

Copyright

by

Nicholas Bacuez

2012

The Dissertation Committee for Nicholas Bacuez
certifies that this is the approved version of the following dissertation:

**Automated Pattern Recognition for Intonation (PInt).
An Essay on Intonational Phonology and Categorization**

Committee:

Jean-Pierre Montreuil, Supervisor

Carl Blyth

Barbara Bullock

Katrin Erk

Rajka Smiljanic

**Automated Pattern Recognition for Intonation (P_RI_{nt}).
An Essay on Intonational Phonology and Categorization**

by

Nicholas Bacuez, M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2012

To the memory of
Bob Dawson (1943-2007)

Acknowledgments

Thank you to Jean-Pierre Montreuil. I have been very fortunate to have a wonderful adviser. It is a privilege to have an indefectible source of support and guidance, both in work and in life.

Thank you to a committee of quality whose insightful advice helped me give shape to this dissertation: Carl Blyth, Barbara Bullock, Katrin Erk, and Rajka Smiljanic. I would also like to include Marta Ortega-Llebaria, who was originally on my committee.

Thank you to inspirational professors: Joaquim Brandão de Carvalho (Paris 8) in phonology/phonetics and Lofti A. Zadeh (UC Berkeley) in Fuzzy Set Theory.

Thank you to the Department of French and Italian, and particularly to Daniela Bini, David Birdsong, Jean-Pierre Cauvin, Bryan Donaldson, Chris Bryce, and Chaz Nailor for always making time for me.

Thank you to Alison Pollack, Giulia Hill, and Elisabeth Spohrer. Through them I was able to meet most of the participants involved in this study.

Thank you to all the participants in the elicitation tasks for their time and their very friendly support to my research. This includes the Alliance Française of San Francisco, the French Consulate of San Francisco, and the faculty of the École Française of Berkeley.

Thank you to the Harry Ransom Center, the Department of French and Italian, and The Graduate School for their financial help during my studies at

the University of Austin.

Thank you to my family, to the family of Allison, and to my friends, who supported me throughout the years, in France and in the USA: Catherine, Michaël, and Benoit Bacuez, Yannick Bacuez and Gabrielle Libong, Kevin Bozelka, Pascal Denis and Sabrina Parent, Emilie Destruel Johnson and Chad Johnson, Al DenBleyker and Patricia Totten, Robert DenBleyker and Heather Leal, Marie-Pierre Etard, Bob Hardgrave, David Heath and Paola Giraudo, Muriel Jechoux, Michael Kersey and Karen Cleary (“Les voisins”), François Lagarde, Marie Magambo, Olivier Maillart and Laetitia Kugler, Charles Mignot, Nicole Milhaud and Theys Willemse (1939-2011), Frédéric Miota and Christelle Banon, Bette Oliver, Clare Perry, Betty Slagle, Margaret Slagle, Jim Stephens (1922-2011), Pepe Vallenás, Josselin Vasseur, Edouard Vatinel and Hervé Dupenloup, Stéphanie Villard, and Seth Wolitz.

Thank you to Adrien Méli, for his friendly wisdom.

Thank you to Andrew Suozzo, for his kindness and support.

Thank you to Daniel Friedman, *mi compañero de piso*.

Thank you to Sean Manning, for letting me use his Star Wars blanket.

Thank you to Allison DenBleyker, for her love and care, her patient help with this project, and comforting *potages*.

NICHOLAS BACUEZ

The University of Texas

December 2012

Automated Pattern Recognition for Intonation (PRInt).

An Essay on Intonational Phonology and Categorization

Publication No. _____

Nicholas Bacuez, Ph.D.
The University of Texas at Austin, 2012

Supervisor: Jean-Pierre Montreuil

This dissertation provides experimental evidence for the validity of an intonational phonology. The widely used Autosegmental-Metrical theory contends that the phonological structure of intonation can be expressed with two tonal targets (L/H tones and derivatives) and retrieved from its phonetic implementations. However, it has not been specifically demonstrated so far in a systematic way. This dissertation argues that this view on intonational phonology considers the phonetic forms of intonation as instances of phonologically structured intonational units forming functionally discrete categories (tones and derivatives).

The model of Pattern Recognition for Intonation (PRInt) applies the concepts of categorization (vagueness, prototype, degrees of typicality) to intonation in order to abstract the phonological structure of intonational categories from the ranking, by degree of typicality, of their variations in phonetic implementation.

First, instances belonging to an intonation category are collected. Second, a pattern recognition module, relying on the 4-layer structure protocol, extracts a feature vector from the phonetic data of each instance: a sequence of structurally organized tones (L/H tones and derivatives).

Third, a fuzzy classifier, using two functions (frequency and similarity), organizes the data from the feature vectors of all instances by degree of typicality (grade of membership of values in multisets) and generates the phonological structure of the intonation category, the prototypical pattern, extracted from all instances, and that subsumes them all. It also re-creates the phonetic implementations of the phonological structure but with their features ranked by degree of typicality. This allows the model to distinguish phonologically distinct structures from phonetic variations of the same phonological structure.

The model successfully extracted the phonological intonation structure associated to three modalities of closed questions in French: neutral, doubtful, and surprised. It found that neutral and doubtful closed questions are phonologically distinct while surprise is a phonetic allocontour of the neutral modality, in line with prior characterizations of these patterns. It demonstrated that a bi-tonal phonological structure of intonation can be retrieved from phonetic variations.

A versatile modeling tool, PRInt will be developed to use its acquired knowledge to evaluate the categorical status of novel instances and to extract multiple phonological units from mixed corpora.

Table of Contents

Acknowledgments	v
Abstract	vii
List of Tables	xv
List of Figures	xix
Chapter 1. Introduction	
A case for intonational phonology	1
1.1 Categorization	4
1.2 Pattern Recognition	5
1.3 Intonational phonology as pattern recognition	6
1.4 Research question and hypotheses	8
1.5 The PRInt model - Pattern Recognition for Intonation	8
1.6 Organization of the dissertation	11
Chapter 2. Categorization	13
2.1 The vagueness of symbols	14
2.2 Vagueness: typicality and prototype	16
2.2.1 Vagueness	16
2.2.2 Category structure: family resemblance, typicality, and prototype	20
2.3 Phonemic categorization	26
2.3.1 Phonemes, phones, allophones	26
2.3.2 Allophonic variation: French velar plosives	28
2.3.3 Gradient features, binary categories: Voice Onset Time	29
2.3.4 The case of English /t/: Taylor (2004)	33
2.3.5 American vowels (1 of 2): Peterson & Barney (1952); Hillenbrand et al. (1995)	36

2.3.6	American vowels (2 of 2): “Perceptual magnets”	39
2.4	Conclusion	43
Chapter 3. Intonation		44
3.1	Intonation	44
3.2	Synthetic and analytic studies of intonation	46
3.2.1	Synthetic studies	48
3.2.2	Analytic studies	49
3.2.2.1	Family resemblance, prototype, vagueness	50
3.2.2.2	Autonomous features and hierarchical structure	52
3.2.2.3	Autosegmental Metrical (AM) model	54
3.2.3	Synthetic vs. analytic	59
3.3	Three intonation contours of French	63
3.3.1	Stress and intonation in French	63
3.3.2	Three contours	65
3.4	Automated Pattern Recognition for Intonation: PRI nt	70
3.4.1	Presentation of the model PRI nt	71
3.4.1.1	Feature extraction - ATLM	71
3.4.1.2	Classification - AFC	73
Chapter 4. Pattern Recognition		76
4.1	Pattern Recognition (PR)	85
4.2	Statistical and structural approaches to pattern recognition	87
4.3	Looking for haystacks in Monet’s painting	88
Chapter 5. Fuzzy Set Theory		97
5.1	Computing with words	97
5.2	Crisp and fuzzy sets	98
5.3	Building a fuzzy function	101
5.3.1	Fuzzification	101
5.3.2	Phonemes as graded categories: a VOT fuzzy function .	103
5.4	Defuzzification	107

Chapter 6. Acquisition and preparation of the data	110
6.1 Participants	110
6.2 Material and setting	111
6.3 Elicitation tasks	111
6.3.1 Unmarked closed question	113
6.3.2 Modalities of question	114
6.4 Sentence format	116
6.5 Data preparation	116
Chapter 7. Automated tonal labeling module The 4-layer structure	119
7.1 Living up with different texts	120
7.2 Thinking human: an analytical approach to tonal labeling	121
7.3 Automated Tonal Labeling Module & the 4-layer structure	126
7.3.1 Layer 1: frame of reference	128
7.3.2 Layer 2: scalar quantization	129
7.3.3 Layer 3: isometric grid and pre-tones	138
7.3.4 Layer 4: tones	147
7.3.5 Tonal isometric grid	156
Chapter 8. Automated Fuzzification Classifier (AFC)	165
8.1 Features	165
8.2 Sets	168
8.3 The Automated Fuzzy Classifier	169
8.3.1 Why fuzzy sets? Why fuzzification?	169
8.3.2 Fuzzification 1 of 2: the frequency principle	171
8.3.3 Fuzzification 2 of 2: the similarity principle	176
8.3.4 Assigning a unified grade of membership	179
8.3.5 Grades' subsets: inner ranking and organization	181
8.3.6 Defuzzification	186

Chapter 9. Intonation contour of unmarked closed questions	190
9.1 Overview	190
9.2 The intonation contour of unmarked closed questions	191
9.2.1 Layer 3: pre-tones and the pattern recognition of the pre-tonal intonation contour	191
Values from layers 1 & 2	191
Pre-tonal contours	192
9.2.2 Layer 4: tones and the pattern recognition of the tonal intonation contour	198
Values from layers 1 & 2	200
Anchoring of tones (1): pre-tones	205
Anchoring of tones (2): tonal co-occurrences	206
Scaled distance between tones	208
9.2.3 Prototypical contour	210
Additional data: velocity, angle	210
Chapter 10. Intonation contour of two modalities: surprise and doubt	213
10.1 The intonation contour of the modality of surprise	213
10.1.1 Layer 3: pre-tones and the pattern recognition of the pre-tonal intonation contour	213
Values from layers 1 & 2	213
Pre-tonal contours	214
10.1.2 Layer 4: tones and the pattern recognition of the tonal intonation contour	217
Values from layers 1 & 2	217
Anchoring of tones: pre-tones, pre-tonal co-occurrences	218
Scaled distance between tones	220
10.1.3 Prototypical contour	221
Additional data: velocity, angle	221
10.2 The intonation contour of the modality of doubt	223
10.2.1 Layer 3: pre-tones and the pattern recognition of the pre-tonal intonation contour	223
Values from layers 1 & 2	223

Pre-tonal contours	224
10.2.2 Layer 4: tones and the pattern recognition of the tonal intonation contour	226
Values from layers 1 & 2	226
Anchoring of tones: pre-tones, pre-tonal co-occurrences .	229
Scaled distance between tones	230
10.2.3 Prototypical contour	231
Additional data: velocity, angle	231
Chapter 11. Comparison of the three intonation contours	233
11.1 Tonal specification of the contours	233
11.2 Prototypes: intonational –vague– intension	234
11.3 Phonetic implementation: intonational –vague– extension (degrees of typicality and allocontour)	239
11.3.1 Prototype (\bar{m})	241
11.3.2 Highest degree of typicality ($m_{(x)} = 1$)	243
11.3.3 Borderline degree of typicality ($m_{(x)} = 0.5$): vagueness .	247
11.4 A note on secondary peaks	249
11.5 A note on primary non-final peaks - Secondary contour	251
11.6 Conclusion	254
Chapter 12. Conclusion	256
12.1 Summary of the project	256
12.1.1 The PRInt model	257
12.1.2 Main results of the application of PRInt	258
12.2 Further developments and ameliorations	260
Additional Chapter 13. Phonetic variation and variation of the frame of reference	265
13.1 Scaling, fuzzification, defuzzification	266
13.1.1 Sentence	266
13.1.2 Syllables	268
13.1.3 Fundamental frequency (F_0)	270
13.2 Re-scaling	275

13.2.1 Sentence	276
13.2.2 Syllables	278
13.2.3 Fundamental frequency (F_0)	280
13.3 Re-scaling of the surprise contour - Results	281
13.3.1 Surprise: scaling, fuzzification, defuzzification	281
13.3.1.1 Sentence	281
13.3.1.2 Syllables	281
13.3.1.3 Fundamental frequency (F_0)	282
13.3.2 Surprise: re-scaling	283
13.3.2.1 Sentence	283
13.3.2.2 Syllables	285
13.3.2.3 Fundamental frequency (F_0)	286
13.4 Re-scaling of the doubt contour - Results	287
13.4.1 Doubt: scaling, fuzzification, defuzzification	287
13.4.1.1 Sentence	287
13.4.1.2 Syllables	287
13.4.1.3 Fundamental frequency (F_0)	288
13.4.2 Doubt: re-scaling	289
13.4.2.1 Sentence	289
13.4.2.2 Syllables	291
13.4.2.3 Fundamental frequency (F_0)	292
Appendix 1 - Elicitation task 1	293
Appendix 2 - Elicitation task 2	298
Bibliography	305
Vita	319

List of Tables

4.1	Pattern recognition system's categorial organization of the canvas relative to their haystack contour	94
7.1	Data for GR72 and scaling. The data for point 10 has been highlighted in gray	132
7.2	Automatic detection of the tones of [JP72] from its pre-tones	150
7.3	The 4-layer structure of [JP72]	155
7.4	The 4-layer structure of [GR65]	156
7.5	Output array of the ATLM for sample [JP72] from the PRInt model	163
7.6	Output array of the ATLM for sample [GR65] from the PRInt model	164
8.1	y : value present in the set n : number of elements bearing value y in the set m : grade of membership of y m_0 : rounded grade of membership of y	174
8.2	Grades of membership by frequency, similarity, and as a mean of both for a few values in the set Y	181
8.3	Unified ranking and organization of the set Y into subsets by grade of membership, from $m_{(y)} = 1$ to $m_{(y)} = 0.1$	182
8.4	Fully expanded fuzzification of the set of y values of pre-tones 15. Grades of membership 1 to 0.1 are ordered left to right. Grades of membership in the subsets are organized top to bottom. The two extreme values are in blackened cells. The defuzzification values are given by subsets and for the set in the last two rows of the table (see Section 8.3.6 p186).	184
9.1	Fuzzification and defuzzification: the scaled time (x) and $F_0(y)$ values are given for each pre-tones (P , 1 to 30 vertically), by grade of membership (1 to 0.1 horizontally), and for the weighted average of all grades (gray column)	192
9.2	Time and F_0 scaled values of the tones by grade of membership (frequency x centrality). These results are for the main pattern, with the main peak on H_3 , and exclude the results of primary peaks realized on H_2	201

9.3	Partial results for the F_0 values of the tones (m_1 and \bar{m}), with P^* as H_2 included and P^* as H_3 excluded. Values in brackets are missing for the grade of membership and they are inferred from the closest grade for which the value is available, indicated in superscript.	202
9.4	Pre-tonal anchoring of the tones ($m_{(x)} = 1$ for frequency x centrality). Each pre-tone corresponds to a fixed position in the syllabic structure.	206
9.5	Pre-tonal co-occurrences by peaks and grades of membership (partial results)	208
9.6	Distance between tones in scaled time and F_0 for the primary and secondary peaks of the question contour	209
9.7	Velocity between tones in scaled F_0 per time for the primary and secondary peak of the question contour, angle (steepness) of the primary peak.	212
10.1	Fuzzification and defuzzification: the scaled time (x) and F_0 (y) values are given for each pre-tones (P, 1 to 30 vertically), by grade of membership (1 to 0.1 horizontally), and for the weighted average of all grades (gray column)	214
10.2	Partial results for the F_0 values of the tones ($m_{(x)} = 1$ and \bar{m}). Top: P^* as H_2 included and P^* as H_3 excluded (patterns a/b). Bottom: P^* as H_2 excluded and P^* as H_3 included (patterns d/e). Values in parentheses are missing for the grade of membership and they are inferred from the closest grade for which the value is available, indicated in superscript.	218
10.3	Time and F_0 values of the tones by grade of membership . . .	219
10.4	Pre-tonal anchoring of the tones ($m_{(x)} = 1$ and \tilde{m} for averaged frequency and centrality). Each pre-tone corresponds to a fixed position in the syllabic structure.	220
10.5	Pre-tonal associations by peaks and grades of membership (partial results)	220
10.6	Distance between tones in scaled time and F_0 for the primary and secondary peak of the surprise contour	221
10.7	Velocity between tones in scaled F_0 per time for the primary and secondary peak of the surprise contour, angle (steepness) of the primary peak.	222
10.8	Fuzzification and defuzzification: the scaled time (x) and F_0 (y) values are given for each pre-tone (P, 1 to 30 vertically), by grade of membership (1 to 0.1 horizontally), and for the weighted average of all grades (gray column)	223

10.9 Partial results for the F_0 values of the tones ($m_{(x)} = 1$ and \bar{m}). Top: P^* as H_2 included and P^* as H_3 excluded (patterns a/b). Bottom: P^* as H_2 excluded and P^* as H_3 included (patterns d/e). Missing values for the grade of membership are inferred from the closest grade for which the value is available, indicated in superscript.	227
10.10 Time and F_0 values of the tones by grade of membership	228
10.11 Pre-tonal anchoring of the tones ($m_{(x)} = 1$ and \tilde{m} for averaged frequency and centrality). Each pre-tone corresponds to a fixed position in the syllabic structure.	229
10.12 Tonal associations by peaks	230
10.13 Distance between tones in scaled time and F_0 for the primary and secondary peak of the doubt contour	231
10.14 Velocity between tones in scaled F_0 per time for the primary and secondary peak of the doubt contour, angle (steepness) of the primary peak.	232
12.1 Evaluation of the degree of typicality of [JP72]	264
12.2 Evaluation of the degree of typicality of [GR65]	264
13.1 Relative duration of sentences by grades of membership	268
13.2 Fuzzification and defuzzification of the duration of syllables (all durations expressed in %). Grade 0.1 serves as an example of the process (see table 13.3 in this section)	270
13.3 Adjustment of fuzzy duration to 100 (all durations expressed as a percentage of the duration of the sentence)	270
13.4 Grades of membership of scaled F_0 values by subject groups (male, female, global) and subcategories (max, min, ran)	276
13.5 Re-scaling, by grade of membership, of the relative duration of syllables to duration in cs, proportionally to the duration of the sentence	279
13.6 F_0 extreme values (top) and re-scaling of the F_0 values for grade $m_{(x)} = 1$	281
13.7 Grades of membership of actual F_0 values (Hz) by subject groups and dimension categories.	281
13.8 Relative differential range of sentence duration by grade of membership	282
13.9 Fuzzification and defuzzification of the duration of syllables (all durations expressed in %)	282

13.10 Adjustment of fuzzy duration to 100 (all durations expressed as a percentage of the duration of the sentence)	283
13.11 Grades of membership of scaled F_0 values by subject groups (male, female, global) and subcategories (max, min, ran)	284
13.12 re-scaling, by grade of membership, of the relative duration of syllables to duration in cs, proportionally to the duration of the sentence	285
13.13 F_0 extreme values (top) and re-scaling of the F_0 values for grade $m_{(x)} = 1$	286
13.14 Grades of membership of actual F_0 values (Hz) by subject groups and dimension categories.	286
13.15 Relative differential range of sentence duration by grade of membership	287
13.16 Fuzzification and defuzzification of the duration of syllables (all durations expressed in %)	287
13.17 Adjustment of fuzzy duration to 100 (all durations expressed as a percentage of the duration of the sentence)	288
13.18 Grades of membership of scaled F_0 values by subject groups (male, female, global) and subcategories (max, min, ran)	289
13.19 Conversion, by grade of membership, of the relative duration of syllables to duration in cs, proportionally to the duration of the sentence	291
13.20 F_0 extreme values (top) and conversion of the F_0 values for grade $m_{(x)} = 1$	292
13.21 Grades of membership of actual F_0 values (Hz) by subject groups and dimension categories.	292

List of Figures

1.1	Two simplified intonation patterns	9
2.1	<i>La trahison des images</i> , 1929, René Magritte, Art Institute of Chicago	13
2.2	<i>Bocal de poissons rouges</i> and <i>Poissons rouges et palette</i> (1914) by Henri Matisse	15
2.3	Two objects to which the term “chair” is applied. “Chair” is vague because there exist cases for which its application is uncertain and/or not typical	20
2.4	Top: Voice onset time distribution: labial stops of two-category languages. From Lisker & Abramson (1964). Bottom: Identification curves as functions of VOT values. From Lisker & Abramson (1970).	30
2.5	Discrimination accuracy between English labials. From Abramson & Lisker (1970)	32
2.6	A network of allophones of /t/, adapted from Taylor (2004) . .	35
2.7	Frequency of second formant versus frequency of first formant for ten vowels by 76 speakers, Peterson & Barney (1952) . .	38
2.8	The prototype /i/ vowel (P) and variants on four orbits surrounding it (open circles) and the non-prototype /i/ vowel (NP) and variants on four orbits surrounding it (closed circles). The stimuli on one vector were common to both sets. From Khul (1991)	40
2.9	Category goodness (typicality) ratings for the prototype /i/ vowel, the non prototype /i/ vowel, and the variants surrounding each of the two vowels. From Khul (1991)	41
3.1	The intonation contours of a declarative (left) and of a closed question (right) over the same sentence <i>Maman va venir</i> . Time and F_0 have been normalized, syllable boundaries are marked with dotted lines.	46
3.2	A hand-drawn <i>ton</i> (intonation contour) by Passy (1887). <i>Et tu préfèrerais avoir cette lettre</i> (“And you would prefer to have this letter”)	46

3.3	Four closed questions produced by a single speaker of French and displayed with the software Praat (Boersma & Weenink, 2012). Syllables have been numbered and separated by dotted lines	51
3.4	Hierarchical intonation structure. From Gussenhoven (2004)	55
3.5	Two instances of the phonological pattern L*+HH% (closed questions) including the word <i>Invalid</i> in sentence final position. Time and F ₀ have been normalized, syllable boundaries are marked with dotted lines.	58
3.6	[LHiLH*] accentual phrases (AC) and <i>arc accentuels</i> . Adapted from Jun & Fougeron (2002)	64
3.7	Four intonation contours applied on the same sentence <i>Maman va venir</i> . Time and F ₀ have been normalized, syllable boundaries are marked with dotted lines.	69
3.8	Organization of the study of intonation with PRInt	74
4.1	A subset of the “seat” category: 0) chair, 1) meditation pillow, 2) Sori Yanagi Butterfly stool, 3) Eames chair, 4) Jean Prouvé armchair, 5) Chippendale armchair	77
4.2	Musical variations or quotations are a creative use of the human pattern recognition ability. Top : Wagner’s opening theme to <i>Tristan und Isolde</i> . Bottom : Debussy’s quotation/variation on the theme in his <i>Children’s Corner</i>	80
4.3	Identification/categorization of objects: distortion, frequency, and similarity	81
4.4	Frequency of second formant versus frequency of first formant for ten vowels by 76 speakers, Peterson & Barney (1952) . . .	82
4.5	Using the dictation function of a modern computer.	84
4.6	12 instances of <i>Les meules</i> by Claude Monet arranged sequentially, according to the number of stacks and their relative distance	89
4.7	Feature extraction. Fig.4.7a : normalization, scalar quantization, and discretization of canvas (2) from Figure 4.6. Fig.4.7b : Numerical and structural feature extraction. The feature vector of the large stack is located and extracted first. The feature vector of the small stack is located relative to that of the large stack.	92
5.1	Binary sets of men’s height	99
5.2	Fuzzy functions of the set of short men (dotted line) and of the set of tall men (dashed line), based on the current world size records established by the Guinness World Records	100

5.3	“Figure 5.3 plots the percentage of “buy” judgments as a function of VOT for a representative young normal-hearing participant. Actual data points are shown (open circles) as well as the PSIGNIFIT fitted function for these data and the derived crossover point (filled square) and endpoints (filled diamonds). The figure shows that this listener categorized the shortest VOT (stimulus 1) as “buy” and the longest VOT (stimulus 7) as pie. The performance of this listener on this stimulus pair was typical of the performance of all of the listeners on most of the stimulus pairs, indicating that participants heard the endpoint stimuli as intended. Typically, the crossover point for all listeners occurred in the region defined by stimuli 3, 4, and 5. (Gordon-Salant et al., 2008)”	105
5.4	Fuzzy functions of the set B of allophones of /b/ (dotted line) and of the set P of allophones of /p/ (dashed line). The grade of membership of a phoneme in either one or the other set is a function of the duration of its VOT.	107
6.1	A distorted instance of a closed question contour: normalized (left) and analyzed by PRInt (right).	113
6.2	Example of a slide used for the task of closed questions. “Some people like to keep souvenirs after a concert. -Are you going to keep the tickets?	114
6.3	Example of a slide used for the task of modality contrast. “Your friend grew up in a farm where sheep and lambs were bred. However, he tells you he never ate lamb. -You never ate lamb?! a) In your question, express your surprise; b) In your question, express your doubt	115
6.4	Implementation of PRInt in Excel	117
7.1	Two simplified intonation contours (1)	122
7.2	Two simplified intonation contours (2)	123
7.3	Two simplified intonation contours (3)	124
7.4	Two actual contours of closed questions in French (1)	128
7.5	Two actual contours of closed questions in French (2)	129
7.6	Two scaled contours of closed questions in French	134
7.7	[JP72] represented as a 100 points by 100 matrix of points. The activated points of the contour are in black.	135
7.8	[GR65] represented as a 100 points by 100 matrix of points. The activated points of the contour are in black.	136
7.9	Syllables and half-syllable frames adjustment of [JP72]	141

7.10	The isometric syllabic grid	143
7.11	Isometric pre-tonal representation of [JP72] (left) and [GR65] (right).	144
7.12	[JP72] represented as a 100 points by 100 matrix of points. The activated points in black are the pre-tones of the contour.	146
7.13	[GR65] represented as a 100 points by 100 matrix of points. The activated points in black are the pre-tones of the contour.	147
7.14	[JP72] (right) and [GR65] (left) on the tonal isometric grid. Both sentences have the tonal structure L#L-H-(L L)-H*H% .	157
7.15	[JP72] represented as a 100 x 100 point matrix. Each layer is a subset of the layer(s) above: {tones} ⊂ {pre-tones} ⊂ {points}	160
7.16	[GR65] represented as a 100 x 100 point matrix. Each layer is a subset of the layer(s) above: {tones} ⊂ {pre-tones} ⊂ {points}	161
7.17	[JP72]: Output of the MOMEL algorithm	162
7.18	[GR65]: Output of the MOMEL algorithm	162
8.1	Comparison of [JP72] and [GR65] on the isometric grid. Pre-tone 15 has been circled on both contours: “the variation in shape of the contour is ultimately and solely conditioned by the variations in F_0 of each pre-tone”	167
8.2	Distribution of the F_0 values of pre-tone 15 in the set Y	172
8.3	Distribution of the f_0 values of pre-tone 15 organized by grades of membership. Each column corresponds to a grade of membership. The height of each column corresponds to the number of values in the subset of the grade of membership. There are 96 values in the set.	175
8.4	Fuzzification of Y according to the similarity principle. 28 is the center \bar{y} of the set. Each circle represents a grade of membership: the innermost circle contains the values that are the closest to the center and whose grade of membership is $m_{(y)} = 1$, the outermost circle contains the values that are the farthest from the center and whose grade of membership is $m_{(y)} = 0.1$	179
8.5	Membership function of subset grade $m_{(y)} = 0.4$. Apart from $y = 3$, the function is symmetrical. The outlier will be penalized in the weighted average defuzzification.	187
8.6	Membership function of the set of F_0 values of pre-tone 15 (set Y). Each point is the crisp value resulting from the defuzzification of a subset (grades of membership 1 to 0.1). The function is almost continuously linear, except for grade $m_{(y)} = 0.8$ (“corrected” as a dotted line)	188

9.1	Closed question contour by grade of membership: frequency.	193
9.2	Closed question contour by grade of membership: centrality.	194
9.3	Closed question contour by grade of membership: frequency and centrality combined.	195
9.4	Closed question pretonal contour: defuzzification	195
9.5	Four samples of pattern variations from the corpus of subject GR: on the left, patterns d (top) and variation e (bottom); on the right, patterns a (top) and variation b (bottom).	204
9.6	Prototypical contour. Left: main contour ($P^*=H_3$). Right: variation of the contour ($P^*=H_2$)	211
10.1	Surprise modality contour by grade of membership: frequency.	215
10.2	Surprise modality contour by grade of membership: centrality.	215
10.3	Surprise modality contour by grade of membership: frequency and centrality combined.	216
10.4	Surprise modality pretonal contour: defuzzification	216
10.5	Prototypical contour. Left: main contour ($P^*=H_3$). Right: contour variation ($P^*=H_2$)	222
10.6	Doubt modality contour by grade of membership: frequency.	224
10.7	Doubt modality contour by grade of membership: centrality.	225
10.8	Doubt modality contour by grade of membership: frequency and centrality combined.	225
10.9	Doubt modality pretonal contour: defuzzification	226
10.10	Prototypical contour. Left: main contour ($P^*=H_3$). Right: variation of the contour ($P^*=H_2$)	232
11.1	The three prototypical contours on the isometric grid.	235
11.2	Scaled crisp (\bar{m}) contours of closed question, surprise, and doubt.	237
11.3	Zone of ideal phonetic implementation (frame of reference) of the prototypical contours (\bar{m}) of closed questions, surprise, and doubt. The prototypical contours have been scaled towards the range of their own frame.	242
11.4	Phonetic implementation of the three contours at their highest grade of typicality ($m_{(x)} = 1$)	244
11.5	Phonetic implementation of the three contours at their medium (vague) grade of typicality $m_{(x)} = 0.5$	248

11.6	Prototypical contours of question and surprise on the isometric grid, secondary peak included	250
11.7	Prototypical contours of question and surprise on the isometric grid, primary non-final peak	252
11.8	Phonetic implementation of question and surprise with their secondary peak (highest grade of typicality ($m_{(x)} = 1$))	253
12.1	A characteristic intonation pattern from a variety of Norman: a declarative ending with a L-H*-L% contour. Pre-tonal representation from the PRInt model, at the highest degree of typicality	262

Chapter 1

Introduction

A case for intonational phonology

Intonation is a half-tamed savage. To understand the tamed or linguistically harnessed half of him one has to make friend with the wild half Dwight (Bolinger, 1978).

This dissertation employs the related principles of categorization, pattern recognition and fuzzy set theory to examine the phonological and phonetic nature of intonation. Compared to the highly systemic nature of segmental phonology, intonational phonology has a somewhat more elusive status. Intonation is at the edge of the linguistic domain because it appears to have at the same time a linguistically discrete structure and a wide range of continuous variation. The goal of this dissertation is to support the argument in favor of an intonational phonology by experimentally separating discrete structure and continuous implementation of the structure.

The units of segmental phonology are conceived as discrete: they can be defined and opposed paradigmatically in terms of articulatory features. Such is not so obviously the case for intonation which can be regarded as “the integral melodic pattern” over a sentence (Bolinger, 1978). An intonation pattern is a complex object made of a physical (acoustic) shape that develops over time: the physical primitives of intonation are fundamental frequency, duration, and intensity (or in perceptual terms: pitch, length, and loudness).

For example, in many languages, the intonation pattern associated with a closed question consists of a F_0 signal that remains more or less consistently flat for a while and then goes up abruptly before suddenly stopping at the top of its height, also corresponding to the end of the sentence over which the pattern is taking place. Given the nature of such objects, and the fact that they are articulatorily produced by many different human beings, the distortion they can potentially undergo is infinite. These dimensions are continuous and consequently a same intonation pattern can physically be “stretched” to accommodate many different sentences. For example, a sentence comprising three syllables and another one comprising seven syllables can both receive the same pattern of intonation leading to the same meaning.

However, the distortion of a pattern can only go so far. For an object to be recognized and meaningfully interpreted, no matter what its nature be, it has to be categorizable. Indeed, in spite of the potentially unlimited range of possible variation, speakers of a language systematically realize intonation patterns in a way that can be interpretable by other speakers of the language. It is important to distinguish between two (concomitant) usages of intonation. One usage of intonation is linguistic, relying on discrete units, to convey a meaning to the sentence over which it is applied. For example, intonation distinguishes between questions and statements in a lot of languages. Speakers systematically produce the same pattern to encode the same meaning. The other usage is para-linguistic, relying on the continuously gradient physical primitives to convey information about the speaker and/or the situation. This untamed part reflects some idiosyncratic and socio-cultural habits. For example, intonation can convey anger, joy, surprise, etc.

A general approach to a systemic description of intonation has been to

abstract away from gradient variations in order to characterize phonetically different sentences as phonologically similar (or different). Basically, the idea is to discretize a continuum into a structurally organized sequence of elements that are part of a limited inventory. One way to achieve this is by segmentation of an otherwise continuous acoustic signal into constitutive morphological chunks or movements: plateau, rise, fall, rise-fall, etc. so that a whole pattern can be expressed as a concatenation of intonational morphemes. Such is, for example, the approach of the British school of intonation. Another way to do it is to only look at the salient points of an acoustic signal, its relatively high and low points instead of movements. This is the approach of various theories and among them is the widely recognized (although not uncontroversial) Autosegmental-Metrical model (AM), devised by Janet Pierrehumbert (1980) in her dissertation. In this model, there are only two levels of fundamental frequency: (F_0) high and low. A point is a high tone (H) or a low tone (L) relatively to the position of the preceding tone(s) or complex of tones in the sequence of the intonation pattern (the AM inventory is more elaborate and is not restricted to the H and L tones). Within this model, two phonetically different sentences can be analyzed as phonologically identical, sharing the same tonal pattern because gradient physical variations have been abstracted away.

It is a fundamental idea of a two-tone model such as the AM that a phonological structure (the discrete pattern) can be abstracted from the observation of its phonetic variations (the gradient implementations of the pattern):

[The approach taken] towards the problem of establishing which intonation patterns are linguistically distinct and which count as variants of the same pattern [...] attempt[s] to deduce a system of

phonological representation from observed features of F_0 (Pierrehumbert, 1980).

However, the AM model has been criticized for its lack of specificity when it comes to explaining how the abstract pattern extracted from the phonetic variations are in turn phonetically implemented on sentences. In other words, once the signal has been encoded into a string of relative tones, some information is lost and the signal cannot be recreated. This has been regarded as a loophole in the model and as an argument in favor of directly mapping communicative functions conveyed by intonation onto articulatory processes, without going through a phonological level of representation (Xu, 2005)

1.1 Categorization

In this dissertation it will be argued that without a phonological level of representation, intonation could not function linguistically but would be instead constrained to para-linguistic, purely phonetic variations. The basis for this argument is the general principle of categorization that is pervasive in human behavior. Categorization is the process by which we recognize that different entities belong to a same class in spite of objective physical differences. All entities in a category, or class, are given the same generic name. For example, the word “bird” covers a wide range of animals varying in many aspects but all displaying a certain degree of “bird-ness”. As discussed in Chapter 2, what keeps the members of the category together is a loose and complex network of similarities, close and distant: a “family resemblance” (Wittgenstein, 1953) which, for example, groups a sparrow and an ostrich in the same category of “birds”. Of course, the application of a generic name

to an entity depends on a subject's own concept under which a category is formed (the intension of the category name) and, in many occasions, one is not sure of the limit of application of the concept (the extension of the category name). Some objects are borderline in the sense that more than one category name could apply to them. This uncertainty in the categorization of borderline objects is referred to as the *vagueness* of the category name. The existence of vagueness and borderline objects implies that among objects of a same category, some are more typical than others and that objects can be ranked for their degree of typicality (Rosch, 1978). Sparrows are somewhat more typical than ostriches in the category of birds. Finally, the prototype of a category is a construct that abstractly represents the intension of a concept. It is, in a way, the ideal object of a category, the center of gravity that subsumes the whole class.

Without categorization, each object in the world would be a singularity. Objectively, without a construct linking them, ostriches and sparrows could not be considered as instances of a single category, no more than an elephant and a rat, or two phonetically different intonation patterns. It is indeed an assumption of the present dissertation that phonetically different intonation patterns can be considered as instances pertaining to categories subsumed by a phonological structure and that they have a degree of typicality within these categories.

1.2 Pattern Recognition

Overlapping the concept of categorization is that of pattern recognition. In order to categorize an object as part of a category or another, this object must first be analyzed as a complex of structurally ordered features:

a pattern or feature vector. This pattern is subsequently matched towards prototypes (or stored examples, depending on the theoretical approach) of pre-existing categories in order to determine its grade of membership in one or more categories, depending on its degree of typicality in each of them. Pattern recognition is the human ability to abstract away from variations in order to extract relevant features for subsequent categorization.

This faculty that is so pervasive in human behavior (and vital for mushroom pickers) that has been artificially emulated to automate numerous processes such as optical character recognition of handwritten postal codes on postal envelopes or logistics and sales sorting using barcodes. Pattern Recognition (PR) as a technological domain comprises a wide range of scientific research and most specifically a large panel of engineering applications. PR applications are commonly ad-hoc systems programmed to discretize the continuous space of the objects they have to classify (or categorize) into smaller parts so that they can extract salient features among these parts. These features are analyzed into a structurally organized pattern that can be matched towards templates pre-entered in the system for the purpose of categorization of new objects. The classifier can also be created to form its own templates from the analysis of all the objects it has been given to analyze (see Chapter 4).

1.3 Intonational phonology as pattern recognition

A phonological approach to intonation that expresses intonation patterns as a sequentially structured set of relatively defined tones is somehow akin to a pattern recognition process applied for the purpose of classifying objects into categories.

When assigning a tone value (L or H) to a point of an intonation pattern, how can a linguist be sure that it is a phonologically distinct tone and not just a phonetic variation of an underlying phonological tone? A model using mainly two categories (H and L) to sequentially label elements of a pattern relies on the fact that high and low F_0 points exist by relative contrast to their direct preceding environment. They cannot be strictly considered as absolute or discrete units in a paradigm. However, high and low points occur or do not occur, they exist as a binary contrast, but their presence or absence depends on their relative degree of realization. Thus, there are going to be borderline cases in which one is not sure if a point is either H or L. There is a need to separate phonological structure from phonetic implementation (Ladd, 2008).

Yet, how is it possible to say something of an underlying phonological structure when all linguists have to see is its superficial phonetic form? It has been noted that if the same sentence is given to a group of trained linguists, there most likely will be some discrepancies among their analysis or categorization, a) regarding what points are the salient points of the contour or tones, b) regarding the categorial status of the tones or complex of tones (Wightman, 2002). Unlike what has been argued by the critics of the AM model (or similar tonal approaches), these discrepancies do not invalidate the existence of a phonological level of description for intonation. Transcribers assign tones according to their own native and scientific knowledge of the language. In other words, transcribers are not “deduc[ing] a system of phonological representation from observed features of F_0 ”, they analyze the instance of an intonation pattern they have in front of them, or part of it, in terms of degree of typicality within the categories they already have formed in their (knowledge of the) linguistic system.

1.4 Research question and hypotheses

The question remains unanswered, and it is the general research question for this dissertation: if intonation patterns have a phonological structure, can this structure actually be abstracted from the observation of its phonetic implementations, as suggested by Pierrehumbert (1980); and furthermore, can it be achieved without prior (meta)linguistic knowledge?

To answer this question, an experimental setting was created to test the following hypotheses:

1. It is possible to abstract a phonological structure from the systematic analysis of a corpus of instances of a given intonation pattern
 2. It is possible to distinguish between phonologically distinct patterns and phonetic variation of the same pattern (allocontours)
- 1+2 In other words, it is not only possible to abstract a phonological structure from the systematic analysis of its phonetic variation, but it is also possible to analyze separately both its phonological structure and its phonetic implementations from the same analysis.

1.5 The PRInt model - Pattern Recognition for Intonation

To test the hypothesis, a computational model has been created to analyze intonation patterns, using duration and fundamental frequency as the physical primitives: a system of Pattern Recognition for Intonation (PRInt). The language used in the present application of the model is French. This is a non trivial choice since French, among other specificities, does not have lexical stress. It makes it an logical choice for the study of intonation patterns

since only syllabic parsing has to be provided to the system by the operator. Three intonation patterns, or contours, have been selected, based on their description provided by Fónagy & Bérard (1973) and Beyssade et al. (2007): closed questions, and two modalities of closed question, doubt and surprise.

Globally, the model analyzes as a tonal sequence a set of instances of a given intonation pattern and then compares all these sequences to come up with the phonological structure that subsumes all instances. Or in more direct way: **the PRInt model extracts the phonological structure of an intonation pattern from the ranking of its phonetic implementations by degree of typicality** (the contrary of a linguist labeling one instance). The Print model is a computerized system consisting of a pattern recognition module coupled to a classifier.

The task of the PR module is to label instances of an intonation contour, which it does by using a protocol called the 4-layer structure. The instances are converted into a feature vector by sequential discretization. Ultimately all instances are analyzed as a string of structurally organized tones, whose position in the structure is solely expressed in relative terms, as L and H tones.

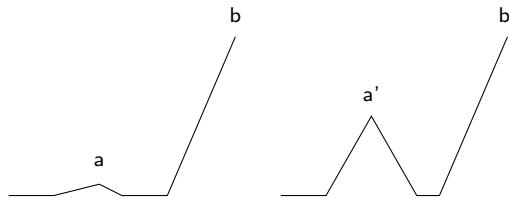


Figure 1.1: Two simplified intonation patterns

However, the main difference between a linguist transcriber labeling a set of sentence and a mathematical model such as PRInt lies in the mechanistic

and blind systematicity of the model. Let the graphs in figure 1.1 be two simplified intonation patterns. A linguist would probably identify the pattern on the left as LLH. In contrast, she would identify the pattern on the right as LHLH. She would probably assume, and rightfully from her prior knowledge of intonation, that the small difference in F_0 displayed by the peak marked (a) is merely a phonetic variation whereas a peak such as (a'), and even more so (b) and (b') are more likely to be linguistically distinctive or part of the phonological structure. The PRInt system would instead label both contours as LHLH, while noting the numerical difference between (a) and (a') but also between (a) and (b) and (a') and (b'). In a systematic approach to pattern recognition, any numerical difference constitutes a difference of tones. The system does not presume that a variation, as small or large as it may be, is *a priori* phonologically or linguistically distinctive or not.

Once the PRInt system has labelled all the instances of a pattern in a corpus, the classifier ranks the values recorded for all features by degree of typicality in the corpus. The classifier broadly relies on the principles of fuzzy logic to assign a grade of membership to values inside of multisets. To do so, it employs two functions: one based on frequency, the other based on similarity to a central tendency. In the output, the most typical values for a given feature are the ones that are both the most frequent and the most central. With all the values of all features organized by degree of typicality (or mathematically by grade of membership), the PRInt model can generate (or calculate) the prototypical pattern from the category, that is, the phonological structure of the pattern, extracted from all its instances and that subsumes them all.

Prior to pattern recognition, the PRInt model normalizes all instances to abstract away from inherent differences in the range of fundamental fre-

quency and of duration. However, the actual dimension of these variations are recorded as well and classified by degree of typicality too. Thus, if the model finds that the phonological structure of two intonation patterns is identical, it can also determine if they are *allocontours* of the same intonation pattern only distinguished by phonetic variation created by the para-linguistic usage of intonation.

1.6 Organization of the dissertation

General background: Chapter 2 presents the concepts of categorization, vagueness and prototypicality in a general manner and then provides two examples of categorization applied to segmental phonology.

Chapter 3 describes intonation and more specifically discusses two main types of approaches to intonation (analytic and synthetic) in regards with the more general principles of categorization. It is argued in favor of intonational phonology, the analytic approach. In the rest of the chapter, specificities of French intonation are presented and the PRInt model is introduced.

Chapter 4 is a general description of the techniques of pattern recognition. This Chapter provides the general background for Chapter 7. An application of PR to a set of paintings exemplifies the application to intonation of Chapter 7.

Chapter 5 is a general description of the principles of fuzzy set theory, with an illustrative application to phonology. This chapter provides the general background for Chapter 8.

The PRInt model: Chapter 6 explains how the data were obtained and prepared for the application with the PRInt model.

Chapter 7 describes the pattern recognition module of the PRInt model: the Automated Tonal Labeling Module (ATLM).

Chapter 8 describes the classifier of the PRInt system: the Automated Fuzzy Classifier (AFC).

Application of the PRInt model: Chapter 9 details how the model analyzes a set of instances for a given intonation pattern and extracts its phonological structure. The intonation pattern is that of closed questions in French.

Chapter 10 presents the results of the same analysis but for the intonation patterns of surprise and doubt.

Chapter 11 is both an analysis and a discussion of the results provided by the PRInt model. It compares the three intonation patterns and determines the phonological and/or phonetic status of each of them.

Chapter 12 contains a summary of the PRInt model and the results of its application to three intonation patterns. It also suggests further development (such as an evaluation module), possible improvements, and further research and applications.

Chapter 13 details how the PRInt model analyzes the phonetic variation of the intonation patterns. The results are used in the discussion of Chapter 10. However, the details of the process have been moved as an appendix to simplify the presentation of the overall results in Chapter 10.

Chapter 2

Categorization



Figure 2.1: *La trahison des images*, 1929, René Magritte, Art Institute of Chicago

A category is a labelled grouping of related objects or concepts (Savas, 1990). Categorization is thus the process of grouping objects under a single label, or category name, that can be shared among speakers of a language. In the classical view, as initiated by Aristotle, categories are specified by a small set of necessary and sufficient properties. If an object has these properties, it is part of the category; otherwise, it is not, and this rule of inclusion in the set is binary. Categorization is chiefly the development of a definition; the necessary and sufficient characteristic, or set of characteristics shared by all the objects in a category. For example, all triangles can be defined by the single characteristic of having three vertices and thus all forms with three

vertices belong to the category of triangles.

An objection to the long-lasting classical view concerns the fact that a strict inclusion rule does not account for most natural categories. This objection has gained some momentum only in the past century starting with the work of Frege in logic, developed philosophically by Wittgenstein (1953), and, finally, experimentally grounded by Rosch & Mervis (1975) in psychology. The classical view is binary. It cannot account for the fact that speakers of a language often disagree on both definition of a term and its application to objects. The concept of vagueness explains this confusion which derives from the human use of symbols, linguistic or otherwise, to communicate experience and knowledge.

2.1 The vagueness of symbols

In 1929, René Magritte created one of his most famous paintings: *La trahison des images*. In this work, a clearly legible caption appears beneath the realistic image of a pipe: *Ceci n'est pas une pipe*, “this is not a pipe.” As Magritte explained: “could you stuff my pipe? No, it's just a representation, is it not? So if I had written on my picture ‘This is a pipe,’ I would have been lying!” (Torczyner, 1977). Alfred Korzybski (1941) pushed this reasoning further when he famously said that “the map is not the territory.” No matter how detailed and sophisticated the medium gets, there will always be a gap between the model and its representation by a symbol. Ultimately, the accuracy of a representation may be subsumed to the degree of perceived realism between the symbolic representation and what it represents. In that sense, the gap in a high definition digital photograph is smaller, making it far more accurate than a picture drawn by a five-year-old.

The existence of a gap between symbols and their referent (vagueness) was discussed 2,400 years ago in Plato's *Res Publica*:

“Those who study geometry and calculation [...] use the visible squares and figures, and make their arguments about them, though they are not thinking about them, but about those things of which the visible are images. Their arguments concern the real square and a real diagonal, not the diagonal which they draw, and so with everything. The actual things which they model and draw [...] they now use as images in their turn, seeking to see those very realities which cannot be seen except by the understanding.” Plato
(translation by A.D. Lindsay, 1976)



Figure 2.2: *Bocal de poissons rouges* and *Poissons rouges et palette* (1914) by Henri Matisse

Matisse, in two paintings he created in 1914 (Figure 2.2), represented the

same object at two moments separated in time by a few weeks, one before the war broke out (left), the other after (right). The same goldfish bowl stands in the middle of the workshop but it is shown from a different perspective not only spatially, but also formally in its treatment. A single referent has been expressed in two different ways using two different pictorial combinations. Matisse's state of mind during the creation of these two paintings was undoubtedly very different, and, one reality is represented differently to express two different moods. It is in fact impossible to grasp their commonality without relying on an ideal concept (the goldfish in their bowl, in Matisse's workshop), abstracted from the two paintings and of which they are both instances. In other words, it is impossible to grasp the relationship of the two objects without a strict definition that unites them.

In a symbolic system, and particularly in human language, different symbols can be used to represent the same reality, and one symbol can be used to represent different realities. This pervasive and well-noted phenomenon generates vagueness.

2.2 Vagueness: typicality and prototype

Vagueness has primarily been studied in the fields of logic, philosophy, and semantics. Therefore, the concept will be discussed with well-known examples from semantics before it is applied to the more abstract units of phonology.

2.2.1 Vagueness

In a talk given in the early 1920s, Bertrand Russell discusses the concept of vagueness, and “propose[s] to prove that all language is vague” (Russell,

1923). Russell's argument concerns symbolic systems of representation and more specifically with natural human language. His initial contention is that the units of language, especially words, are not in a one-to-one relation with what they refer to. There can be many symbols to designate one object (many-to-one) and one symbol to designate many objects (one-to-many).

many-to-one In a work Russell cites as one of the sources of his reasoning, Gottlob Frege gives the example of the planet Venus, which has been called “the morning star” or the “evening star” depending on the time of its apparition in the sky. Frege distinguishes between the *reference*, the thing to which a sign refers (Venus) and the *sense*, the way this thing is presented symbolically (“morning star” or “evening star”). “François Hollande” and the “current president of France” are two senses of the same reference.

Frege’s distinction between sense and reference corresponds to the distinction between **intension** and **extension**, two related concepts found in logic, philosophy, and semantics. A symbol’s set of features or properties is its intension. The classical view of categorization, states that the dictionary definition of a term corresponds to its intension, being an attempt to delimit the set of sufficient and necessary features shared by all objects to which the term is applied. In the Merriam-Webster dictionary, a chair is defined as “a seat typically having four legs and a back for one person.” Notice the use of “typically,” since a category name such as “chair” may be applied to objects whose features may differ from those of the “typical” chair. The extension of a term is the set of objects which possess these features and can be referred to by the term.

one-to-many In a later work citing Russell's talk, Black gives the example of the extension with the word "chair" and how it undermines the definiteness of its intension in the classical sense:

"... think of arm chairs and reading chairs and dining-room chairs, and kitchen chairs, chairs that pass into benches, chairs that cross the boundary and become settees, dentist's chairs, thrones, opera stalls, seats of all sorts, those miraculous fungoid growths that encumber the floor of the arts and crafts exhibitions, and you will perceive what a lax bundle in fact is this simple straightforward term. In co-operation with an intelligent joiner I would undertake to defeat any definition of chair or chairishness that you gave me" (Wells, 1945).

It follows that vagueness is not to be equated with generality (Bolinger, 1961; Black, 1937), the fact that one term is a category name and can refer to a class of generic objects. The word "chair" is not vague because it can refer to both the kitchen chair present in the physical space where a conversation takes place and a chair mentioned in the conversation that is distant from the other in space, time, and shape. In such a situation, the chair in "Could you pass me this chair?" and the chair in "Do you remember grandpa's favorite chair?" would never be confused since language specifies one as the *hic et nunc* chair (*this chair*) and the other one as a referred object from some family's history.

Vagueness is not ambiguity, either (Bolinger, 1961; Black, 1937). Ambiguity resides in the polysemy of a term. In French, the word "bureau" as in "*Jean a un grand bureau*" is ambiguous because it can mean "Jean has a large office" or "Jean has a large desk." The word "chair" is vague because

it has multiple meanings. Ambiguity requires resolution from a context or situation for comprehension, vagueness does not. “Bureau” must be clarified for the intended meaning to be retrieved, but the case of “chair” can be left unspecified because more precision is not needed for understanding and could possibly be cumbersome.

As noted by Black, vagueness applies especially to terms or propositions that refer to the physical world, or natural categories, “all whose application requires the recognition of sensible qualities” (Black, 1937). Thus, vagueness comes from the existence of objects for which the application of a term is uncertain. There exist “borderline cases or doubtful objects [...] to which we are unable to say whether the class name does or does not apply” (Black, 1937). There exist some objects about which we are unable to say whether the word “chair” applies or not. Vagueness is a characteristic of symbols – of language – not of the objects they represent. There is no such thing as a vague object, a vague chair. Vagueness lies in the decision of the speaker to apply the term to an object or not. The two objects below (Figure 2.3) should belong to the category of “chair” since both have been designed as such by their creators. However, the one on the left (a), a kitchen chair made for Ikea, would cause no uncertainty about the application of the word “chair” to it, whereas the one on the right (b), a Loopita model by Victor Aleman, constitutes a borderline case to which the application of “chair” is far from typical. The case of these two objects reveals how both the intension/definition of a term and its extension/application are vague. If a sign is intentionally vague, its set of characteristic features is uncertain in terms of number or specificity (its definition is loose). If the extension of a sign is vague, there are borderline cases of its application (its application is loose). The two type



Figure 2.3: Two objects to which the term “chair” is applied. “Chair” is vague because there exist cases for which its application is uncertain and/or not typical

of vagueness influence each other, loose definition leading to loose application and vice versa. Therefore, there are various types of vagueness and the degree of vagueness of a term varies idiosyncratically and contextually.

2.2.2 Category structure: family resemblance, typicality, and prototype

Ludwig Wittgenstein developed the concept of vagueness further through what he called *Familienähnlichkeit* (“family resemblance”) in his *Philosophical Investigations* (Wittgenstein, 1953). In his view, objects in a category do not necessarily share common features but are linked together by a “complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of details” (Wittgenstein, 1953), as, for ex-

ample, the category of *game* that Wittgenstein discussed at length¹. Objects are in a category because of many types of relationships such as associative chains, one object resembling another, itself resembling another, and so, on until the link between the first and the last object of the chain is nothing but the sequence itself:

One can imagine an exhibition in some unlikely museum of applied logic of a series of ‘chairs’ differing in quality by least noticeable amounts. At one end of a long line, containing perhaps thousands of exhibits, might be a Chippendale chair: at the other, a small non descript lump of wood (Black, 1937).

¹§66 – Consider for example the proceedings that we call “games.” I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? –Don’t say: “There must be something common, or they would not be called ‘games’ ”but look and see whether there is anything common to all. –For if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat: don’t think, but look! –Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost. –Are they all ‘amusing’? Compare chess with noughts and crosses. Or is there always winning and losing, or competition between players? Think of patience. In ball games there is winning and losing; but when a child throws his ball at the wall and catches it again, this feature has disappeared. Look at the parts played by skill and luck; and at the difference between skill in chess and skill in tennis. Think now of games like ring-a-ring-a-roses; here is the element of amusement, but how many other characteristic features have disappeared! And we can go through the many, many other groups of games in the same way; can see how similarities crop up and disappear.

And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail.

§67 – I can think of no better expression to characterize these similarities than “family resemblance”; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. –And I shall say: ‘games’ form a family. (Wittgenstein, 1953)

In the series of chairs, the similarity between the Chippendale chair and the lump of wood is found in the perceived contextual sequence of similarity, not in any direct similarity between the two objects. Encountered separately outside of the exhibition, they could not belong to the same category. In Wittgenstein's work, categories are non-discrete and context-dependent constructs. The category of "chair" changes between individuals and also during the course of an individual's existence, depending on experience. Thus, not only is a category a loose network of objects, but it is also not finite: it has fuzzy edges and objects can be added to it and removed from it over time. Depending on the category, some features may be common to all objects, some may not, and the aspect of these features may change from object to object.

The idea of family resemblance, as exemplified by the concepts of "game" or "chair", entails that a category created by the application of a term to a set of objects is structured in terms of typicality, that is, in terms of centrality and periphery. The applications of the word "chair" to the two objects in Figure 2.3 and to the Chippendale model and the lump of wood are not equivalent. The kitchen chair naturally seems to be a closer match for the word than the designer chair, or more precisely, the kitchen chair is a more typical or central instance of the category formed by the word "chair" than the designer chair, which is more peripheral. Although speakers apply the same category name to all the objects in a category, these objects are not equal in terms of representativeness. Some have a higher degree of typicality than others.

The idea of the graded typicality of objects in a category led to the theory of *prototype*, started by Rosch & Mervis (1975) (see also Rosch, 1973, 1975a,b, 1976, 1973). Prototype theory is feature-based. In line with Wittgen-

stein's family resemblance, the theory posits that subjects categorize objects in a network of relationships between their constitutive features, without any feature(s) being sufficient or necessary. Rosch (1975b) asked a group of American students to grade 60 objects as to how typical they were in the category of "furniture," from 1, very good example, to 7, very bad example. Rosch found that a chair ranked first as the most typical object of the category of furniture. In Rosch's approach to categorization, prototypes are the clearest instances of a category. The chair is the prototype of the category furniture, and the kitchen chair is the prototype of the category of chairs. The group's prototype was statistically calculated from the individual judgements of prototypicality: it is the most frequently cited best example in a given population. "To speak of a prototype at all is simply a convenient grammatical fiction; what is really referred to are judgments of prototypicality" (Rosch, 1978). In other words, the experiment did not ask for subjects to grade objects but to grade the application of the category name "furniture" to these 60 objects in order to analyze how linguistic vagueness (intensional and extensional) leads subjects to structure a set of physical objects according to how they conceive of a term's intension and extension. As indicated by the high degree of agreement between subjects involved in the task, ranking objects under a linguistic label was found to be a meaningful and natural task. Rosch and other researchers conducted many such experiments with other classes of objects, all leading to the same results: graded categorization is a natural psychological process.

Furthermore, judgments of prototypicality were not influence by whether the categories were *natural* or *nominal*. "Chair" or "bird" are natural or taxonomic categories because they refer to a set of real-world objects sharing

characteristic features. Nominal categories are conceptual categories that can be defined analytically, by a function or a use, such as “furniture” or “seat.” In every experiment, natural and nominal category names led to the same prototype effect and graded categorization of objects.

In Rosch’s experiment, a bench or a stool ranked in the middle of the set, next to a lamp and a buffet. The lowest-ranking objects were an ashtray, a fan, and a telephone. In her results, it is notable that a chair was more frequently ranked first than a bench or a stool that were ranked well after a group of high-ranking objects including “chair, sofa, couch, table, easy chair, dresser, rocking chair, coffee table, china closet, etc.” If only commonality and proximity of features were at work, surely a stool and a bench, which are physically related to a chair, should be closer to the chair in the ranking, at least before the “china closet.” Although similarity is a central factor in category formation, objects are also ranked according to experience of how they are found in association in the real world, for example, furniture in a given room, such as a living room in the case of the American students in the study. In fact, the furniture ranking varies with the ranking population, as shown by Dirven’s duplication of Rosch’s study on a group of German students. These students graded “bed” first in the category instead of “chair,” and they also moved “closet” from 56th position in the American ranking to 5th. Because categorization is context-dependent, the term “furniture” as used by the American group is not equivalent to *möbel* as used by the German group; the two categories are ordered differently, reflecting two visions of the world.

In the category of “furniture,” no feature is saliently common to all objects, nor can a set of sufficient and necessary features be extracted from

the observation of objects in the category. What loosely holds all the objects together is the vagueness, both intensional and extensional, of the term “furniture,” not the physical properties of the objects. Ultimately, the degree of prototypicality of a given object in a category such as “furniture” derives from contextual and idiosyncratic experience of the world, that is, from usage; Echoes of this idea can be found in more recent works on the effect of frequency and rich memory on language acquisition and category formation, as discussed by Bybee (2001, 2007, 2010).

The concept of prototype as defined Rosch, focuses on those few objects that are the most typical and does not attempt to take into account the organization of the category as a whole. What allows subjects to grade objects in a category is not only a sense of typicality but also the fact that other objects, despite not being as typical, are nonetheless included in the category. Following Wittgenstein’s insight, another way of conceiving of a prototype is to think of it as a central tendency that derives from the observation of all instances, each contributing to a certain extent to the abstract concept (Posner & Keele, 1968a,b; Goldstone et al., 2003). Just as vagueness is not a quality of an object, prototypicality is not the property of an object in a category, not even its best example. In this work, a categorical prototype is defined as a construct, abstracted from all objects to which a concept applies and of which each object in the category is an instance. In this definition, a prototype is a central tendency and, intension and extension are co-extensive, emanating from a loose network of associations built on chains of similarity and graded categories. This definition of a prototype is the foundation of this work. It is distinct from but not incompatible with exemplar theory, which holds that subjects actually form categories by storing individual exemplars

in their memory. New instances are compared to previously stored instances for categorization (Nosofsky, 1988a,b).

2.3 Phonemic categorization

This section relates the concepts of categorization, prototypicality, and vagueness to the concept of phoneme. As a phonological unit, a phoneme exhibits the properties of a category whose members' typicality can be graded and ranked according to a set of features, whether articulatory (production) or acoustic (perception).

2.3.1 Phonemes, phones, allophones

An inventory of phones, as documented in human languages is provided by the International Phonetic Alphabet (IPA). The IPA is a symbolic system of representation designed to transcribe the many aspects of the sounds of natural human languages with a high level of accuracy. In this system, the phones (vowels and consonants) are represented by a set of letters, and their variations (in terms of features) by diacritics. The following transcriptions are by Hayes (2008):

1. [d̪̩ɪs ɪz ə f̪̩nɛt̪̩i̪k̪̩ t̪̩s̪̩h̪̩.l̪̩ɛ̪̩n̪̩.sk̪̩.ɪ̪̩p̪̩f̪̩ɪ̪̩n̪̩]

This is a phonetic transcription

2. /d̪ɪs ɪz ə f̪̩nɪ̪̩mɪ̪̩k̪̩ træ̪̩n̪̩.sk̪̩.ɪ̪̩p̪̩f̪̩ə̪̩n̪̩/

This is a phonemic transcription

The phonetic transcription (1) is intended to precisely transcribe all aspects of the sounds, phones, and variations of phones, as they were produced in the utterance. The phonemic transcription (2) retains only those

features that are functional in the language, those that are perceived as categorically distinct. The phonological symbol /ð/ of *this* has phonetically been pronounced as [d̪̩ɪs], a glottalized (') affricate consonant whose stop part, the alveolar /d/, has been realized with a dental quality marked by („).

The presence or absence of these additional features may be contextually conditioned and predictable or sociolinguistically conditioned. These variations in form are not semantically meaningful or distinctive. They are generally not consciously perceived by speakers, unless they mark social or idiosyncratic particularities:

$$/\ddot{\text{ð}}\text{i}s/ = \{[\text{d̪̩}\ddot{\text{ð}}\text{i}s], [\text{d̪̩}\ddot{\text{ð}}\text{i}s], [\ddot{\text{ð}}\text{i}s]\}$$

Paul Passy, one of the founders of the International Phonetic Association, described the IPA as “an alphabet based on the principle *one symbol for each sound*. At the same time the same symbol may be used for representing several sounds which are very much alike, and the grouping together of which under one symbol cannot give rise to practical inconvenience” (Passy, 1887, 1907). In this way, Passy defines the phoneme as an abstract construct subsuming a range of phonetic instantiations or *allophones*: a phoneme is a category name for a group of related phones whose distribution is contextually and/or extra-linguistically constrained. The concept of *phoneme* is closely related to the concept of a category with a prototype, as defined in the previous section, in the following characterization given by Jones (1964), a phoneme is:

a family of sounds consisting of an important sound of the language (generally the most frequently used member of that family) together with other related sounds which “take its place” in particular sound-sequences or under particular conditions of length or stress or intonation. (Jones, 1964: cited by Taylor (2004))

Each phoneme's instantiation is contextually and situationally constrained, and, in spite of being phonetically different with each instantiation, a phoneme is categorized systematically as the same unit. This behavior pertains to the general domain of categorization. In the subsequent examples, it will be shown how prototypicality and vagueness apply to phonemes.

2.3.2 Allophonic variation: French velar plosives

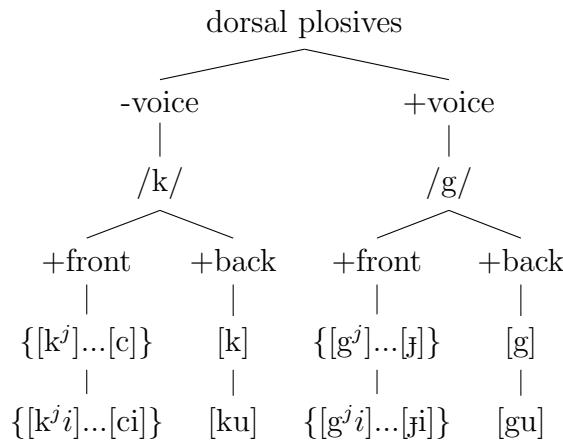
A well-documented case of allophony is that of the French velar plosives denoted by /k/ and /g/, presented in the following paradigm:

	front vowel	back vowel
voiceless	/ki/, <i>qui</i> , “who”	/ku/, <i>cou</i> , “neck”
voiced	/gi/, <i>gui</i> , “mistletoe”	/gu/, <i>goût</i> , “taste”

French speakers do not consciously perceive a difference between the consonants whether they are followed by a front or a back vowel. However, when phones are concatenated in a string, the articulation of a phone alters and is altered by the neighboring phones in a phenomenon described as co-articulation. In French, /k/ and /g/ are palatalized before a front vowel (/i/ and /y/). The point of articulation of the velars shifts forward in direction of the palatal position in anticipation of the following vowel. As can be expected, the importance of this shift varies and the articulation of the consonant can be realized anywhere on a continuum from very close to the velar position to the palatal position itself: {[k^j]...[c]} and {[g^j]...[ʃ]}. The velar consonants remain velar when they are followed by a back vowel such as /u/, since no adjustment is needed for co-articulation.

Furthermore, since the velar and the palatalized varieties are always found in different contexts, they are said to be allophones of the same phoneme

and to be in complementary distribution. It has even been suggested that since velar and palatal plosives are not phonemes but allophones in French; they can be grouped under the larger category of dorsals, which is the only functional and distinctive feature among plosives, as opposed to dentals and labials (Walter, 1977). The dorsal category of French can be represented by the following diagram:



A phoneme is a category. It is a set of objects whose values for one or more features differ from those of other objects in the set by gradient steps. In the case of the French velars, allophones of a given phoneme vary in their articulation between the two points that constitute the limit of a continuum of possible variation. In the previous example of /g/ and /k/, the focus was on intra-categorical variation. Separate phonemic categories can also be distinguished from one another by continuous features, such as voice onset time.

2.3.3 Gradient features, binary categories: Voice Onset Time

Lisker & Abramson (1964) showed that voiced stops /b, d, g/ and voiceless stops /p, t, k/ could be distinguished by a single feature: “the interval

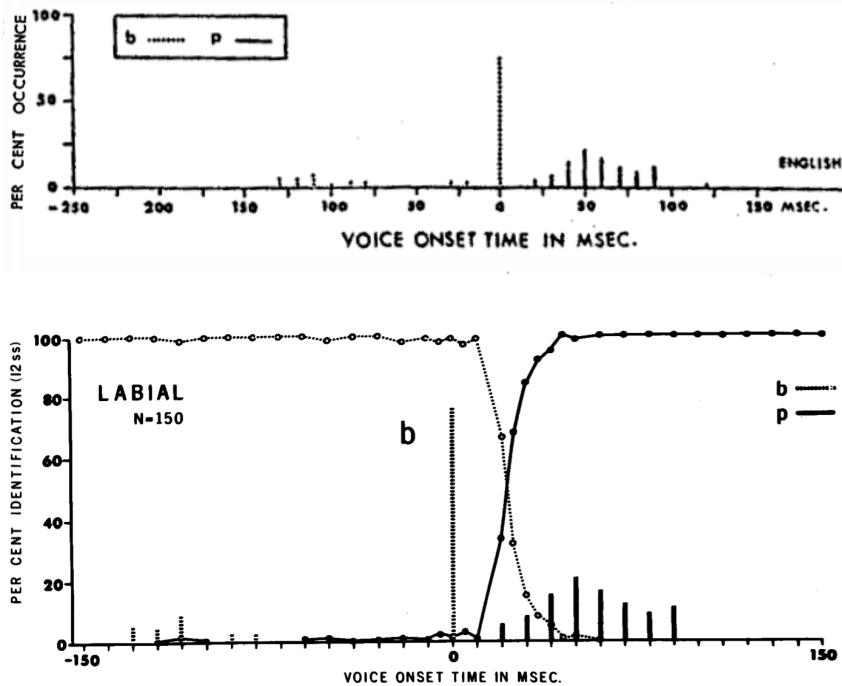


Figure 2.4: **Top:** Voice onset time distribution: labial stops of two-category languages. From Lisker & Abramson (1964).

Bottom: Identification curves as functions of VOT values. From Lisker & Abramson (1970).

between the release of the stop and the onset of the glottal vibration” or voice onset time (VOT). Figure 2.4 (top) presents the distribution of VOT for labials for a group of speakers of English (from Lisker & Abramson, 1964) on a [-150 ms, +150 ms] continuum. As reported by the authors, VOT values “cluster near some favored or ‘modal’ value.” This mode is at 0 ms for /b/ and at +50 ms for /p/. For /b/, other VOT values are relatively infrequent and located before 0. For /p/, VOT values are normally distributed around the mode. Similar results were found for apical and velar stops, and in various other

languages as well, although the location of the modes differs slightly. Most importantly, VOT values were produced in a categorical way; the typicality of a VOT value for a phoneme can thus be expressed in terms of frequency of production.

In subsequent perception studies by Lisker & Abramson (1970), subjects were presented with a series of stimuli varying in VOT by small increments on a [-150 ms, +150 ms] continuum. The authors found that in an identification task, the subjects' perceived phonemic boundary of subjects did not exactly match the distribution of produced VOTs. As shown in Figure 2.4 (bottom), the function curves for the identification of the two phonemes have a crossover point around $\simeq +25$ ms.

In a third round of experiments, Abramson & Lisker (1970) had subjects discriminate between triads of stimuli on a [-150 ms, +150 ms] continuum (Figure 2.5). Two of the stimuli were identical in their VOT and the third was different by a 2, 3, or 4-step increment (20, 30, or 40 ms). Results indicate that discrimination accuracy dramatically rises on the phonemic boundary ($\simeq +25$ ms).

The categorical behavior of stop consonants' VOT has been shown to be a cross-linguistic property. The perception of the category boundary was also found to develop very early on in infants (Eimas et al., 1971). This development depends on experience, as infants could actually be trained to perceive phonemic categories not present in their surrounding language (Aslin et al., 1981). When trained, animals such as chinchillas (Khul & Miller, 1978) and quails (Kluender et al., 1987) were also able to discriminate between voiced and voiceless stop consonants. They categorized gradient stimuli in reference to the same $\simeq 25$ ms threshold. The animals accomplished this categorization

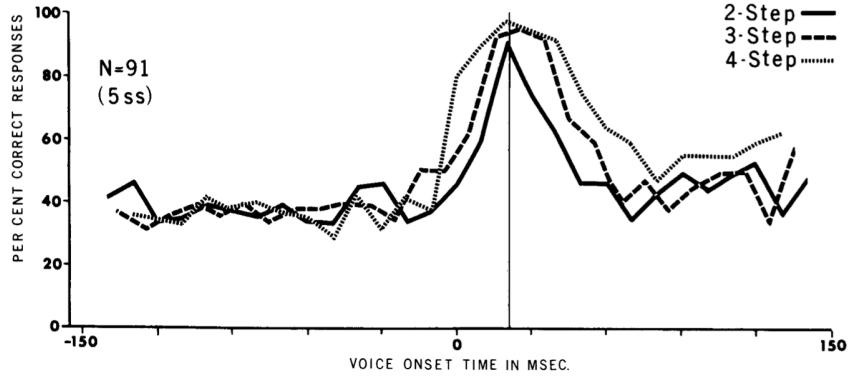


Figure 2.5: Discrimination accuracy between English labials. From Abramson & Lisker (1970)

in pure physical or phonetic terms since they do not possess a linguistic system.

However, Pisoni (1977) noticed that intra-category discrimination is always inferior to inter-category discrimination. Pisoni claimed that in spite of the gradience between the VOT stimuli, subjects were mostly sensitive to the presence or absence of a feature to discriminate among phonemes. As will be discussed in the present work, the presence or the absence of features are themselves a categorization of gradient stimuli. For the VOT of labial stops, the temporal boundary of about 25 ms loosely separates two groups of values. Below this value, VOT is phonologically absent; above it, it is phonologically present.

In Figure 2.4 (bottom), the VOT borderline zone spans the [+10 ms, +50 ms] interval, in the middle of the graph, where the functions rise and fall. On both sides of this zone, identification functions are binary, constantly at the maximum (100%) or constantly at the minimum (0%). Outside of the borderline zone phones are invariably categorized as one or the other phoneme.

For values between +10 ms and +40 ms, identification consistency drops. Around 25 ms, “one must look for possible perceptual instability” (Lisker & Abramson, 1964): identification can lean toward one or the other phoneme (50%). In Chapter 5 of this dissertation, the application of fuzzy set theory to the VOT continuum of /b/ and /p/ leads to the same result of 25 ms. This value is the categorical threshold above which category membership increases and below which it decreases.

VOT is a phonetic feature whose gradient variation enables phonological categorization. VOT is just one among many others. In their work on quails, Khul & Miller (1978) described “phonetic categories [as] examples of polymorphous concepts with no single necessary or sufficient condition for class membership. Instances of such concepts can at best be described as having a family resemblance.” The next section illustrates how the ideas of Wittgenstein apply to phonemes when they are considered as a network of features.

2.3.4 The case of English /t/: Taylor (2004)

Taylor uses the phoneme /t/ in English to illustrate the range of possible articulatory realizations encompassed by a single phonemic unit (see Taylor, 2004: Chapter 13). In the category denoted /t/, Taylor finds allophonic instantiations that can be contextual, stylistic, and regional:

$$/t/ = \{t, \text{t̪}, t^h, t^s, t^{s'}, t^r, \text{t̫}, \text{t̬}, t^l, \text{s}, \text{t̪}, \text{d̪}, d, r, \text{x̪}, \tilde{t}, t^l, \text{no aspiration}\}$$

Taylor analyzes the phonemic category in structural terms, characterizing allophones by their phonetic features. Although it is tempting to posit /t/ as the best exemplar of the /t/ category, the set of allophones of the /t/ category is strikingly reminiscent of the idea of family resemblance: a network

of objects linked by complex relationships and emanating from usage. These 18 allophones can be only loosely related to one another since no one phonetic feature is common to all these phones. Some phones share some features ([t], [d], and [r] are alveolar), some do not at all (r and ?).

Unlike furniture or even VOT, it might prove difficult to ask speakers to rank the perceived typicality of the various allophones of a phonemic category by their articulatory features, and to order the allophones in the category, like the items in the list of furniture. Speakers of a language are, for the most part, unaware of allophonic variation and they do not naturally differentiate allophones. Even if asked to consciously do so in an experimental setting, subjects might not be at all successful (see Walter & Hacquard, 2005: for example).

Speakers of a language are aware of allophony when it functions as sociolinguistic or idiosyncratic marker. In these cases, the variation itself is secondary to what it signals. A native speaker of English will identify the flap [r] in [bərər] (“better”) as a form of /t/ and, more importantly, will be informed of the American origin of her interlocutor.

A study of the statistical frequency of usage of the allophones would probably reveal that the ranking in terms of typicality varies with individuals and regions, much like for the category of furniture. The flap [r] is more widely used in the USA than in Britain, for example.

The allophones could also be organized as a sequential chain of similarity, in a “network” based on feature closeness. Figure 2.6 is adapted from Taylor (2004). As pointed out by Taylor, such a diagram can account only for structural similarity, not for the development of the network in time nor for where it started. This, again, varies from individual to individual, although

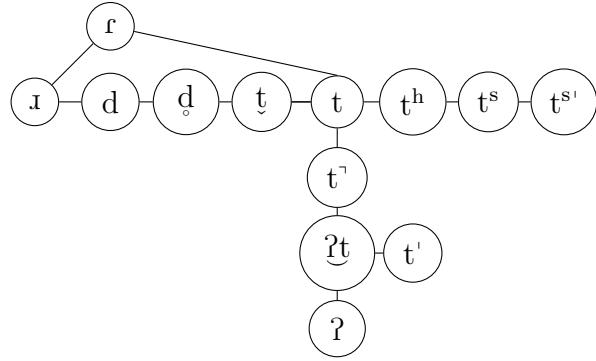


Figure 2.6: A network of allophones of /t/, adapted from Taylor (2004)

/t/ is structurally a good candidate, as all sub-chains originate from it. All three sub-chains are context specific. The chain to the left of /t/ would be intervocalic allophones, adding more or less voicing to the voiceless /t/. The chain to the right of /t/ would be the chain of aspiration, gradually turning into a dental sibilant. The chain below /t/ is that of the coda position associated with possibly gradual types of release (see Taylor, 2004: Chapter 13 for details).

For a given individual, the inclusion and exclusion of phones in an allophonic network changes over time: first as it expands during acquisition, then as the individual is confronted to or surrounded by changes in her usage. A British speaker may see her /t/ category change if she moves to the USA, as it expands to include the allophonic flap [r]. A phonemic category has fluctuating borders, typical of a category whose association with a concept is vague. Such vagueness can create uncertainty about the phonemic category of an allophone. Phonologically, /d/ and /t/ distinguish the minimal pair /raudər/ (“rider”) and raitər/ (“writer”) by an opposition of voice. However, in American English, the /t/ is more likely to be realized with some voicing,

anywhere between [raɪtər] and [raɪrər]. It is often difficult, if not impossible, to assign the phone to a category without the larger context of the sentence that contains the word. In this case, the application of the category names /t/ or /d/ for a phone is vague since the phone could be either in American English.

Phonemes may now be characterized as category names subsuming a range of allophones in a class whose boundaries are fluctuating and overlapping. These categories, the phones they include, and how these phones are structurally organized vary from individual to individual and derives from each individual's experience. Thus, for each speaker of a language, there exist borderline allophones for which the application of a category name is uncertain. Although the category name of a phonemic category is an allophone taken from the category, the structural prototype of a phonemic category is a construct abstracted from all the allophones and of which each allophone is an instance.

2.3.5 American vowels (1 of 2): Peterson & Barney (1952); Hillenbrand et al. (1995)

Another approach to analyzing phonemic categories is not structural but quantitative, based on acoustic data. Research in acoustics has shown that vowel formants are realized within a rather wide range of f_1 and f_2 combinations. Peterson & Barney (1952) studied instances of 10 vowels among 76 speakers (see also Hillenbrand et al., 1995). In Figure 2.7, each point corresponds to an allophonic instance of a phoneme (f_2 as a function of f_1). Peterson & Barney found that the realization of vowels varies widely but that allophones are roughly contained in a zone (the ellipses on the figure) and that these al-

lophonic zones overlap. Notable as well is the convergence of the points to the center of the figure, toward a central vocalic tendency for all phonemes. In the global category of phonemes, each phonemic category is a point in a chain of similarity: roughly from /i/ to /u/. Uncertainty due to vagueness is more likely to occur between allophones of close phonemic categories (/i/ and /ɪ/, /ɪ/ and /ɛ/, etc.) than between distant categories (/i/ and /ɔ/, /ɛ/ and /u/, etc.). Depending on their dialectal origin, some subjects have trouble differentiating [i] and [ɛ], as in [tm]/[tɛn] or [pɪn]/[pɛn]. They interpret both phonemes as /ɛ/.

French speakers do not distinguish between /i/ and /ɪ/, which are allophonic in French. To a French ear, “heat” and “hit” are equivalent. French learners of English have to actively work on their perception and production to divide into two allophonic zones their single native category. In a similar fashion, when given proper training, Japanese subjects were able to distinguish instances of English /l/ and /r/ although the two phonemes are allophonic in Japanese (see for example Logan et al., 1991). In the concept of phoneme as a network of allophones, what might characterize a “foreign accent” is the pronunciation of certain phonemes as allophones that diverge from those naturally present in a native subject’s network. Peterson & Barney concluded their paper by stating that “the data [...] reveal that both the production and identification of vowel sounds by an individual depend on his previous language experience” (Peterson & Barney, 1952). In that sense, a native-like accent is a vague concept as well. The “strength” of a foreigner’s accent would depend on the native speaker’s experience and ability to comprehend foreign allophonic deviation from natural variation.

According to the results of Peterson & Barney and those of Hillen-

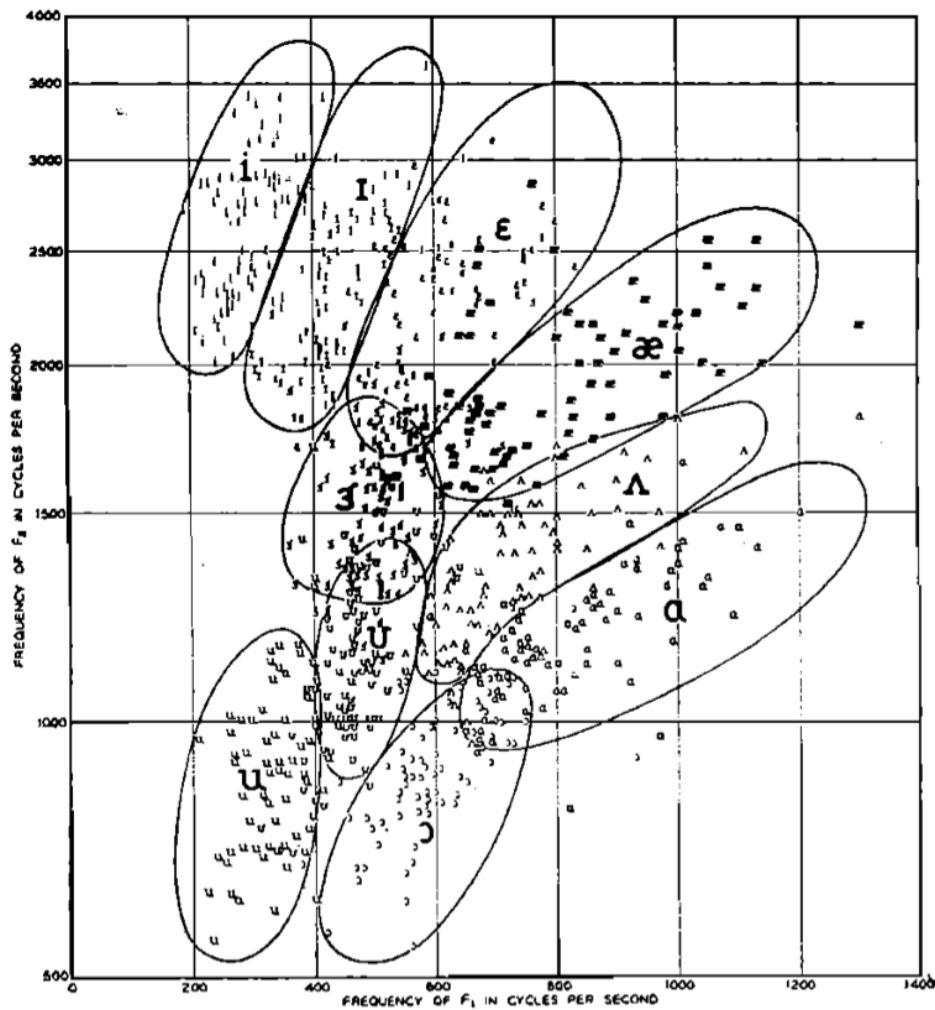


Figure 2.7: Frequency of second formant versus frequency of first formant for ten vowels by 76 speakers, Peterson & Barney (1952)

brand et al., phonemes constitute a category name which subsumes a range of allophones in a category with fluctuating and overlapping boundaries. The boundaries vary from individual to individual, such that there exist borderline

allophones for which the application of a category name is uncertain.

2.3.6 American vowels (2 of 2): “Perceptual magnets”

Khul (1991) used the concepts of categorization and prototypicality to develop her work on American vowels from the study of Peterson & Barney. She conceived of vowels as graded categories within which certain allophonic variations are more typical than others. These more central allophones constitute the core of a phonemic category: its *perceptual magnet*.

In a set of allophonic instances of the American English vowel /i/, Khul selected two allophones whose f_1 and f_2 combinations were ranked by native subjects, one as the highest for its typicality in the category, the other one as the lowest for its typicality in the category. The highest ranking allophone was assumed to be the prototypical center of the phonemic category; the lowest ranking allophone was assumed to be the non-prototypical center of the phonemic category. Around these centers, Khul organized a set of synthesized vowels /i/ whose f_1 and f_2 varied from each other by 30 points on the perceptual mel scale. Khul’s analysis of the set covered the range of the /i/ vowels for male speakers originally from Peterson & Barney (1952) is presented in Figure 2.8.

Khul asked subjects to grade the stimuli for their *goodness* as members of the phonemic category of /i/ in English, on a scale from 7 (typical) to 0 (not typical). Subjects were presented with stimuli in a random order and they were not aware of the prototypical/non-prototypical status of the central stimuli. In Figure 2.9, each circle corresponds to one of the formant combinations presented in Figure 2.8. The size of the circles is relative to the grading of the allophones by the subjects. The south-east branch of the top left figure

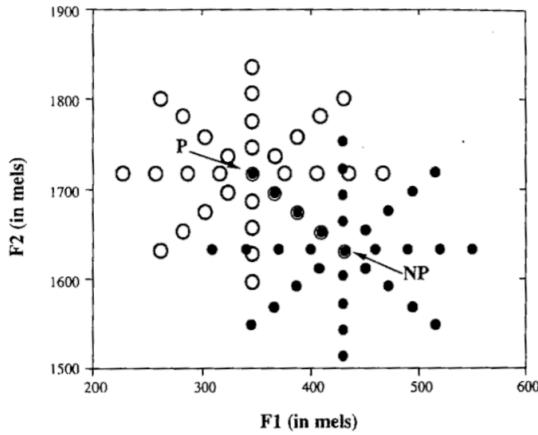


Figure 2.8: The prototype /i/ vowel (P) and variants on four orbits surrounding it (open circles) and the non-prototype /i/ vowel (NP) and variants on four orbits surrounding it (closed circles). The stimuli on one vector were common to both sets. From Khul (1991)

corresponds to the north-west branch of the bottom right figure. The goodness rating decreases as the distance of the stimulus from the prototypical center increases and the distance from the non-prototypical center decreases. For Khul, the consistency of rating among subjects suggested “that adult listeners, at least those who speak the same dialect of English, have an internal standard for the vowel /i/ that is quite similar.” Khul claimed that phonemes were conceived of as internally graded categories with a central tendency that serves as a reference point for the evaluation of other stimuli: in other words, a prototype.

Within the framework of categorization, phonemes have been described structurally as a network of allophones defined by gradient articulatory features varying along continua (Taylor, 2004) or quantitatively as contiguous regions along acoustic continua (Lisker & Abramson, 1964; Peterson & Bar-

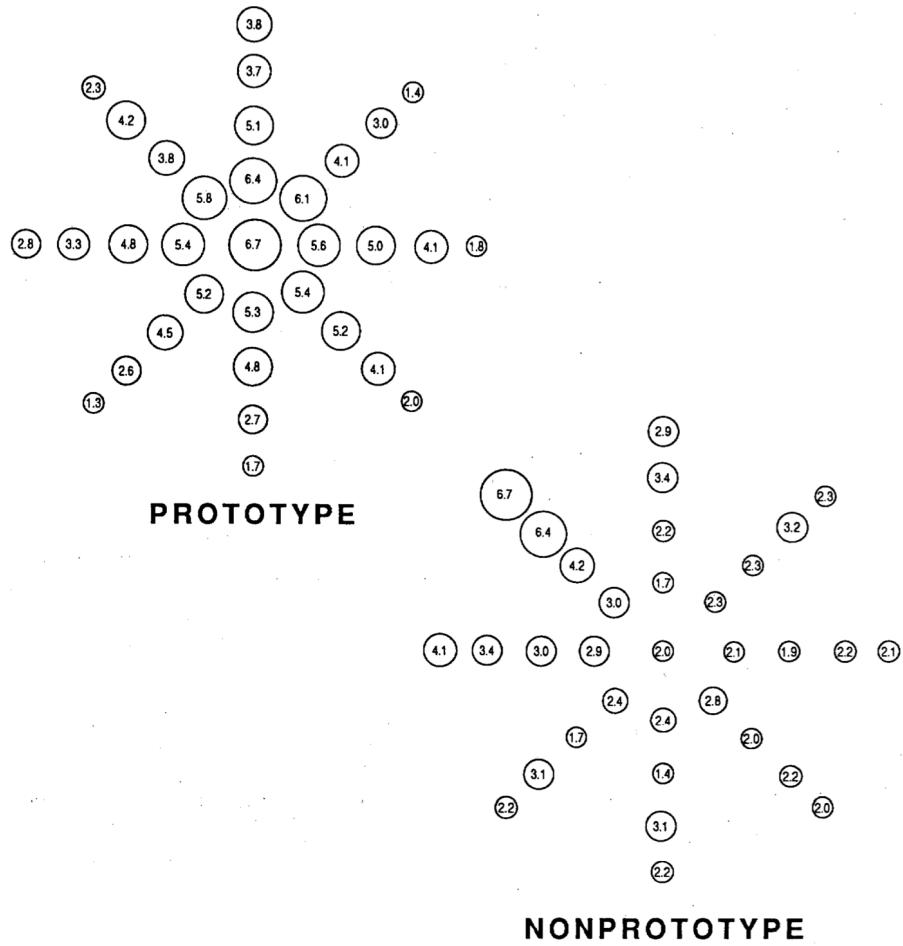


Figure 2.9: Category goodness (typicality) ratings for the prototype /i/ vowel, the non prototype /i/ vowel, and the variants surrounding each of the two vowels.

From Khul (1991)

ney, 1952; Khul, 1991). Both approaches lead to defining phonology as the discretization of the phonetic continuum into categories functioning as discrete units, the phonemes. These phonemes are groups of ordered allophones (or feature vectors). Phonology is thus the symbolic representation of the oth-

erwise mostly unconscious process of categorization at work among speakers of a language.

Through the process of categorization – that is by classification and identification of phonetic features into phonemic categories – contextual distortion (co-articulation, speech rate, style, etc.) is extracted (mostly by normalization) and communication through discrete functional units is made possible.

However, it has been shown that these contextual distortions are not discarded. Assessing Kuhl's work, Lively & Pisoni (1997) found similar results but also found that prototypically varies contextually: “the goodness of a category member is a *relational property* rather than an absolute one.” Therefore, a theory of the phoneme in terms of prototype must be large and flexible enough to encompass normalization and contextual variation.

Phonetic variation, whether it is contextual, sociolinguistic, or idiosyncratic, is kept in memory and facilitates subsequent identification and categorization of novel stimuli. Nygaard & Pisoni (1998) suggest that the normalizing process that extracts high-level linguistic features from speakers' variations is co-extensive with an analysis of the low-level variation. The authors named this idiosyncratic variation the *indexical information*: the unique acoustic properties of an individual's voice and speech. This information is retained by other speakers in their memory.

Experimental settings can train subjects to focus on specific perceptual cues that they might not retain as efficiently in a natural conversation. Therefore, Goldinger (1996) proposes that the memorization of low-level variation is selective. Systematic memorization of all variations seems unnecessary and, possibly, physiologically impossible. Selective memorization would apply to

variation only if it is linked to a meaningful distinction. Variation is therefore stored in a structured category of its own, in relation to which a speaker can evaluate novel instances' variation.

The PRInt model is based on this idea; it extracts high-level linguistic features in a graded category, and, in parallel, it analyzes the phonetic variation as a graded category. This allows an analysis of the phonological variation leading to a difference of meaning on the one hand, and phonetic variation leading to para-linguistic differences on the other hand.

2.4 Conclusion

Graded categorization is a natural human ability that allows subjects to abstract concepts and categories from a group of objects related by a network of features. The prototype of a category is not a typical object or best exemplar from a category but rather a conceptual center, structural and quantitative, of the category's organization. Phonemes have been described as graded categories with such a prototypical center. In the next section, intonation is presented within the framework of categorization. It will be argued that, if categorization is a general principle of human behavior, intonation should be described phonologically and phonetically in terms of vagueness and prototypically, from the systematic analysis of phonetic instances of intonation contours.

Chapter 3

Intonation

H and L are phonological abstractions, comparable to phonemes, and there is no reason to expect them to be realized always in the same way. Rather, the phonetic realization of H and L – like the realization of any other phoneme – is subject to a variety of conditioning factors, which may make any occurrence of H and L come out phonetically in a quite different way from some other occurrence (Ladd, 2008).

3.1 Intonation

The domain of research of this study is intonation. Intonation is defined by the three following characteristics, adapted from Ladd (2008). Intonation is:

Suprasegmental Suprasegmental features include fundamental frequency (F_0), intensity, and duration. This study focuses on F_0 and duration, leaving intensity for further investigations.

Phrasal The meaning of intonation is “appl[ied] to phrases or utterances as a whole” (Ladd, 2008). The phrase *intonation contour* refers to “the integral melodic pattern” (Bolinger, 1978) of a sentence, also called the intonation phrase. More generally, intonation applied to group smaller

than the sentence will be referred as *intonation units*.

Linguistically structured The features of intonation are organized in a system of categories and relations (such as phonemes). This definition implies that there exists a phonology of intonation.

In this study, intonation refers to the change of fundamental frequency (F_0) as a function of time, associated with a particular communicative function. The same phrase or sentence can have a totally different meaning when the intonation unit or contour changes. Bolinger (1961) used typographical levels to represent the intonation contour of a sentence:

I don't want to go to ^{mo}row I don't want to go tomorrow

The intonation of the first sentence indicates that the speaker might want to go another day but not tomorrow. The intonation of the second sentence indicates that the speaker does not want to do what has been previously discussed. In French, under certain conditions, a final raising tone of voice can make of any unit a question, be it a single word (*moi ?*, “me?”) or a whole utterance, as in Figure 3.1 (*Maman va venir ?*, “Is mom going to come?”).

Without any electronic resources, early scholars such as Passy (1887) relied on their perception to draw intonation contours by hand. In the contours Passy drew (figure 3.2) in the late 19th century, he marked voiced segments by a solid line and voiceless segments by a dashed line. He also aligned the turning points of the curve with specific segments of the sentence. For anyone familiar with the concept of tones and interpolation, Passy seems like a distant precursor unbeknownst to himself.

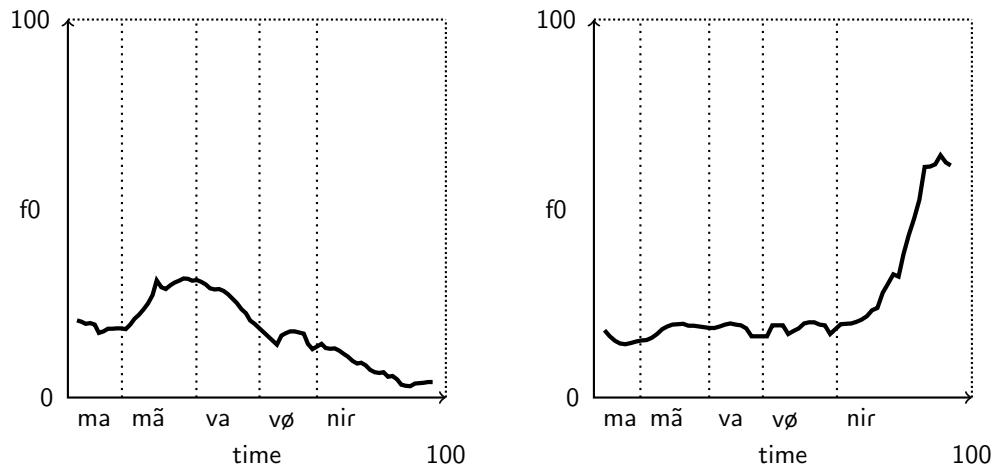


Figure 3.1: The intonation contours of a declarative (left) and of a closed question (right) over the same sentence *Maman va venir*. Time and F_0 have been normalized, syllable boundaries are marked with dotted lines.

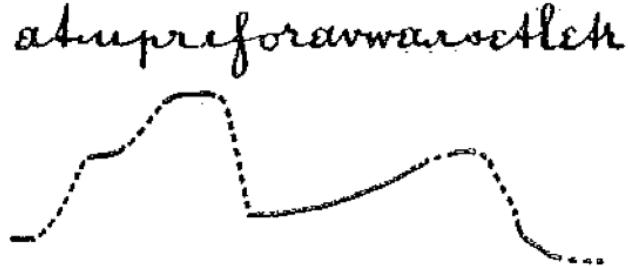


Figure 3.2: A hand-drawn *ton* (intonation contour) by Passy (1887).
Et tu préférerais avoir cette lettre (“And you would prefer to have this letter”)

3.2 Synthetic and analytic studies of intonation

In the research on intonation, two main groups of approaches can be distinguished, with some overlap. The first type of approach to intonation is

synthetic, and its goal is to model or artificially recreate the F_0 signal of an intonation contour as accurately as possible. The other approach to intonation is analytic, and its goal is to describe intonation in terms of features grouped in an organized structure. This distinction is explained within the larger domain of categorization.

Contrary to phonemes (i.e. phonemic categories) intonation has a peculiar nature: the category name of intonation units can only be evoked by a symbol taken out of the category (by a phrase) or by an instance of the category itself but implemented on other units (phonemes). The access to intonation units is not independent from other linguistic units. Xu (2011) refers to this phenomenon as the *lack of reference problem*: “a pivot that serves as both a starting point of inquest and a point that one can comfortably fall back on.”

Someone can pronounce an instance of the phoneme /a/ to refer to the category name for the phoneme. If someone says: “please pronounce a series of 10 a’s” to another person, this one will actually produce 10 different instances of [a].

Someone cannot pronounce an intonation contour of a question to refer to the category name for the contour. If someone says: “please pronounce 10 ‘questions’” to another person, this one will have to actually implement the contour on 10 different sentences. “Question” is the category name for an articulatory behavior that cannot be accessed, other than by naming it with a phrase or by realizing it as an instance of the contour applied to an utterance.

3.2.1 Synthetic studies

Synthetic approaches rely on instrumental research to model intonation. Their primary goal is to generate (synthesize) the F_0 signal of an intonation unit or contour by artificially emulating the biological mechanisms responsible for its production. The physiology of the vocal tract and the articulatory organs, as well as their connection to the brain, are studied in great details so that the level of precision in the modeling be very high. Engineered speech synthesis is a direct application of such research, with an accent on accuracy and naturalness of the generated intonation. Such models include notably the *command-response* model of Fujisaki (Fujisaki, 1983; Fujisaki et al., 2005), and the PENTA model of Yi Xu and colleagues which will be briefly presented here (see for example Xu, 2004a, 2007; Prom-on et al., 2009)

The PENTA model assumes the existence of communicative functions at various levels (lexical, sentential, etc.) that are directly encoded as a complex of articulatory orders that sequentially executed to reach a series of targets, forming the intonation unit and/or contour of an utterance. The model synthesizes the surface signal by synchronizing the target to be reached with syllables of the utterance over which the contour is implemented. Depending on the utterance, the targets can only be partially reached, but the model continues to reach the next target and keeps doing so until the end of the utterance. The idea underlying the PENTA model and similar models is that a communicative function is directly encoded bio-mechanically in the speaker's brain and vocal apparatus. There is no need for any phonology since the goal is not to describe intonation from a linguistic point of view but to generate highly accurate and natural sounding units and contours, a task at which the models of Xu and Fujisaki are very successful.

In a model such as PENTA, the acoustic form of the contour is the output, the highest level to be reached whereas it is the input or lowest level of a phonological representation. Because they are so accurate, such models treat each sentence as a singular, although complex, event instead of looking at it as an instance of an abstract concept or category. Arvaniti & Ladd (2009) pointed out that “any complete theory of intonation also needs an abstract description that accounts for the linguistic aspects of the system and allows for predictions and generalizations based on this description.”

Synthetic approaches are sophisticated computational systems with no phonological agenda because they are not designed as a symbolic representation of language but as a way of mimicking natural processes in order to obtain similar results. “The map is not the territory” (Korzybski, 1941) and an accurate reproduction of the mechanics does not mean that anything has been said about the cognitive organization that underlies it. From the point of view of categorization, these models fail to explain how instances, once they have been generated, are identified and classified into categories, that is, how an intonation unit or contour is linguistically functional. In other words, these models are strictly limited to their highly accurate synthesis function.

3.2.2 Analytic studies

Analytical approaches to intonation preceded the synthetic approaches, principally due to technological progress enabling more sophisticated and extended instrumental studies in recent times. Analytical studies derive in large part from the principles of categorization presented in Chapter 1. The goal of analytical studies is the description of intonation into a structured and organized linguistic system. These approaches are interested in finding a phonology

of intonation by discretizing the phonetic continuum of the acoustic signal into a finite set of discrete phonological features organized in categorizable patterns. From the point of view of categorization, these approaches, best exemplified by the Autosegmental Metrical (AM) model of Pierrehumbert (1980), have the advantage to be feature-based and thus make identification and classification of intonation units or contours possible. This ensures that intonation is described in a linguistically functional way.

3.2.2.1 Family resemblance, prototype, vagueness

The four utterances presented in Figure 3.3 have all been produced as a closed question by a single native speaker of French. However, the four utterances of this contour clearly differ in pattern, although it is also possible to recognize a common identity among them: a family resemblance. The common salient feature is a final rising movement of the F_0 curve. The size of this movement is different for each sentence, but it is always contained within the last syllable. In Figure 3.3, (A) crucially differs from the three other sentences by an absence of movement before the final rise. (B) and (D) both contain the verb phrase *vous voulez* (“you want”) and both are characterized by the presence of an F_0 peak on the third syllable, the end of the verb phrase. (C) also contains the verb phrase, but it is altered by the inclusion of the pronoun *en* (“some”). This pronoun and the end of the verb *vouloir* (“to want”) occur in the second and fourth syllables and are the locations of F_0 peaks.

If the intonation contours of Figure 3.3 are conceived categorically, as instances of a same prototype, the pattern variation of the utterances is in fact expected as long as they remain related to each other in some way, more or less, depending on the category. In the case of a closed question, the meaning

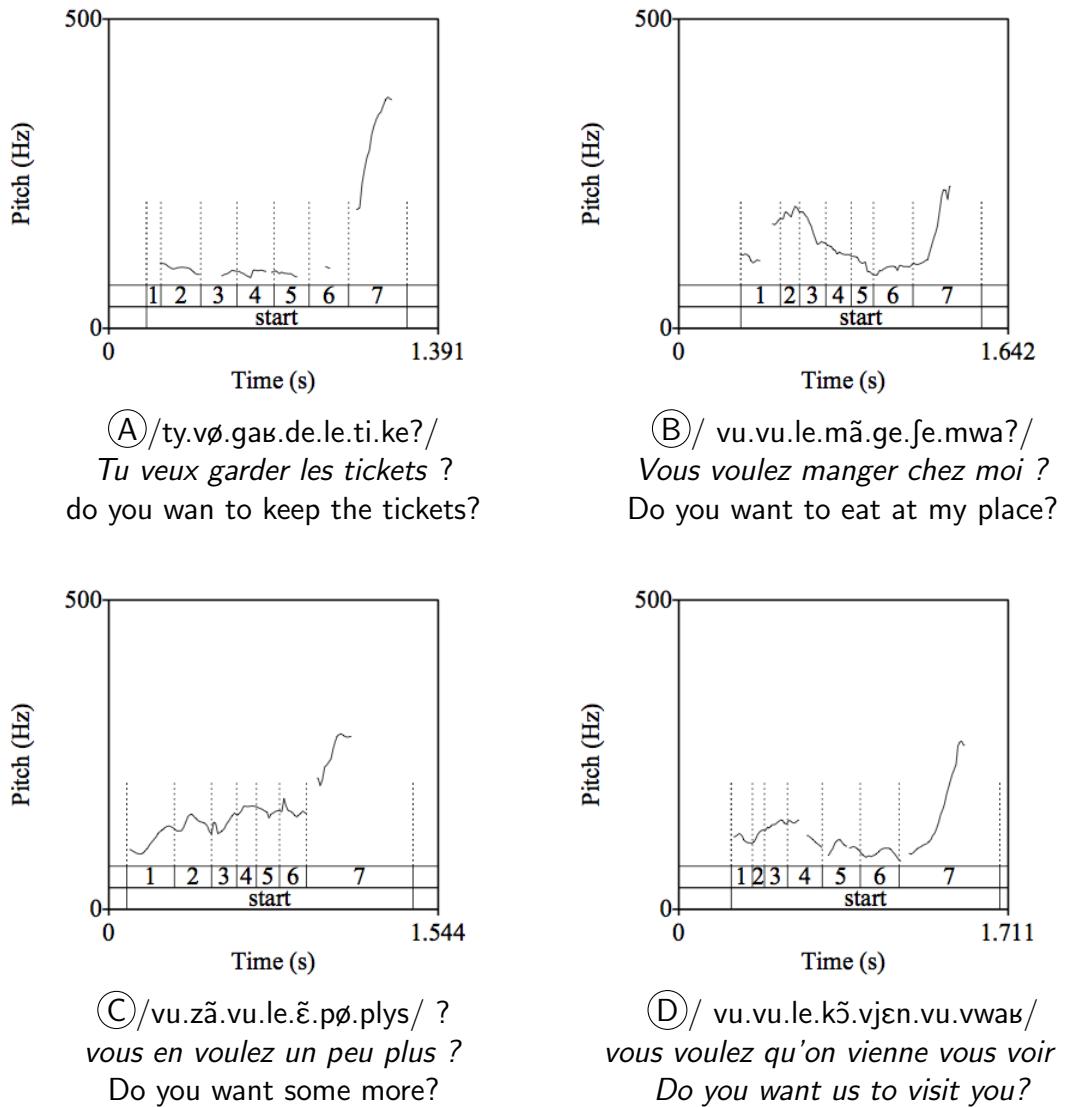


Figure 3.3: Four closed questions produced by a single speaker of French and displayed with the software Praat (Boersma & Weenink, 2012). Syllables have been numbered and separated by dotted lines

of the contour is fixed: either a question is asked or not. The implementation of the contour varies depending on contextual variables such as the nature of the phonemes, some intrinsically altering the F_0 curve of the intonation contour, the presence of phrases receiving an additional pitch accent (end of the verb phrase), or because of purely idiosyncratic alterations (e.g. the mood of the speaker, the fact she is eating while talking, etc.)

Finally, in some cases, a category name for intonation contours (e.g. question or irony) may prove uncertain. For example, if someone uses a particular intonation contour to mark the irony of her statement, her interlocutor, depending on her experience with irony and the other person, might mis-categorize the contour and interpret the meaning of the sentence literally as a statement in the absence of other cues. The sentence

Harry is a real genius

can convey its literal meaning, or the opposite, depending on the intonation contour and how it is categorized by the addressee (Cutler, 1974). The categorization of the sentence depends on the experience of both speakers and cannot be expressed in absolute terms. The concept of irony is vague and its instances might be mis-categorized, that is, misinterpreted. It is an assumption of this study that all intonational categories are vague to a certain degree.

3.2.2.2 Autonomous features and hierarchical structure

Discretization of the continuum, Pike (1945) Pike (1945) observed that speakers of a language use the same contours in the same situations and that their variations have to be “semi-standardized or formalized” to enable communication. The “patterns of variation” are only obscured by the multiplicity of superficially complex instances but intonation contours are not “su-

perimposed to lexical meanings” in a “whim or fancy”, they can be analyzed in some organized structure. Pike implied in his comments that intonation contours are linguistic categories, and, most interestingly, he inferred from his observations that intonation contours (as category names for types of contours) are “abstracted characteristic melodies”, an idea recalling that of a prototype.

Pike stated that in order to describe and compare – and therefore categorize – intonation contours adequately, they should be conceived as a series of linked pitch points forming a pattern of pitch points connected by lines. Most importantly, he noticed that the position of these points, in the utterance and in pitch, should be described relative to the position of the other points in the contour. Thus, gradient differences were abstracted away from actual F_0 values and pitch height was expressed on four levels from 1 (highest) to 4 (lowest). Latent in Pike’s study is the idea that by discretizing continua (time and pitch range), utterances of different length and F_0 range can be compared if they are interpreted as an ordered sequence of features on relative scales. In Pike’s work, a conception of intonation as a relation of acoustic events aligned to elements on a metrical grid is already present. Below is an example of Pike’s notation:

a)	Tom my !?	telephone number!?
	<u>2-</u> -4-3	<u>2-</u> -4 -3
b)	Has he gone ?	Where has he gone?
	<u>3-</u> °3-2	<u>2-</u> °2-4

The same patterns can be stretched on two sentences varying in length while remaining the same pattern, although distorted, aligned with metrical elements of the utterance, stressed syllables in the case of English. The two points of pattern (a) are aligned with both edges of the sentences in each case, in spite of the two extra syllables of the sentence on the right. In the

examples of (b), stressed syllables have been indicated by a degree symbol (\circ). The prominence of the last syllable is the same, but the pitch contour changes between a closed question (left) and a wh- question (right).

Number of features In Pike's and other structuralists' view (Wells, 1945; Trager & Smith, 1951), the features were treated as phonemic pitch points realized on one of four levels (Low, Mid, High, and Overhigh) and aligned with stressed syllables. However, in the previous example of Pike's notation (a), only ranking arbitrates that the utterance starts at value 2, the second point is given value 4 for being the highest pitch point, the last point gets value 3 for being lower than the second point, and finally the first point is assigned value 2 for being lower than the final syllable. This begs the question: why not use only three levels, or do the four levels have some sort of phonetic validation? A paradigm of 4 levels (or more) actually maintains some of the gradience of the acoustic signal¹. Bruce (1977) and Pierrehumbert (1980) reduced to two the number of necessary levels : high (H) and low (L). If pitch points, called hereafter *tones* are expressed relatively to all other tones in the sentence, then intermediate levels might seem necessary. If they are expressed relatively only to the adjacent group of tones, then the intermediary levels turn out to be superfluous in the economy of a truly discrete phonological description.

3.2.2.3 Autosegmental Metrical (AM) model

As in Pike's model, in the model of Pierrehumbert (1980) the tones are autosegmental, which means that segments and tones are placed on two

¹Bolinger (1951) suggested, instead of these 4 levels that he argued to be inadequate, a set of “configurations” or types of movements, such as “rise” or “fall”, in what has been termed the “levels vs. configuration” debate.

separate tiers of analysis. The model is also metrical because the elements of each tier are contained by phonological units in a hierarchical structure represented by Figure 3.4. The type of units and the number of levels in the structure varies with language and theory but there is always at least an intermediary level between the level of tones and that of the larger unit under the sentence.

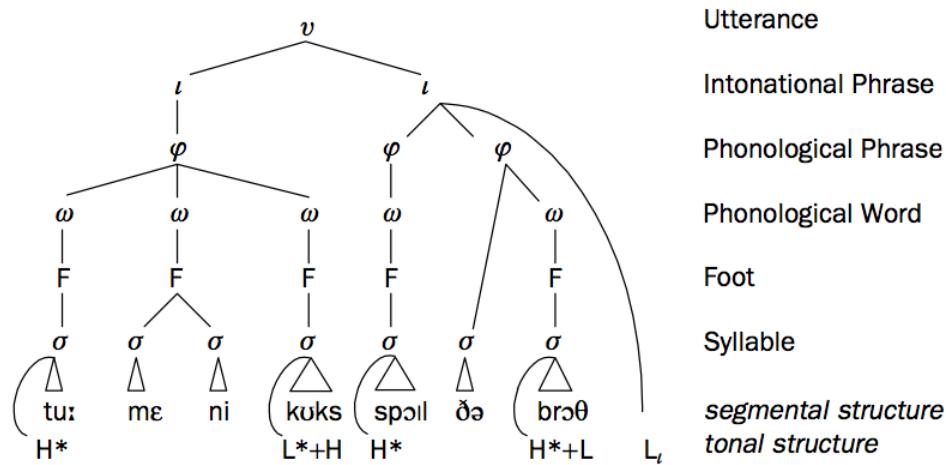


Figure 3.4: Hierarchical intonation structure. From Gussenhoven (2004)

The AM theory and the derived transcription system ToBI (Tones and Break Indices) (Silverman et al., 1992; Beckman et al., 2005)² model intonation as a tonal sequence composed of two types of intonation units, the *pitch accents* and the *edge tones*. Pitch accents are composed of a sequence of L and H tones and the central tone, associated with the prominent stressed syllable, is

²For further reference, the website of the ToBI community is maintained by The Ohio State University Department of Linguistics: <http://www.ling.ohio-state.edu/tobi/>

indicated by a star (*). A plus sign (+) unites the two tones of bi-tonal pitch accents. There are two categories of edge tones. *Boundary tones* are single tones associated with the beginning or the end of an intonation phrase and marked with a percent sign (%). *Phrase accents* (or *phrase tones*) are single tones between a pitch accent and a boundary tone, marked by a high hyphen (-).

Pitch accents and *edge tones* are category names for two groups of intonation units related by common features. All pitch accents have in common a central tone associated with a prominent syllable. They vary in number of tones and their configuration. Edge tones (both boundary and phrase tones) are associated with an extremity of a phonological unit and thus mark one of the two edges of the unit. Edge tones vary in their placement in an utterance. The two categories of intonation units of the AM model can be organized as follows:

$$\text{Intonation units} \left\{ \begin{array}{l} \text{Pitch accents} \left\{ \begin{array}{l} H^*, H^* + L, H + L^* \\ L^*, L^* + H, L + H^* \end{array} \right\} \\ \text{Edge tones} \left\{ \begin{array}{l} \text{Boundary tones} \left\{ H\%, L\% \right\} \\ \text{Phrase tones} \left\{ H^-, L^- \right\} \end{array} \right\} \end{array} \right\}$$

Starred tones of pitch accents are associated with salient units of the sentence (stressed syllables in English). Edge tones are associated with boundaries of higher level phonological constituents (phonological or intonational phrases). In the phonetic implementation, the alignment of the tones is not necessarily precisely on the associated elements; it can be before or after, depending on the sentence. This explains the variation in surface forms of utterances for a similar intended contour.

Prior to any categorization, a tone can possibly be assigned only two values (H or L) and can only be parsed for this value from left to right, within

a “window” (Pierrehumbert, 1980). The value (H or L) of a tone can only be relative to the domain around the tone: tonal values are assigned sequentially. Figure 3.5 presents the same pitch accent realized in two different ways depending on the position of the stressed syllable with which it is associated. The AM model dictates that the two level tonal representation ensures the same pattern can be phonologically expressed in the same manner, independently of how it is “stretched” in its various phonetic implementations. In Figure 3.5, the pattern, also corresponding to the meaning of a closed question, is noted L*+HH% in the two sentences, even though their shape superficially differs. The syllabic position of the L* tone varies: it is on the antepenultimate syllable in “is he an invalid?” and on the penultimate in “is the answer invalid?” The span between L* and H is also greater in the former than it is in the latter. The two sentences belong to the same general category even though the phonological pattern L*+HH% is phonetically different. Since stress is lexical, that is part of the native knowledge of the language, the position of the L* tone is phonologically equivalent because it is anchored on syllables belonging to the same category (stressed syllables). The two tone phonological system ensures the linguistic functionality of intonation.

A similar example is presented by Hualde (2003) in Spanish, also a language with lexical stress. The sequence /nu.me.ro/ can be stressed on any of the three syllables, leading to three different meanings. When a pitch accent with the meaning of question is applied to the words, the contour is aligned differently with the syllables but the contour is phonologically the same. All three phrases belong to the same category:

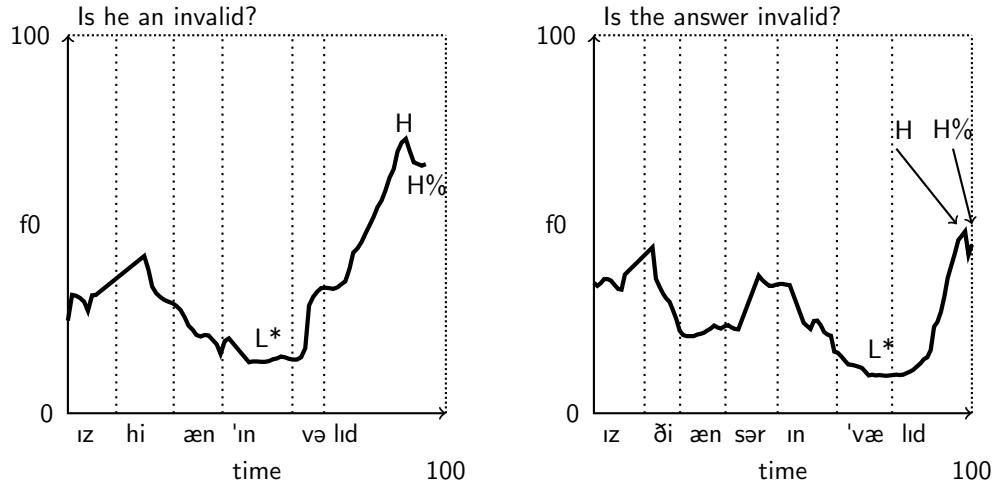


Figure 3.5: Two instances of the phonological pattern L*+HH% (closed questions) including the word *Invalid* in sentence final position. Time and F₀ have been normalized, syllable boundaries are marked with dotted lines.

/'nu.me.ro/	“number”	<u>número</u>
<i>número</i>	noun	H* L%
/nu.'me.ro/	“to number”	<u>numero</u>
<i>numero</i>	1st pers. sing. present	H* L%
/nu.me.'ro/	“to number”	<u>numeró</u>]
<i>numeró</i>	3rd pers. sing. preterit	H* L%

An other example from Hualde (2003) presents the reverse situation, where the same phrase is given a different meaning by a change in its intonation pattern: a) declarative and b) interrogative. The word *manera* is accentuated on its second syllable /ma.'ne.ra/ and the central tone of each pitch accent is associated with it:

a) declarative	b) interrogative
a mi manera] H* L%	a mi manera] L* H%

Overall, the AM theory assumes that between the intended communicative meaning and phonetic implementation, there exists a phonological level: tones and intonation units are organized in an autonomous and hierarchical system, a grammar, of which the AM model is a representation. This grammar ensures the intonation of an utterance is well-formed, categorizable and thus linguistically functional. According to this assumption, speakers of a language apply the tones of an intonation contour by fitting this contour over the hierarchically organized constituents of an utterance (*tune-to-text*). This has caused the model to be criticized for its *circularity*.

3.2.3 Synthetic vs. analytic

Circularity On the one hand, the AM model does not systematically address the phonetic implementation of the tones, pitch accents, and contours over a sentence, especially in articulatory terms. Phonetic implementation is conceived as interpolation between tonal targets either straight (Bruce, 1977) or sagging (Pierrehumbert, 1980) lines. But on the other hand, phonological units must be postulated from the observation of the phonetic surface form only. Phonology and phonetics seem to be co-extensive, without any external reference (see the *lack of reference problem* discussed in Xu, 2011). It has been noted by several authors that even among experts disagreement is common. Xu (2007) reported results by Wightman (2002) stating that “pitch accents, widely accepted as the basic units of intonation, have an inter-labeler consistency of no more than 50% even by experts performing repeated visual and

auditory examinations” (see also Hirst, 2005).

However, transcriber agreement for the detection of prominence and edge tones is fairly high according to the same results by Wightman: between 81 and 92%. The success rates drop in the categorization of the type of edge tones and pitch accent. The method does not fail since transcribers do locate the tones both by auditory and visual inspection of graphic representations. They discretize the phonetic continuum into target levels or tones. What is problematic is the organization of the sequence in smaller units: pitch accent and edge tones. There is a lack of consistency in the choice made to symbolically represent the pattern, not in the existence of these patterns. Again, vagueness is in the symbol not in the object. Intonation units such as pitch accent or edge tones are still extremely vague concepts, a vagueness that generates much research to better define these concepts. Vagueness is constitutive of linguistic categorization, and inter-transcriber uncertainty (or lack of consistency) is only to be expected. The intension and extension of the units proposed by the AM model might very well need adjustment. In the framework of categorization as presented in Chapter 2, the intension of a concept is abstracted from a complex network of loosely related features. Intension and extension are co-extensive: the application of the intension is vague and the limits of the extension fluctuate. Such is the case for intonation categories as well. The AM categories can never be clear-cut and inter-transcriber uncertainty reflects this vagueness, especially because, compared to phonemes, intonation concepts are extra vague in their lack of independent mode of symbolic representation and instantiation.

Because a phonological model of intonation seems prone to so much uncertainty or not useful enough for speech technology (Hirst, 2005), synthetic

approaches are avoiding any phonological speculation by encoding communicative functions directly onto (artificially modeled) phonetico-articulatory processes. By doing so, synthetic approaches treat each instance of intonation into a singular object of micro-prosodic sophistication and therefore render intonation uncategorizable, that is, communicatively null since only a phonological level ensures that phonetically different intonational units be categorized. Without a phonological level of analysis, intonation is excluded from the general principles of (linguistic) categorization. Nothing obviously supports the exclusion of intonation as categorizable instances of a prototype and members of categories unified under a concept (a prototype); even if this concept has to be evoked by an external symbolic category name. Furthermore, the consistency with which human beings produce and interpret intonation in spite of variation clearly indicates that categorization is at play in the use of intonation and that a phonology of intonation must exist (see works on tones by Earle (1975); Liberman et al. (1993), see also Faure (1973); Liberman & Pierrehumbert (1984)).

Furthermore, the AM model is not meant to be universal. Among the several disclaimers listed on the website dedicated to ToBI, it is clearly stated that the system must be adapted to the language to transcribe. Neither the AM model nor the ToBI system have been systematically applied to French, especially because other models have been suggested for this language in other theoretical frameworks (see Welby (2003) for references and Section 3.3 in this chapter). Welby (2006) discusses the categorization of two tonal movements in French within the AM categories: an “early phrase accent” (noted LH) and a late “pitch accent” (noted LH*) which have been suggested in earlier works by Jun & Fougeron (2000). As Welby notices, these two movements are not

functionally distinct but only structurally different, one marking the left edge of a phrase, the other marking the right edge of a phrase. They cannot be strictly categorized one as “pitch accent” and the other as “edge tone.”

[...] Intonational units vary in their properties across languages. Treating the French LH late rise as a pitch accent reflects the fact that this rise shares properties with pitch accents in other languages, but does not assert that it shares all those properties. Consider an example from animal taxonomy. Robins and penguins are two types of birds. Robins and most other birds can fly. Penguins can't fly, but they are like their fellow birds in many other respects (they have feathers and wings, build nests, and lay eggs). Penguins are a different kind of bird from robins, but they are still birds. And just as penguins are not robins, French pitch accents are not English pitch accents.

Remarkably, Welby makes a reference to the oft-cited case of birds as a vague concept. Not only are intonation units vague concepts in one language but even more so across languages. “H and L are phonological abstractions, comparable to phonemes” that are realized each time differently depending on the context (see quote from Ladd, 2008: at the top of this chapter). Accordingly, larger intonation units should be even more context-dependent since they convey a communicative function, local or global, in a sentence. Just as the two phonemic categories /i/ and /ɪ/ of English are allophonic in French, it cannot be expected that intonation units of one language are categorized similarly in another language. Both form and function vary within large categories. Just like phonemes, intonation units (tones and compound of tones) are graded categories with a prototypical center in the sense developed in Chapter 2. They

are a group of related objects (instances), more or less loosely related by a network of features. The prototype of an intonation category is not a typical instance or the best exemplar from the category but a conceptual center, subsuming the category organization. This prototype is abstracted from all the objects in the category, and all objects are instances of this prototype.

3.3 Three intonation contours of French

3.3.1 Stress and intonation in French

Stress Lexically, stress is not functionally distinctive in French (Beckman, 1986); there are no minimal pairs distinguished only by stress (or word accent) as it is the case in other languages (e.g., in English, 'in.va.lid vs in.'va.lid). Stress systematically falls on the last syllable of French words. Because of the status of stress in French, the existence of a pitch accent associated with it is not a consensus among linguists. Even for functions such as focus marking, French naturally resorts to grammatical constructions (cleft, dislocation) rather than accentuation (*'Paul did it, not Mary* vs. *C'est Paul qui l'a fait, pas Marie*). As noticed by Welby (2006), it is preferable to limit the application of tones and tonal accents to a demarcative function at the phrasal level (see also Fónagy, 1979; Di Cristo, 2000; Jun & Fougeron, 2002).

Intonational accent There is a consensus on the existence of two main intonational phrasal accents in French, marked by one tone or a combination of two tones. The primary accent is obligatory and associated with the right edge of a phrase, with the final full syllable (excluding a schwa) of a prosodic phrase: if the phrase is not the last of an utterance, an F_0 rise occurs on the last syllable(s) of the phrase. This accent is usually accompanied by the lengthening of

the accented syllable. The secondary accent is optional and loosely associated with the left edge of a phrase, on one of the first syllables. Together, the accents form what Fónagy (1979) termed an *arc accentuel* (“accentual arc”), which prosodically groups the constitutive parts of phrases. The size of these arcs has been studied by Jun & Fougeron (2002). They found that the average size of an arc depends on the type of words, phrases and segmental content but generally corresponds to three or four syllables with a limit around seven syllables before the group gets broken in two (3-4 or 4-3). In their model, each intermediary intonation constituent (*accentual phrase* (AC)) is composed of an initial secondary accent (LHi) and a final obligatory accent (LH*). In the figure below (Figure 3.6), the arcs have been superimposed onto the original graph from Jun & Fougeron (2002)

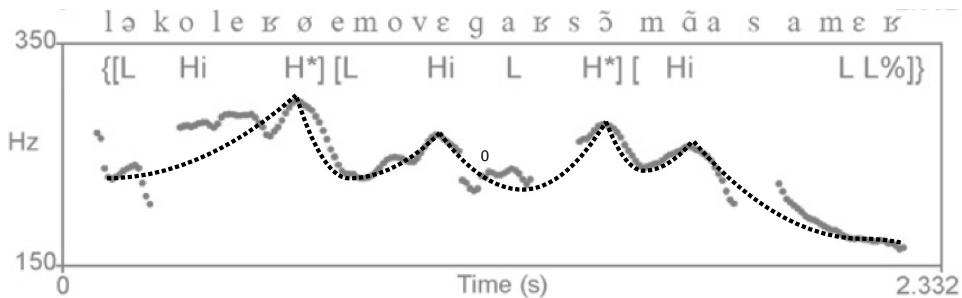


Figure 3.6: [LHiLH*] accentual phrases (AC) and *arc accentuels*.
Adapted from Jun & Fougeron (2002)

As mentioned in the previous example, research on French intonation has suggested a hierarchical structure over the segmental level. The Instint model developed by Hirst & Di Cristo (1998) (see also Di Cristo, 1999, 2000; Di Cristo et al., 2000) and the model of Jun & Fougeron (2000) assume an intermediate tier of phonological constituents between the tier of the intonation phrase and that of the tones: Tonal Units (TU) for the former, Accentual

Phrase (AC) for the latter. They both roughly correspond to a phonological phrase in the hierarchical model of the AM framework presented by Figure 3.4, p.55. These models, although they differ in many aspects from the AM model, have in common with it the ability to analyze intonation in terms of features (tones) within a metrical grid. More importantly for this work, these approaches, and especially those with an automated modeling system use a bottom-up approach that derives phonological units from the observation of the phonetic information, whether it is the model developed by Hirst and Di Cristo (INSTINT and Momel) or the Prosogram model by Alessandro and Mertens(Alessandro & Mertens, 1995; Mertens, 2004, 2012).

3.3.2 Three contours

Closed questions The first contour to be examined is that of unmarked closed question (also called “polar question,” “yes/no question,” or “declarative question”). In standard metropolitan French, closed questions can be indicated by (1) syntactical inversion of the subject-verb order (SV→VS), (2) a morpheme placed before a declarative sentence (*est-ce que*), or (3) intonation alone.

Declarative form:

Tu vas garder les tickets. (You are going to keep the tickets.)

Question forms:

- (1) inversion: ***Vas-tu garder les tickets ?*** (Are you going to keep the tickets?)
- (2) morpheme: ***Est-ce que tu vas garder les tickets ?*** (Are you going to keep the tickets?)

(3) intonation:  Tu vas garder les tickets ? (Are you going to keep the tickets?)

The representation in (3) is adapted from Delattre (1966). He devised a 4-level scale on which he described ten contours of French from the most ascending to the most descending. Closed question is the most ascending contour, rising from 2 to 4 (3-1 in Pike's notation).

This usage of intonation is preponderant and it is taught in French language classrooms, along with its morphological and syntactic counterparts. It is usually assumed that the three forms have a stylistic, if not social, value attached to them, inversion being the most formal and intonation the least formal. Martinet (1960) wrote: “En français, par exemple, il est fréquent que le caractère interrogatif de l'énoncé ne soit marqué que par la montée mélodique de la voix sur le dernier mot. On distingue fort bien ainsi entre l'affirmation *il pleut* et l'interrogation *il pleut ?* Ce dernier est l'équivalent de *est-ce qu'il pleut ?*³” In a survey of French *boulevard* plays, characteristic for their use of everyday language, Terry (1967) found that “of the total yes-no questions (3,016) only 97 were formed with *est-ce que* (3.22%), while 339 used inversion (11.24%) and 2,580 used simple change in terminal intonation (85.54%).”

Because of the pragmatic frequency of closed questions in language, their intonation contour is easy to elicit. The contour can also carry additional meanings such as disbelief or irony. The intonation of a sentence can

³In French, [...] it is frequent that the interrogative meaning of a sentence is only marked by a melodic bitonal rise of the voice on the last word. Thus, the distinction between the declaration *il pleut* (It is raining) and the question *il pleut ?* (is it raining?) is perfectly clear. The latter is the equivalent of *est-ce qu'il pleut ?* (is it raining?) (Translation mine. N.B.)

be marked for these meanings in various combinations. However, assertion and question differ from doubt and irony in the sense that a variation in the degree of realization of the contour does not mean a variation in the degree of their meaning. One can be more or less ironic but one cannot ask a question more or less. Finally, the use of unmarked closed questions ensures that the output of the automated system can be controlled: while new findings are expected, a radically different contour would lead to question parts or all of the methodology.

Using Pierrehumbert’s notation system, the contour, characterized by a plateau and a final bitonal rise, is noted L% L+H* H% (Beyssade et al., 2007). The high tone of the closed question contour is the highest of the sentence (H*) and is also usually merged with the last high boundary tone (H%). The low tone of the terminal rise is associated with the boundary of the penultimate and ultimate syllable. The bitonal rise takes place in the last syllable of the utterance, sometimes starting in the penultimate syllable (Faure, 1973). In their study on French interrogative sentences, Fónagy & Bérard (1973) noted that “[among] sentences identified as questions in more than 80% of the cases, one salient feature is noticeable: a continuous rise of the pitch from the beginning to the end of the last syllable” (Translation mine. N.B.)⁴. The height of this rise varies from subject to subject as does F₀ range.

Disbelief: two modalities of the unmarked closed question A speaker of French can express her disbelief concerning what she has just heard by using intonation alone. Disbelief can be intonationally expressed in mainly two

⁴ “[parmi] les phrases identifiées dans plus de 80% des cas comme interrogatives, on peut relever un trait saillant: une montée continue du début jusqu’à la fin de la dernière syllabe”

forms, characterized by Fónagy & Bérard (1973) as *question étonnée* (“astonished” or “surprised question”) and *question incrédule* (“doubtful” or “incredulous question”). In writing, both intonation contours would be indicated by the joint use of an exclamation mark with a question mark: *really, you had dinner with the Queen !?*

In the case of a *question étonnée* (*surprise* hereafter), the subject expresses her disbelief because of the unexpectedness of what has just been said. Something she thought to be impossible has happened. It is an emotional response. The shape of the contour is the same, but the ratio F_0/time (the velocity) of the final rise is greater for surprise than for unmarked question. There can be an imperceptible fall after the rise, and the overall amplitude of the F_0 range is greater than for neutral questions as well.

In the case of a *question incrédule* (*doubt* hereafter), the subject expresses her doubt relatively to the propositional content of what has just been said. The contour is characterized by a triangular shape over the penultimate syllable. The rise is usually larger than the fall. The F_0 peak is associated with the penultimate syllable but the rise and fall can overlap with the adjacent syllables.

These two modalities are not clear cut categories. During a preliminary study for this dissertation, an elicitation task was given to speakers, asking them to express disbelief in the utterances. The speakers produced both types of contour, surprise and doubt. Only the contour of doubt had been expected. The choice of a modality over the other depended on the subjects’ interpretation of the context as being whether more doubtful or unexpected.

In all three contour types, there can be a secondary peak on a syllable before the antepenultimate, depending on whether the contour was realized

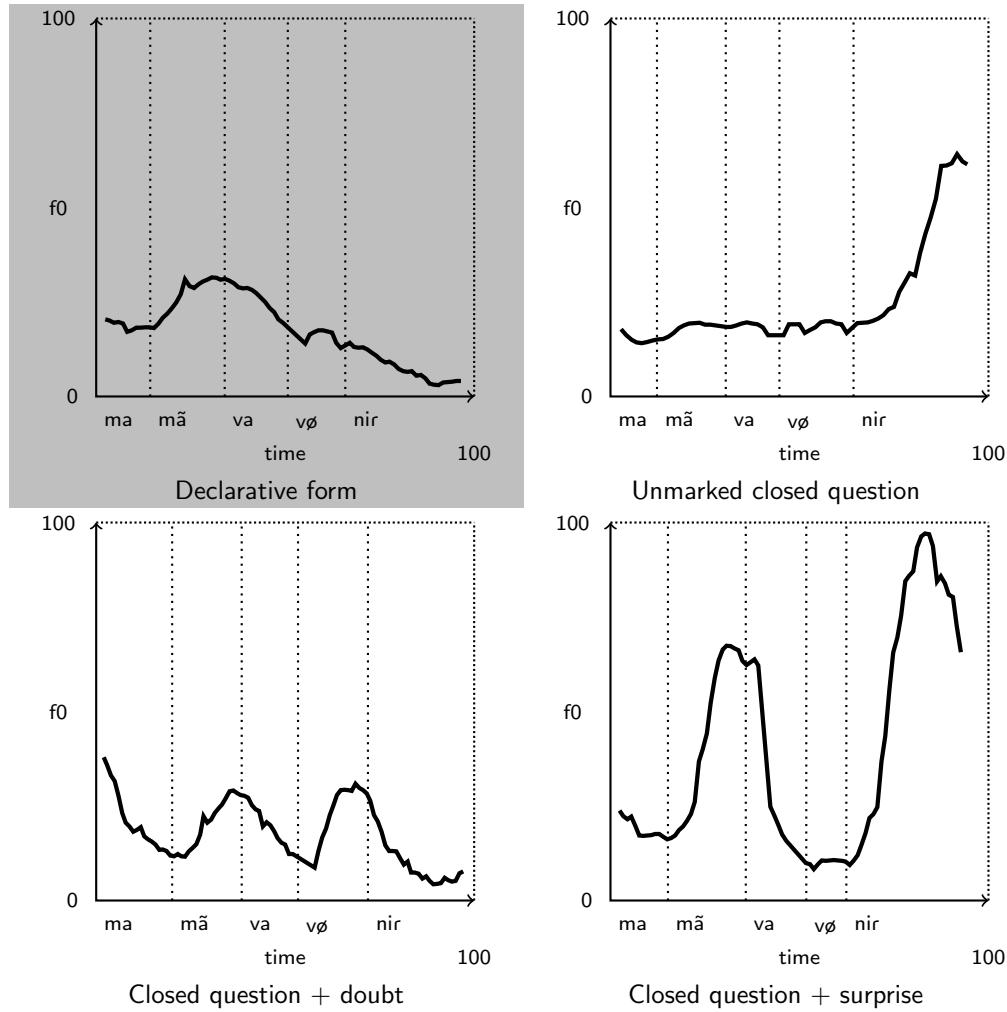


Figure 3.7: Four intonation contours applied on the same sentence *Maman va venir*. Time and F_0 have been normalized, syllable boundaries are marked with dotted lines.

as one or two accentual arcs. The three contours are illustrated on Figure 3.7 using the earlier example of *Maman va venir* (“Mom is going to come”).

It can be argued that the three contours have a family resemblance and therefore constitute a general category because of their relation of meaning

and form. Intensionally, the three contours are related by a general meaning of interrogation. Extensionally, they are related by the presence of salient rise on the end of the utterance. Corpora of these three contours are the material used for the implementation of the model developed in this work.

3.4 Automated Pattern Recognition for Intonation: PRInt

Fonagy called his own study of the modalities of question *une analyse vectorielle* (a vectorial analysis) because he conceived contours as patterns distorted by production but perceptually categorizable as the intended patterns. A vectorial approach is the foundation of the model presented in this study, which takes Pierrehumbert's words to the letter:

A theory framed in terms of target levels is attractive because it affords good facilities for describing how the same intonation pattern lives up with different texts; the crucial points in the contour, the F_0 targets, can be lined up with crucial points in the text, with stretches in between computed accordingly. The behavior of a given contour under changes in pitch range can be modeled in a similar fashion, by transforming the target points. (Pierrehumbert, 1980)

In these lines, Pierrehumbert characterizes intonation as patterns or feature vectors (a series of structurally organized tones) that undergo distortion to fit any context, phrasal or sentential. Conversely, and an assumption on which the AFP is based, the prototypical pattern of an intonation contour can be abstracted from the observation of a group of instances (distorted implementations of a single pattern or variations) belonging to the category defined by

the prototype. In line with the AM model, but also the model of Hirst and Di Cristo, (INSTINT, Momel), or the model of Piet Mertens (Prosogram), PRInt is a bottom-up approach that seeks to “deduce a system of phonological representation from observed features of F_0 contours” (Pierrehumbert, 1980).

3.4.1 Presentation of the model PRInt

The model developed in this work (PRInt, hereafter) is a system of analytical pattern recognition for intonation contours. It comprises two modules. The first module, the Automated Tonal Labeling Module (ATLM) extracts features by applying linguistic fuzzy quantifiers (high-low for F_0 , before-after for metrical alignment). The second module, the Automated Fuzzy Classifier (AFC), organizes the features by grade of membership to their category.

3.4.1.1 Feature extraction - ATLM

The ATLM normalizes both time (sentences and syllables) and F_0 range locally for each individual sentence, towards the extreme values of each utterance independently (maximum and minimum in time and F_0). Initially, the value of tones (H and L) and their alignment to syllables is evaluated strictly against internal structural relation, not comparing to any outside values. The pitch range of individual speakers is also analyzed but does not enter in the computation of sentence tonal structure.

The ATLM does not assume a pre-established hierarchical structure of larger constituents; it does not resort to an intermediate tier between tones and the intonational phrase. The model is strict in its iterative use of only two relative tones (L and H) at two levels of analysis.

First, the ATLM uses a metrical grid divided into “windows” (half-

syllables) to systematically locate L and H tones at the level of the syllable and to discretize the continuous “physical and phonetic levels” into an intermediate level of micro-prosodic and pre-phonological units (Di Cristo, 1999). The value of syllabic tones is calculated purely as a F_0 difference between the maximum (H) and minimum (L) of the time window. The syllabic tones are called the *pre-tones* because they are used in the subsequent identification of actual intonational tones.

Second, the ATLM sequentially and recursively locates three tonal compounds of the form L-H-L (hyphens only indicate that the three tones are in the same compound). The tonal compounds can be complete or incomplete. Incomplete tonal compounds can have the structure L-H or H-L only. The first tonal compound comprises the highest H of all syllabic H's and two associated L's (the lowest points immediately before (L-) and after (-L) the H tone). Then, the ATLM locates other compounds to the right and to the left of the highest one. The ATLM can be set to iteratively find as many compounds as desired but the utterances contained in the corpora used for this work have maximally 3 compounds, most have two or one only. The decision of using seven-syllable utterances is based on the limit of an *arc accentuel*.

The ATLM joins the tonal compounds to create the tonal pattern of the sentence: ordered position of the tones in the metrical grid and relative F_0 position. This constitutes the macro-prosodic level of discrete phonological units (Di Cristo, 1999). The patterns of all sentences are stored together and PRInt passes them on to the next module.

3.4.1.2 Classification - AFC

The second module of PRInt is the Automated Fuzzy Classifier (AFC). The classifier is *fuzzy* because it relies on the general principles of fuzzy set theory to assign a grade of membership to the features of each contour in a corpus. It computes the degree of typicality of the instances of features in their category. The degree of typicality of an object in a category (an instance of a feature) is computed twice by different methods: frequency (mode) and similarity (median). When all features of all utterances in a category have been assigned a grade of membership, the data in the corpus has been *fuzzified*, or organized as a structured category by level of typicality from 0.1 to 1. The AFC can then *defuzzify* the sets: it computes the central tendency in the data as an average of all values weighted by their grade of membership/typicality (method of centroid). By doing so, the AFC extracts the prototypical contour of the category, abstracted from all contours and of which each contour in the category is an instance. The AFC uses the extension of an intonation contour (all objects categorized as such) to extract its prototypical (intensional) centroid.

In parallel, the AFC also ranks the variation of the outer dimensions of the instances. It can provide the degree of typicality of gradient parameters that have been abstracted away for the pattern recognition of the phonological structure. The PRInt system takes into account the variation of the phonetic implementation of the contours and organizes them by degree of typicality as well.

The parallel fuzzification of the phonological structure and its phonetic variation makes it possible to compare contours and to determine their phonological (intensional) and/or phonetic (extensional) status.

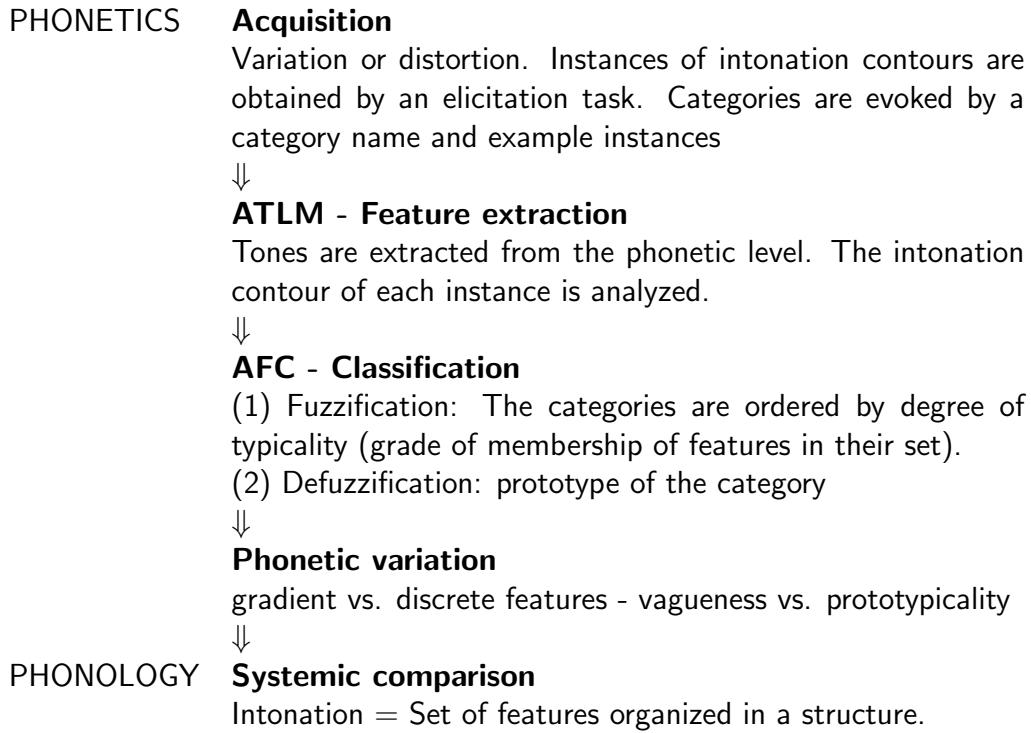


Figure 3.8: Organization of the study of intonation with PRInt

By assuming that intonation has a phonological level of representation, contours are made analyzable into structurally ordered feature vectors. This position places intonation into the more general and unified principles of categorization, thus explaining linguistic variation and gradience in terms of vagueness and prototypicality. Intonation is more vague than other linguistic categories since the category name of intonation units can only be evoked by a symbol taken out of the category (by a phrase) or by an instance of the category itself.

Intonation units (tones and larger constituents) are vague concepts. In phonological studies, they are similar to feature vector or patterns grouped into

categories by family resemblance. The PRInt system relies on two modules to extract the prototypes of intonation categories from a corpus of instances. Before turning to the PRInt modules themselves, the next two chapters introduce the basic principles on which each module is based. The first module, the ATLM, is a pattern recognition application. Chapter 4 is an overview the principles of pattern recognition. The second module, the AFC, is a fuzzy classifier. Chapter 5 is a presentation of the principles of fuzzy set theory, with examples applied to linguistics. PRInt combines pattern recognition and fuzzy set theory, which are technical approaches to the concepts of categorization, vagueness, and prototypicality.

Chapter 4

Pattern Recognition

The ease with which we recognize a face, understand spoken words, read handwritten characters, identify our car keys in our pocket by feel, and decide whether an apple is ripe by its smell belies the astoundingly complex processes that underlie these acts of pattern recognition. Pattern recognition - the act of taking in raw data and taking an action based on the “category” of pattern - has been crucial for our survival, and over the past tens of millions of years we have evolved sophisticated neural and cognitive systems for such tasks (Duda, 2001).

Pattern recognition refers mainly to two related concepts. The first one is the human natural ability to discretize one’s surroundings into distinct entities and to classify these entities. The second one is the process of imitating this ability artificially to integrate it to automated systems.

Human pattern recognition The first process is that described by Duda, in the quote opening this chapter. It is the human cognitive capability to distinguish objects (physical or conceptual) from each other and from their environment and to categorize them according to some classification process, or in a more technical way, to identify the membership class of any given object

(Nadler & Smith, 1993). It is “crucial for our survival” since an inappropriate categorization can be problematic at least and fatal at most. Mushroom pickers can suffer or even die for not accurately distinguishing one species from another. In a more trivial way, it is not necessarily obvious to categorize objects in the category of “seats”, as presented in Figure 4.1 since the category contains many objects whose shapes may have very little in common. Most



Figure 4.1: A subset of the “seat” category: 0) chair, 1) meditation pillow, 2) Sori Yanagi Butterfly stool, 3) Eames chair, 4) Jean Prouvé armchair, 5) Chippendale armchair

human beings would recognize these six objects as “seats” in spite of their seemingly unrelated aspects. What makes one classify these objects into a single large class might well be their primary function as “object to sit on,”

rather than their shape. Even the two most chair-like objects ③ and ⑤ are not necessarily what one might think of a prototypical chair, made of a flat plane on top of four posts and to which a back of some sort is attached, something one could characterize as a “kitchen chair”, such as image ①. Indeed, most “objects to sit on” are also categorized for a secondary function or location: meditation ①, working (② and ③), resting ④, living-room ⑤, office ③, etc. Beyond man-designed objects, many a flat stone or a log has become an improvised seat. Minimally, the only physical feature required by the primary function is that the object have a somewhat flat surface on which to place one’s bottom. Each of the six objects have that flat surface in a shape or another (see next paragraph on *distortion*). Secondary functions require more features: a back to rest in the armchair, a low seating for a meditation pillow, etc. Therefore, in many ways, pattern recognition is related to categorization, it is the always active categorization of our surroundings to make them usable for us. The set of objects of Figure 4.1 could be subdivided into subsets: pillows ①, stools ②, chairs (①, ③), and armchairs (④, ⑤). As pointed out by Duda and illustrated by the simple category of “seats,” the natural ability to identify object to make use of them is an extremely sophisticated process so pervasive in our life that we never notice it, except in case of categorial uncertainty. Is ④ more of a chair or an armchair? It shares some physical features of both sub-categories (the posts and the arms). Do armchairs require posts (④)? Form and function are co-extensive and it is would be necessary to also divide between objects that have been specifically designed to be sat on and those that can be used to sit on.

Distortion Distortion is a crucial issue for pattern recognition. It is the materialization of the concept of vagueness. Patterns, or objects belonging to

a single class or category, are variations of one another rather than variations from a single exemplar that would serve as the class “model” from which the variation of the others can be calculated. There is no ideal or nominal seat in the category of “seats”. Rather, the ideal seat is an abstract and plastic construct made of all the various features of all seats ever encountered by a subject in her life. Experience plays a central role in the construction of classes and in the way human beings classify objects into classes. Classes are fuzzy in that some objects seem to be more central to the class, closer to an abstract ideal than others. Objects in a class do not all have the same degree of typicality. The kitchen chair seems to be a more central and more frequently cited sit-able object (Rosch, 1978) than a meditation pillow. For the pillow to gain consideration in someone’s category, it has to be part of that person’s surroundings and activities so that the pillow is associated with the shape-function co-extensive relation discussed earlier. The pillow is a sit-able object inasmuch as the relation is activated in one’s mind. An object from one class can be used as an object from another class because some of its defining properties (intension) makes it part of the extension of another category: one can sit on a coffee table since its shape enables the sitting function. Thus, frequency of exposure and similarity are fundamental to class formation and identification of individual objects, two co-dependent processes. Figure 4.3 illustrates how our categories influence the identification of objects. The ability to recognize a pattern in spite of its distortion is so ingrained in human activities that it naturally finds its place in cultural creation, especially in art, for which deviation from the nominal object becomes a creative process. In music, it takes the form of variations on a theme, a genre by itself: *Goldberg Variations* by Bach, *Diabelli Variations* by Beethoven, *Variations Symphoniques* by César Franck, etc. A theme is introduced and the composer displays his technique by stretching the



Figure 4.2: Musical variations or quotations are a creative use of the human pattern recognition ability. **Top:** Wagner’s opening theme to *Tristan und Isolde*. **Bottom:** Debussy’s quotation/variation on the theme in his *Children’s Corner*

theme to the limits of recognition. Beethoven went as far as finally transforming Diabelli’s waltz into a minuet. Musical quotation also refers to a part from another piece, generally from a different composer. Debussy mockingly quoted the opening theme from Wagner’s *Tristan und Isolde* in the playful *Golliwog’s cake-walk* of his *Children’s Corner*. From surrounding physical objects, to abstract concepts and to cultural creations, pattern recognition is a natural and pervasive process in all human activities, and especially language.

Bybee (2007, 2010) cites instances of the individual and/or social nature of category membership in natural language and how, over time, new exemplars to a category modify the prior organization of this category. Bybee discussed the usage of apical [r] and dorsal [R] among a group of speakers of

The object in the center of the picture is an instance of a mass produced object. Almost anyone who grew up in the USA is able to recognize it, even without knowing how to read. This object is encountered extremely (too?) frequently. Each instance of this object is an exact copy of any other, without any distortion at all. The object at the top of the picture is a unique object since it has been handcrafted, but it unmistakably belongs to the class of spoon by its high degree of similarity with other members of the category, although its categorical features (material, shape, size) are realized with a clear distortion from those of more central instances of the category. The categorial membership of the object at the bottom right of the picture is uncertain. Its overall shape makes it an uncanny shaving brush but individual features such as the bristles and the handle are rather odd. One could attribute a brushing or whisking function to this object, based on its resemblance with other objects. The highly specific and categorizing function of the object is to whisk powdered green tea in the Japanese tea ceremony. The last object is a unique object. It is a piece from a painting frame in the shape of an angel's face. Although anyone could interpret the form figuratively as a face, its function would be impossible to retrieve for almost anyone without knowing the history of this object.



Figure 4.3: Identification/categorization of objects: distortion, frequency, and similarity

French in Montreal. The distribution between “variable” and “categorical” (allophonic or phonemically distant) usage of the two phones varied between speakers and over time for individuals. Phonemes are ranges of formant compounds and not fixed frequency measures. There is a large span of variation between neutralization of contrast, allophonic overlap, and maximum contrast. Bybee indicated that the change in phonemic categorization “may be due to the particular social situation the person is in” and/or to “individual differences” (idiolectal). Independently from what triggers it in each given category, categorical plasticity is a pervasive phenomenon and binary class membership is more the exception than the rule in natural classes or languages.

The graph of American vowels by Peterson & Barney (1952) (Figure

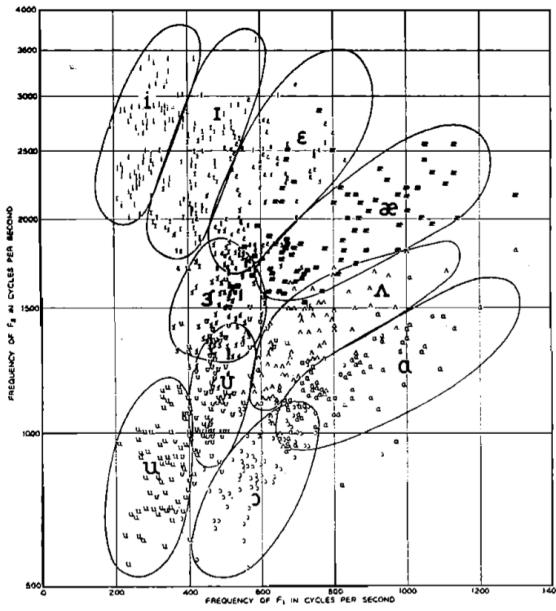


Figure 4.4: Frequency of second formant versus frequency of first formant for ten vowels by 76 speakers, Peterson & Barney (1952)

4.4) shows how a phoneme does not exist per se as a singular nominal unit but as a group of instances. These instances are variations of each other but grouped by proximity and frequency in a cluster that has a more densely represented central region. A zone of dispersion extends from the dense center with more and more distortion from central values as the distance from this center increases. Borderline instances are in two phonemic zones simultaneously, making them more difficult to categorize without recourse to the context. In terms of category membership, this implies that a borderline instance of a phoneme belongs to two categories but may be more strongly associated with one category over the other.

It will be shown in the next chapter that distortion applies to intonation

contours as well, in a manner akin to musical variation or phonemic dispersion.

Automated pattern recognition (PR) Emanating from human pattern recognition, Pattern Recognition (or PR) as a field of research is the attempt to achieve artificially what humans do naturally by developing computational systems “that imitate in some way human sensory processes” (Nadler & Smith, 1993). Today, machines capable of pattern recognition are found everywhere, with different levels of accuracy: face recognition module on a camera, optical character recognition (OCR) for scanner, DNA sequencing, speech processing, factory sorting, weather forecasting, blood cell counting, cancer detection, etc. (Friedman & Kandel, 1999). Among these applications, speech processing is crucial to the domain of human-machine communication. Oral language is by far the most efficient and powerful way of communication among humans. It is efficient in terms of speed compared to gesturing, typing, or reading (Fink, 2008). Thus, speech pattern recognition modules (in the form of software) are found more and more on computers and even phones to allow users to communicate with the machine by voice rather than by typing on an interface. The effectiveness of these applications is variable, depending on the choices of techniques and desired goals. Figure 4.5 illustrates a commercial speech recognition software that only tolerates so much distortion. A native French speaker is not interpreted as well as a native American speaker.

To create a system capable of categorizing objects as “seats” amounts to make a system capable of deciding of the sit-ability of any given object. This leads to a ranking issue, whether it is decided that sit-ability is a binary (sit-able or not) or gradient feature (object being more or less sit-able). In order to artificially recreate a particular process of pattern recognition and

The dictation softwares implemented on modern computer software and phones have issues if the user does not natively speak the language set in the application. As an experiment, a native speaker of American English and a native speaker of French both read the first few lines of James Joyce's Ulysses into a dictation software. Comparing what the software recognized of the French speaker's pronunciation to the original text, the software apparently matches the pronunciation and prosody of both speakers to some sort of templates of American English. However, it is very sensitive to distortion, as its recognition of the American speaker's pronunciation suggests.

James Joyce Stately, plump Buck Mulligan came from the stairhead, bearing a bowl of lather on which a mirror and a razor lay crossed. A yellow dressing gown, ungirdled, was sustained gently-behind him by the mild morning air. He held the bowl aloft and intoned: – Introibo ad altare Dei.

French speaker Sticky bun back Monegan came from this to head bearing both leather on reaching mirror and originally crossed that you don't dressing gun and girl that was assisting gently be hanging in my morning you had the bold enough and doing: in queen who at the AAB

American speaker Stately plump Buck Mulligan came from Mr. head buried in a bowl of letter from which American recently crossed the yellow dressing gown undergrowth was sustained gently behind him on the mild morning there he held the ball a lot and intones: enjoyable I don't daddy.

Figure 4.5: Using the dictation function of a modern computer.

determine whether it is a binary or gradient process, the process should be studied among humans and then programmed into machines. For example, robotics attempt to recreate natural motion, such as walking, by observing how this motion in nature. Unlike mechanical processes,

perception is something everyone experiences but no one really understands. Introspection has not proved as helpful in discovering the nature of perception as one might hope, apparently because most everyday perceptual processes are carried out below the conscious level. Paradoxically, we are all expert at perception, but none of us knows much about it. (Duda, 1973)

The development of any pattern recognition system depends on the end goal it must achieve. As is the case for much research in cognitive science, phonology included, the processes or structures that one is looking for are not directly observable; they are below the conscious level and must be inferred from

their output at the conscious level. Thus, to design a pattern recognition system, one must first analyze the output of the human process of pattern recognition, how objects have been identified as classes of membership or categories. Then one can devise a series of methods to form an automated system capable of achieving the same categorization.

4.1 Pattern Recognition (PR)

The general sequence of pattern recognition consists of three main phases (Duda, 2001; Nadler & Smith, 1993; Pal & Mitra, 2004):

1. **Acquisition** Data are acquired with a (set of) sensor(s), adapted to the nature of the task: camera, microphone, thermometer, probe, etc. Usually, the analog data is converted into digital data, whether within the sensor or in the computer to which it is attached. For intonation patterns, the data is acquired with microphones which convert the analog acoustic pressure waves into an electronic digital signal that is passed to software on a computer (Praat in this work).
2. **Feature extraction** “Features are functions of the measurements performed on a class of objects that enable that class to be distinguished from other classes in the same general category” (Nadler & Smith, 1993). Features are the traits or attributes distinguishing objects from each other in a group of otherwise identical objects, such as the size or the color in a series of otherwise identical shirts. Thus, a feature has an associated value, whether numerical (size in centimeters) or discrete (S/M/L/XL, blue/red/yellow). To measure these values, data must be prepared. A scale or space of reference toward which each object in a

category can be evaluated must be created (grid of size in centimeter, hue chart).

Segmentation, or low-level feature extraction, is the process that discretizes a continuous space into distinct constituents. These can be the letters of a word, the region of a satellite view, or the phonemes in a string of words. Most of the time though, low-level features are not descriptive but numerical. An initial quantization converts the continuous data into a finite set of data (the *pattern space*), in which quantifiable and interpretable measurements can be made. The choice of the quantization function influences the degree of resolution or precision in the output¹. For example, an image can be divided into a grid so that its subparts (localization, size, surface, etc.) can be measured. Thus, the large set of raw numbers of the data can be transformed into (finite) sets of organized numbers corresponding to the various parts of the pattern. Feature extraction (high-level features) consists of selecting which of the low-level features are actually necessary and which are redundant in the task of recognizing the pattern. Imagine that there are several geometric shapes drawn on a piece of paper. Low-level features are the series of all the contiguous points making the lines of the contours of these figures. High-level features will be the set of vertices characterizing each figures (e.g. the four corners of a square), with their localization on the plane. Thus, feature extraction reduces the complexity of the pattern space and makes features simpler to process by the classifier.

3. Classification Once all objects have been transformed into a *feature*

¹For example, the size of pixels on a given screen: the smaller the size of the pixels, the greater their number, the more precise the image, but also the heavier the processing and the data load.

vector, “a set of features arranged in an ordered set” (Nadler & Smith, 1993), such as the 4 coordinates of the vertices of a square, their data are passed on to a classifier. The classifier’s task is to analyze statistically or structurally how the features are organized among the objects in the set and then to classify the objects into categories based on the found organization of features. For example, if there are triangles, rectangles, and squares on the aforementioned piece of paper, the classifier will create a category for triangles, based on their number of vertices (3 vs. 4 for the other shapes). Then, the classifier would have to distinguish rectangles from squares for having two opposite sides longer than the others.

4.2 Statistical and structural approaches to pattern recognition

There are two general approaches to feature extraction that are important to distinguish. In the statical approach, the order of the feature in a vector is arbitrary with regard to the structure of the object. This structure is not crucial to the pattern recognition process, and this approach is purely numerical. Conversely, the premise of the structural approach is that the structure of the pattern is central to its recognition.

In the structural approach, the features are ordered in a way that reflect the organization and the relations of the parts of the object’s structure. Structural features are ordered relationally and hierarchically. Structural extraction has to take into account this order in the way it is processed. Structural features may have a discrete number of values (small/tall, very small/small/medium/tall/very tall, left/center/right, etc.) or they may be attached to numerical values, in which case they are continuous or graded along some sort

of continuum (size in inches, position in degrees, minutes, seconds).

These two approaches are not exclusive and they can be integrated into a single system of pattern recognition, as is the case for the one developed in this work. The statistical approach is used for the low-level features and the structural approach is used to extract high-level features, based on the values of the low-level features.

The creation of a system of pattern recognition involves the development of ad-hoc techniques specifically for the type of patterns for which it is designed. “As with segmentation, the task of feature extraction is much more problem- and domain-dependent than is classification proper, and thus requires knowledge of the domain” (Duda, 2001). As presented in chapter 7, the segmentation and extraction processes designed specifically for intonation contours are based on prior knowledge and structural assumptions. The classification process presented in chapter 8 is also domain-dependent but to a lesser extent.

4.3 Looking for haystacks in Monet’s painting

To illustrate how a pattern recognition system works, images are typically better examples, especially on printed paper. A series of paintings by Claude Monet constitutes the set of object to classify by the fictitious system presented in this section.

In the 1890’s, Claude Monet undertook painting a series of haystacks at various times of day and year, and within various framing. This work is reminiscent of his own series on the Cathedral of Rouen or the *Thirty-six views of Mount Fuji* by Hokusai. Figure 4.6 is a subset of 12 of these haystack



Figure 4.6: 12 instances of *Les meules* by Claude Monet arranged sequentially, according to the number of stacks and their relative distance

paintings. The number of stacks, their size, their relative position, and the background (seasonal mostly) change from painting to painting. The twelve canvases can be classified in this manner:

1 stack = {07, 08, 09, 10}

2 stacks = {01, 02, 03, 04, 05, 06, 11, 12}

Position of the smaller stack relative to the larger stack

left = {01, 02, 03, 04, 05, 06}

right = {11, 12}

The distance between the two stacks decreases from canvas 01 to 06

and increases from canvas 11 to 12.

Although it is a rather philistine statement, the landscape behind the stacks and the assumed season or moment of the day are irrelevant to the classification of the canvas in terms of haystacks, they might even constitute some *noise* to be filtered out before processing the haystacks patterns. For the sake of the example, let's assume that the pattern recognition system is able to separate the contour of the stacks (foreground) from the rest of the canvas (background). Its remaining task is to find and classify the patterns of stacks in the set of paintings. The process of pattern recognition applied to each canvas is illustrated for canvas 02, as shown on Figure 4.7.

Segmentation The continuous space of canvas 02 is discretized (see figure 4.7a). Similarly, all canvases are scaled to a square plane of $60 \times 60 = 360$ points. This process of scalar quantization normalizes the varying sizes of canvases to the same relative size and the same number of 360 discrete partitions (analogous to large size pixels on a screen). The contour of the haystacks is consequently discretized and segmentation generates the initial information about the haystacks by extracting their constitutive subset of a few dozen points out of the 360 partition of the normalized canvas. The position of each

constitutive points is a coordinate pair on the plane. The pattern recognition system could compare the haystacks' contours among canvases as their subsets of points but that would not lead to meaningful results in terms of patterns being organized into a category. Since these points are coordinates indicating a distance from a fixed origin $(0, 0)$, they pertain to purely numerical low-level feature vectors. More specifically, on a small number of instances, a pattern recognition system would not find enough (and maybe no) recurring points among canvases to identify a pattern. Segmentation prepares the canvas for high-level feature extraction.

Feature Extraction High-level features are a subset of the discretized points forming the haystacks' contours. In the proposed categorization, canvases are distributed first by their number of stacks. This should be a non-problematic issue for a pattern recognition system since it can simply count the occurrences of haystacks per canvas. More difficult is the characterization of the canvas with two haystacks, one being smaller than the other and being left or right of the larger stack. One solution is to determine the position of each haystack relative to the horizontal length of the frame. If the canvas is scanned left to right, it could be divided into two symmetrical frames, to the left and and to the right of a central line separating the canvas in two (Figure 4.7b). It would work well for canvases ① to ④, and probably ② but not for ⑤, ⑥, and ⑦, for which both stacks are on one side of the canvas. That is why a structural approach is preferable. It integrates human-like linguistic quantifiers and works sequentially in a hierarchical order that leads to meaningful high-level feature vectors. Ideally, a robust pattern recognition system would include the following steps.

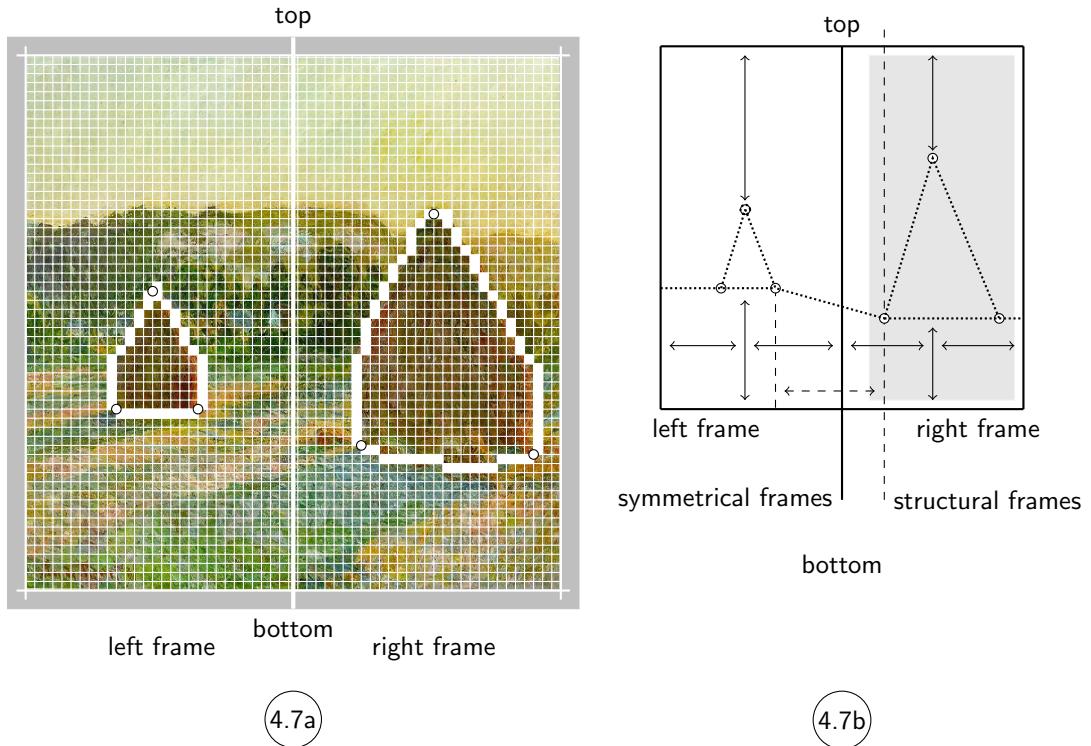


Figure 4.7: Feature extraction. **Fig.4.7a:** normalization, scalar quantization, and discretization of canvas (2) from Figure 4.6. **Fig.4.7b:** Numerical and structural feature extraction. The feature vector of the large stack is located and extracted first. The feature vector of the small stack is located relative to that of the large stack.

1. The pattern recognition system would locate the highest points among all coordinate values of the constitutive points of all haystacks on a canvas. This is the top of the highest haystack.
2. From this point, the pattern recognition system would locate the lowest point to the right of and the lowest point to the left; these points correspond to the extremities of the base of the haystack. The contour is

simplified to a feature vector of 3 points, as in Figure 4.7b. The position of the three points is thus expressed relative to that of the initial top point. The larger haystack is now located on the canvas and its outer shape has been reduced to an ordered feature vector.

3. The pattern recognition system locates the highest point in the canvas to the right or to the left of the contour of the large stack. It finds the top of the smaller stack if there is one, otherwise it stops. It extracts the 2 other points of the stack as it did for the larger stack.
4. The position of the smaller stack can be defined by the distance from the top of the smaller stack to that of the larger stack. If the value is positive the small stack is to the left of the large one and to its right if the value is negative. Similar results can be obtained if the inter-stack distance is calculated from the closest corners of each stack, as represented by the dashed lines on Figure 4.7b. The position of the small stack is calculated from the edge of the large stack.
5. To complete the foreground line drawn by the haystacks against the background, the pattern recognition system should link the bottom corners the stacks' sides that are facing each other and the other sides to the frame (dotted line on Figure 4.7b)

Classification When the process has been applied to all canvases, each feature of the haystack contour on a canvas has a corresponding feature in other canvases, realized or not (binary opposition) or realized but with different values (gradient distortion). The feature “top of the large stack” exists on each canvas although its position on the normalized canvas varies from canvas to canvas. This variation can be calculated and graded to organize the haystack contours in the category in terms of pattern distortion. The same goes for the

other two features of the large stack. When there is a small stack, the same variation in position and shape of the feature vector can be analyzed among all canvases. Finally, the distance and the relative position of the two stacks can be similarly analyzed. The pattern recognition system can eventually organize the instance in the haystack canvas category in various ways, looking at the nature of the contrast, whether it is binary (number of stacks, position of the small stack relative to the large one) or gradient (size of the stacks, position in the canvas, distance between each other). Table 4.1 is a simplified organization of the canvas by the pattern recognition system, with the distance between haystacks distributed among six values (Table 4.1).

binary contrasts			
2 stacks (C_2)		1 stack (C_1)	
	small on the left (C_{2l})	small on the right (C_{2r})	
graded distance	1	(05), (06)	(07)
	2	-	(08)
	3	(04)	(09)
	4	(03)	(10)
	5	(02)	-
	6	(01)	-

Table 4.1: Pattern recognition system's categorial organization of the canvas relative to their haystack contour

The pattern recognition system can organize the instances of the canvases into subsets of contours depending on their common features. The set C of all canvas contains 12 instances of haystack contours:

$$C = \{(01), (02), (03), (04), (05), (06), (07), (08), (09), (10), (11), (12)\}$$

Set C can be divided into two subsets: the subset of contours with one stack C_1 and the subset of contours with two stacks C_2 :

$$C = C_1 \cup C_2 = \{\textcircled{07}, \textcircled{08}, \textcircled{09}, \textcircled{10}\} \cup \{\textcircled{01}, \textcircled{02}, \textcircled{03}, \textcircled{04}, \textcircled{05}, \textcircled{06}, \textcircled{11}, \textcircled{12}\}$$

The subset C_2 can itself be divided into two subsets: the subset C_{2l} with the small stack on the left of the large stack and C_{2r} with the small stack on the right of the large stack.

$$C_2 = C_{2l} \cup C_{2r} = \{\textcircled{01}, \textcircled{02}, \textcircled{03}, \textcircled{04}, \textcircled{05}, \textcircled{06}\} \cup \{\textcircled{11}, \textcircled{12}\}$$

The instances in subset C_1 of contours with one stack all have the same grade of membership since they are all exactly the same in term of position. The instances in subset C_{2l} and C_{2r} vary in terms of distance between the small and large stack. The pattern recognition system should be able to grade each instance in the subset as a function of this distance. However, the system lacks a reference point to do so. Should the larger distances or the smaller ones have the highest grade of membership in the sub-category? Similarly, should contours with one stack or two stacks be ranked higher in C ? There is a need for ranking the instances the set or subsets contain since they do not all share the same values for one or more features.

A set within which all instances have the same grade of membership is a binary set. Objects or elements are in the set (membership = 1) or not (membership = 0), they cannot be ranked: binary membership is noted $m_{(x)} = \{0, 1\}$. An example of binary set is the subset C_1 . A set within which all instances do not have the same grade of membership is a graded or fuzzy set. Objects or elements are in the set with various grades of membership on the interval between 1 and 0 included: binary membership is noted $m_{(x)} = [0, 1]$.

Set C , subsets C_{2l} , and subset C_{2r} are fuzzy sets that can be graded in terms of number of stacks and subsequently by the distance between stacks.

Some general principles of fuzzy set theory are implemented into the classifier of the PR system so that it can grade each instance in the set of haystack contours and generate an abstract nominal figure extracted from the comparison (i.e. grading) of all instances.

Chapter 5

Fuzzy Set Theory

Clearly, the ‘class of all real numbers which are much greater than 1,’ or ‘the class of beautiful women,’ or ‘the class of tall men,’ do not constitute classes or sets in the usual mathematical sense of these terms. Yet, the fact remains that such imprecisely defined ‘classes’ play an important role in human thinking, particularly in the domain of pattern recognition, communication of information, and abstraction (Zadeh, 1965).

5.1 Computing with words

The paper by Zadeh (1965) titled *Fuzzy Sets* is held to be a seminal paper for research in non binary set theory and non binary logic. Its main postulate is that, unlike what happens in the realms of logic or mathematics, most real-world categories are vague categories. Zadeh transferred the concept of linguistic vagueness to the realm of mathematics by extension of crisp sets into what he named fuzzy sets.

Central to the theory, and as indicated by the quote by Zadeh at the beginning of this chapter, fuzzy sets represent common linguistic vague quantifiers such as *short/tall, lower/higher, before/after, cheap/expensive, young/old, etc.*. These words cannot be mapped onto a binary function since objects in

the real world are neither small nor not small but small to a certain degree, depending on the category of the object itself (there is no small skyscraper), the context (a skyscraper can be smaller than another one) and the person(s) applying the word to an object (a New Yorker is more used to skyscrapers than a bedouin in the Sahara). Linguistic categorization is the result of individual experience and knowledge. Thus a linguistic concept does not exactly cover the same range of meaning among speakers (Klir & Yuan, 1995). The concept of mapping words onto functions for decision making processes has been tagged *computing with words*. Fuzzy logic is part of a larger domain called *soft computing*, which also comprises neural networks and probabilistic reasoning. As such, fuzzy logic “ha[s] been used in image-understanding applications, such as detections of edges [contours], feature extractions, classification, and clustering” Dubois & Prade (2001).

5.2 Crisp and fuzzy sets

In a crisp or binary set A , an element x has full membership (m) or no membership at all, with $m_{(x)} = 1$ if $x \in A$ and $m_{(x)} = 0$ if $x \notin A$. For example, in the set of all odd numbers, the rule of inclusion and exclusion in the set is strictly binary and the classifying property “is odd” can only have two values: “is odd” ($m_{(x)} = 1$) or “is not odd” ($m_{(x)} = 0$). Thus, the grade of membership for value 5 is $m_{(x)} = 1$ (included), that for value 6 is $m_{(x)} = 0$ (excluded).

In a fuzzy set, instead of all elements being equal and having the same grade of membership, $m_{(x)} = 1$, each element has a grade of membership ranging between zero and one, such as for example, $m_{(x)} = 0.4$. For example, in the set of {tall men}, the linguistic quantifier “tall” is vague. If it was binary

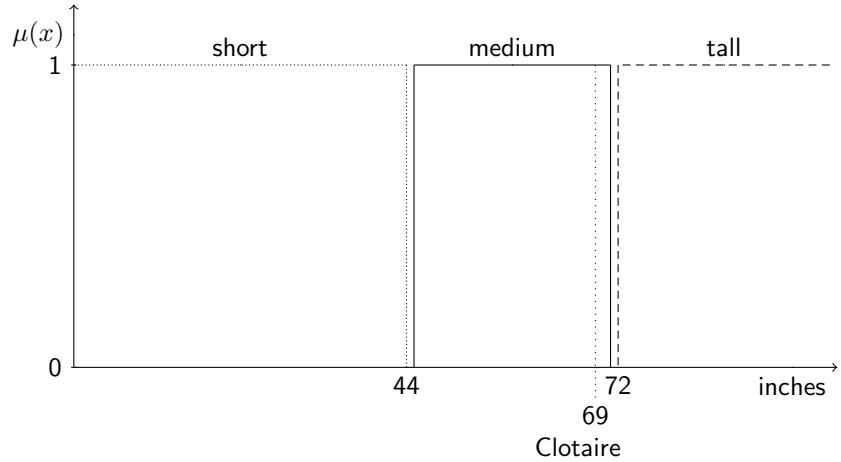


Figure 5.1: Binary sets of men's height

or crisp, there would be a numerical boundary, say 72 inches, corresponding to a binary function that would divide the set of all men between {tall men} and {not tall men}, as shown by Figure 5.1. Three groups could be created, each delimited by a height in inches: short (below 44 inches), medium (between 44 and 72 inches), and tall (above 72 inches). If an individual, called Clotaire, were 69 inches tall, then Clotaire would belong exclusively to the class of {medium men}, even though he is clearly “rather” tall. “Tallness” is vague because there are “borderline cases”, i.e., individuals to which it seems impossible either to apply or not to apply the term” Black (1937). If Clotaire were 69 inches tall, would he be not so short or would he be rather tall? On the other hand, if Clotaire were 101 inches tall, he would be undeniably tall. Thus, vague quantifiers do not allow to distribute objects into two crisps distinct sets. A fuzzy set such as the set of {tall men} is characterized by a function associating each element in the fuzzy set to a real number (a grade m) between 0 and 1 (Zadeh, 1965), as in Figure 5.2. In a fuzzy distribution of

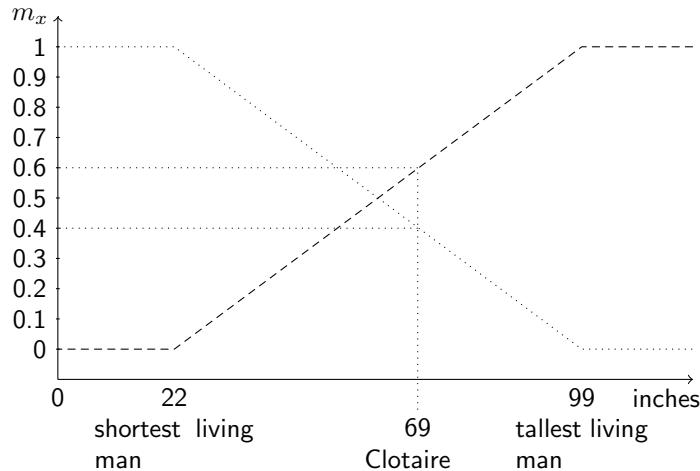


Figure 5.2: Fuzzy functions of the set of short men (dotted line) and of the set of tall men (dashed line), based on the current world size records established by the Guinness World Records

men according to their size, Clotaire belongs to the set of {short men} with a grade of membership of $m_{Clotaire} = 0.4$ and to the set of {tall men} with a grade of membership of $m_{Clotaire} = 0.6$. Clotaire's size is a borderline case since he is not short but he is not really tall either. Perhaps a third class of size could be added for medium sized people. Nonetheless, it would create borderline cases between short and medium, and, between medium and tall.

A grade of membership is a numerical value and the interpretation of a membership grade is context dependent, just as is the choice of the linguistic concept to quantify objects in a category. For example, in a decision-making application, m conveys the degree of preference towards an object in the category. In a classification application, m conveys the degree of proximity to the more prototypical instances in a category (Liao et al., 2003). With a quantifier such as “tall,” the grade of membership in the set of values corresponds to the

comparison of a value towards reference values (extreme heights in the case of Figure 5.2). A grade of $m_{(x)} = 1$ means full membership, $m_{(x)} = 0$ means total exclusion, and $m_{(x)} = 0.5$ is the limit of categorical status. Above $m_{(x)} = 0.5$ the categorical inclusion goes up increasingly with the grade of membership. Below $m_{(x)} = 0.5$ the categorical inclusion goes down increasingly with the grade of membership. In the example of Clotaire, he is more included in the set of {tall men} than he is in the set of {short men}. Naturally, since the two functions map two opposite linguistic concepts, the degree of inclusion of Clotaire in one set is exactly opposite to his degree of inclusion in the other set. This is an example of how fuzzy set theory captures contrast in real-world categories.

5.3 Building a fuzzy function

5.3.1 Fuzzification

“Fuzzy set theory provides a framework within which the process of knowledge acquisition takes place and in which the elicited knowledge can effectively be represented” (Klir & Yuan, 1995). Because linguistic quantifiers are context dependent and vague, the task of developing an adequate fuzzy application to capture linguistic concepts is left to expert individuals in the domain of the application. For example, the concepts “normal count of white blood cells” and “dangerous speed in snowy weather” should not be mapped onto fuzzy functions by the same group of experts. In many cases, the elements or objects to which the highest and lowest grades of membership (1 and 0) must be attributed are clearly associated to absolute inclusion quantifiers such as “not at all a...” (for 0), “fully a...” (for 1) or any context dependent quantifier such as “the tallest man in France” or “the average size of men in France”, and

this, “by reference to properties given in a theoretical literature or supplied by expert judges” (Smithson & Verkuilen, 2006).

The two functions of human height, as presented in Figures 5.1 and 5.2 have been built on purely numerical and statistical data. The two extreme heights create a natural limit to the two sets and, with more than 7 billion people in the world, chances are that almost any size can be found in the world population as a continuum between these two values. However, these limits are not context-dependent but purely numerical. On both sides of the continuum, extremely short people and extremely tall people form a very small group that is not representative of the world population. Indeed the range and the average height vary from place to place, depending on multifarious factors. Thus, the functions plotted on Figure 5.2 map the linguistic concept of height in a numerical and arbitrary way instead of integrating its context dependency. As a French man, Clotaire’s sense of height is different from that of a pygmy. The linguistic concepts “short” and “tall” can only be elicited by a survey of a given population. Individuals in a population should probably be asked to rank sizes on a graded continuum and from the results obtained, it might be possible to find some statistical tendencies of association of sizes with linguistic quantifiers. If a poll were run among a population, the height value most frequently associated with the quantifier “tall” would be attributed the grade of $m_{(x)} = 1$ and the grade of all other values would be the ratio of their frequency to that of the size most frequently qualified as tall.

Furthermore, the set of {tall men} is doubly problematic because not only is “tall” vague but so is “men”. What property(ies) must an individual have to be a man, especially what age? If Clotaire is 17, is he a man, an adolescent, a child? Again, it mostly depends on the regional and socio-cultural

contexts. And beyond that, imagine a moment that Clotaire claims to be transexual. Not everyone would classify him/her in the same gender category, depending on one's conception of gender. Clotaire might well be a tall woman for some and a medium young man for others.

Other sets of real-world objects are easier to map onto a function. Engineers creating an automated fuzzy system to control water pressure in a factory can input values in the fuzzy controller that come from experience and robust knowledge of material resistance and point of fracture. Fuzziness itself is a vague concept: a set can be more or less fuzzy depending on how many values the quantifier can take: two values (binary, e.g., odd numbers), a finite set of values (e.g., school grades: A,..., F), or a continuum of values (e.g., human height). The joint processes of creating a fuzzy function and assigning a grade of membership to the objects of a category or the elements of a set is called *fuzzification*. The next section is an illustration of fuzzification applied to a phonological category.

5.3.2 Phonemes as graded categories: a VOT fuzzy function

It is easy to conceive of phonemes as graded categories, with objects, i.e, instances, inside of these categories being mapped onto a fuzzy function. From the first experiments of Peterson and Barney in 1952 to the recent duplication of this work by Hillenbrand, it has been shown, at least for some varieties of English, but there is no reason to think it is otherwise in other languages, that vowels “are more properly modeled not as points in formant space but as trajectories through formant space” (Hillenbrand et al., 1995). Other researchers, such as Massaro, Olden, or Miller, work directly with the concepts of fuzzy logic or linguistic categorization (Barth-Weingarten, 2011;

Miller, 1994; Allen et al., 2003; Olden & Massaro, 1978; Massaro, 1989; Taylor, 2004).

Voice onset time (VOT) is the duration of the period of time between the release of a plosive and the beginning of vocal fold vibration (see Chapter 2 for references and discussion). Let VOT be the property of inclusion in the sets of allophonic instances of /p/ and /b/. As any quantifier, VOT is context-dependent. In English, voiceless plosives have a positive VOT value and voiced plosives have a VOT equal to zero. The situation is exactly the contrary in French: voiceless plosives have a VOT value equal to zero and voiced plosives have a negative VOT values. If the phonemic distribution is binary, the following mapping rules can be written:

1. B is the set of all allophones of /b/
2. P is the set of all allophones of /p/
3. x is the allophone to be classified:

Thus :

English rules: if $\text{VOT}_{(x)} > 0 \rightarrow x \in P$ & if $\text{VOT}_{(x)} = 0 \rightarrow x \in B$

French rules: if $\text{VOT}_{(x)} = 0 \rightarrow x \in P$ & if $\text{VOT}_{(x)} < 0 \rightarrow x \in B$

The VOT binary function can be mapped as the absence (0) or presence (>0 or <0) VOT.

In experimental settings, it has been shown that far from being so clear cut, VOT actually varies in duration, for the same phoneme, from speaker to speaker (see Chapter 2 but also Allen et al. (2003) or Gordon-Salant et al. (2008) for references). In a perception task led by Gordon-Salant et al. (2008) (see Figure 5.3), the VOT “was altered in the buy/pie continuum by varying the duration of the aspiration [release-vibration interval] in the natural token,

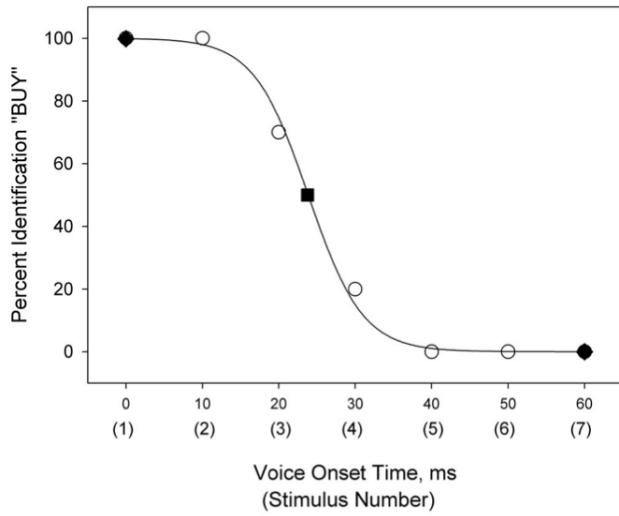


Figure 5.3: “Figure 5.3 plots the percentage of “buy” judgments as a function of VOT for a representative young normal-hearing participant. Actual data points are shown (open circles) as well as the PSIGNIFIT fitted function for these data and the derived crossover point (filled square) and endpoints (filled diamonds). The figure shows that this listener categorized the shortest VOT (stimulus 1) as “buy” and the longest VOT (stimulus 7) as pie. The performance of this listener on this stimulus pair was typical of the performance of all of the listeners on most of the stimulus pairs, indicating that participants heard the endpoint stimuli as intended. Typically, the crossover point for all listeners occurred in the region defined by stimuli 3, 4, and 5. (Gordon-Salant et al., 2008)”

pie, from 0 ms (buy) to 60 ms (pie) in 10 ms steps.” Thus there were seven different VOT durations in the created continuum, noted (1) to (7) on x-axis of Figure 5.3. Two groups participated in the experiment, divided into groups of young and elderly listeners. While listeners categorized phonemes by the presence or absence of VOT, the VOT is present or absent in a binary sense (exact match). Absence is perceived as a true absence (0 ms) or a short VOT (10 ms) while presence is perceived as a long VOT (40, 50, 60 ms). It was also found that, overall, elderly listeners needed more contrast

between presence and absence. This is due to the existence of borderline cases that are neither short enough (>10 ms) or long enough (<40 ms) to be categorized either as an absence or a presence of VOT, or, as noted by the authors: “typically, the crossover point for all listeners occurred in the region defined by stimuli 3, 4, and 5.” It is remarkable that this region of uncertainty occurs precisely on these VOT durations (20, 30, 40 ms) that form the center of the continuum. Durations on both extremities of the continuum are unproblematically categorized. In fact, the *s*-shaped buy/pie function of Figure 5.3 is reminiscent of a fuzzy function.

The fuzzy functions of /b/ and /p/ can be calculated by using the frequency of judgement as the variable to assign membership to VOT values in each set. In the set B , stimuli with a VOT of 0 ms are classified by listeners as /b/ 100% of the time. Thus for each fuzzy function, if the frequency of judgment is equal to 100, the grade of membership of the VOT value is 1. Conversely, if the frequency of judgment is equal to 0, the grade of membership of the VOT value is 0. The two fuzzy sets of VOT durations, as inferred from Figure 5.3, are signified below as {VOT duration value/grade of membership}:

1. $B = \{0/1, 10/1, 20/0.7, 25/0.5, 30/0.2, 40/0, 50/0, 60/0\}$
2. $P = \{0/0, 10/0, 20/0.3, 25/0.5, 30/0.8, 40/1, 50/1, 60/1\}$

The attribution of grade of membership to VOT values can also be presented as an array, with values in the set ordered similarly by descending grades of membership in the interval [1,0] :

m	[1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1]
B	{0, 10,
P	{40, 50, 60 30, 25, 20, 10, 0}

Finally, the two functions $m_{(P)}$ and $m_{(B)}$ can be plotted on a single graph, presented in Figure 5.4. The two functions are symmetrical and the “crossover” point is the point where the functions meet. This point, which corresponds to a VOT of 25 ms is graded 0.5, the exact center of the $[1, 0]$ interval of grade of membership. Thus, it is the edge for categorical status: before this point the VOT is more likely to have the consonant perceived as /b/, after this point, it is more likely to have the consonant perceived as /p/. The further away from this point, the clearer the categorical status of the consonant.

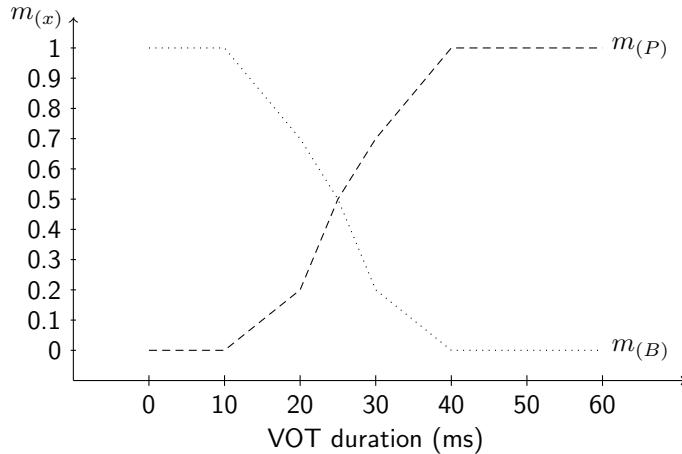


Figure 5.4: Fuzzy functions of the set B of allophones of /b/ (dotted line) and of the set P of allophones of /p/ (dashed line). The grade of membership of a phoneme in either one or the other set is a function of the duration of its VOT.

5.4 Defuzzification

Through fuzzification, the elements in a set have been given a grade of membership m . The input of the fuzzification process is a set of values which are all considered for full membership. Initially all values are fully included

in the set with $m_{(x)} = 1$ and the output is a set in which values have been ranked according to a fuzzy function (representing a linguistic concept) and ordered in the set by grade of membership from 0 to 1, like B and P in the VOT example.

The defuzzification consists of taking the aggregated output fuzzy set and reducing it to a single *crisp value* that captures the graded organization of the set by the fuzzy linguistic quantifier. There are many ways to calculate the defuzzified or crisp value of the set (see Dubois & Prade (2001); Klir & Yuan (1995); Sivanandam et al. (2007); Yan (1994) among others for a review of defuzzification methods). The most common method is the *center of gravity*, by which the geometric center of the function is determined. This center corresponds to a mean of all values of the set weighted by their grade of membership (weighted mean):

$$\bar{m} = \frac{\sum m_{(x)} \cdot x}{\sum m_{(x)}}$$

The resulting crisp value represents the entire set; weighting ensures that all elements in the set contribute to the final output according to their grade of membership. As an example, the sets B and P of VOT values are defuzzified below:

$$\bar{m}_B = \frac{(0 + 10) \cdot 1 + 25 \cdot 0.5 + 30 \cdot 0.2 + (40 + 50 + 60) \cdot 0}{2 \cdot 1 + 1 \cdot 0.5 + 1 \cdot 0.2 + 3 \cdot 0} = 12.5 \text{ ms}$$

$$\bar{m}_P = \frac{(40 + 50 + 60) \cdot 1 + 30 \cdot 0.7 + 25 \cdot 0.5 + 20 \cdot 0.3 + (0 + 10) \cdot 0}{3 \cdot 1 + 1 \cdot 0.5 + 1 \cdot 0.3 + 2 \cdot 0} = 41.85 \text{ ms}$$

The crisp values of B and P are 12.5 ms and 41.85 ms. These two values correspond roughly to the limiting points of full membership for the two functions in Figure 5.4. With a VOT duration below 12.5 ms, the consonant is

unequivocally categorized as a /b/. With a VOT duration over 41.85 ms, the consonant is unequivocally categorized as a /p/. Between these values the categorical status of a consonant varies from the central value of 25 ms, which is equally categorizable as a /b/ or as a /p/. Below 25 ms, the degree of categorization as a /b/ increases as the VOT diminishes. Over 25 ms, the degree of categorization as a /p/ increases as the VOT augments. This value of 25 ms also corresponds to the crossover point found in previous studies of VOT discussed in Chapter 2.

Chapter 6

Acquisition and preparation of the data

The PRInt model abstracts the prototype of intonation contours from a category's group of instances. A group of speakers was recruited to take part in an elicitation task that was designed to generate the original audio recorded data. The recordings were manually parsed into syllables and processed with Praat, creating text files for input to the PRInt model.

6.1 Participants

No personal information about the participants was recorded. The only requirement for participation in the task was to be a native speaker of French.

The study involved 22 participants (12 women, 10 men), all of them native speakers of French. Some lived in France, in the cities of Paris and Cahors, others were living in the regions of either Austin, Texas, or San Francisco, California. The origin of the participants was quite diverse, especially among the group of expatriates who came from various regions of France and various social backgrounds. Participants were between the age of 20 and 55. For those living in the USA, they were over 20 when they first came to the country.

6.2 Material and setting

- The recordings took place in a quiet, closed space. For the type of analysis conducted in this study, a sound-proof environment was not necessary.
- Participants were recorded directly onto a computer using a USB condenser microphone AudioTechnica 2020. An external device was also used for back up, a portable microphone Zoom H4. Both microphones were set to encode the signal into .wav files (sampling frequency of 44,100 Hz).
- Participants were wearing headphones (AKG K701) for the duration of the task.
- The material was presented on a computer screen as a sequence of slides.

6.3 Elicitation tasks

For the application of the PRInt model presented in this study, three intonation contours were analyzed: unmarked closed question, surprise, and doubt, the latter two being considered as modalities or variations of unmarked closed questions. Accordingly, three corpora were acquired using elicitation tasks.

There were two elicitation tasks with the same general setting. On a computer screen, participants were presented a set of slides with a voiced over tutorial explaining how to perform the task and also how to control its progression. Most importantly, the task was not timed so participants controlled the pace. These audio instructions could be accessed at any time during the elicitation task.

In the instructions, the intonation category was named and defined in a short sentence. Three audio instantiated examples of the contour were provided for participants to listen to. Examples could be heard on demand. The idea was to activate a conceptual category in the mind of the participants so that they would produce more instances of the same category on the provided sentences.

Each corpus must contain instances of the same contour, realized with more or less felicitous results. As discussed in Chapter 3, intonation concepts are vague, both in their intension and extension. Their category name can only be evoked by a phrase or some sort of definition, or by giving an example by instantiating the intonation over some material such as a phrase or a sentence. This is how the elicitation tasks were designed. Participants were given a context and a sentence over which they were asked to apply one of the three contours. The type of contour they had to apply was presented as both a phrase and an example of the contour over a sentence. Their own realization of the contour over the prompted sentence depended on their own categories and linguistic system. It also depended on how fitting, in their own mind, the intonation they were asked to produce was with the context, and within the relation between the context and the prompted sentence.

Participants were asked to consciously produce contours based solely on the vague intension they have been provided with in the form of a phrase or an instantiated example. The format of the task had a direct effect on the participants. In the example above (Figure 6.1), the instructed contour was that of a closed question. As shown by the F_0 curve, after starting her utterance very high in the F_0 range, the speaker had to readjust her intonation contour to match the intended one. She lowered her F_0 to match that of the

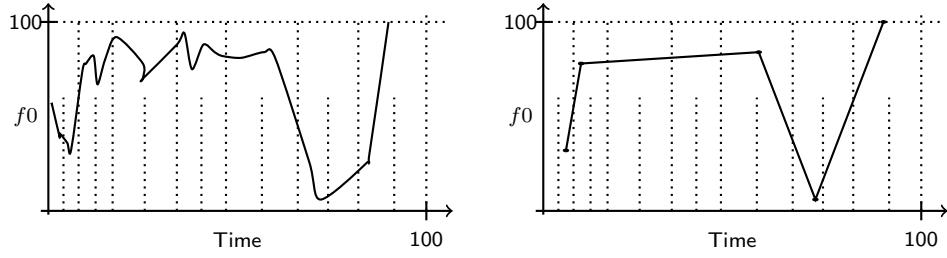


Figure 6.1: A distorted instance of a closed question contour: normalized (left) and analyzed by PRInt (right).

expected contour on the sixth syllable. Otherwise, she would have been forced to raise her F_0 out of range and may not have been able to realize the contour.

6.3.1 Unmarked closed question

In each task, there were 34 sets of contexts and sentences presented three times in random order (see Appendices 1 and 2 for a list of contexts and sentences).

The first elicitation task required participants ($n=20$) to produce intonation contours associated with the meaning of closed question. Only the word “question” was used in the instructions with no further definition. Emphasis had been placed on the fact that intonation was the only cue for the intended meaning. Closed questions are non problematic because of their highly frequent use and pragmatic transparency. In other words, the concept of a question is easily accessible for speakers, although its application still remains subject to contextual variations. Unmarked closed questions are thus a good choice for developing the PRInt model since the output is controllable to a certain extent. On each slide, a context was given and it was followed by a sentence to be pronounced as a closed question by the participants. Figure

6.2 is an example of a slide presented during the task.

CONTEXTE 1

Il y a des gens qui aiment garder des souvenirs après un concert.

→ **Tu vas garder les tickets ?**

Figure 6.2: Example of a slide used for the task of closed questions.

"Some people like to keep souvenirs after a concert. -Are you going to keep the tickets?

6.3.2 Modalities of question

For the second task, participants ($n=22$) were instructed to produce two contrastive contours: one for surprise (*la question étonnée*) and one for the doubt (*la question incrédule*).

The short contexts provided on each slide were prepared to generate disbelief: the participant is placed in a context in which someone reveals a fact that goes against expectation. As was observed during an initial pilot study, the use of the general concept of disbelief generated both contours. Thus it was decided to ask participants to try, as well as they possibly could, to produce a contrastive pair of intonation contours. In many cases, participants struggled in separating the contours and produced contours falling in between the two modalities. This is a direct consequence of the vagueness of both the definition of this type of contour and their realization on request. However, it fits the general idea of the PRInt model: peripheral instances will be graded lower in typicality anyway.

Participants were instructed to read the context presented on each slide and then to read the sentence immediately below as a question, but also to

CONTEXTE 3

Votre ami a grandi dans une ferme où on élève des moutons et des agneaux.
Pourtant il vous dit n'avoir jamais mangé d'agneau.

a . Dans votre question, exprimez votre **étonnement**:

→ **t'as jamais mangé d'agneau ?!**

b . Dans votre question, exprimez votre **doute**.

→ **t'as jamais mangé d'agneau ?!**

Figure 6.3: Example of a slide used for the task of modality contrast.

“Your friend grew up in a farm where sheep and lambs were bred. However, he tells you he never ate lamb. -You never ate lamb?!”

a) In your question, express your surprise; b) In your question, express your doubt

indicate – as well as they could and in the way they thought appropriate – first their surprise, and second, their doubt, relative to the “unbelievable” information someone is revealing in the context.

Each context was followed twice by the question to be read. Before the first instance of the question, participants were instructed: *Dans votre question, exprimez votre étonnement*, “in your question, express your surprise”. Before the second instance of the question, participants were instructed: *Dans votre question, exprimez votre doute*, “in your question, express your doubt.”

Figure 6.3 is an example of a slide used during the elicitation task.

6.4 Sentence format

The sentences selected for the study have three common characteristics.

1. The current implementation of the PRInt model imposes a fixed number of syllables for all sentences. This number has been set to seven, the observed limit for an accentual arc (see Chapter 3). This length also allows participants to realize the contours as more than one arc, that is with more than one intonation unit. Variation in the number of intonational phrases is also analyzable through categorization of the instances of a contour by PRInt.
2. The second constraint derives from the first one. All syllables must contain a full vowel and exclude any possible unpronounced schwa [ə]. Beyond this restriction, the segmental content of the syllables was not controlled and varies in composition from syllables containing only a single vowel to syllables having complex consonantal onset and/or coda.
3. The sentences are morphologically and syntactically unmarked for the meaning of question. Without intonation, they are declarative sentences.

6.5 Data preparation

The total number of instances in each category of contour included: 2040 closed questions, 2244 instances of doubt, 2244 instances of surprise. All elicited utterances were manually annotated and parsed into syllables using the software Praat (Boersma & Weenink, 2012). From the initial annotations and parsing, three types of data are obtained with a series of script:

1. Duration of sentences

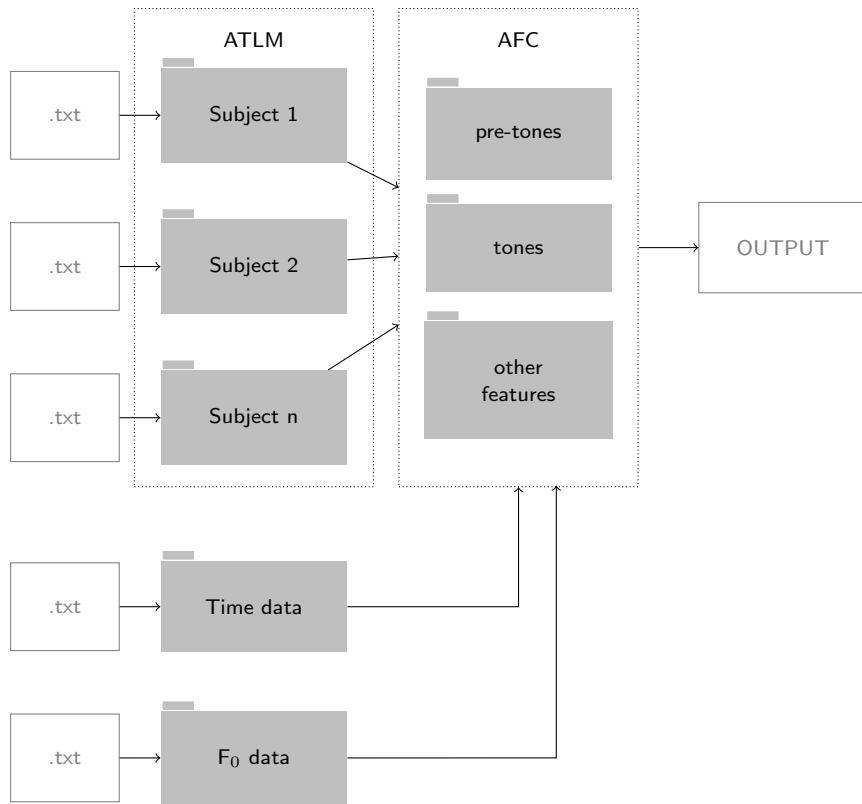


Figure 6.4: Implementation of PRInt in Excel

2. Duration of syllables in each sentence
3. F₀ as a function of time in each sentence

These data are stored as text files and input to the PRInt model.

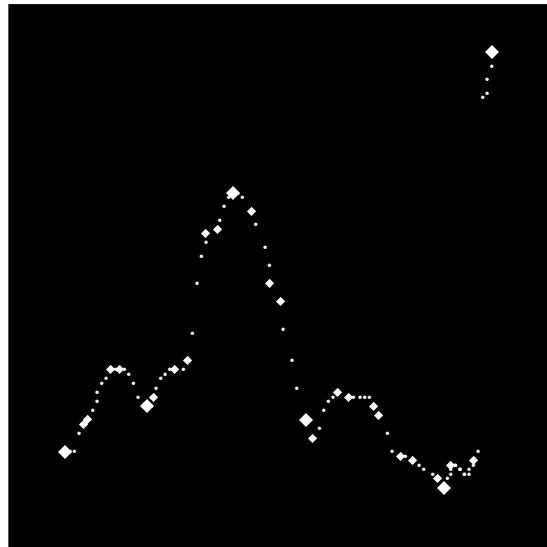
The PRInt model itself is implemented on a series of inter-connected Excel workbooks. The structure of the model is presented in Figure 6.4. Each Excel workbook is represented in gray and text files are in white. There is a workbook for each participant, in which the ATLM is implemented. The data of each sentence are analyzed as a feature vector and the results for all sen-

tences are consolidated into one table. The results of all individual workbooks are sent to a group of workbooks in which the AFC has been implemented. Each of these workbooks performs the fuzzification, defuzzification and conversion calculations at various levels of analysis but especially pre-tones, tones, and tonal relations. Time data (duration of sentences and syllables) and F_0 data are treated separately in two workbooks in which are standalone version of the AFC has been installed. Their results are sent to the main AFC workbooks and integrated into the calculation. Finally, the AFC generates the output of the analysis in the form of tables that can be exported or converted to graphs. The next two chapters describe the PRInt model. Chapter 7 is dedicated to the labeling module (ATLM) and Chapter 8 to the classifier (AFC).

Chapter 9, provides a detailed description of the various calculations implemented in the workbooks. Chapters 9 and 10 present the results of the analysis of three intonation patterns by the PRInt model (closed questions, surprise, and doubt). Chapter 11 is a comparison of the three contours and an analysis of their phonological/phonetic status.

Chapter 7

Automated tonal labeling module The 4-layer structure



/vuvulekõvjεnvuvwar/

Any pattern which can be classified in some category must possess a number of *features*. The first step in the process of classification is to consider the problem, what features to select and how to *extract* (measure) them (Friedman & Kandel, 1999).

7.1 Living up with different texts

In a class or a corpus of instances of a contour, all instances are different. They sound alike, and graphically look alike, but, although the human mind effortlessly and naturally recognizes a recurring pattern among the instances, there is a lot of physical variation in size and shape of this unique pattern. To look for pattern(s) among instances of a contour, this contour must be defined as a vector of features that the PRInt model can find in each sentence. The PRInt model is a somewhat engineeringly driven approach to the analysis of intonation contours and their variations. Rather than running sophisticated calculations or sophisticated polynomial functions over each instance, the PRInt relies on a human-like sequential process to interpret all instances of the contour as the same vector of features whose implementation and dimensions change with the instance. In a more linguistic way, the PRInt model extracts the units of the macro-prosodic (coarse grained phonological features) level from those of the micro-prosodic level (fine grained phonetic features): in a sequence of simple calculations, it labels the instance as a string of tonal targets or tones from the F_0 information. The theoretic point of departure of the process is found in Janet Pierrehumbert's dissertation, repeated here for convenience (and for appreciation):

To our mind, a theory framed in terms of target levels is attractive because it affords good facilities for describing **how the same intonation pattern lives up with different texts**; the crucial points in the contour, the F_0 targets, can be lined up with crucial points in the text, with stretches in between computed accordingly. The behavior of a given contour under changes in pitch range can

be modeled in a similar fashion, by transforming the target points (Pierrehumbert, 1980).

The labeling module of the PRInt, or Automated Tonal Labeling Module (ATLM), is an attempt to implement these ideas into an automated system for the description of intonation contours and their variations. This module is a bottom-up system and operates reversely from Pierrehumbert's statement. From the acoustic data of a sentence (section 7.3.1), it computes the *stretches* (section 7.3.2, locates the *crucial points* (section 7.3.3), and finds the F_0 *targets* (section 7.3.4). However, the labeling module is blind to the meaning of the sentences in that it operates on the acoustic signal (F_0 and time data) and the phonological structure (phonemes and syllables) only, regardless of the meaning of the sentence.

7.2 Thinking human: an analytical approach to tonal labeling

Before entering the description of the labeling module (ATLM), it is necessary to present the human analytical approach to the tonal labeling of an intonation contour from its F_0 contour, as this contour appears on the window of a software for acoustic research such as Praat or WinPitch. Human beings (linguists included) naturally analyze physical objects at several levels of resolution. More specifically, they can distinguish between the general shape of the object and the set of features that constitute it, and how these features are organized in a structure. The following description of two intonation contours might seem very elaborate for something that seems so benign and easy to interpret for a human ear or eye, but to create an automated system that can distinguish between micro and macro-prosody, a system that can identify the

crucial points of a contour from all the available points, it is necessary to try to formalize, even if in a mechanistic way, the process of feature selection for an intonation contour from the F_0 /time data.

In Figure 7.1 below, (a) and (b) are two intonation contours. There is no indication of their respective duration ($=x$) and F_0 span ($=y$), only that they are on the same scale. Their associated meaning is unknown:

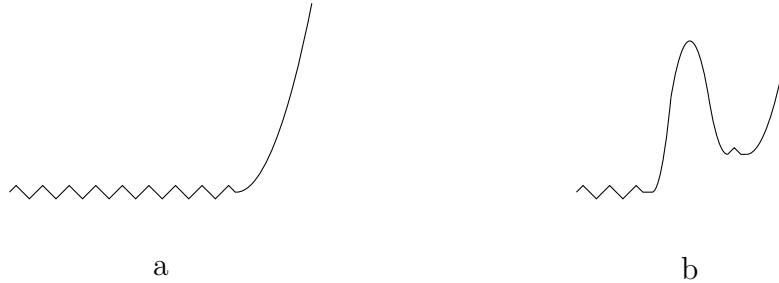


Figure 7.1: Two simplified intonation contours (1)

Even without more information, one can informally characterize them by 1) their overall relative dimensions, and 2) the general pattern they form as a series of vectors between points:

- [1] (a) is broader than (b), both horizontally and vertically
- [2a] contour (a) can be decomposed into 3 points and 2 parts : an uneven plateau followed by a rise
- [2b] contour (b) can be decomposed into 5 points and 4 parts: an uneven plateau followed by a first rise, followed by a fall, and a final second rise.

Three related remarks can be made about these descriptions. First, the smaller movements of the plateaus (the zigzagging lines) do not prevent the analysis of the contours in their general shape. The contrasts are relative

and the large amplitude movements make small amplitude movements appear secondary, or not relevant, to the general shape of the contour. Second, it is possible to distinguish the two peaks in contour (b) based on their relative *quantity*. The points are both high compared to the baseline of the figure but the high point on the left is higher than the one on the right. Thus, just looking at these two graphic representation of contours, one would naturally perceive a gradient contrast between high and low points, with those forming a maximum contrast serving as a point of reference for the others. To use the technical names, the plateaus, the rises, and the falls that characterize the overall contour as its constitutive parts correspond to the macro-prosodic level. The movements of smaller amplitudes correspond to the micro-prosodic level. Third, one would also think of the highest of the two peaks of contour (b) as occurring before the lowest one: points are distinguished by their occurrence in time, or their position, left to right on the graph. If the movements of lesser amplitude are abstracted from the figures, contours (a) and (b) are interpreted as in Figure 7.2 below.

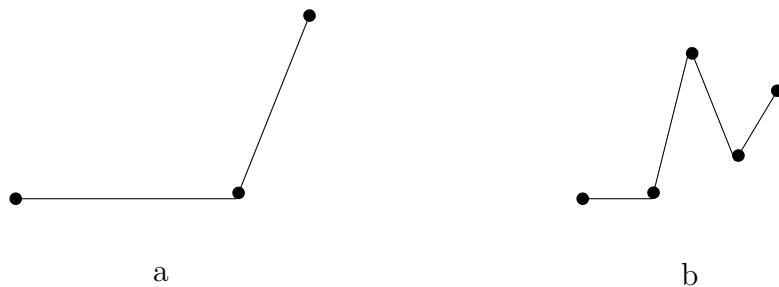


Figure 7.2: Two simplified intonation contours (2)

Let $\{T\%, L, H\}$ be a limited set of labels for the points of an intonation contour. The points $T\%$ mark the left and right boundaries of the contour,

its start and end. L is low relatively to H, or, L is lower than H. A hyphen (-) between labels indicates this relation from one point to the next: e.g, in the [L-H] relation, L- is lower than the following H or H is higher than the preceding. L and H mark turning points in the contour. With this set of labels, the two contours can be transcribed as in Figure 7.3 below.

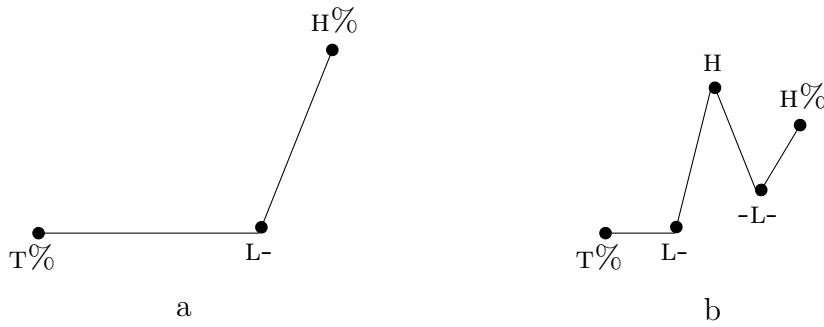


Figure 7.3: Two simplified intonation contours (3)

contour a T%[L-H] T_# → T% L-H%, H and T% being merged

contour b T% T [L-H] [H-L] [L-H] T% → T% L-H-L-H%, some points being merged.

This limited notation resembles the Pierrehumbert or ToBI notations and captures, even if imperfectly, the global pattern of the contours. The goal of this simplified example is to show that human beings (or at least linguists) naturally analyze a complex object as the organization of smaller relevant constituents and distinguish these constituent from local (the micro-prosody) or global alterations of the pattern (the overall dimension of the contours). At an abstract level, the contours are analyzed as a series of points whose

categorization as (H or L) is defined relationally (-) by vague quantifiers (point x is lower or higher than point y) and by their position on a temporal grid (point x occurs before or after point y).

The position of the points in the contours can be expressed relatively to the other points of the sentence (internally) or relatively to the position of corresponding points in a set of sentences (externally). For contour (a) and (b), these contrasts are:

- External contrasts:
 - (b) has two H and (a) has only one.
 - The H of (a) is higher than either H of (b).
 - The contour final H of (a) is higher than the contour final H of (b).
- Internal contrast:
 - In (b), the non contour-final H (on the left) is higher than the contour final one (on the right). The span $L-H$ of the non final rise is broader than the span of the contour final rise.
 - the two H of (b) occur sequentially from left to right. The highest of the two occurs first.

Higher vs. *lower*, *before* vs. *after* are typical vague quantifiers and points of reference, both context and subject dependent. They work perfectly well as descriptive tools for the relation between the points of the contour from a human perspective. For a computational system, there is a need to precisely quantify these relations (how much higher or lower a point is relatively to another, how much before or after two points are, respectively to each other) while preserving the gradient aspect of the the contrast between the two.

The Automated Fuzzification Process relies on the principles of fuzzy logic to find the patterns of an intonation contour among sentences, and to organize these patterns as the graded variations of the contour. In order to do so, and based on what has been described so far, the PRInt must, at the sentence level, be able to:

1. quantify the contrasts (or *stretches*) (sections 7.3.1 & 7.3.2)
2. determine which contrasts are the most relevant (or *crucial*) to the overall contour and which are less relevant (macro vs. micro-prosody) (sections 7.3.3 & section 7.3.4)
3. encode the intonation contour of the sentence as a string of points (or F_0 *targets*) that can be compared to the string of points of other sentences (section 7.3.4)

The methodology developed for the PRInt to analyze individual sentences seeks to match as closely as possible the human use of natural vague quantifiers (higher, lower, before, after), without the recourse to a sophisticated mathematical function but with a sequential, analytical and logical approach that mimics the visual analysis one would do using a software such as PRAAT. This methodological tool is the *4-layer structure*, implemented in the Automated Tonal Labeling Module of the PRInt.

7.3 Automated Tonal Labeling Module & the 4-layer structure

The analysis of sentences as a 4-layer structure, as implemented in the ATLM, is a systematical, bottom-up approach to the issue of labeling tones in a sentence.

Layer 1 The first layer is that of the original data, the series of time and F_0 coordinates that serve as the numerical equivalents to the acoustical signal. It is entered in the ATLM as it is provided by PRAAT: a two column tab delimited .txt file with the time values on the left and the corresponding F_0 values on the right.

Layer 2 - Low-level features The second layer is that of scalar quantization. The ATLM scales the sentences to convert the absolute position of the points to positions relative to the same frame of reference for all sentences. Thus, contrasts in F_0 and time can be expressed uniformly throughout the corpus (sections 7.3.1 & 7.3.2)

Layer 3 - Mid-level features At the third level, sentences are all decomposed into the same set of attributes (pre-tones) anchored to a fixed syllabic grid and defined by local contrasts (micro-prosody or local highs and lows) (section 7.3.3)

Layer 4 - High-level features At the fourth level, sentences are all decomposed into the same subset of features (tones) anchored to a fixed tonal grid and defined by global contrasts (macro-prosody or sentence highs and lows) (section 7.3.4)

The selection of points as pre-tones, from layer 2 to 3, is purely binary. The selection of pre-tones as tones, from layer 3 to 4, is gradient.

To illustrate the 4-layer structure analysis of a sentence, two samples from the corpora used in this research are presented in Figure 7.4 as contours (c) or [JP72], from its label in the corpus, and (d) or [GR65].



Figure 7.4: Two actual contours of closed questions in French (1)

7.3.1 Layer 1: frame of reference

What contour (c) and (d) have in common are points of reference. The first layer of the structure contains the original data, a series of time and F_0 coordinates, and among these values, two pairs can be found in the data of all sentences because they are defined structurally by their position in the data set. The values of these pairs and the distance between the two points of each pair vary but they are present in the data of all instances, and constitute the abstract and fixed *frame of reference* or pattern space of any given instance:

1. In the horizontal dimension (time or x)
 - (a) a starting point T%, corresponding the minimum time value (MIN)
 - (b) an ending point T%, corresponding the maximum time value (MAX)
 - (c) the span between the starting and ending points is the duration of the sentence (RAN).
2. In the vertical dimension (F_0 or y)
 - (a) a F_0 baseline, corresponding to the minimum F_0 value (MIN) of the sentence.

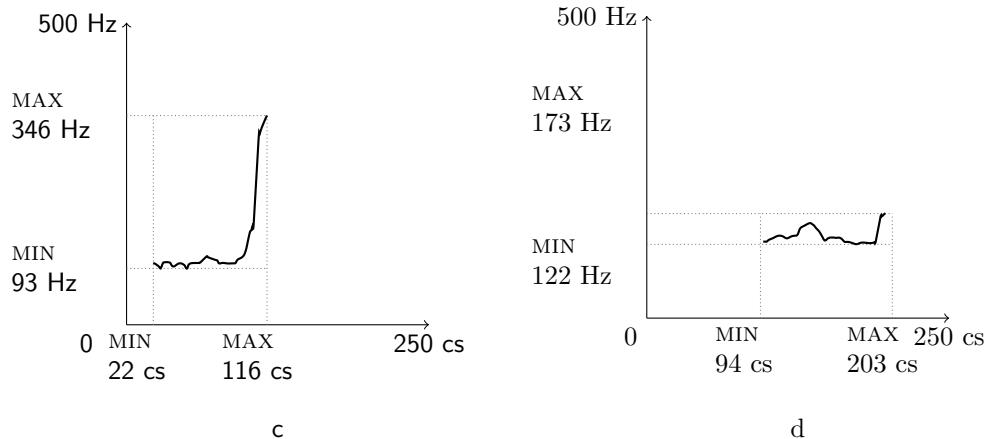


Figure 7.5: Two actual contours of closed questions in French (2)

- (b) a F_0 excursion corresponding to the maximum F_0 value (MAX) of the sentence
- (c) the span between the baseline and the top of the excursion is the F_0 range (RAN) of the sentence.

In Figure 7.5, (c) and (d) are presented on the same scale and with the values of their frame of reference. On the abscissa is time, in centiseconds (cs) from 0 at to 250 cs. On the ordinate is F_0 , in Hertz (Hz) from 0 at to 500 Hz: Note that the continuous line of the graphs should not be interpreted as continuous voicing, but as the graphical way to represent the shape of the contour.

7.3.2 Layer 2: scalar quantization

The second layer of the structure is that of the segmentation. “Segmentation algorithms isolate discrete objects from a grouping of objects for further analysis and classification” Nadler & Smith (1993). In the case of intonation contours, the continuous signal is segmented into a chain F_0 values as

a function of time. Before detailing the quantization process, which segment the input data into a smaller set of points, it is important to note that what is called raw data is not the continuous acoustic signal but an already discretized signal. The continuous acoustic signal from the microphone is sampled at regular time interval in the computer, leading to a series of close, but discrete, point in time with there associated F_0 .

Quantization is a common process in signal processing, used for example in digital music format such as MP3. Mathematically, it is the mapping of a large set of values (possibly infinite and uncountable, such as time (cs.) or frequency (Hz) values) into a smaller set (finite and made of integers). The size of the frame of reference varies from sentence to sentence. In the time dimension, the silence before the actual start of the sentence depends on how the sample has been cut from the rest of the recording, and, from their starting point, all sentences also vary in duration. In the F_0 dimension, the baseline and the excursion change for the same speaker from sentence to sentence, and among speakers as well. The frame of reference of all sentences has a $(0, 0)$ origin but has no x or y limits. The sets of time and F_0 values are infinite sets of positive non-integers $\{0, +\infty\}$. Keeping in mind that the goal of the PRInt is to find patterns, the nature of these sets poses a crucial problem: the chance that a pair of coordinate (x, y) occurs more than once is very low, and the chance that several pairs of coordinates form a recurring pattern is even lower.

Therefore, instead of expressing the coordinates of a point in absolute values, they can be expressed relatively to the limits of the frame of reference, in a process called scaling. Scaling absolute values to relative ones retains the relative positions between points (higher, lower, before, after) while limiting

the range of values in which they are expressed and making them comparable among sentences throughout the corpus. A F_0 span of 50 hz and a F_0 span of 100 Hz are different in absolute terms but relatively equivalent if the overall F_0 range of the sentence is 100 Hz for the former and 200 Hz for the latter.

The data of each sentence is scaled to 100, graphically matching a cartesian plane of 100 points by 100 points. This plane is also called the *hyperspace*, the abstract projection of the data on a graphical plane to facilitate the explanation of the process. As a second form of quantization, values are rounded to the nearest integer. Scaling and rounding turn the infinite set of non-integers to a finite set of integers. The choice of 100 is arbitrary but it follows the common habit of expressing relative values as percentages. Scaling the data to a higher value will lead to a higher resolution by reducing the number of values binned together as a single percentage. The extreme time and F_0 values of a sentence are reset to 0 for the minimum and 100 for the maximum in both dimensions¹. The values between are scaled accordingly. Graphically, the minimum values of time and F_0 of a sentence become the origin of the plane and the rightmost and upper limits of the plane correspond to the maximum values of time and F_0 respectively.

To illustrate the process, the time and F_0 data attached to sample GR72 [contour (c)], as extracted with Praat and exported to Excel, are provided in Table 7.1, in columns 2 and 3 (raw data). Time and F_0 are scaled with the same function. The scaled value (f_x) is the product of: the difference between the value to be converted (x) and the minimum value of the dimension (MIN),

¹For technical reasons, the minimum F_0 value is scaled with a minimum value corresponding to the minimum value of the sentence minus one: working MIN = actual MIN-1

	(1) ↓	(2) ↓	(3) ↓	(4) ↓	(5) ↓	(6) ↓		(1) ↓	(2) ↓	(3) ↓	(4) ↓	(5) ↓	(6) ↓		
frame #	raw cs	data Hz		time % pass 1		pass 2		frame #	raw cs	data Hz		time % pass 1		pass 2	f0 %
1	22	103		1	3	4		10	70	110		51	65		7
1	23	101		2	6	4		10	71	109		52	67		7
2	24	100		3	10	3		10	72	108		53	68		6
2	25	99		4	13	3		10	73	108		55	71		6
3	26	98		5	15	2		11	74	107		56	72		6
3	27	95		6	16	1		11	75	106		57	73		6
3	28	93		7	17	0		11	76	102		58	74		4
3	29	98		8	18	2		11	77	102		59	75		4
3	30	102		9	20	4		11	78	102		60	76		4
3	31	103	10	21	4			11	79	102		61	76		4
4	32	104	11	22	5			11	80	102		62	77		4
4	33	104	12	23	5			11	81	102		63	78		4
4	34	104	13	24	5			12	82	102		64	79		4
4	35	103	14	25	4			12	83	102		65	80		4
4	36	102	15	27	4			12	84	101		66	81		4
4	37	99	16	28	3			12	85	101		67	81		4
5	38	97	18	30	2			12	86	101		68	82		4
5	39	97	19	32	2			12	87	102		69	83		4
5	40	98	20	33	2			12	88	102		70	84		4
5	41	99	21	35	3			12	89	102		71	85		4
5	42	100	22	36	3			13	90	102		73	86		4
6	43	101	23	37	4			13	91	106		74	87		6
6	44	102	24	39	4			13	92	108		75	87		6
6	45	101	25	40	4			13	93	109		76	88		7
6	46	101	26	42	3			13	94	110		77	88		7
7	47	100	27	43	3			13	95	111		78	89		8
7	48	98	28	44	2			13	96	113		79	89		8
7	49	94	29	44	1			13	97	115		80	90		9
7	50	93	30	45	0			13	98	120		81	90		11
7	51	99	31	46	3			13	99	127		82	91		14
7	52	101	32	47	3			13	100	137		83	91		18
7	53	102	33	48	4			13	101	146		84	92		22
7	54	102	34	48	4			13	102	154		85	92		25
7	55	102	35	49	4			14	103	157		86	93		26
8	56	102	37	51	4			14	104	163		87	93		28
8	57	102	38	51	4			14	105	169		88	94		31
8	58	102	39	52	4			14	109	307		93	96		86
8	59	102	40	53	4			14	110	316		94	97		90
8	60	103	41	54	4			14	111	323		95	97		92
8	61	103	42	54	4			14	112	328		96	98		94
9	66	113	47	59	8			14	113	333		97	98		96
9	67	112	48	61	8			14	114	338		98	99		98
9	68	111	49	62	8			14	115	342		99	99		100
9	69	110	50	64	7										

Table 7.1: Data for GR72 and scaling. The data for point 10 has been highlighted in gray

and, the ratio of 100 to the range of the dimension (RAN).

$$f_x = (x - \text{MIN}) \cdot (100/\text{RAN})$$

The 10th point of sentence [JP72] (highlighted in gray in table 7.1) serves as an example of the scaling function. Its coordinates in time and F₀ are (31cs, 103 Hz).

Time (x) (Pass 1) The points of reference (in cs) of sentence [JP72] are:

$$116(\text{MAX}) - 22(\text{MIN}) = 94(\text{RAN})$$

Point 10 of sentence [JP72] has a time value of 31 cs. Its scaled value is calculated with the points of reference:

$$f_{31} = (31 - 22) \cdot (100/94) = 10\%$$

The results of time scaling for all points is in column 4 (time %, pass 1) of table (7.1).

F₀ (y) The points of reference (in Hz) of sentence [JP72] are:

$$346(\text{MAX}) - 93(\text{MIN}) = 253(\text{RAN})$$

Point 10 of sentence [JP72] has a F₀ value of 103 Hz. Its scaled value is calculated with these points of reference:

$$f_{103} = (103 - 93) \cdot (100/253) = 4\%$$

The results of F₀ scaling for all points is in column 6 (F₀%) of table (7.1).

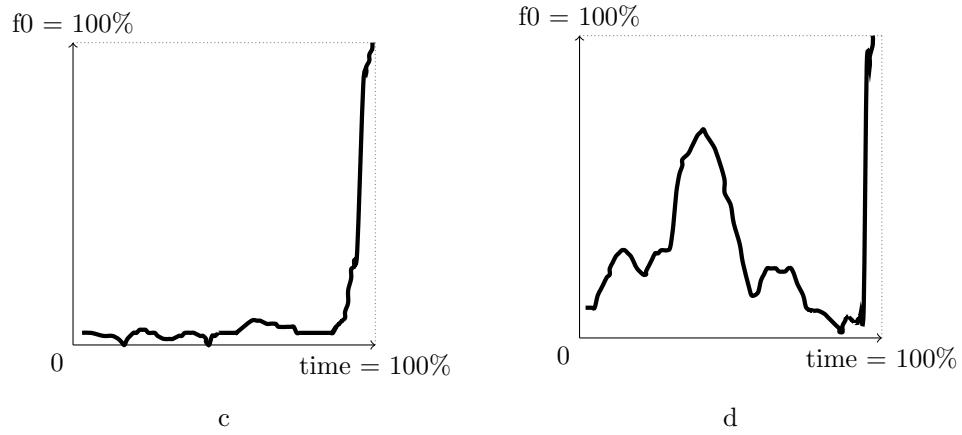


Figure 7.6: Two scaled contours of closed questions in French

The graphic representation of the scaling process for sentences [JP72] and [GR65] is presented in Figure 7.6. At the end of the scaling process, all sentences in a corpus have been turned into a matrix of one hundred points horizontally and one hundred points vertically, or $100 \times 100 = 10,000$ potential coordinates pairs. Figure 7.7 and 7.8 are a graphic representation of contours (c) and (d) as such a matrix. The sentences are 100 points long in time on the abscissa. Each of these time points can be or not be realized. When they are realized, the dot at the intersection with the F_0 value is marked in black on the figure.

Scaling converts an open set of non-integers into a finite set of integers. Yet, at the level of layer 2 of the sentence analysis, the number of possible points (100^2) and their combinations (100^{100}) is still too high to be practicable for some meaningful pattern recognition by the PRInt. However, the relative position of the $\#T$, L , H , $T_{\#}$ points can be expressed on the same scale in both dimensions between contours.

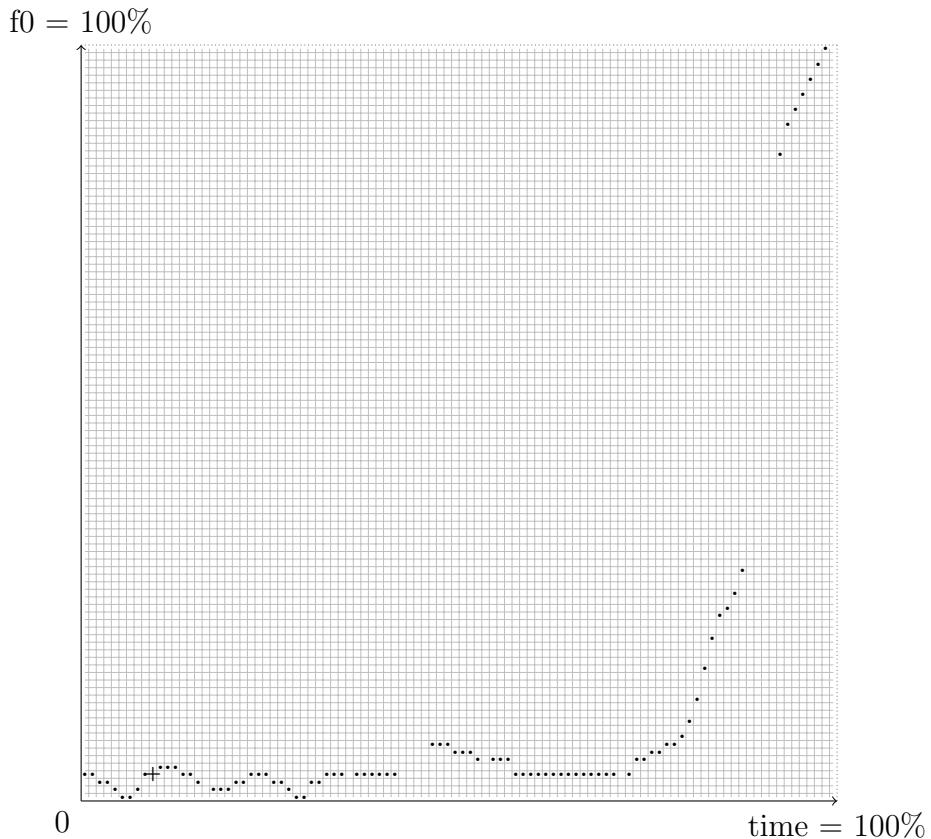


Figure 7.7: [JP72] represented as a 100 points by 100 matrix of points.
The activated points of the contour are in black.

The number of potential time points that have been realized is 87 for [JP72] and 99 for [GR65]. The initial point $\#T$ of contour [JP72] is realized as (1, 3) and as (2, 9) for [GR65]. These points are now abstractly almost at the origin of the sentence ($x = 0$). The highest H point of contour [JP72] is realized as (99, 100) and as (97, 100) for [GR65]. These points are now abstractly almost identical on a relative scale, both in time and F_0 . Very striking visually as well is the size of the non final H of contour [GR65] compared to

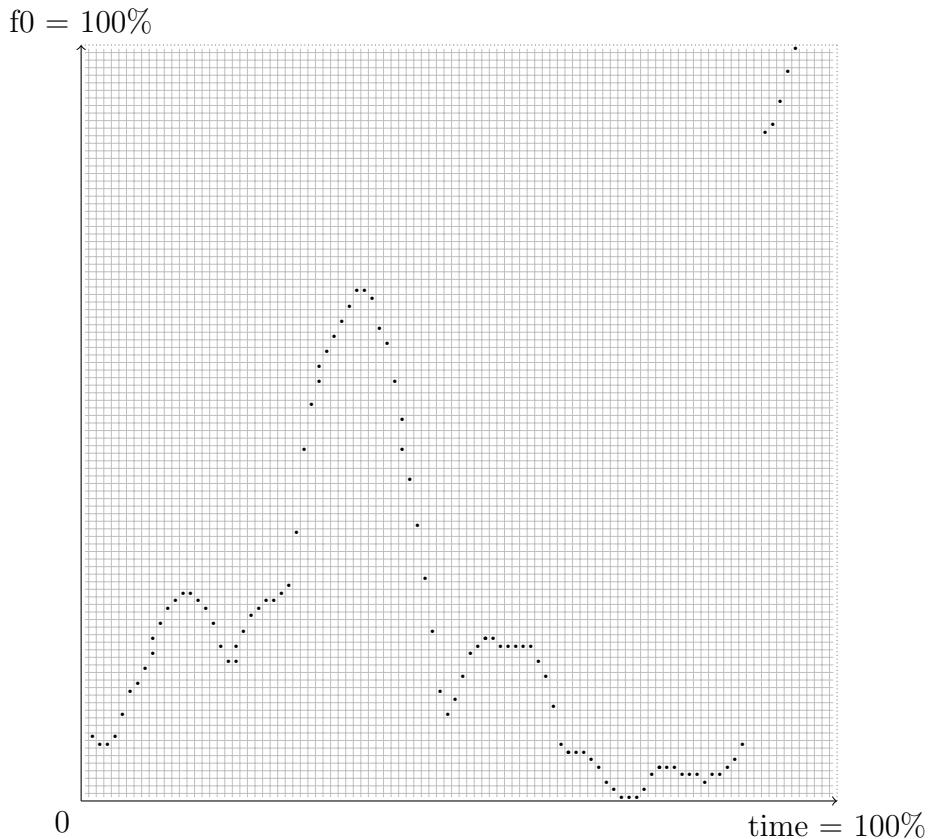


Figure 7.8: [GR65] represented as a 100 points by 100 matrix of points.
The activated points of the contour are in black.

its final H. Their relative proportion is magnified by scaling. The non-final H is realized as (41, 68), 32 points under the final H. Scaling makes possible the quantification of the relation between points (higher and lower, before and after) and comparable between sentences.

At this stage, if each contour were to be transcribed left to right as a string of L and H, the sentences would be a sequence of 87 and 99 points respectively. L being lower than the preceding point, H being higher than the

preceding point, and (-) being no change, the contours' transcription would look like that:

[JP72]:

```
#T L - L - L L H H H - - - L L L - H - H - - - L - L - H H - - -
- - - H - H L - - L - - L - L - - - - - - - - - H H - H -
H H H H H H H H H H H H H H H H H H H H H H H H H H H H H H H H H #
```

[GR65]:

```
#T - - - H H H H H H H H H H - - - L L L L - H H H H H - - H H H
H H H H H H H H - L L L L L L L L L L L L H H H H H - L - - -
- L L L L L - - L L L L L L - - H H H H - - L - L - - H H H H H H
H H H #
```

This poses two related problems. The most important one is that not only sentences do not have the same number of points but these points do not correspond from sentence to sentence. As a consequence of scaling (binning values as the same percentage by stretching or compression), point $x=36$ is realized in sentence [GR65] but not in sentence [JP72], and conversely point $x = 10$ is realized as two y values in [GR65] but only one in [JP72]. Secondly, even if the first problem was ignored or bypassed, the number of possible patterns (sequences of L and H) is far too large to allow the necessary rate of recurrence for pattern recognition. All sentences would be a different combination of a different number of hyphen-linked L and H points. Therefore, there is a need for the selection of a smaller subset of points that addresses these two issues: these points must be chosen for their nature or quality so that they form a class of points that can be found in all sentences, and, the number of these

points must be limited so that their possible combinations generate a higher, workable rate of recurrence of patterns.

7.3.3 Layer 3: isometric grid and pre-tones

At the second layer of the 4-layer structure, L and H points constitute two main classes: L is defined as having a lower F_0 value than the preceding point and H is defined as having a higher F_0 value than the preceding. In other words, the class of each point is defined by the preceding one, as a result of intonation being relative in nature. To create two classes of L and H points of that can be found in all sentences, the definition of the point entering these class must be external to the points. Instead of comparing points one by one sequentially left to right to determine whether each one is higher or lower than the points immediately before and after it, points can be compared in a larger time frame than the span of two adjacent points. Within each larger time frame, only the lowest point in the frame is L and only the highest point in the frame is H. Thus, the L or H nature of a points is defined by its position in a temporal structure rather than by adjacency.

The first issue is to determine the span of the time frame. An arbitrary fraction, as for example 10 points out of the 100 points, seems to be free of any theoretical assumption. However, there exists a more natural and non-arbitrary candidate for the choice of the frame span: the syllable. The choice of this number can be argued or changed, and more flexibility can be added. For the time being, a fixed number of syllables, whatever this number be, ensures that all sentences have phonologically the same structure at the level above the segments or phonemes. The segmental content of two sentences from a corpus used in this study, [JP72] and [GR65], is provided below. [JP72]

comprises 14 phonemes and [GR65] comprises 18 phonemes. These phonemes are spread differently among the seven syllables of each sentence.

[JP72]

/vuzabiteabɔ̃do/ (14 phonemes)

Vous habitez à Bordeaux, “you live in Bordeaux”

[GR65]

/vuvulekɔ̃vjenvuvwar/ (18 phonemes)

Vous voulez qu'on vienne vous voir, “you want us to visit you”

The syllabification of the sentences is as this:

	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	
[JP72]	→ [vu] $_{\sigma}$	[za] $_{\sigma}$	[bi] $_{\sigma}$	[te] $_{\sigma}$	[a] $_{\sigma}$	[bɔ̃] $_{\sigma}$	[do] $_{\sigma}$	
[GR65]	→ [vu] $_{\sigma}$	[vu] $_{\sigma}$	[le] $_{\sigma}$	[kò] $_{\sigma}$	[vjɛn] $_{\sigma}$	[vu] $_{\sigma}$	[vwar] $_{\sigma}$	

The segmental content of syllables differs in quality and quantity from syllable to syllable in a sentence and by equivalent syllable from sentence to sentence. Accordingly, the time span of each syllable varies. Note for example that even syllables with the same segmental content and pronounced by the same speaker can vary greatly in duration (syllable 1):

	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	
[JP72]	→ 4cs	12cs	10cs	18cs	9cs	16cs	26cs	
[GR65]	→ 16cs	13cs	13cs	10cs	14cs	11cs	31cs	

To give all sentences the same comparable isometric syllabic structure, the syllables are scaled to an equal fraction of the 100% of the duration of the

sentence or $\frac{100}{n \text{ syllables}}\%$. The adjustment of syllable duration to a unique value is crucial for fuzzification. The second pass of time scaling is achieved with the following function:

$$\text{Scaled } x = ((x - \sigma\text{-MIN}) \cdot (\sigma\text{-OUT}/\sigma\text{-IN})) + (\text{CUMUL.})$$

x : scaled time point expressed in %

$\sigma\text{-MIN}$: left boundary of the scaled syllable expressed in time %

$\sigma\text{-OUT}$: duration of the adjusted syllable expressed in %

$\sigma\text{-IN}$: input duration of the scaled syllable expressed in %

cumul.:cumulative adjusted duration of the preceding syllables

To understand the function it is necessary to have a few elements present in mind. First, the result of the second pass of time scaling is called the *adjusted* value of a point to distinguish it from the result of the first pass of time scaling, called the *scaled* value of a point. Second, the value to be adjusted is the output of the first pass of time scaling and already a relative value (a percentage of the actual value). Third, the *syllable* refers to the syllable containing the scaled point to be adjusted. Finally, the adjusted duration is equal to $\frac{100}{n \text{ syllables}}$ (or 14.28 scaled points in the case of the two examples). Therefore, the adjusted value (f_x) is the product of: the difference between the scaled value of the point and the scaled value of the left boundary of the syllable, and, the ratio of the adjusted duration of the syllable to its scaled duration, plus, the cumulated duration of all syllables before the current syllable.

Depending on the original size of the syllable, the points inside of it can be spread out if the syllable had a duration shorter than $\frac{100}{n \text{ syllables}}\%$, compressed if the syllable had a duration longer than $\frac{100}{n \text{ syllables}}\%$. Figure 7.9 below represents the two pass process of time scaling for [JP72]: on top, the actual durations of the syllables; in the middle, the sentence scaled to 100 points, at the bottom, the syllables adjusted to $\frac{100}{n \text{ syllables}}\%$.

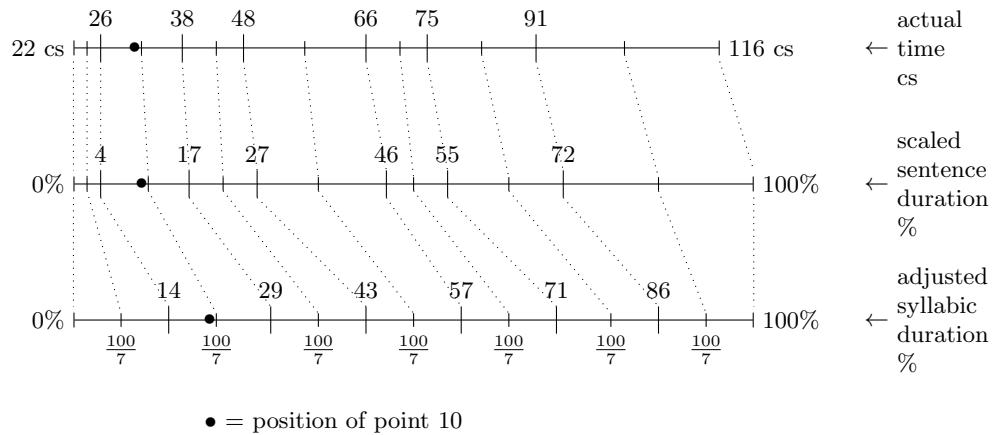


Figure 7.9: Syllables and half-syllable frames adjustment of [JP72]

Figure 7.9 illustrates the process. The time coordinates of point 10 of [JP72], marked as a dot on the figure, is scaled in the sentence (pass 1) and then adjusted in its syllable (pass 2). Its actual time value is 31 cs. It is located within the second syllable, 5 cs after its left boundary (= 26 cs). In the application of the model presented in this study, all sentences are seven syllable long. When the sentence is scaled to 100, the time value 31 cs of point 10 is scaled to 10 % of the duration of the sentence. The adjusted time value, in the adjusted syllable, is calculated as follows:

$$f_{10} = ((10 - 4) \cdot (\frac{100}{7}/12)) + (\frac{100}{7}) = 21$$

First the ATLM calculates the distance of the points from the left boundary of the syllable. The location of the boundary represents 4% of the sentence, spans 12% of it, and ends at 16% of it. Thus, point 10 is located at a distance equal to 6 % of the syllable from its left boundary. The ATLM calculates the proportion to which this scaled distance corresponds in the adjusted syllable by applying the ratio of the syllable duration to its projected scaled duration

$(\frac{100}{7} / 12)$: the distance of the point in its syllable is $6 * (\frac{100}{7} / 12) = 7\%$. The distance between the point and the left boundary of the adjusted duration of the syllable is 7%. This value is added to the cumulative duration of all the syllables preceding the one in which the point to be adjusted is located. The point is located in the second syllable and there is only one syllable before. Its position in the sentence is $7 + \frac{100}{7} = 21\%$. The results for this second pass of scaling, or syllabic adjustment, are in column 5 of Table 7.1.

At the third layer of the structure, all sentences are represented on the same *isometric syllabic grid*. Within each syllable, the ATLM selects the highest and the lowest points available, in the order they occur. The two points are selected for their relative height within the local temporal limit. These points are called *pre-tones* and are anchored to the temporal grid. To increase the resolution of the time anchoring by reducing dispersion, each syllable is divided into two equal *frames* (half-syllables). Instead of selecting two points per syllable, the ATLM selects four, one L/H pair in each syllable, two pairs per syllable. In those frames where there is no contrast between points, the ATLM selects the points on each edge of the frames. Having the syllables cut in two frames minimizes the dispersion of points in time. The range of possible positions for the two points in a frame is extremely limited by the short span of the frames. For example, in the present application, any of the two pre-tones in a frame can maximally take about 7 possible positions, since a frame is $\frac{100}{14} \simeq 7$ points long. The isometric grid, composed of a sequence of equal frames (14 for this study), is represented on Figure 7.10 as a comb-like structure in which each *prong* is a pre-tone, the highest or lowest point of a frame:

On the isometric grid each pre-tone ideally occupies only one position.

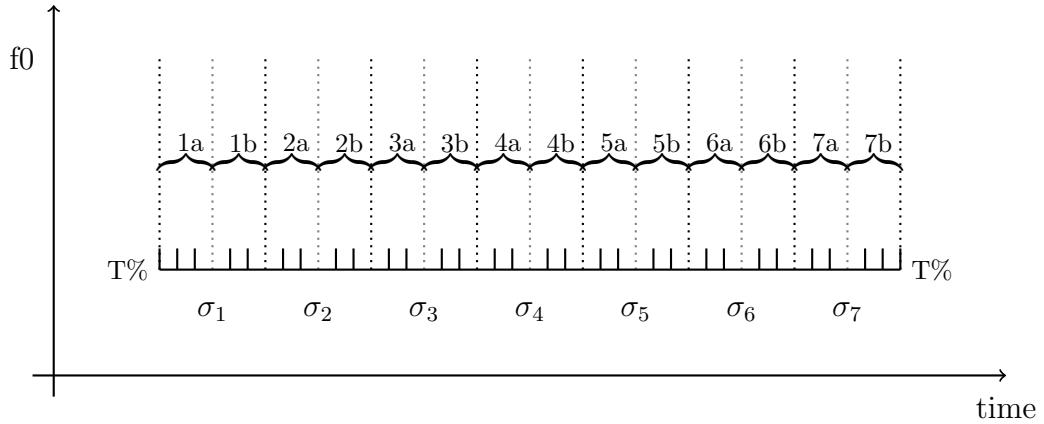


Figure 7.10: The isometric syllabic grid

Rather than being expressed as a percentage point on the [1, 100] interval of an instance scaled duration, the position of a pre-tone is attached to the phonological syllabic structure of the instance. On the isometric grid, all pre-tones are ideally equidistant because a pre-tone cannot be realized elsewhere but in the frame of the syllable to which it is anchored. Furthermore, the position of a pre-tone in its frame is irrelevant because first it is very limited and second, and more importantly, it is confined to that frame uniquely, that is, to that half-syllable only. In Figure 7.11, [JP72] and [GR65] are represented on the isometric syllabic grid, disregarding the time variation of the pre-tones in their frames.

The ATLM gives to each sentence in a corpus the same temporal structure made of an isometric syllabic grid of seven syllables and 14 frames, containing a total number of 30 pre-tones. There are two pre-tones per frame, plus a start pre-tone T% and an end pre-tone T%. Figures 7.12 and 7.13 are the contours of [JP72] and [GR65] after the ATLM has adjusted the syllables to an equal duration and after it has selected the two most contrastive points,

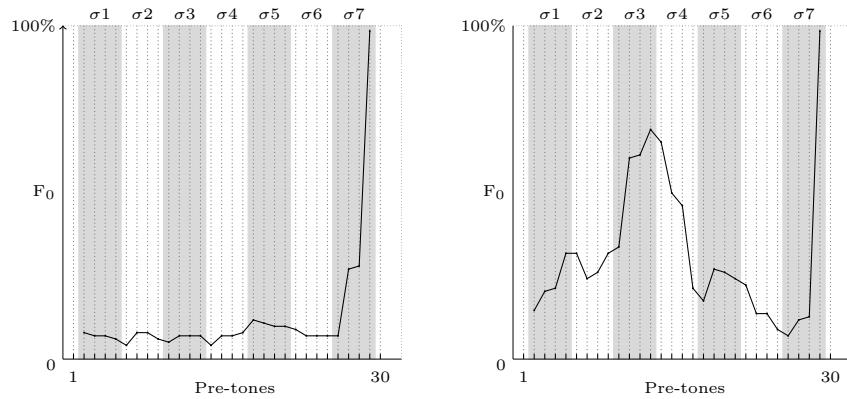


Figure 7.11: Isometric pre-tonal representation of [JP72] (left) and [GR65] (right).

the pre-tones, of each half-syllable, or frame. The pre-tones of [JP72] and [GR65] are presented in the tables below. In the first row, the two frames of each syllable are labelled with the number of the syllable followed by *a* (first frame) and *b* (second frame): 1*a*, 1*b*, ..., 14*b*. The time value of the pre-tone is in the second row, its F₀ value is in the third row.

[JP72]

	#T	1a	1b	2a	2b	3a	3b	4a								
	#T	H	L	H	L	L	H	H	L	H						
time	3	3	6	10	13	17	21	22	28	30	36	37	37	44	47	
F ₀	5	5	4	4	3	1	5	5	3	2	4	4	4	1	4	
		4b	5a	5b	6a	6b	7a	7b	T#							
		L	H	H	L	-	-	H	L	-	-	L	H	L	H	T#
time	51	51	54	59	61	65	65	72	74	79	79	86	92	93	99	99
F ₀	4	4	5	9	8	7	7	6	4	4	4	4	25	26	100	100

[GR65]

	#T	1a	1b	2a	2b	3a	3b	4a								
	#T	H	L	H	L	L	H	H	L	H						
time	3	3	6	10	13	17	21	22	28	30	36	37	37	44	47	
F ₀	5	5	4	4	3	1	5	5	3	2	4	4	4	1	4	
		4b	5a	5b	6a	6b	7a	7b	T#							
		L	H	H	L	-	-	H	L	-	-	L	H	L	H	T#
time	51	51	54	59	61	65	65	72	74	79	79	86	92	93	99	99
F ₀	4	4	5	9	8	7	7	6	4	4	4	4	25	26	100	100

[JP72] and [GR65] (and all sentences in a corpus) have been reduced to a string of 30 pre-tones. This class of points is defined by their relative position to the other points within a temporal limit, the frame. Pre-tones are a subset of scaled points that are either the lowest or the highest of their time frame. Contrary to scaled points whose position is defined by direct adjacency with other points, the relative position of a pre-tone is defined externally, by the syllabic (phonological) structure of the sentence. Therefore, a pre-tone in one sentence has a counterpart in all other sentences (provided all instances have the same number of syllables, or are aligned on one or the other sentence boundary - most likely the right edge of sentences). The pre-tonal combinations of [JP72] and [GR65] are comparable globally, by syllable, frame

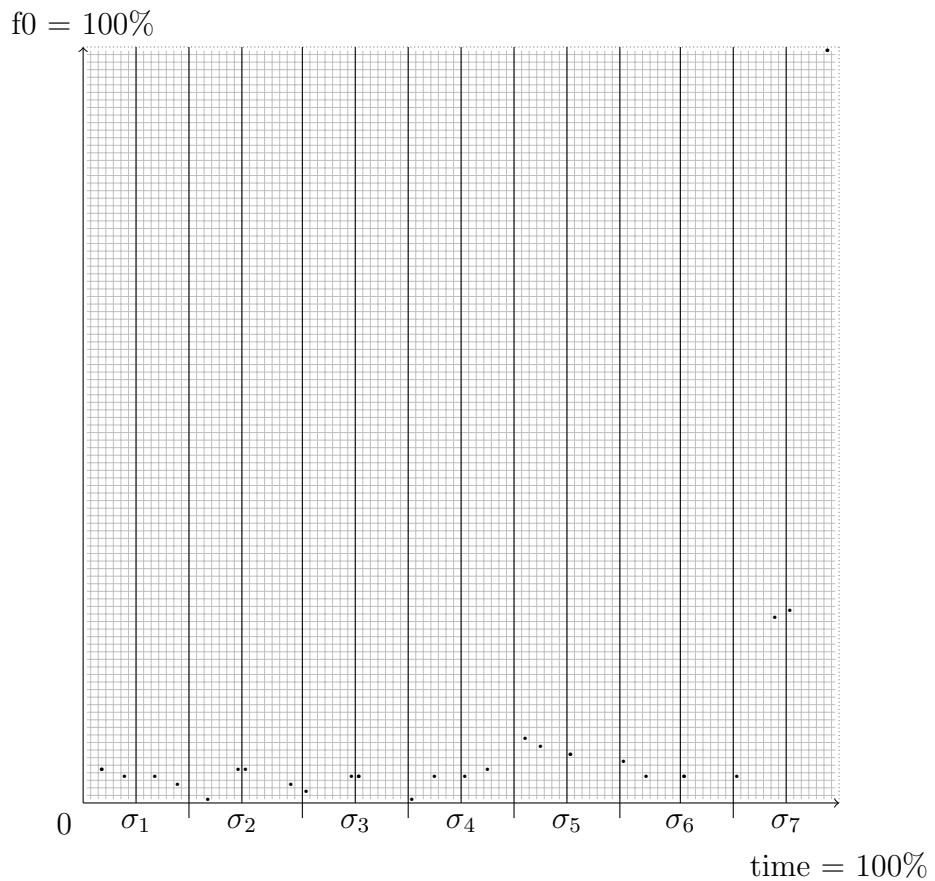


Figure 7.12: [JP72] represented as a 100 points by 100 matrix of points. The activated points in black are the pre-tones of the contour.

and pre-tonal class of each pre-tone. The number of possible combinations has been reduced to 100^{30} . It is still a large number but the ATLM can process the data of these pre-tones for pattern recognition.

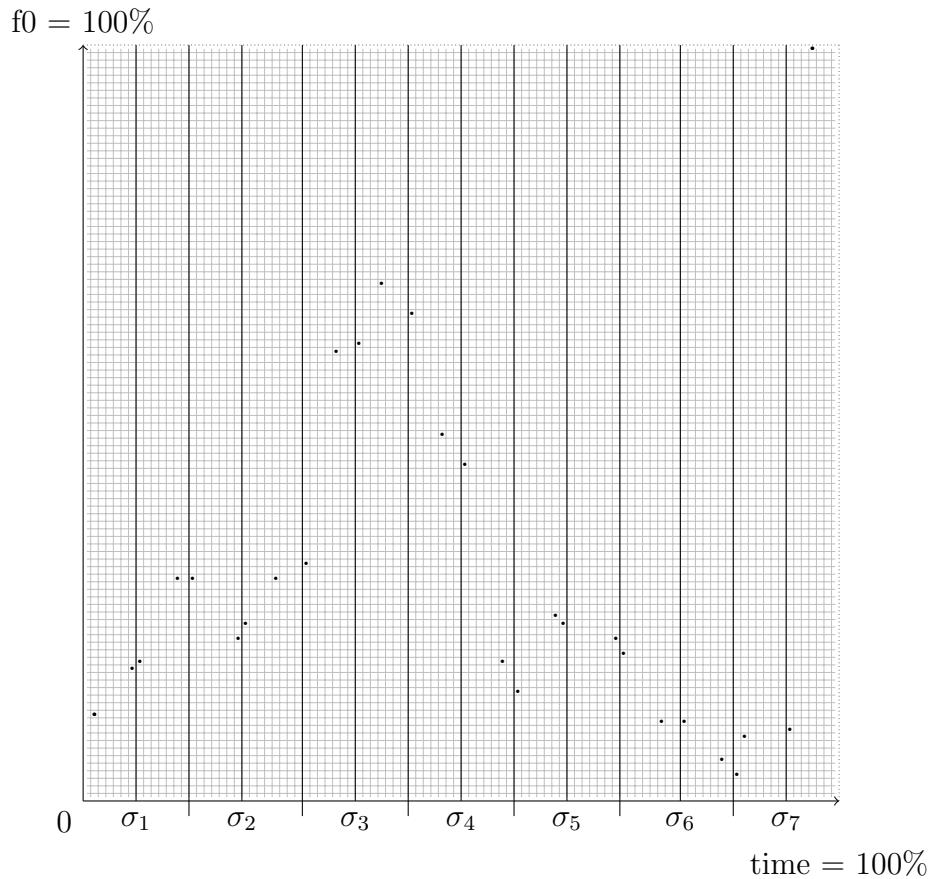


Figure 7.13: [GR65] represented as a 100 points by 100 matrix of points. The activated points in black are the pre-tones of the contour.

7.3.4 Layer 4: tones

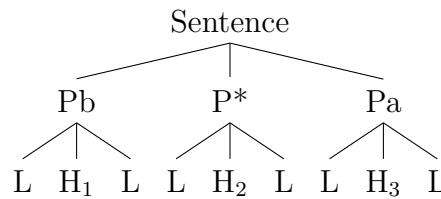
The last layer of the structure combines the information obtained in lower layers to detect the global string of *tones* that forms the intonation contour of the sentence. The tones are a subclass of pre-tones, themselves a subclass of points:

$$\{\text{tones}\} \subset \{\text{pre-tones}\} \subset \{\text{points}\}$$

or less formally: $\{\text{points} \ \{\text{pre-tone} \ \{\text{ tones }\}\}\}$

From the information generated by layers 2 and 3, the ATLM can find the subset of pre-tones corresponding to the tonal string of the sentence by using the fuzzy quantifiers. Layer 2 provides the information for the global quantification of *lower* vs. *higher* and layer 3 provides the information for the global quantification of *before* vs. *after*.

In the current application the ATLM looks for three tonal compounds in the intonation contour because sentences are only seven syllables long and no more than two main intonation movements are expected. Each compound is a combination of three pre-tones forming a peak and noted L-H-L. A subscript letter indicates the order of the contour, from 1 to 3. Accordingly, the ATLM searches for the three pre-tones with the three highest values. The highest and maximum F_0 peak of the string is noted P^* , the highest but non maximum F_0 peak after P^* is noted P_a , the highest but non maximum F_0 peak before P^* is noted P_b . The ATLM also searches for the lowest points before and after each of the peaks. The basic sentence structure at the tonal level (layer 4) is represented as follows:



The possible combinations of the peaks over the sentence depends on position of the primary peak P^* (as indicated by the pre-tonal anchoring to

the syllabic grid) and the presence or absence of the secondary peaks Pa and Pb. These combinations and their simplified graphic representations are:

	H_1	H_2	H_3		H_1	H_2	H_3		
$*(?)$	-	-	-	—	(d)	-	-	P^*	— \wedge
(a)	-	P^*	-	— \wedge	(e)	-	Pb	P^*	— $\wedge\wedge$
(b)	-	P^*	Pa	— $\wedge\wedge$	(f)	P^*	-	-	\wedge —
(c)	Pb	P^*	Pa	$\wedge\wedge\wedge$	(g)	P^*	Pa	-	$\wedge\wedge$ —

The presence of the first contour, noted $*(?)$ is unlikely but it completes the inventory. Contours (d) and (e) can also be realized as —/ and — \wedge / respectively, if there is no final downward slope. The same is true for contours (b) and (c): \wedge — and $\wedge\wedge$ —.

Application of the automated tone finder system to [JP72] In Table 7.2 results have been imported from the pre-tonal analysis of sentence [JP72]. This data will be the example to explain how the tonal string of the sentence is extracted from the information from the lower layers of the 4-layer structure analysis. The table has 11 columns and it is a copy of the Excel file, edited for legibility, on which the automated system has been implemented.

- ① This column contains the labels of the frames, from 1a to 7b.
- ② This column contains the labels of the pre-tones, from 1 to 30.
- ③ This column contains the label that indicates the position of each pre-tone relatively to its frame, T%, L, H, or T%.
- ④ & ⑤ These two columns contain the time and F_0 coordinates of each pre-tone, scaled (layer 2) and adjusted (layer 3).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
#	#	class	time	F ₀	{1,0}	↑	↓	Pa	Pb	tones	↓
	1	#t	3	5	-	0	5		5	T%	
1a	2	H	3	5	0	0	4		5		
	3	L	6	4	0	0	3		4		
1b	4	H	10	4	0	0	2		4		
	5	L	13	3	0	0	1		3		
2a	6	L	17	1	0	0	0		1		
	7	H	21	5	1	1	3		5		
2b	8	H	22	5	0	0	2		5		
	9	L	28	3	0	0	1		3		
3a	10	L	30	2	0	0	0		2		
	11	H	36	4	1	1	3		4		
3b	12	-	37	4	0	0	2		4		
	13	-	37	4	0	0	1		4		
4a	14	L	44	1	0	0	0		1		
	15	H	47	4	1	1	1		4		
4b	16	L	51	4	0	0	0		4	L-b	
	17	H	54	5	1	1	0		5		
5a	18	H	59	9	1	2	8		9	H-b	
	19	L	61	8	0	0	7		8		
5b	20	-	65	7	0	0	6		7		
	21	-	65	7	0	0	5		7		
6a	22	H	72	6	0	0	4		6		
	23	L	74	4	0	0	3		4		
6b	24	-	79	4	0	0	2		4		
	25	-	79	4	0	0	1		4		
7a	26	L	86	4	0	0	0		4	-L-b, L-	
	27	H	92	25	1	1	0				
7b	28	L	93	26	1	2	0			H-L	
	29	H	99	100	1	3	1				
	30	t#	99	100	0	0	0			-L, T%	

Table 7.2: Automatic detection of the tones of [JP72] from its pre-tones

Binary strings, up-string (\uparrow), and down-string (\downarrow)

To find the movements of large amplitude in the sentence, the PRInt calculates the local change from pre-tone to pre-tone by comparing the F₀ value of each pre-tone to the preceding one (column (6)). Then the ATLM scans these binary strings downward for ascending large movements (column (7)) and then right to left the descending large movements (column (8)).

- ⑥ The ATLM assigns a binary mark {0, 1} to each pre-tone, depending on its F_0 value relative to the preceding pre-tone and disregarding frame boundaries. If the F_0 value of the pre-tone is greater than that of the preceding pre-tone, it is assigned value 1. If its F_0 value is equal or lower than that of the preceding pre-tone, it is assigned value 0.

$$\left. \begin{array}{ll} y > y_{-1} \rightarrow 1 \\ y \leq y_{-1} \rightarrow 0 \end{array} \right\} \text{if } y > y_{-1} \text{ then 1 else 0}$$

- ⑦ Up strings (\uparrow): The ATLM scans the values of ⑥ downward (first to last pre-tone). When it encounters a 1, it starts a summation of a consecutive 1 until a zero occurs. The summation starts again when a 1 occurs. Pre-tones are assembled into strings as long as their F_0 keeps increasing or remains at the same high value.
- ⑧ Down strings (\downarrow): The ATLM scans the values of ⑥ upward (last to first pre-tone). When it encounters a 0, it starts a summation of a consecutive 0 until a 1 occurs. The summation starts again when a 0 occurs. Pre-tones are assembled into strings as long as their F_0 keeps decreasing or remains at the same low value.

[Note: the process is shifted by one cell above to account for the fact that binary marks are only assigned downward in column ⑥. The mark for a pre-tone actually reflects the binary evaluation of the pre-tone below. For example, pre-tone 19 is marked 1 in column ⑥ because its F_0 value is higher than that of pre-tone 17. If the upward summation of zeroes in column ⑧ was not shifted up by one cell, it would start at pre-tone 26 up to pre-tone 19 and would be reset to 0 at pre-tone 18. The downward pre-tonal strings would be shifted by one pre-tone up.]

In column ⑦ the PRInt has formed two main ascending strings: 3 pre-tones (16-18) and 4 pre-tones (26 -29). The second string is longer than

the first one. In column ⑧ the ATLM has formed four main descending strings: 6 pre-tones (1-6), 4 pre-tones (7-10), 4 pre-tones (11-14), and 9 pre-tones (18-26). The last string is the longest of all.

Peaks and associated tonal strings (L-H-L)

The ATLM is set to find the three peaks P^* , Pa , and Pb sequentially. Each peak is composed of a high tone H , a preceding lower tone $L-$, and a following lower tone- L . These nine tones (three per peak) are a subset of the pre-tones. With the information contained in Table 7.2, the ATLM can find the largest movements based on their relative size. The ATLM uses the binary information from layer 2 and especially layer 3 to form the binary strings from which it computes the gradient size of the movements (in time and F_0) from sentence to sentence. Thus, the variations in shape and size of the intonation contour from sentence to sentence can be analyzed for patterns.

For each of the three peaks, once the pre-tonal position of H has been calculated, the ATLM calculates the position of the two associated low tones $L-$ and $-L$ in the same way. Before turning to the example, here is the general mode of calculation. The terms of the calculations are:

1. h is the number of the pre-tone on which H is realized
2. $l-$ is the number of the pre-tone on which $L-$ is realized
3. $-l$ is the number of the pre-tone on which $-L$ is realized
4. u is the value corresponding to h found in column ⑦ (binary up-string)
5. d is the value corresponding to h found in column ⑧ (binary down-string)

Pre-tone h is labelled u in the up-strings column ⑦ and d in the down-strings column ⑧. This means that:

1. The pre-tone with the lowest F_0 value before pre-tone h is n pre-tonal positions before pre-tone h . $h - u = l_-$.
2. The pre-tone with the lowest F_0 value after pre-tone h is d pre-tonal positions after pre-tone h . $h + d = l_+$.

The application of these calculations to example [JP72] illustrates how the ATLM proceeds to determine what pre-tones are the tones of the intonation contour of a sentence.

Primary peak P^* The ATLM finds the pre-tone with the highest F_0 value ($F_0 \text{ MAX}$, with $F_0 = 100$ as a result of scaling at layer 2). In the example, this P^* is pre-tone 29, in frame 7b. From the information in columns ⑦ and ⑧, the ATLM calculates which pre-tones correspond to L_- and $-L_+$. Pre-tone 29 is labelled 3 in the up-strings column ⑦ and 1 in the down-strings column ⑧. This means that:

L_- : the pre-tone with the lowest F_0 value before pre-tone 29 is three pre-tonal positions before pre-tone 29. $29-3 = 26$, L_- is realized as pre-tone 26.

$-L_+$: the pre-tone with the lowest F_0 value after pre-tone 29 is zero pre-tonal positions after pre-tone 29. $29+1 = 30$, $-L_+$ is realized as pre-tone 30.

The (L_-H-L_+) string of P^* is realized as pre-tones 26-29(-29). (H) and ($-L_+$) are merged.

Secondary tonal string after the main string (Pa) After the ATLM has calculated the pre-tonal positions (L-H-L) of the primary peak P^* , it calculates the positions of the secondary peak after P^* . To do so, the ATLM only looks into the values of the pre-tones after the last tone -L of P^* (= pre-tone 30)

- ⑨ In this column, only the F_0 values of the pre-tones after the last pre-tonal position of the primary peak (i.e. tone -L of P^*) are copied, including this last pre-tone. For [JP72], there is no pre-tone left after the final low tone (= pre-tone 30) of the primary movement. Pa is not realized in [GR72]

Secondary tonal string before the main string (Pb) After the ATLM has calculated the pre-tonal positions (L-H-L) of the primary peak P^* and of the secondary peak Pa, the ATLM calculates the positions of the secondary peak before P^* . To do so, the ATLM only looks into the values of the pre-tones before the first tone L- of P^* (= pre-tone 26)

- ⑩ In this column, only the F_0 values of the pre-tones before the first pre-tonal position of the primary peak (i.e. tone L- of P^*) are copied, including this first pre-tone. The H tone of Pb is pre-tone 18 since its F_0 value is the maximum value ($F_0 \text{ MAX} = 9$) of all pre-tones before or equal to pre-tone 26. From the information in columns ⑦ and ⑧, the PRInt calculates which pre-tone(s) correspond to L- and -L of Pa. Pre-tone 18 is labelled 2 in the up-strings column ⑦ and 8 in the down-strings column ⑧. This means that:

L-: the pre-tone with the lowest F_0 value before pre-tone 18 is 2 pre-tonal positions before pre-tone 18. $18-2 = 16$, L- is realized as pre-tone 16.

-L : the pre-tone with the lowest F_0 value after pre-tone 18 is 8 pre-tonal positions after pre-tone 18. $18+8 = 26$, -L is realized as pre-tone 26.

The (L-H-L) string of Pb is realized as pre-tones 16-18-26.

- (11) In the last column of the table, all the tones of the intonation contour of [JP72] have been labelled. The ATLM assembles the string of tones according to the three possible positions for the main peak P^* in the sentence: initial (H_1), central (H_2), or final (H_3). The position of the peak corresponds to pattern (e), a sentence final primary peak with a secondary peak before the final peak. Table 7.3 is a summary of the 4-layer structure of [JP72]:

[JP72] _ Pb P^* 

		H_1				H_2				H_3					
Position															
Layer 4	Peaks														
	T%	T%	L-	H	-L	L-	H	-L	L-	H	-L	T%			
	Layer 3	Pre-tones	1	-	-	-	16	18	26	26	29	29	30		
Layer 2		x	3	-	-	-	51	59	86	86	99	99	99		
		y	5	-	-	-	4	9	4	4	100	100	100		

Table 7.3: The 4-layer structure of [JP72]

The H and -L tones of P^* are merged since they are realized as the same pre-tone (= 29). Furthermore, the time and F_0 values of these last two tones of P^* and those of T% are merged. Thus, H, -L, and T% are merged as a single tone in the contour. The same is true for the -L of Pb and the L- of P^* . They are merged in the same pre-tone (= 26). Table 7.4 is a summary of the 4-layer structure of [GR65]. The intonation contour [GR65] also corresponds

to pattern (e), a sentence final primary peak with a secondary peak before the final peak. The three last tones L-, H, and T% are merged:

[GR65] _ Pb P* 

	Position		H_1			H_2			H_3			
Layer 4	Peaks		-			Pb			P*			
	Tones	T%	L-	H	-L	L-	H	-L	L-	H	-L	T%
Layer 3	pre-tones	3	-	-	-	7	13	18	26	29	30	30
Layer 2	x	2	-	-	-	21	40	57	86	97	97	97
	y	12	-	-	-	22	69	15	4	100	100	100

Table 7.4: The 4-layer structure of [GR65]

7.3.5 Tonal isometric grid

Finally, with this information, the two tonal contours can be described and compared by equivalent features on the tonal isometric grid.

The tonal isometric grid enables the PRInt model to express the position of the tones in a sequence without resorting to actual time and F_0 anymore but as a relational structure abstracted from the actual values by the use of vague quantifiers: lower/higher, after/before. The pre-tonal isometric grid is completed by the addition of three tonal levels. H^* is the level of the sentence's highest point from which the position of all other points in the pattern is determined as lower and before/after. H is the level of the highest points lower than H^* before (H_b) or after (H_a). L is the level of the lowest points before and after the H and H^* points. The two sentences [JP72] and [GR65] are represented on the tonal isometric grid (Figure 7.14). Both sentences have the same tonal structure: $L\% \text{ L-H-(L L)-H}^*\text{H}\%$, where the bracketed tones

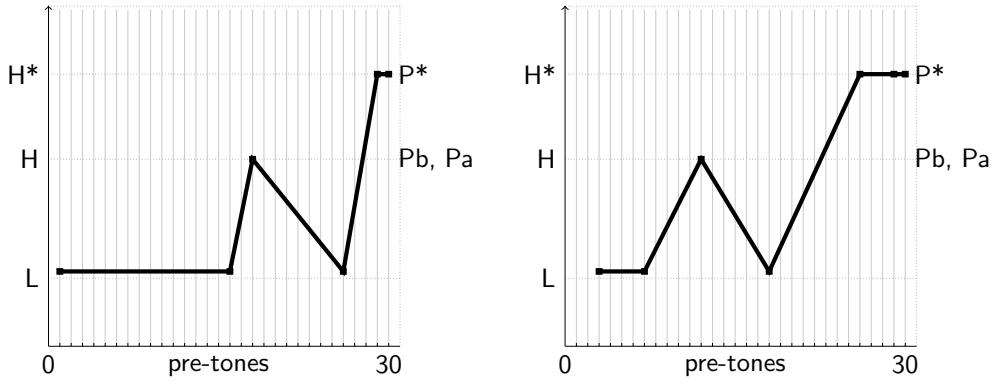


Figure 7.14: [JP72] (right) and [GR65] (left) on the tonal isometric grid. Both sentences have the tonal structure $L\#L-H-(L)-H^*H\%$

indicate they are the same tone serving as -L and -L.

However, although the ATLM attributes the same tonal structure to the two contours, it also records the variations in F_0 and time amplitude or how a same string of tone is phonetically implemented on different sentences. The last peak P^* of both sentences is anchored on pre-tones 26 and 29: over the span of the last syllable. The secondary movements differ in timing (pre-tonal anchoring) and in F_0 amplitude. The secondary peak of [JP72] is anchored between the middle of the fourth syllable and the beginning of the seventh syllable with its peak on the beginning of the fifth syllable. The secondary peak of [GR65] is anchored earlier than for [JP72]: between the middle of the second syllable and the beginning of the fifth syllable with its peak on the end of the third syllable. The F_0 amplitude of the secondary movement of [GR65] ($= 69\%$) is much greater than that of [JP72] ($= 9\%$).

Summary The ATLM converts all utterances in a corpus into comparable vectors of features, or in Pierrehumbert’s words, *the crucial points in the contour, the F₀ targets*. The PRInt can now analyze ***how the same intonation pattern lives up with different texts***, by analyzing the variations in the organization of the features of the contour in terms of F₀ height and alignment with the segmental level (how F₀ targets are *lined up with crucial points in the text, with stretches in between computed accordingly*). To conclude this chapter, here is a summary of how the ATLM, as implemented on Excel, operates with each sentence in a corpus:

Layer 1 The initial data from Praat is entered in the ATLM module. These data include the F₀ listing (time/F₀), and the time of the syllable boundaries.

Layer 2 The ATLM resets the sample origin from its values in the raw cutting from the recording to (0, 0) and scales its time and F₀ values to 100.

Layer 3 The ATLM calculates the span of each syllable and frame (half syllable). It labels all time/F₀ coordinates according to the syllable that contains them. It re-scales these coordinates to adjust them to isometric syllables, each equal to $\frac{100}{7}$ points out of the 100 points of the total sample. In each isometric frame the ATLM finds the highest and lowest F₀ point. These 14 pairs of coordinates are the pre-tones of the sentence. Two additional pre-tones are then located: the starting and ending points of the sentence. The 30 pre-tones are labelled 1 to 30.

Layer 4 The ATLM finds 11 tones in the sentence. The starting and ending pre-tones are the two boundary tones of the sentence T%. The other nine tones are those pre-tones whose positions contrast the most at the sentence level: the highest of the 30 pre-tones, and the lowest pre-tones

before and after it, constitute the primary L-H-L string or peak P^* . The highest peak before and after this P^* constitute the two secondary peaks P_b and P_a , each of them also a L-H-L string. The ATLM merges the information of the three peaks to form the tonal sequence of the sentence on the tonal isometric grid.

The successive extraction of 1) the subset of pre-tones from the subset of points, and 2) the subset of tones from the subset of pre-tones, is illustrated for sentences [JP72] and [GR65] by Figures 7.15 and 7.16. On these figures, each layer is sequentially extracted from the lower level. As a point of comparison, the two sentences have been passed through the MOMEL algorithm of Hirst & Espresser (1993) (as implemented on Praat as a script). This algorithm relies on a quadratic spline function to model intonation (fundamental frequency) automatically. The results are presented in Figures 7.17 and 7.18. The macro-melodic analysis of the algorithm is superimposed over the F_0 curve as a polynomial function (the extra segment of the spline beyond the F_0 contour must be disregarded). The ATLM performs a few additional calculations. For each peak P^* , P_b , and P_a , the ATLM calculates the distance in scaled time, scaled F_0 , and number of pre-tones between each of the three tones L-, H, and -L. The ATLM also calculates the velocity of the F_0 movement from L- to H and from H to -L. Finally, it calculates the angle or sharpness of the L-H part of the primary peak P^* . The output of the ATLM is not a graph but arrays of values containing the information of the intonation contour of the sentence in terms of pre-tones and tones. The array of [JP72] and [GR65] are presented in Tables 7.5 and 7.6, transcribed from the PRInt model results. Graphs are plotted from these arrays.

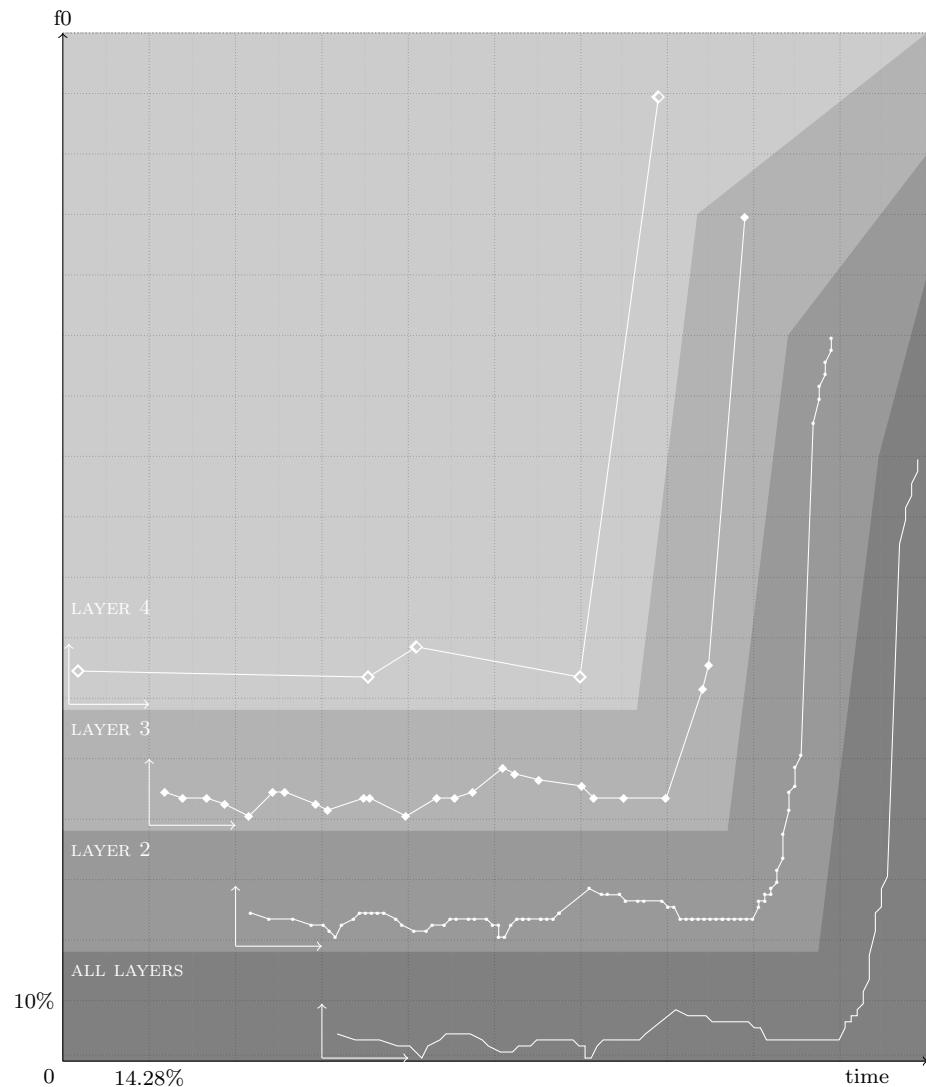


Figure 7.15: [JP72] represented as a 100×100 point matrix. Each layer is a subset of the layer(s) above: $\{\text{tones}\} \subset \{\text{pre-tones}\} \subset \{\text{points}\}$

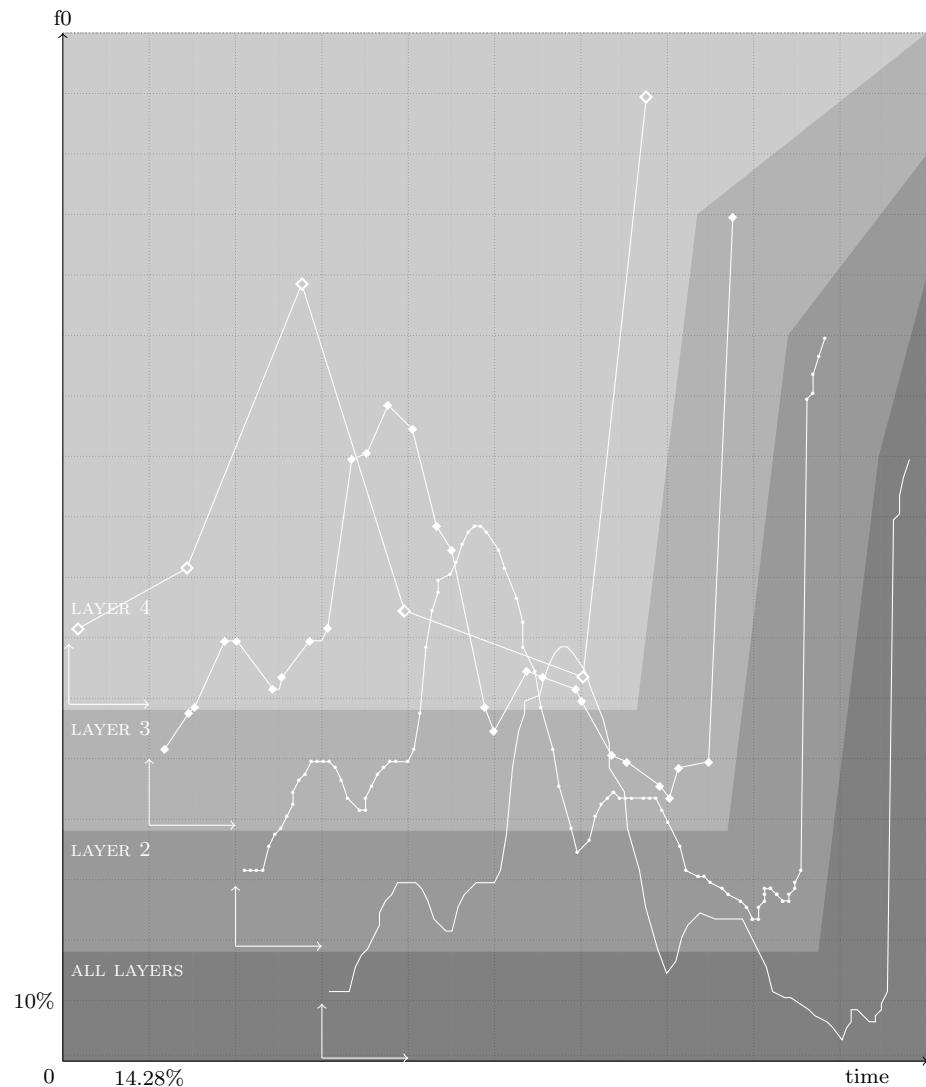


Figure 7.16: [GR65] represented as a 100×100 point matrix. Each layer is a subset of the layer(s) above: $\{\text{tones}\} \subset \{\text{pre-tones}\} \subset \{\text{points}\}$

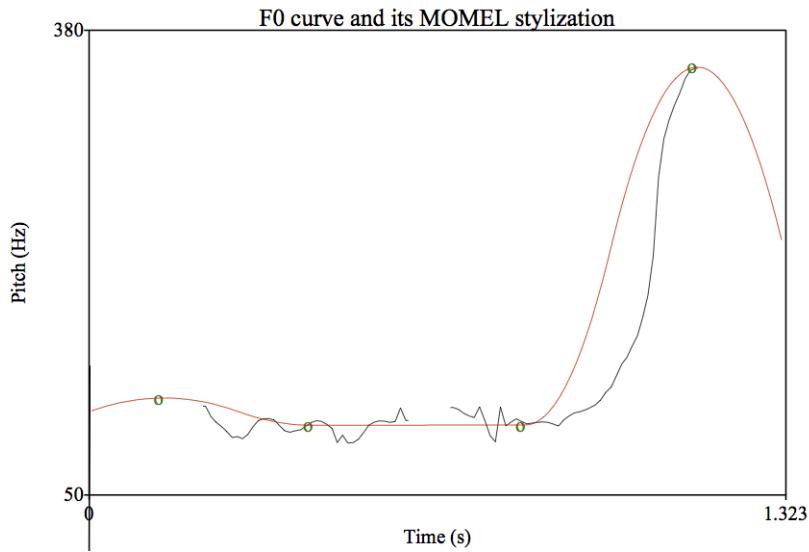


Figure 7.17: [JP72]: Output of the MOMEL algorithm

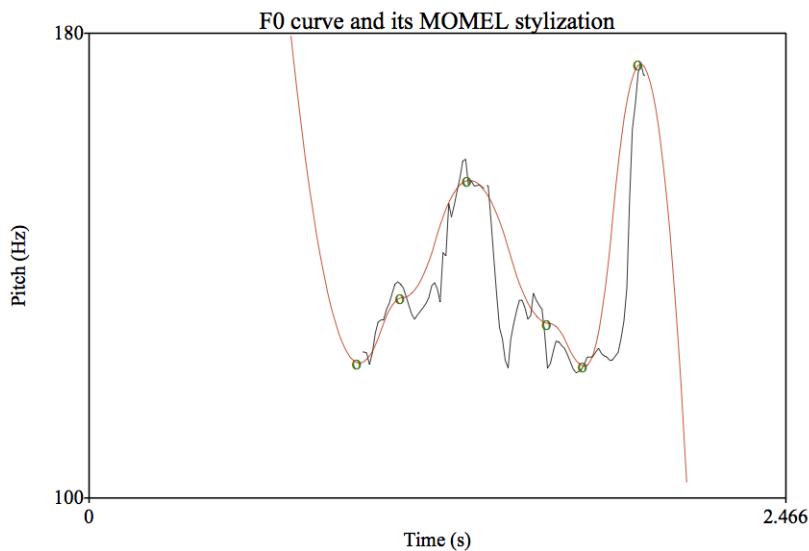


Figure 7.18: [GR65]: Output of the MOMEL algorithm

Time	Hz	pre-tones	P*	Pa	Pb	T
3	5	1				T
3	5	2				
6	4	3				
10	4	4				
13	3	5				
17	1	6				
21	5	7				
22	5	8				
28	3	9				
30	2	10				
36	4	11				
37	4	12				
37	4	13				
44	1	14				
47	4	15				
51	4	16			L	
54	5	17				
59	9	18			H	
61	8	19				
65	7	20				
65	7	21				
72	6	22				
74	4	23				
79	4	24				
79	4	25				
86	4	26	L		L	
92	25	27				
93	26	28				
99	100	29	H			
99	100	30	L			T

		Distance			velocity	Angle
		time	F ₀	pretones		
Pb	L-H	8	5	2	1	
	H-L	27	5	8		
P*	L-H	13	96	3	7	82
	H-L	0	0	1		
Pa	L-H					
	H-L					

Table 7.5: Output array of the ATLM for sample [JP72] from the PRInt model

Time	Hz	pre-tones	P*	Pa	Pb	T
2	12	1				T
2	12	2				
7	18	3				
8	19	4				
13	30	5				
15	30	6				
21	22	7			L	
22	24	8				
26	30	9				
30	32	10				
34	60	11				
36	61	12				
40	69	13			H	
44	65	14				
48	49	15				
50	45	16				
56	19	17				
57	15	18				
63	25	19				
64	24	20				
71	22	21				
72	20	22				
77	11	23				
78	11	24				
85	6	25				
86	4	26	L		L	
88	9	27				
93	10	28				
97	100	29	H			
97	100	30	L			T

		Distance			velocity	Angle
		time	F ₀	pretones		
Pb	L-H	19	47	6	2	
	H-L	17	54	5	3	
P*	L-H	11	96	3	9	83
	H-L	0	0	1		
Pa	L-H					
	H-L					

Table 7.6: Output array of the ATLM for sample [GR65] from the PRInt model

Chapter 8

Automated Fuzzification Classifier (AFC)

8.1 Features

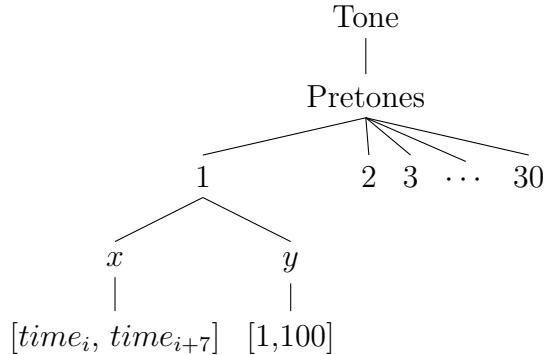
Features are functions of the measurements performed on a class of objects that enable that class to be distinguished from other classes in the same general category (Nadler & Smith, 1993).

Once the ATLM has processed all instances in a corpus, they become an identical feature vector whose dimension vary from instance to instance. These features are summarized in an array similar to Table 7.5 and 7.6, pages (163 and 164, Chapter 7). In the 4-layer structure, the potential range of variation of each feature is numerically expressed as the number of features of the lower layer.

Mid-level features Pre-tones are a subset of time and F_0 coordinates (low-level features). Thus, a pre-tone can vary in time and F_0 values.

- In time, it can take any of the seven values in the interval between the two boundaries of its time frame, $x = [t_n, t_{n+7}]$, with t as the first time value in the time frame (each time frame is equal to $\frac{100}{14} \simeq 7$ points). The implementation of the isometric grid renders time variation almost irrelevant.
- In F_0 , it can take any value in the interval between 0 and 100, $y = [1,100]$.

High-level features Tones are a subset of pre-tones. Thus, a tone can vary in terms of what pre-tone it is implemented on. It can be any pre-tone in the set of pre-tone = {1,..., 30}. Variation in the pre-tonal anchoring of tones is important since it encodes the alignment of tones on the syllabic structure of the instance. Finally, peaks are the special subset of tones that includes only the three H tones. The position of the primary and secondary peaks is crucial to the contour as a whole. The variation from tones to points can be schematically represented by the following tree:



Minimally, each instance of an intonation contour is a vector of 44 layered features: 3 peaks, 11 tones, and 30 pre-tones. Aside from the basic features, relations of features include the L-H and H-L distance of each peak, the velocity of F_0 in the L-H and H-L segment of each peak, and the angle of the L-H segment of the primary peak.

The Automated Fuzzy Classifier (AFC) does not analyze an intonation contour from the variation of its instances in a corpus. Instead, the AFC computes separately the variation of each feature of an intonation contour among its instances in a corpus. The model assumption is that what distinguishes instances as a whole is the addition of the individual differences in the value of

each feature. The two contours [JP72] and [GR65] are presented on the pre-tonal isometric grid (Figure 8.1 below) so that their features can be visually contrasted:

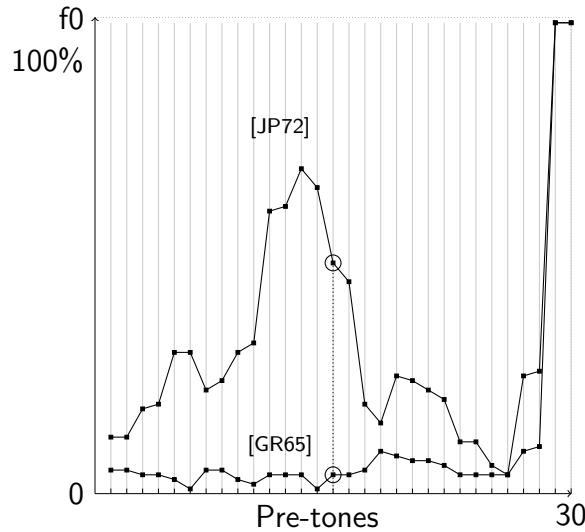


Figure 8.1: Comparison of [JP72] and [GR65] on the isometric grid.
Pre-tone 15 has been circled on both contours: “the variation in shape of the contour is ultimately and solely conditioned by the variations in F_0 of each pre-tone”.

Note how the two instances of contour exactly match pre-tone by pre-tone on the iso-metric grid. The most striking aspects of this comparison are the perfect match of the final movements and the obvious contrast in the presence of the secondary peak on [GR65]. This graph summarizes how, **with the crucial analysis of all instances of a contour as a vector of pre-tones implemented on an isometric grid, the variation in shape of the contour is ultimately and solely conditioned by the variations in F_0 of each pre-tone. Once the variation of each pre-tone is known,**

the variation of the contour is the concatenation of the variation of each pre-tone. The variation in alignment and F_0 of the features in the higher layer, the tones, are inherited directly from the variation of the F_0 value of the pre-tones of which they are a subset.

To calculate the variation of the contour, the values for each pre-tones must be gathered into sets so that they can be entered into the AFC.

8.2 Sets

The output data of the ATLM is consolidated into one set of tables in Excel. For each corpus of instances, the AFC creates a total of 110 multisets (sets within which values can occur more than once).

1. 99 multisets of values relative to features:
 - (a) time values (x) of each pre-tone → 30 sets
 - (b) F_0 values (y) of each pre-tone → 30 sets
 - (c) pre-tones realized for each tone → 11 sets
 - (d) P* whether it is realized as an initial, central, or final peak → 3 sets (3 positions)
 - (e) L-H and H-L distances in time value → 6 sets (3 peaks * 2 segments)
 - (f) L-H and H-L distances in F_0 value → 6 sets (3 peaks * 2 segments)
 - (g) L-H and H-L distances in pre-tones → 6 sets (3 peaks * 2 segments)
 - (h) velocity of F_0 in the L-H and H-L → 6 sets (3 peaks * 2 segments)
 - (i) angle of the L-H segment → 1 set
2. 11 multisets of values relative to the frame of reference:
 - (a) total duration of instances → 1 set

- (b) duration of syllables → 7 sets
- (c) maximum F_0 value of instances → 1 set
- (d) minimum F_0 value of instances → 1 set
- (e) span of F_0 of instances → 1 set

8.3 The Automated Fuzzy Classifier

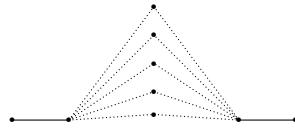
A multiset from one of the corpora used in this study is used to illustrate the methodology. It is the multiset of the F_0 values of all instances of pre-tone 15, that is, the y values of pre-tone 15 in all instances in the corpus of the intonation contour. This pre-tone has been circled in instances [JP72] and [GR65] on figure 8.1. It is chosen arbitrarily for the sake of the demonstration (and it is located in the middle of the set of pre-tones).

8.3.1 Why fuzzy sets? Why fuzzification?

The F_0 value of pre-tone 15 in one instance can only be one of the value in the interval [1, 100]. The set Y of the F_0 values of all the instances of pre-tone 15 in the corpus comprises 1981 elements.

In a classical binary set, all elements have the same grade of membership. They are inside the set (1) or they are not (0). Each element has the same importance, the same weight in the set. For the multiset of F_0 values of a point that is part of an intonation contour, it is inconceivable that all values affect the overall contour similarly. Consider a simple contour made of five pre-tones. The F_0 value of all pre-tones is 1, except for the third pre-tone, in the middle of the contour, that will be conveniently called pre-tone 15. The value of pre-tone 15 can take any value in the multiset Y , i.e. almost any value between 2 and 100. On the following image, the pre-tone 15 assumes an

F_0 value of 2, 25, 50, 75, and 100% of the F_0 range:



The shape of the contour changes dramatically from the lowest value to the highest one, leading to, and depending on the range of variation: 1) a (free) variation of the contour not affecting its meaning (stylistic, idiosyncratic, accidental from a particular phoneme), 2) a meaningful variation of the contour (gradient meaning) or 3) another contour altogether, with a different meaning. Independently of its cause, it is clear that a change in F_0 of any pre-tone alters the general shape of the contour. Therefore, values in a set are not equal, they cannot share the same grade of membership. However, it must be decided what values are the expected values for a given pre-tone in a contour, which ones are accidental or contrastive, and how the other values are organized in between.

The Automated Fuzzy Classifier relies on the assumption that there is an organization in the multiset of values and that not all elements in a multiset are necessarily equal in term of membership but can be ranked. The participants in the study were asked to target an intonation contour and consequently each instance they produce is a variation of this contour. There might be a general trend of realization and some variations around this trend, whether intentional or accidental. Since the ATLM analyzes the instances of a contour as a feature vector, each feature follows the trend or lack of accordingly. The AFC relies on two principles to identify the trends among the elements of the multisets, paralleling the assumed trend among speakers. First, it ranks the values according to their frequency in the multiset, the most frequent values being in the range of those the speakers are targeting. Second, it ranks the values towards the central tendency, the speakers targeting a certain range of

value and other values being more peripheral. The AFC attributes a grade of membership or typicality to each value in its multiset according to this two ranking principles. The former is called the *frequency principle*, the latter is called the *similarity principle*. After the AFC has graded all values in a multiset by frequency and similarity, the AFC assigns them the mean of these two grades as their a unique grade.

8.3.2 Fuzzification 1 of 2: the frequency principle

The AFC calculates the grade of membership of the values in the set according to their frequency of recurrence. The procedure comprises the following sequence of steps:

1. For each recurring value in the multiset S, the AFC counts how many times n it occurs, $N\{n_a, n_b, n_c, \dots\}$.
2. The AFC find the F_0 value(s), noted y , of the multiset with the largest count of instances (n_{max}) or the mode.

$$\exists x: n_y = n_{max}$$

3. The AFC assigns the highest grade of membership $m_{(y)} = 1$ to the most frequent value(s).

$$n_y = n_{max} \rightarrow m_{(y)} = 1$$

If the distribution has more than one mode, all modes are equally ranked within the set.

4. The AFC calculates the grade of membership of the other values in the set as a the ratio of their frequency to that of the most frequent value.

$$m_{(y)} = n_{max} / n_y$$

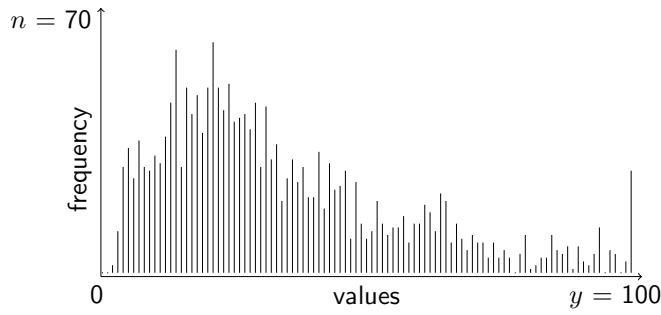


Figure 8.2: Distribution of the F_0 values of pre-tone 15 in the set Y

The set Y of all the F_0 value y of pre-tone 15 in the corpus of unmarked closed questions will serve to illustrate the procedure. The distribution of F_0 values of pre-tone 15, or multiset Y , is presented in Figure 8.2, and the numbers are provided in Table 8.1. The procedure comprises the following sequence of steps:

1. The AFC counts how many time each value is reoccurring in the multiset Y . This count (n) defines the frequency of each represented value. Table 8.1 below indicates the count (n) of each value (y) in the set.
2. The AFC finds that the frequency of value $y = 21$ is $n = 61$ ($y = 21$ represents 61 elements in the set) and $y = 21$ is the most frequent value in the multiset Y .

$$n_{max} = n_{21} = 61$$

3. The AFC assigns the highest grade of membership to value y .

$$n_{21} = n_{max} \rightarrow m_{(y)} = 1$$

4. The AFC assigns a grade of membership to all other values in the multiset. This grade is the ratio of the frequency of the values to that of the most frequent value.

$$m_{(y)} = n_y/n_{21}$$

For example, values $y = 4$ has a count of $n = 28$ and $y = 63$ has a count of $n = 11$. The grade of membership of these value is the ratio of their frequency ($n_4 = 28$, $n_{63} = 11$) to that of the most frequent value ($n_{21} = 61$):

$$m_{(4)} = n_4/n_{21} = 28/61 = 0.459$$

$$m_{(63)} = n_{63}/n_{21} = 11/61 = 0.18$$

The resulting values of fuzzification, the grades of membership, are non-integer values from 0 to 1. In order to obtain more interpretable results, the grades of membership are binned together by rounding them to the nearest one decimal point in the interval [0.1, 1]. For example, the grade of membership of $y = 4$ and $y = 63$ are $m_{(4)} = 0.459$ and $m_{(63)} = 0.18$ and are thus rounded to $m_{(4)} = 0.5$ and $m_{(63)} = 0.2$

As a intermediary output, the AFC organizes the values by grade of membership, thus creating 10 subsets of values sharing the same grade of membership. The organization of the values of multiset Y according to the frequency principle is presented in Figure 8.3 below. Each of the 10 columns of the table is the subset corresponding to a grade of membership, from 1 to 0.1 left to right. The height of the columns depends on the number of values sharing the same grade of membership.

In this first procedure, the AFC assigns the grades of membership to the values in the set as a function of their frequency among elements compared to the frequency of the most frequent value(s). However, it seems, if not intuitive at least sensible to group close F_0 values around identical or close grades of membership. It is improbable that values 43, 15 and 4, all sharing a grade of

y	2	3	4	5	6	7	8	9	10	11	12	13
n	2	11	28	33	25	35	28	27	31	29	36	45
m	0.033	0.18	0.459	0.541	0.41	0.574	0.459	0.443	0.508	0.475	0.59	0.738
$m_{0.x}$	0	0.2	0.5	0.5	0.4	0.6	0.5	0.4	0.5	0.5	0.6	0.7
y	14	15	16	17	18	19	20	21	22	23	24	25
n	59	28	49	42	47	37	49	61	49	43	50	40
m	0.967	0.459	0.803	0.689	0.77	0.607	0.803	1	0.803	0.705	0.82	0.656
$m_{0.x}$	1	0.5	0.8	0.7	0.8	0.6	0.8	1	0.8	0.7	0.8	0.7
y	26	27	28	29	30	31	32	33	34	35	36	37
n	41	42	38	45	28	44	30	34	19	25	30	24
m	0.672	0.689	0.623	0.738	0.459	0.721	0.492	0.557	0.311	0.41	0.492	0.393
$m_{0.x}$	0.7	0.7	0.6	0.7	0.5	0.7	0.5	0.6	0.3	0.4	0.5	0.4
y	38	39	40	41	42	43	44	45	46	47	48	49
n	28	20	20	32	17	29	22	23	27	9	24	13
m	0.459	0.328	0.328	0.525	0.279	0.475	0.361	0.377	0.443	0.148	0.393	0.213
$m_{0.x}$	0.5	0.3	0.3	0.5	0.3	0.5	0.4	0.4	0.4	0.1	0.4	0.2
y	50	51	52	53	54	55	56	57	58	59	60	61
n	9	11	19	13	10	12	12	15	8	13	13	18
m	0.148	0.18	0.311	0.213	0.164	0.197	0.197	0.246	0.131	0.213	0.213	0.295
$m_{0.x}$	0.1	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.3
y	62	63	64	65	66	67	68	69	70	71	72	73
n	16	11	21	19	8	13	9	6	10	8	8	4
m	0.262	0.18	0.344	0.311	0.131	0.213	0.148	0.098	0.164	0.131	0.131	0.066
$m_{0.x}$	0.3	0.2	0.3	0.3	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.1
y	74	75	76	77	79	80	81	82	83	84	85	86
n	8	4	6	4	5	10	1	2	4	4	10	6
m	0.131	0.066	0.098	0.066	0.082	0.164	0.016	0.033	0.066	0.066	0.164	0.098
$m_{0.x}$	0.1	0.1	0.1	0.1	0.1	0.2	0	0	0.1	0.1	0.2	0.1
y	87	88	89	90	91	92	93	94	96	97	99	100
n	5	7	1	7	3	2	5	12	6	5	3	27
m	0.082	0.115	0.016	0.115	0.049	0.033	0.082	0.197	0.098	0.082	0.049	0.443
$m_{0.x}$	0.1	0.1	0	0.1	0	0	0.1	0.2	0.1	0.1	0	0.4

Table 8.1: y : value present in the set

n : number of elements bearing value y in the set

m : grade of membership of y

$m_{0.x}$: rounded grade of membership of y

$m_{(x)} = 0.5$ in Figure 8.3 have the same impact on the overall intonation contour if they are chosen alternatively to 21 or 14 that have both been assigned the highest grade of membership $m_{(x)} = 1$. More specifically, a difference from 14% of the F_0 span to 15% might not be perceived if the F_0 span is quite narrow and, more importantly, since the two values are so close, it is unlikely that the difference is relevant at the level of the intonation contour; it is very unlikely that a variation of 1% in the F_0 span can differentiate two contours. In Figure

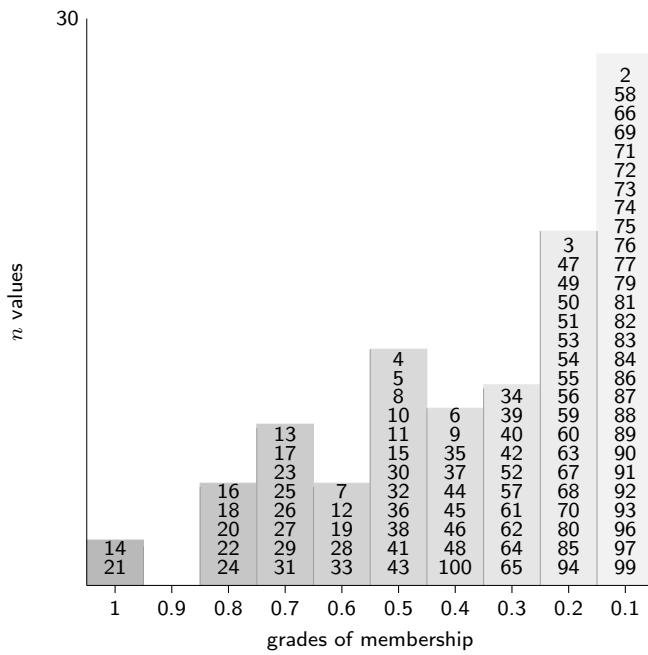


Figure 8.3: Distribution of the f_0 values of pre-tone 15 organized by grades of membership. Each column corresponds to a grade of membership. The height of each column corresponds to the number of values in the subset of the grade of membership. There are 96 values in the set.

8.3, many close values (such as $y = 2, 3$, and 4) are dispersed among various grades of membership, depending on their respective frequency. Therefore, the fact that two or more values are close, like 14 and 15, must be taken into account in the ranking of the values in the set, in spite of the fact that the frequency might differ (in the corpus). This is achieved by ranking and organizing the values in the set according to the similarity principle.

8.3.3 Fuzzification 2 of 2: the similarity principle

The AFC calculates a second grade of membership for the values in the set, this time according to the degree of similarity of their distance to the center of the set: “the measure of central tendency for ordinal data is the median” (Gries, 2010). This measure is best adapted for skewed distribution such as that of the features of an intonation contour and especially the pre-tones (see Figure 8.2 for the distribution of the F_0 values of pre-tone 15). It also prevents outliers from excessively weighting the calculation of the central tendency of the set. The procedure comprises the following sequence of steps:

1. The AFC must first locate the center of the multiset as its median value, noted \bar{y} .
2. For all other values, the AFC calculates the distance (δ) to the center \bar{x} :

$$\forall y : \delta_y = |\bar{y} - y|$$

3. The AFC finds the range r as the distance from the center (\bar{y}) to the most distant value from the center in the multiset:

$$r = \delta_x \max$$

4. Using the center ($\bar{y} = 28$) and the maximum distance ($r = 72$), the AFC can calculate the grade of membership of each value in the multiset according to its location relative to these two extreme points, expressed as a fraction of the range subtracted from 1:

$$\forall x : m_y = 1 - (\delta_y \cdot \frac{1}{r})$$

The procedure is now applied to multiset Y of the F_0 values (y) of pre-tone 15. The results of the fuzzification are in Figure 8.4.

1. First the AFC looks for the median value of the multiset or its center.
 - (a) The AFC orders all the elements in the multiset according to their numerical value.

$$Y \{y_1, y_2, y_3, \dots, y_n\}$$

- (b) The AFC counts the occurrences of each value (Table 8.1) and sums them as N :

$$N = n_{y_1} + n_{y_2} + \dots + n_{y_i} = 2 + 11 + \dots + 27 = 1981$$

- (c) The AFC determines the median value of the set:

$$\bar{y} = \frac{1981+1}{2} = 991^{\text{th}} \text{ value of the ordered set}$$

The AFC counts 991 values from the first value in the set and finds the 991^{th} value of the ordered set to be $y = 28$. The center \bar{y} of the set is 28

2. For all other values, the AFC calculates their distance (δ) to the center \bar{y} . For example, the distance of values $y = 4$ and $y = 63$ to \bar{y} are:

$$\delta_4 = |28 - 4| = 24$$

$$\delta_{63} = |28 - 63| = 35$$

3. The AFC finds the range r as the maximum distance to the center among the values in the multiset:

$$r = \delta_{100} = |28 - 100| = 72$$

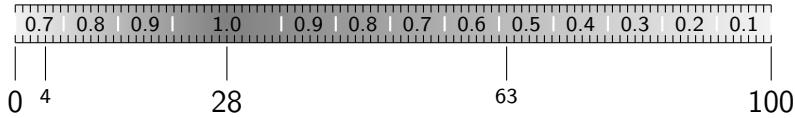
4. Using the center \bar{y} and the maximum distance r , the AFC calculates the grade of membership of the values in the multiset according to their location relative to these two points, expressed as a fraction of the range subtracted from 1. As for the first procedure, the results are rounded

to the nearest decimal point on the interval $[0.1, 1]$. For example, the grade of membership m of 4 and 63 are:

$$\text{for } 4 : m_{(y)} = 1 - (24 \cdot \frac{1}{72}) = 0.667 = 0.7$$

$$\text{for } 63 : m_{(y)} = 1 - (35 \cdot \frac{1}{72}) = 0.514 = 0.5$$

5. On the graph below, the grades of membership are implemented according to the distance to the center $\bar{y} = 28$ on a graded ruler. Values $y = 4$ and $y = 63$ are shown and their grade of membership can be read:



As a intermediary output, the AFC organizes the values by grade of membership, thus creating 10 subsets of values sharing the same grade of membership. The organization of the values of multiset Y according to the similarity principle is presented in Figure 8.4 below. The figure is composed of 10 concentric level of values, one for each grade of membership, from the closest to the center ($m_{(y)} = 1$) to the furthest from it ($m_{(y)} = 0.1$). In the organization of the multiset according to the similarity principle, values $y = 14$ and $y = 15$ are ranked with the same grade of membership $m_{(y)} = 0.8$ because they are almost similarly distant to the center.

At this point, the AFC has assigned two grades of membership to all values in the multiset, one according to its frequency, the other according to its distance to the arithmetic center of the multiset. The next step in the fuzzification process is to consolidate these results into a single grade of membership for each value.

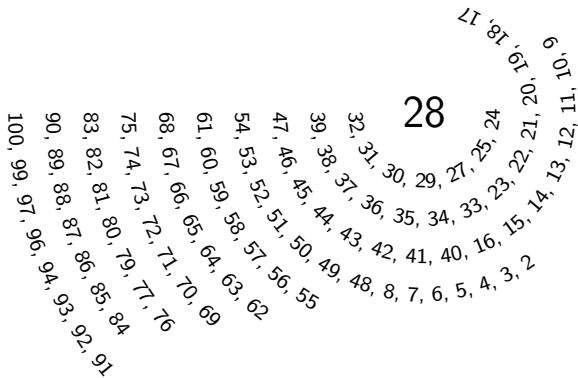


Figure 8.4: Fuzzification of Y according to the similarity principle. 28 is the center \bar{y} of the set. Each circle represents a grade of membership: the innermost circle contains the values that are the closest to the center and whose grade of membership is $m_{(y)} = 1$, the outermost circle contains the values that are the farthest from the center and whose grade of membership is $m_{(y)} = 0.1$.

8.3.4 Assigning a unified grade of membership

The final step in the process of ranking and organizing values in the set is to give them a unified grade that incorporates the information from the two ranking principles. The two grades (frequency and similarity) are averaged. This method is a departure from the principles of fuzzy set theory. However, fuzzy operators (union and intersection more specifically) have the disadvantage to select only a part of the data to which fuzzy functions have assigned a grade of membership.

The two principles are equal in importance and one should not outweigh the other: they both derive from the frequency of the values in the set. By averaging the two grades into one single grade, both principles have a reciprocal effect on the final unified membership function. For values with high

frequency and close proximity the central value of the set, its unique grade will remain high. If a value is not frequent but is close to the center of the set, its low ranking in term of frequency will be compensated by its closeness to the central value. By nature of the mode and the median, there should not be a value that is far from the center but with the highest frequency. If there is an outlier, a value distant from the center but that happens to have a relatively high frequency, its distance will bring its overall grade down from that gained by frequency.

The unrounded values of the grades of membership obtained with the two ranking principles are used in the calculation of the unique grade of membership. The result of this mean is rounded:

$$m = \frac{m_{FREQUENCY} + m_{SIMILARITY}}{2}$$

In Table 8.2 below, the grades of membership of a few selected values are presented to illustrate the interaction of the two ranking principles in the unified grade of membership assigned to each value. Value $y = 16$ has an identical grade for both rankings and its unique grade remains the same, $m_{(16)} = 0.8$. $y = 14$ is the highest ranking value in terms of frequency and still ranks high in terms of distance to the center. Its unique grade remains high, $m_{(14)} = 0.9$. $y = 15$ is close to the highest ranking value in terms of distance ($m_{(15)} = 0.8$) but its frequency is much lower ($m_{(15)} = 0.5$): the unique grade of $y = 15$ is higher than its frequency grade but does not reach that of $y = 14$. Thus, the AFC balances one principle by the other in the final ranking. $y = 28$ is the highest ranking value in terms of distance to the center, since *it is* the center ($m_{(28)} = 1$). However, its unified grade is $m_{(28)} = 0.8$ because of because of its lower frequency. The results of the unified ranking by grade average is

value	grades		
	frequency	similarity	unique
4	0.5	0.7	0.6
14	1	0.8	0.9
15	0.5	0.8	0.6
16	0.8	0.8	0.8
28	0.6	1	0.8
63	0.2	0.5	0.4
100	0.4	0.1	0.2

Table 8.2: Grades of membership by frequency, similarity, and as a mean of both for a few values in the set Y

provided in Table 8.3 below. The AFC has organized the values according to two ranking principles and then averaged these results into a unified ranking. The AFC has calculated what the most frequent values and the values closest to the center were for a given feature, thus determining the main trend among speakers, what range of values they are targeting and how the variation from this trend is distributed among grades of membership, from the closest to the trend ($m_{(x)} > 0.5$) to the furthest from the trend ($m_{(x)} < 0.5$). Accordingly, for each grade of membership, there is a subset of values such that the set of all subsets by grade of membership is the set of all values. Next, the AFC ranks and organizes the values within each grade subset, since even within a grade subset, not all value are equal in the decimal interval $[0.x_i, 0.x_j]$

8.3.5 Grades' subsets: inner ranking and organization

Unlike the multiset, subsets by grade do not contain recurring values since their values result from the sorting of the multiset. Thus, only the similarity principle is applied to the ranking of values in subsets.

1.0	{21}
0.9	{31, 29, 27, 24, 22, 20, 14}
0.8	{33, 28, 26, 25, 23, 19, 18, 17, 16, 13}
0.7	{41, 38, 36, 35, 32, 30, 15, 12, 7}
0.6	{48, 46, 45, 44, 43, 42, 40, 39, 37, 34, 11, 10, 9, 8, 6, 5, 4}
0.5	{52, 49, 47}
0.4	{65, 64, 63, 62, 61, 60, 59, 58, 57, 56, 55, 54, 53, 51, 50, 3}
0.3	{74, 72, 71, 70, 69, 68, 67, 66, 2}
0.2	{100, 94, 88, 86, 85, 84, 83, 82, 81, 80, 79, 77, 76, 75, 73}
0.1	{99, 97, 96, 93, 92, 91, 90, 89, 87}

Table 8.3: Unified ranking and organization of the set Y into subsets by grade of membership, from $m_{(y)} = 1$ to $m_{(y)} = 0.1$

1. The AFC locates the center of the subset \bar{y} as it does for the set, only the number of values is much smaller. The subset for $m_{(y)} = 0.4$, as shown in table 8.3, is used as the example for the calculation. The subset contains an even number of values, $N = 8$, and the split has been indicated by a set of double bars:

$$m_{(y)} = 0.4 \{3, 50, 51, 53, 54, 55, 56, 57 || 58, 59, 60, 61, 62, 63, 64, 65\}$$

The formula to locate the median value of an even number of values is of course slightly different than the one used previously for an odd number of values, the median being the mean of the two values on each side of the split:

$$\bar{y} = ((\frac{N}{2})^{th} + (\frac{N}{2} + 1)^{th})/2$$

With the $\frac{8}{2} = 4^{th}$ and $\frac{8}{2} + 1 = 5^{th}$ values of the ordered subset being 57 and 58 the center or median value of the subset is:

$$\bar{y} = (57 + 58)/2 = 57.5$$

2. For all values, the AFC calculates their distance (δ) to the center \bar{y} . For example, the distance of values $y = 3$ and $y = 65$ to \bar{y} are:

$$\delta_3 = |57.5 - 3| = 54.5$$

$$\delta_{65} = |57.5 - 65| = 7.5$$

3. The AFC determines the range r as the maximum distance to the center among values in the subset:

$$r = \delta_3 = |57.5 - 3| = 54.5$$

4. Using the center \bar{x} and the maximum distance r , the AFC calculates the grade of membership of the values in the subset according to their location relative to these two points, expressed as a fraction of the range subtracted from 1. As for other procedures, the results are rounded to the nearest decimal point in the interval [0.1, 1]. For example, the grades of membership m of 3 and 65 are:

$$m_{(3)} = 1 - (54.5 \cdot \frac{1}{54.5}) = 0 \rightarrow 0.1^1$$

$$m_{(65)} = 1 - (7.5 \cdot \frac{1}{54.5}) = 0.862 = 0.9$$

5. The result of the fuzzification of subset $m_{(y)} = 0.4$ is presented in Table 8.4 along with the fuzzification of all grade subsets. In each subset, the values have been ranked according to their distance to center of their subset and sorted by grades of membership.

Table 8.4 shows how the AFC ranks and organizes all the values in the set Y . Each value has a grade of membership at the set level and a grade of

¹0 → 0.1: since $y = 3$ is in the subset, it cannot have a grade of membership that excludes it from the subset ($m_{(x)} = 0$). The result is correct but must be adjusted to the lowest level of membership $m_{(x)} = 0.1$ to penalize the value without excluding it.

		set membership																	
		0.1					0.2												
		#	m	#	m	#	m	#	m	#	m	#	m						
21	1	24	1	23	0.8	32	1	37	1	49	1	60	1	72	1	82	1	92	1
22	0.8	19	0.8	35	0.9	40	0.9	47	0.3	59	1	71	1	84	0.9	93	0.9		
27	0.7	18	0.8	30	0.9	39	0.9	52	0.1	58	1	70	1	83	0.9	91	0.9		
20	0.6	25	0.7	38	0.8	34	0.9	57	1	69	1	81	0.9	90	0.7				
29	0.5	17	0.7	36	0.8	45	0.8	56	1	68	1	80	0.9	89	0.6				
31	0.3	26	0.6	41	0.6	44	0.8	55	1	67	1	86	0.8	96	0.4				
14	0.1	16	0.6	15	0.3	43	0.8	65	0.9	66	1	85	0.8	97	0.3				
		28	0.4	12	0.2	42	0.8	64	0.9	74	0.9	79	0.8	87	0.3				
		13	0.3	7	0.1	48	0.7	63	0.9	2	0.1	88	0.7	99	0.1				
		33	0.1			46	0.7	62	0.9			77	0.7						
						11	0.2	61	0.9			76	0.7						
						10	0.2	54	0.9			75	0.6						
						9	0.2	53	0.9			73	0.5						
						8	0.1	51	0.9			94	0.3						
						6	0.1	50	0.9			100	0.1						
						5	0.1	3	0.1										
						4	0.1												
\bar{m}		subset membership					set membership												
subsets		21	24	21	32	38	49	57	69	82	92								
set		36																	

Table 8.4: Fully expanded fuzzification of the set of y values of pre-tones 15. Grades of membership 1 to 0.1 are ordered left to right. Grades of membership in the subsets are organized top to bottom. The two extreme values are in blackened cells. The defuzzification values are given by subsets and for the set in the last two rows of the table (see Section 8.3.6 p186.

membership within the subset of the set level grade of membership. At the set level, the target for the F_0 values of pre-tone 15 are those with a high grade of membership, close to the left of the table. Within the subsets, the target values are also those with a high grade of membership towards the top of the table. Thus in terms of trend, $y = 21$ is the best-graded target value, with the best combination of frequency and centrality. At the opposite is $y = 99$, as the least well-graded value (black cells in the table). All the other values' grades of membership fall between these extremes. This continuum represents the graded variation of the F_0 value of the pre-tone.

The AFC fuzzifies the sets of all features of a given intonation contour. In a fuzzy set, values are not indeterminate or equal in term of membership as in a binary set, they are organized in a meaningful way. The fuzzification leads to the organization of the values of a set as the graded variation of a single feature, from most favored to least favored in terms of realization of this feature in the intonation contour. The variation of the intonation contour as a whole is the concatenation of the variation of its features, as captured into graded sets. As a way to integrate all the graded variations of the features within a single representative value, the AFC *defuzzifies* the organized set of graded values into a unique *crisp* value. The crisp value of a feature, such as pre-tone 15, is the value that stands for its entire set, ranked and organized as in Table 8.4. The crisp value would not be calculable without the preliminary organization of the set into graded levels. The crisp intonation contour is made of the concatenation of the crisp values of each feature in the vector.

8.3.6 Defuzzification

Defuzzification reduces the collection of membership function values in to a single sealer quantity (Yan, 1994).

When a set has been completely fuzzified (set and subsets), there is a need for an exploitable value that encapsulates all values in the set. The process of calculating the crisp value is called *defuzzification* and the result is noted \bar{m} . In the present work, this crisp value is calculated as the mean of all the values weighted by their respective grades.

$$\bar{m} = \frac{\sum m_x \cdot x}{\sum m_x}$$

There are several methods to defuzzify sets but this method is the most practical in terms of technical implementation in the automated system. Also, and very importantly, it is adapted to symmetrical and linear fuzzy functions such as those in the sets and subsets (with a trapezoidal or triangular shape).

Defuzzification of subsets The AFC first defuzzifies all subsets in a set. The calculation is shown for subset $m_{(y)} = 0.4$ of the set Y . The subset of values and their respective grades of membership is the following: $m_{(y)} = 0.4 \{3/0.1, 50/0.9, 51/0.9, 53/0.9, 54/0.9, 55/1, 56/1, 57/1, 58/1, 59/1, 60/1, 61/0.9, 62/0.9, 63/0.9, 64/0.9, 65/0.9\}$ (see Table 8.4). Figure 8.5 represents the membership function of the subset $m_{(y)} = 0.4$. As can be seen on the graph, apart from $y = 3$, the function is symmetrical. The outlier will be penalized in the weighted average defuzzification.

The AFC calculates the crisp value of the subset $m_{(y)} = 0.4$ as:

$$\bar{m}_{m_{(y)}=0.4} = \frac{3 \cdot 0.1 + (50 + 51 + 53 + 54 + 61 + 62 + 63 + 64 + 65) \cdot 0.9 + (55 + 56 + 57 + 58 + 59 + 60) \cdot 1}{1 \cdot 0.1 + 9 \cdot 0.9 + 6 \cdot 1} = 36.115$$

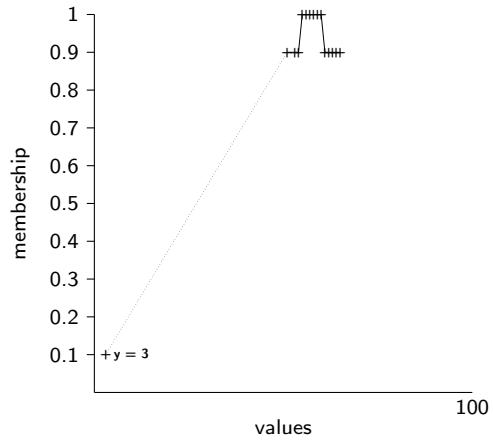


Figure 8.5: Membership function of subset grade $m_{(y)} = 0.4$. Apart from $y = 3$, the function is symmetrical. The outlier will be penalized in the weighted average defuzzification.

The crisp value of subset $m_{(y)} = 0.4$ is 57.465, rounded to 57. The AFC performs the same calculation for each of the 9 other subsets. Results are in the second to last row of Table 8.4 labelled “ \bar{m} subsets” plotted from Figure 8.6.

Each subset corresponds to a grade of membership in the set. Therefore, the crisp value of each subset is part of the membership function of the set of F_0 values of pre-tone 15. The results, as shown in Figure 8.6, indicate that the speakers are targeting values around 21 ($m_{(y)} = 1$) as the F_0 value of pre-tone 15 in the contour of closed questions. The variation in the set is expressed in terms of membership: the higher the grade, the better the fit of the value as the F_0 of pre-tone 15, and conversely, the lower the grade, the less good the fit as the value for that feature.

set m	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
subsets \bar{m}	21	24	21	32	38	49	57	69	82	92

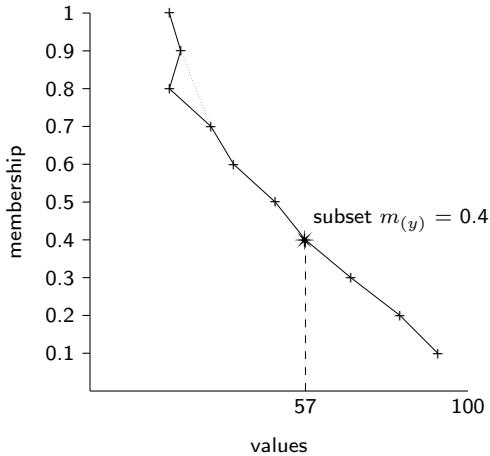


Figure 8.6: Membership function of the set of F_0 values of pre-tone 15 (set Y). Each point is the crisp value resulting from the defuzzification of a subset (grades of membership 1 to 0.1). The function is almost continuously linear, except for grade $m_{(y)} = 0.8$ (“corrected” as a dotted line)

Defuzzification of the set Because each subset corresponds to a grade of membership in the set, the AFC uses the crisp value of each subset to calculate the crisp value of the set. It relies on the same weighted average formula than for the subsets. The calculation of the crisp value of set Y of the F_0 values of pre-tone 15 is:

$$\bar{m}_Y = \frac{21 \cdot 1 + 24 \cdot 0.9 + 21 \cdot 0.8 + 32 \cdot 0.7 + 38 \cdot 0.6 + 49 \cdot 0.5 + 57 \cdot 0.4 + 69 \cdot 0.3 + 82 \cdot 0.2 + 92 \cdot 0.1}{5.5} = 36.115 \rightarrow 36$$

The crisp value of the set Y is 36.

The crisp value of a fuzzy set corresponds to the weighted average of the weighted averages of the ten subsets by grade of membership of the set.

It is thus the ideal F_0 value for pre-tone 15, as extracted from the analysis of all the values in the set in terms of their frequency and distance to the a central tendency. It is the “sealer” (Yan, 1994) that encapsulates all the processes of looking for a trend, or recurring pattern(s) among the realization of the speakers. The next chapters describe how the automated process of fuzzification is applied to three intonation contours. The Automated Tonal Labeling Module extracts the values of the feature vector of all instances of a given contour and passes them on to the Automated Fuzzification Module that fuzzifies and defuzzifies the sets. Next, it concatenates the results of all features to create the intonation contour abstracted from the analysis, organization, and ranking of all of its parts. The system generates the crisp contour and its ranked variations.

Chapter 9

Intonation contour of unmarked closed questions

9.1 Overview

This chapter presents an application of the PRInt model to the intonation contour of closed questions in French. This application to a corpus allows to describe the data for an intonation contour that extracted from a corpus of instances by the ATLM (Chapter 7) and fuzzified by the AFC (Chapter 8). These data correspond to the different tiers of the 4-layer structure: duration (sentence and syllables) and F_0 (maximum, minimum, range) for the first and second layers, pre-tones for the third layer, and tones for the fourth layer (including positional and relational data). The data of the first two layers are treated separately by the AFC for subsequent use in the analysis of the contours (see Chapter 11 and 13). The data for the third and fourth layers are presented in this chapter. The section for each layer has two parts: the preparation of the data prior to fuzzification and the fuzzification/defuzzification of the data. The preparation of the data consists in turning infinite data multisets $\{0, +\infty\}$ into finite multisets of integers $\{x, y, z\}$ that can be fuzzified efficiently. The data is fuzzified and defuzzified by the AFC, and the results of both the fuzzification and defuzzification are presented. Finally, the prototype of the contour is established.

9.2 The intonation contour of unmarked closed questions

9.2.1 Layer 3: pre-tones and the pattern recognition of the pre-tonal intonation contour

Values from layers 1 & 2: scaling, fuzzification, defuzzification The scaling process of the F_0 and time coordinate values of the pre-tones was explained in Chapter 7. Prior to the fuzzification of pre-tones coordinates, all syllables have been scaled to an equal duration (1/7th or 14.28%) of the sentence. All sentences comprise 30 pre-tones with the same syllabic anchoring from sentence to sentence. Accordingly, 30 pairs of multisets have been created to receive the original scaled values of the coordinates of each pre-tone in each sentence. For example, there is a set for the scaled time values of the first pre-tone of all sentences, a set for the scaled F_0 values of the first pre-tone of all sentences, etc, down to pre-tone 30.

The elements in these 60 sets are fuzzified and defuzzified according to the general methodology presented in Chapter 8. The data presented in Table 9.1 serve as the core of the category information. Each column from 1 to 0.1 corresponds to a grade of membership, the gray column labelled \bar{m} corresponds to the weighted average (the centroid noted \bar{m}) of all grades. Each row from 1 to 30 corresponds to a pre-tone (noted P in the header of the column). Pre-tones are grouped by syllables (1 to 7, horizontal solid lines) and half-syllables (first frame of the syllable σ_a or second frame of the syllable σ_b , horizontal dotted lines). For example, the defuzzified values of the coordinates of pre-tone 18 are (58, 34) on the 100 by 100 Cartesian plane. This means that the first pre-tone of the 5th syllable is typically realized with a distance of 34 points out of 100 from the F_0 baseline (= 0) and at a distance of 58 points out of 100 after the beginning of the sentence (= 0).

σ	$m(x)$	\bar{m}	1		0.9		0.8		0.7		0.6		0.5		0.4		0.3		0.2		0.1		
	P		x	y																			
%T	1	4	23	2	6	2	11	2	14	2	18	4	22	5	30	7	42	12	56	13	71	17	86
1a	2	4	22	3	7	4	11	3	12	3	17	4	21	5	29	6	40	8	56	11	71	11	86
	3	6	21	5	7	6	10	6	12	5	15	5	21	5	29	5	39	7	55	9	70	7	86
1b	4	8	21	8	8	9	10	9	12	8	13	8	20	7	28	7	37	8	54	12	68	8	86
	5	12	21	12	8	12	9	12	12	11	13	11	20	10	28	11	38	13	54	15	68	14	86
2a	6	15	24	15	11	16	11	16	14	15	16	14	24	14	32	14	41	16	57	16	70	14	86
	7	19	27	19	14	19	14	20	16	19	21	18	29	18	37	19	46	20	60	20	72	21	87
2b	8	22	30	22	17	23	17	24	18	23	24	22	32	22	40	23	50	22	63	21	74	21	87
	9	26	29	26	15	27	16	28	17	27	23	25	31	26	40	26	49	26	62	27	74	28	87
3a	10	30	30	29	16	30	17	31	17	30	22	29	32	29	40	29	50	28	62	28	74	28	87
	11	33	31	33	18	34	19	34	19	34	22	33	35	32	43	33	52	34	65	34	76	35	89
3b	12	37	32	37	20	38	19	38	19	37	19	37	36	35	44	37	53	37	68	35	78	35	89
	13	41	32	40	21	42	18	41	19	41	18	41	37	39	46	41	53	42	69	41	78	42	90
4a	14	44	33	44	22	45	20	45	19	45	19	44	37	43	47	43	54	45	69	43	80	42	91
	15	48	35	47	21	48	24	49	21	49	32	47	38	47	49	47	57	49	69	48	82	49	92
4b	16	51	36	51	22	52	24	52	22	52	27	51	41	51	50	49	58	51	70	51	82	50	92
	17	55	35	55	22	55	22	55	20	55	23	54	41	55	49	53	58	55	71	57	82	56	92
5a	18	58	34	59	24	59	21	59	17	58	18	58	41	58	48	57	57	58	73	58	82	57	91
	19	62	35	62	25	62	21	62	18	61	18	62	41	62	49	60	58	63	73	62	82	63	92
5b	20	65	34	66	24	66	21	66	18	65	17	65	40	65	49	64	57	65	71	64	80	64	92
	21	69	33	69	22	70	20	70	18	69	17	69	38	68	46	69	56	69	68	70	79	71	91
6a	22	73	33	72	21	73	21	74	18	73	20	72	38	71	46	73	54	72	67	71	78	71	89
	23	76	33	76	21	77	21	78	19	77	22	76	38	74	46	77	54	78	66	77	77	78	89
6b	24	80	33	79	21	81	20	81	18	80	25	79	39	79	47	80	54	79	66	78	77	78	89
	25	83	33	83	20	85	19	84	19	83	22	83	39	83	47	82	56	84	67	84	77	84	89
7a	26	87	39	87	21	88	26	88	28	87	48	87	48	87	52	85	59	86	58	85	59	83	63
	27	91	56	91	46	91	53	92	57	91	70	91	59	90	61	89	65	91	49	91	41	89	35
7b	28	94	74	94	74	94	81	95	85	95	91	94	72	93	75	94	72	93	44	92	28	91	13
	29	96	87	96	100	97	99	97	99	97	90	96	84	94	83	94	76	95	50	91	33	78	16
T%	30	96	89	98	100	98	100	98	99	98	93	97	90	95	89	95	78	94	54	87	36	67	18

Table 9.1: Fuzzification and defuzzification: the scaled time (x) and $F_0(y)$ values are given for each pre-tones (P , 1 to 30 vertically), by grade of membership (1 to 0.1 horizontally), and for the weighted average of all grades (gray column)

Pre-tonal contours A graphic representation of the ten grades of membership of the contour is obtained by transferring onto a plane the coordinates of the pre-tones by grade of membership. The advantage of a graphic representation over the table is that it gives a visual structure to the membership ranking and the weight of each principle (frequency or similarity) that is easier to comprehend than rows of numbers.

Each column of Table 9.1 corresponds to one graph of Figure 9.3, from membership 1 to 0.1. In actual sentences, each pre-tone has its distinct grade of membership independently of other pre-tones. For this reason, the graphs in this section are not actual sentences but abstractions or prototypical assem-

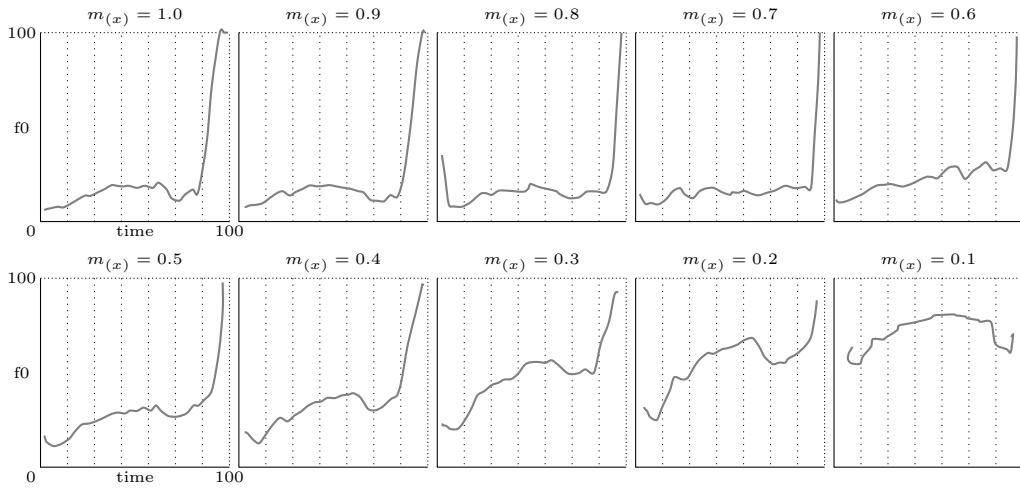


Figure 9.1: Closed question contour by grade of membership: frequency.

blages of pre-tones whose coordinates all share the same grade of membership. The reader should keep in mind that the PRInt model discretizes sentences into features and processes these features separately. If best exemplars were to be found, it would be at the level of the single value, not of the whole sentence. For example, it could be argued that value 3 is the best exemplar for the time value of pre-tone 1 since its grade of membership is ($m_{(x)} = 1$).

On each graph of Figures 9.1, 9.2, and 9.3, scaled time [0,100] is placed on the abscissa and scaled F_0 [0,100] is placed on the ordinate. All syllables have the same duration (1/7th or 14.28% of the sentence). The first set of graphs (Figure 9.1) corresponds to the fuzzification with the frequency principle, the second set of graphs (Figure 9.2) corresponds to the fuzzification with the similarity principle, and the third set of graphs (Figure 9.3) corresponds to the weighing of both principles and the results in table 9.1.

Graphs for the grades of membership under 0.5 display combinations

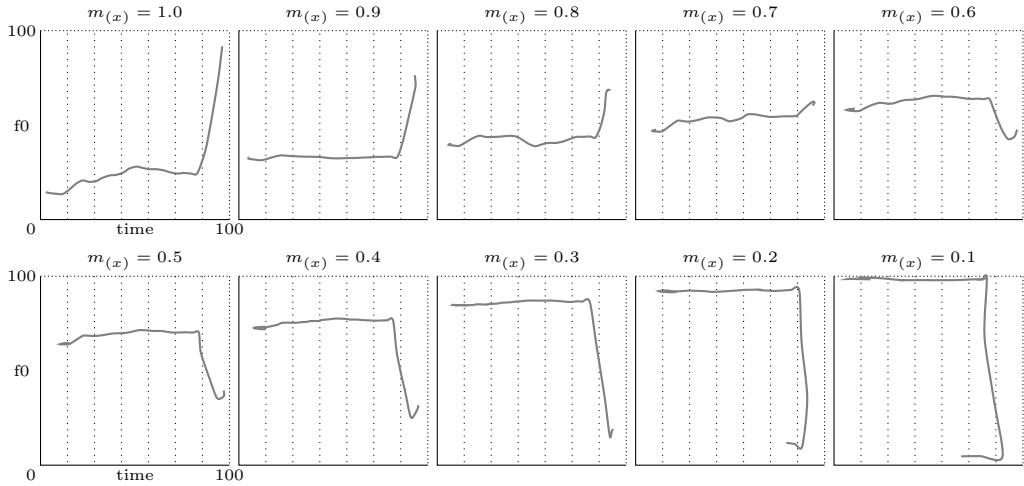


Figure 9.2: Closed question contour by grade of membership: centrality.

of pre-tonal values with low frequency and low similarity to the rest of the elements in their sets. Membership 0.1 regroups the outliers: the least frequent and the least central elements. It is important to emphasize again the fact that because all pre-tones in a graph have an equivalent grade of membership, they are abstract assemblages: the contour line of $m_{(x)} = 0.1$ for in Figure 9.3 goes back in time at its end. Such a contour cannot actually exist among instances. The PRInt model processes the data of each pre-tone separately and then concatenates the results by grade of membership. Thus, the graph for $m_{(x)} = 0.1$ in Figure 9.3 accounts for the fact that pre-tones 23 and 29 both have a value of 78 points in time at grade 0.1, but it does not imply that these two pre-tones co-occurred with the same grade of membership in the corpus. The value 78 is ranked 0.1 in terms of centrality for pre-tone 23 and 29. In the corpus, when pre-tone 23 occurred at 78 points in time in a subset of sentences, pre-tone 29 necessarily occurred later in these sentences (between 79 and 100 points). Conversely, if pre-tone 29 occurred at 78 points in time in a subset of

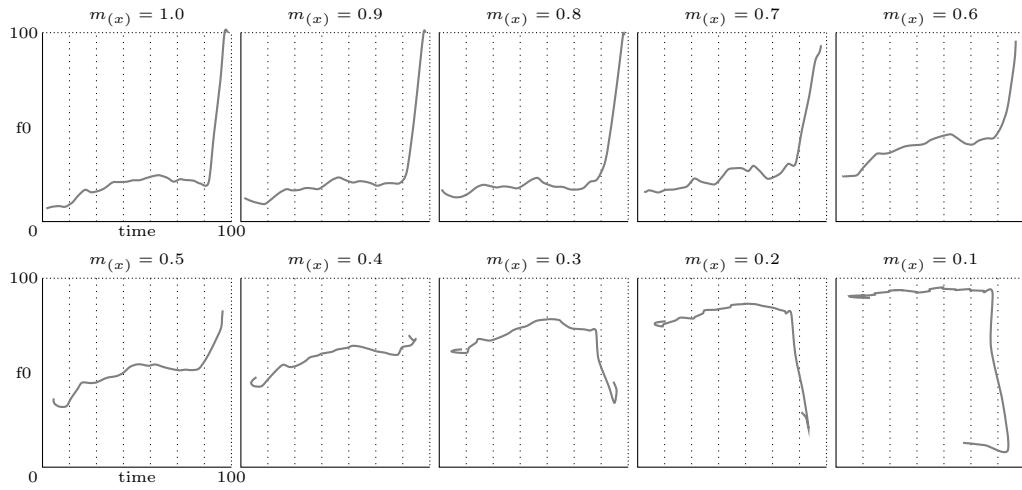


Figure 9.3: Closed question contour by grade of membership: frequency and centrality combined.

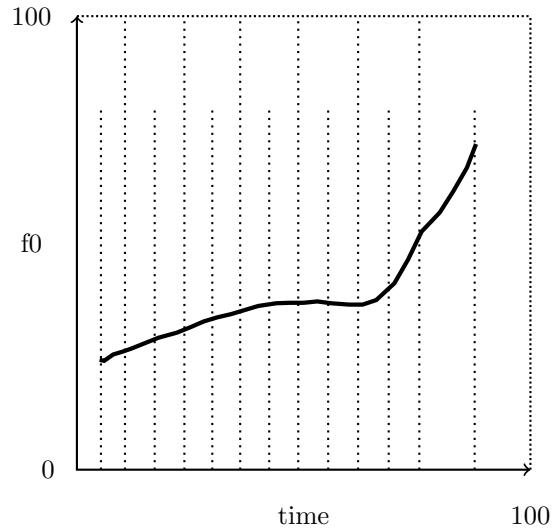


Figure 9.4: Closed question pretonal contour: defuzzification

sentences, pre-tone 23 necessarily occurred earlier in these sentences (between 0 and 77 points). This situation is addressed by the addition of the tonal layer.

The results of the defuzzification (\tilde{m}) constitute the PRInt model categorical output. It takes into account all values in the sets after they have been ranked and weighted. Figure 9.4 is the graphic representation of the intonation contour of closed questions as extracted by the model after it has processed all the data and merged the results. It closely matches earlier descriptions of this contour. There does not exist a true plateau in this defuzzified contour. The contour has a slight but constant rising slope through the first six syllables followed by a sharp rise on the last syllable.

The highest grade of membership $m_{(x)} = 1$ provides the highest level of precision for the alignment of events since it only retains the “best values,” those that are both the most frequent and the most central. Other levels provide additional crucial information concerning the frequency and centrality of the secondary patterns and information as to what values are on the edge of the category (low grades).

In the graphs for the frequency function (Figure 9.1), as the grade of membership lowers, the pre-tones in the plateau progressively rise to become more distant from the F_0 baseline while the pre-tones in the final syllable lower to approach it. This behavior indicates that a maximum contrast is favored between the constitutive parts of the contour (a low plateau and a sharp high rise in the case of closed question). As the contrast gradually diminishes, so does the categorical membership.

In the graphs for the centrality function (Figure 9.2), there is a noticeable mirror effect from the most central values of grade $m_{(x)} = 1$ to the least central grade $m_{(x)} = 0.1$. As it is expected from the similarity principle, $m_{(x)} = 0.1$ values are the least similar to the $m_{(x)} = 1$ values. $m_{(x)} = 0.9$ values are close to $m_{(x)} = 1$ values, $m_{(x)} = 0.8$ values are closer to $m_{(x)} = 0.9$ values

than $m_{(x)} = 1$ values and so on until the most radical difference is obtained. The PRInt model constructs the intonation category differently depending on the ranking principle.

With the frequency principle, the categorical contrast vanishes along a gradient continuum [1,0] from absolute presence of contrast ($m_{(x)} = 1$) to near absence of it ($m_{(x)} = 0.1$). The categorical contrast goes from absolute contrast $m_{(x)} = 1$ to near absence of it ($m_{(x)} = 0.7$) and to absolute dissimilarity or opposite contrast ($m_{(x)} = 0.1$). With centrality, the PRInt model goes further than with frequency in that it computes what would be the opposite contour from the expected one, a contour that, unlike that of the frequency model, is completely unlikely.

The similarity principle, based on central tendency, gives more weight to values that are not the most frequent when there is a strong mode. In this case, apart from the mode, the distribution of the rest of the values might be spread out, each value occurring just a few times and sometimes only once. Thus in the graph for $m_{(x)} = 1$ in Figure 9.3, the values of each pre-tone are at the same time mode and median. As the mode becomes smaller (that is the frequency of the most frequent value lowers), the median become skewed to one side or the other of the mode.

For example, in graph $m_{(x)} = 1$ of the frequency principle (Figure 9.1), the F_0 scaled value corresponding to pre-tone 29 is 100 because it is the most frequent F_0 scaled value for this pre-tone. Other high values close to 100 are frequent as well for this pre-tone. Thus, at the lowest grade of membership $m_{(x)} = 0.1$, the F_0 values of the last pre-tones are still high. In graph $m_{(x)} = 1$ in Figure 9.2 showing the centrality principle, the median is close if not identical to the mode and thus the contour looks like the one of the

frequency principle (Figure 9.1). The centrality-based $m_{(x)} = 1$ has a narrower F_0 since the median and mode do not necessarily match. At the lowest grade of membership $m_{(x)} = 0.1$, mode and median are much more distant and the least frequent values do not correspond to the least central, hence the growing difference between the graphs of the frequency and centrality principles as the grade of membership decreases from 1 to 0.1.

In the graphs of Figure 9.3, both rankings are included with an equal weighing. At grade $m_{(x)} = 1$, the profile of the frequency principle most closely matches that of the centrality principle because the modal and medial values are close. From $m_{(x)} = 0.9$ to $m_{(x)} = 0.2$, as the grade of membership becomes smaller, the distance between mode and median increases. The values of the centrality ranking more heavily influence the combined profile in 9.3 than those of the frequency ranking. At grade $m_{(x)} = 0.1$ the centrality principle is at its most influential. The PRInt model combines both principles: from the most expected values (most frequent and central membership) to the least expected ones (least frequent and central membership). The shape of the contour contrasts at both ends of the continuum: a low plateau with a final sharp rise at one end of the continuum and a high plateau with a sharp fall at its other end.

9.2.2 Layer 4: tones and the pattern recognition of the tonal intonation contour

Previous works (Fónagy & Bérard, 1973; Beyssade et al., 2007; Vion & Colas, 2002: among others) characterized the contour of closed question as mainly a final rise on the last syllable (or possibly starting on the left edge of the penultimate syllable) of a sentence. The string of tones includes (1) an

initial low tone (%L) marking the beginning of a low plateau, (2) a low tone marking a turning point at the end of the plateau, also called the *elbow* of the contour (L-), and (3) a high, final tone (H*). This last tone is generally merged with the final tone of the sentence ($H^* = T\% \rightarrow H^*\%$).

The PRInt model proceeds from the bottom layer (actual data) of the sentence 4-layer structure to the top layer (tones), using at each level the results obtained from the previous level. The PRInt model is now ready to analyze the sentences in the corpus in terms of patterns. It has converted all sentences into feature vectors, strings of tones whose positions can be moved on an abstract grid. The PRInt model finds how many times each pattern, that is, each possible arrangement of the tones on the grid, is occurring, which is the most frequent, and the variations.

In this particular application, the ATLM was set to find three peaks (P^* , P_a , P_b) per sentence independently of any pre-established phonological knowledge (see Chapter 7). First, it finds the highest point of the sentence P^* , noted H^* ($= \max F_0$), and the lowest points before (L-) and after (-L) this point. Second, it looks for a lower peak H (P_a) after the lowest point (-L) following H^* . Third, it looks for a lower peak H (P_b) before the lowest point (L-) preceding H^* . Depending on the sentence, all three peaks may not necessarily be present and, for each existing peak, associated low points (preceding L- and following -L) may or may not exist. The -L of one peak and the L- of the following peak can be merged. When the highest peaks (P^*) occurs close to the end of the sentence, as it does with unmarked questions, there is no time after this peak for another high peak (P_a) to occur; the PRInt model may only find one more peak (P_b) before the main one (P^*).

The PRInt model seeks a P_b - P^* - P_a sequence in all sentences. Thus

each sentence is a grid with three peak positions on which P_b , P^* , and P_a can be anchored: H_1 , H_2 , and H_3 . The PRInt model counts the number of occurrences of P^* on each of the positions H_1 , H_2 , and H_3 throughout the corpus. In the output of the fuzzification, the PRInt model found that the main peak (P^*) is a final peak ($P^* = H_3$; patterns d, e) in 79% of the sentences. In 21% of the sentences, P^* was non-final ($P^* = H_2$; patterns a, b, c).

In those cases when P^* is realized as a final peak on H_3 (patterns d,e) and there is no space after it for a subsequent movement, the results in Table 9.2 indicate that the time and F_0 values of H , -L, and T% are actually merged (see last four columns on the right of the table). Looking at grade 1 of membership, H , -L, and T% are 98, 98, and 99 respectively in time points; 99, 99, and 99 respectively in F_0 points. This merge occurs more or less at all levels of membership.

In those cases when P^* is realized as a non final peak on H_2 (patterns a, b, c) and there is space after it for a subsequent movement, the final movement (P_a on H_3) is realized with some values below that of the primary peak, that is, with a grade of membership less than 1. However, the H points are not necessarily much lower than the H^* point. Actually, H^* being equal to 100, an H can potentially be as high as 99.

Values from layers 1 & 2: scaling, fuzzification, defuzzification To calculate the amplitude of the non-final secondary peak in the case of a primary final peak, (pattern e), the 437 occurrences of the non-final primary peak, P^* as H_2 , must be removed from the data. The results in Table 9.2 account for the modification. The results for time are in the top part of the table and the results for F_0 are in the bottom portion of the table. The four most interesting

$m_{(x)}$	T%	L	H	L	L	H	L	L	H	L	T%	
TIME		H ₁			H ₂			H ₃				
Fuzzification	1	1	28	-	-	42	-	-	85	98	98	99
	0.9	-	14	35	-	43	-	<u>85</u>	-	-	99	-
	0.8	0	13	20	43	-	<u>50</u>	-	-	96	-	-
	0.7	-	29	50	28	37	53	71	86	96	97	-
	0.6	3	28	34	44	41	53	64	82	92	95	97
	0.5	6	25	35	46	46	63	65	82	87	90	93
	0.4	8	26	34	47	44	70	64	73	78	83	89
	0.3	14	53	50	57	59	23	38	53	60	65	-
	0.2	19	65	66	69	67	13	21	41	44	49	-
	0.1	24	73	74	75	80	7	7	25	31	40	44
\bar{m}		5	27	37	45	45	50	65	77	87	90	94
F0		H ₁			H ₂			H ₃				
Fuzzification	1	7	-	57	-	11	-	-	20	99	99	99
	0.9	10	11	-	52	12	<u>32</u>	<u>23</u>	9	-	-	-
	0.8	14	10	52	-	16	39	22	14	-	-	-
	0.7	17	5	29	13	19	32	19	20	-	-	-
	0.6	21	21	56	48	30	47	41	32	98	96	-
	0.5	29	34	54	33	40	55	50	40	88	87	91
	0.4	40	43	66	30	49	59	60	51	77	77	80
	0.3	55	61	27	74	62	75	71	64	52	52	55
	0.2	69	70	13	84	74	86	80	78	35	36	38
	0.1	84	84	5	90	86	96	89	88	18	18	20
\bar{m}		22	24	47	43	27	47	38	29	83	82	81

Table 9.2: Time and F₀scaled values of the tones by grade of membership (frequency x centrality). These results are for the main pattern, with the main peak on H₃, and exclude the results of primary peaks realized on H₂.

lines are the results for grade 1 and for \bar{m} , for both time and F₀. The two peaks, secondary on H₂ and primary on H₃ are clearly distinct in F₀ magnitude. One can interpolate the missing data by using the number immediately under them (underlined in the table). At grade 1 of membership, there is no value for H

f0 (FREQUENCY X CENTRALITY)											
	%T	H ₁		H ₂			H ₃		T%		
m=1	7	(11) ^{0.9}	57	(52) ^{0.9}	11	(53) ^{0.5}	(21) ^{0.7}	20	(98) ^{0.9}		
\bar{m}	22	24	47	43	27	51	40	28	81	66	70

Table 9.3: Partial results for the F₀values of the tones (m_1 and \bar{m}), with P* as H₂ included and P* as H₃ excluded. Values in brackets are missing for the grade of membership and they are inferred from the closest grade for which the value is available, indicated in superscript.

and -L of H₂. They can be approximated as 50 ($m_{(x)} = 0.8$) and 85 ($m_{(x)} = 0.9$) for time, 32 ($m_{(x)} = 0.9$) and 21($m_{(x)} = 0.7$) for F₀, keeping in mind that these approximations improve as the grades of membership get closer.

Table 9.3 presents partial results of the fuzzification of the data but with the inclusion of the occurrences of P* as H₂ (when the H of the H₂ is realized as 100% of the F₀ range) and with the exclusion of the occurrences of P* as H₃. Overall (defuzzification of both functions), the only difference is in the value of H under H₂: it goes from 47 to 51, a negligible difference. However, looking at the frequency function, H₂ and H₃ display the same merging tendency for both F₀ and time values at grade 1 of membership but not at the defuzzied level. Contrary to H₃, when H₂ is a high peak (P*), it is followed by a falling slope, followed by the H₃ final rise.

The semantics of the sentences used in the elicitation task expectedly led to the observed variation in their interpretation and in the placement of the peaks. To illustrate these variations of the contour, four samples from the corpus of a participant (coded GR) are presented in Figure 9.5. Participant GR realized his sentences as two main subsets: with a primary sentence final high peak (figures on the left) or with a primary non sentence-final peak. Within

each subset, participant GR also realized some sentences as two intonation phrases with a primary peak and a secondary peak. The phrasal division usually consists of two syntactical phrases (noun phrase or verb phrase). The divided sentences have two high tones and the highest tone can occur on the first (pattern b) or second phrase (pattern e). These two patterns correspond to two variations of the phenomenon described by Fonagy as the Parisian double rise, with the position of the peaks alternating between patterns b and e. Sample [GR75] is an example of pattern (d). It is realized with P^* as H_3 on the final syllable (σ_7). The primary peak is both phrase-final and sentence-final. There is no other peak.

Sample [GR72] is an example of pattern (e). It is realized with P^* as H_3 on the final syllable (σ_7). The primary peak is both phrase-final and sentence-final. The secondary peak P_b is realized as H_3 at the boundary of syllables (σ_3) and (σ_4). It is phrase-final (main verb phrase *vous voulez*, “you want”).

Sample [GR91] is a (rare) example of pattern (a): it is realized with P^* as H_2 on syllable σ_4 . This peak is on the second syllable of the adverb *beaucoup* (“much/a lot”) which can arguably modify the verb or qualify the following non phrase. Its ambiguous status might be the reason why it receives the highest peak rather than the verb. It might reflect the decision to group the adverb with the verb rather than the noun. Notwithstanding the choice of subject GR, the intonation contour is not common among the sentences of the corpus (and among speakers, for that matter). After the peak, the F_0 contour slowly drops over the course of the noun phrase.

Sample [GR65] is an example of pattern (b): it is realized with P^* as H_2 at the boundary of syllables (σ_4) and (σ_5). It is phrase-final (verb phrase

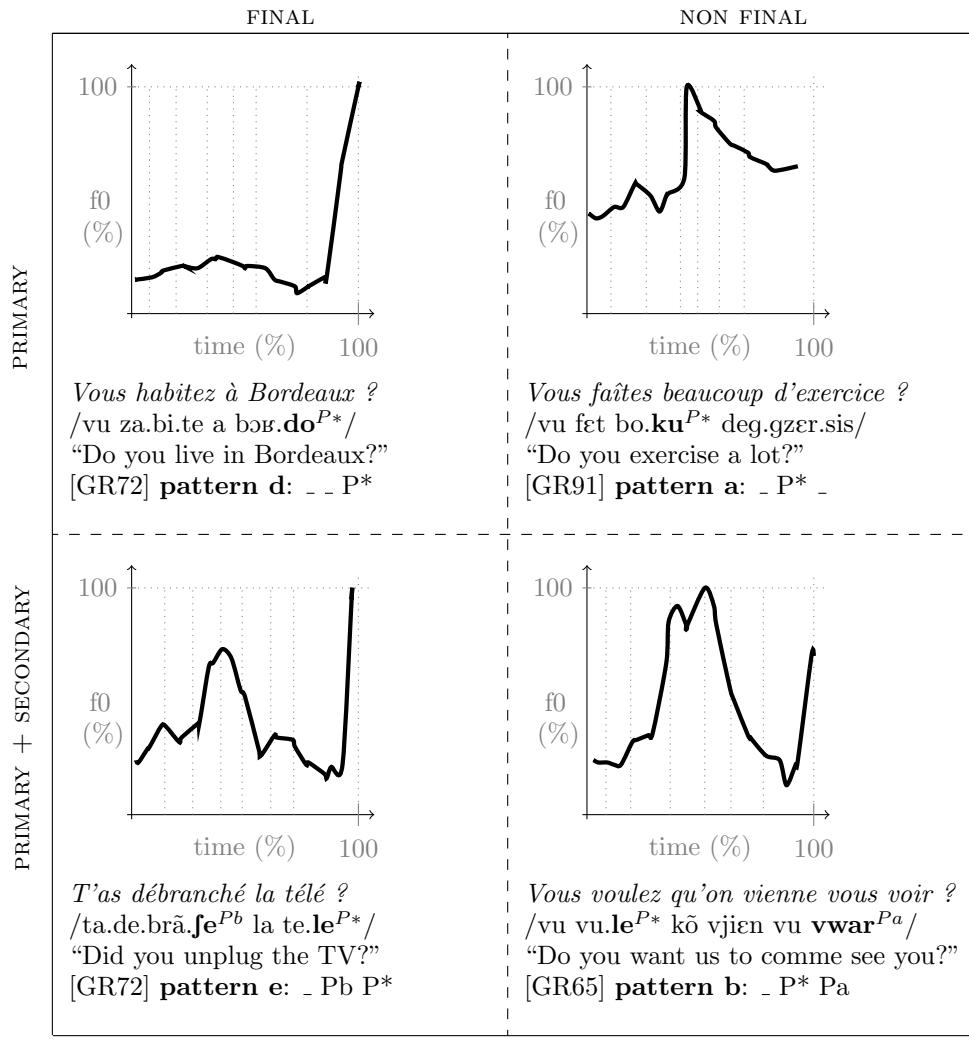


Figure 9.5: Four samples of pattern variations from the corpus of subject GR: on the left, patterns d (top) and variation e (bottom); on the right, patterns a (top) and variation b (bottom).

vous habitez, “you live”). The secondary peak P_a is realized as H_3 on the final syllable ($\sigma 7$). It is sentence-final and phrase-final (noun phrase *à Bordeaux*, “in Bordeaux”).

In conclusion of this analysis from the data provided by PRInt, the contour of closed question is primarily realized as pattern (d), or with a possible variation, as pattern (e)(79% of the sentences). Alternatively, it can be realized as pattern b (21% of the sentences).

$$\begin{aligned} P^* = H_3 \rightarrow \%T \overline{L-H-L}^1 \overline{L-H-L}^2 \overline{L-H-L}^3 T\% &= L\% (\overline{L-H-L}^2) \overline{L-H\%}^3 & (d/e) \\ P^* = H_2 \rightarrow \%T \overline{L-H-L}^1 \overline{L-H-L}^2 \overline{L-H-L}^3 T\% &= L\% \overline{L-H\%}^2 (\overline{L-H\%}^3) & (b) \end{aligned}$$

At this stage, the PRInt model has identified the main pattern(s) of the tonal contour of closed questions. It still has to provide more specific information relative to the alignment of the tones of the contour with the syllabic structure. The pre-tonal isometric grid constitutes the abstract frame of reference common to all sentences, independently of their true range. To determine the syllabic alignment of the tonal contour, the PRInt model has to determine onto which pre-tones the tones of each peak have been anchored for each instance of the contour in the corpus.

Anchoring of tones (1 of 2): pre-tones The PRInt model processes the data of the pre-tonal anchoring of tones as the rest of the data. No preparation is needed since each tone can only be anchored to one of the 30 pre-tones. Results in Table 9.4 show the highest ranking positions of pre-tonal anchoring ($m_{(tonal\ anchoring)} = 1$) of each tone (L-, H, -L) of the three peaks (P_a , P^* , P_b).

The tones of the main contour and their corresponding pre-tones have been highlighted in gray. The plateau (%TL-) extends from the first half of the first syllable (pre-tone 3) to the left edge of syllable 7 (pre-tone 26). The final rising movement (%L-H) is anchored on the last syllable. Pre-tone 26 is the first pre-tone after the left boundary of the 7th syllable and pre-tone 29 is the last pre-tone before the right boundary of the 7th syllable. These results coincide

Peaks		H ₁			H ₂			H ₃			
Tones	%T	L-	H	-L	L-	H	-L	L-	H	-L	T%
Pretones	3	8	3	13	14	17	17	26	29	29	30

Table 9.4: Pre-tonal anchoring of the tones ($m_{(x)} = 1$ for frequency x centrality). Each pre-tone corresponds to a fixed position in the syllabic structure.

with the results for the time and F_0 scaled values of the tones. The secondary movement is typically anchored on the 4th syllable (pre-tones 14 for L- and 17 for H-L). Although pre-tonal anchoring provides some information about the implementation of the contour on the syllabic structure of the sentence, data have been processed separately for each tone. It does not take into account the relation from one tone to the next, the co-occurrence of one part of a tonal movement with the next.

Consider the tones of the first peak H₁ (which is not strictly relevant for the contour of closed questions): the initial L- is anchored on pre-tone 8 and H is anchored on pre-tone 3. These results are not faulty but reveal a limit of the PRInt model when it fuzzifies features of the contour separately. These results indicate that L- is realized typically as pre-tone 8 while H is realized typically as pre-tone 3. They do not indicate on which pair of pre-tones L- and H are typically co-occurring.

Anchoring of tones (2 of 2): pre-tonal pairing The PRInt model processes more elaborate features, or relations of features, if they are properly prepared to enable the ranking of elements in a finite set. As discussed in the previous section, it is important to know on what pairs of pre-tonal anchors the tones forming a movement occur, instead of only getting the information

for each tone individually.

For each of the three peaks P^* , P_a , and P_b , the PRInt model finds the pre-tones on which each of the constituting tones (L-, H, and -L) occurs. For each sentence, tonal co-occurrences are encoded as two pairs for each peak: L-H (upward) and H-L (downward). If in one instance, the initial L- of a were to takes place on pre-tone 26 and its H on pre-tone 29, the pair would be labeled (26.29). Here are two examples of sentences, SV66 and FL28, whose pre-tonal co-occurrences have been prepared for fuzzification:

Peaks	Initial data								Association labels								
	P^*			P_a			P_b			P^*			P_a			P_b	
Tones	L	H	L	L	H	L	L	H	L	L-H	H-L	L-H	H-L	L-H	H-L	L-H	H-L
SV66	14	17	17	28	29	29	4	12	14	14.17	17.17	28.29	29.29	4.12	12.14		
FL28	26	29	30	-	-	-	13	14	18	26.29	29.30	-	-	13.14	14.18		

Pre-tonal co-occurrences have only been fuzzified by frequency because they are compounds, not single values. Consequently, results could not be defuzzified. As shown in Table 9.5, there exists a large range of possible pairings but only a few recur more than twice.

The most frequent pairings for the main peak (P^*) are (26.29) for L-H and (29.29) for H-L. These results support those found for the individual pre-tones and indicate a somewhat binary constraint on the alignment of the main peak. The final movement occurs on the last syllable and the H and -L tones are merged. There is no intermediate results between grades 1 and 0.4 of membership for L-H and none between grades 0.9 and 0.2 for H-L. Notice also that the anchoring possibilities of L- all take place in the 7th syllable or within the end of the 6th syllable (pre-tone 24 and 25) but that all the anchorings of H only occur on the end of syllable 7.

	P*		Pa		Pb	
	L-H	H-L	L-H	H-L	L-H	H-L
1	26.29	29.29	26.29	29.29	25.26	26.26
0.9		29.30			18.19	
0.8				29.30	6.7	
0.7				3.3	13.14	
0.6				3.6	22.23	
0.5			21.22			24.26
0.4	25.29		29.29	22.22	24.26	15.15
0.3	26.28		29.30	27.27	24.25	23.26
0.2	24.29	28.30	28.30	28.30	22.26	25.26
0.1	28.29	27.27	28.28	24.24	23.26	26.27

Table 9.5: Pre-tonal co-occurrences by peaks and grades of membership
(partial results)

The range of possibilities is much greater for the secondary peak (P_b), especially for its first part H-L. It seems that for the primary and secondary peaks, the alignment constraint is stronger on the second part of the movement (H-L), if it occurs at all. The secondary peak can take place anywhere between the 3rd and 6th syllable. Its anchoring is gradient and varies with sentences.

Scaled distance between tones The PRInt model calculates the distance in scaled F_0 and scaled time (%) between the tones of the three peaks it has identified in each sentence. These values do not need adjustment since they are already included in a finite set of possible values [1,100]. Multisets of values for each dimension are created, and their elements are fuzzified and defuzzified according to the general methodology. Results are presented in Table 9.6.

The highest grade of membership $m_{(x)} = 1$ represents the most frequent and most central pattern among all subjects. At this grade of membership, the

Peaks →	Scaled time (%)				Scaled F_0 (%)			
	Primary		Secondary		Primary		Secondary	
Tones →	L-H	H-L	L-H	H-L	L-H	H-L	L-H	H-L
1	12	0	6	-	-	0	-	8
0.9	10	-	5	-	89	-	0	7
0.8	11	-	9	-	87	-	-	10
0.7	7	-	10	-	79	-	11	18
0.6	16	3	13	10	65	2	14	26
0.5	13	6	18	11	65	9	28	33
0.4	15	11	25	18	44	24	39	40
0.3	27	23	33	31	37	43	56	52
0.2	32	32	45	38	24	62	69	67
0.1	38	40	67	45	4	85	85	86
$\bar{m}.$	14	9	14	19	68	16	12	22

Table 9.6: Distance between tones in scaled time and F_0 for the primary and secondary peaks of the question contour

sentence-final primary peak corresponds to an upward movement L-H of the contour of 12 points in time and 89 in F_0 . There is no downward movement, the F_0 and time values of L-H are zero. These results match what the PRInt model has found so far about the contour: the primary sentence-final peak is not followed by any movement. The F_0 value of the L-H movement goes down with the grade of membership. Low F_0 are at the limit of the category. The contrary is true for the H-L movement: values higher than zero are graded low. Interestingly, there is no value before grade 0.6 and this value is 2. The same observations stand for time values. However, in this dimension, the L-H movement has a more limited range of variation: the values do not vary much above grade of membership 0.4. The H-L movement displays a similar behavior for time values as for F_0 . There is no value over zero before grade 0.6 and the value is 3. The typical last syllable is a final rise with no subsequent

downward movement. When a downward movement do occur at the end, the contour is graded a lower membership.

Accordingly, the PRInt model finds that the F_0 value for the secondary peak is zero: there is no secondary peak in the main pattern of the contour. When the secondary peak is realized, its time and F_0 values are typically small, higher values being graded lower.

In summary, the results from the PRInt model reveals three strong constraints on the realization of the closed question contour: (1) the presence of a sharp final rise (markedly high in F_0 , relatively short in time), (2) the absence of a subsequent downward movement, and (3) the absence of a secondary peak. If the secondary peak exists, it is markedly smaller in amplitude than the primary peak.

The defuzzified results (\bar{m}) support these findings and give the general proportion of the two peaks as found in the corpus: the primary peak must be much higher in F_0 but shorter in time than the secondary peak. The defuzzified results take into account all grades of membership. Consequently, the results of the second movement of the primary peak (H-L) are smaller than that of the first movement because the second part occurs less frequently in the corpus.

9.2.3 Prototypical contour

From the information gathered by the PRInt model, it is possible to establish the prototypical contour of the category of closed questions, in relational terms (higher/lower, before/after), from the position of the H* tone on the isometric grid. The result is presented in figure 9.6

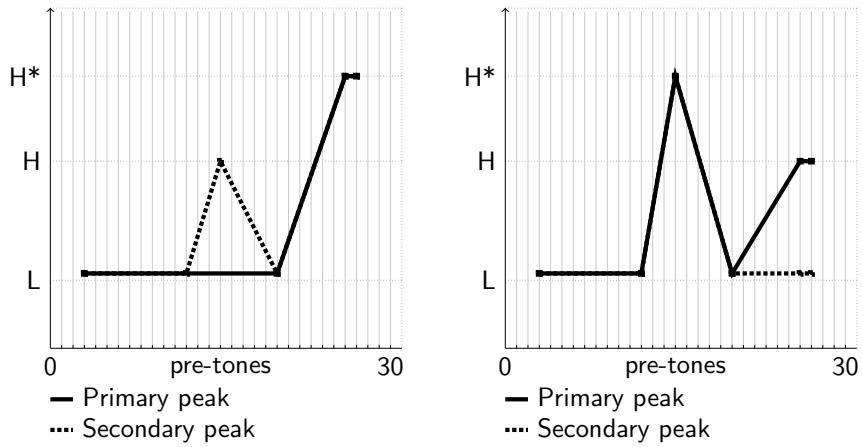


Figure 9.6: Prototypical contour. **Left:** main contour ($P^*=H_3$). **Right:** variation of the contour ($P^*=H_2$)

Additional data: velocity and angle Finally, the PRInt model calculate the slope of the rise of the primary peak (F_0 to time ratio) and the angle formed by the L-H* (considered as a straight line) line and the time axis. These dimensions will be used to compare the contour of closed question with contours of different categories.

Peaks →	Velocity (f0/time)		Angle	
	Primary	Secondary	Primary	
Tones →	L-H	L-H + H-L	L-H	
1	6	2	1	82
0.9	4	-	-	81
0.8	6	3	2	84
0.7	10	4	0	79
0.6	7	5	3	77
0.5	14	9	5	72
0.4	22	15	8	64
0.3	45	24	14	50
0.2	-	39	21	36
0.1	85	51	31	21
Defuzz.	12	9	5	74

Table 9.7: Velocity between tones in scaled F_0 per time for the primary and secondary peak of the question contour, angle (steepness) of the primary peak.

Chapter 10

Intonation contour of two modalities: surprise and doubt

Overview

In this chapter, the PRInt model is applied to two additional intonation contours: the modality of surprise and the modality of doubt. Since the protocol for using the PRInt model with a contour has been described in the previous chapter, this chapter will be more succinct and will simply provide the results of the fuzzification of the data and the prototypical contours of each modality.

10.1 The intonation contour of the modality of surprise

10.1.1 Layer 3: pre-tones and the pattern recognition of the pre-tonal intonation contour

Values from layers 1 & 2: scaling, fuzzification, defuzzification The results of the PRInt model for the pre-tonal coordinates are presented in Table 10.1. Each column from 1 to 0.1 corresponds to a grade of membership. The gray column labelled \bar{m} corresponds to the weighted average (the centroid labelled \bar{m}) of all grades. Each row from 1 to 30 corresponds to a pre-tone (labelled P in the header of the column). Each row from 1 to 30 corresponds to a pre-tone (labelled P in the header of the column). Pre-tones are grouped by syllables (1 to 7, horizontal solid lines) and half-syllables (first frame of the

syllable σa and second frame of the syllable σb , horizontal dotted lines).

σ	$m(x)$	\bar{m}	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
	P	x y	x y	x y	x y	x y	x y	x y	x y	x y	x y	x y
%T	1	6 24	2 7	2 9	2 14	2 16	9 26	8 34	14 45	17 60	22 76	31 92
1a	2	6 23	3 8	4 9	3 13	3 15	7 25	6 33	10 44	13 59	16 75	20 92
	3	6 23	5 7	6 9	5 12	5 15	5 24	4 33	7 43	8 58	9 74	8 93
1b	4	8 21	9 5	9 8	9 11	8 13	8 22	6 32	8 42	9 57	12 73	7 95
	5	12 21	12 4	12 8	11 10	11 12	12 22	11 31	13 42	12 58	16 72	14 97
2a	6	15 22	16 5	16 8	15 12	14 11	15 23	14 33	16 44	14 59	19 72	14 95
	7	19 24	19 6	20 10	19 15	18 14	19 27	18 37	20 48	20 62	23 74	21 93
2b	8	23 25	22 5	24 11	23 15	21 16	22 27	21 39	23 50	21 63	26 75	21 91
	9	26 26	26 5	27 11	26 14	26 18	26 30	25 40	27 52	27 65	30 78	28 91
3a	10	30 31	30 13	30 16	30 17	30 19	30 38	29 47	30 58	30 70	31 82	29 92
	11	33 39	33 23	34 24	32 24	34 22	33 49	33 58	34 67	36 78	35 87	35 94
3b	12	37 47	37 35	38 34	36 32	37 26	36 54	36 65	36 74	37 85	38 91	36 96
	13	40 50	40 41	41 40	40 39	40 33	40 53	40 64	40 73	41 83	44 93	42 82
4a	14	44 51	44 43	45 42	45 42	44 37	43 53	44 61	43 71	44 82	45 93	43 80
	15	48 51	48 42	48 41	49 41	48 40	47 56	47 61	48 70	49 82	49 92	49 81
4b	16	51 50	52 40	52 38	52 38	51 36	51 56	50 64	50 72	52 84	50 92	49 96
	17	55 50	55 41	55 35	56 37	55 34	54 56	54 66	55 74	57 85	56 92	56 95
5a	18	58 48	58 37	58 34	60 36	59 34	58 54	58 65	58 73	59 84	59 92	57 96
	19	62 46	61 31	62 31	63 33	62 34	62 54	62 61	62 71	63 83	65 90	64 96
5b	20	65 42	65 25	66 26	65 29	65 31	65 50	66 56	65 67	64 79	66 86	64 96
	21	69 37	69 22	70 20	68 23	68 27	69 44	70 51	70 62	70 74	70 84	71 94
6a	22	73 31	74 16	73 15	73 18	72 23	72 36	74 45	72 55	72 68	72 80	71 93
	23	77 28	76 12	77 12	77 16	77 23	76 31	77 40	77 50	77 65	77 77	78 94
6b	24	80 25	80 7	80 10	81 15	80 20	79 28	79 37	79 46	78 63	80 76	78 93
	25	83 27	82 13	84 8	84 15	83 18	83 32	83 41	84 50	84 67	86 79	85 93
7a	26	87 38	87 43	88 22	87 13	87 26	87 51	86 53	87 56	86 60	88 63	85 66
	27	90 56	91 77	92 53	90 26	89 47	91 73	91 67	91 61	92 51	90 46	91 37
7b	28	93 71	94 100	94 86	92 49	93 60	95 86	93 77	93 60	94 36	90 26	92 10
	29	95 79	96 100	97 100	94 78	95 75	96 85	94 77	94 61	94 36	90 27	90 11
T%	30	96 81	98 100	98 100	96 91	98 76	97 83	94 77	93 61	92 35	88 26	85 10

Table 10.1: Fuzzification and defuzzification: the scaled time (x) and F_0 (y) values are given for each pre-tones (P, 1 to 30 vertically), by grade of membership (1 to 0.1 horizontally), and for the weighted average of all grades (gray column)

Pre-tonal contours On each graph of figures 10.1, 10.2, and 10.3, scaled time [0,100] is placed on the abscissa, and scaled F_0 [0,100] is placed on the ordinate. All syllables have the same duration (1/7th or 14.28% of the sentence).

The first set of graphs (Figure 10.1) corresponds to the fuzzification with the frequency principle, the second set of graphs (Figure 10.2) corresponds to the fuzzification with the similarity principle, and the third set of graphs (Figure 10.3) corresponds to the average of both principles and the results in Table 10.1. Figure 10.4 is the graphic representation of the crisp or defuzzified into-

nation contour of the modality of surprise as extracted by the model after it has processed all the data and merged the results.

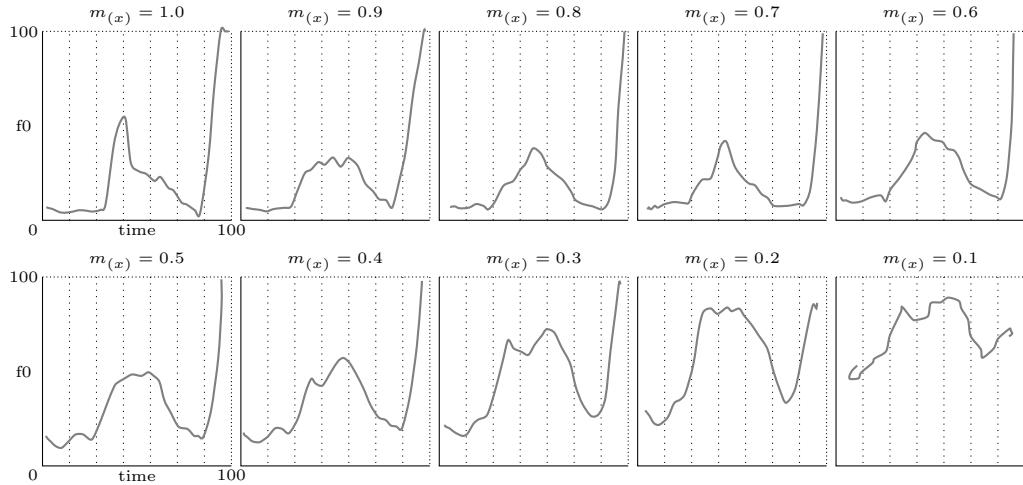


Figure 10.1: Surprise modality contour by grade of membership: frequency.

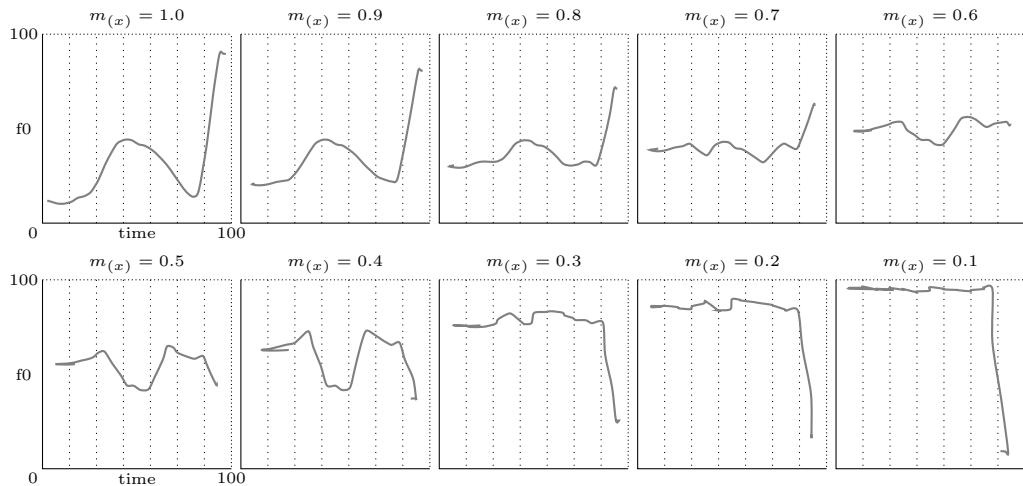


Figure 10.2: Surprise modality contour by grade of membership: centrality.

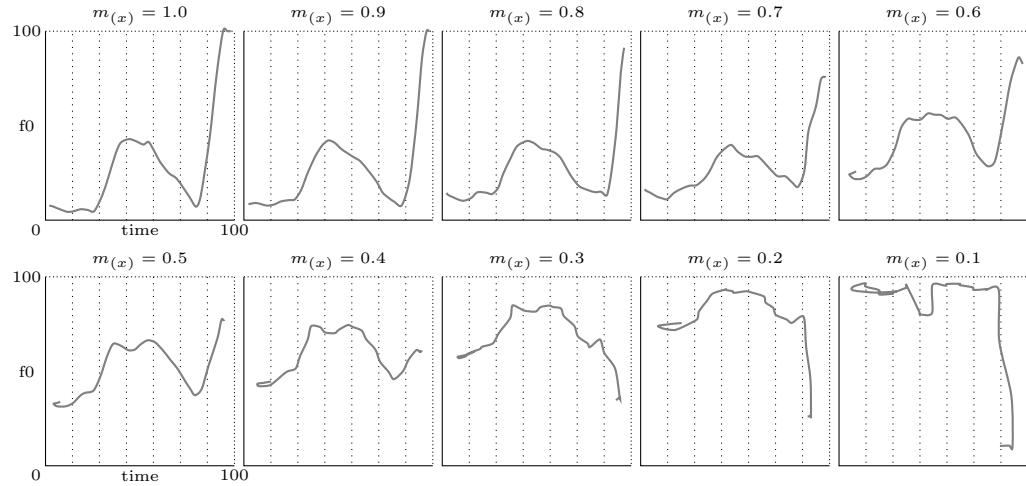


Figure 10.3: Surprise modality contour by grade of membership: frequency and centrality combined.

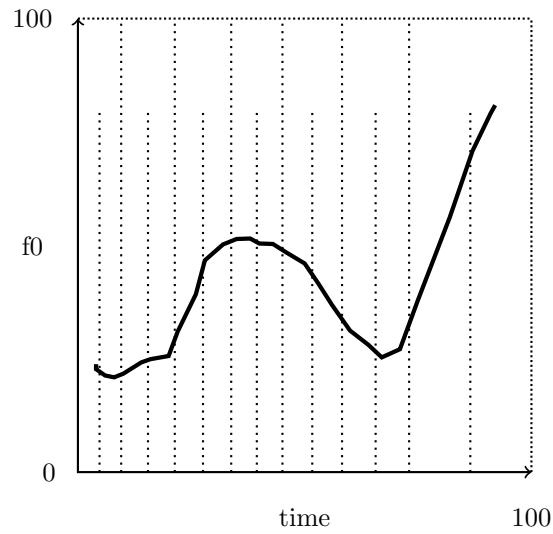


Figure 10.4: Surprise modality pretonal contour: defuzzification

10.1.2 Layer 4: tones and the pattern recognition of the tonal intonation contour

Values from layers 1 & 2: scaling, fuzzification, defuzzification The results presented in this section take into account the presence of both possible peaks. Patterns (d) and (e) – corresponding to a final peak, preceded (pattern e) or not (pattern d) by a secondary lower peak – account for 62% of the sentences. Patterns (a) and (b) – corresponding to a primary high peak, followed (pattern b) or not (pattern a) by a secondary lower high peak – account for 38% of the sentences. Results in Table 10.3 are undifferentiated and incorporate both peaks (as the pre-tonal contours do).

In Table 10.2, results have been separated by groups of patterns ((a/b) and (d/e)). This table has two tiers. In the top tier are the results for patterns (a) and (b) only; the data for patterns (d) and (e) have been removed. In the bottom tier are the results for patterns (d) and (e) only; the data for patterns (a)/(b) have been removed. Even with the data for pattern (d) and (e) removed, the secondary final peak is still almost at the level of the primary peak. Conversely, patterns (d) and (e) do contrast with the global figure and with pattern (b). Patterns d and e are the main contour of the surprise modality, with a variation as (b): the final peak is constant, always realized close to the F_0 range limit and the secondary non-final peak varies in amplitude.

With the data provided by PRInt, the contour of surprise can be characterized as patterns (d), or with a possible variation as pattern (e)(62% of the sentences). Alternatively, it can be realized as pattern (b) (38% of the sentences). Although surprise patterns like unmarked question in terms of its tonal string, the proportion of the different possible patterns is not the same: the corpus of surprise contains twice as many instances of pattern (b) as the

F ₀ (FREQUENCY X CENTRALITY)											
	%T	H ₁			H ₂			H ₃			T%
- P* - m _(x) = 1	7	(4) ^{0.9}	73	(31) ^{0.8}	(3) ^{0.9}	(100) ^{0.9}	(35) ^{0.6}	(3) ^{0.9}	(96) ^{0.8}	(99) ^{0.9}	100
- P* - \bar{m}	21	22	63	46	20	75	46	19	68	75	80
- - P* m _(x) = 1	7	(4) ^{0.9}	73	(31) ^{0.8}	(3) ^{0.9}	65	(3) ^{0.8}	(3) ^{0.9}	100	(100) ^{0.9}	100
- - P* \bar{m}	21	22	63	46	20	55	30	19	82	79	80

Table 10.2: Partial results for the F₀ values of the tones ($m_{(x)} = 1$ and \bar{m}). **Top:** P* as H₂ included and P* as H₃ excluded (patterns a/b). **Bottom:** P* as H₂ excluded and P* as H₃ included (patterns d/e). Values in parentheses are missing for the grade of membership and they are inferred from the closest grade for which the value is available, indicated in superscript.

corpus of unmarked question.

$$\begin{aligned} P^* = H_3 &\rightarrow \%T \overline{L-H-L}^1 \overline{L-H-L}^2 \overline{L-H-L}^3 T\% = L\% (\overline{L-H-L}^2) \overline{L-H*L}^3 \quad (d/e) \\ P^* = H_2 &\rightarrow \%T \overline{L-H-L}^1 \overline{L-H-L}^2 \overline{L-H-L}^3 T\% = L\% \overline{L-H*L}^2 (\overline{L-H}^3) \quad (b) \end{aligned}$$

At this stage, the PRInt model has identified the main pattern(s) of the tonal contour of the modality of surprise. It must still provide more specific information about the alignment of the tones of the contour with the syllabic structure. The pre-tonal isometric grid constitutes the abstract frame of reference common to all sentences, independent of their actual size. To determine the syllabic alignment of the tonal contour, the PRInt model must examine the anchoring of the tones of each peak on pre-tones, for each instance of the contour in the corpus.

Anchoring of tones: pre-tones, pre-tonal co-occurrences Results in Table 10.4 are the highest ranking ($m_{(x)} = 1$) and defuzzified (\tilde{m}) pre-tonal anchoring of the tonal string. Results in Table 10.5 are the pre-tonal co-occurrences of the tones for each peak. As for the question contour, the final

L-H* rise is anchored between the left and right edge of the last syllable.

$m_{(x)}$	T%	L	H	L	L	H	L	L	H	L	T%
TIME											
H ₁											
1	2	29	41	57	-	-	-	86	96	100	100
0.9	-	-	-	58	29	-	-	-	94	-	99
0.8	-	21	40	51	43	-	75	-	99	-	-
0.7	2	21	37	51	44	54	71	87	99	97	-
0.6	4	32	44	52	35	49	67	83	92	95	98
0.5	8	33	45	57	39	53	74	85	88	91	95
0.4	13	33	51	67	49	57	44	71	82	86	93
0.3	23	52	33	31	59	39	35	55	65	67	90
0.2	31	62	16	21	72	28	26	41	55	54	87
0.1	51	71	15	9	78	8	12	24	43	37	79
\bar{m}	9	31	40	52	42	48	61	77	90	90	96
F ₀											
H ₁											
1	7	-	73	-	-	65	-	-	100	-	100
0.9	9	4	82	-	3	67	-	3	-	100	-
0.8	12	3	72	31	5	69	3	5	-	-	-
0.7	16	9	72	31	7	54	4	8	-	-	-
0.6	21	22	63	38	18	49	31	14	-	92	92
0.5	29	30	51	41	28	48	32	25	91	83	82
0.4	38	41	40	61	38	44	53	36	80	71	71
0.3	52	61	24	82	57	27	73	56	56	49	48
0.2	67	72	15	88	70	22	82	72	38	34	34
0.1	86	86	8	93	85	12	92	88	19	17	17
\bar{m}	21	22	63	46	20	55	30	19	82	79	80

Table 10.3: Time and F₀ values of the tones by grade of membership

Peaks		H_1			H_2			H_3			
Tones	%T	L-	H	-L	L-	H	-L	L-	H	-L	T%
m_1	3	9	13	13	10	13	17	27	29	30	30
\bar{m}	3	11	13	15	13	15	19	15	26	26	27

Table 10.4: Pre-tonal anchoring of the tones ($m_{(x)} = 1$ and \tilde{m} for averaged frequency and centrality). Each pre-tone corresponds to a fixed position in the syllabic structure.

	P*		Pa		Pb	
	L-H	H-L	L-H	H-L	L-H	H-L
1	26.29	29.30	29.29	29.30	10.13	13.13
0.9						
0.8			29.30			
0.7	26.28					
0.6					16.17	17.17
0.5	25.28	28.30		27.27	14.17	
0.4			26.27	29.29	10.11	26.26
0.3	28.29	27.30	25.27	27.29	13.17	17.25
0.2	24.28	27.29	19.20	28.30	9.12	13.15
0.1	25.26	27.28	15.16	28.28	7.13	13.19

Table 10.5: Pre-tonal associations by peaks and grades of membership (partial results)

Scaled distance between tones The values for the rise L-H* in the primary peak are the reverse of the values for the fall H*-L. The primary final rise is typically realized without subsequent fall (= 0). For the secondary peak, medium values (around 50) are more typical during both the rise and the fall; the secondary peak symmetrically rises and falls before and after H. These values are almost identical to those of closed question.

Peaks →	Scaled time (%)				Scaled F ₀ (%)			
	Primary		Secondary		Primary		Secondary	
Tones →	L-H	H-L	L-H	H-L	L-H	H-L	L-H	H-L
1	9	0	-	-	-	-	-	-
0.9	10	-	-	0	97	1	46	-
0.8	9	3	11	-	96	-	47	43
0.7	13	3	11	-	76	2	34	23
0.6	13	5	13	-	83	16	41	30
0.5	18	8	16	10	74	27	31	35
0.4	21	11	21	16	63	37	22	48
0.3	28	19	27	26	45	58	77	75
0.2	36	25	30	32	31	71	84	81
0.1	49	34	34	39	16	84	91	89
\bar{m}	14	6	16	12	78	22	45	42

Table 10.6: Distance between tones in scaled time and F₀ for the primary and secondary peak of the surprise contour

10.1.3 Prototypical contour

From the information gathered by the PRInt model, it is possible to establish the prototypical contour of the modality of surprise, relative to the position of the H* tone on the isometric grid (higher/lower, before/after). The result is presented in Figure 10.5

Additional data: velocity, angle Table 10.7 provides additional data gathered by the PRInt model. These data will be used to compare contours in Chapter 11.

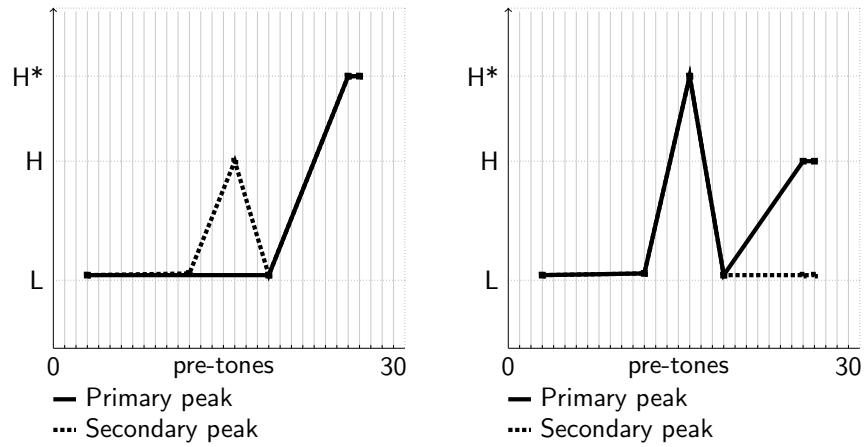


Figure 10.5: Prototypical contour. **Left:** main contour ($P^*=H_3$).
Right: contour variation ($P^*=H_2$)

Peaks →	Velocity (F_0/time)		Angle Primary
	Primary	Secondary	
Tones →	L-H	L-H + H-L	L-H
1.0	6	2 2	84
0.9	8	3 1	84
0.8	5	3 3	80
0.7	9	5 -	83
0.6	12	6 2	78
0.5	15	8 6	74
0.4	21	12 9	65
0.3	34	20 15	48
0.2	46	28 20	37
0.1	73	39 32	21
\bar{m}	13	7 5	75

Table 10.7: Velocity between tones in scaled F_0 per time for the primary and secondary peak of the surprise contour, angle (steepness) of the primary peak.

10.2 The intonation contour of the modality of doubt

10.2.1 Layer 3: pre-tones and the pattern recognition of the pre-tonal intonation contour

σ	$m(x)$	\bar{m}	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
	P	x y	x y	x y	x y	x y	x y	x y	x y	x y	x y	x y
%T	1	6 48	2 44	3 44	2 45	2 32	3 47	8 51	11 52	19 68	32 83	36 95
1a	2	6 46	3 43	4 42	4 42	4 36	3 42	7 44	9 44	13 72	20 84	23 95
	3	6 44	5 39	6 38	6 39	6 36	5 39	6 45	7 47	7 74	7 86	9 95
1b	4	9 42	9 35	9 33	9 34	9 36	8 37	7 45	10 52	8 76	8 86	13 95
	5	12 42	12 30	12 30	12 32	12 31	12 40	10 52	12 64	13 74	13 85	20 94
2a	6	15 40	15 28	16 27	17 29	15 31	15 40	14 50	15 61	15 72	14 84	20 94
	7	19 40	19 28	20 27	20 29	19 28	18 42	18 51	19 62	21 73	20 84	23 94
2b	8	23 39	22 27	24 25	24 28	23 26	22 42	23 51	22 61	24 74	22 84	23 93
	9	27 39	26 28	27 25	27 27	26 22	26 43	26 52	26 63	28 75	27 85	28 94
3a	10	30 38	30 26	31 23	31 25	30 21	30 43	30 50	28 60	30 76	31 86	29 94
	11	34 37	33 24	34 22	34 27	34 21	33 43	33 51	33 57	34 77	37 89	35 95
3b	12	37 35	37 21	39 22	38 27	37 21	37 39	36 49	36 50	35 76	37 89	36 95
	13	41 35	40 21	42 23	41 26	41 25	40 37	39 46	40 50	41 75	41 88	42 94
4a	14	44 36	44 21	45 23	45 24	44 28	44 37	43 45	42 53	44 75	43 88	43 94
	15	48 36	47 21	48 22	49 24	48 27	47 39	47 46	46 57	49 76	48 89	49 95
4b	16	51 35	51 20	52 21	53 23	51 22	51 37	51 48	50 57	51 78	51 90	50 96
	17	55 35	54 21	55 21	56 24	55 17	55 39	55 49	54 60	55 80	58 91	57 96
5a	18	58 36	57 21	58 21	59 25	59 19	58 37	58 49	58 62	57 81	61 90	57 96
	19	62 36	61 21	62 22	62 26	62 20	62 39	61 50	62 62	62 81	65 89	64 95
5b	20	65 35	65 21	66 21	67 22	65 22	66 36	65 47	64 57	65 77	66 85	64 94
	21	69 36	69 20	70 29	70 28	68 20	69 39	69 44	69 56	71 77	70 85	71 94
6a	22	73 51	72 46	74 54	74 53	72 38	72 51	74 49	72 54	73 63	72 65	72 67
	23	77 68	76 74	77 82	78 79	76 60	76 68	77 62	77 55	76 51	78 45	78 39
6b	24	80 82	79 100	81 100	82 93	80 79	80 82	80 77	80 60	78 40	79 28	78 13
	25	84 77	83 90	84 100	85 74	84 73	83 76	83 82	84 65	83 43	83 31	85 15
7a	26	87 61	86 70	88 71	88 49	87 52	86 61	86 72	87 66	86 54	84 48	84 39
	27	91 43	90 40	92 39	92 27	91 29	90 43	89 57	90 61	91 65	89 65	90 66
7b	28	94 31	93 20	94 12	95 22	94 22	93 40	92 48	93 57	92 62	91 63	91 72
	29	96 33	97 12	97 18	97 32	97 32	96 44	95 49	95 52	93 55	92 61	91 71
T%	30	97 36	98 13	98 21	99 37	98 43	97 49	96 53	95 50	92 46	89 55	86 65

Table 10.8: Fuzzification and defuzzification: the scaled time (x) and F_0 (y) values are given for each pre-tone (P, 1 to 30 vertically), by grade of membership (1 to 0.1 horizontally), and for the weighted average of all grades (gray column)

Values from layers 1 & 2: scaling, fuzzification, defuzzification The results of the PRInt model for the pre-tonal coordinates are presented in Table 10.8. Each column from 1 to 0.1 corresponds to a grade of membership. The gray column labelled \bar{m} corresponds to the weighted average (the centroid labelled \bar{m}) of all grades. Each row from 1 to 30 corresponds to a pre-tone (labelled P in the header of the column). Pre-tones are grouped by syllables (1 to 7, horizontal solid lines) and half-syllables (first frame of the syllable σa

and second frame of the syllable σb , horizontal dotted lines).

Pre-tonal contours On each graph of figures 10.6, 10.7, and 10.8, scaled time [0,100] is placed on the abscissa and scaled F_0 [0,100] is placed on the ordinate. All syllables have the same duration (1/7th or 14.28% of the sentence). The first set of graphs (Figure 10.6) correspond to the fuzzification with the frequency principle, the second set of graphs (Figure 10.7) corresponds to the fuzzification with the similarity principle, and the third set of graphs (Figure 10.8) corresponds to the average of both principles and the results in Table 10.8. Figure 10.9 is the graphic representation of the crisp or defuzzified intonation contour of the modality of doubt as extracted by the model after it has processed all the data and merged the results.

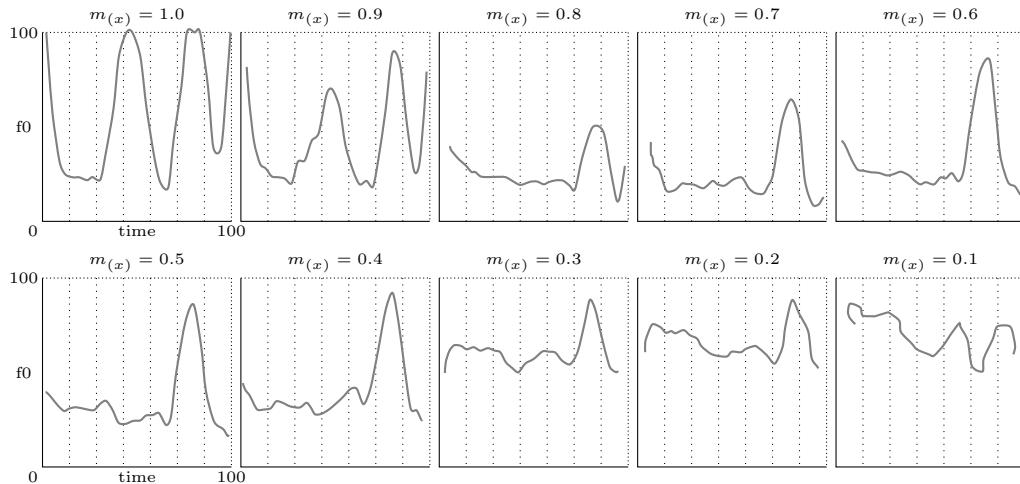


Figure 10.6: Doubt modality contour by grade of membership: frequency.

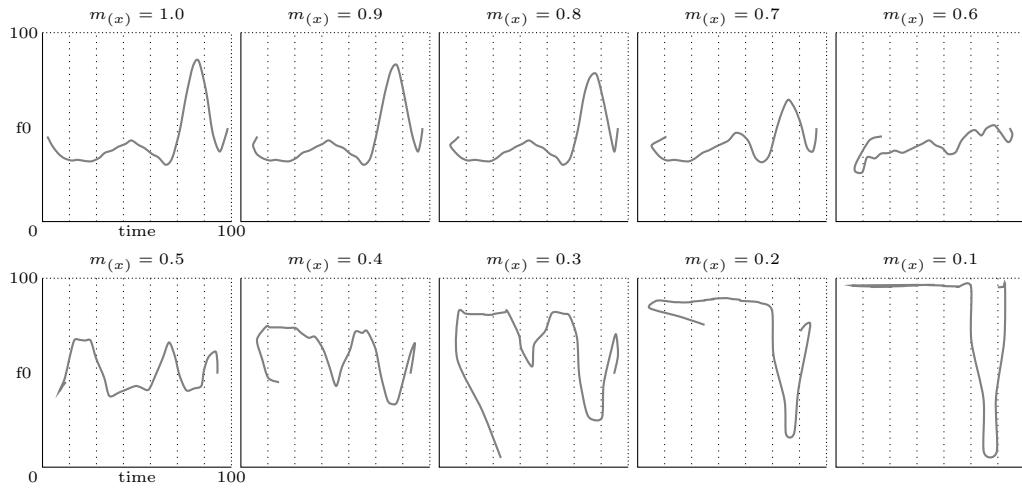


Figure 10.7: Doubt modality contour by grade of membership: centrality.

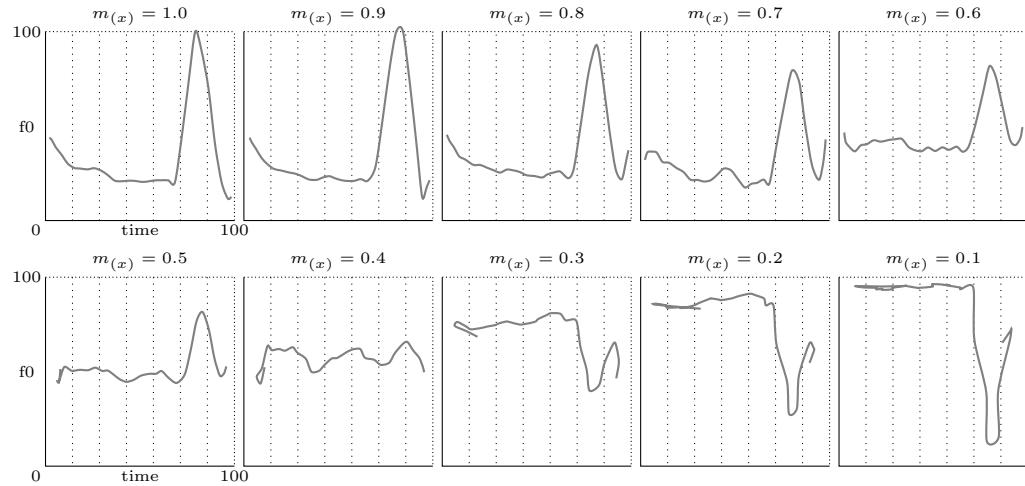


Figure 10.8: Doubt modality contour by grade of membership: frequency and centrality combined.

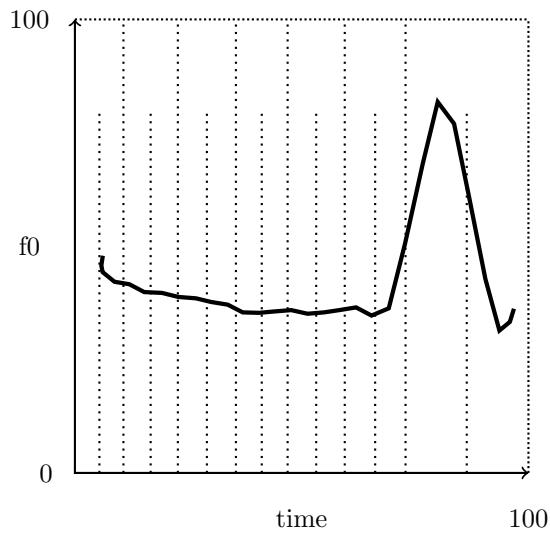


Figure 10.9: Doubt modality pretonal contour: defuzzification

10.2.2 Layer 4: tones and the pattern recognition of the tonal intonation contour

Values from layers 1 & 2: scaling, fuzzification, defuzzification The results presented in this section take into account the presence of both possible peaks. Patterns (d) and (e) – corresponding to a central peak, followed (pattern b) or not (pattern a) by a secondary lower peak – account for 87% of the sentences. Patterns (d) and (e) – corresponding to a primary high peak, preceded (pattern e) or not (pattern d) by a secondary lower high peak – account for 13% of the sentences. Results in Table 10.10 are undifferentiated and incorporate both peaks (as the pre-tonal contours do). The opposition between the two variations (central vs. final peak) is more acute for doubt than it is for question or surprise. The contour of doubt is crucially different in that its primary peak is not sentence-final.

In Table 10.9, results have been separated by groups of patterns ((a)/(b) and (d)/(e)). This table has two tiers. In the top tier are the results for patterns (a) and (b) only; the data for patterns (d) and (e) have been removed. In the bottom tier are the results for patterns d and e only; the data for patterns (a) and (b) have been removed. Even with the data for patterns (a) and (b) removed, the secondary central peak and the primary final peak are almost at the same height: even in secondary position, the central peak remains high. The contour is duplicated in the “double montée” phenomenon. Patterns (d) and (e) stand in contrast to patterns (a) and (b). Patterns (a) and (b) are the main contour of the doubt modality, with a variation as (d) or (e): the central peak is a constant, always realized close to the F_0 range limit and the secondary non-final peak varies in amplitude.

F ₀ (FREQUENCY X CENTRALITY)												
	%T		H ₁			H ₂			H ₃		T%	
- P* $m_{(x)} = 1$	45 ^{0.8}	33 ^{0.8}	67 ^{0.9}	41	21 ^{0.8}	99	38 ^{0.5}	9 ^{0.8}	69 ^{0.9}	50 ^{0.8}	9 ^{0.9}	
- P* \bar{m}	47	48	54	53	37	83	50	34	57	52	47	
- - P* $m_{(x)} = 1$												
- - P* \bar{m}	45 ^{0.8}	33 ^{0.8}	67 ^{0.9}	41	15 ^{0.9}	98 ^{0.8}	9 ^{0.8}	9 ^{0.8}	99 ^{0.8}	50 ^{0.7}	49 ^{0.7}	
- - P* \bar{m}	47	48	54	53	35	76	32	34	70	46	54	

Table 10.9: Partial results for the F_0 values of the tones ($m_{(x)} = 1$ and \bar{m}). **Top:** P* as H_2 included and P* as H_3 excluded (patterns a/b). **Bottom:** P* as H_2 excluded and P* as H_3 included (patterns d/e). Missing values for the grade of membership are inferred from the closest grade for which the value is available, indicated in superscript.

With the data provided by PRInt, the contour of doubt can be characterized as patterns (a) or (b) (87% of the sentences). Alternatively, it can be realized as pattern (d) or (e) (13% of the sentences). Doubt does not pattern like questions or surprise. Although doubt and surprise share a double rise variation, peaks in any doubt variation are fully realized as an L-H-L rise-fall.

$m_{(x)}$	T%	L	H	L	L	H	L	L	H	L	T%	
TIME				H ₁			H ₂			H ₃		
Fuzzification	1	1	-	-	71	85	92	92	98	99	99	
	0.9	-	-	-	28	-	83	-	-	98	-	
	0.8	-	6	-	42	-	81	-	-	98	-	
	0.7	0	29	20	37	33	79	90	91	95	-	
	0.6	3	23	31	36	69	76	86	90	87	94	
	0.5	8	25	35	37	64	73	80	82	83	87	
	0.4	16	26	25	35	57	60	69	72	74	79	
	0.3	30	54	56	64	36	39	47	54	58	60	
	0.2	54	65	68	68	25	28	34	41	46	47	
	0.1	60	77	76	78	11	14	19	24	29	31	
\bar{m}		10	27	35	40	55	74	78	81	86	87	
F ₀				H ₁			H ₂			H ₃		
Fuzzification	1	-	-	-	41	-	99	-	-	-	-	
	0.9	-	-	67	46	-	-	-	69	-	-	
	0.8	45	33	73	42	21	-	-	9	77	50	
	0.7	32	31	63	30	16	-	-	11	64	49	
	0.6	42	42	54	60	31	-	-	30	63	48	
	0.5	43	49	47	64	37	92	38	35	53	46	
	0.4	43	66	33	79	51	83	32	55	40	33	
	0.3	70	77	24	86	72	58	64	77	21	74	
	0.2	83	86	17	93	85	42	76	86	14	87	
	0.1	94	94	11	96	94	24	91	96	5	98	
\bar{m}		47	48	54	53	37	83	50	34	57	46	
54												

Table 10.10: Time and F₀ values of the tones by grade of membership

$$\begin{aligned}
 P^* = H_2 &\rightarrow \%T \overline{L-H-L}^1 \overline{L-H-L}^2 \overline{L-H-L}^3 T\% = L\% \overline{L-H*-L}^2 (\overline{L-H-L}\%)^3 \quad (a/b) \\
 P^* = H_2 &\rightarrow \%T \overline{L-H-L}^1 \overline{L-H-L}^2 \overline{L-H-L}^3 T\% = L\% (\overline{L-H-L}^2) \overline{L-H-L}\%^3 \quad (d/e)
 \end{aligned}$$

At this stage, the PRInt model has identified the main pattern(s) of the tonal contour of the modality of doubt. It must still provide more specific information about the alignment of the tones of the contour with the syllabic

structure. The pre-tonal isometric grid constitutes the abstract frame of reference common to all sentences, independent of their actual size. To determine the syllabic alignment of the tonal contour, the PRInt model must examine the anchoring of the tones of each peak on pre-tones, for each instance of the contour in the corpus.

Anchoring of tones: pre-tones, pre-tonal co-occurrences Results in Table 10.11 are the highest ranking ($m_{(x)} = 1$) and defuzzified (\tilde{m}) pre-tonal anchoring of the tonal string. Results in Table 10.12 are the pre-tonal co-occurrence of the tones for each peak. The H tone of the primary peak is typically anchored on the middle of the antepenultimate syllable, with the L-tone on the left edge of the fifth syllable and the -L tone on the right edge of the last syllable.

Peaks		H_1			H_2			H_3			
Tones	%T	L-	H	-L	L-	H	-L	L-	H	-L	T%
m_1	(2)	(2)	(2)	(12)	21	24	27	24	28	28	29
\bar{m}	2	8	10	12	18	20	23	20	24	25	26

Table 10.11: Pre-tonal anchoring of the tones ($m_{(x)} = 1$ and \tilde{m} for averaged frequency and centrality). Each pre-tone corresponds to a fixed position in the syllabic structure.

	P*		Pa		Pb	
	L-H	H-L	L-H	H-L	L-H	H-L
1	22.25	25.27	28.29	29.30	21.22	22.22
0.9		25.28		29.29		23.23
0.8					22.23	
0.7	21.25					
0.6		26.28	27.29		22.22	
0.5		25.25			13.14	
0.4	22.26	26.27			10.11	
0.3	21.24	24.27			16.16	3.7
0.2	20.25	25.26	27.28	28.30	15.16	3.7
0.1	20.24	24.26	29.30	28.29	14.14	10.10

Table 10.12: Tonal associations by peaks

Scaled distance between tones The time span of the L-H* rise (= 15) of the primary peak is slightly longer than the following H*-L fall (= 10). Rise and fall are equal in terms of F_0 ($L-H^*=+65$; $H^*-L=-62$). The same is true for the secondary peak, though it is shorter in time ($L-H=10$; $H-L=11$) and 3 times as low in F_0 ($L-H=+22$; $H-L=-17$).

Peaks →	Scaled time (%)				Scaled F ₀ (%)			
	Primary		Secondary		Primary		Secondary	
Tones →	L-H	H-L	L-H	H-L	L-H	H-L	L-H	H-L
1	12	-	-	-	-	-	-	-
0.9	9	0	0	0	83	-	1	1
0.8	12	-	-	-	-	90	3	-
0.7	5	7	-	-	63	88	7	2
0.6	17	9	6	7	65	66	25	7
0.5	19	11	10	10	55	65	34	17
0.4	24	17	14	16	42	44	43	29
0.3	30	23	23	25	21	28	62	52
0.2	37	29	29	33	14	12	71	67
0.1	53	34	43	40	4	8	84	82
\bar{m}	15	10	10	11	62	65	22	17

Table 10.13: Distance between tones in scaled time and F₀ for the primary and secondary peak of the doubt contour

10.2.3 Prototypical contour

From the information gathered by the PRInt model, it is possible to establish the prototypical contour of the category of the modality of doubt, relative to the position of the H* tone on the isometric grid (higher/lower, before/after). The result is presented in figure 10.10

Additional data: velocity, angle Table 10.14 provides additional data gathered by the PRInt model. These data will be used to compare contours in Chapter 11.

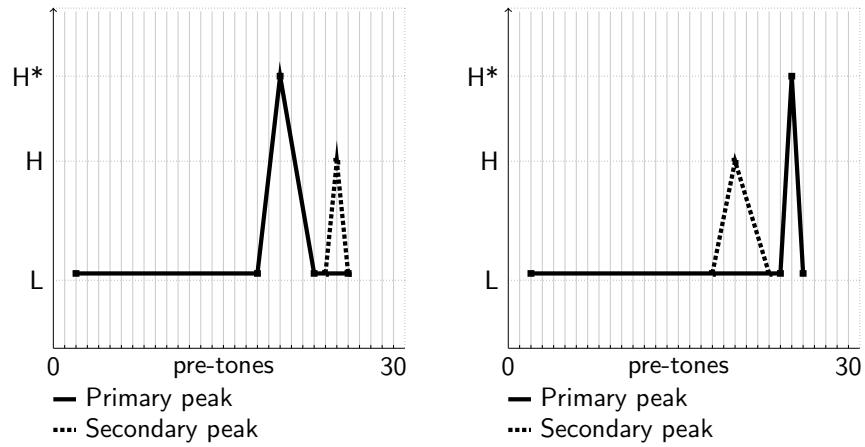


Figure 10.10: Prototypical contour. **Left:** main contour ($P^*=H_3$). **Right:** variation of the contour ($P^*=H_2$)

Peaks →	Velocity (F_0/time)		Angle Primary
	Primary	Secondary	
Tones →	L-H	H-L	L-H
1	5	5	81
0.9	2	5	80
0.8	5	8	81
0.7	5	8	79
0.6	10	12	74
0.5	15	16	68
0.4	24	24	61
0.3	41	47	50
0.2	57	-	39
0.1	95	96	26
\bar{m}	13	13	73

Table 10.14: Velocity between tones in scaled F_0 per time for the primary and secondary peak of the doubt contour, angle (steepness) of the primary peak.

Chapter 11

Comparison of the three intonation contours

From the analysis of corpus data by the PRInt model, it was possible to characterize the intonation contours of three communicative functions of French as phonological tonal strings: unmarked closed question and two modalities of closed question, one expressed with doubt and one expressed with surprise. In this chapter, the results obtained with the PRInt model are used to determine whether these strings are phonologically distinct or phonetic allophones. This is possible because, with the implementation of the 4-layer structure, the PRInt model separates the pattern recognition process of the tonal sequences (phonological intension) from that of their physical variations (phonetic extension).

11.1 Tonal specification of the contours

In this application of the PRInt model, the tonal pattern of sentences has been analyzed as three possible L-H-L compounds: one L-H-L compound for the primary peak (P*), one compound for the precedent peak (Pb), and one compound for the following peak (Pa). Pa and Pb may or may not be found. The three contours are compared for the placement of their primary compound only: secondary contours with more than one peak are variations from the main pattern and represent a subset of instances. From the results of the PRInt model, the three contours have been described as the following

phonological tonal strings:

Question L% L-H* H%

Surprise L% L-H* H%

Doubt L% L-H*-L L%

The tonal specification of closed question found by the PRInt model is similar to that given elsewhere in the literature (e.g. Beyssade et al., 2007), and the model worked well for this contour. The modalities of doubt and surprise have not been as systematically studied as closed question but the tonal specifications provided by the PRInt model correspond to the description of the contours given by Fónagy & Bérard (1973). Closed question and surprise have the same tonal specification: they should be considered as a single phonological unit. The PRInt model provides the quantitative evidence of the phonetically allophonic status of surprise compared to closed question. Thus, not only does the PRInt model extract the phonological structure of intonation from the analysis of variation, but it can also re-generate this variation, organized by degrees of typicality.

11.2 Prototypes: intonational –vague– intension

Each prototype generated by PRInt is the symbolic representation of its category and its internal structure. This representation is abstracted from all the phonetic instances to which the category name applies and of which each utterance in the category is an instance (in accordance with the definition of a prototype discussed in Chapter 2 about categorization). In Figure 11.1, the three prototypical contours have been plotted on the isometric grid for comparison, abstracting away from the variation in syllabic duration. They

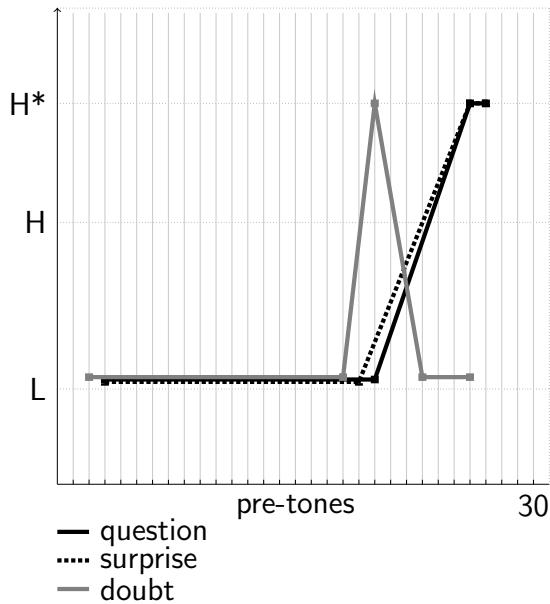


Figure 11.1: The three prototypical contours on the isometric grid.

indicate the “vague” implementation of the contours over a sentence. They provide the placement of the tones within the pattern in relational terms from the highest tone in the chain. They are the intonational intension or phonological representation of the concepts (question, surprise, doubt) and represent the abstract implementation of the contours on sentences.

With the isometric grid, a prototypical intonation contour can be characterized without any reference to specific time or F_0 values but only as a string of tones (H^* , H , L , T%) relative to possible syllabic anchoring (pre-tones). The prototypical contour of doubt is clearly distinct from the other two. Its primary compound is not sentence-final and comprises all three tones ($L-H^*-L$, rise-fall), spanning six pre-tones; tonal anchoring should take place over at most three syllables from the end of the sentence. The $-L$ tone should

be anchored at the maximum to the left edge of the last syllable. This specification is due to the presence of a final plateau that spans four pre-tones (thus anchored over one or at most two syllables) and that leads to the final L% tone.

The primary compound of both closed question and surprise have the same characteristics. It is sentence-final and only comprises two tones (L-H*, rise), the last -L being merged with the H tone. The compound spans 7 or 8 pre-tones; tonal anchoring should take place over at most three syllables, the -L tone being anchored at the maximum to the right edge of the last syllable. The final boundary tone H% immediately follows the H* tone.

In Figure 11.2 the prototypical contours have been plotted within the scaled frame of reference (100 X 100). In scaled time and F_0 , the PRInt model records a difference between the contours of closed question and surprise in the position of the L- tone of the rise (the turning point between the plateau and the rise); the L- tone of the surprise contour occurs one pre-tone earlier and lower than the L- tone of the closed question contour. The contours of closed question and surprise track closely in F_0 versus time, except that the L- tone for the surprise contour has a lower relative F_0 . This difference is irrelevant at the level of tonal specification (both contours are phonologically identical) but indicates a difference in phonetic implementation.

The three hatched rows at the top of the figure represent the relative duration of the syllables in each contour. The three colored zones represent the range of the dataset for each contour, limited by the extreme values of the membership continuum [0.1, 1]. The alignment of the tones with the syllabic grid is indicated by numbers corresponding to pre-tonal position. The alignment of the tones in relative time, as shown on the figure, differs from

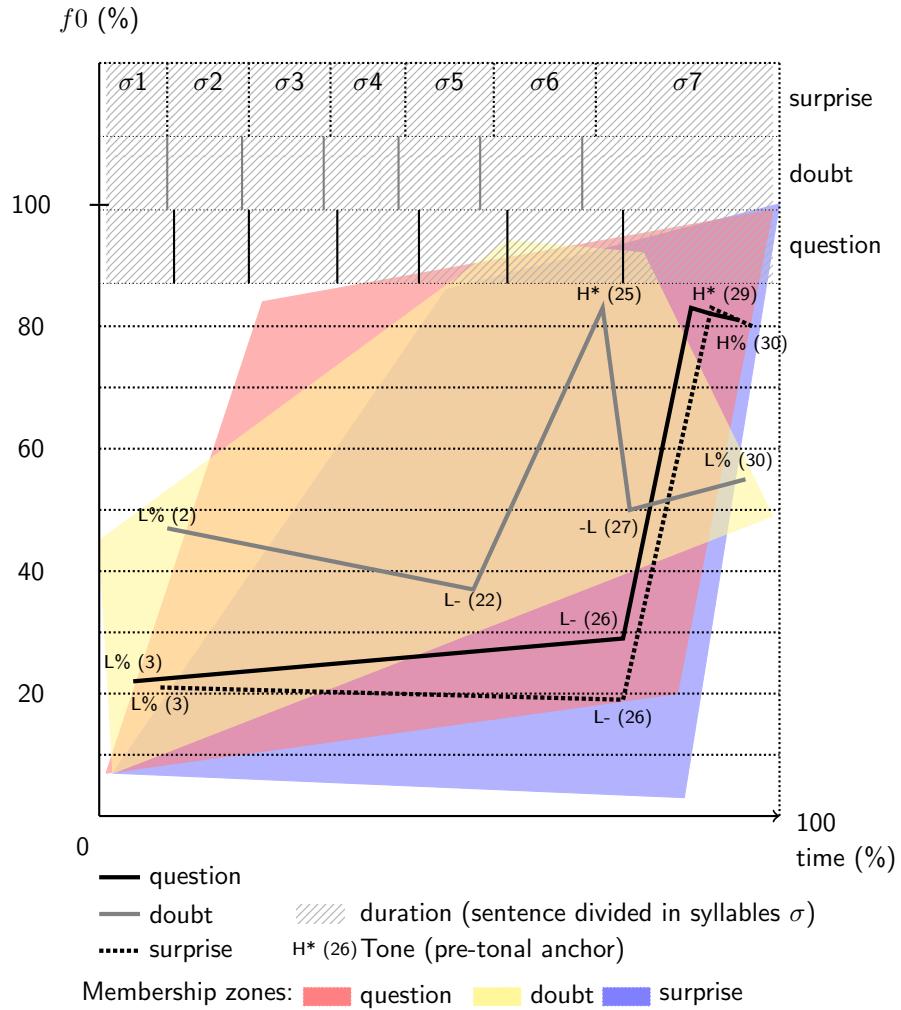


Figure 11.2: Scaled crisp (\bar{m}) contours of closed question, surprise, and doubt.

the pre-tonal alignment.

The H^* tone of the contour of doubt is within the penultimate syllable, the L^- tone is in the antepenultimate, and the $-L$ tone is in the last syllable, together with the plateau and the final $L\%$ tone. This last syllable is actually

almost three times as long as the average duration of the other six syllables, making the duration of the final plateau proportionally almost as large as the L-H*-L compound itself .

The similarity between the contours of closed question and surprise is apparent and is confirmed by the convergence of data obtained separately for each dimension of the contours that have been analyzed by the PRInt model:

- the L-H* bi-tones are on the last syllable, the L- tones are near or on the left syllable boundary
- the heights of the H* tones are identical (82 percent points for closed question, 83 for surprise)
- the time spans between the L- and H* tones are identical (14 percent points)
- the velocity of the F_0 excursions between L- and H are identical (12 for closed question, 13 for surprise)
- the angles formed by the plateau and the rise are identical (74° for closed question, 75° for surprise)
- the last syllable, on which the bi-tones occur, is about twice as long as the average duration of the other six syllables in both cases.

Another striking point of similarity between the two contours is the fuzzy distributions by grade of membership of the relative time and F_0 values of the L-H* bi-tones: they are almost exactly the same, except for the high grades of membership (above 0.5) of L- tones marking the end of the plateau and the start of the rise. This single difference leads to a slight incline of the closed question's plateau (+ 10 percent points), while the surprise's plateau remains flat. This correlates with a difference in the span of the final rise (analyzed

as a separate variable by PRInt: +68 percent points for closed question, +78 for surprise). However, in a study on the recognition of the patterns of closed question vs. declarative, Vion & Colas (2002) have shown that subjects were unable to reliably categorize the contours solely from the first part of a utterance prior to the tonal movement itself. In the absence of the distinctive rising bi-tones, subjects categorize the initial contour of the question as a declarative in 71% of the cases. If only the bi-tone on the last syllable is discriminating for the contours, then, according to the results of the PRInt analysis, closed question and surprise have the same contour; the difference in F_0 span in rise between questions and surprise is not phonologically distinct. According to the literature, surprise is a para-linguistic phenomenon that is not phonologically discrete (Ladd, 2008). Surprise indicates the “emotional” response of the speaker to what have just been said (Fónagy & Bérard, 1973). Question and surprise are not phonologically distinct, but are non-discrete allophonic variations (other parameters, such as intensity or velocity, might also be relevant for the distinction of these contours). From the phonetic implementation of the prototypical contours, the PRInt model provides evidence of the allophonic nature of surprise compared to closed question.

11.3 Phonetic implementation: intonational –vague– extension (degrees of typicality and allocontour)

The structure and dimensions of a sentence’s string of tones are expressed relative to internal points of reference: the extreme values of the sentence that constitute its frame of reference. For the purposes of categorization and extraction of prototypical patterns, variations in the size of the frame of reference have been abstracted by scalar quantization so that the features of

all instances in a category may be compared on the same scale. However, as the ATLM abstracts from actual values to tones using to the 4-layer structure, it does not discard information from layer 1 of the structure: it stores the actual physical dimensions of the frame of reference of all utterances.

For each intonation contour, the size of the phonetic prototypical frame of reference can be calculated along with its graded range of variation. The phonetic prototypical frame of reference corresponds to the ideal space, expressed in centiseconds (cs) and Hertz (Hz), within which the prototypical tonal structure is phonetically implemented. From a geometric perspective, each utterance is contained in a rectangle delimited by the extreme values of the utterance (min-max time (cs); min-max F_0 (Hz)). It is thus possible to calculate the prototypical dimension of this rectangular space (or frame of reference) among sentences of a category by re-employing the same data that have been used for the calculation of the tonal structure of each sentence.

These data are arranged into multisets: {max time} and {min time} for sentences and syllables, {max F_0 } and {min F_0 } by subject, gender, and as a combined dataset. They are fuzzified and defuzzified with the AFC module as has been done with the rest of the data. The prototypical phonetic frame of reference is obtained for each category of contour along with the variation in the overall duration of the sentence, the duration of individual syllables, and extreme F_0 values, all organized by grade of typicality on the [0.1, 1] continuum.

The dimensions of the prototypical contours expressed so far in relational terms or as percentages, can be re-scaled to the dimensions of the prototypical phonetic frame of reference expressed in Hertz and centisecond. This transformation provides an external point of reference since all contours

are then expressed on the same scale outside of their own category. This procedure enables the analysis of the phonetic implementation of the contours at various levels of typicality and the study of the status of each subcategory (i.e. closed question, surprise and doubt) in the general category of intonation contours of French.

11.3.1 Prototype (\bar{m})

The phonetic implementation of the prototypical contours is presented in Figure 11.3. The colored rectangles represent the prototypical spaces for the phonetic implementation of each contour in centiseconds and Hertz, as has been extracted from the fuzzy analysis of the frames of reference of all utterances. The prototypical frame of reference of surprise (purple rectangle) is noticeably higher, and larger in both dimensions than the frames of the two other contours:

1. The F_0 span of surprise is one third larger than that of question and two thirds larger than that of doubt. Its lower and higher limits are also higher than those of the other two contours.
2. The overall duration of the sentence is longer for surprise and doubt than for closed question. The first syllable of each contour has about the same duration. In all three contours, the last syllable ($\sigma 7$) has a longer duration than any other syllable. However, the last syllable is much longer for doubt and surprise than it is for closed question. In the case of closed question, $\sigma 7$ is not twice as long (x1.7) as the mean duration of preceding syllables. This contrast is stronger for surprise and doubt. For these contours, $\sigma 7$ is more than twice as long as the mean duration of preceding syllables (x2.4 for doubt; x2.2 for surprise).

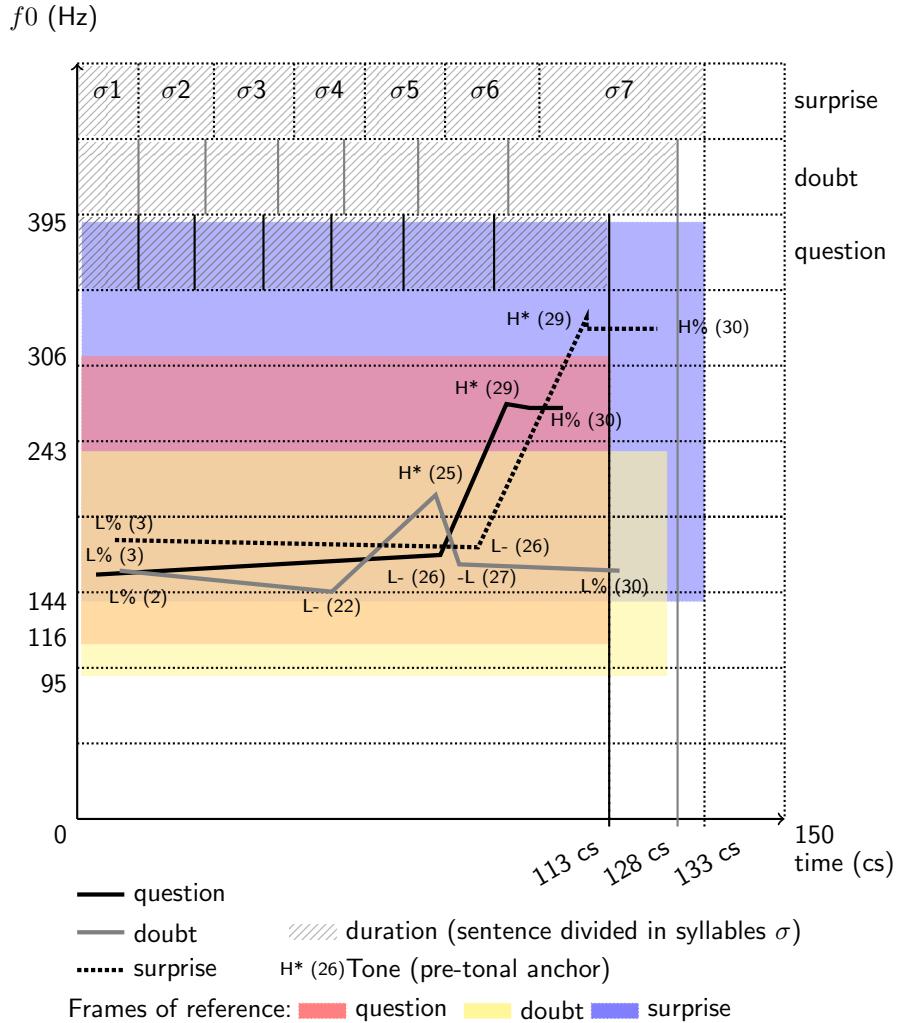


Figure 11.3: Zone of ideal phonetic implementation (frame of reference) of the prototypical contours (\bar{m}) of closed questions, surprise, and doubt. The prototypical contours have been scaled towards the range of their own frame.

3. Finally, the distance between the two tones L- and H* is larger for surprise (23 cs) than closed question (14 cs), but since the F0 span is also larger for surprise, the F0/time ratio is the same (7Hz/cs). Although the

two contours vary quantitatively in the actual dimensions of their frames, because they are relatively the same (see Figure 11.2 in the previous section), the L-H* compounds are stretched proportionally, with surprise being a phonetically scaled up version of question, an allocontour.

11.3.2 Highest degree of typicality ($m_{(x)} = 1$)

The phonetic implementation of the tonal strings with values corresponding to the highest grade of typicality ($m_{(x)} = 1$) reveals more of the phonetic contrast. Figure 11.4 displays the reconstructed three contours at grade 1 of typicality, implemented in their corresponding frames of references. This represents the aggregation of all the highest ranking values, as calculated separately for each feature analyzed by the PRInt model, in what would be the best phonetic implementation of the prototypical contour extracted by the analysis by the PRInt model of all utterances among all participants in the study. Figure 11.4 is the phonetic ideal expected from the instantiation of the prototype when all features are realized with the values that were the most frequent and the most central among all users. Whereas the prototypical contour is a centroid figure subsuming all degrees of typicality and representing the category as a whole, the contours for specific degrees of typicality provide more detailed information about the category structure itself.

For speakers, the highest degree of typicality corresponds to the creation of maximum phonetic contrast for otherwise phonologically similar contours. This prevents the effect of vagueness that would render pattern identification and communication problematic.

The frames of reference of closed question and doubt are almost identical. They have about the same F_0 span (151 and 163 Hz, respectively), their

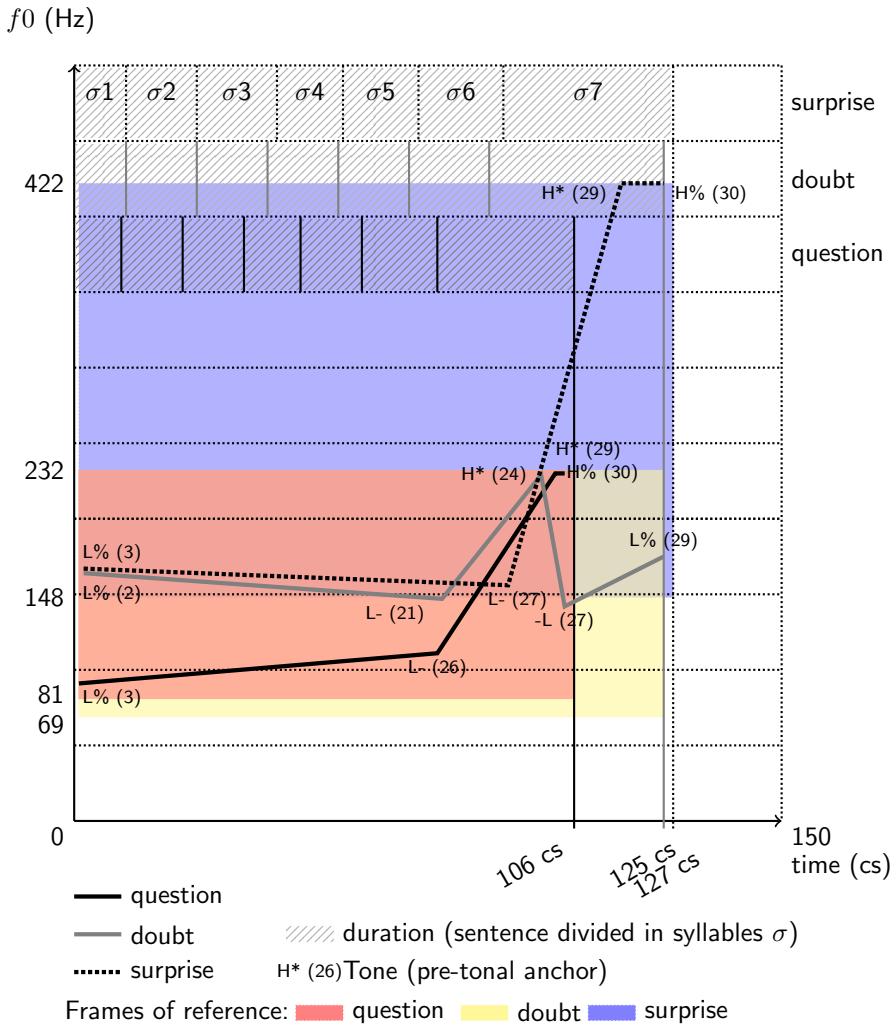


Figure 11.4: Phonetic implementation of the three contours at their highest grade of typicality ($m_{(x)} = 1$)

lower limits are close (81 Hz for question; 69 Hz for doubt), and their higher limits are identical (232 Hz). By contrast at this level of typicality, the F₀ span of surprise (274 Hz) is much larger than closed question or doubt. First, the lower level of the contour of surprise (148 Hz) corresponds to the middle

range of the two other contours, and second, the higher limit of surprise (422 Hz) is almost twice as high as the higher limit of the two other contours. Since closed question and doubt are phonologically distinct, they need not be contrasted by phonetic parameters. The overlapping frames of reference of closed question and doubt can be considered as the “normal” F_0 range. The speakers’ F_0 range for surprise is strikingly larger than closed question for maximum contrast. Furthermore, the corpus of closed question and that of contrastive modalities (doubt vs. surprise) were recorded separately at different times. The fact that the frames of reference of closed question and doubt match in spite of their distance in time is another indicator of the existence of a “normal” F_0 range within which phonologically distinct contours are realized.

At the highest level of typicality, the anchoring of the tones to the pre-tonal structure is the preferred one for the participants in the study. The implementation of the contour of doubt conforms to the intension of the prototype. Its L-H*-L tonal compound is not sentence-final and spans three syllables. In summary:

L% is anchored to the left edge of the 1st syllable (pre-tone 2)

L- is anchored to the right edge of the 5th syllable (pre-tone 21)

H* is anchored to the center of the 6th syllable (pre-tone 24)

-L is anchored to the left edge of the 7th syllable (pre-tone 27)

L% is anchored to right edge of the 7th syllable (pre-tone 29)

When calculated as co-occurring pairs, the positions of tones are slightly different than when calculated individually. The favored anchoring is over two syllables: the L- tone is inside the sixth syllable, on its left edge (pre-tone 22),

and the H* tone is on the right edge of the same syllable. The position of the -L tone remains the same (pre-tone 27). In any case, the H* tone is within the penultimate syllable.

The implementation of the contour of closed question and surprise conform to the intension of the prototype. Their L-H* tonal compound is final and spans one syllable. In summary:

L% is anchored to the left edge of the 1st syllable (pre-tone 3)

L- is anchored to the left edge of the 7th syllable (pre-tone 26 for closed question/ 27 for surprise)

H* is anchored to the right edge of the 7th syllable (pre-tone 29)

H% is anchored to right edge of the 7th syllable (pre-tone 29 for closed question/ 30 for surprise). However, pre-tones 29 and 30 are usually merged.

When calculated as co-occurring pairs, the positions of tones are also slightly different for these two contours. The favored anchoring is over one syllable: the L- tone is anchored to the left edge of the last syllable (pre-tones 26/27) for both contours and the H* tone remains anchored to the right edge of the same syllable (pre-tone 29).

The F₀ velocity in the L-H* bi-tones of closed question and surprise, calculated as a separate dimension, is the same for both contours on a relative scale (= 6 F₀ percent points per time percent point) but it is different for their phonetic implementation: 5 Hz/cs for closed question and twice as much for surprise or 11 Hz/cs. The difference is in the amplitude of the F₀ span and not in the duration of the movement (about 24 cs in both cases). There is a tradeoff between making the F₀ difference as contrastive as possible between

closed question and surprise, and maintaining the ratio between time and F_0 to reduce distortion of the scaled up pattern of surprise. The contrast at the highest level of typicality between closed question and surprise is marked by a sharp difference in steepness that presumably have a perceptual impact in order to facilitate the identification of the emotional state of the speaker in the question and so as to ensure that both the meaning of question and the emotion are conveyed.

11.3.3 Borderline degree of typicality ($m_{(x)} = 0.5$): vagueness

Indeed, if the contours of closed question and surprise were to be not only phonologically identical but also phonetically the same, it would lead to a case of vagueness, wherein no phonetic cues would make closed question and surprise distinguishable. In a graded category (or a fuzzy set), 0.5 is the grade of membership corresponding to the tipping point for typicality. An object with a grade of 0.5 can equally belong in two categories without the classifier being able to make a membership decision. Figure 11.5 is the phonetic implementation of contours at grade 0.5 of typicality.

At grade 0.5, the L-H*-L compound of the contour of doubt occurs earlier in the sentence, with its H* tone anchored to the right edge of the 5th syllable and the following -L tone anchored to the left edge of the 6th syllable. As a result, the plateau leading to the sentence-final boundary tone L% spans two syllables, compared to one syllable in the prototype or grade $m_{(x)} = 1$. The F_0 range is considerably narrower than at grade 1 of typicality: 64Hz, or a decrease in size of -61% from grade 1.

More interestingly, the contours of closed question and surprise are merging. The L- tone of the bitonal rise is anchored earlier at grade 0.5 than

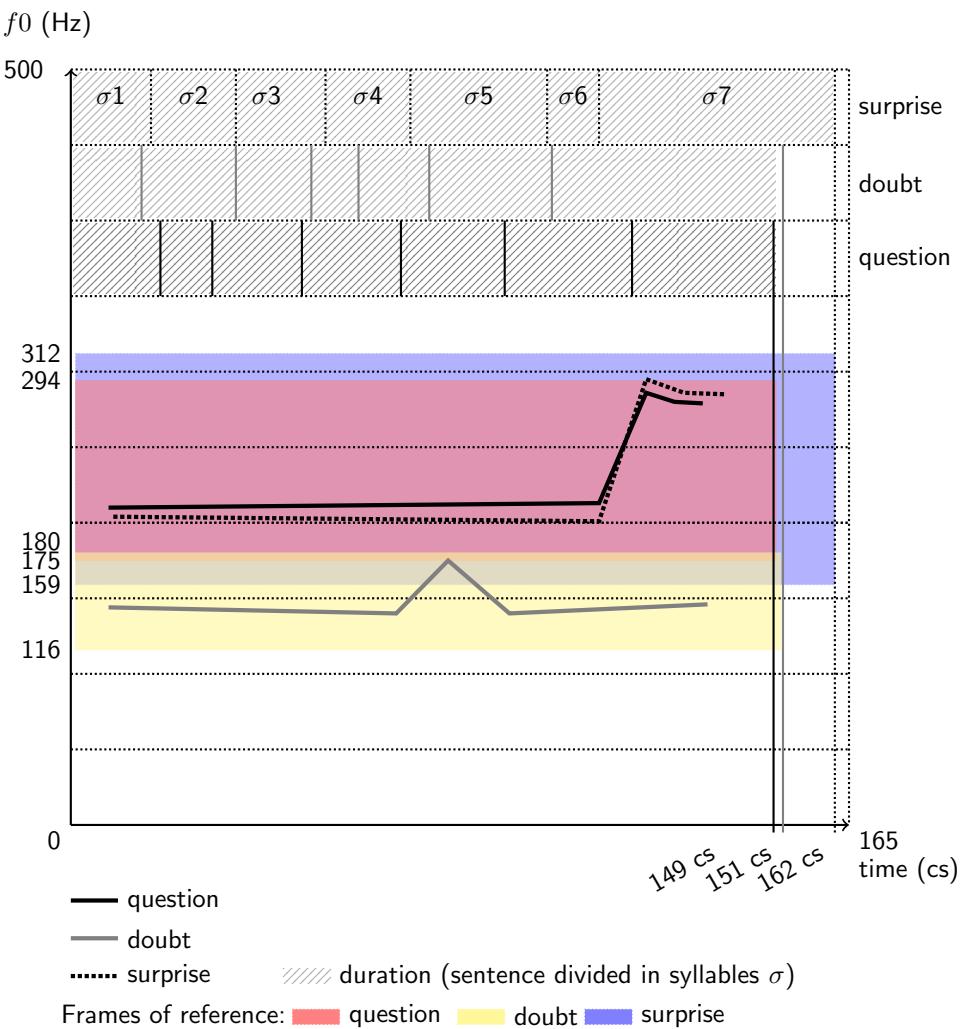


Figure 11.5: Phonetic implementation of the three contours at their medium (vague) grade of typicality $m_{(x)} = 0.5$

at grade 1 of typicality, to the left edge of the penultimate syllable. The H* tone is also anchored earlier, to the left edge of the last syllable. Overall, the whole bi-tone occurs earlier, by the span of one syllable. The L-H* rise crosses the boundary between the penultimate and the last syllable, creating the space

for a short final high plateau H* H% in the last syllable.

At grade 0.5, the frame of reference of closed question is narrowed (-21% from grade 1) and translated up along the Hz axis. The frame of reference of surprise is narrowed even more (-44% from grade 1) and translated down along the Hz axis. Remarkably, both frames end up being vertically aligned as the central F_0 value is about 235 Hz for both.

The tonal contours of closed question and surprise are also almost identical. The initial plateaus are flattening out, the L- tone of surprise being translated up and the one of close question being translated down. The time distance between the L- and H* tones is identical (about 10 cs) between the contours and the spans of the F_0 excursions are closer in value than they are at grade 1: 73 Hz for close question and 94 Hz for surprise. Accordingly, the time to F_0 ratios are not contrastive anymore at grade 0.5 of typicality: 7 for closed question and 9 for surprise ($\times 1.3$ from closed question to surprise, compared to $\times 2.2$ at grade 1 of typicality). The steepness of the bitonal L-H* rise is the same for the contours of closed question and surprise at grade 0.5; it was twice as large for surprise as for closed question at grade 1. There are no longer any phonetic cues to distinguish between close question and surprise at the 0.5 level of typicality. The “vagueness” present in the implementation of the phonetic cues of closed question and surprise will most likely lead to the identification of the contour of the surprised question uniquely as a closed question, without its additional para-linguistic information.

11.4 A note on secondary peaks

Fónagy (1979) noticed the existence of *accentual arcs* in French, or groups of syllables between two accents (high tones). These accents are asso-

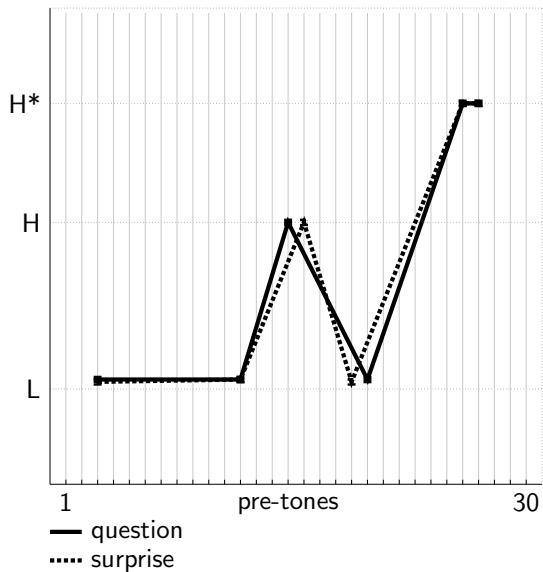


Figure 11.6: Prototypical contours of question and surprise on the isometric grid, secondary peak included

ciated with phrase final syllables and are generally regarded as the prosodic bracketing of syntagmatic groups. According to studies on these accentual arcs (Jun & Fougeron, 2002), they usually span three or four syllables.

In the corpora used in this study, the secondary peak phenomenon was observed for closed question and surprise (Figure 11.6), and, to a lesser extent, for doubt. In his description of closed question in contemporary French, Fónagy discusses the case of the *double montée* (“double rise”) that corresponds to the duplication of the L-H* rise. The duplicate rise takes place where an accent of an accentual arc occurs (the end of a phrase). In the present study, the secondary peak would be expected to occur on the third or fourth syllable. Since the PRInt model is designed to find several H tones, the secondary peak of closed question and surprise have been compared.

Two main results were found. First, in accordance with Fónagy's description, the H* tone of the secondary peak prototypically occurs on the 4th syllable: pre-tone 16 for closed question, pre-tone 15 for surprise, around the middle of the 4th syllable. Second, the secondary peak follows the trend of the primary peak in that, if it is realized, it is much higher for surprise than it is for closed question. Although the F_0 excursion in the secondary peak L-H rise is markedly different between the two contours (+31 Hz for question; +170 Hz for surprise), because of the parallel difference in time span of the rise (8 cs for question; 27 cs for surprise), the F_0 to time ratios differ only slightly: 4 Hz/cs for closed question and 6 Hz/cs for surprise. However, there is a sharp difference in the H-L falls. The H-L bi-tone is almost flat for question (-0.4 Hz/cs, H-L decrease = -13 Hz). By contrast, it is somewhat steeper than the rise for surprise (-8 Hz/cs, H-L decrease = -170 Hz). Fónagy (1979) described the duplication phenomenon as a "linguistically well-tuned tonal configuration", a contour leading to a choppy and emphatic redundancy in the melody when the speakers are more "animated." Overall, the presence of an accentual arc is marked by a normal rising of F_0 on the last syllable of a group in the case of closed question and it is emphatically realized in the case of a surprise.

11.5 A note on primary non-final peaks - Secondary contour

A secondary contour was found in 21% of the instances of closed question and in 38% of the instances of surprise: the primary peak was not sentence-final (H_2) and was followed by a secondary sentence-final rise (H_3) of smaller amplitude. The tonal string of this secondary contour of closed question and surprise is %L L-H*-L L-H% (Figure 11.7).

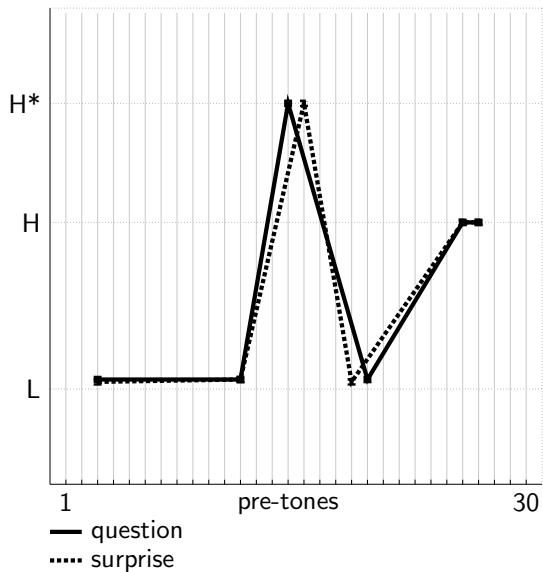


Figure 11.7: Prototypical contours of question and surprise on the isometric grid, primary non-final peak

The PRInt system assigns the H^* tone to the highest F_0 value ($= 100$ percent points) and then continues on to find lower H peaks. For the Print model, lower values starts at 99, and secondary H peaks can actually be very close or very far in value from the primary peak H^* .

Like all other features, the F_0 values of the primary non-sentence-final H^* are ranked for frequency (mode) and centrality (median), and then these two rankings are averaged. In the case of surprise, the results are similar for both rankings and the highest grade of typicality for the this H^* tone is assigned to $F_0 = 100$: mode and median of the distribution are close. This is not the case for closed question. The F_0 values of H^* are more spread over the continuum [1, 100]: mode and median of the distribution do not coincide. In the overall ranking, the high frequency of $F_0 = 100$ is matched

with its centrality in the distribution for surprise, but not for closed question. This leads to the situation presented in Figure 11.8. In line with Fónagy's

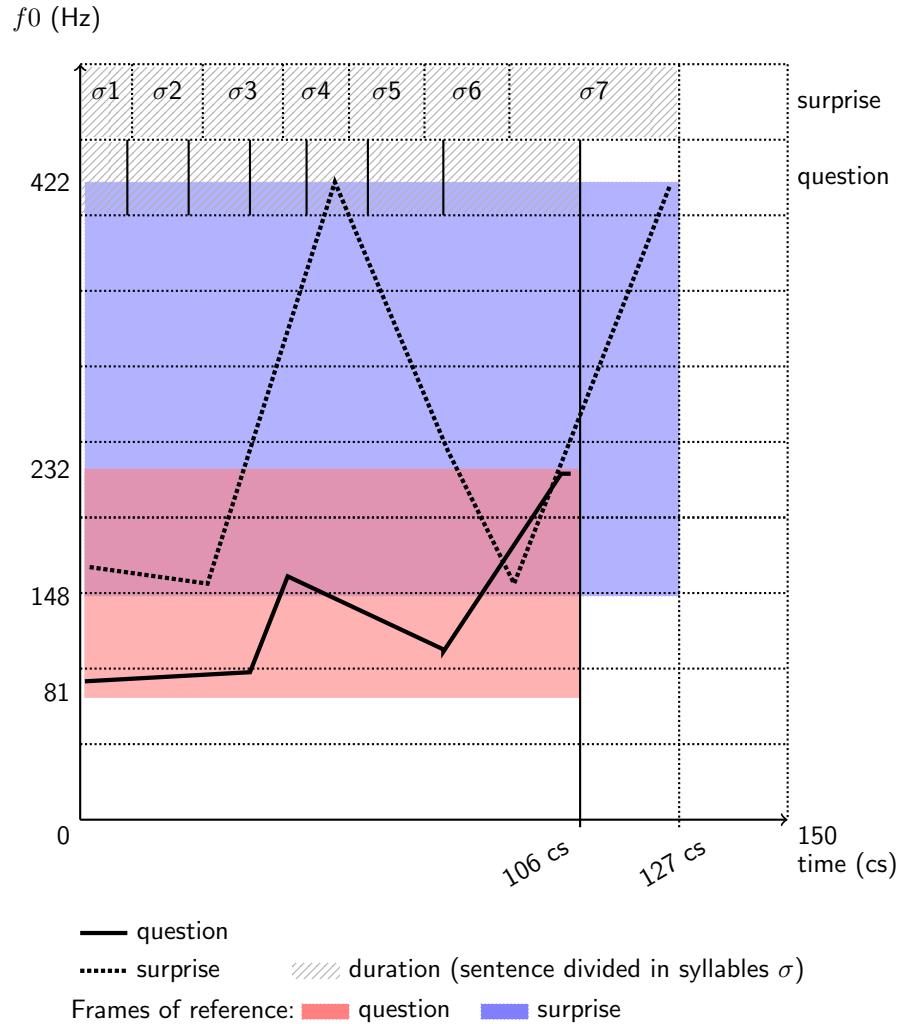


Figure 11.8: Phonetic implementation of question and surprise with their secondary peak (highest grade of typicality ($m_{(x)} = 1$))

interpretation of the double rise as an “animated” contour, the participants

in the study realized the surprise contour more often and more markedly as a double rise than they did for closed question. As a result, the PRInt model extracts a similar tonal pattern for both contours, but with a different phonetic implementation, as was the case for the primary contour. The H* tone of the contour of surprise is phonetically realized as the highest point of the contour and is followed by a slightly lower L-H% final rise. The H* tone of the contour of closed question is phonetically realized as lower than the final H% tone. The case of the non-sentence-final primary peak of these two contours exemplifies how the PRInt model uses all the information in the categories both to extract the prototypes and to re-construct the contours phonetically at various level of typicality.

11.6 Conclusion

The PRInt model extracted the phonological tonal structures of three modalities of questions (i.e. unmarked, with doubt, and with surprise). It also produced the prototypical contours of each modality on the isometric grid. This grid represents the intension of the contours that indicates in a “vague” way how the contours should be implemented phonetically. The grid provides the placement of the tones in the pattern in relational terms from the highest tone in the chain. The PRInt model calculated that only the contours of closed question and doubt are phonologically different. Surprise is a phonetic variation of closed question, an emphatic allocontour of closed question with larger ranges in F_0 and time.

By analyzing separately the variation of the intonation contours and the variation of their frames of reference, the PRInt model calculated that closed question and doubt are realized within the same “normal” F_0 range, while

surprise is realized in a higher and much larger range. Consistent with the principles of categorization and fuzzy set theory, the contour of the allophonic variation of surprise is merging with that of the unmarked closed question at the categorical threshold of 0.5. At this point, the categorical status of the contour is vague in regard with the “surprise” modality.

The PRInt model also analyzed the variation of the contours of closed question and surprise with a secondary peak and it was found that the secondary peak patterns like the primary peak: more F_0 amplitude leading to a phonetic contrast between the two contours, especially in the H-L fall of this secondary peak.

Finally, from the results of PRInt, it is possible to describe the case of a secondary contour for closed question and surprise, with a primary non-sentence-final peak, as a case of “double montée”, which is also much more phonetically marked for surprise than it is for closed question.

Chapter 12

Conclusion

The aim of this dissertation was to experimentally support the case for an intonational phonology and to test one of the core ideas of the AM model, according to which it is possible to extract the phonological structure of intonation patterns from the observation of their phonetic variations.

12.1 Summary of the project

To verify this hypothesis, it was first necessary to characterize intonation in the larger domains of categorization and pattern recognition. It was argued in chapters 2 and 3 that language relies on both processes. Without categorization, every object in the world would be singularly different from any other. Human beings naturally group together objects that present more or less similar features, and create categories for them. Categories are thus loosely defined by a concept that takes the shape of a generic name. Objects in a category can be ranked by degree of typicality, and it was argued that a prototype constitutes the abstract center of gravity of all objects in the category: an abstraction that subsumes the category and of which each object is an instance. These ideas were illustrated by the phonemic categories of /t/ and of vowels. It was discussed that categorization applies to intonation patterns as well, and more specifically to their physical phonetic implementations.

Thus, a parallel was made between the phonological structure of an intonation pattern and the prototype of a category. To experimentally test the hypothesis, a computerized model (the PRInt model) was created to artificially recreate a process of categorization.

12.1.1 The PRInt model

Three corpora containing a large number of instances of French intonation patterns were acquired by elicitation. The three patterns were those associated with unmarked closed question and two modalities of closed question: one pattern expressing doubt, the other expressing surprise. These instances were manually parsed for syllables, converted into data sets in text files, and entered in the PRInt model.

The PRInt model consists of two modules. First, a pattern recognition system encodes each instance in a category as a string of tones, in a way analogous to that developed in the AM model bi-tonal approach to intonation (7). The procedure included scaling, segmentation, and feature extraction.

Second, the string of tones of all sentences within a category were compared to one another by a fuzzy classifier (chapter 8). This classifier organized the values of the features by degree of typicality in their categories. The classifier used a function of frequency and a function of central tendency to compute these degrees of typicality. As output, the PRInt model generated the prototypical/phonological structure of the intonation pattern for each category. It also generated the variation of the phonological structure by degree of typicality: from the most typical (most frequent and most central) to the least typical (least frequent and least typical).

In parallel, the PRInt model also analyzed the actual physical variation

of the phonetic implementations of the intonation pattern. It organized these observed values by degree of typicality as well (chapter 13).

12.1.2 Main results of the application of PRInt

From the results provided by the PRInt model, and within the limits of the chosen parameters (sentence duration, syllable durations, F_0), it was possible to determine the phonological and phonetic status of three contours: unmarked closed question, doubt, and surprise (chapter 11).

- The PRInt model found that unmarked closed question and doubt have a distinct phonological structure. Both extracted structures also match the previous descriptions of the contours in the literature:
 - The contour of unmarked closed question is characterized by a sentence-final H^* peak and the sequence $L-H^*H\%$, where H^* is the highest point of the sentence associated with the right boundary of the last syllable and $L-$ is associated with the left boundary of the last syllable.
 - The contour of doubt is characterized by a non-sentence-final H^* peak and the sequence $L-H^*-L L\%$, where H^* is the highest point of the sentence, associated with the penultimate syllable and the L tones are associated with the right and left boundaries of the preceding and following syllables, respectively. A plateau spans the duration of the last syllable.
- The PRInt model found that the phonological structure of surprise was identical to the structure of unmarked closed question. Therefore, the contours were compared for their phonetic implementations.

- It was found that, at the highest degree of typicality, the pattern of unmarked closed question is implemented within the “normal” F_0 range of the subjects while the pattern of surprise is implemented within an F_0 range that is both higher and much larger than the “normal” F_0 range. Thus, the pattern of surprise is stretched out both in duration and F_0 . The intonation contour of surprise is distinguished only by its phonetic implementation. It has been characterized as an allocontour of the pattern of unmarked closed question, or to use the word of Fónagy & Bérard (1973), an “emphatic” version of the contour of unmarked closed question. However, even though it is gradient in nature, the contrast between the two implementations at the highest level of typicality is so sharp that it almost seems binary.
- This status of allocontour was further supported by the comparison of the phonetic implementation of the two patterns towards the categorical threshold (a grade of membership of 0.5). The two contours are realized within the same central and narrow F_0 range, midway between the typical range of questions and the typical range of surprise. At this level, the two contours are phonetically identical (or almost so).

During the second elicitation task, the participants were instructed to produce contrastive patterns for doubt and surprise. According to the results of the PRInt model, they intuitively separated the two contours by their characteristic features: a particular phonological structure for doubt, and a particular range of phonetic implementation for surprise.

The PRInt model found the presence of a phrasal accent, as described by Fónagy (1979) and Jun & Fougeron (2002). This accent is larger in the case of surprise.

Finally, for all three contours, the last syllable was consistently longer than the preceding syllables. It twice as long as the highest degree of typicality. This result also matches the description of French in the literature.

In conclusion, the PRInt model demonstrated that it is possible to extract the phonological structure of an intonation pattern from the observation of its phonetic implementations. Furthermore, the PRInt model analyzed phonetic variation and its range. The model separated the phonological and phonetic levels of analysis of intonation from the ranking of the phonetic implementations by degree of typicality, as hypothesized in chapter 1. These results support the case for an intonational phonology based on a two-tone approach (AM model).

12.2 Further developments and ameliorations

The PRInt model has not covered many aspects of intonation.

Intensity Like F_0 and duration, intensity is physically gradient. The analysis of intensity could easily be implemented in the PRInt model, in exactly the same manner as F_0 . Its role in the phonological structure of intonation patterns might, *a priori*, be limited. But only a thorough and systematic analysis of its variation should lead to any conclusion. The role of intensity in the phonetic implementation seems more obvious. As shown by Fónagy & Bérard (1973), the intensity in an instance of surprise is markedly higher than in an instance of a closed question, especially in the last syllable.

Lexical stress The language used in this first application of the PRInt model was French. French does not have a functionally distinctive lexical accent or stress as, for example, English or Spanish do. The input for the PRInt model needs only to specify the syllable boundaries. For languages with lexical accents, this information should be entered, along with the syllabic information, into the PRInt model. The model’s analysis of some intonation patterns that relies on lexical stress might give some insight regarding the association of tones (phonological level) and their actual phonetic alignment (phonetic implementation), prototypically and by degree of typicality.

Variable number of syllables and reduced corpus One limitation of the PRInt model, as it has been conceived so far, is in the number of syllables in the input. It is necessary to develop the model so that it can analyze instances of intonation patterns with different numbers of syllables. Furthermore, the number of instances in the input was intentionally very large to ensure that the model would have enough material to analyze an intonation pattern. However, it has been suggested that human beings can actually create categories from a very low number of instances, and can modify the category (ranking and prototype) when new instances are encountered and are incorporated to the category (Feldman, 1997). These two issues were partially addressed in a study by Montreuil & Bacuez (2012). An intonation pattern of a variety of Norman (a dialect of Northern France) was chosen for its peculiarity. Certain declaratives are realized with a L-H-L% tonal sequence over the last two syllables, with the position of the H* tone moving from before to after the syllabic boundary. The corpus contained only 83 instances of the pattern, implemented over sentences comprising 4 to 11 syllables. The ad-hoc solution to the different number of syllables was to align all instances by the end of

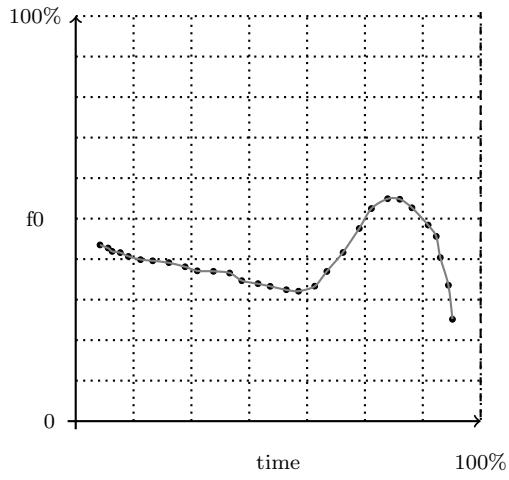


Figure 12.1: A characteristic intonation pattern from a variety of Norman: a declarative ending with a L-H*-L% contour. Pre-tonal representation from the PRInt model, at the highest degree of typicality

the tonal compound (the L% tone), which, in this particular case, also corresponds to the end of most instances. Thus, this intonation pattern adequately fit the model. More importantly, from the small batch of instances, the model generated a phonologically prototypical structure for the intonation pattern that matches the previous characterization (see figure 12.1). In spite of these results, the PRInt model requires adjustments to analyze instances with more flexibility, accommodating both a smaller set of instances and instances varying in syllable count.

Development of an additional module for the model: evaluation of new instances The PRInt model was designed to extract the phonological structure of an intonation pattern from the analysis of its phonetic variation. Once it has achieved this goal, it has acquired some “knowledge” about the

intonation pattern itself. It can apply it towards the classification of new instances not belonging to the original corpus. This will be the task of an additional module (under development), which will assess the degree of typicality of an instance of an intonation pattern (or part of it) in any category that the PRInt model will have previously analyzed. As an illustration, the degree of typicality of the two sentences from chapter 7, [JP72] and [GR65], have been evaluated towards the data stored in the model (tables 12.1 and 12.2). These two sentences were assigned the same tonal structure: L%L-H-L-H*H%. However, the relative values of these tones were not identical in terms of pre-tonal/syllabic alignment and relative F_0 value. Most distinctively, the secondary H tone was low for [JP72] (=9) and high for [GR65] (=69). The pre-tonal and relative F_0 values of the two sentences have been assigned the degree of typicality of their closest match among the ranked data. Crucially, [JP72] and [GR65] differ in the degree of typicality of their secondary peak in the category of closed questions: the H tone of [JP72] scores 1 for its relative F_0 value while the H tone of [GR65] scores a low 0.3. The mean degree of typicality of [JP72] and [GR65] is almost identical: 0.86 and 0.88, respectively. What makes [JP72] a more typical instance of the intonation pattern of closed questions than [GR65] is its mean degree of typicality for relative F_0 : 0.95 vs. 0.8 (on a scale of 0.1 to 1).

From the examples of Norman and of the evaluation module, it is obvious that the PRInt model needs more flexibility in its mode of operation. Otherwise it will have to always rely on very large, closely formatted corpora to organize the category of an intonation pattern and extract its phonological structure (and also to evaluate new sentences). In a way, it should be developed to become a better learner (especially so that it can “generalize” from the

[JP72] _ Pb P* 

	Tones	#T	L-	H	-L	L-	H	-L	L-	H	-L	T#
Layer 4												
Layer 3	Pre-tones	1	-	-	-	16	18	26	26	29	29	30
	$m_{(x)}$	1	-	-	-	0.5	0.9	0.5	1	1	1	1
Layer 2	relative F ₀	5	-	-	-	4	9	4	4	100	100	100
	$m_{(x)}$	1	-	-	-	1	1	0.7	0.9	1	1	1

Table 12.1: Evaluation of the degree of typicality of [JP72]

[GR65] _ Pb P* 

	Tones	#T	L-	H	-L	L-	H	-L	L-	H	-L	T#
Layer 4												
Layer 3	pre-tones	3	-	-	-	7	13	18	26	29	30	30
	$m_{(x)}$	1	-	-	-	0.4	0.7	0.9	1	1	1	1
Layer 2	relative F ₀	12	-	-	-	22	69	15	4	100	100	100
	$m_{(x)}$	0.9	-	-	-	0.7	0.3	0.7	0.9	1	1	1

Table 12.2: Evaluation of the degree of typicality of [GR65]

observation of a small set of instances). One way to achieve this is by incorporating more sophisticated tools from computer science and computational linguistics into the model.

Additional Chapter 13

Phonetic variation and variation of the frame of reference

Scaling the frame of reference (scalar quantization) gives it the same relative dimensions among all instances, which allows the parallel scaling of the patterning features (pre-tones and tones). Time and F_0 are scaled into a Cartesian plane of 100 on 100 points. Each of the syllables has a duration equal to the same fraction of the utterance duration. However, the frame of reference varies in dimension. An important step in the pattern recognition process is to capture the scope of the variation of the grid onto which the feature vector is anchored. The grid's variable dimensions are the overall duration of the sentence, the duration of the syllables, the F_0 baseline and its maximum excursion. For each dimension, the PRInt model calculates the range between the minimum and maximum values found in the sets and ranks each value in the set relative to the difference between these two extremes.

The analysis of the variation of the frame of reference is presented using the data from the corpus of unmark closed question. The first part of the procedure consists in ordering each category of values by reference to the extreme values of the categories. This is achieved by scaling the values towards the extreme values and then by fuzzfying/defuzzifying the scaled values. Thus, scaled values are ordered by grade of membership, from 0.1 to 1. In the second part, ordered scaled values are returned to actual values, Hertz and

centiseconds.

13.1 Scaling, fuzzification, defuzzification

13.1.1 Sentence

Preparation Sentence duration is the first set of measures to be processed by the PRInt model. First, a set is created in which the values for the duration of all sentences are collected. The elements (sentence durations) in this set can have any value above zero: $\{0, +\infty\}$. Even if the segmental information of the sentences limits the sentence duration to some degree (no seven syllable sentence is likely to be seven seconds long, for example), the number of potential durations remains close to limitless within a small range. Consider for example the duration of the first ten sentences produced by one of the subjects (in seconds): 1.134691822, 1.13308915, 1.117004294, 0.953860571, 1.053926706, 1.076705108, 1.026795584, 1.010026309, 1.17956664, 1.056480607, 1.332923083. The probability of finding two sentences with exactly the same duration is near to zero.

To enable the fuzzification procedure, it is necessary to scale the values of the elements contained in the set so that they all are included in the finite range of integers $\{0, \dots, 100\}$. The longest (MAX) and shortest (MIN) durations found in the set form the limits of the observed range of variation (RAN), expressed in centiseconds (cs). In the set of closed questions, the maximum duration is 189 cs, the minimum duration is 63 cs, and the difference between the two (RAN) is 126 cs ($= 189 - 63$). This difference between MIN and MAX, labelled RAN, is called the *differential range* hereafter. MIN and RAN are the values of reference for the purpose of scaling the values of all the elements in the data set.

The PRInt model calculates the difference in duration between each sentence and the shortest one (MIN). This difference is then expressed as a percentage of the differential range (RAN), to the nearest integer. For example, sentence CC01 has a duration of 114 cs. CC01 is 51 cs longer than the minimum value ($114 - 63 = 51$). This difference is then expressed as a percentage of the differential range : 51 is 40% of 126.

$$\begin{aligned} \text{relative duration (\%)} &= ((x - MIN)/RAN) * 100 \\ CC01 &= ((114 - 63)/126) * 100 \\ CC01 &= 40\% \text{ of } RAN \end{aligned}$$

This transformation bins together close values. CC01 (1.134691822 s) and CC02 (1.13308915 s) are both longer than the shortest sentence by 40% of the differential range. Thus, instead of counting CC01 and CC02 as two single elements with two distinct values, the PRInt model will count them as two elements with the same value. All sentence durations are now expressed as values between 0 and 100.

Fuzzification The AFC fuzzifies and defuzzifies the set of relative values according to the general procedure described in Chapter 8. Results are presented in Table 13.1 on two lines: on the upper line are the grades of membership; on the lower line, the values represent the portion of the differential range that the sentence spans at each given grade of membership. At the highest grade of membership, the sentence is longer than the shortest sentence by 34% of the differential range.

$m_{(x)}$	0.5	0.6	0.7	0.8	0.9	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
%	1	7	14	20	27	34	42	48	55	61	68	75	82	88	96

Defuzzification value: 39%

Table 13.1: Relative duration of sentences by grades of membership

13.1.2 Syllables

Preparation First, one set is created for each of the seven syllables. In each syllable set, the duration values of that syllable are collected for all sentences. For example, the durations of all first syllables in the corpus are grouped in the first syllable set. Similarly to sentence duration, syllable duration can have any value, $\{0, +\infty\}$, with the same phonemic limitation noted for sentences; a syllable is unlikely to be five seconds long. But again, the number of potential values is too large to expect significant repetitions throughout the corpus. To enable the fuzzification procedure, it is necessary to scale the values so that they are all included in the finite range of integers $\{0, \dots, 100\}$. For each sentence, the PRInt model converts the durations of the seven syllables into a percentage of the total duration of that sentence. This transformation preserves the relative proportion of each syllable in its sentence and it makes corresponding syllables comparable among sentences of different durations. To illustrate the process, the duration data of two sentences are displayed below:

Sentence	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total
ED51 (cs)	14	12	9	17	12	28	31	122
(%)	11	10	7	13	10	23	25	(47)
PC64 (cs)	8	7	10	13	12	13	22	85
(%)	9	8	12	16	14	15	25	(17)

Samples ED51 and PC64 have a duration of 122 cs and 85 cs respectively, which represents 47% of the global sentence differential range for the

former and 17% for the latter. In both cases, the 7th syllable equals 25% of the sentence duration, even though this syllable is physically a third shorter in sample PC64 than it is in sample ED51. The syllable sets now contain relative durations from 0 and 100.

Fuzzification The AFC fuzzifies and defuzzifies the set of relative values according to the general procedure described in Chapter 8. Results are presented in Table 13.2.

Since the PRInt model processes each of the seven syllable sets independently, the sum of all syllables for each level of membership grade does not equal 100. The PRInt model must subsequently adjust the values to match a total of 100 at each grade of membership. Thus, values are made comparable between grades of membership. For example, the value for the first syllable at grade 0.1 of membership is 27% of the sentence duration and the total of all syllables is 192% (see the row highlighted in light gray in table 13.2). The value is adjusted as follow:

$$\begin{aligned} \text{adjusted duration (\%)} &= x(\%) * (100 / \text{sum of 7 syllables}) \\ (m_{\text{syllable1}} = 0.1) &= 27 * (100 / 192) \\ &= 14 \end{aligned}$$

In Table 13.3, the PRInt model has adjusted the results of Table 13.2 for each syllable and at each grade of membership. Durations of sentences and syllables have now been completely scaled and fuzzified.

$m_{(x)}$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total	
Fuzzification	1.0	9	12	12	11	12	14	25	95
	0.9	9	11	12	11	12	14	25	94
	0.8	9	11	12	11	11	15	24	93
	0.7	9	10	12	11	12	17	24	95
	0.6	15	11	16	18	18	25	26	128
	0.5	19	11	19	21	22	27	29	146
	0.4	21	20	22	22	24	29	32	168
	0.3	22	21	24	24	26	31	21	169
	0.2	25	23	25	25	28	34	18	177
	0.1	27	25	27	28	31	36	19	192
\bar{m}	13	13	15	15	16	20	25	117	

Table 13.2: Fuzzification and defuzzification of the duration of syllables (all durations expressed in %). Grade 0.1 serves as an example of the process (see table 13.3 in this section)

$m_{(x)}$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total	
Fuzzification	1.0	10	12	13	12	13	15	26	100
	0.9	10	11	13	12	13	15	27	100
	0.8	10	12	12	11	12	16	26	100
	0.7	9	11	13	12	13	18	25	100
	0.6	12	9	13	14	14	19	20	100
	0.5	13	8	13	14	15	18	20	100
	0.4	12	12	13	13	14	17	19	100
	0.3	13	12	14	14	15	19	13	100
	0.2	14	13	14	14	16	19	10	100
	0.1	14	13	14	15	16	19	10	100
\bar{m}	11	11	13	12	13	17	23	100	

Table 13.3: Adjustment of fuzzy duration to 100 (all durations expressed as a percentage of the duration of the sentence)

13.1.3 Fundamental frequency (F_0)

Three F_0 values undergo fuzzification: the maximum value of sentences, the minimum value of sentences, and the F_0 span (max-min) of sentences.

Preparation (1): values of reference The elements in the three sets of F_0 values (max, min, ran) can have any value: $\{0, +\infty\}$. To enable the fuzzification procedure, it is necessary to scale the values so that they all are included in the finite range of integers $\{0, \dots, 100\}$. The first step towards scaling is to find the reference values of the data: the extreme values of F_0 throughout all the data. The maxima and the minima are collected at three different levels: first at the level of individual subjects, second, at the level of genders (female/male), and third, at all levels, as a pooled dataset.

SUBJECT LEVEL: For each subject's corpus of sentences, the PRInt model analyzes each sentence to collect its maximum F_0 value (max), its minimum F_0 value (min), and its F_0 range (ran = max - min). It then finds the maximum and the minimum of each type of value: the largest max value (MAXmax) and the smallest max value (MINmax), the largest min value (MAXmin) and the smallest min value (MINmin), the largest ran value (MAXran) and the smallest ran value (MINran). In the table below, max, min, and ran values are given for six sentences (FF1 to FF102), representative of the corpus of subject FF. The extreme values for the entire corpus are in the two rightmost columns (labelled MAX and MIN).

Sentence	FF1	FF66	FF67	FF70	FF73	FF102	MAX	MIN
max	332	247	406	226	508	336	508	226
min	170	179	242	51	151	165	242	51
ran	162	68	164	175	357	171	357	68

In the corpus of subject FF, the maximum of all sentence maxima (MAXmax) is 508 Hz. The minimum of all sentence minima (MINmax) is 242 HZ, etc.

GENDER LEVEL: When all subjects' corpora have been processed, the PRInt model finds the extremes for each type of reference values by gender. The largest MAXmax for men (M-MAXmax), the smallest MINmax for men (M-MINmax), etc., the largest MAXmax for women (F-MAXmax), the smallest MINmax for women (F-MINmax), etc. :

Gender	MAXmax	MINmax	MAXmin	MINmin	MAXran	MINran
Female	599.5	123.7	251.8	18.3	575.8	38.8
Male	599.8	109.0	151.2	15.3	542.3	17.4

Notice that female and male sentence extremes are similar: MAXmax is around 599 Hz, MINmin are 18 Hz for female and 15 Hz for males. Not surprisingly, the MAXmin measure indicates that females have a broader range for their baseline, which can be noticeably higher than for men (251 Hz for females, 151 Hz for males). There is also a noticeable difference of about 30 Hz between the possible F_0 range of females and males (MAXran is 575 for females and 542 Hz for males). Finally, the MINran measure reveals that some sentences are produced with a very narrow range, most likely borderline contours with very little contrast. For example, sentence GR28's minimum and maximum F_0 values are 119 Hz and 148 Hz, leading to a narrow range of 29 Hz.

GLOBAL LEVEL: Finally, the PRInt model finds the global extremes across genders: G-MAXmax, G-MINmax, G-MAXmin, etc.:

G-max		G-min		G-ran	
MAX	MIN	MAX	MIN	MAX	MIN
599.8	109.0	251.8	15.3	575.8	17.4

Preparation (2): scaling With the values of reference now available, the F_0 values of each sentence are first scaled relative to the gender of their producers, and then scaled globally across genders.

GENDER LEVEL: The max, min, and ran values of each sentence are individually scaled relative to the reference values of the gender of the subject who produced the sentence. The scaling formula is the usual percentage re-scaling. The targeted F_0 value is scaled to a percentage relative to the range of its subcategory (max, min or ran) rounded to the nearest integer. For example, sentence RR22 has a maximum of 154 Hz. It is first adjusted to the minimum of the subcategory max for males ($x - M\text{-MINmax}$) and then compared with the range; this range corresponds to the difference between the largest and the smallest values of each F_0 subcategory:

[max] Maximum F_0 values (max) of both genders are scaled towards the minimum and the maximum of their respective sets of maximum values: F-MAXmax and F-MINmax for females, M-MAXmax and M-MINmax for males. The ranges of these sets are RAN = F-MAXmax - F-MINmax and RAN = M-MAXmax - M-MINmax.

[min] Minimum F_0 values (min) of both genders are scaled towards the minimum and the maximum of their respective sets of minimum values: F-MAXmin and F-MINmin for females, M-MAXmin and M-MINmin for males. The ranges of these sets are RAN = F-MAXmin - F-MINmin and RAN = M-MAXmin - M-MINmin.

[ran] F_0 range values (ran) of both genders are scaled relative to the minimum and the maximum of their respective sets of range values:

F-MAXran and F-MINran for females, M-MAXran and M-MINran for males. The ranges of these sets are $RAN = F\text{-MAXran} - F\text{-MINran}$ and $RAN = M\text{-MAXran} - M\text{-MINran}$.

Here are two examples, one of the maximum F_0 of a sentence produced by a man (sentence RR22), the other of the minimum F_0 value of a sentence produced by a woman (sentence EM76):

$$\begin{aligned}
 \text{relative } f_0 (\%) &= ((x - MIN/RAN) * 100 \\
 RAN \text{ of max for male} &= M\text{-MAXmax} - M\text{-MINmax} \\
 &= 599.8 - 109 = 490.8 \\
 RR22 \text{ max } (\%) &= ((154 - 109)/490.8) * 100 \\
 &= 9\% \text{ of the male max range} \\
 \\
 RAN \text{ of min for female} &= F\text{-MAXmin} - F\text{-MINmin} \\
 &= 251.8 - 18.3 = 233.5 \\
 EM76 \text{ min } (\%) &= ((176 - 18.3)/233.5) * 100 \\
 &= 68\% \text{ of the female min range}
 \end{aligned}$$

The maximum F_0 value of sentence RR22 represents 9% of the possible range of male max values. The minimum F_0 value of sentence EM76 represents 68% of the possible range of female min values.

GLOBAL LEVEL: The F_0 values are also scaled globally relative to the absolute extremes of each subcategory (max, min, ran) across all subjects: A-MAXmax and A-MINmax for max values, A-MAXmin and A-MINmin for min values, A-MAXran and A-MINran for ran values. Similarly to gender

subcategories, the ranges of absolute max, min and ran are the difference between the largest and smallest values of each subcategory.

$$\begin{aligned}
 RAN \text{ of max for both genders} &= A\text{-MAXmax} - A\text{-MINmax} \\
 &= 599.8 - 109 = 490.8 \\
 RR22 \text{ max (\%)} &= ((154 - 109)/490.8) * 100 \\
 &= 9\% \text{ of the overall max range}
 \end{aligned}$$

$$\begin{aligned}
 RAN \text{ of min for both genders} &= A\text{-MAXmin} - A\text{-MINmin} \\
 &= 251.8 - 15.3 = 236.5 \\
 EM76 \text{ min (\%)} &= ((176 - 15.3)/236.5) * 100 \\
 &= 68\% \text{ of the overall min range}
 \end{aligned}$$

Fuzzification When all max, min, and ran values of all sentences of all subjects have been scaled both by gender and overall, nine sets have been created: a set of max values, a set of min values, and a set of ran values for each of the three groups (males, females, all). The AFC fuzzifies and defuzzifies these sets according to the general procedure. The results are in Table 13.4: as percentages of a range they are rather abstract and uninformative but they will be necessary in the re-scaling procedure.

13.2 Re-scaling

Each type of value has been scaled and graded by degree of typicality by the AFC. Next, the AFC calculates the actual dimensions (in cs and Hz) of the

$m_{(x)}$	MALE			FEMALE			BOTH			
	max	min	ran	max	min	ran	max	min	ran	
Fuzzification	1	38	49	62	48	22	20	25	28	28
	0.9	27	49	57	76	21	19	28	34	22
	0.8	28	44	60	76	13	21	40	33	20
	0.7	22	43	62	51	10	21	32	35	17
	0.6	21	45	69	61	21	31	43	66	24
	0.5	21	70	60	43	49	40	38	68	27
	0.4	36	77	69	45	55	54	71	65	34
	0.3	62	13	29	32	63	79	70	48	57
	0.2	66	43	25	41	58	71	58	35	70
	0.1	66	66	44	41	57	56	60	57	67
\bar{m}		32	49	58	56	28	31	39	43	28

Table 13.4: Grades of membership of scaled F_0 values by subject groups (male, female, global) and subcategories (max, min, ran)

ranked scaled values for the dimension of the frame of reference: the sentence duration, syllables duration and F_0 values for each grade of membership.

13.2.1 Sentence

The equation used to calculate the relative duration of sentences for the creation of the fuzzy set is applied in reverse to calculate durations in centiseconds at each grade of membership. For example, at the maximum grade of membership ($m_{(duration)} = 1$), the duration of the sentence spans 34% of the differential range. The proportion of the range (in cs) that this value represents is calculated and the minimum duration value (in ms) is added to

it:

$$\begin{aligned}
 \text{actual duration (cs)} &= (x * RAN/100) + MIN \\
 (m_{(duration)} = 1) &= (34 * 126/100) + 63 \\
 &= 106\text{cs}
 \end{aligned}$$

The highest-ranking duration of closed questions ($m_{(duration)} = 1$) spans 34% of the range or 106 cs. A third line has been added to the results of the fuzzification of sentence duration. It contains the calculated durations in centiseconds. Note that, as the sentence duration approaches the extremities of the membership scale, the sentence duration approaches the extremities of the actual range of the set (63 cs - 189 cs):

$m_{(x)}$	0.5	0.6	0.7	0.8	0.9	1	0.9	0.8	0.7	0.6	0.5	0.4
%	1	7	14	20	27	34	42	48	55	61	68	75
ms	64	72	80	88	97	106	116	124	133	140	149	158
Defuzz												
								0.3	0.2	0.1		
								82	88	96		39
								166	176	184		113

The global value obtained by defuzzification of all grades of membership is 39% of the range or $\bar{m} = 113$ cs. The range of sentence duration that is categorically sound – i.e. with a grade of membership equal or over $m_{(x)} = 0.5$ – is 64 cs to 149 cs. Under or over this range, the sentence can be misclassified. A count of occurrences by grade of membership reveals that sentences with a duration between 97 cs and 106 cs and corresponding to a grade of membership between $m_{(x)} = 0.9$ and $m_{(x)} = 1$ account for over 46% of the sentences. Only 3% of the sentences have a duration with a grade of

membership under $m_{(x)} = 0.5$, that is, under the category threshold. All other sentences are between grade $m_{(x)} = 0.5$ and $m_{(x)} = 0.8$

m(duration)	range (cs)	portion of the set
>0.9	$97 < x < 116$	46%
<0.9 and ≥ 0.5	$64 < x < 97$	51%
	$116 < x < 149$	
<0.5	$x > 149$	3%

The results obtained for the 10 grades of membership of sentence duration serve as the reference for the calculation of syllable duration.

13.2.2 Syllables

The equation used to calculate the relative duration of syllables for the creation of the fuzzy set is applied in reverse to calculate the duration of syllables in centiseconds at each grade of membership. The point of reference for each grade of membership is the sentence duration result found at the previous step.

The value for grade $m_{(x)} = 1$ of sentence duration is 106 cs. The duration in centiseconds of each syllable corresponds to the proportion of the total duration that its percentage value represents. Note that numbers have been rounded to the nearest integer. The value 10 is in fact 9.52, hence the result:

$$\begin{aligned}
 \text{actual duration (ms)} &= (x * \text{Sentence}/100) \\
 (m_{(\text{syllable 1})} = 1) &= (10 * 106/100) \\
 &= 10 \text{ cs}
 \end{aligned}$$

The relative duration of each syllable at grade $m_{(x)} = 1$, as calculated

$m_{(x)}$	σ 1	σ 2	σ 3	σ 4	σ 5	σ 6	σ 7	Total	
Fuzzification	0.5	8	5	8	9	9	12	13	64
	0.6	8	6	9	10	10	14	15	72
	0.7	8	9	10	9	10	14	20	80
	0.8	9	11	11	10	11	14	23	88
	0.9	9	11	12	11	12	15	26	97
	1.0	10	13	13	12	13	16	28	106
	0.9	11	13	15	14	15	18	31	116
	0.8	12	15	15	14	15	20	33	124
	0.7	13	15	17	15	17	23	33	133
	0.6	16	12	18	19	20	27	29	140
	0.5	19	11	19	21	22	27	30	149
	0.4	19	18	21	21	22	27	30	158
	0.3	22	21	23	24	25	31	21	166
	0.2	25	22	25	25	28	33	18	174
	0.1	26	23	26	27	30	34	18	184
\bar{m}		13	12	15	14	15	19	24	113

Table 13.5: Re-scaling, by grade of membership, of the relative duration of syllables to duration in cs, proportionally to the duration of the sentence

in section 13.1.2, is copied on the upper line below. Under the value corresponding to the relative size of each syllable (in %), the actual part of the total duration (106 cs) represented by each of these percentages has been calculated:

%	10	12	13	12	13	15	27	100
cs	10	13	13	12	13	16	28	106

The complete results are in Table 13.5 and will be used in the next sections to calculate the position of the pre-tones and the tones on the syllabic grid.

13.2.3 Fundamental frequency (F_0)

The equation used to calculate the relative F_0 values of sentences for the creation of the fuzzy set is applied in reverse to calculate the F_0 values (max, min, ran) of the contour in Hz at each grade of membership. For example, at the maximum grade of membership ($m_{(A-MAXmax)} = 1$), the maximum of the sentence represents 40% of the observed range (A-MAXmax = 599.8, A-MINmax = 109, RAN = 599.8 - 109 = 490.8). The part of the range (in Hz) that this value represents is calculated and the minimum duration value (in Hz) is added to it:

$$\begin{aligned} \text{actual } F_0 \text{ (cs)} &= (x * RAN/100) + MIN \\ (m_{(F_0)}) = 1 &= (38 * 490.8/100) + 109 \\ &= 294 \text{ Hz} \end{aligned}$$

The extreme values of the corpus have been copied below for reference in the upper part of Table 13.20. In the lower part of the table, the relative results for the highest grade of membership ($m_{(x)} = 1$, upper row) are re-scaled to values proportional to the reference values (lower row) from the upper table.

The results of the re-scaling for all grades of membership and for the defuzzification are in Table 13.7.

	F ₀ MALE			F ₀ FEMALE			F ₀ BOTH		
	max	min	ran	max	min	ran	max	min	ran
MAX	599.8	151.2	542.3	599.5	251.8	575.8	599.8	251.8	575.8
MIN	109	15.3	17.4	123.7	18.3	38.8	109.0	15.3	575.8
RAN	490.8	136.0	524.9	475.8	233.5	537.0	490.8	236.5	558.4

$m_{(x)}=1$									
(%)	38	49	62	48	22	20	25	28	28
Hz	294	99	133	355	130	146	232	81	172

Table 13.6: F₀ extreme values (top) and re-scaling of the F₀ values for grade $m_{(x)} = 1$

	MALE			FEMALE			BOTH			
	max	min	ran	max	min	ran	max	min	ran	
Fuzzification	1	294	99	133	355	130	146	232	81	172
	0.9	242	93	130	355	196	138	244	96	138
	0.8	246	97	87	333	195	151	305	92	131
	0.7	215	99	68	328	138	151	264	97	112
	0.6	212	109	126	340	160	205	319	170	151
	0.5	212	97	273	456	118	254	294	175	168
	0.4	283	109	307	489	124	329	456	170	209
	0.3	415	55	347	185	94	463	455	129	333
	0.2	431	49	322	326	115	419	394	98	407
	0.1	431	75	315	436	115	337	404	151	391
	\bar{m}	265	95	164	357	150	206	306	116	174

Table 13.7: Grades of membership of actual F₀ values (Hz) by subject groups and dimension categories.

13.3 Re-scaling of the surprise contour - Results

13.3.1 Surprise: scaling, fuzzification, defuzzification

13.3.1.1 Sentence

13.3.1.2 Syllables

$m_{(x)}$	0.4	0.5	0.6	0.7	0.8	0.9	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
%	2	4	7	10	15	19	23	23	29	32	38	44	52	63	82	100

Defuzzification value: $\bar{m} = 26\%$

Table 13.8: Relative differential range of sentence duration by grade of membership

$m_{(x)}$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total	
Fuzzification	1	8	12	13	11	12	14	27	95
	0.9	8	10	13	13	10	14	27	94
	0.8	10	10	10	12	12	12	27	92
	0.7	6	11	15	8	16	14	28	97
	0.6	8	15	13	15	12	18	29	110
	0.5	11	12	13	12	19	7	32	105
	0.4	13	12	13	9	14	21	34	114
	0.3	17	18	17	11	13	26	26	128
	0.2	21	19	13	20	24	29	23	149
	0.1	24	23	23	25	28	36	18	177
\bar{m}	10	12	13	12	14	15	28	104	

Table 13.9: Fuzzification and defuzzification of the duration of syllables
(all durations expressed in %)

13.3.1.3 Fundamental frequency (F_0)

GENDER LEVEL:

Gender	MAXmax	MINmax	MAXmin	MINmin	MAXran	MINmax
Female	599.8	253.9	289.2	17.1	528.5	108.9
Male	598.5	131.1	178.7	16.6	507.6	44.4

GLOBAL LEVEL:

$m_{(x)}$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total	
Fuzzification	1	8	12	13	11	13	14	28	100
	0.9	9	11	13	13	11	15	29	100
	0.8	11	11	11	12	12	13	29	100
	0.7	6	11	16	8	16	14	28	100
	0.6	7	14	11	14	11	16	27	100
	0.5	11	11	12	11	18	7	31	100
	0.4	11	10	11	7	13	18	30	100
	0.3	13	14	14	9	10	20	20	100
	0.2	14	13	8	13	16	19	16	100
	0.1	14	13	13	14	16	20	10	100
\bar{m}	10	12	12	11	13	15	27	100	

Table 13.10: Adjustment of fuzzy duration to 100 (all durations expressed as a percentage of the duration of the sentence)

G-max		G-min		G-ran	
MAX	MIN	MAX	MIN	MAX	MIN
599.8	131.1	289.2	16.6	528.5	44.4

13.3.2 Surprise: re-scaling

13.3.2.1 Sentence

$m_{(x)}$	0.4	0.5	0.6	0.7	0.8	0.9	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
%	2	4	7	10	15	19	23	23	29	32	38	44	52	63	82	100
ms	93	98	102	107	115	<u>122</u>	<u>127</u>	129	138	144	153	162	176	194	224	254

Overall, the surprise modality is realized with a duration of approximately 1.3 seconds; the global value obtained by defuzzification of all grades of membership is 26% of the differential range for a sentence duration of 133 cs. A count of occurrences (results below) shows that there is a lot of variation in the duration of sentences along the grade continuum, particularly over 0.5. Sen-

$m_{(x)}$	MALE			FEMALE			BOTH			
	max	min	ran	max	min	ran	max	min	ran	
Fuzzification	1	26	49	15	43	43	47	62	48	37
	0.9	33	46	40	68	41	53	66	34	41
	0.8	39	51	34	59	45	50	64	35	47
	0.7	32	40	19	43	55	45	48	35	45
	0.6	50	64	44	59	50	37	52	49	50
	0.5	66	37	61	38	55	67	39	52	77
	0.4	78	45	77	42	58	38	42	56	66
	0.3	47	60	53	50	40	41	45	78	67
	0.2	53	50	40	60	50	51	78	78	78
	0.1	-	53	45	55	46	59	78	78	78
\bar{m}		42	49	38	50	47	48	56	47	51

Table 13.11: Grades of membership of scaled F_0 values by subject groups (male, female, global) and subcategories (max, min, ran)

tences with a duration between 122 cs and 129 cs, corresponding to a grade of membership between 0.9 and 1, account for 16% of the sentences. Sentences with a duration whose grade of membership is equal to or above 0.5 and below 0.9 account for 75% of all sentences. Only 9% of the sentences have a duration with a grade of membership below 0.5, that is, under the category threshold.

m_(duration)	range (cs)	portion of the set
>0.9	$122 < x < 129$	16%
<0.9 and ≥ 0.5	$98 < x < 122$	75%
	$129 < x < 62$	
<0.5	$x < 98 \& x > 162$	9%

13.3.2.2 Syllables

$m_{(x)}$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total	
Fuzzification	0.4	10	9	10	7	12	17	28	93
	0.5	10	11	12	11	17	7	30	98
	0.6	7	14	12	14	11	17	27	102
	0.7	7	12	17	9	17	15	31	107
	0.8	13	13	13	14	14	15	34	115
	0.9	10	13	16	16	13	18	35	122
	1	11	15	17	14	16	18	36	127
	0.9	11	14	17	17	14	19	37	129
	0.8	15	15	15	17	17	17	41	138
	0.7	9	16	22	12	23	20	41	144
	0.6	11	21	17	21	17	25	41	153
	0.5	17	18	19	18	29	11	50	162
	0.4	19	18	19	13	22	32	53	176
	0.3	26	27	26	17	20	39	39	194
	0.2	31	29	19	30	36	44	35	224
	0.1	35	33	33	35	41	52	26	254
\bar{m}	13	16	17	15	17	20	36	133	

Table 13.12: re-scaling, by grade of membership, of the relative duration of syllables to duration in cs, proportionally to the duration of the sentence

13.3.2.3 Fundamental frequency (F_0)

	F ₀ MALE			F ₀ FEMALE			F ₀ BOTH		
	max	min	ran	max	min	ran	max	min	ran
MAX	598.5	178.7	507.6	599.8	289.2	528.5	599.8	289.2	528.5
MIN	131.1	16.6	44.4	253.9	17.1	108.9	131.1	16.6	44.4
RAN	467.4	162.1	463.2	345.9	272.1	419.6	468.7	272.6	484.1

$m_{(x)}=1$									
(%)	26	49	15	43	43	47	62	48	37
Hz	252	96	114	403	133	307	422	148	224

Table 13.13: F₀ extreme values (top) and re-scaling of the F₀ values for grade $m_{(x)} = 1$

Fuzzification	MALE			FEMALE			BOTH		
	max	min	ran	max	min	ran	max	min	ran
1	252	96	114	403	133	307	422	148	224
0.9	283	92	230	487	127	330	440	110	245
0.8	313	99	200	459	138	320	432	112	274
0.7	282	82	134	402	167	298	355	113	264
0.6	364	120	250	459	154	263	374	150	288
0.5	437	77	325	386	168	391	312	159	419
0.4	494	89	402	312	176	267	327	170	364
0.3	352	114	291	426	125	281	342	228	367
0.2	376	97	227	462	153	324	495	228	420
0.1	-	103	251	444	143	355	495	228	420
\bar{m}	328	96	219	427	146	312	395	144	293

Table 13.14: Grades of membership of actual F₀ values (Hz) by subject groups and dimension categories.

13.4 Re-scaling of the doubt contour - Results

13.4.1 Doubt: scaling, fuzzification, defuzzification

13.4.1.1 Sentence

$m_{(x)}$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
%	4	7	11	15	19	22	25	28	29	34	33	41	47	54	63	73	89

Defuzzification value: $\bar{m} = 30\%$

Table 13.15: Relative differential range of sentence duration by grade of membership

13.4.1.2 Syllables

$m_{(x)}$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total	
Fuzzification	1.0	8	11	12	12	12	14	29	97
	0.9	8	10	12	11	12	14	28	95
	0.8	10	11	12	9	12	14	26	93
	0.7	9	11	12	11	15	14	30	101
	0.6	9	14	12	15	14	20	28	111
	0.5	11	14	12	7	11	18	34	106
	0.4	11	12	12	7	13	21	35	109
	0.3	19	14	20	17	13	15	30	127
	0.2	23	19	23	20	20	15	33	153
	0.1	33	23	32	25	24	21	33	191
\bar{m}	10	12	13	12	13	16	30	105	

Table 13.16: Fuzzification and defuzzification of the duration of syllables (all durations expressed in %)

$m_{(x)}$	σ 1	σ 2	σ 3	σ 4	σ 5	σ 6	σ 7	Total
Fuzzification	1.0	8	11	12	12	14	30	100
	0.9	8	11	12	11	13	15	100
	0.8	11	11	12	10	13	15	100
	0.7	8	10	12	11	15	14	100
	0.6	8	13	10	14	12	18	100
	0.5	10	13	11	7	10	17	100
	0.4	10	11	11	7	11	19	100
	0.3	15	11	16	13	10	11	24
	0.2	15	13	15	13	13	10	22
	0.1	17	12	17	13	13	11	17
\bar{m}	10	11	12	11	12	15	28	100

Table 13.17: Adjustment of fuzzy duration to 100 (all durations expressed as a percentage of the duration of the sentence)

13.4.1.3 Fundamental frequency (F_0)

GENDER LEVEL:

Gender	MAXmax	MINmax	MAXmin	MINmin	MAXran	MINmax
Female	599.7	185.6	190.8	15.8	550.1	51.1
Male	599.9	88.6	168.6	17.0	559.2	22.3

GLOBAL LEVEL:

G-max		G-min		G-ran	
MAX	MIN	MAX	MIN	MAX	MIN
599.9	88.6	190.8	15.8	559.2	22.3

$m_{(x)}$	MALE			FEMALE			BOTH			
	max	min	ran	max	min	ran	max	min	ran	
Fuzzification	1	8	50	29	25	41	17	28	31	32
	0.9	14	38	13	41	71	17	29	50	14
	0.8	17	43	13	40	70	19	29	50	34
	0.7	11	30	7	41	69	22	18	43	15
	0.6	14	43	9	31	75	25	19	61	10
	0.5	23	36	18	46	80	28	18	57	13
	0.4	50	52	19	42	66	29	39	44	10
	0.3	63	68	65	42	31	52	71	19	21
	0.2	56	52	52	47	54	67	62	45	27
	0.1	60	60	60	73	52	60	64	53	70
\bar{m}		21	44	21	38	63	25	30	45	21

Table 13.18: Grades of membership of scaled F_0 values by subject groups (male, female, global) and subcategories (max, min, ran)

13.4.2 Doubt: re-scaling

13.4.2.1 Sentence

$m_{(x)}$	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
%	4	7	11	15	19	22	25	28	29	34	33	41	47	54	63	73	89
ms	90	94	100	106	112	116	120	125	126	133	132	144	152	162	176	190	213

Overall, the doubt modality is realized with a duration of approximately 1.25 seconds; the global value obtained by defuzzification of all grades of membership is 30% of the differential range or a sentence duration of 127 cs. A count of occurrences (results below) shows that there is a lot of variation in the duration of sentences along the graded continuum, particularly over 0.5. Sentences with a duration between 120 cs and 126 cs, corresponding to a grade of membership between 0.9 and 1, account for 14% of the sentences. Sentences with a duration whose grade of membership is equal to or above 0.5 and below 0.9

account for 71% of all sentences. Only 15% of the sentences have a duration with a grade of membership below 0.5, that is, under the category threshold.

m(duration)	range (cs)	proportion of the set
>0.9	$120 < x < 126$	14%
<0.9 and ≥ 0.5	$100 < x < 120$	71%
	$126 < x < 152$	
<0.5	$x < 100 \ \& \ x > 152$	15%

13.4.2.2 Syllables

$m_{(x)}$	$\sigma 1$	$\sigma 2$	$\sigma 3$	$\sigma 4$	$\sigma 5$	$\sigma 6$	$\sigma 7$	Total	
Fuzzification	0.3	14	10	14	12	9	10	21	90
	0.4	9	10	10	6	11	18	30	94
	0.5	10	13	11	7	10	17	32	100
	0.6	8	13	11	14	13	20	27	106
	0.7	9	12	13	12	17	16	33	112
	0.8	12	13	14	11	15	17	33	116
	0.9	10	13	15	14	15	18	36	120
	1	10	14	15	15	15	17	37	125
	0.9	11	13	15	14	16	19	37	126
	0.8	14	15	16	13	17	20	38	133
	0.7	11	14	15	14	20	18	39	132
	0.6	11	18	15	19	18	27	36	144
	0.5	15	20	16	10	15	26	49	152
	0.4	16	17	17	11	19	31	52	162
	0.3	27	20	28	23	17	20	41	176
	0.2	28	24	29	24	25	19	41	190
	0.1	37	26	36	28	27	23	37	213
\bar{m}	13	15	15	14	16	19	36	127	

Table 13.19: Conversion, by grade of membership, of the relative duration of syllables to duration in cs, proportionally to the duration of the sentence

13.4.2.3 Fundamental frequency (F_0)

	F ₀ MALE			F ₀ FEMALE			F ₀ BOTH		
	max	min	ran	max	min	ran	max	min	ran
MAX	599.9	186.6	559.2	599.7	190.8	550.1	599.9	190.8	559.2
MIN	88.6	17	22.3	185.6	15.8	51.1	88.6	15.8	22.3
RAN	511.3	169.6	536.9	414.1	175	499	511.3	175	536.9

	$m_{(x)}=1$								
(%)	8	50	29	25	41	17	28	31	32
Hz	127	93	180	290	87	136	232	69	192

Table 13.20: F₀ extreme values (top) and conversion of the F₀ values for grade $m_{(x)} = 1$

	MALE			FEMALE			BOTH			
	max	min	ran	max	min	ran	max	min	ran	
Fuzzification	1	127	93	180	290	87	136	232	69	192
	0.9	160	75	91	357	140	136	234	103	95
	0.8	174	82	91	352	139	148	238	103	206
	0.7	145	63	60	357	137	161	181	92	104
	0.6	160	83	71	313	146	177	185	123	77
	0.5	209	72	119	376	156	191	180	116	92
	0.4	343	96	125	357	130	198	286	93	78
	0.3	410	121	372	357	69	310	452	50	136
	0.2	376	96	303	379	110	387	403	95	169
	0.1	-	108	343	489	107	352	418	109	400
\bar{m}		198	84	134	344	126	177	243	95	137

Table 13.21: Grades of membership of actual F₀ values (Hz) by subject groups and dimension categories.

Appendix 1 - Elicitation task 1

List of sentences (emphasized in bold face) to be read with a simple closed question contour, preceded by their context.

1. Il y a des gens qui aiment garder des souvenirs après un concert.

⇒ **Tu veux garder les tickets ?**

2. Je vois que vous avez amené votre livre.

⇒ **Vous allez nous en parler ?**

3. Il est déjà 2 heures du matin. C'est un peu tard pour rentrer.

⇒ **Tu veux rester pour la nuit ?**

4. C'est une petite ville ici. Il n'y a pas grand-chose à faire, j'imagine.

⇒ **Vous avez beaucoup d'amis ?**

5. On est sur la route pour venir chez toi. On va passer à l'épicerie.

⇒ **Tu veux qu'on ramène du vin ?**

6. Je ne suis pas sûr des papiers qu'il faut pour aller en Belgique.

⇒ **Il faut avoir un passeport ?**

7. J'ai fait un boeuf bourguignon hier et il m'en reste pas mal. Ça serait bête de le gâcher.

⇒ **Vous voulez manger chez moi ?**

8. On m'a dit que votre fils était allergique à certains aliments.

⇒ **On peut lui donner du lait ?**

9. C'est embêtant cette histoire de divorce. Il faut que tu trouves un logement provisoire en attendant que ton nouvel appartement soit libre.

⇒ **Tu vas rester chez ta soeur ?**

10. Votre ami m'a dit que vous habitez une grande ville du sud ouest.

⇒ **Vous habitez à Bordeaux ?**

11. Je vais me faire rembourser pour les frais de notre voyage.

⇒ **T'as gardé toutes les factures ?**

12. On va passer à Paris quand nous viendrons en France à Noel.

⇒ **Vous voulez qu'on vienne vous voir ?**

13. Je ne retrouve plus mes affaires. Je ne sais plus ce que j'en ai fait.

⇒ **Tu m'as rendu mon stylo ?**

14. Cette élève a beaucoup de problèmes et je l'ai envoyée voir le conseiller d'éducation.

⇒ **Elle a suivi ses conseils ?**

15. Je vois que vous appréciez mon boeuf bourguignon. Il en reste beaucoup.

⇒ **Vous en voulez un peu plus ?**

16. On n'avait pas prévu que tes amis viendraient au pique-nique avec nous.
Il faut qu'on leur fasse des sandwichs maintenant.
⇒ **Il nous reste encore du pain?**
17. Vous avez l'air en forme.
⇒ **Vous faites beaucoup d'exercice ?**
18. On n'est pas d'ici et on cherche à se rendre chez un ami.
⇒ **Vous connaissez bien la ville ?**
19. On doit aller à Lisbonne pour rencontrer des clients. On m'a dit qu'ils ne parlaient pas bien anglais.
⇒ **Tu sais parler portugais ?**
20. Si j'avais su, je me serais préparé. Je ne pensais pas que tu arriverais si tôt.
⇒ **T'es passé par l'autoroute ?**
21. Il est presque 4 heures. Les enfants ne vont pas tarder à rentrer de l'école.
⇒ **T'as préparé leur goûter ?**
22. Ton ami m'a dit que tu donnais un cours de philosophie à la fac de Caen cette année.
⇒ **T'as des étudiants sympa ?**

23. Il y a beaucoup de bruit chez toi à cause des travaux. Ça doit être difficile pour tes recherches.

⇒ **Tu veux travailler chez moi ?**

24. Il est bon votre curry.

⇒ **Vous avez mis du cumin ?**

25. On est au café avec quelques amis. On va y rester un bon moment.

⇒ **Vous allez passer plus tard ?**

26. Marie m'a dit que vous avez amené des draps pour le lit dans la chambre d'amis.

⇒ **Vous avez pris la bonne taille ?**

27. Si tu veux, je peux appeler Pierre, j'ai mon téléphone sur moi.

⇒ **Tu connais son numéro ?**

28. Henri m'a dit que votre mère était en convalescence à l'hôpital Saint-Julien.

⇒ **Nous pouvons passer la voir ?**

29. Il paraît que Bruno veut que tu t'occupes de sa chienne le weekend prochain.

⇒ **Tu vas pouvoir la garder ?**

30. On va voir Antoine dans quelques minutes pour discuter de nos progrès.

⇒ **T'as imprimé ton rapport ?**

31. Quand on part pour longtemps, c'est mieux de ne rien laisser branché.

⇒ **T'as débranché la télé?**

32. La procédure pour obtenir un prêt est assez facile. Il suffit de remplir ce formulaire et d'y joindre quelques justificatifs.

⇒ **Vous avez les documents ?**

33. Je sais qu'Antoine est très occupé en ce moment mais je pense qu'il aimerait discuter de votre projet avec vous.

⇒ **Vous avez pu lui montrer ?**

34. Il commence à se faire tard. On attend Edouard pour aller faire un tour.

⇒ **Il a fini son travail ?**

Appendix 2 - Elicitation task 2

List of sentences (emphasized in bold face) to be read with two contrastive contours (doubt and surprise), preceded by their contexts.

1. Votre ami vous raconte comment il a pu rencontrer Madonna après un concert et comment il l'a embrassée.
⇒ **t'as embrassé Madonna ?!**
2. Lors d'une journée très pluvieuse, un ami vous rend visite. Il arrive trempé et en disant qu'il ne pensait pas qu'un parapluie serait nécessaire.
⇒ **t'as pas pris ton parapluie ?!**
3. Votre ami a grandi dans une ferme où on élève des moutons et des agneaux. Pourtant il vous dit n'avoir jamais mangé d'agneau et vous avez du mal à le croire.
⇒ **t'as jamais mangé d'agneau ?!**
4. Tous les tickets pour le concert de votre artiste préféré ont été vendus et un de vos amis vous apprend qu'il vient d'en obtenir deux gratuitement.
⇒ **t'as eu des tickets gratuits ?!**
5. On vous apprend que votre amie Isabelle part en vacances avec le Club Med alors qu'elle vous a toujours dit qu'elle trouvait les voyages organisés déprimants

⇒ elle va partir au Club med' ?!

6. Vous avez un ami cinéphile qui va au cinéma plusieurs fois par semaine.

Il vous apprend qu'il n'y va plus depuis bientôt un an.

⇒ tu vas plus au cinéma ?!

7. Votre ami vous a répété souvent qu'il aimait les légumes. Pourtant il refuse les brocolis que vous lui servez en disant qu'il n'aime pas beaucoup les légumes.

⇒ tu n'aimes pas les brocolis ?!

8. Vous pensez que votre ami est un socialiste convaincu. Pourtant il vient de vous apprendre qu'il a voté pour Nicolas Sarkozy aux élections présidentielles.

⇒ t'as voté pour Sarkozy ?!

9. Lors d'une conversation, vous entendez que quelqu'un est allé à Juan les Pins par le train. Vous pensez que Juan les Pins est une petite ville et qu'il est improbable qu'il s'y trouve une gare

⇒ y a une gare à Juan les Pins ?!

10. Un de vos amis déteste votre ami Jean-Pierre. Pourtant, il vient de vous dire qu'il a diné avec Jean-Pierre la veille.

⇒ t'as diné avec Jean-Pierre ?!

11. A la fin d'une soirée que vous avez organisée pour des amis on vous apprend que votre meilleure amie est partie avec Pierre alors que vous

pensiez qu'elle ne l'aimait pas beaucoup.

⇒ **elle est partie avec Pierre ?!**

12. Quelqu'un déclare connaître la recette d'un plat que vous appréciez dans un restaurant réputé.

⇒ **vous connaissez cette recette ?!**

13. Une personne de votre entourage ne jure que par les nouvelles technologies. Pourtant elle vous parle d'une émission qu'elle a entendue à la radio. Vous ne pouvez pas croire qu'elle écoute un medium aussi peu moderne.

⇒ **vous écoutez la radio?!**

14. Vous surprenez un de vos collègues en pleine discussion sur les cours boursiers et son portfolio. Vous n'auriez jamais imaginé qu'il fût à ce point intéressé par la bourse.

⇒ **vous suivez les cours boursiers ?!**

15. Deux de vos amis parlent de leur footing matinal alors que vous pensiez qu'ils refusaient catégoriquement toute forme d'exercice physique.

⇒ **vous aimez faire du footing ?!**

16. Vous apprenez que votre collègue Paul, avec qui vous devez terminer un important projet, est parti en congés sans vous avertir.

⇒ **il est parti en vacances ?!**

17. Quelqu'un de votre famille s'est préparé longuement à son permis de

conduire et a même pratiqué avec brio la conduite dans votre propre automobile. Vous avez pensé que l'examen serait un succès évident et pourtant on vous apprend que c'est un échec

⇒ **il a raté son permis?!**

18. On vous raconte qu'un policier porte toujours son arme malgré plusieurs incidents qui ont clairement montré qu'il ne maîtrisait pas ses réactions et qu'il était dangereux pour la société.

⇒ **on lui a laissé son arme ?!**

19. Des amis se sont moqués de vous pour leur avoir proposé d'aller à l'opéra alors qu'ils détestent ça. Vous apprenez maintenant qu'ils veulent vous accompagner.

⇒ **ils veulent venir avec moi ?!**

20. Un ami vous emmène au musée et vous y entrez sans payer. Vous êtes habitué(e) de l'endroit et vous avez toujours payé l'entrée

⇒ **on peut entrer sans payer?!**

21. En voyage avec un ami, vous avez juste le temps de vous rendre à la gare. Pourtant votre ami vous dit qu'il ne faut pas oublier de passer chez sa sœur. Vous ne vous souvenez pas que cela fasse parti du plan.

⇒ **on doit passer chez ta sœur ?!**

22. Vous êtes convaincu que deux de vos meilleurs amis se connaissent. Pourtant lors d'une soirée, l'un d'eux vous dit n'avoir jamais rencontré l'autre.

⇒ **t'as jamais rencontré Paul ?!**

23. Le fils d'un ami a raté plusieurs fois le bac. Le temps passant, vous pensiez qu'à force de persévérance, il avait finalement réussi. Pourtant on vous apprend qu'il se prépare à le passer de nouveau.

⇒ **il a toujours pas son bac ?!**

24. En route pour les vacances avec des amis, le conducteur annonce qu'il va s'arrêter pour faire le plein. Vous avez l'impression que vous venez juste de partir et cela vous semble étrange de consommer autant d'essence.

⇒ **on a déjà plus d'essence ?!**

25. Un collègue raconte à tout le monde une histoire aberrante au sujet de sa rencontre avec Catherine Deneuve. Vous entendez deux stagiaires qui discutent de cette histoire comme si elle était vraie.

⇒ **vous avez cru son histoire ?!**

26. Un ami qui a des problèmes financiers veut vendre la collection d'art de ses parents. On lui dit que c'est une mauvaise idée et qu'il regretterait plus tard d'avoir bradé le trésor de ses parents pour payer ses dettes. Pourtant on vous apprend que la collection est en vente.

⇒ **il va quand même la vendre ?!**

27. C'est le 4 février et on vous dit que c'est l'anniversaire d'une amie. Vous êtes convaincu(e) qu'elle est née le 4 mars et non le 4 février.

⇒ **elle est née en février ?!**

28. On vous apprend qu'un meurtrier particulièrement violent vient d'être

acquitté.

⇒ **on l'a pas mis en prison ?!**

29. Lors d'une conversation, vous apprenez que votre ami Pierre, qui a 28 ans, vient de faire établir son testament. Le connaissant un peu, vous ne voyez pas de bonne raison pour laquelle Pierre voudrait faire un testament si jeune.

⇒ **il a fait un testament ?!**

30. Vous offrez une boîte de 36 macarons à un ami et le lendemain il vous dit les avoir tous mangés. Cela vous paraît impossible sans être dangereux.

⇒ **t'as fini les macarons ?!**

31. Lors d'une étude de linguistique particulièrement pénible, vous aimeriez bien arrêter mais on vous dit que c'est impossible. Pourtant il vous semble que vous pourriez simplement partir.

⇒ **y a pas moyen d'arrêter ?!**

32. Votre ami vient d'adopter un chat et on vous raconte qu'il le promène partout en laisse.

⇒ **il promènd son chat en laisse ?!**

33. Lors d'un tournoi local d'échecs, vous êtes admiratif de la qualité des joueurs de l'une des équipes. On vous apprend que c'est votre ami Sébastien qui en est le chef. Vous ne pensiez pas qu'il fût si doué pour les échecs.

⇒ **c'est l'équipe à Sébastien ?!**

34. Lors d'une soirée, vous entendez quelqu'un affirmer que Charles de Gaulle était breton
⇒ **Charles de Gaulle était breton ?!**

Bibliography

- Abramson, A. S. & L. Lisker. 1970. Discriminability along the voicing continuum: cross language tests. *In Proc. Int. Congr. Phon. Sci., 6th. Prague* 569–573.
- Alessandro, C. d' & P. Mertens. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language* 9(3). 257–288.
- Allen, J.S., J.L. Miller & M. DeSteno. 2003. Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America* 113(1). 544–552.
- Arvaniti, A. & R. Ladd. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26. 43–74.
- Aslin, R.N., D.B. Pisoni, B.L. Hennessy & A.J. Perey. 1981. Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development* 52. 1135–1145.
- Barth-Weingarten, D. 2011. *The fuzziness of intonation units: Some theoretical considerations and a practical solution* (*url: <http://www.inlist.uni-bayreuth.de/issues/51/inlist51.pdf>*) Interaction and Linguistic Structures, vol. 51. InLiSt.
- Beckman, M. 1986. *Stress and non-stress accent*. Dordrecht: Foris.
- Beckman, M., J. Hirschberg & S. Shattuck-Hufnagel. 2005. The original ToBi system and the evolution of the ToBi framework. In S.-A. Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 9–54. Oxford University Press.

- Beckman, M. & J. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 15–70.
- Beyssade, C., E. Delais-Roussarie & J.M. Marandin. 2007. The prosody of interrogatives in French. *Nouveaux cahiers de linguistique française* 28. 163–175.
- Black, M. 1937. Vagueness. An exercise in logical analysis. *Philosophy of Science* 4(4). 427–455.
- Boersma, P. & D. Weenink. 2012. *Praat: doing phonetics by computer (version 5.3 to 5.3.24) [computer program]*. retrieved September 2011-september 2012 from <http://www.praat.org/>.
- Bolinger, D. 1961. *Generality, gradience and the all-or-none*. The Hague: Mouton.
- Bolinger, D. 1978. Intonation across languages. In J. Greenberg (ed.), *Universals of human language, volume 2 (phonology)*, 471–524. Stanford: Stanford University Press.
- Bruce, G. 1977. *Swedish word accents in sentence perspective*: Lunds Universitet dissertation.
- Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. 2007. *Frequency of use and the organization of language*. Oxford University Press.
- Bybee, J. 2010. *Language, usage and cognition*. Cambridge University Press.
- Caelen-Haunont, G. 2008. Labelling and structuring F0 prominences using an automatic segmentation and annotation tool (MELISM), and statistical results. In *Proceedings of O-COCOSDA*, 7–12. Hanoï, Vietnam.
- Cutler, A. 1974. On saying what you mean without meaning what you say. *Chicago Linguistic Society* 10. 117–127.

- Dainora, A. 2001. *An empirically based probabilistic model of intonation in American English*: University of Chicago dissertation.
- Dainora, A. 2002. Does intonational meaning come from tones or tunes? Evidence against a compositional approach. In *Proceedings of Speech Prosody*, 47–57. Aix-en-Provence, France.
- Delais-Roussarie, E. 2005. *Phonologie et grammaire : Études et modélisation des interfaces prosodiques*: Université de Toulouse-le-Mirail dissertation.
- Delais-Roussarie, E & B. Post. 2008. Unités prosodiques et grammaire de l'intonation : vers une nouvelle approche. In *Actes des XXVII^e Journées d'Études sur la Parole (JEPTALN 2008), Avignon* .
- Delattre, P. 1966. Les dix intonations du français. *The French Review* 40. 1–14.
- Di Cristo, A. 1999. Vers une modélisation de l'accentuation du français: première partie. *French Language Studies* 143–179.
- Di Cristo, A. 2000. Vers une modélisation de l'accentuation du français: deuxième partie. *French Language Studies* 10. 27–44.
- Di Cristo, A., P. Di Cristo, E Campione & J. Veronis. 2000. A prosodic model for text-to-speech synthesis. In A. Botinis, G. Kouroupetroglou & G. Carayannis (eds.), *Intonation: Models, analysis and applications*, 321–355. Athens: European Speech Communication Association.
- D'Imperio, M., R. Espesser, H. Lvenbruck, C. Menezes, N. Nguyen & P. Welby. 2007. Are tones aligned with articulatory events? Evidence from Italian and French. In J. Cole & J.I. Hualde (eds.), *Papers in laboratory phonology 9*, 577–608. Berlin: Mouton de Gruyter.
- Dubois, D. & H. Prade. 2001. Fuzzy logic fundamentals. In D. Dubois & H. Prade (eds.), *Fundamentals of fuzzy sets*, Boston: Kluwer Academic.

- Duda, R.O. 1973. *Pattern classification and scene analysis*. New-York: John Wiley & Sons, Inc.
- Duda, R.O. 2001. *Pattern classification*. New-York: John Wiley & Sons, Inc.
- Earle, M. A. 1975. An acoustic phonetic study of Northern Vietnamese tones. *Speech Communication Research Laboratory*.
- Eimas, P.D., E.R. Siqueland, P. Jusczyk & J. Vigorito. 1971. Speech perception in infants. *Science* 171. 303–306.
- Fass, D. & J. Feldman. 2002. Categorization under complexity: a unified MDL account of human learning of regular and irregular categories. *Advances in Neural Information Processing Systems*.
- Faure, G. 1973. La description phonologique des systèmes prosodiques. In A. Grundstrom & P.R. Léon (eds.), *Studia phonetica* 8, 1–18. Paris: Didier.
- Feldman, J. 1997. The structure of perceptual categories. *Journal of Mathematical Psychology* 41. 145–170.
- Feldman, J. 2003. The simplicity principle in human concept learning. *Current Directions in Psychological Science* 12. 227–232.
- Feldman, J. 2004. How surprising is a simple pattern? quantifying “eureka!”. *Cognition* 93. 199–224.
- Fink, G.A. 2008. *Markov models for pattern recognition*. Berlin: Springer.
- Fónagy, I. 1979. L’accent français : accent probabilitaire. *Studia Phonetica* 15. 123133.
- Fónagy, I. & E. Bérard. 1973. Questions totales simples et implicatives en français parisien. In A. Grundstrom & P.R. Léon (eds.), *Studia phonetica* 8, 53–98. Paris: Didier.
- Fougeron, C & S.-A. Jun. 1998. Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics* 26. 45–69.

- Friedman, M. & A. Kandel. 1999. *Introduction to pattern recognition. Statistical, structural, neural and fuzzy logic approaches*. London: World Scientific.
- Fu, K.S. (ed.). 1976. *Digital pattern recognition*. Berlin: Springer-Verlag.
- Fujisaki, H. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In P.F. MacNeilage (ed.), *The production of speech*, 39–55. Springer-Verlag.
- Fujisaki, H. 2003. Prosody, information, and modeling with emphasis on tonal features of speech. In B. Bel & I Marlien (eds.), *Proceedings of speech prosody 2004, 23-26 March*, Nara.
- Fujisaki, H., C. Wang, S. Ohno & W. Gu. 2005. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command-response model. *Speech Communication* 47. 59–70.
- Goldinger, S.D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology* 22 (5). 1166–1183.
- Goldstone, R.L., M. Steyvers & B.J. Rogosky. 2003. Conceptual interrelatedness and caricatures. *Memory & Cognition* 31(2). 169–180.
- Gordon-Salant, S., G. Yeni-Komshian & P. Fitzgibbons. 2008. The role of temporal cues in word identification by younger and older adults: Effects of sentence context. *Journal of the Acoustical Society of America* 124(5). 3249–3260.
- Gries, S.T. 2010. *Statistics for linguistics with r: A practical introduction* Mouton Textbook Series. De Gruyter.
- Gussenhoven, C. (ed.). 2004. *The phonology of tone and intonation*. Cambridge, UK ; New York: Cambridge University Press.
- Hayes, B. 2008. *Introductory phonology* Blackwell Textbooks in Linguistics. Wiley.

- Hillenbrand, J., L. Getty, M. Clark & K. Wheeler. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97(5). 3099–3111.
- Hirst, D. 2005. Form and function in the representation of speech prosody. *Speech Communication* 46. 334–347.
- Hirst, D. & al. 2007. On the probabilistic modeling of the form-function articulation for prosodic phenomena. *Mathématiques et Sciences Humaines (Mathematics and Social Sciences)* 180(4). 113–126.
- Hirst, D. & A. Di Cristo (eds.). 1998. *Intonation systems : a survey of twenty languages*. Cambridge, UK: Cambridge University Press.
- Hirst, D. & R. Espresser. 1993. Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l’Institut phonétique d’Aix* 15. 71–85.
- Hualde, J.I. 2003. El modelo métrico y autosegmental. In P. Prieto (ed.), *Las teorías lingüísticas de la entonación*, 155–184. Barcelona: Ariel.
- Jones, D. 1964. *An outline of English phonetics, 9th edition*. Cambridge: Heffer.
- Jun, S.-A. & C. Fougeron. 2000. A phonological model of French intonation. In A Botinis (ed.), *Intonation: Analysis, modeling and technology*, 200–242. Dordrecht : Kluwer Academic Publishers.
- Jun, S.-A. & C. Fougeron. 2002. Realizations of accentual phrase in French intonation. *Probus* 14. 147–172.
- Kaminskaia, S. 2009. *La variation intonative dialectale en français. Une approche phonologique (LINCOLM Studies in French Linguistics 07)*. Lin-colm.
- Khatchadourian, H. 1962. Vagueness. *The Philosophical Quarterly* 12. 38–152.
- Khul, P.K. 1991. Human adults and human infants show a “perceptual magnet

- effec” for the prototypes of speech categories, monkeys do not. *Perception Psychophysiology* 50. 93–107.
- Khul, P.K. & J.D. Miller. 1978. Speech perception by the chinchilla: identification functions for synthetic vot stimuli. *Journal of the Acoustical Society of America* 63. 905–917.
- Klir, G. & B. Yuan. 1995. *Fuzzy sets and fuzzy logic. Theory and applications*. Upper Saddle River, NJ, USA: Prentice Hall PTR.
- Kluender, K.R., R.L. Diehl & P.R. Killeen. 1987. Japanese quail can learn phonetic categories. *Science* 237. 1195–1197.
- Kochanski, G & C Shih. 2003. Prosody modeling with soft templates. *Speech Communication* 39. 311–352.
- Kohler, K.J. 2004. Prosody revisited. Function, time, and the listener in intonational phonology. In B. Bel & I Marlien (eds.), *Proceedings of speech prosody 2004, 23-26 March*, Nara.
- Korzybski, A. 1941. *Science and sanity : an introduction to non-Aristotelian systems and general semantics*. New York: Lancaster, Pa. : International Non-Aristotelian Library Pub. Co; Science Press Printing Co.
- Kruschke, J. 1992. Alcove: An exemplar-based connectionist model of category learning. *Psychological Review* 99. 22–44.
- Ladd, R. 1996. *Intonational phonology. First edition*. Cambridge, UK: Cambridge University Press.
- Ladd, R. 2008. *Intonational phonology. Second (revised) edition*. Cambridge, UK: Cambridge University Press.
- Liao, T.W., Aivars K. Celmins & R.J. Hammell. 2003. A fuzzy c-means variant for the generation of fuzzy term sets. *Fuzzy Sets and Systems* 135. 241–257.
- Liberman, M. & J. Pierrehumbert. 1984. Intonational invariance under changes

- in pitch range and length. In A Aronoff & R.T. Oehrle (eds.), *Language sound structure*, 157–233. Cambridge: MIT Press.
- Liberman, M., J.M. Schultz, S. Hong & V. Okeke. 1993. The phonetic interpretation of tone in Igbo. *Phonetica* 50. 147–160.
- Liberman, M. Y. 1975. *The intonational system of English*: MIT dissertation.
- Lindsay, A. D. 1976. *The Republic by Palto*. London: Everyman's university library.
- Lisker, L. & A.S. Abramson. 1964. A crosslanguage study of voicing in initial stops: acoustical measurements. *Word* 20. 384–422.
- Lisker, L. & A.S. Abramson. 1970. The voicing dimension: some experiments in comparative phonetics. *In Proc. Int. Congr. Phon. Sci., 6th. Prague* 563–567.
- Lively, S.E. & D.B. Pisoni. 1997. On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology* 23. 1665–1679.
- Logan, J., S. Lively & D. Pisoni. 1991. Training japanese listeners to identify english /r/ and /l/: a first report. *Journal of the Acoustical Society of America* 89 (2). 874–886.
- Maddieson, I. 2006. In search of universals. In *Linguistic universals*, 80–100. Cambridge.
- Marandin, J.M. 2006. Contours as constructions. In D. Schoenfeld (ed.), *Constructions all over; case studies and theoretical implications*, <http://www.constructions-online.de/articles/specvol1/>.
- Martin, P. 2000. Intonation and syntax: Another point of view. <http://www.univie.ac.at/wissenschaftstheorie/srb/cyber/Martin1.html>.
- Martinet, A. 1960. *Eléments de linguistique générale*. Paris: Armand Colin.

- Massaro, D.W. 1989. Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology* 21. 398–421.
- Mertens, P. 2004. The Prosogram : Semi-automatic transcription of prosody based on a tonal perception model. In B. Bel & I Marlien (eds.), *Proceedings of speech prosody 2004, 23-26 March*, Nara.
- Mertens, P. 2012. Transcription of tonal aspects in speech and a system for automatic tonal annotation. In *Advancing Prosodic Transcription Workshop at Laboratory Phonology 2012, July 30, 2012*, Stuttgart.
- Miller, J.L. 1994. On the internal structure of phonetic categories: a progress report. *Cognition* 50. 271–285.
- Montreuil, J.P. & N. Bacuez. 2012. Deux approches d'un contour dialectal. In *Colloque du Réseau Français de Phonologie (rfp). June 26-27, 2012*, Paris.
- Nadler, M. & E.P. Smith. 1993. *Pattern recognition engineering*. New-york: John Wiley and Sons Inc.
- Nelson, C.A. 2006. *Neuroscience of cognitive development*. Hoboken: John Wiley and Sons Inc.
- Nosofsky, R.M. 1988a. Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14. 700–708.
- Nosofsky, R.M. 1988b. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14. 54–65.
- Nygaard, L. & D.B. Pisoni. 1998. Talker-specific learning in speech perception. *Perception and Psychophysics* 60 (3). 355–376.
- Olden, G.C. & D.W. Massaro. 1978. Integration of featural information in speech perception. *Psychological Review* 8(5). 172–191.
- Pal, S.K. & P. Mitra. 2004. *Pattern recognition algorithms for data mining*

- : *Scalability, knowledge discovery and soft granular computing*. Chapman and Hall/CRC.
- Passy, P. 1887. *Les sons du français*. Paris: Firmin Didot.
- Passy, P. 1907. *The sounds of the French language*. Oxford: Clarendon Press.
- Peterson, G. E. & H.L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24. 175–184.
- Pierrehumbert, J. 1980. *The phonology and phonetics of english intonation*: MIT dissertation.
- Pierrehumbert, J. 2001. Stochastic phonology. *GLOT* 5:6. 1–13.
- Pierrehumbert, J. 2003. Exemplar theory. Probability theory in linguistics. In *Proceedings of the LSA conference 2003*, LSA.
- Pierrehumbert, J. & M.E. Beckman. 1988. *Japanese tone structure*. Cambridge, Mass.: MIT Press.
- Pike, K. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan publications. Linguistics 1.
- Pisoni, D.B. 1977. Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America* 61. 1352–1361.
- Posner, M. I. & S. W. Keele. 1968a. On the genesis of abstract ideas. *Journal of Experimental Psychology* 77(3). 353–363.
- Posner, M. I. & S. W. Keele. 1968b. On the genesis of abstract ideas. *Retention of abstract ideas* 83(2). 304–308.
- Post, B. & al. 2006. Développer un système de transcription des phénomènes prosodiques. *Bulletin PFC* .
- Post, B. & E. Delais-Roussarie. 2006. Vers un système multilinéaire de transcription des variations intonatives. In *Actes des XXVIèmes Journées d'étude sur la parole*, JEP.

- Prom-on, S., Y. Xu & B. Thipakorn. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America* 125. 405–424.
- Rosch, E. 1973. Natural categories. *Cognitive Psychology* 4(3). 328–350.
- Rosch, E. 1975a. Cognitive reference points. *Cognitive Psychology* 7. 532–547.
- Rosch, E. 1975b. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104. 192–233.
- Rosch, E. 1976. Basic objects in natural categories. *Cognitive Psychology* 8(3). 382–439.
- Rosch, E. 1978. Principles of categorization. In E. Rosh & B.B. Lloyd (eds.), *Cognition and categorization.*, Hillsdale, NJ Erlbaum.
- Rosch, E. & C. B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7(4). 573–605.
- Rosch, E. & C. B. Mervis. 1983. Fuzzy set theory and class inclusion relations in semantic categories. *Journal of Verbal Learning & Verbal Behavior* 22(5). 509–525.
- Rúa, Paula López. 2003. *Birds, colours and prepositions. the theory of categorization and its applications in linguistics.* Munchen: Lincom.
- Russell, B. 1923. Vagueness. *The Australasian Journal of Psychology and Philosophy* 1. 84–92.
- Savas, D. 1990. *Meanings and prototypes.* London: Routledge.
- Schalkoff, R.J. 1992. *Pattern recognition: Statistical, structural and neural approaches.* New-York: John Wiley & sons, Inc.
- Silverman, K. E., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman & P. Prica. 1992. ToBi: a standard for labeling English prosody. In *Proceedings of the 2nd international conference on the processing of spoken language*, 867–870.

- Sivanandam, S.N., S. Sumathi & S.N. Deepa. 2007. *Introduction to fuzzy logic using matlab*. Berlin Heidelberg: Springer-Verlag.
- Smithson, M. & J. Verkuilen. 2006. *Fuzzy set theory. Application in the social sciences*. Thousand Oaks, CA: Sage Publications.
- Taylor, J.R. 2004. *Linguistic categorization*. Oxford, UK: Oxford University Press.
- Terry, R. 1967. The frequency of use of the interrogative formula *est-ce que*. *The French Review* 40. 814–816.
- Theodoridis, S. 2008. *Pattern recognition*, 4thed.. Burlington: Elsevier.
- Torczyner, H. 1977. *René Magritte: signes et images*. Paris: Draeger.
- Trager, G.L. & H.L. Smith. 1951. *An outline of English structure*. Norman, OK: Battenburg Press.
- Vion, M. & A. Colas. 2002. La reconnaissance du pattern prosodique de la question. *Travaux Interdisciplinaires Parole et Langage (TIPA)* 21. 153–177.
- Walter, H. 1977. *La phonologie du français*. Paris: PUF.
- Walter, M.A. & V. Hacquard. 2005. Conditioned allophony in speech perception: An MEG study. In *Second Old World Conference in Phonology (ocp2)*, Tromso.
- Webb, A.R. 2011. *Statistical pattern recognition*. Chichester: Wiley.
- Welby, P. 2003. *The Slaying of Lady Mondegreen, being a study of french tonal association and alignment and their role in speech segmentation*: Ohio State University dissertation.
- Welby, P. 2004. The structure of french intonational rises: A study of text-to-tune alignment. In B. Bel & I Marlien (eds.), *Proceedings of speech prosody 2004, 23-26 March*, Nara.

- Welby, P. 2006. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34. 343–371.
- Welby, P. & H Loevenbruck. 2006. Anchored down in anchorage: Syllable structure and segmental anchoring in French. *Italian Journal of Linguistics. Special issue on “Autosegmental-metrical approaches to intonation in Europe: tonal targets and anchors”* 18. 74–124.
- Wells, R. 1945. The pitch phonemes of English. *Language* 21. 27–40.
- Wightman, C. 2002. Tobi or not tobi? In *1st internat. conf. on speech prosody. aix en provence.*, .
- Wittgenstein, L. 1953. *Philosophical Investigations*. Wiley-Blackwell.
- Xu, Y. 2004a. Transmitting tone and intonation simultaneously - the parallel encoding and target approximation (penta) model. In *Proceedings of international symposium on tonal aspects of languages: With emphasis on tone languages*, 215–220.
- Xu, Y. 2004b. Understanding tone from the perspective of production and perception. *Language and Linguistics* 5. 757–797.
- Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46. 220–251.
- Xu, Y. 2007. Speech as articulatory encoding of communicative functions. In *Proceedings of The 16th International Congress of Phonetic Sciences, Saarbrucken* 25–30.
- Xu, Y. 2009. Timing and coordination in tone and intonation - an articulatory-functional perspective. *Lingua* 119. 906–927.
- Xu, Y. 2011. Speech prosody: a methodological review. *Journal of Speech Sciences* 85–115.
- Xu, Y. & F. Liu. 2006. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics* 18. 125–159.

- Xu, Y. & M. Wang. 2009. Modeling tone and intonation in mandarin and english as a process of target approximation. *Journal of the Acoustical Society of America* 405–424.
- Xu, Y. & S. Xuejing. 2002. Maximum speed of pitch change and how it may relate to speech. *J. Acoust. Soc. Am* 111. 1399–1413.
- Yan, J. 1994. *Using fuzzy logic*. Prentice Hall.
- Zadeh, L.A. 1965. Fuzzy sets. *Information and Control* 37. 338–353.
- Zadeh, L.A. 1994. Fuzzy logic, neural networks, and soft computing. *Communication of the A.C.M.* 8. 77–84.
- Zadeh, L.A. 1997. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems* 90. 111–127.
- Zadeh, L.A. 1998. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. *Soft Computing* 2. 23–25.

Vita

At the Université Paris 5 - René Descartes, Nicholas Bacuez graduated with a Bachelor of Arts (2003) and a Master of Arts (2005), both in Linguistics. Nicholas entered the graduate program in French Linguistics at the University of Texas at Austin in August 2005. He worked as a Graduate Research Assistant, an Assistant Instructor, and was awarded the Graduate School Continuing Fellowship in 2011.

Permanent address: nicholasbacuez@yahoo.com

This dissertation was typeset with L^AT_EX[†] by the author.

[†]L^AT_EX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's T_EX Program.