomane ou ma lagune |

APPENDIX E

PERCEPTION EXPERIMENT 2 INSTRUCTIONS

The experimental instructions were presented to participants on the computer screen immediately prior to the experiment. The program used (E-Prime) did not allow us to type accented letters (é, à, etc.). The instructions were therefore presented to participants in capital letters, for which accents are optional. A series of asterisks indicates the beginning of a new screen. An English translation is given after the instructions.

Instructions

ECRIVEZ “ LISTE 1 ” SUR VOTRE FICHE-REPONSE.

APPUYEZ SUR “ ENTER ” POUR CONTINUER.

INSTRUCTIONS

VOUS ALLEZ ECOUTER DES PHRASES QUI CONTIENNENT DES MOTS QUE NOUS AVONS INVENTES POUR CETTE EXPERIENCE. TOUS CES MOTS SE TROUVENT DANS LA PHRASE

“ ET X POURRAIT ETRE UNE EXPRESSION UTILISEE PAR LES FRANCAIS ”.

VOUS DEVEZ ECRIRE L'EXPRESSION (LE OU LES MOTS) QUE VOUS ENTENDEZ EN VOUS SERVANT DE LA FICHE-REPONSE DEVANT VOUS.

APPUYEZ SUR “ ENTER ” POUR CONTINUER.

PAR EXEMPLE, SI VOUS ENTENDIEZ

“ ET MON GAFRAUDON POURRAIT ETRE UNE EXPRESSION UTILISEE PAR LES FRANCAIS ”,

VOUS ECRIRIEZ “ MON GAFRAUDON ”.

SI VOUS ENTENDIEZ

“ ET GATINAUBAS POURRAIT ETRE UNE EXPRESSION UTILISEE PAR LES FRANCAIS ”,

VOUS ECRIRIEZ “ GATINAUBAS ”.

APPUYEZ SUR “ ENTER ” POUR CONTINUER.

VOUS ALLEZ COMMENCER PAR UN ESSAI TRES COURT QUI VOUS FAMILIARISERA AVEC LE PRINCIPE DE L'EXPERIENCE.

APPUYEZ SUR " ENTER " POUR CONTINUER.

AVEZ-VOUS DES QUESTIONS ?

QUAND VOUS SEREZ PRET(E), VEUILLEZ VERIFIER QUE VOS ECOUTEURS SONT BIEN EN PLACE.

LE CHERCHEUR RESTERA ICI PENDANT L'ESSAI POUR REpondre A VOS QUESTIONS.

APPUYEZ SUR " ENTER " POUR COMMENCER

L'ESSAI EST TERMINE. ETES VOUS PRET(E) A COMMENCER L'EXPERIENCE ? SI VOUS AVEZ DES QUESTIONS, N'HESITEZ PAS A LES POSER AU CHERCHEUR. APPUYEZ SUR " ENTER " POUR COMMENCER L'EXPERIENCE.

L'EXPERIENCE EST TERMINEE.

MERCI DE VOTRE PARTICIPATION !

VEUILLEZ OUVRIR LA PORTE DE LA CHAMBRE SOURDE POUR REJOINDRE LE CHERCHEUR.

English translation of instructions

Write "List 1" on your answer sheet.

Press "Enter" to continue.

Instructions

You are going to listen to sentences that contain words that we have invented for this experiment. All these words are found in the sentence:

"Et X pourrait être une expression utilisée par les Français."

'And X could be an expression used by the French.'

You should write the expression (the word or words) that you hear using the answer sheet in front of you.

Press "Enter" to continue.

For example, if you heard, "Et mon gafradon pourrait être une expression utilisée par les Français", you would write "mon gafradon".

If you heard, "Et gatinaubas pourrait être une expression utilisée par les Français", you would write "gatinaubas".

Press "Enter" to continue.

You are going to start with a very short practice session that will familiarize you with the experiment.

Press “Enter” to continue.

Do you have any questions?

When you are ready, please check to make sure that your headphones are in place.

The experimenter will stay here during the practice to answer your questions.

Press “Enter” to begin.

The practice is finished. Are you ready to start the experiment? If you have any questions, do not hesitate to ask the experimenter.

Press “Enter” to begin the experiment.

The experiment is finished.

Thank you for your participation.

Please open the sound booth door to rejoin the experimenter.

APPENDIX F

PERCEPTION EXPERIMENT 2 SAMPLE ANSWER SHEET

The format for the answer sheets for Perception Experiment 2 is shown below. Participants listened to the stimuli and “filled in the blank” with the word or words they heard between *Et* and *pourrait* in the carrier phrase:

Et _____ pourrait être une expression utilisée par les Français

‘And _____ could be an expression used by the French’.

Each participant heard 56 items (36 critical items and 20 fillers). The actual answer sheets therefore had 56 lines.

- 1) Et _____ pourrait être une expression utilisée par les Français.
- 2) Et _____ pourrait être une expression utilisée par les Français.
- 3) Et _____ pourrait être une expression utilisée par les Français.
- 4) Et _____ pourrait être une expression utilisée par les Français.
- 5) Et _____ pourrait être une expression utilisée par les Français.

APPENDIX G

PERCEPTION EXPERIMENT 2 MATERIALS

For each critical item, the possible segmentations with and without a boundary at the critical region are given. Segmentations of the types listed were the most common, although other segmentations were possible (e.g. *mané rinal* was a possible “no boundary” response and *ma né rinal* was a possible “boundary” response.)

.	no boundary at critical region	boundary at critical region
1)	manérinale	ma nérinale
2)	lamasoulée	la masoulée
3)	mélamondine	mes lamondines
4)	létrisinale	les trisinales
5)	nobrémateur	nos brémateurs
6)	délinosite	des linosites
7)	maribonelle	ma ribonelle
8)	lanorimelle	la norimelle
9)	lamésomite	la mésomite
10)	légéminelle	les géminelles
11)	déménotaire	des ménotaires
12)	mondoliant	mon doliant
13)	télimailé	tes limailés
14)	ménologame	mes nologames
15)	nomanoulant	nos manoulants
16)	mélouvélant	mes louvéants
17)	lanumineuse	la numineuse
18)	méjaminole	mes jaminoles
19)	léminarole	les minaroles
20)	sénéraliane	ses néralianes
21)	mondandaleur	mon dandaleur
22)	déléomane	des léomanes
23)	mondémulard	mon démulard
24)	macuréaille	ma curéaille
25)	lézonulaire	les zonulaires
26)	maralomie	ma ralomie
27)	lavanadrice	la vanadrice
28)	mondobliment	mon dobliment
29)	dégounaleur	des gounaleurs
30)	téméridelle	tes méridelles
31)	lanéburance	la néburance
32)	magelliforme	ma gelliforme
33)	malunagure	ma lunagure
34)	léramolène	les ramolènes
35)	dénilagards	des nilagards
36)	larainolade	la rainolade

APPENDIX H

PERCEPTION EXPERIMENT 3 INSTRUCTIONS

The experimental instructions were presented to participants on the computer screen immediately prior to the experiment. A series of asterisks indicates the beginning of a new screen. An English translation is given after the actual instructions.

Instructions

Vous allez écouter des phrases contenant des mots inventés. Tous ces mots se trouvent dans la phrase : “ On a lu _____ mollement ”. Votre tâche consiste à choisir entre les deux réponses qui vous sont proposées en cliquant sur votre choix.

Parfois vous entendrez un seul mot.

Par exemple, vous pourriez entendre: “ On a lu maténaubas mollement ”.

Dans ce cas, vous verriez deux choix :

maténaubas

ma ténaubas

et vous cliqueriez sur “ maténaubas ”.

Parfois vous entendrez deux mots : un petit mot comme “ mon ”, “ la ”, etc. suivi d’un mot inventé.

Par exemple, vous pourriez entendre: “ On a lu les périlounes mollement ”.

Dans ce cas, vous verriez deux choix :

lépérilounes

les périlounes

et vous cliqueriez sur “ les périlounes ”.

Quand vous serez prêt(e), veuillez vérifier que votre casque est bien en place. Si vous avez des questions, n’hésitez pas à les poser au chercheur.

Vous allez commencer par un essai très court qui vous familiarisera avec le principe de l’expérience.

L’essai est terminé.

Etes-vous prêt(e) à commencer l’expérience ?

Si vous avez des questions, n’hésitez pas les poser au chercheur.

Veuillez ouvrir la porte de la chambre sourde quand vous aurez fini.

L'expérience est terminée.

Merci de votre participation précieuse.

English translation of instructions

You are going to hear sentences containing invented words. All these words are found in the sentence: "On a lu _____ mollement". 'We read _____ listlessly.'

Your task is to select between the two answers that are given to you by clicking on your choice.

Sometimes you will hear one word.

For example, you could hear: "On a lu maténaubas mollement".

In this case, you would see two choices:

maténaubas

ma ténaubas

and you would click "maténaubas".

Sometimes you will hear two words: a little word like "mon" 'my', "la" 'the', etc.

For example, you could hear: "On a lu les périlounes mollement".

In this case, you would see two choices:

lépérilounes

les périlounes

and you would click "les périlounes".

When you are ready, please check to make sure that your headphones are in place. If you have any questions, do not hesitate to ask the experimenter.

You are going to start with a very short practice session that will familiarize you with the experiment.

The practice is finished.

Are you ready to start the experiment?

If you have any questions, do not hesitate to ask the experimenter.

Please open the door to the sound booth when you are finished.

Click to begin the experiment.

The experiment is finished.

Thank you for your participation.

APPENDIX I

PERCEPTION EXPERIMENT 3 MATERIALS

For each critical item, the one-word and two-word segmentations are given. Since this experiment used a forced choice task, these were the only two segmentations available to participants.

	one-word segmentation	two-word segmentation
1)	laminoulée	la minoulée
2)	lanolimelle	la nolimelle
3)	laménoline	la ménoline
4)	lanyomale	la nyomale
5)	lanémulène	la némulène
6)	lanainolie	la nainolie
7)	laneulinaie	la neulinaie
8)	lanamilée	la namilée
9)	lamunimale	la munimale
10)	léminaillole	les minailloles
11)	lémunolame	les munolames
12)	lénamolène	les namolènes
13)	lélonillau	les lonillaux
14)	lénélagnon	les nélagnons
15)	manéminaie	ma néminaie
16)	malimonelle	ma limonelle
17)	manénaline	ma nénaline
18)	manémulée	ma némulée
19)	manalomie	ma nalomie
20)	maménilome	ma ménilome
21)	malunamue	ma lunamue
22)	maleulinaie	ma leulinaie
23)	manumulaille	ma numulaille
24)	méloménie	mes loménies
25)	méluninale	mes luninales
26)	ménolomane	mes nolomanes
27)	mélaminole	mes laminoles
28)	mémalouneu	mes malouneux
29)	ménanayé	mes nanayés
30)	mélominaille	mes lominailles
31)	mélamondine	mes lamondines
32)	nomalunant	nos malunants
33)	nolinalot	nos linalots
34)	nolélomame	nos lélomames
35)	noniliment	nos niliments
36)	nomalomune	nos malomunes