This dissertation is organized in six chapters, which conceptually comprise four parts. Following the introduction, which contains a brief review of the literature regarding tonal scaling contrasts and states the objectives of the study, the first part discusses the prosodic distinction between wh-questions and yes-no questions, examining not only the difference in tonal scaling related to the nuclear region (Chapter 1), but also how listeners exploit the accentual properties of the prenuclear pitch accent when the nuclear intonational properties are neutralized (Chapter 2).

In the second of the four parts, we present data from a production experiment (Chapter 3) and a perception experiment (Chapter 4) laying out the prosodic distinctions between information-seeking yes-no questions (questions for which the speaker has no particular bias towards the answer s/he expects) and confirmation-seeking yes-no questions (questions for which the speaker has some bias based on belief, expectations, world knowledge or information that has become available in the discourse context).

In the third part, which consists of Chapter 5 (the only chapter of the dissertation which depicts a scaling contrast within the boundary domain), we present a perception experiment dealing with the contrast between statements of the obvious and counterexpectational questions.

In the fourth and final part, Chapter 6 features a production and perception experiment addressing the relevance of three prosodic parameters (alignment, duration and scaling) in conveying narrow contrastive focus in three Romance languages: Catalan, Italian and Spanish. Finally, the major findings of the study are summarized in the conclusions section.
A mumpare i a mumare,
flama encesa
The covers of this dissertation were designed by Clau&Cat.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td><strong>Part I: Yes-no questions and wh- questions</strong></td>
<td>13</td>
</tr>
<tr>
<td>Chapter 1: The importance of nuclear tonal scaling in interrogative interpretation</td>
<td>15</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>15</td>
</tr>
<tr>
<td>2. Methodology</td>
<td>18</td>
</tr>
<tr>
<td>2.1. Participants</td>
<td>19</td>
</tr>
<tr>
<td>2.2. Materials</td>
<td>19</td>
</tr>
<tr>
<td>2.3. Procedure</td>
<td>20</td>
</tr>
<tr>
<td>2.3.1. Identification task</td>
<td>20</td>
</tr>
<tr>
<td>2.3.2. Discrimination task</td>
<td>21</td>
</tr>
<tr>
<td>2.4. Statistical analyses</td>
<td>22</td>
</tr>
<tr>
<td>3. Results</td>
<td>23</td>
</tr>
<tr>
<td>3.1. Identification results</td>
<td>23</td>
</tr>
<tr>
<td>3.2. Reaction Time results</td>
<td>26</td>
</tr>
<tr>
<td>3.3. Discrimination results</td>
<td>27</td>
</tr>
<tr>
<td>3.4. Models of discrimination performance</td>
<td>31</td>
</tr>
<tr>
<td>3.4.1. Relating discrimination responses to identification responses</td>
<td>31</td>
</tr>
<tr>
<td>3.4.2. Signal Detection Theory</td>
<td>34</td>
</tr>
<tr>
<td>4. Discussion and conclusions</td>
<td>37</td>
</tr>
<tr>
<td><strong>Chapter 2: The role of prenuclear accents on interrogative recognition</strong></td>
<td>41</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>41</td>
</tr>
<tr>
<td>2. Experiment 1: Identification task</td>
<td>45</td>
</tr>
<tr>
<td>2.1. Methodology</td>
<td>45</td>
</tr>
<tr>
<td>2.1.1. Participants</td>
<td>45</td>
</tr>
<tr>
<td>2.1.2. Materials</td>
<td>45</td>
</tr>
<tr>
<td>2.1.3. Procedure</td>
<td>47</td>
</tr>
<tr>
<td>2.1.4. Statistical analyses</td>
<td>48</td>
</tr>
<tr>
<td>2.2. Results</td>
<td>48</td>
</tr>
<tr>
<td>2.2.1. Identification results</td>
<td>48</td>
</tr>
<tr>
<td>2.2.2. Reaction Time results</td>
<td>50</td>
</tr>
<tr>
<td>3. Experiment 2: Gating task</td>
<td>52</td>
</tr>
<tr>
<td>3.1. Methodology</td>
<td>53</td>
</tr>
<tr>
<td>3.1.1. Participants</td>
<td>53</td>
</tr>
<tr>
<td>3.1.2. Materials</td>
<td>53</td>
</tr>
<tr>
<td>3.1.3. Procedure</td>
<td>54</td>
</tr>
<tr>
<td>3.1.4. Statistical analyses</td>
<td>55</td>
</tr>
<tr>
<td>3.2. Results</td>
<td>55</td>
</tr>
<tr>
<td>4. Discussion and conclusions</td>
<td>60</td>
</tr>
</tbody>
</table>
Part II: Information-seeking questions and confirmation-seeking questions

Chapter 3: Information and confirmation yes-no questions in Eastern Catalan: production data

1. Introduction
   1.1. What does interrogativity mean?
   1.2. Interrogativity and syntax
   1.3. Interrogativity and semantics
   1.4. Interrogativity and pragmatics
   1.5. Goals

2. Methodology
   2.1. Participants
   2.2. Materials
   2.3. Procedure
   2.4. Analysis

3. Results
   3.1. Central Catalan
   3.2. Majorcan Catalan
   3.3. Minorcan Catalan
   3.4. Ibizan and Formenteran Catalan

4. Discussion and conclusions

Chapter 4: Intonation as an encoder of speaker certainty: perception data

1. Introduction

2. Methodology
   2.1. Participants and general procedure
   2.2. Congruity test methodology
   2.3. Rating test methodology
   2.4. Tests based on the classical CP paradigm
   2.5. Statistical analyses

3. Results
   3.1. Congruity test results
   3.2. Rating test results
   3.3. Classical CP test results
      3.3.1. Identification results
      3.3.2. Reaction Time results
      3.3.3. Discrimination results

4. Discussion and conclusions

Part III: Counterexpectational questions and statements of the obvious

Chapter 5: Evidence for mid boundary tones in Catalan

1. Introduction

2. General methodology
   2.1. Experimental tasks
2.1.1. Experiment 1: Congruity task 115
2.1.2. Experiment 2: Identification task 115
2.1.3. Experiment 3: Discrimination tasks 115
2.1.4. Experiment 4: Imitation task 117

2.2. Speech materials 117
   2.2.1. Stimuli for Experiment 1 117
   2.2.2. Stimuli for Experiments 2-4 119

2.3. Participants 121
2.4. Statistical analyses 122

3. Results 122
   3.1. Experiment 1: Congruity task 122
   3.2. Experiment 2: Identification task 123
   3.3. Experiment 3: Discrimination tasks 125
      3.3.1. AX Discrimination Task 125
      3.3.2. AX Discrimination Task with increased step size 127
      3.3.3. AX Discrimination Task with increased ISI 128
      3.3.4. Four-Interval Discrimination Task 129
      3.3.5. Semantic Discrimination Task 130
   3.4. Experiment 4: Imitation task 131

4. Discussion and conclusions 134

Part IV: Narrow contrastive focus and broad focus declaratives 141

Chapter 6: The contribution of alignment, duration and scaling to the perception of contrastive focus in Catalan, Italian and Spanish 143
1. Introduction 143
2. Experiment 1: Production experiment 146
   2.1. Methodology 146
      2.1.1. Participants 146
      2.1.2. Experimental materials and elicitation method 146
      2.1.3. Analysis 148
      2.1.4. Statistical analyses 151
   2.2. Results 151
      2.2.1. Alignment results 151
      2.2.2. Pitch range and pitch scaling results 152
      2.2.3. Duration results 155
   2.3. Discussion 155
3. Experiment 2: Perception experiment, identification task 156
   3.1. Methodology 157
      3.1.1. Participants 157
      3.1.2. Materials 157
      3.1.3. Procedure 158
      3.1.4. Statistical analyses 159
   3.2. Results 159
      3.2.1. Identification results 160
      3.2.2. Reaction Time results 160
      3.2.3. Logistic regression results 163
4. Discussion and conclusions 164

Conclusions 167

Appendix 173

Curriculum Vitae 175

References 177
Actually this dissertation is also the story of a passion, and the most heartfelt gratitude in these acknowledgments go to those who have accompanied me on this journey and shared with me this passion for sounds and language. Like all love stories, it deserves a sound track. I would suggest that you follow the musical path that I propose below in order to better understand this journey that started in 1998 with Magdalena Capellà, Llorenç Carrió and myself as crew members.

My first song is *Aquest camí tan fi* (a poem by Josep Carner sung by Arianna Savall), a song which is a reflection about the different paths that one can choose in life and the doubts and dilemmas of life. I dedicate this song and give my profound thanks to Pilar Prieto for her constant guidance but especially for offering me the possibility of breaking fresh ground, overcoming fears and obstacles and stimulating day after day my desire to learn. Thanks also for having provided me the possibility of working with and getting to know such a long list of people who have all contributed to the improvement of my research and to compensating for my shortcomings: I am grateful to Lluïsa Astruc, Stefan Baumann, Aoju Chen, Valter Ciocca, Laura Dilley, Mariapaola D’Imperio, Gorka Elordieta, Eva Estebas, Ingo Feldhausen, Alexander L. Francis, Sónia Frota, Barbara Gili-Fivela, Martine Grice, José Ignacio Hualde, Frank Kügler, Ignasi Mascaró, Marta Ortega, Elinor Payne, Caterina Petrone, Brechjte Post and Elina Savino for their helpful advice and discussion of different aspects of my work as well as for providing me with important papers and references.

I am very grateful to my PhD co-advisor Teresa Cabré for helping me to move forward in my understanding of Catalan phonology and morphology and much more than that for her friendship and constant support. It was Teresa who first started me out on my adventure with truncated vocatives. The song *Paisatge de l’Ebre* (by Teresa Rebull) is dedicated to her. It reminds me of the long hours spent with her in our office at the Universitat Autònoma de Barcelona talking about so many topics: linguistics, music, food and life in general.

Thanks also to the Departament de Filologia Catalana at the UAB and the Departament de Traducció i Ciències del Llenguatge at the UPF for making me feel welcome from the very first moment and to the teachers and researchers of my first year of doctoral studies and the Centre de Lingüística Teòrica (CLT) for initiating me in such exciting fields. Thank you for these years of not only academic but also personal support. To all of you I would like to dedicate a traditional Majorcan song, *So de Pastera* (as sung by Música Nostra). I have chosen this song because I think that it is so representative of my Majorcan origins.

I am very grateful to the researchers, PhD students, grant-holders and collaborators of the Grup d’Estudis de Prosòdia (GrEP) for so many things: for stimulating...
discussions on linguistics, intonation and acquisition over the last few years, for technical support, for sharing ideas on language, culture and life but, especially, for the tremendous sense of humor that characterizes you and for always feeling like having fun. The best song for you is *Datemi un martello* (by Rita Pavone). It is a good song to play down the importance of the difficulties that one has to face when doing research. I remind you that we still lack the choreography.

Big thanks go to the Centro di Ricerca Interdisciplinare sul Linguaggio (CRIL) where I did my research stay during the first term of 2010 and also to the friends and colleagues that I met in Lecce during this period, people not necessarily related to linguistics. The song for them is *Baciù 'nvelanatu* (a traditional song interpreted by Zimbaria). They all know that though I had a difficult time in Lecce I felt supported, accompanied and loved by every one of the members of CRIL. Special thanks to Barbara and Antonio, who introduced me to the world of articulatory phonology, but especially because they have become like a family to me.

All the subjects that participated unselfishly in the experiments as listeners and speakers deserve a special mention not only for their participation but for having replied to the “subject recruitment” e-mails and calls so fast. I would like to thank all the friends and students who were “suddenly” involved in a spiral of experiments and recordings. The song for them is *Gonna fly now* (by Bill Conti).

I would also like to thank the people that have been close to me despite the distance that separated us: my friends from Mallorca, Madrid and Sardinia, my sister and brother and their families. Thanks too to those that have been close to me geographically as well as spiritually. I am particularly grateful to the members of the *Les Illes* project and its group of fans. *Les Illes* was an unrepeatable experience at the personal and musical level. With this project I recovered the feeling that everything is possible. The traditional Majorcan song that we sang in *Les Illes* with lyrics by Mestre Lleó is for all of you. We have to defend ourselves… *it is not only the legitimate right of defending what is ours or defending the threatened spirit of a people, we are talking about defending the world, which in essence is plurality and exuberance. Every time a custom, a language or a species is lost, the planet but also the heart of man is impoverished* (Antoni Vidal Ferrando: *Amors i laberints*).

Many thanks indeed to Felip, my partner in life. My song for him is *Dóna'm la mà* (a poem by Joan Salvat-Papasseit sung by Arianà Savall). Thanks for understanding me, for trusting me, for offering me always the possibility of choosing but above all for sharing this adventure with me hand in hand. Thanks also for your critical awareness and for listening to my linguistic digressions at any moment. More thanks for having replied so many times and so quickly to my insistent vocatives —“Felip, O Felip, come here right now and look at this really interesting pitch accent!”

I owe everything to my mother, but especially her vitality and strength. My song for her is *Mercè* (by Maria del Mar Bonet). *Mercè* is a song that Maria del Mar Bonet devoted to her mother to assure her that she would come back home. During my father’s terrible decline, she gave us a life lesson in overcoming adversity but also a lesson in unconditional love. It is very difficult to put my feelings in words but all I can say is that I feel extremely lucky to have a person like her with me, a person who is life in its
purest form, and that I will devote this time that I am away from home to spreading her lessons about life.

This last paragraph is for my father and my song for him is *Per a ser cantada en la meva nit* (a poem by Salvador Espriu as sung by Arianna Savall). First of all, let me apologize for not having been with you when you needed me most. Thanks so much for allowing me to share that precious time during your last moments of lucidity last Christmas when we said so much to each other. I have always known that we had a special place outside of memory in which understanding between us was possible, a place in which we met and we talked about us. My last words are in defence of all those suffering from Alzheimer’s disease. Each man is not only just memory but also feeling, will, sensibility, a moral self and subject. The restoration of the integrity is possible through art, communion and contact with the human spirit, and this possibility can be preserved even when “in principle” there is nothing left but utter neurological devastation (thanks to Oliver Sacks for lending me his words).
Introduction

During the past thirty years a considerable amount of research has been devoted to the description and analysis of intonation of Catalan. The first thorough studies of Catalan intonation were Bonet (1984) and later Prieto (2002b), who focused on describing the intonation patterns found in different sentence types in Central Catalan. Also, several studies have analyzed specific Catalan intonation contours (mostly declaratives and interrogatives), both for Central Catalan (Recasens 1977; Virgili Blanquet 1971; Bonet 1986; Salcioli 1988; Prieto 1995, 2001, 2002b; Estebas-Vilaplana 2003, 2009; Font 2007; and others) and for other dialectal varieties (Mascaró i Pons 1986, 1987; Prieto 2001; Prieto & Pradilla 2002; Martínez-Celdrán, Fernández & Carrera 2005; Martínez-Celdrán, Fernández, Salcioli-Guidi, Carrera & Es Puny 2005; Fernández, Carrera, Román & Martínez-Celdrán 2006; Fernández, Martínez-Celdrán, van Oosterzec, Salcioli, Castellví & Szmidt 2007; Vanrell 2006; Prieto & Rigau 2007; Crespo-Sendra 2008; and others). In the last decade, relevant research has been conducted within the Autosegmental-Metrical framework focusing on a variety of issues related to the typology of pitch accents and boundary tones as well as tonal contrasts in Catalan (Estebas-Vilaplana 2003, 2009; Astruc 2005, Astruc & Nolan 2007; Prieto 2005, 2007; Prieto, D’Imperio & Gili-Fivela 2005; Borràs-Comes, Vanrell & Prieto 2010; Crespo-Sendra, Vanrell & Prieto 2010). The present work represents an addition to the body of research which is grounded in the framework of Autosegmental-Metrical framework (henceforth AM) and which has the goal of investigating the role of tonal scaling in the intonational grammar of Catalan both from a production and a perception point of view.

The AM model is one of the most widely accepted phonological frameworks for analyzing intonation. In this framework, the F0 contour of an utterance is described as a sequence of high (H) and low (L) tones, with an additional mid (M) or downstepped H (!H) tone in certain languages. The tones are of two kinds, pitch accents and boundary tones. Pitch accents are tonal events that are associated with the metrically prominent syllables in a sentence, and they can be either monotonal (e.g., H*, L*) or bitonal (e.g., L+H*, L*+H, H+L*). The association of tones with the metrically strong syllable is indicated through the use of an asterisk (*) after the specific tone. In bitonal pitch accents, the starred tone (the one corresponding to the stressed syllable) is either preceded by a leading tone or followed by a trailing tone. Boundary tones are tonal events that are associated with the edges of prosodic phrases. They can be high (H) or low (L). The boundary tones associated with the right edges of intonational phrases (IP) are marked with a ‘%’ sign following the tone (e.g., H%, L%). An IP can have more than one pitch accent, and the final one is usually referred to as the nuclear pitch accent; the remaining pitch accents are referred to as the prenuclear pitch accents. An important point related to this system is that it is motivated by phonetic and phonological considerations. In other words, the linguistic categories should be related to differences in sound but also to differences in semantic interpretation like in the segmental domain, where “[p] is different from [b] because they are pronounced differently, and because

The most recent version of the application of the AM system to the Catalan language (or Cat_ToBI system\(^1\)) is based on early work on Catalan within the field of intonational phonology (Prieto et al. 2009, Prieto in press). It can also be found at the website of the Cat_ToBI Training Materials (Aguilar, De-la-Mota & Prieto (coords.) 2009-2011). This version is based on the analysis of Central Catalan, meaning that the tonal units and tonal contrasts proposed have been attested at least (yet not solely) in the Central dialect. The Cat_ToBI system\(^2\) proposes seven basic pitch accents for Catalan: two monotonal pitch accents (H* and L*) and five bitonal pitch accents (H+L*, L*+H, L+H*, L+>H* and L+¡H*). The symbol “>” indicates that the maximum in F0 is attained outside the stressed syllable (Prieto et al. 2005). With respect to boundary tones, fourteen boundary tones are attested: eight types of boundary tones at the end of IPs (L%, !H%, H%, HH%, HL%, LH%, L!H% and LHL%); five types of boundary tones at the end of intermediate phrases, or “ips” (L-, !H-, H-, HH- and LH-); and one type of initial boundary tone (%H).\(^4\) The main difference between this version and previous versions (Prieto et al. 2009) is the presence of an additional pitch accent, L+¡H*, and a labelling difference with respect to the mid boundary tone, which will be dealt with in Part III of this dissertation. The proposal that the L+¡H*\(^5\) pitch accent is part of the intonational grammar of Catalan partially anticipates the goal of this PhD dissertation, which seeks to present experimental evidence for discrete intonational contrasts in pitch range variation.

The scope of this dissertation is twofold. On the one hand it addresses the issue of whether pitch range variation gives rise to distinctive linguistic categories in Catalan. On the other hand, it seeks to contribute to this discussion by assessing the appropriateness of different perceptual and production experimental methods such as the DCT (Discourse Completion Test, see Part II, Chapter 3 of the dissertation) or the imitation task for production (see Part III), and the congruity task, identification task and different types of discrimination tasks for perception (see Part II and Part III).

Although the variation in pitch height has been assumed to be paralinguistic by the standard AM model (Pierrehumbert 1980, Pierrehumbert & Beckman 1988), that is, its

---

\(^1\) The Cat_ToBI labelling scheme is a tool for the annotation of Catalan speech corpora that contains prosodic, phonetic and intonational information (Aguilar, de-la-Mota & Prieto (coords.) 2009-2011).

\(^2\) See the Appendix for a schematic representation of the pitch accents and boundary tones found in Catalan.

\(^3\) In Aguilar, de-la-Mota and Prieto (coords.) (2009-2011) the label M% is used when a sustained boundary tone is found, realized either as a rising movement to a mid tone target after a low tone or as a mid level plateau after a high tone. See Part III of this dissertation for an explanation of the proposed replacement of the label M% by the label !H%.

\(^4\) See the Appendix for a schematic representation of the tonal inventory of the Cat_ToBI system.

\(^5\) Borràs-Comes, Vanrell and Prieto (2010) tested the nature of a possible three-way tonal scaling contrast (narrow focus statements vs. contrastive focus statements vs. counterexpectational questions). They argued that counterexpectational questions differ from both narrow focus statements and contrastive focus statements in a discrete way. They thus included the category L+¡H* in the Cat_ToBI phonological analysis of tones and recommended a reassessment of the concept of upstep and downstep within the AM framework of intonation.
function is assumed to exclusively express phonetic differences in emphasis or prominence, some studies have shown that the difference in pitch height may also lead to categorical effects. Ladd and Morton (1997) applied the Categorical Perception paradigm (CP) to a contrast between two pitch-accents in English, the normal high accent and the emphatic extra high accent. Evidence was found for a shift between these two categories in identification but a related discrimination peak was not. As the main assumption of classical CP is that discrimination is easier at the category boundary and more difficult within categories, they concluded that this contrast was not categorically perceived. From the results of Ladd and Morton (1997), Chen (2003) argued that the absence of a discrimination peak might be related not to the inexistence of categorical perception but rather to a hypothetical unsuitability of the CP paradigm as applied to a pitch height contrast. Thus, Chen (2003) proposed an alternative method to examine the nature of two intonational contrasts between normal high accent vs. emphatic high accent and early peak alignment vs. late peak alignment, which she called the reaction Time (RT) measurement. Noting that mean RT was shortest for within-category identification and longest for across-category identification only for the peak height continuum and not for the peak alignment contrast, and basing herself on previous studies (Pisoni & Tash 1974), Chen concluded that this pitch height contrast is discrete.

In a production and perception experiment, Calhoun (2004) found that themes and rhemes in English are marked by distinctive pitch accents and that the most reliable cue to the theme and rhyme accents is pitch height.

Work on languages other than English has also revealed that pitch range variation can convey discrete intonational contrasts. Prieto (2004) studied the effects of sentence type on scaling variation of the first pitch accent peak in Castilian Spanish, and found evidence that pitch height is not solely related to paralinguistic variation, since sentence-type information also has a strong effect on H1 scaling in Spanish. Face (2005) carried out a gating perception experiment to examine the disambiguating role of intonation in the perception of two sentence types in Castilian Spanish. It was found that the different scaling of the first F0 peak between declaratives and absolute interrogatives is the cue that leads to 95% accuracy in the perception of sentence type in Castilian Spanish. Savino and Grice (2007, 2011) tested whether the pitch range of a rising nuclear configuration was responsible for the difference between two types of questions in Bari Italian. In this language the same rising pitch accent is used for two types of questions, namely those seeking information and those challenging what has been said. However, they differ in their pitch range: while information-seeking questions are produced with a compressed pitch range (L+H*), challenging-echo questions have an expanded pitch range (L+¡H*). The results of a semantically motivated identification task provided clear evidence for the categorical use of pitch range variation in Bari Italian question interpretation. In a similar fashion, Borràs-Comes et al. (2010) performed three types of perception experiments, namely identification tasks with a dual response, identification tasks with a triple response and a congruity task to test the nature of a possible three-way tonal scaling contrast (narrow focus statements vs. contrastive focus statements vs. counterexpectational questions). They concluded that counterexpectational questions differ from both narrow focus
statements and contrastive focus statements in a discrete way. In addition, it seems that this specific pattern (L+¡H*) found in counterexpectational questions for Central Catalan (Borràs-Comes et al. 2010) and Bari Italian (Savino & Grice 2007, 2011) is also attested in Friulian (Roseano, Vanrell & Prieto 2011). These investigations reveal that pitch range variation can be perceived in a discrete fashion in some languages, which strengthens the arguments in favour of treating pitch range differences in phonological terms. The idea of enriching the traditional High-Low dichotomy with a finer differentiation of pitch range is not new and has been advocated by researchers such as Ladd (1994). As Ladd (1994: 60) pointed out, “the Bruce-Pierrehumbert approach to intonational phonology must be enriched with a notion of categorical distinctions of pitch range. We need to get rid of the idea that any distinction that is orthogonal to the basic opposition between High and Low tones is ipso facto gradient: both gradient factors and categorical ones play a role in the vertical scale of any given tone”.

In the boundary tone domain, similar contrasts in H scaling have been reported. Thus, in languages like English, Greek, German, Spanish, Korean or Portuguese a mid level boundary tones is needed in order to account for linguistic contrasts (see Beckman & Ayers-Elam 1997 for American English; Grabe 1998 for British English and German; Post 2000 for French; Beckman et al. 2002 for Spanish; Lee 2003 for Korean; Arvaniti & Baltazani 2005 for Greek; Grice, Baumann & Benzmüller 2005 and Peters 2006 for German; and Frota in press-b for European Portuguese).

A number of experimental methods have investigated what is paralinguistic and gradient in the intonation domain (see a review in Gussenhoven 2004, 2006; see also Dilley 2010), including imitation tasks (Pierrehumbert & Steele 1989; Dilley 2005, 2007; Dilley & Brown 2007), congruity tasks (Rathcke & Harrington 2010; Borràs-Comes et al. 2010; Crespo-Sendra et al. 2010), prominence judgments (Gussenhoven & Rietveld 1988) and semantic judgments (Gussenhoven & Rietveld 2000; Calhoun 2006). Several studies have applied the complete categorical perception paradigm to the study of intonation contrasts, be they between boundary tones, e.g., L% vs. H% contrasts that prototypically mark the difference between questions and statements (Remijsen & van Heuven 1999 for Dutch; Schneider & Linfert 2003 for German; Cummins, Doherty & Dilley 2006 for English; Falé & Hub Faria 2006 for European Portuguese) or pitch accents (Kohler 1987 for German; Ladd & Morton 1997 and Dilley 2010 for English; Gili-Fivela 2008 for Italian). In these articles, while identification curves show clear discrete effects, discrimination results are less clear cut and thus these articles leave several questions unanswered. Although claims are made of ‘categorical perception’ for a particular contrast, out of the abovementioned articles, only three show clear evidence of categorical perception, with a clear discrimination peak at the expected position (Kohler 1987, Remijsen & van Heuven 1999, Schneider & Linfert 2003). The explanations for this lack of a clear peak in discrimination functions are varied. Some researchers (e.g., Remijsen & van Heuven 1999) argue that this absence of a discrimination peak could be due to the fact that intonational units, like vowel phonemes, are encoded over long time intervals, which can favour a continuous rather than a categorical perception. Other researchers like Gili-Fivela (2009) argue that the mixed results obtained by the use of the CP paradigm for studying intonation might be
due to the “complex meaning intonation may have and the possible interpretations subjects may give it” (p. 12). In her view, meanings involving a linguistic distinction such as the one between questions and statements or yes-no and wh-questions would be more easily categorized than those involving a linguistic meaning related to focus since “focus may easily become part of the paralinguistic domain, involving degrees of emphasis rather than categorical changes of meaning/interpretation” (p. 13). In the work by Chen (2003), Kohler (2006), Niebuhr and Kohler (2004) and Frot a (in press-a), it is claimed that discrimination tasks are not suitable for investigating discreteness in intonational contrasts because these tasks are too focused on the acoustic perceptive properties of the stimuli rather than on tonal categories. Thus, according to these studies, the absence of a clear peak should be related not to the nonexistence of a phonological distinction but rather to the unsuitability for different reasons of applying the CP paradigm to intonational contrasts. By comparing the results of congruity tasks, imitations tasks, identification tasks and different types of discrimination tasks, we hope to be able to test the applicability of this paradigm to the analysis of tonal contrasts.

In sum, this dissertation addresses the issue of how Catalan speakers use pitch height to convey differences in pragmatic meaning associated with different sentence types such as yes-no questions, wh-questions or statements. The first goal of the study is to investigate by means of production and perception experiments the nature (i.e., whether discrete or gradient) of certain pitch scaling differences found in Catalan intonation. An important question concerns how the ToBI prosodic annotation system in place for Catalan (Cat_ToBI: Prieto et al. 2009; Prieto in press; Aguilar et al. (coords.) 2009-2011) can capture the differences that will be shown to be of a categorical nature as well as how this system can be integrated into a common transcription system for Romance languages. The second goal of the study is to test the convergence and degrees of appropriateness of different perceptual and production experimental methods. Importantly, we will investigate the interplay existing between production and perception of prosody in an attempt to increase the comprehension of the principles involved both in production and perception. Finally, we will discuss the typological implications of the ways different languages exploit pitch range and relate these typological implications to the prosodic annotation systems for these different languages.

The language under study is Catalan. Parts I and II of this dissertation deal with two intonational distinctions that have only been consistently attested in Majorcan Catalan. However, other Balearic subdialects, Alguerese Catalan and some varieties of Northwestern and Central Catalan use the same phonological contrast to distinguish between yes-no questions and wh-questions or information-seeking questions and confirmation-seeking questions (Prieto & Cabrè (coords) 2007-2010). Part III analyses the prosodic difference between statements of the obvious and counterexpectational questions which is found in Central Catalan but is also attested in Northwestern and Valencian Catalan. Finally, the contrast between broad and narrow contrastive focus described in Part IV is found in the general Catalan linguistic domain with the exception of Northern and Alguerese Catalan. Though there is important dialectal variation in Catalan with respect to lexicon, morphology and segmental phonetics, dialects are
mutually intelligible. As far as suprasegmental phonetics is concerned, the most important differences mostly centre around interrogatives, with small or non-existent differences with respect to other sentence types like statements, imperatives or vocatives (Prieto & Cabré (coords.) 2007-2010). Both Central and Majorcan Catalan belong to the Eastern Catalan dialect group along with the other Balearic subdialects, Northern Catalan and Alguerese. Central Catalan is considered to be the standard form of the language and has the highest number of speakers. It is spoken in populated areas such as the province of Barcelona, the eastern half of the province of Tarragona and most of the province of Girona. Majorcan Catalan is spoken in Majorca, the largest of the Balearic Islands, with a population of roughly 860,000 inhabitants. The most prominent differences between Central and Majorcan Catalan involve lexicon (e.g., al·lot for noi ‘boy’, moix for gat ‘cat’, etc.), morphosyntax (e.g., Balearic varieties preserve the sa definite article derived from Latin ipse/ipsa, the first person singular present indicative has no morpheme representation in contrast with Central Catalan, etc.) and phonetics (Majorcan has a vowel system of eight stressed vowels /a ə e i o ɔ u/ reduced to four [ɔ i o u] in unstressed position, while Central Catalan has a vowel system of seven stressed vowels /a ə e i o ɔ u/ per three unstressed vowels [ɔ i u]. Once again, the differences between these varieties are not significant enough to prevent the mutual comprehension between speakers of Central and Majorcan Catalan. The following map represents the two major dialect blocks, Eastern and Western Catalan, as well as the individual dialects into which each block can be subdivided.

Map of the Catalan linguistic domain.
In Part IV, in addition to Majorcan Catalan, the varieties of Lecce Italian and Madrid Spanish are analyzed. Lecce Italian is the variety of Italian spoken in the province of Lecce (Apulia region), which has a population of approximately 814,000 inhabitants. Madrid Spanish belongs to the variety known as Castilian Spanish, which is spoken in north and central Spain. It is the national standard for radio and TV broadcasts.

This study is organized in six chapters, which conceptually comprise four parts. Following this introduction, which contains a brief review of the literature regarding tonal scaling contrasts and states the objectives of the study, the first part discusses the prosodic distinction between wh- questions and yes-no questions, examining not only the difference in tonal scaling related to the nuclear region (Chapter 1), but also how listeners exploit the accentual properties of the prenuclear pitch accent when the nuclear intonational properties are neutralized (Chapter 2). The results obtained in this first part have important implications not only for the notion of ‘pitch range variation’ within the AM framework of intonation but also for the idea of ‘compositional approach’ advocated for this model. The findings of Chapter 2 are a clear indication that listeners evaluate perceptually all parts of the contour and integrate them while searching for pragmatic meaning.

In the second of the four parts, we present data from a production experiment (Chapter 3) and a perception experiment (Chapter 4) laying out the prosodic distinctions between information-seeking yes-no questions (questions for which the speaker has no particular bias towards the answer s/he expects) and confirmation-seeking yes-no questions (questions for which the speaker has some bias based on belief, expectations, world knowledge or information that has become available in the discourse context). The results of this second part lead us to conclude that intonation constitutes a well-established linguistic strategy to mark the grammatical category of evidentiality across languages, and that its role in marking certainty deserves to be further investigated.

In the third part, which consists of Chapter 5 (the only chapter of the dissertation which depicts a scaling contrast within the boundary domain), we present a perception experiment dealing with the contrast between statements of the obvious and counterexpectational questions. The conclusions are that Catalan listeners perceive the contrast between the two levels of pitch height in a discrete fashion. We also conclude that a combination of semantically motivated tasks like congruity and identification tasks (along with Reaction Time measurements) can be profitably used to investigate intonational categories. In the fourth and final part, Chapter 6 features a production and perception experiment addressing the relevance of three prosodic parameters (alignment, duration and scaling) in conveying narrow contrastive focus in three Romance languages: Catalan, Italian and Spanish. The results presented in this final chapter show that while these three prosodic parameters are active in production, in terms of perception language-specific preferences for particular prosodic parameters are found, which are explained by the particular Italian marker of salience, that is, the rising-falling movement aligned with the accented syllable. Finally, the major findings of the study are summarized in the Conclusions section.

Finally, it is important to add that readers can access sound files for many of the examples from this dissertation by downloading a compressed folder from the web site http://optimitza.cat/mvanrell/publicacions.html. Files that are available on the web page
are marked by the symbol arehouse with the filename appearing next to the symbol. Permission to freely access the data was given by the speakers that participated in the various production and perception experiments presented in this dissertation.
Part I
Yes-no questions and wh- questions
CHAPTER 1

The importance of nuclear tonal scaling in interrogative interpretation

This first chapter addresses the question of how differences in pitch height can be related not only to paralinguistic meaning but also to linguistic meaning. Majorcan Catalan provides a good test case to show this because it offers pairs of phrases (such as yes-no question and wh-questions) that are homophonous at the segmental level and in which tonal scaling features presented at the nuclear region play an important role in disambiguating them. The goal of the study is to test through an experimental procedure based on the Categorical Perception paradigm (CP) whether listeners make categorical linguistic use of F0 scaling differences in the pretonic nuclear syllable in perceiving yes-no questions as opposed to wh-questions in Majorcan Catalan. The prediction is that if listeners’ perception of differences in pitch height is categorical, that will mean that pitch scaling has a phonological character in Majorcan Catalan interrogatives.

1. Introduction

In Majorcan Catalan, yes-no questions and wh-questions are characterized by a falling nuclear accent H+L*, that is, an H leading tone aligned with the pretonic syllable and a L* tone associated with the word-final stressed syllable. Yes-no questions can be headed by the unaccented interrogative particle que\(^6\) (‘that’) (Figure 1, upper panel), and this can be compared with wh-questions headed by the accented wh-particle què (‘what’) (Figure 1, bottom panel). Results from a Map Task recording by Payà and Vanrell (2005) and Vanrell (2006) demonstrated that in yes-no questions the H leading tone is upstepped (Figure 1, upper panel). As can be seen in Figure 1, the difference in tonal height of the H leading tone consequently involves a difference in the intonation pattern of these two types of questions in terms of realization. Thus, even though both types of questions are characterized by a falling final intonation, wh-questions show a steady high tone which extends from the beginning of the sentence to the last stressed syllable, with the falling pitch movement aligned with the last stressed syllable in the utterance (i.e., -ri-, in duries)\(^7\). By contrast, yes-no questions have a well-defined rising slope over the three syllables preceding the nuclear accent (i.e., -ri-).

This intonational difference is particularly relevant in Majorcan Catalan because both sentences in Figure 1 are homophonous at the segmental level, since the two interrogative

---

\(^6\) See Prieto and Rigau (2007) for a syntactic and prosodic description of Catalan yes-no questions headed by interrogative marker que.

\(^7\) A typical realization of the nuclear pitch accent of wh-questions is that of a H leading tone that is downstepped (!H+L*) with respect to the first H tone (H*) of the sentence which is aligned with the wh-word què (‘what’). However a plain H leading tone is also a possible realization of wh-questions (Vanrell 2006) that would coincide with the realization of the confirmation-seeking questions (see Chapter 4 of this dissertation).
particles (accented and unaccented) are pronounced as [ˈkə]/[kə]. This does not occur in other varieties of Catalan, in which the accented interrogative particle què is pronounced as [ˈke] and the unaccented interrogative particle que is pronounced as [kə].

Figure 1: Waveform, spectrogram and fundamental frequency contours of the yes-no question Que l’hi duries? (‘Would you bring it to her/him?’), top panel [Chap1Fig1top.wav], and the wh- question Què li duries? (‘What would you bring to her/him?’), bottom panel [Chap1Fig1bottom.wav].

Pitch height variation has been assumed to be paralinguistic by the standard Autosegmental-metrical (AM) model (Pierrehumbert 1980, Pierrehumbert & Beckman 1988), that is, it exclusively expresses a change in the degree of meaning corresponding usually to differences in emphasis or prominence. However, some studies have shown that the difference in pitch height can also trigger categorical effects (Ladd & Morton...
1997, Chen 2003, Calhoun 2004 for English; Prieto 2004 for Spanish; Savino & Grice 2007, 2011 for Bari Italian; Borràs-Comes et al. 2010 for Central Catalan). In this context, Majorcan Catalan provides a good test case to show that degrees of pitch height may not be uniquely related to paralinguistic variation, but can play a decisive role in differentiating two different utterance types: yes-no questions and wh-questions.

The Categorical Perception paradigm (CP) has been one of the most commonly used methods to examine the nature of tonal contrasts and has been applied to differences in peak alignment (Kohler 1987, D’Imperio & House 1997, Chen 2003, Gili-Fivela 2008, 2009) and differences in pitch height in both tonal languages (Francis et al. 2003, Francis & Ciocca 2003) and intonational languages, for boundary tones (Remijsen & van Heuven 1999, Post 2000, Schneider & Linfert 2003, Cummins et al. 2006, Falé & Faria 2006, Schneider, Dogil & Möbius 2009) as well as for pitch accents (Ladd & Morton 1997, Chen 2003, Gili-Fivela 2008, 2009, Dilley 2010).

The application of the CP paradigm involves the presence of asymmetries in discrimination results. Asymmetries in tonal perception occur when the discrimination of a tonal change presented in one direction is easier compared to the same change presented in the reverse direction. These asymmetries have been found repeatedly in the application of CP to the perception of not only intonational languages (Kohler 1987, Ladd & Morton 1997, Remijsen & van Heuven 1999, Schneider & Linfert 2003, Cummins et al. 2006, Falé & Faria 2006, Prieto, Estebas-Vilaplana & Vanrell 2010) but also tonal languages (Francis & Ciocca 2003). Studies that report asymmetries in the application of the CP paradigm to pitch height contrast seem to agree that two different contours are more successfully discriminated when the second one—whatever two points in a contour are compared—has a higher pitch.

Accordingly, the primary goal of the present investigation is to test through an experimental procedure based on the Categorical Perception paradigm (CP) whether Majorcan Catalan listeners make categorical linguistic use of F0 scaling differences in perceiving yes-no questions as opposed to wh-questions.

Since asymmetries have been reported repeatedly in the literature on the application of the CP paradigm to pitch height contrasts, a second goal of the study is to verify the presence of asymmetries. It is predicted that these effects do occur, such that it will be easier to discriminate between pairs of stimuli when the direction of change is upwards in pitch level than when it is downwards (Ladd & Morton 1997, Remijsen & van Heuven 1999, Schneider & Linfert 2003, Cummins et al. 2006, Falé & Faria 2006, Prieto et al. 2010).

Chapter 1 comprises four sections. The first section contains this introduction, which presents the contours studied, that is, yes-no questions and wh-questions in Majorcan Catalan, reviews briefly the literature regarding the categorical effects of pitch height (see also the general Introduction of this dissertation) as well as a preliminary discussion on the implication of applying the CP paradigm to intonational contrasts; the second section presents the methodology used in this study, that is, the Categorical Perception paradigm; section three states the results obtained for both the identification and the discrimination tasks as well as discusses different models of discrimination
performance, and, finally, section four presents a summary of the major findings of this experiment and its general conclusions.

2. Methodology

In the CP paradigm, subjects perform two tasks: an identification task and a discrimination task. In the identification task, subjects listen to randomly ordered stimuli constructed from a continuum and judge which of two categories each stimulus represents. In the discrimination task, the subjects listen to the stimuli again, but this time they are asked to identify the test stimulus in terms of a reference stimulus. In the pair-wise AX task, the subjects hear the test stimulus and a single reference stimulus, and decide whether the two stimuli are two instances of the same stimulus or different stimuli. The patterns of expected results are shown in Figure 2. The idealized functions of responses to the identification task have an S-shape (solid lines), i.e., an abrupt shift from one category to the other rather than a gradual transition. Figure 2 also shows the idealized discrimination function (dashed line). If perception is categorical, discrimination is easier when the two stimuli straddle the boundary between the categories than when the two stimuli are from within the same category. A stimulus continuum is typically considered categorical if listeners’ responses match two criteria (Francis, Ciocca & Kei Chit Ng 2003): first, identification proportions should predict discrimination accuracy (Liberman, Harris, Hoffman, & Griffith 1957); second, peaks of discrimination should correspond to the location of the category boundaries determined by identification (Repp, Healy & Crowder 1979). Typically, discrimination results are predicted through a formula taken originally from Liberman et al. (1957) and the boundary between categories in sigmoid response curves is calculated by Probit analysis, which fits a cumulative normal curve to probability estimates as a function of stimulus level by the method of least squares (Finney 1971), estimating the mean (which marks the identification crossover) and standard deviation for each distribution.

In this study the CP paradigm has not been applied in the classical sense. Firstly, in the identification task, reaction time as well as the response rate is measured. Secondly, another formula is used in addition to the formula taken from Liberman et al. (1957) to predict the discrimination performance from the identification proportions. And finally, the statistical analyses employed are different from the ones that researchers typically use in this type of study (see details on these alternative or additional tools in the relevant subsections).
2.1. Participants

Forty-two listeners (twenty-five women, seventeen men), between 16 and 41 years old, participated in the experiment. All participants were native speakers of Majorcan Catalan, coming from different areas of the island. None of them reported a history of hearing disability. Subjects had to achieve a pre-established level of identification accuracy whereby 80% of the base stimuli had to be recognized. The responses of those listeners who failed to identify 80% of the base stimuli were rejected. The data from 10 subjects was discarded for that reason. Finally, the data of 32 subjects were analyzed.

2.2. Materials

One token of the yes-no question *Que l’hi duries?* (‘Would you bring it to her/him?’) and one token of the wh- question *Què li duries?* (‘What would you bring her/him?’), based on the production results of Vanrell (2006), were produced by the author of this dissertation, which is a native speaker of Majorcan Catalan. Both tokens are homophonic at the segmental level (see Figure 1). In the yes-no question token the leading tone was 263 Hz while in the wh- question token it was 203 Hz. A linear stylization of the rising-falling movement was carried out. Three points were interpolated: a point at rising onset L1, a point at the peak H, and a point at the falling offset L2. L1 was aligned in both tokens at the onset of the pretonic syllable *du-* (*duries*), H at the offset of the vowel of the pretonic syllable *-ri-* (*duries*), and L2 at the offset of the vowel of the syllable *-ri-* (*duries*). From these two base tokens, ten stimuli were created by means of PSOLA synthesis: two synthesized base tokens (one from the yes-no question token and one from the wh- question token), four stimuli created by shifting the peak downwards from the yes-no question synthesized token (Figure 3, left panel) and four stimuli by shifting the peak upwards from the wh- question synthesized token (Figure 3, right panel) in four steps of 15 Hz each. Stimulus 1 is always 203 Hz,

---

8 Please note that the contrast presented in Figure 1 for yes-no questions and wh- questions is a general distinction attested throughout the entire island (Prieto & Cabrè (coords.) 2007-2010).
stimulus 2 is 218 Hz, stimulus 3 is 233 Hz, stimulus 4 is 248 Hz and stimulus 5 is 263 Hz (see Table I). The choice of 15 Hz for the step-size used in this study was motivated by the verification through a pilot test that this difference in Hz would not be too easily perceptible, especially in the discrimination task, in which the listeners reported perceiving hardly any difference between the two stimuli in a pair.

Figure 3: Schematized continua of the stimuli created from yes-no question base stimulus (left panel) and from wh-question base stimulus (right panel).

### 2.3. Procedure

Subjects were seated in front of a laptop in a quiet room and heard the stimuli over headphones. The perception test was played by means of PERCEVAL (Ghio, André, Teston & Cavé 2003), software for performing computerized auditory and visual perception experiments, which also records reaction times (henceforth RTs). As we were also interested in RT measurements, the listeners were instructed to rest their hands near the keyboard and to press the keys as fast as they could. The identification task preceded the discrimination task and there was no break between the two tasks. In both tasks, subjects were given written instructions about how they were to respond. There was a practice block before both the identification and the discrimination test block. The full test lasted approximately 30 minutes.

#### 2.3.1. Identification task

The materials for the identification task consisted of 4 repetitions of each of the 10 stimuli (five stimuli from the yes-no question token and five stimuli from the wh-question token). These 40 stimuli (5 stimuli x 2 question types x 4 repetitions —see Table I) were presented in blocks of 10 in random order. There was a practice block before the test session made up of the two synthesized base stimuli plus the eight stimuli created from the synthesized base tokens by shifting the peak. There was no break between the blocks. The subjects were asked to respond after each stimulus according to
how they would answer the question in a real situation. In other words, if they perceived the yes-no question *Que l’hi duries?* (‘Would you bring it to her/him?’), they were to press the “S” key on the keyboard (for *Sí* = “Yes”), whereas if they perceived the wh-question *Què li duries* (‘What would you bring her/him?’), they were to press the “A” key (for *Això* = “That”).

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>F0 value of the peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>203 Hz</td>
</tr>
<tr>
<td>2</td>
<td>218 Hz</td>
</tr>
<tr>
<td>3</td>
<td>233 Hz</td>
</tr>
<tr>
<td>4</td>
<td>248 Hz</td>
</tr>
<tr>
<td>5</td>
<td>263 Hz</td>
</tr>
</tbody>
</table>

5 stimuli

5 stimuli x 2 base stimuli x 4 repetitions + 10 practice stimuli = 50 stimuli (presented in blocks of 10 in random order).

Table I: Stimuli for the identification task.

2.3.2. Discrimination task

The materials for the discrimination task consisted of pairs of stimuli taken from the identification task. Eight pairs of stimuli were created in AB order, meaning that stimulus B is always higher in frequency than stimulus A (four from the yes-no-question-based continuum and four from the wh-question-based continuum): pair 203-218 Hz, pair 218-233 Hz, pair 233-248 Hz and pair 248-263 Hz. Eight pairs of stimuli were created in BA order in which stimulus A was lower in frequency than stimulus B (four from the yes-no-question-based continuum and four from wh-question-based continuum): pair 218-203 Hz, pair 233-218 Hz, pair 248-233 Hz and pair 263-248 Hz. Additionally, 10 AA pairs were created which contained two identical stimuli (five from the yes-no-question-based continuum and five from the wh-question-based continuum): pair 203-203 Hz, pair 218-218 Hz, pair 233-233 Hz, pair 248-248 Hz, pair 263-263 Hz. Two repetitions of all these stimuli were randomized. The practice block before the test session was based on the yes-no-question-based continuum for half the listeners and on the wh-question-based continuum for the other half. Note that there were no blocks made up of pairs of stimuli created from the yes-no question and wh-question base stimuli all together in order to avoid blocks of 26 pairs of stimuli. It was felt that this would have tired the subjects and that it was important to keep the duration of the experimental session under an hour without breaks. Thus, each subject heard a total of 65 pairs of stimuli (4 AB pairs + 4 BA pairs + 5 AA pairs x 2 question types x 2 repetitions + one practice block of 13 pairs —see Table II). Subjects were asked to decide whether they heard the pair of stimuli as same or different. If the two stimuli
sounded the same, they were to press the “I” key on the keyboard (for Igual = “Same”) and if the stimuli sounded different, they were to press the “D” key (for Diferent = “Different”). The interval between the two stimuli in each pair was 0.5 seconds. The order of the blocks was counterbalanced, that is, the half of the listeners whose practice block was based on the yes-no question base stimulus, started with the block based on the wh-question base stimulus and went on to the block based on the yes-no question base stimulus, alternating thus until they had listened to the four test blocks; the other half of the listeners heard a practice block based on the wh-question and thereafter heard the test blocks in the reverse order relative to the first group of subjects.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>4 pairs</td>
<td></td>
</tr>
<tr>
<td>Pair 1_2</td>
<td>203-218 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 2_3</td>
<td>218-233 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 3_4</td>
<td>233-248 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 4_5</td>
<td>248-263 Hz</td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>4 pairs</td>
<td></td>
</tr>
<tr>
<td>Pair 2_1</td>
<td>218-203 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 3_2</td>
<td>233-218 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 4_3</td>
<td>248-233 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 5_4</td>
<td>263-248 Hz</td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>5 pairs</td>
<td></td>
</tr>
<tr>
<td>Pair 1_1</td>
<td>203-203 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 2_2</td>
<td>218-218 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 3_3</td>
<td>233-233 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 4_4</td>
<td>248-248 Hz</td>
<td></td>
</tr>
<tr>
<td>Pair 5_5</td>
<td>263-263 Hz</td>
<td></td>
</tr>
</tbody>
</table>

13 pairs of stimuli x 2 base stimuli x 2 repetitions + 13 pairs of practice stimuli = 65 pairs of stimuli (presented in blocks of 13 in random order).

**Table II:** Pairs of stimuli for the discrimination task.

### 2.4. Statistical analyses

We used non-parametric tests because they do not require the assumption of a normal population. The statistical test used was the Wilcoxon matched pairs signed rank test (see e.g., Blalock 1979). Results were obtained by SPSS statistics software (SPSS for Windows). The Wilcoxon matched pairs signed rank test was used to: (1) compare the identification rate between two conditions (e.g., stimulus 1 on the continuum vs. stimulus 2 on the continuum) in determining the location of the boundary shift, (2) compare the values of RT between two conditions (e.g., stimulus 1 on the continuum vs. stimulus 2 on the continuum) in determining whether subjects were significantly faster at identifying stimuli across categories relative to stimuli within categories, (3) compare the “different” response rate between two conditions (e.g., pair 1_2 in the continuum vs. pair 2_3 in the continuum) in each order of presentation in determining where the real

---

9 The use of different statistic methods along the dissertation (non-parametric and parametric methods) in cases in which the same assumptions are made about the data to be used captures the evolution made during the PhD period. However, in each chapter an explanation of the reasons for using a specific statistic test is provided.
discrimination peak was located in the discrimination results, (4) analyze the rate of “different” responses between AB (low-high) and BA (high-low) orders of presentation for each pair of stimuli (e.g., pair 1_2 vs. pair 2_1), (5) compare the discrimination rate predicted by the formulas taken from Liberman et al. (1957) and Godfrey et al. (1981) with the actual discrimination rate, and (6) compare the difference between hit rate and false alarm rate (d prime) between AB and BA orders of presentation for each pair of stimuli. The tables that report the results from the Wilcoxon matched pairs signed rank tests include the test statistic (denoted by the letter $T$ and the smallest of the two sums of ranks), its significance and the effect size. Since in most cases the Wilcoxon matched pairs signed rank statistic was used for multiple tests, the post-hoc Bonferroni correction was applied by adjusting the $p$ values.

3. Results

3.1. Identification results

Figure 4 shows the identification rate for the two continua created from the yes-no question (in black) and wh- question (in grey) base stimuli. The “identification rate” is defined as the number of “yes-no question” responses (in yes-no-question-based stimuli) or “wh- question” responses (in wh- question-based stimuli) over the total responses. As can be seen, the functions present the expected S-shape. The frequency range from 203 Hz to 218 Hz (i.e., stimuli 1-2) seems to correspond to the “wh-question” category and the 248-263 Hz range (i.e., stimuli 4-5) to the “yes-no question” category. Thus, the transition between the two categories would correspond to the 218-248 Hz range (i.e., stimuli 2-4). However, if we examine Figure 4 carefully, we see that the exact boundary between the two categories may be located at the specific range values between 218 Hz (stimulus 2) and 233 Hz (stimulus 3) because it is between these two stimuli where the biggest difference in identification rate is observed for both continua created. That the midpoint between stimuli 2 and 3 is the crossing point or boundary between the two categories is also apparent by the fact that it is in this point where the two categories are harder to identify, shown by the identification rate of 0.6.

It is worth noting that the responses to the two continua behave differently. This is particularly noteworthy in two respects. Firstly, in the yes-no-question-based continuum, stimulus 3 (233 Hz) triggered a high rate of yes-no question responses (0.76), whereas for the wh-question-based continuum the identification rate is around 0.5 for this stimulus. Secondly, we see that stimulus 5 (263 Hz) created from the wh-question base stimulus triggered a higher rate of identification as wh-question than would be expected. We should recall that, as noted above, the real boundary between

---

The effect size is calculated through the equation shown in (1) in which $z$ is the z-score that SPSS produces, and $N$ is the size of the study, that is, the number of total observations on which $z$ is based:

\[ r = \frac{Z}{\sqrt{N}} \]
the categories would be located between the range of 218 and 233 Hz. Thus, stimulus 1 (203 Hz) and stimulus 2 (218 Hz) would correspond to the “wh- question” category while stimulus 3 (233 Hz), stimulus 4 (248 Hz) and stimulus 5 (263 Hz) lie in principle within the “yes-no question” category. This indeed seems to be the case at least for the yes-no question based stimuli. However, for the wh- question-based continuum, it is not so clear that stimulus 3 (233 Hz) belongs to the “wh-question” category, since its identification rate as “wh-question” is higher than the identification rate as yes-no question of stimulus 3 in the yes-no-question-based continuum. Consequently, stimulus 3 receives identification rates that set it neither in the wh-question category neither in the yes-no-question category. That is, it receives identification rates that make it lie in between stimulus 2 (identified clearly as wh-question) and stimulus 4 (clearly yes-no-question, or non-wh-question). It is speculated that this difference in the location of the boundary in the wh-question-based continuum may be caused by the presence of the accented interrogative particle *qué* ‘what’, which would weaken the effect of peak height and would lead listeners to identify stimulus 3 more as “wh-question” than the identification rate as yes-no question of stimulus 3 in the yes-no-question-based continuum.

Likewise, the high identification rate as wh-question for stimulus 5 created from the wh-question base stimulus could be explained by the accented interrogative particle exerting a similar effect. This explanation is supported by the results of Vanrell (2006), in which the identification results were analyzed according to the subjects’ gender and musical background, since gender differences have been reported in production and perception experiments (Jensen & Carlin 1981, Johnson, Strand & D’Imperio 1999, Rogers 2003), as well as differences in the accuracy of perception depending on the degree of musical training (Schellenberg 2002, Cummins et al. 2006). The identification results for qualified musicians displayed a clear frontier region between the two categories, that is, between stimulus 2 (218 Hz) and stimulus 3 (233 Hz) in both continua, and there was no high identification rate as wh-question for either stimuli 3 or stimulus 5 in the wh-question-based continuum. In trying to provide an explanation it
was suspected that, because of their occupational training, these subjects were more attuned to changes in pitch height and during the experiment focused on tonal changes, while paying little attention to the presence of the accented interrogative particle.

Although there are clear differences between the respective responses to the two continua, it is important to emphasize that the boundary shift in identification is very robust: the functions are unmistakably S-shaped in spite of the effect on subjects’ perception of the accented interrogative particle.

Table III reports the standard error of the mean for every stimulus. Stimulus 1 and stimulus 5 for both base stimuli have lower standard errors. This means that for stimulus 1 and 5 subjects agreed in their responses because these stimuli represent the canonical categories. On the other hand, stimulus 3, which would be the crossover stimulus particularly in the wh-question-based continuum, has a higher standard error, which shows less agreement among subjects. However, for the yes-no-question-based continuum it is stimulus 2 that gets the highest standard error of all stimuli. Note that identification rates get higher standard errors as they get closer to the most ambiguous rate, 0.5 (shaded values). Stimulus 2 in yes-no-question-based continuum receives an identification rate of approximately 0.3, so it shows a higher standard error than stimulus 2 of wh-question-based continuum, which gets a rate of approximately 0.8-0.9. Stimulus 3 in yes-no gets a rate of almost 0.8, so its standard error is not as high as that of stimulus 2, or as high as that of stimulus 3 in wh-question-based continuum, which gets a rate between 0.4 and 0.5.

Table IV shows the results of four Wilcoxon matched pairs signed rank test comparing the identification rate between adjacent stimuli (e.g., stimulus 1 vs. stimulus 2, stimulus 2 vs. stimulus 3, and so on) for each continuum. According to the CP paradigm, there should be significant differences between the response rate for stimulus 3 and stimulus 2, as they belong to different categories in the case of yes-no-question-based continuum. However, in the case of the wh-question-based continuum, since stimulus 3 is the crossover stimulus, there should be significant differences between stimuli 2 and 3 and between stimuli 3 and 4. As can be seen in Table IV, the results are nearly what we would expect: stimulus 3 is significantly different from stimulus 2 in the yes-no-question-based continuum and significantly different from both stimulus 2 and stimulus 4 in the wh-question-based continuum (shaded values).\textsuperscript{11}

\[
\begin{array}{cccccc}
\text{Stimulus 1} & \text{Stimulus 2} & \text{Stimulus 3} & \text{Stimulus 4} & \text{Stimulus 5} \\
\hline
\text{Yes-no-question-based stimuli} & 0.023 & 0.040 & 0.038 & 0.030 & 0.023 \\
\text{Wh-question-based stimuli} & 0.015 & 0.032 & 0.044 & 0.036 & 0.032 \\
\end{array}
\]

**Table III**: Standard error of the mean of the identification rate for yes-no-question- and wh-question-based stimuli.

\textsuperscript{11} Note that we find significant differences between stimuli 1-2 in both continua, although the difference in identification rate between them is small and that they can be assumed to fall in one category, separate from stimulus 3.
Table IV: Results of four Wilcoxon matched pairs signed rank tests comparing the identification rate between adjacent stimuli for yes-no-question- and wh- question-based stimuli.

<table>
<thead>
<tr>
<th></th>
<th>Stimulus 1-2</th>
<th>Stimulus 2-3</th>
<th>Stimulus 3-4</th>
<th>Stimulus 4-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes-no-question-</td>
<td>$T = 34,$</td>
<td>$T = 234,$</td>
<td>$T = 142,$</td>
<td>$T = 92,$</td>
</tr>
<tr>
<td>based stimuli</td>
<td>$p &lt; .01,$</td>
<td>$p &lt; .01,$</td>
<td>$p &gt; .05,$</td>
<td>$p &gt; .05,$</td>
</tr>
<tr>
<td>$r = -.38,$</td>
<td>$r = -.20,$</td>
<td>$r = -.65,$</td>
<td>$r = -.38,$</td>
<td></td>
</tr>
<tr>
<td>Wh- question-</td>
<td>$T = 170,$</td>
<td>$T = 1767,$</td>
<td>$T = 1020,$</td>
<td>$T = 93.50,$</td>
</tr>
<tr>
<td>based stimuli</td>
<td>$p &lt; .01,$</td>
<td>$p &lt; .01,$</td>
<td>$p &gt; .05,$</td>
<td>$p &gt; .05,$</td>
</tr>
<tr>
<td>$r = -.30,$</td>
<td>$r = -.60,$</td>
<td>$r = -.37,$</td>
<td>$r = -.13,$</td>
<td></td>
</tr>
</tbody>
</table>

3.2. Reaction Time results

In order to claim the categoriality of a contrast, it is not enough that identification results have a clear category boundary, since this category boundary could be task-induced. For that reason, some researchers (Chen 2003, Falé & Faria 2006 for intonation) propose a Reaction Time approach to test the hypothetical discreteness of a contrast. According to Chen (2003), if the categories that result from the identification task are not task-induced, it is expected that the subjects will need approximately the same time to identify the stimuli that belong to the same category, while subjects will require more time to identify the stimuli that are in the crossover region between categories, the across-category stimuli. Thus, the within-category stimuli will be less demanding than the across-category stimuli in terms of cognitive load. Figure 5 plots the mean of RT measurements of the peak height in continua created from the yes-no question (black bars) and wh- question (grey bars) stimuli. For this specific experiment, Reaction Time measurements were taken from the onset of the stimulus to the subject’s reply. As in Chen (2003), the results show that listeners are faster at identifying stimuli within categories and slower at identifying stimuli across categories. Observe that while there is a peak in RT measurements corresponding to the stimulus 3 for the continuum created from the wh- question base stimulus, in the continuum created from the yes-no question base stimulus RT measurements of stimuli 2 and 3 are balanced, so we find a sort of plateau instead of a clear peak. This agrees with the results from the identification task, where it was found that the category boundary would be located between stimuli 2 and 3 in yes-no-question-based stimuli but specifically at stimulus 3 in wh- question-based stimuli. Consequently, stimuli 2 and 3 (from yes-no-question-based continuum) require the same time to be identified because they both flank the frontier region while stimulus 3 (from wh- question-based continuum) requires more time to be identified because it lies in the frontier region.

Table V shows the results of four Wilcoxon matched pairs signed rank test comparing mean RTs between adjacent stimuli for yes-no-question- and wh-question-based continua. As we can see, there are no differences between mean RTs of the stimuli 2 and 3 from the yes-no-question-based continuum because they both are located at the sides of the frontier region. By contrast, we find significant differences between stimuli 2 and 3 in the wh-question-based continuum, since stimulus 3 is the across-category stimulus (shaded values).
Table V: Results of four Wilcoxon matched pairs signed rank tests comparing mean RTs between adjacent stimuli for yes-no-question- and wh-question-based continua.

<table>
<thead>
<tr>
<th></th>
<th>RT 1-2</th>
<th>RT 2-3</th>
<th>RT 3-4</th>
<th>RT 4-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes-no-question-</td>
<td>T = 4932</td>
<td>T = 4617</td>
<td>T = 4514</td>
<td>T = 2287</td>
</tr>
<tr>
<td>based stimuli</td>
<td>p &lt; .01</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.16</td>
<td>r = -.11</td>
<td>r = -.11</td>
<td>r = -.38</td>
</tr>
<tr>
<td>Wh-question-</td>
<td>T = 5856</td>
<td>T = 5168</td>
<td>T = 3123</td>
<td>T = 4887</td>
</tr>
<tr>
<td>based stimuli</td>
<td>p &lt; .01</td>
<td>p &lt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.38</td>
<td>r = -.23</td>
<td>r = -.19</td>
<td>r = -.15</td>
</tr>
</tbody>
</table>

Figure 5: Mean RTs of the continua created from yes-no question and wh-question base stimuli.

3.3. Discrimination results

The materials for the discrimination task was made up of pairs of stimuli in AB and BA order (see Methodology section) and pairs in which two stimuli were identical. Figure 6 shows the rate of “different” responses (number of “different” responses over the total responses) to the various pairs of stimuli corresponding to the continuum created from the yes-no question base stimulus (peak at 263 Hz). AB hits are shown in light grey (“different” responses to dissimilar pairs where the second stimulus, stimulus B, had a pitch value 15 Hz higher than the first stimulus, stimulus B: 203 vs. 218 Hz, 218 vs. 233 Hz, 233 vs. 248 Hz, 248 vs. 263 Hz), BA hits in dark grey (“different” responses to dissimilar pairs where the second stimulus, stimulus A, has a pitch value 15 Hz lower than the first stimulus, stimulus B: 218 vs. 203 Hz, 233 vs. 218 Hz, 248 vs. 233 Hz, 263 vs. 248 Hz) and false alarms in black (“different” responses to identical pairs: 203 vs. 203 Hz, 218 vs. 218 Hz, 233 vs. 233 Hz, 248 vs. 248 Hz, 263 vs. 263 Hz). The actual “different” responses rates are indicated by the intermediate points between two adjacent stimuli. AB order should
be interpreted from left to right, whereas BA order should be interpreted from right to left.

We find two discrimination peaks at the pairs 218 vs. 233 Hz (AB pair) and 233 vs. 218 Hz (BA pair). The identification results suggested that this frequency range did indeed represent the crossover between the categories. The most striking feature of these results has to do with the BA hits. Note that the discrimination peak in AB pairs in which the second stimulus has a higher peak than the first has a higher rate of “different” responses. These results suggest that listeners have more trouble discriminating between pairs of stimuli presented in BA order (the pairs in which the second stimulus has a lower peak than the first).

![Yes-no-question-based stimuli](image)

**Figure 6:** Rate of “different” responses for pairs that were actually different (hits) and pairs that were identical (false alarms) corresponding to the continuum created from the yes-no question base stimulus. Error bars represent standard error of the mean.

Table VI shows the results of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the response rate between two different conditions: pair 1_2 vs. pair 2_3, pair 2_3 vs. pair 3_4, and so on. Table VII shows the results of four Wilcoxon matched pairs signed rank tests comparing the response rate between the conditions AB vs. BA for each pair of stimuli. From the results of table VI we can infer that the most important discrimination peak is in the AB function because pair 2_3 (218-233 Hz), where the discrimination peak is located, is significantly different from both pair 1_2 and pair 3_4 (shaded values). Although we observe differences in discrimination with regards to order of presentation (Figure 6), these differences are not statistically significant (Table VII).

Results of the discrimination task using stimuli created from the wh- question base stimulus (peak at 203 Hz) are plotted in Figure 7 as the rate of AB hits in light grey (203 vs. 218 Hz, 218 vs. 233 Hz, 233 vs. 248 Hz, 248 vs. 263 Hz), BA hits in dark grey (218 vs. 203 Hz, 233 vs. 218 Hz, 248 vs. 233 Hz, 263 vs. 248 Hz) and false alarms in black (203 vs. 203 Hz, 218 vs. 218 Hz, 233 vs. 233 Hz, 248 vs. 248 Hz, 263 vs. 263 Hz).
A major discrimination peak can be seen at AB pair 218 vs. 233 Hz (the rate of “different” responses reaches nearly 0.7), which agrees with the results of the identification task (we found the shift between categories around 233 Hz). Note that there is hardly any difference between the rate for BA hits and AA false alarms. These results confirm the findings shown in Figure 6, that it appears that subjects have trouble discriminating between stimuli when the direction of change in frequency is downwards.

<table>
<thead>
<tr>
<th>AB order</th>
<th>1_2 vs. 2_3</th>
<th>2_3 vs. 3_4</th>
<th>3_4 vs. 4_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes-no-question-based stimuli</td>
<td>$T = 105$</td>
<td>$T = 260$</td>
<td>$T = 216$</td>
</tr>
<tr>
<td></td>
<td>$p &lt; .01$</td>
<td>$p &lt; .01$</td>
<td>$p &lt; .01$</td>
</tr>
<tr>
<td></td>
<td>$r = -.47$</td>
<td>$r = -.37$</td>
<td>$r = -.33$</td>
</tr>
</tbody>
</table>

Table VI: Results of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the “different” response rate between adjacent pairs for yes-no-question-based stimuli.

<table>
<thead>
<tr>
<th>BA order</th>
<th>2_1 vs. 3_2</th>
<th>3_2 vs. 4_3</th>
<th>4_3 vs. 5_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes-no-question-based stimuli</td>
<td>$T = 34.50$</td>
<td>$T = 304$</td>
<td>$T = 225$</td>
</tr>
<tr>
<td></td>
<td>$p = .01$</td>
<td>$p &gt; .05$</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td></td>
<td>$r = -.42$</td>
<td>$r = -.15$</td>
<td>$r = -.30$</td>
</tr>
</tbody>
</table>

Table VII: Results of four Wilcoxon matched pairs signed rank tests comparing the “different” response rate between two conditions (AB order and BA order) for each pair of yes-no-question-based stimuli.

Figure 7: Rate of “different” responses for pairs that were actually different (hits) and pairs that were identical (false alarms) corresponding to the continuum created from the wh-question base stimulus. Error bars represent standard error of the mean.
Table VIII shows the results of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the response rate between two different conditions: pair 1_2 vs. pair 2_3, pair 2_3 vs. 3_4, and so on. Table IX shows the results of four Wilcoxon matched pairs signed rank tests comparing the response rate between conditions AB vs. BA for each pair of stimuli. The discrimination peak is located again at the 218 vs. 233 Hz interval in AB order. This can be seen in the results shown in Table VIII, where the differences between pair 2_3 and adjacent pairs in the AB order are significant (shaded values). By contrast, differences between pair 3_2 and adjacent pairs in BA order are not significant. Effects of order of presentation are confirmed from the results shown in Table IX, where there are significant differences between pairs 2_3 vs. 3_2 and 3_4 vs. 4_3 (shaded values).

<table>
<thead>
<tr>
<th>AB order</th>
<th>1_2 vs. 2_3</th>
<th>2_3 vs. 3_4</th>
<th>3_4 vs. 4_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh-question-based continuum</td>
<td>$T = 540$</td>
<td>$T = 180$</td>
<td>$T = 87$</td>
</tr>
<tr>
<td>$p &lt; .01$</td>
<td>$p &lt; .01$</td>
<td>$p &lt; .01$</td>
<td></td>
</tr>
<tr>
<td>$r = -.52$</td>
<td>$r = -.31$</td>
<td>$r = -.37$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BA order</th>
<th>2_1 vs. 3_2</th>
<th>3_2 vs. 4_3</th>
<th>4_3 vs. 5_4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh-question-based continuum</td>
<td>$T = 294.5$</td>
<td>$T = 135$</td>
<td>$T = 99$</td>
</tr>
<tr>
<td>$p &gt; .05$</td>
<td>$p &gt; .05$</td>
<td>$p &gt; .05$</td>
<td></td>
</tr>
<tr>
<td>$r = -.18$</td>
<td>$r = -.25$</td>
<td>$r = -.08$</td>
<td></td>
</tr>
</tbody>
</table>

Table VIII: Results of three Wilcoxon matched pairs signed rank tests for each order of presentation comparing the “different” response rate between adjacent pairs for wh-question-based stimuli.

This difference with respect to the significance of the effects of order of presentation could be related to the nature of the original stimuli. When listeners hear the accented interrogative particle in the wh-question-based stimuli, subjects expect a high tone on the syllable preceding the one with the nuclear accent. In the order AB, in the second token of the pair the pitch in this pretonic syllable is even higher in frequency than expected, so this difference is very noticeable. By contrast, after the unaccented interrogative particle in the yes-no-question-based stimuli, listeners would expect the super-high variant of the high tone in the syllable preceding the nuclear accent, but only the second token has a pitch level on that syllable that could be close to what is expected. Perhaps this difference would not be so noticeable because the pitch height of the prenuclear syllable of the first token is contradictory with the absence of accent in the interrogative particle. This could explain not only why we get significant effects of order of presentation only for the wh-question-based continuum, but also why we get
generally more sharply differentiated discrimination results in the wh-question-based stimuli.

3.4. Models of discrimination performance

3.4.1. Relating discrimination responses to identification responses

Identification and discrimination results in this study fulfill one of the criteria which according to Francis et al. (2003) should obey in order to consider that the contrast tested is categorical, namely, the peaks of discrimination correspond to the location of the category boundary as determined by identification. The other criterion is that identification results should predict discrimination accuracy (Liberman et al. 1957). It can be said, hence, that the extent to which discrimination performance can be predicted from classification is what is referred to as categorical perception. Thus, in order to determine whether discrimination performance can be predicted by identification results, two formulas for predicting discrimination were applied. The first formula was taken from Liberman et al. (1957), who used it to predict the results of an ABX discrimination task. However, Pollock and Pisoni (1971) showed that the same equation can also be used to predict performance in a same/different discrimination task. The equation is:

\[ P(\text{disc}_{12}) = 0.5[1 + (p_1 - p_2)^2] \]

\( p_1 \) is the probability of identifying Stimulus 1 as category A and \( p_2 \) is the probability of identifying Stimulus 2 as category A. This formula assumes that when listeners do not hear a difference they respond “same” or “different” randomly, so that performance is by chance. As pointed out by Macmillan, Kaplan, Howard and Creelman (1977): “If the resulting classification led to a decision (i.e., if A and B were classified differently and X as one of them—in an ABX discrimination task), the observer would respond as indicated; if it did not lead to a decision, he would guess, choosing each response with probability 0.5”.

The second formula is taken from Godfrey, Syrdal-Laskym, Millay and Knox (1981) and is a more general formula that predicts discrimination on the basis of phonetic categorization without guessing probabilities. This formula was also used by other studies of children’s categorical perception such as Wolf (1973) and Brandt and Rosen (1980) that employed same/different discrimination tasks:

\[ \text{Proportion discriminated} = (P_{1a} \times P_{2b}) + (P_{1b} \times P_{2a}) \]

\( P_{1a} \) = proportion of time that stimulus 1 was identified as “a”,
\( P_{2b} \) = proportion of time that stimulus 2 was identified as “b”,
\( P_{1b} \) = proportion of time that stimulus 1 was identified as “b”,
\( P_{2a} \) = proportion of time that stimulus 2 was identified as “a”.
Figures 8 and 9 show the obtained (black) and predicted (grey) discrimination functions as a result of the application of formula (1). The rate of correct discrimination for each pair was calculated as in Francis and Ciocca (2003), that is, as the average of the rate of “different” responses for different pairs and the rate of “same” responses for same pairs. For example, the rate of “correct” responses for the 1-2 pair was the average of the rate of “different” responses for the 1-2 and 2-1 pairs and the rate of “same” responses for the 1-1 and 2-2 pairs.

![Yes-no-question-based stimuli](image1)

**Figure 8:** Predicted and obtained discrimination ratios for the continuum created from the yes-no question base stimulus. Error bars represent the standard error of the mean.

![Wh-question-based stimuli](image2)

**Figure 9:** Predicted and obtained discrimination ratios for the continuum created from the wh-question base stimulus. Error bars represent the standard error of the mean.
Four Wilcoxon matched pairs signed rank tests were carried out for each continuum in order to compare the obtained and predicted discrimination rates (Table X) for each pair of stimuli. In order to claim that the formula accurately predicts the discrimination function, predicted and obtained discrimination functions cannot be significantly different. The results show that the difference between predicted versus obtained discrimination is significant for pair 1_2 and 2_3 with stimuli created from the yes-no question base stimulus and for pair 3_4 with stimuli created from the wh-question base (shaded values). Thus, it was concluded that discrimination results cannot be predicted accurately from the identification results if we assume that listeners respond randomly when they do not hear a difference.

<table>
<thead>
<tr>
<th></th>
<th>Obtained 1_2 vs. predicted</th>
<th>Obtained 2_3 vs. predicted</th>
<th>Obtained 3_4 vs. predicted</th>
<th>Obtained 4_5 vs. predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes-no-question-based stimuli</td>
<td>$T = 391$</td>
<td>$T = 884.5$</td>
<td>$T = 489$</td>
<td>$T = 267$</td>
</tr>
<tr>
<td>$p &lt; .05$</td>
<td>$p &lt; .05$</td>
<td>$p &gt; .05$</td>
<td>$p &gt; .05$</td>
<td></td>
</tr>
<tr>
<td>$r = -.36$</td>
<td>$r = -.39$</td>
<td>$r = -.22$</td>
<td>$r = -.19$</td>
<td></td>
</tr>
<tr>
<td>Wh-question-based stimuli</td>
<td>$T = 199$</td>
<td>$T = 650.5$</td>
<td>$T = 871.5$</td>
<td>$T = 211$</td>
</tr>
<tr>
<td>$p &gt; .05$</td>
<td>$p &gt; .05$</td>
<td>$p &gt; .05$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r = -.03$</td>
<td>$r = -.12$</td>
<td>$r = -.37$</td>
<td>$r = -.17$</td>
<td></td>
</tr>
</tbody>
</table>

**Table X:** Results of four Wilcoxon matched pairs signed rank tests for each continuum comparing the obtained and predicted discrimination rate for each pair of stimuli.

Figures 10 and 11 show the obtained (in black) and predicted (in grey) discrimination functions as a result of the application of formula (2). The real discrimination values have been calculated as the rate of different pairs which were correctly called “different”, as in Godfrey et al. (1981), who use this formula with the same kind of task as in this study.

![Yes-no-question-based stimuli](image)

**Figure 10:** Predicted and obtained discrimination ratios for the continuum created from the yes-no question base stimulus Error bars represent the standard error of the mean.
Figure 11: Predicted and obtained discrimination ratios of the continuum created from the wh-question base stimulus. Error bars represent the standard error of the mean.

The results of four Wilcoxon matched pairs signed rank tests (Table XI) for each continuum show that the differences between predicted versus obtained discrimination is significant only for pair 1 with stimuli created from the wh-question base stimulus. This means that formula (2) is suitable for our data and that discrimination data can be predicted from identification data only on the basis of phonetic categorization, without making assumptions about guessing.

<table>
<thead>
<tr>
<th>Obtained vs. predicted</th>
<th>Obtained vs. predicted</th>
<th>Obtained vs. predicted</th>
<th>Obtained vs. predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_2</td>
<td>2_3</td>
<td>3_4</td>
<td>4_5</td>
</tr>
<tr>
<td>Yes-no-question-based stimuli</td>
<td>Yes-no-question-based stimuli</td>
<td>Yes-no-question-based stimuli</td>
<td>Yes-no-question-based stimuli</td>
</tr>
<tr>
<td>(T = 648)</td>
<td>(T = 945.5)</td>
<td>(T = 884.5)</td>
<td>(T = 648)</td>
</tr>
<tr>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
</tr>
<tr>
<td>(r = -.12)</td>
<td>(r = -.03)</td>
<td>(r = -.03)</td>
<td>(r = -.06)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wh-question-based stimuli</th>
<th>Wh-question-based stimuli</th>
<th>Wh-question-based stimuli</th>
<th>Wh-question-based stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T = 214.5)</td>
<td>(T = 725)</td>
<td>(T = 1170)</td>
<td>(T = 221)</td>
</tr>
<tr>
<td>(p &lt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
</tr>
<tr>
<td>(r = -.12)</td>
<td>(r = -.12)</td>
<td>(r = -.12)</td>
<td>(r = -.15)</td>
</tr>
</tbody>
</table>

Table XI: Results of four Wilcoxon matched pairs signed rank tests for each continuum comparing the obtained and predicted discrimination rate for each pair of stimuli.

3.4.2. Signal Detection Theory

The model of discrimination performance described above by formula (1) presupposes that when listeners do not hear a difference or do not know how to respond, they
respond “same” or “different” at random. But it is by no means certain that listeners act in this way. According to Keating (2004), some subjects may tend to give a “different” response most of the time while, on the other hand, other subjects may be very conservative and only give a “different” response when they are completely sure that they hear a difference. This means that in the former case the results for same pairs are not reliable and in the latter case the same will be true for different pairs. The point is that the percentage of correct discriminations between different pairs is highly susceptible to subjects who tend to give only “different” (or “same”) responses all the time, and it should be interpreted in terms of the listener’s response bias, that is, his or her tendency to qualify stimuli pairs as “same” or “different”.

Signal Detection Theory attributes responses to a combination of sensitivity and bias. Sensitivity is the variable that is being investigated and bias is what we must take into account so that the sensitivity measure is meaningful. The statistical expression $d'$ (d prime) is a measure of the difference between the hit rate (proportion of different pairs to which subjects responded “different”) and false alarm rate (proportion of same pairs to which subjects responded “different”). However, $d'$ is not just the difference between the hit rate and the false alarm rate, rather, it is defined in terms of $z$, the inverse of the normal distribution functions, as shown in (3):

$$d' = z(H) - z(F)$$

$d'$ has been used in the discrimination literature for obtained and predicted discrimination by Best, Morrongiello and Robson (1981) and in addition to percentage of correct / percentage of different responses to different pairs by Francis and Ciocca (2003). In order to validate the results plotted in Figures 6 and 7 related to the presence of order of presentation effects, Signal Detection Theory was applied to our data. Figures 12 and 13 show the discrimination results presented as $d'$ for each stimulus pair in low-high order (black lines) and high-low order (grey lines). $d'$ scores were calculated on the basis of “different” responses to the pairs that were truly different (hits) and “different” responses to the pairs that were actually the same (false alarms). Following Macmillan and Creelman (1991), $d'$ was calculated using roving methods\(^\text{12}\) (using Table A5.4, pp. 338-354). This table was generated by varying the response threshold (the value of $k$) in equations shown in (4):

$$\begin{align*}
H &= \left[\frac{(-k + d')}{\sqrt{2}}\right] + \left[\frac{(-k - d')}{\sqrt{2}}\right] \\
F &= 2\phi(-k\sqrt{2})
\end{align*}$$

\(^{12}\) According to Macmillan and Creelman (1991), there are two different strategies used by subjects in responding to “same-different” experiments and therefore different appropriate models for $d'$ depending on how stimuli are across trials in a block. In roving designs, the two stimuli vary from trial to trial and “subjects are likely to apply a differing strategy, applying a threshold of difference to decide if 2 stimuli are different enough to count as ‘different’” (Keating 2004).
Figures 12 and 13 show that Majorcan listeners are more sensitive to F0 differences in speech stimuli when the second stimulus in a pair has higher F0. Observe that this difference in sensitivity is especially important in the pair that represents the boundary between the categories and which corresponds to the discrimination peak, pair 2_3.

![Yes-no-question-based stimuli](image1)

**Figure 12:** $d'$ calculated for both orders of presentation of pairs from the continuum created from the yes-no question base stimulus.

![Wh-question-based stimuli](image2)

**Figure 13:** $d'$ calculated for both orders of presentation of pairs from the continuum created from the wh-question base stimulus.

Four Wilcoxon matched pairs signed rank tests (Table XII) were carried out on hit rate minus false alarm rate for each pair with the conditions AB and BA orders of presentation. Significant differences were found only for the second and third pairs for
stimuli created from the wh- question base stimulus (shaded values). For stimuli created from the yes-no question base stimulus, the differences in order of presentation were significant for none of the pairs. Thus, the application of d’ in addition to the rate of “different” responses to different/same pairs has confirmed the presence of order of presentation effects which are statistically significant only for pair 2_3 vs. pair 3_2 and pair 3_4 vs. pair 4_3 for stimuli created from the wh- question base stimulus.

<table>
<thead>
<tr>
<th></th>
<th>Pair 1_2 vs.</th>
<th>Pair 2_3 vs.</th>
<th>Pair 3_4 vs.</th>
<th>Pair 4_5 vs.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pair 2_1</td>
<td>pair 3_2</td>
<td>pair 4_3</td>
<td>pair 5_4</td>
</tr>
<tr>
<td>Yes-no-question-based stimuli</td>
<td>T = 70</td>
<td>T = 152</td>
<td>T = 103</td>
<td>T = 80</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.21</td>
<td>r = -.33</td>
<td>r = -.14</td>
<td>r = -.12</td>
</tr>
<tr>
<td>Wh-question-based stimuli</td>
<td>T = 51</td>
<td>T = 18</td>
<td>T = 23.5</td>
<td>T = 42.5</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.01</td>
<td>r = -.40</td>
<td>r = -.48</td>
<td>r = -.04</td>
</tr>
</tbody>
</table>

Table XII: Results of four Wilcoxon matched pairs signed rank tests comparing the difference between hit rate and false alarm rate for each pair with the conditions AB and BA orders of presentation.

4. Discussion and conclusions

The present study has provided evidence that Majorcan Catalan listeners make categorical linguistic use of F0 scaling differences in perceiving yes-no questions as opposed to wh- questions in Majorcan Catalan. This evidence comes from different sets of results. The identification results show that it is possible to switch the perceived category by manipulating the pitch height of the leading tone from an H tone to a super-high tone and vice versa. We observed in Figure 4 that the presence/absence of the accented wh- particle in the two continua does not interfere in the categorical perception of this contrast, that is, the identification functions appear unmistakably S-shaped with an identification rate that goes from 0.85 to about 0.2 (in the case of the continuum created from the wh- question base stimulus) and from less than 0.27 to about 0.87 (in the case of the continuum created from the yes-no question base stimulus) within 2 steps of the 5-step continuum.

Evidence of this linguistic contrast is also provided by RT measurements. A mean RT peak/plateau can be observed in Figure 5 at the identification boundary for both continua; hence, the mean RTs were shorter for within categories and longer for across categories. According to Chen (2003), these are essential properties of linguistically real categories and that are not task-induced.

Moreover, by comparing through statistical analyses the magnitude of difference in the identification rate for adjacent stimuli in the identification task, it has been observed that, although there is a shift from one response to the other in the range of values between 218 and 248 Hz, the threshold between the “yes-no question” and the “wh-question” category would be located between 218 and 233 Hz. This is actually the pair of stimulus that is best discriminated and, hence, where the discrimination peak is
situated. However, for classic CP it is not enough that an abrupt shift in the identification function is observed and that this shift in the identification function corresponds to the discrimination peak; rather, discrimination results should also be predicted from identification results. As it turns out, the application of the Haskins formula (the formula taken from Liberman et al. 1957) and the more general formula taken from Godfrey et al. (1981) do indeed permit us to predict the shape of the discrimination curve. From Figures 8, 9, 10 and 11 and from statistical analyses, it can be said that the formula used by Godfrey et al. (1981) based simply on phonetic categorization, without assumptions about guessing, is the one that best fits our data. Because the results of the Wilcoxon matched pairs signed rank tests show that the differences between predicted and obtained discrimination are not significant, it is concluded that this contrast is categorically perceived.

On the basis of this evidence, we claim that the difference in pitch height of the leading tone of the nuclear accent is a strong perceptual cue that Majorcan listeners use when distinguishing yes-no questions from wh- questions. This does not mean that pitch height is the only cue; on the contrary, identification results seem to suggest that there is a supplementary cue, the accented interrogative particle. The effect of the accented interrogative particle is especially noticeable in the wh- question-based continuum, in which it appears to delay the switch from one category to the other. This effect can be observed not only in our identification results but also in the results of RT measurements, in which the difference in the crossover boundary between the two kinds of stimuli can be seen in the presence of a mean RT plateau in the yes-no-question-based continuum on the one hand and a mean RT peak in the wh- question-based continuum on the other. Thus, it would be of great interest to know how listeners would respond to the continua if the effect of accentuation was neutralized. It is expected that we would obtain results similar to the identification performance of qualified musicians. In any case, this indicates the need for further research to confirm that the effect of the interrogative particle is exactly the effect suggested by these results (see Chapter 2).

Finally, it is worth trying to account for the discrimination asymmetries that our results report, namely that it is easier to perceive differences between two stimuli when the second one has a higher pitch than the first (our results agree with previous studies: Kohler 1987, Ladd & Morton 1997, Remijsen & van Heuven 1999, Schneider & Linfert 2003, Cummins et al. 2006, Falé & Faria 2006, Prieto et al. 2010). These asymmetries have been related in the literature to the F0 declination or downdrift, the gradual declination of fundamental frequency over the course of an utterance (Pierrehumbert 1979, Gussenhoven & Rietveld 1988). F0 declination has been argued to be a universal characteristic of speech production. Evidence for compensation of this declination effect has been provided for American English listeners (Pierrehumbert 1979), Dutch listeners (Gussenhoven & Rietveld 1988) and Cantonese listeners (Wong 1999). According to Francis & Ciocca (2003), these asymmetries may be explained in terms of a compensation for an expected declination in F0 over the course of an utterance. Thus, listeners are able to compensate for this decline by taking into account the position of the accent within the utterance so that the meaning conveyed by the speaker is correctly identified. Given two tokens, when the second token has lower pitch
than the first, this compensation would ensure that the two tokens sound identical; by contrast, when the second token has a higher pitch than the first, this raising in pitch of the second token would enhance the perception of the difference between the two tokens. Notice that for our case a putative effect of declination would have to be interpreted as applying across utterances, rather than within utterances. However, further research is necessary to test whether declination exists in Majorcan Catalan or whether Majorcan Catalan listeners compensate perceptually for this expected declination.

To sum up, the results of this perception study confirm the categorical perception of the difference in pitch height between (downstepped) H and upstepped H within the H+L* nuclear accent and consequently the phonological role of scaling of the H leading tone, since it permits listeners to distinguish yes-no questions from wh- questions in Majorcan Catalan. Our results also indicate that there are discrimination asymmetries that depend on the direction of change, it being easier to distinguish between the stimuli pair when the direction of change is upwards.
CHAPTER 2

The role of prenuclear accents on interrogative recognition

In the previous chapter we saw that in Majorcan Catalan differences in the scaling of the H leading tone of the falling nuclear pitch accent H+L* are used to perceptually distinguish yes-no questions from wh- questions. Yet the role of the pitch accentual properties of the homophonous interrogative marker or wh- word [kə] que/què has not yet been investigated. In the present chapter, identification tasks and the gating paradigm will be used to investigate whether Catalan listeners use the pitch accent cues in the prenuclear part of the contour in online recognition of yes-no questions and wh-questions.

1. Introduction

Intonation contours have traditionally been broken down into at least two parts, namely, the “nucleus” or “nuclear configuration” (i.e., the intonation region which carries maximal prominence, and which generally extends from the utterance’s last pitch-accented syllable until the end of the intonation phrase) and the “prenuclear” region (i.e., the stretch of the pitch contour preceding the nuclear pitch accent). Autosegmental-Metrical (AM) studies on intonational meaning have implicitly tended to regard the nucleus as a crucial part of the pitch contour (see, for example, Pierrehumbert & Hirschberg 1990). Yet while the contribution of nuclear pitch contours to tune meaning has been the object of extensive study, the contribution of the prenuclear contour and/or prenuclear pitch accents has received somewhat less attention.

In recent years, a variety of studies from different languages have investigated the role of the prenuclear part of intonation contours in the recognition of sentence-type (mostly interrogative utterances). They have found that small differences in F0, alignment and scaling as well as the presence of distinct pitch accents and boundary tones in the prenuclear region help listeners in indentifying yes-no questions in early portions of the sentence with a fairly high degree of confidence (Hadding-Koch & Studdert-Kennedy 1964; Vion & Colas 2006; Face 2005; Petrone 2008, 2010; Petrone & D’Imperio 2010; Crespo-Sendra 2011; and others).

In Hadding-Koch and Studdert-Kennedy’s (1964) classic study, the intonation contour of the utterance För Jane (‘for Jane’) was systematically varied at three different locations (peak, turning point and end point) and presented for judgement to Swedish and American subjects, who were asked to classify the contours as (1) questions or statements and (2) having terminal rises or falls. The results showed that listeners used not only terminal rises and falls but also the pitch height of the preceding peak and its turning point to distinguish questions from statements.
In Vion and Colas (2006), the gating paradigm was used to determine whether subjects listening to French NP utterances containing three stress groups could accurately predict whether the utterance was a statement or question before listening to the final rise. Listeners were able to recognize with mid-level confidence ratings (between 50%-75% of responses were accurate) the intonation of a question when they listened to the downstepped tones preceding the final rise but not the final rise itself. In addition, the authors reported the presence of a bias in the responses in favour of the answer ‘statement’ rather than ‘question’. According to the authors, this tendency could be due to the different frequency of the two types of sentences in everyday language use.

Castilian Spanish yes-no questions have also been traditionally reported to have higher initial peaks in yes-no questions in comparison with statements (Navarro Tomás 1944, Prieto 2004, Sosa 1999). To examine the potential effect of this prenuclear scaling difference between statements and questions as well as the effect of the other three primary intonational differences located at different parts of the sentence, Face (2005) performed a gating task and a full utterance perception task in which the F0 patterns were manipulated to contain all possible combinations of cues to sentence type. For the gating task, sixteen subjects had to decide whether they perceived each gate as having a declarative or an interrogative interpretation. For complete utterances, subjects had to identify the sentence type and also rate the naturalness of the production. The results for the first task showed that after the presentation of the first cue to sentence type, that is, the tonal scaling difference related to the initial F0 peak, listeners were highly accurate (87.5% accurate identifications) in perceiving sentence type. With respect to the full utterance perception task, the results indicated that even when the final F0 movement was presented to listeners, the earlier cues had a significant effect on the listeners’ rating of the naturalness of the stimuli.

More recently, Petrone (2008) and Petrone and D’Imperio (2010) investigated the potential role of prenuclear contours in the perceptual distinction between narrow focus statements and yes-no questions in Neapolitan Italian, testing the potential relevance of a tone which appears to be inserted at the right edge of the Accentual Phrase in the prenuclear contour (H in questions and L in statements). Nine Neapolitan listeners responded to a forced-choice identification task with gated stimuli. The results showed that the presence of prenuclear accent information was indeed important: the mean ‘question’ score for question base stimuli was above chance (67%), while it was around 37% for statement base stimuli. Moreover, when the prenuclear edge tone was present in the gate, question scores decreased for statement base stimuli (20%), suggesting that the presence of this tone plays an important role in sentence type identification.

Petrone (2010) investigated the relevance of the prenuclear region in distinguishing statements and yes-no questions in Cosenza Italian. To this end, Petrone carried out a forced-choice identification task with the gated stimulus *Marinella vuole* (‘Marinella wants’) to prevent listeners from exploiting the nuclear information in distinguishing.

13 For example, Navarro Tomás pointed out that “in the first accented syllable [of the interrogative utterance] pitch is increased by three or four semitones over the average level” (Navarro Tomás 1944: 141; our translation).
statements from questions. Then, the following manipulations were carried out on the stimulus: rise-fall timing, L1 timing, H timing, rise-fall scaling, L2 height, rise-fall steepness and rise shape. Fourteen Cosenzan listeners participated in the experiment. After listening to each fragment, participants had to decide whether the speaker was going to formulate a question or a statement. The results demonstrated that Cosenza Italian listeners were able to identify the contrast between yes-no questions and statements by exploiting phonetic differences in the implementation of rise-fall prenuclear contour; specifically, the early rise-fall alignment was exploited by Cosenza listeners to accurately identify utterances as statements 52% of the time.

Finally, Crespo-Sendra (2011) investigated whether Central and Valencian Catalan listeners relied differently on prenuclear and nuclear cues to identify the contrast between information-seeking questions and incredulity questions by means of several perception experiments, namely, a congruity task, a gating identification task and a set of identification tasks. The results confirmed that there was indeed a contrast between how Central Catalan listeners perceptually processed these two different sentence types compared to Valencian Catalan listeners. While a difference in pitch scaling of the boundary tone was the strongest cue for perceptually distinguishing between the two interpretations in Central Catalan, Valencian Catalan listeners relied more on prenuclear pitch range differences. For Central Catalan, the results revealed an average accurate identification rate of 89% and 97% —meaning that stimuli were correctly recognized as fragments of their respective original utterances—for the third and fourth gates respectively, whereas the mean rate of accurate identification for the first and second gates was 51% and 49% respectively. By contrast, in Valencian Catalan, the results revealed an average accurate identification rate of 91%, 92% and 93% for the second, third and fourth gates respectively (the average rate for the first gate was 52%).

From the results presented above it is clear that the importance of the prenuclear part of the pitch contour for the perception of interrogative meaning varies according to the study since reported identification rates range from chance level (Petrone 2010) to well above chance level (Petrone 2008, Petrone & D’Imperio 2010) and even to a highly accurate level (Face 2005, Crespo-Sendra 2011). Yet interestingly, all these studies share a common finding, namely, that we should acknowledge to a small or great extent the contribution of the prenuclear part of the pitch contour to the perception of intonational meaning.

The main goal of the experiments is to contribute to the investigation of the role of prenuclear accentual/tonal cues to interrogative utterance interpretation in a language that seems to rely almost exclusively on the nuclear part of the contour to convey the interrogative interpretation. In this specific case, we would like to test whether reliance on the prenuclear part of the contour becomes greater when the nuclear part of the contour is more ambiguous. If this hypothesis is confirmed, we would have a clear indication that the processing of intonational contours is an active process of temporal
integration\textsuperscript{14} (Nguyen & Hawkins 2003) that takes into account all components of the target intonation contour (i.e., both prenuclear and nuclear).

In the previous chapter, the Categorical Perception paradigm was applied to a pitch accent contrast (two types of falling nuclear pitch accents), which distinguishes yes-no from wh- questions in Majorcan Catalan (the contrast is illustrated in Figure 1, previous chapter). Forty-two Majorcan Catalan listeners participated in this two-part experiment consisting of an identification task and a discrimination task. Two pitch scaling continua of the nuclear leading tone were created from two natural tokens of the sentences *Qué li duries?* ‘What would you bring to her/him?’ vs. *Que l’hi duries?* ‘Would you bring it to her/him?’. The results from the identification task showed that it was possible to switch the category perceived by the listener simply by manipulating the pitch height of the leading tone of the falling nuclear accent. Discrimination results revealed that the shift in the identification function corresponded to the peak in the discrimination functions. A comparison of obtained and predicted discrimination results indicated that the location of the discrimination peak could be predicted from identification results. These results confirmed that the difference in pitch height of the leading tone in nuclear accent for yes-no and wh- questions in Majorcan Catalan is discrete and has a phonological character.

Although the scaling difference of the leading tone of the falling nuclear accent was shown to be the perceptual cue that Majorcan listeners used to distinguish yes-no questions from wh- questions, identification results also revealed a small effect of the base stimulus.\textsuperscript{15} This seemed to suggest that the acoustic information located at the prenuclear region might play a role in the identification task. In Majorcan Catalan, yes-no questions can be headed by the unaccented interrogative particle *que* [kə] ‘that’ and wh- questions are headed by the accented wh- word *qué* [ˈkə] ‘what’ words which are homophonous at the segmental level.

Chapter 2 will be organized as follows. Following this introduction, which contains a review of the literature on the importance of the prenuclear part in the recognition of sentence-type and states the purpose of this chapter, section 2 describes the methodology and results of experiment 1 (the identification task), and section 3 presents the methodology and results of experiment 2 (the gating task). Finally, section 4 discusses the main findings of the study and states its overall conclusions.

\textsuperscript{14} Nguyen and Hawkins (2003) refer to temporal integration as “how chunks of information arriving at the ears at different times are linked together by the listener in mapping speech sounds onto meaning”.

\textsuperscript{15} While in the yes-no question-based continuum the boundary between the two different categories was located between stimuli 2 and 3, in the wh- question-based continuum the boundary between the categories coincided with stimulus 3. In other words, the results suggested that the accented interrogative particle delayed the switch from one category to the other, which would also weaken the effect of peak height. Likewise, we explained the high identification rate as wh- question for stimulus 5 (which we expected would be identified as a yes-no question) by the accented interrogative particle exerting a similar effect.
2. Experiment 1: Identification task

The aim of this experiment was to test whether Majorcan Catalan listeners are sensitive to the subtle difference in prenuclear cues between a yes-no question and a wh-question, namely, the presence of an accented \textit{qué} [’kə] ‘what’ vs. an unaccented \textit{que} [kə] ‘that’. The methodological innovation of this work lies in our use of combinations of prenuclear cues with an acoustic continuum of the typical nuclear pitch accents in each sentence type. This allowed us to test whether prenuclear cues would become more relevant when the nuclear information was ambiguous. Our expectation was that the presence of prenuclear cues to accentuation (pitch, duration and F1 differences) would help listeners to decide which type of question they were listening to, especially when the cues located at the nuclear region were neutralized. In this first experiment, an identification task was performed in which the cues related to the nuclear region and those related to the prenuclear region were manipulated. Participants were presented with combinations of congruent and incongruent prenuclear/nuclear components, in order to be able to test whether the prenuclear cues would become more important when a category changed. In addition, we sought to determine which combinations of cues would cause listeners to be faster in categorizing the stimuli.

2.1. Methodology

2.1.1. Participants

Twenty-one subjects (thirteen women, eight men) aged between 18 and 40 (median = 32) participated in the experiment. All participants were native speakers of Majorcan Catalan, coming from different areas of the island. None had any specific training in linguistics or reported a history of hearing disability. In order for their input to qualify for inclusion in the study, subjects had to achieve a pre-established level of identification accuracy whereby 80% of the base stimuli in the training block (which consisted of 10 original trials, 5 for each base stimulus) had to be correctly identified. Five listeners failed to achieve this and their input thus had to be rejected. In the end, the data from 16 subjects were analyzed. Of these sixteen subjects, nine were women and seven were men.

2.1.2. Materials

Two target stimuli were created from a set of tokens of the sentence \textit{Qué l’hi duries?} (‘Would you bring it to her/him?’) and \textit{Què li duries?} (‘What would you bring to her/him?’) produced by the author of this dissertation, who is a native speaker of Majorcan Catalan. They were recorded in a quiet room at the Universitat Autònoma de Barcelona using a TASCAM TEAC Professional Division digital audio tape recorder and a SHURE PG81 microphone. The recorded utterances were digitized at a 44100 Hz sample rate using Goldwave software. The two tokens were homophonic at the
segmental level. The selection of the base tokens from among several samples and their manipulations were informed by a previous systematic production study reported in Vanrell (2006). In the two utterance types under investigation here, the respective vowels of the unstressed que ‘that’ and the stressed què ‘what’ interrogative particles display significant differences in pitch and vowel quality and optional differences in duration. Specifically, the stressed/accented vowel in the interrogative particle has a higher mean pitch, an increase in F1 value (but not in F2 value) and longer vowel duration compared to the unstressed/unaccented particle. In the recorded base tokens finally selected, the mean values of the stressed/accented vowels and unstressed/unaccented vowels respectively were as follows: average pitch at the centre of the vowel was 227/212 Hz, average F1 at the centre of the vowel was 620/526 Hz and average duration of the vowel was 54/43 ms. Following the procedure used in Vanrell (2006), the interrogative particle of the two recorded base tokens was then manipulated in terms of pitch, duration and first formant values so that they would reproduce the mean values found in production. The pitch manipulation was done by lowering the F0 of the què (‘what’) base token by 15 Hz and raising the F0 of the que (‘that’) base token by the same amount (15 Hz); the F1 of the què (‘what’) token was lowered by 94 Hz while the F1 of the que (‘that’) token was raised by 94 Hz; and, finally, for the duration manipulation we subtracted 20 ms from the què (‘what’) token and added 20 ms from the que (‘that’) token. The Computer Speech Lab (CSL) 4300 (Kay Elemetrics Corp., Lincoln Park, NJ, USA) program was used for the manipulation of F1 and manipulation of duration and pitch was performed using the Praat (Boersma & Weenink 2009) program. Table I shows a summary of the manipulations in pitch (F0), F1 and duration carried out on the two original recorded tokens.

<table>
<thead>
<tr>
<th>Original recorded token</th>
<th>Yes-no question-base stimulus</th>
<th>Wh-question-base stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>40 ms</td>
<td>53 ms</td>
</tr>
<tr>
<td>F0</td>
<td>213 Hz</td>
<td>225 Hz</td>
</tr>
<tr>
<td>F1</td>
<td>526 Hz</td>
<td>616 Hz</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Manipulated token</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>60 ms</td>
<td>33 ms</td>
</tr>
<tr>
<td>F0</td>
<td>228 Hz</td>
<td>209 Hz</td>
</tr>
<tr>
<td>F1</td>
<td>620 Hz</td>
<td>522 Hz</td>
</tr>
</tbody>
</table>

Table I: Summary of manipulations in pitch, F1 and duration made in the two original tokens.

The goal of this manipulation was to obtain two tokens for the yes-no question: one with a congruently unaccented ‘that’ particle and the other with an incongruently accented ‘that’ particle. Likewise we ended up with two tokens for the wh- question: one ‘what’ that was congruently accented and the other which was incongruently unaccented.

16 Thanks so much to the Laboratori de Fonètica Experimental (Universitat de Barcelona) and to Marianna Nadeu for help with the manipulation of F1 and the Computer Speech Lab (CSL) 4300 (Kay Elemetrics Corp., Lincoln Park, NJ, USA) program.
Separately from these manipulations of the interrogative particles, we also manipulated the nuclear region following the procedure described in Chapter 1, section 2.1. of this dissertation. Then, using PSOLA synthesis, we created two continua by shifting the peaks upwards from the wh-question base stimuli and two more continua by shifting the peaks downwards from the yes-no question base stimuli, in each case in ten steps of 6 Hz each. Figure 1 shows two diagrams illustrating the manipulations performed in the prenuclear and nuclear regions for the wh-question base stimulus. For this experiment, subjects heard a total of 220 tokens (11 stimuli x 2 base stimuli x 2 interrogative particles x 5 blocks) presented in blocks of 44 randomized items.

![Figure 1](image)

**Figure 1:** Schematized diagrams of the ten tokens resulting from manipulation of the wh-question base stimuli (F0 movements are in bold). Manipulations of the congruent version with appropriate accented particle *qué* (‘what’) are shown in the left panel and manipulations of the incongruent version with unaccented particle are shown in the right panel. The creation of the pitch range continuum from a wh-question to a yes-no question are also shown in the graph. White boxes represent the consonants and shaded boxes the vowels.

### 2.1.3. Procedure

Subjects were seated at a laptop in a quiet room and the stimuli were played back through headphones. Since we were interested in not only utterance type identification choices but also Reaction Time measurements, listeners were instructed to maintain their hands near the keyboard and press the keys as fast as they could.

Subjects were asked to respond after each stimulus according to how they would answer the question in a real situation. In other words, if they perceived the yes-no question *Que l’hi duries?* (‘Would you bring it to her/him?’) they were to press the ‘S’ key on the keyboard (for *Sí* = ‘Yes’, an appropriate response to a yes-no question), whereas if they perceived the wh-question *Què li duries?* (‘What would you bring to her/him?’), they were to press the ‘A’ key (for *Això* = ‘That’, an appropriate response to a wh-question). The experiment lasted approximately 30 minutes and there was a break of 10 seconds between the blocks.

The presentation of the stimuli was controlled by an Acer TravelMate 4050 laptop, equipped with a Realtek AC’97 Audio sound card and the stimuli were presented through Sennheiser HD 202 headphones. The experimental situation was set up using E-prime version 1.2 (Psychology Software Tools Inc. 2009).

### 2.1.4. Statistical analyses
Although our data did not follow the normal distribution, ANOVA tests were used because of the following reasons: (a) they are generally considered to be robust to violation of the normality assumption meaning that the data might deviate quite a bit from normality but the test will still lead us to the right conclusion about our null hypothesis, (b) they reduce the unsystematic variability in the design and provide greater power to detect effects, and (c) they are widely used in the field of intonational phonology. Specifically repeated measures ANOVA tests were performed because we have a situation in which the same people take part in different experimental conditions (and not different people taking part in different experimental conditions). A repeated measures ANOVA was performed on the data for each type of question (yes-no question and wh-question), with Response as the dependent variable, and with two between-subjects factors, namely Stimulus (11 steps in the pitch range continuum) and Manipulated/Unmanipulated interrogative particle (2 values). When the Mauchly’s test indicated that the assumption of sphericity had been violated, degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity. The results from the repeated measures ANOVA tests include the $F$-ratio, the degrees of freedom from which it was calculated and the significance value.

2.2. Results

2.2.1. Identification results

Figure 2 and Figure 3 show the identification rate for the continua created from the yes-no question base stimulus and the wh-question base stimulus when the particle was either unmanipulated (black bars) or manipulated (grey bars). The ‘identification rate’ is defined as the number of ‘yes-no question’ responses (to yes-no question-based stimuli) or ‘wh-question’ responses (to wh-question-based stimuli) over the total number of responses. We observe that the four identification curves have a clear S-shape and interestingly, at least in the case of the yes-no question-based continua, the prenuclear region seems to have an effect on category identification by delaying its shift. Thus, when participants heard stimuli that contained the manipulated accented particle they tended to give more ‘wh-question’ answers than when the stimuli contained the unaccented particle, and this is especially noticeable when the nuclear scaling difference is in the ambiguous range (steps 5 to 8 on the x-axis). By contrast, the effect of the prenuclear region on utterance type identification is not so clear in the wh-question-based continua. The repeated measures ANOVA tests revealed statistically significant effects of Stimuli for both question types ($F(3.595, 25.110) = 104.092, p < .0001$ for the yes-no question-based-stimuli and $F(3.788, 34.096) = 111.766, p < .0001$ for the wh-question-based-stimuli). There was also a significant effect of Manipulated/Unmanipulated interrogative particle for the yes-no question-based stimuli ($F(1, 112)=5.926, p < .05$) but not for the wh-question-based stimuli ($F(1, 9) = .234, p > .05$).
Yes-no-question-based continua

Figure 2: Identification rate for the continua created from the yes-no question base with unmanipulated (black lines) or manipulated particle (grey lines). Error bars represent the standard error of the mean.

Wh-question-based continua

Figure 3: Identification rate for the continua created from wh-question base token with unmanipulated (black lines) or manipulated particle (grey lines). Error bars represent the standard error of the mean.

In order to compare mathematically the two identification curves obtained under the two conditions (manipulated/unmanipulated interrogative particle, for both the yes-no-question- and wh-question-based continua), the set of data points was fitted to a logistic function in SPSS program. From the SPSS fitted logistic curves, we obtained the “b1” and the “b0” values. While the term “b1” is related to the slope (with lower values reflecting steeper curves), the parameter “b0” does not contain any variable and, therefore, has constant value. The boundary is computed from these two terms given for the logistic curve for our data and the following formula to solve for $x$ when $y = 0.5$, 

$$x = \frac{\ln\left(\frac{b1}{b0}\right)}{\ln\left(\frac{1}{y} - 1\right)}$$
that is, when speakers identify the stimulus in question with the same number of “yes-no” and “wh-question” responses:

\[
\text{boundary} = -\frac{\ln b0}{\ln b1}^{17}
\]

For the yes-no question continuum with the unmanipulated particle, when ‘y’ equals 0.5, ‘x’ is 5.97; and when the stimuli have the manipulated wh- particle, when ‘y’ equals 0.5, ‘x’ is 6.43. For the wh-question continuum with the unmanipulated particle, when ‘y’ equals 0.5, ‘x’ is 5.11; and when the stimuli have the manipulated particle, when ‘y’ equals 0.5, ‘x’ is 5.61. These results agree with the previous statistical results in the sense that while the manipulation of the interrogative particle does not exert any effect on the wh-question-based continua, it does exert an effect in the yes-no question-based continua. In other words, the category shift is always between stimuli 5 and 6 in the yes-no question-based continua but it is delayed one stimulus in the yes-no question-based stimuli when the particle is manipulated.

Results from the analyses of the identification curves show that (a) the main predictor of the listeners’ responses is the scaling of the leading tone in the nuclear pitch accents, and (b) the effects of the manipulated/unmanipulated interrogative particle depend on the type of base token. Specifically, the presence of an accented interrogative particle in yes-no questions significantly favours wh-question judgements. This result might be attributed to the fact that it is perceptually more noticeable to hear an accented particle when an unaccented one would be expected than the other way around.

### 2.2.2. Reaction Time results

Figure 4 and Figure 5 show the averaged RT responses in ms for all subjects for the yes-no question-based stimulus series and the wh-question-based stimulus series with unmanipulated (black lines) or manipulated (grey lines) interrogative particles (for this specific experiment, the RTs did not include the duration of the stimulus). First, we can observe that for the continua with the unmanipulated interrogative particle there is a clear peak/plateau in RT that corresponds to the category shift calculated through the logistic regression. Thus, the peak/plateau for the yes-no question-based continuum is located between stimuli 6 and 7 (with the category shift occurring at point 5.97), whereas for the wh-question-based continuum the peak/plateau occurs between stimuli 4 and 5 (which roughly coincides with the boundary calculated from the logistic regression, that is, 5.11). By contrast, the identification curves corresponding to stimuli with manipulated interrogative particles have a peak in RT around the continuum edges (e.g., at the stimuli 3 and 11 for the yes-no question-based continuum and at stimuli 3 and 7-8 for the wh-question-based continuum). To some extent, we could say that the curves of the RT means obtained for the continua with the unmanipulated interrogative particle have an inverted U-shape while the curves of the RT means obtained for the

---

17 This formula as well as the procedure to calculate the location of category boundary are taken from Keating (2004).
continua with the manipulated interrogative particle have a U-shape. Interestingly, the peak of the RT curves for the continua with the unmanipulated interrogative particle coincides with the valley of the RT curves for the continua with the manipulated interrogative particle (circled in Figures 4-5 with a broken line). This indicates that listeners were slower at utterance type identification across categories when the stimuli contained an unmanipulated particle, and they were faster when the prosodic cues of the particle were those appropriate for the correct identification of the category located at the right edge of the continua. In other words, the RT measurements show that the prenuclear region was helpful when participants heard ambiguous stimuli and for that reason they were faster in identifying the question type.

For the yes-no question-based stimuli, the repeated measures ANOVA analysis revealed significant effects of Stimulus ($F(10, 150) = 5.847, p < .0001$) and Manipulated/Unmanipulated interrogative particle ($F(1, 15) = 186.692, p < .0001$) and also an interaction between Stimuli and Manipulated/Unmanipulated interrogative particle ($F(10, 150) = 16.079, p < .0001$). In the wh- question-based continuum we do find statistically significant effects of Stimuli ($F(10, 110) = 3.991, p < .0001$) and Manipulated/Unmanipulated interrogative particle ($F(1, 11) = 339.237, p < .0001$) as well as significant interaction between the two ($F(10, 110) = 6.795, p < .0001$).

![Figure 4: Averaged Reaction Time (RT) responses (in ms) for the yes-no question-based continua with unmanipulated (black lines) or manipulated interrogative particle (grey lines).](image)

Figure 4: Averaged Reaction Time (RT) responses (in ms) for the yes-no question-based continua with unmanipulated (black lines) or manipulated interrogative particle (grey lines).
In sum, we can conclude from these results that the prenuclear region plays a supplementary role in question identification in Majorcan Catalan. Importantly, we find an increase of its effect in the frontier region between the two categories, where the scaling properties are perceptually ambiguous. Given the secondary role of the prenucleus in distinguishing these two different types of questions, we now propose to test with a gating procedure whether the prenuclear part, taken in isolation, is sufficient in itself to allow Majorcan listeners to identify the specific type of question.

3. Experiment 2: Gating task

As we saw in the previous experiment, the effect of the prenuclear region on the identification of question type was limited to the frontier region between the categories in which the scaling difference was ambiguous. Now, our goal is to test whether there can be an increment in the perceptual impact of the prenuclear region on question type identification when presented in isolation, that is, when listeners cannot use the information located from the nuclear region. To this end, a gating procedure was used in which participants were presented with segments (gates) of the base token of gradually increasing duration, such that the shortest segment consisted of only one syllable (the syllable related to the interrogative particle) while the longest segment consisted of the full utterance. This technique was intended to allow us to evaluate the relevance of the prenuclear region for question identification in the absence of the nuclear tonal pattern. We used the same pair of recorded base tokens used in Experiment 1 (a yes-no question and a wh- question), and our initial hypothesis was that in any incomplete utterance the prenuclear
acoustic information in isolation would be used by listeners to try to identify the question type, but once they had the whole contour they would no longer pay attention to the interrogative particle, thus confirming the secondary role played by the prenuclear configuration.

3.1. Methodology

3.1.1. Participants

Fifteen native speakers of Majorcan Catalan (eight women, seven men) aged between 21 and 40 (median = 30) participated in this experiment. All participants had been born in Mallorca and came from different locales on the island. Again, they reported having no previous training in linguistics or history of hearing disability.

3.1.2. Materials

The materials were the base tokens used in Experiment 1, again with manipulated and unmanipulated question particles. Thus, we had four different base stimuli, as follows: (a) yes-no question base *Que l’hi duries?* (‘Would you bring it to her/him?’); (b) wh- question base *Qué li duries?* (‘What would you bring to her/him?’); (c) yes-no question with manipulated interrogative particle *que* (that is, accented *que* [ˈkə]); and (d) wh- question with manipulated wh- word *qué* (that is, unaccented *qué* [kə]). Since we wanted to isolate the contribution of the particle to the identification of the question type, each of the 4 stimuli was gated at 5 different locations in the utterance, namely, at the end of each syllable. The four conditions are shown in Table II. Conditions that contain the manipulated interrogative particle are shown in the shaded cells.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Yes-no question base</th>
<th>Wh- question base</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATE 1</td>
<td><em>Que</em></td>
<td><em>Qué</em></td>
</tr>
<tr>
<td>GATE 2</td>
<td><em>Que l’hi</em></td>
<td><em>Qué li</em></td>
</tr>
<tr>
<td>GATE 3</td>
<td><em>Que l’hi du-</em></td>
<td><em>Qué li du-</em></td>
</tr>
<tr>
<td>GATE 4</td>
<td><em>Que l’hi duri-</em></td>
<td><em>Qué li duri-</em></td>
</tr>
<tr>
<td>GATE 5</td>
<td><em>Qué li duries?</em></td>
<td><em>Qué li duries?</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Yes-no question with manipulated particle</th>
<th>Wh- question with manipulated particle</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATE 1</td>
<td><em>Qué</em></td>
<td><em>Que</em></td>
</tr>
<tr>
<td>GATE 2</td>
<td><em>Qué l’hi</em></td>
<td><em>Qué li</em></td>
</tr>
<tr>
<td>GATE 3</td>
<td><em>Qué l’hi du-</em></td>
<td><em>Qué li du-</em></td>
</tr>
<tr>
<td>GATE 4</td>
<td><em>Qué l’hi duri-</em></td>
<td><em>Qué li duri-</em></td>
</tr>
<tr>
<td>GATE 5</td>
<td>* Qué l’hi duries?*</td>
<td><em>Qué li duries?</em></td>
</tr>
</tbody>
</table>

Table II: Materials used in the gating experiment (5 gates x 2 types of interrogatives x manipulated vs. unmanipulated interrogative particle).
Participants heard a total of 10 tokens (5 gates x 2 base stimuli x 2 interrogative particles x 5 blocks) presented in blocks of 20 randomized items. Figure 6 shows an example of the different gates of the yes-no question *Que l’hi duries?* (Would you bring it to her/him?) with the unmanipulated interrogative particle [’kə].

![Figure 6](image)

**Figure 6**: Waveform, spectrogram, orthographic transcription, prosodic transcription and different gates of the yes-no question *Que l’hi duries?* (‘Would you bring it to her/him?’) with the unmanipulated interrogative particle [’kə].

### 3.1.3. Procedure

In this experiment we followed the same general procedure as in Experiment 1 (see section 2.1.3.), that is, subjects were seated at a laptop in a quiet room and the stimuli were played back through headphones. Reaction Time measurements were also recorded and for that reason listeners were asked to respond as fast as they could.

The classical gating paradigm (Grosjean 1996) was followed. Participants were presented with the different gates of the 4 stimuli in a randomized fashion and had to identify the sentence they heard as either a yes-no question or a wh- question. They responded by clicking either the ‘S’ key (for *Sí* = ‘Yes’ as an appropriate response to a yes-no question) or the ‘A’ key on the keyboard (for *Això* = ‘That’, as an
appropriate response to a wh-question). In addition, participants were told that since it was a difficult task (given the truncated nature of some of the tokens they were about to hear), they could express their uncertainty about the utterance type they had heard by means of a 5-point confidence rating scale (1 = totally uncertain; 5 = totally certain). These two tasks are based on the gating paradigm and on the distinction between isolation and recognition points proposed by Grosjean (1996). The isolation point is “the size of the segment needed to identify the stimulus” (Grosjean 1996: 598), while the recognition point is obtained “by taking the segment size needed, after the isolation point, to reach a particular confidence level (e.g., 80% confidence for Tyler & Wessels 1983)” (Grosjean 1996: 599). The experiment lasted approximately 15 minutes and there was a break of 10 seconds between blocks.

Stimuli were presented on a MSI U100 Wind Notebook laptop equipped with a Realtek HD sound card and Sennheiser HD 202 headphones using E-prime version 1.2 (Psychology Software Tools Inc. 2009).

3.1.4. Statistical analyses

Repeated measures ANOVAs were performed on the data collected in this experiment because of the same reasons explained in section 2.1.4. of this chapter. Three repeated measures ANOVAs were performed on the data for each type of question (yes-no question and wh-question), with Response, Confidence Rating and RT as the dependent variables, and with two between-subjects factors, namely Gate (5 values) and Manipulated/Unmanipulated interrogative particle (2 values). The same procedure as in the previous experiment was followed with respect to the assumption of sphericity and the statistic values.

3.2. Results

Following Grosjean (1996), we calculated the isolation (portion of the segment needed to identify the stimulus) and recognition points (segment needed to obtain a particular confidence level). Figure 7 and Figure 8 show the cumulative response rates for gates of the yes-no question (top) and wh-question (bottom). Black bars represent unmanipulated gates and grey bars manipulated gates. We observe that the results for gates 1 and 2 (the first two syllables in the sentence, before the syllable carrying the leading tone of the nuclear pitch accent) display an identification rate of correct responses that is at chance level for the yes-no question (with an average rate of 46% for the first and second gates when the interrogative particle has not been manipulated) but above chance for the wh-question gates (with an average rate of 56% and 59% for the first and second gates respectively when the interrogative particle has not been manipulated). This suggests that although the information contained in the two first gates is not sufficient to clearly distinguish yes-no questions from wh-questions, it does prevent listeners from responding totally at random at least in the specific case of the wh-question gates. Again, as in the case of the identification results, we find that hearing an accented particle (wh-question
gates) yields higher scores than hearing an unaccented particle (yes-no question gates). Correct judgements increase sharply at the third gate, which corresponds to the third syllable (Que/què l’hi du-), that is, the syllable carrying the leading tone H of the nuclear pitch accent. On the other hand, we also see in the graph that the manipulated wh- particle does not affect the listeners’ judgements. In terms of the isolation point, that is, the size of the segment needed to correctly identify the stimulus, Figures 7 and 8 clearly confirms that listeners were able to do this at above chance level at the third gate.

**Yes-no question Gates**

Figure 7: Response rates for yes-no question gates. Black bars represent unmanipulated and grey bars manipulated interrogative particles.

**Wh- question Gates**

Figure 8: Response rates for wh- question gates. Black bars represent unmanipulated and grey bars manipulated interrogative particles.
For the yes-no question, the repeated measures ANOVA analysis showed significant effects of Gate ($F(4, 24) = 56.895, p < .0001$) but non-significant effects of Manipulated/Unmanipulated particle ($F(1, 6) = 1.000, p > .05$). No significant interaction between Gate and Manipulated/Unmanipulated particle was found ($F(4, 24) = 0.388, p > .05$). For the wh- question, again the analysis revealed a significant effect of Gate ($F(4, 24)=15.864, p < .0001$), no significant effects of Manipulated/Unmanipulated particle ($F(1, 6) = 1.636, p > .05$), and no interaction between Gate and Manipulated/Unmanipulated particle ($F(4, 24) = 1.186, p > .05$).

Finally, Figure 9 and Figure 10 show the mean confidence ratings given by the subjects at the 5 gates. Black lines represent unmanipulated stimuli and grey lines manipulated stimuli. The mean confidence rates show that it is not until the fourth gate that listeners feel truly confident about their judgements, as their decisions go over the 80% confidence (numbers 4 and 5 in the confidence scale) at that point.

For the yes-no question, the repeated measures ANOVA analysis revealed statistically significant effects of Gate ($F(4, 56) = 32.862, p < .0001$), no effect of Manipulated/Unmanipulated particle ($F(1, 14)= .051, p > .05$), and no interaction between Gate and Manipulated/Unmanipulated particle ($F(4, 56) = 2.075, p > .05$). For the wh- question, the statistical results were parallel, that is, a significant effect for the Gate factor ($F(4, 56) = 37.163, p < .0001$), no effect of Manipulated/Unmanipulated particle ($F(1, 14) = 3.500, p > .05$), and no significant interaction between Gate and Manipulated/Unmanipulated particle ($F(4, 56) = 2.025, p > .05$).

**Figure 9:** Mean confidence ratings for yes-no question gates. Black lines represent unmanipulated gates and grey lines manipulated gates.
Figures 11 and 12 show average Reaction Times (RT) in ms for the yes-no question and wh- question gates. Grey lines indicate the gates in which the interrogative particle was manipulated and black lines the gates in which the interrogative particle was not changed. As predicted, the data show a tendency to have higher RT values for the gates that contain a contradictory interrogative particle (especially for the 2nd gate), since listeners heard an unexpected interrogative particle. Another important effect is that the longer the gate, the lower the RT scores, meaning that the increase in correct judgements goes hand in hand with a decrease in the time needed to identify the stimuli. For the yes-no interrogative, the repeated measures ANOVA the analysis revealed a significant effect of Gate on RT \( F(4, 24) = 12.189, p < .0001 \), no effect of Manipulated/Unmanipulated particle \( F(1, 6) = .884, p > .05 \), and no significant interaction between Gate and Manipulated/Unmanipulated particle \( F(4, 24) = 1.313, p > .05 \). For the wh- question interrogative, we obtained the same pattern of results: a significant effect of Gate \( F(4, 20) = 25.620, p < .0001 \), no effect of Manipulated/Unmanipulated particle \( F(1, 5) = .077, p > .05 \), and no significant interaction between Gate and Manipulated/Unmanipulated particle \( F(4, 20) = 1.143, p > .05 \).

In sum, the results of the gating task show that when Majorcan Catalan listeners hear the accented ['kɔ] vs. unaccented [kə] version of the interrogative particle in isolation, that is, the first gate, this is not sufficient information for them to distinguish a yes-no question from a wh- question. As seen in figures 7 and 8, it is not until the third gate that participants are able to correctly identify the question type above chance level. In addition, Figure 9 and Figure 10 show that an 80% confidence level is obtained only at the fourth gate, namely, only after the syllable carrying the nuclear pitch accent is present in the stimulus. An important methodological finding related to this set of
results is that although little importance has been attached thus far to confidence rating in intonational studies in which the gating procedure was used, we claim that it is only by means of confidence rating results that we are really able to confirm that listeners recognize the question type.

**Figure 11:** Average Reaction Time (RT) in ms for yes-no question gates. Black lines represent unmanipulated gates and grey lines manipulated gates.

**Figure 12:** Average Reaction Time (RT) in ms for wh- question gates. Black lines represent unmanipulated gates and grey lines manipulated gates.
4. Discussion and conclusions

The goal of Experiment 1 (identification task) was to test two hypotheses, namely, (a) whether Majorcan Catalan listeners would be sensitive to the accentual properties of the interrogative particles *que* [ka] ‘that’ vs. *què* [’ka] ‘what’ in the identification of yes-no questions and wh-questions; and (b) whether Majorcan Catalan listeners would give more weight to the prenuclear information when the nuclear information was ambiguous. We tested this by combining the prenuclear information with a continuum of pitch scaling in the nuclear pitch accent. Our initial expectation was that the presence of prenuclear accentuation (expressed through pitch, duration and F1 differences in the particles *que*/*què*) would help the listeners to identify question type, and that this role would be enhanced when the nuclear information was ambiguous. In addition, according to the results of Chapter 1, we expected that prenuclear acoustic cues would facilitate perception of the category change and that, in addition, it would enable listeners to be faster in categorizing stimuli. As expected, results revealed a significant effect of the prenuclear region on utterance type identification in those stimuli in which the scaling properties of the nuclear pitch accent were ambiguous (i.e., in-between categories, in the mid part of the pitch range continuum). This result is an indication that listeners process intonational contours by making an active use of both the prenuclear and nuclear parts of the pitch contour information.

In addition, the results of Experiment 1 revealed a perceptual asymmetry depending on whether the manipulated/unmanipulated interrogative particle was accentuated or unaccented. In essence, the presence of an incongruent accented interrogative particle with yes-no questions favoured more wh-question judgements than the reverse combination (unaccented interrogative particle with wh-questions). We interpret this as being due to the greater perceptibility of accented particles. The RT results point in the same direction. As expected, a peak in RT is found corresponding to the identification boundary calculated for both continua in which the interrogative particle is not manipulated. By contrast, in the continua in which the interrogative particle *was* manipulated, a mean RT minimum can be observed that corresponds exactly to the peak obtained for the other continua. Thus, Catalan listeners are slower at the frontier region between the categories because the prenuclear cue is not present and the nuclear difference is ambiguous; by contrast, listeners are faster when the interrogative particle is manipulated because it serves to resolve the ambiguity of the nuclear region.

The goal of Experiment 2 (the gating task) was to investigate whether there would be an increase in the effect of the prenuclear region on question type identification when the cues related to the prenuclear accent were presented in isolation. First, the identification point is located at the third gate. This is especially interesting because the third gate is the first portion that includes the syllable which is associated to the H leading tone of the nuclear pitch accent. The mean confidence ratings show that although subjects can identify the question type at the third gate, it is not until the fourth gate that they feel more than 80% confident about their judgements. Moreover, we observed that for the wh-question gates the results at the first and second gates are well
above chance (56% and 59% correct identifications for the first and second gate respectively). This is an indication that listeners are using the information that is available in very early portions of the sentence. Based on our gating task results, we can conclude that even though Catalan listeners are unable to accurately identify question type simply by listening to the prenuclear portion of questions, they can and do exploit this portion of the intonation contour to some extent in making judgements.

When comparing our gating results with those obtained by previous studies, we note that they are in line with some of the results obtained for French (Vion & Colas 2006), Italian (Petrone 2010 for Cosenza Italian) and Central Catalan (Crespo-Sendra 2011) since all these studies also obtained identification scores that were just above chance level. In our gating experiment, the first two gates obtained a correct identification score that was at chance level or above it (0.46 for the first gates of the yes-no question base stimuli and 0.56 and 0.59 for the first and second gates respectively of the wh- question base stimuli). On the other hand, other studies have obtained moderate results for identification rates (e.g., Petrone 2008, and Petrone & D’Imperio 2010, for Neapolitan Italian). Finally, results for other languages (or language varieties) like Spanish or Valencian Catalan show that the prenucleus can have a key role in sentence-type identification and yield very high identification rates. For example, in Face’s (2005) study, the identification rate of yes-no questions in Spanish at the prenuclear region was 87.5%. Similarly, in Crespo-Sendra (2011), the identification rate for incredulity yes-no questions in Valencian Catalan was 91%.

Seen as a whole, the abovementioned studies together with the present one suggest that there are cross-linguistic differences in the amount of weight assigned to the prenuclear part of intonation contours for the purposes of sentence-type identification. These varying perceptual enhancement properties of the prenuclear region seen across studies may very well depend on the perceptual strength of the specific acoustic cues being investigated (which can be perceptually more salient in some specific contours), and on the assignment of sentential prominence to the earlier parts of the contour. The present study has confirmed that Catalan listeners can also exploit the prenuclear part of the contour in question identification tasks, and, more importantly, that the enhancement properties of the prenuclear part of the contour become greater when the nuclear part of the contour is more ambiguous. This is a clear indication that the processing of intonation contours involves an active task of temporal integration between the different parts of the intonation contours.

The results obtained in this chapter have some implications for the idea of “compositional approach” advocated by the Autosegmental-Metrical model of intonation. The notion of compositionality in the meaning of intonation contours was first proposed by Pierrehumbert and Hirschberg (1990), and has been maintained in subsequent work within the Autosegmental-Metrical framework, as well as in other frameworks (Ladd 2008, Gussenhoven 1984, Bartels 1999, Steedman 2003, Marandin et al. 2004, and others). In essence, this theory argues that the meaning of an intonational contour is the sum of the ‘independent’ contribution of each of its tonal morphemes (pitch accents and edge tones) (Pierrehumbert & Hirschberg 1990). The findings of the present chapter related to the perceptual integration and dependence
between the two parts of the contour are difficult to accommodate to a general compositional view of intonational meaning. In essence, the fact that the prenuclear part of the contour becomes more relevant perceptually when the nuclear part of the contour is ambiguous (or is not present) is a clear indication that listeners evaluate perceptually all parts of the contour and integrate them while searching for pragmatic meaning. It is this perceptual integration that may explain the interaction found in this research. At this juncture, these results are difficult to reconcile with a view that attributes the general meaning of an utterance to the sum of the separate meanings of each one of the tonal morphemes (i.e., pitch accents and boundary tones).
Part II
Information-seeking questions and confirmation-seeking questions
CHAPTER 3

Information and confirmation yes-no questions in Eastern Catalan: production data

This chapter presents an empirical investigation into the relationship between intonational phonology and pragmatics using empirical data from Eastern Catalan (specifically from the varieties of Central and Balearic Catalan), a language which can signal the distinction between (or function of) information- and confirmation-seeking question by means of prosody. The motivation for studying these varieties of Catalan is that quite often dialectal variation is found in the intonation of interrogatives (see general Introduction), and we felt that it would be of considerable interest to test how this dialectal variation interacts with speaker certainty. This chapter addresses the issue of the interplay between dialectal variation in intonation and the expression of epistemic modality\(^{18}\) while testing the reliability of an intonation survey of the sort known as a Discourse Completion Test in the pragmatics literature (Kasper & Dahl 1991, Cohen 1996, Beebe & Cummings 1996, Nurani 2009).

1. Introduction

1.1. What does interrogativity mean?

If we conceive of communication as one of the main goals of human language, interrogativity is the speech act that accomplishes this goal in the most optimal way. The term “question” is defined in Diec2 (Institut d’Estudis Catalans) as follows: “Acció de demanar a una persona informació sobre alguna cosa que es desitja saber, un aclariment, la solució d’algun dubte”.\(^{19}\) That means that even though the act of asking is an effective way of making people communicate relevant information, the intention that is behind the speaker is not always that of getting information, but may instead be confirming previous assumptions, formulating an offer, expressing surprise, etc. When we talk about interrogatives, it is not sufficient to consider the sentential force or sentence modality, which is the form of a specific sentence; we must also refer to its illocutionary force, that is, its function. Modality or sentential force refers to the pragmatic force associated with a specific type of sentence (Chierchia & McConnell-Ginet 1990) whereas the illocutionary force is related to the intentions of the speaker. Thus, we observe that (1) is a declarative with the illocutionary force of a question (example taken from Portner & Zanuttini 2003), while (2) is formally a question but has

\(^{18}\) Epistemic modality concerns an estimation of the likelihood that a certain state of affairs is true (or false) in the context of the possible world under consideration.

\(^{19}\) ‘The action of asking somebody for information about something, a clarification, the resolution of a doubt’ (the translation is ours).
the illocutionary value of a statement (example taken from Prieto & Cabré (coords.) 2007-2010):

1. Avrei bisogno di sapere come andare all’aeroporto. (Italian)
   ‘I would like to know how I can get to the airport.’

2. Hi haurà mai tranquil·litat en aquesta casa? (Catalan)
   ‘Is there ever going to be peace and quiet in this house?’

Thus it is clear that not all interrogative syntactic forms function as questions and that sentences that function as questions can present a non-interrogative form (Kiefer 1981). In fact, some scholars have proposed that the terms “question” and “interrogative” should even be regarded as having quite separate meanings. Thus, the term “interrogative” should correspond to the syntactic form, while the term “question” should refer to the pragmatic function. This distinction will be very important in this chapter. For that reason, a first step will be to consider the terms “question” and “interrogative” from the perspective of three different levels of linguistic analysis: syntax, semantics and pragmatics (Haan 2001).

1.2. **Interrogativity and syntax**

Sentences have been traditionally classified according to their syntactic form, there being three different types: declaratives, interrogatives and imperatives. Interrogatives in Germanic languages present a more rigid structure compared to Romance languages, this structure being characterized by subject-verb inversion (3) or auxiliary forms (4), although there exist interrogatives with neither inversion nor auxiliary (5), and it is also possible to find auxiliaries in declarative sentences (6) or even inversion in declarative sentences (7).

3. Is David sick?

4. Did Jim leave early?

5. Jim left early?

6. Jim did leave early.20

7. Came a day when for forty hours he had not tasted food.21

With respect to Romance languages and specifically with respect to Catalan, there are three factors that make less clearcut the one-to-one correspondence between sentence type and the form of the surface structure: the absence of an auxiliary verb, a

---

20 According to personal communication with Jill Thorson, these examples are possible in American English although (5) would be restricted to a counterexpectational interpretation.

freer word order and the possibility of not expressing the syntactic subject. Let’s see how word order is expressed in Catalan interrogatives.

According to Payrató (2002), polar interrogatives (i.e., yes-no questions) can present either a SVO order, as in *El jutge va dictar tres sentències?* (‘The judge passed three sentences?’) or a VOS order, as in *Va dictar tres sentències el jutge?* (‘Passed three sentences the judge?’) (according to this author, the second version is more common). Interrogatives with dislocated subjects (and therefore belong to a different group in terms of prosody) are also possible, according to this author, although they would have a marked interpretation since the speaker would thereby be placing special emphasis on one particular element or another: *Va dictar, el jutge, tres sentències?* (‘Did he pass, the judge, three sentences?’), *El jutge, va dictar tres sentències?* (‘The judge, did he pass three sentences?’) or *Va dictar tres sentències, el jutge?* (‘Did he pass three sentences, the judge?’). Also according to this author, wh- questions in Catalan present more visible word order changes compared to yes-no questions, since the interrogative modality marks are more evident and the subject, when it is present, is dislocated: *Quantes sentències va dictar, el jutge?* (‘How many sentences did he pass, the judge?’) and also constitute a different prosodic group.

Following Rigau (2002), in the interrogative modality the position of the subject is peripheral, that is, outside the sentence, both in yes-no questions: *Que treballa, el noi?* (‘Is he working, the boy?’) and in wh- questions: *Què feia, l’àvia?* (‘What did she use to do, the grandmother?’).

By applying the Map Task methodology22 to elicit oral data from native speakers of Catalan with the goal of determining the role of information structure in the intonation of yes-no questions, Payà and Vanrell (2005) found two different orders: (S)VO (*Bé, tu has passat per davant ca na Janera?,* ‘Well, have you been past Janera’s house?’) and either (S),VO or VO,(S) (with dislocated subject) (*Tens una barberia, tu?,* ‘Have a barber’s shop, [do] you?’). According to the same study, the former is more typical in interrogatives that merely seek to confirm a pre-existing assumption on the part of the speaker (confirmation-seeking questions) while the latter is more common in interrogatives that are asking for information of which the speaker is ignorant at the time of asking (information-seeking questions). Consistent with Hernanz and Rigau (2006), tag questions were interpreted to be confirmation-seeking questions which presuppose either an affirmative or negative answer e.g., *Damunt sa immobiliària i Caixa Catalunya, no?* (‘Above the real estate agency and Caixa Catalunya, no?’).

Prieto and Rigau (2007) contrast the presence of left- or right-dislocated subjects in information-seeking yes-no interrogatives like in *En Joan, (que) viu a Barcelona?* (‘Joan, is he living in Barcelona?’) with the initial and non-dislocated position of the subject in biased questions (i.e., confirmation-seeking questions or counterexpectational

---

22 The Map Task is a validated technique in which two subjects cooperate to complete a specified task. Each of the two subjects has a map of an imaginary town marked with buildings and other specific elements such as fountains and monuments. A route is marked on the map of one of the two participants and that person has the role of the instruction-giver. The version of the same map held by the other participant differs from that of the instruction-giver in that it does not show the route to be followed. The second participant therefore has to ask the instruction-giver questions in order to be able to reproduce the same route on her or his own map.
questions) like in *Oi que en Pere no va a Barcelona?* (‘Peter isn’t going to Barcelona, is he?’).

In sum, just as the syntactic property of inversion and the presence of an auxiliary is not an immutable feature of interrogatives with respect to Germanic languages, the presence of a specific word order in Romance languages is not a formal and fixed characteristic for an interrogative, since, particularly in the case of Catalan, the dislocation of a subject will clearly only be key when the subject is present (Catalan is a null subject language). For instance, while the Catalan interrogative shown in (8) would be interpreted as an information-seeking yes-no question (i.e., a question for which the speaker has no particular bias with respect to the answer s/he expects) the one shown in (9) would be a biased yes-no question (in the sense that the speaker has some bias about the reply based on belief, expectations, world knowledge or discourse context), and (10) would be uninterpretable altogether, at least on the basis of syntax alone. We must conclude, then, that syntactic criteria are not sufficient to identify interrogative sentences and to distinguish them from other question types.

(8) *Ve, la Maria?*

*Comes, pers. article fem. Maria?*

‘Is Maria coming?’

(9) *La Maria ve?*

*Pers. article fem. Maria comes?*

‘Maria is coming?’

(10) *Ve?*

*Comes?*

‘Is s/he coming?’

1.3. **Interrogativity and semantics**

Most semanticists agree with the idea that interrogatives are distinguished from declaratives in the sense that interrogatives do not contain full propositions. For that reason, they cannot be assigned a truth value (Hiz 1978, Kiefer 1983, Higginbotham 1995). In order to go beyond this problem, it has been proposed that a semantic analysis of interrogatives should include, apart from the question, the corresponding answer (Hiz 1978, Kiefer 1983, Higginbotham 1995). But the problem again is that interrogatives

---

23 We feel that this difference in interpretation with respect to the position of subject (whether dislocated or not) in neutral yes-no questions between Payrató (2002) on one hand and the other studies referred to in this section on the other might be due to the fact that Payrató argues that topic and comment work in a similar way in questions and declaratives. Here we claim that in questions the speaker focalizes a presupposed comment with the aim of checking whether it is a correct comment with respect to other potential comments. Hence, in the neutral interrogative version of the sentence *[La Maria]TOPIC [menja mandarines]COMMENT ([Maria]TOPIC [eats tangerines]COMMENT)*, the speaker will focalize the comment *menja mandarines* and will dislocate the rest, i.e., *La Maria*. The result will be a question like *Menja mandarines, la Maria?* or *La Maria, menja mandarines?*
are not always followed by an answer (see underlined questions in (11a) and (11b)). Hence, this approach would not be completely valid.

(11a) EULÀLIA: Això mateix. Parlem-ne. Tu, Mercè, on aniries?
EULÀLIA: That’s it. Let’s talk about it. Where would you go, Mercè?
MERCÈ: I per què he de començar jo?
MERCÈ: And why do I have to start?
NÚRIA: Perquè ets la més gran!
NÚRIA: Because you are the eldest!
MERCÈ: Fugiu! A tu et porto dos mesos i a l’Eulàlia...
MERCÈ: Come on! I’m two months older than you and with respect to Eulàlia...

From Joan Oliver: Ball robat (1958)

(11b) LISBETH: Per què no comença? On és Walter?
LISBETH: Why aren’t they starting? Where is Walter?
SCHMID: Fillà, Walter no podrà tocar la flauta.
SCHMID: Daughter, Walter won’t be able to play the flute.
LISBETH: I per què? La plaça està plena de gent que l’espera. Tothom està molt content.
LISBETH: And why not? The square is full of people waiting for him. Everybody is very happy.

From Jordi Teixidor: El retaule del flautista (1970)

Hence, if a question does not carry a full proposition, we will need to resort to the semantic notion of “presupposition”. It is assumed that interrogatives depend on presuppositions, that is, they depend on some declaratives which the speaker believes to be true (Droste 1972, Kiefer 1980, Groenendijk & Stokhof 1996). Thus, a wh- question presupposes a declarative sentence in which the wh- word will be replaced by an expression like la Maria when, for example, we use the wh- word qui (‘who’). By contrast, yes-no questions can be based on different presuppositions depending on what is being asked for (in other words, whether what is questioned is the full proposition, i.e., broad focus, or only a part, i.e., narrow focus)24. Propositions can also be expressed contextually. For that reason, we must now address the question of pragmatics.

1.4. Interrogativity and pragmatics

Pragmatics is mainly interested in the context in which utterances are produced, that is, language use with respect to the communicative context. Thus, utterances can represent a wide range of speech acts such as warnings, assertions, requests, promises, etc. (Searle 1976). It is claimed that there is no one-to-one correspondence between speech acts and

---

24 An example of broad focus interrogative would be [La Maria menja mandarines?],F (“Is Maria eating tangerines?” [Maria is.eating tangerines?]_{F} and a narrow focus interrogative, [La Maria],F [menja mandarines?],F ([Maria],F [is.eating tangerines?]_{F} or [Menja mandarines],F [la Maria?],F. ([Is.eating tangerines.]_{F} [Maria?]_{F}).
syntactic categories. In the standard theory of speech acts, interrogatives have been considered “petitions”, since the speaker is producing a petition and the hearer is expected to provide relevant and correct information to answer this request.

Escandell (1993, 1998) characterizes interrogative sentences according to whether the speakers' discursive goals are *transactional* (i.e., the speaker’s intentions are mainly informative) or *interactional* (i.e., the speaker seeks to cooperate and socialize). In some cases, when the discursive intentions of the speaker are mainly transactional, the interrogative variable is an informative gap that the hearer can fill with a certain piece of information. Yet importantly, according to Escandell, how much a speaker can choose from a variety of questions depends on her/his own available knowledge and how much knowledge s/he assumes the hearer to have. Figure 1 below (from Escandell 1993, as adapted by Prieto 2002a) schematically represents the different degrees of knowledge on the part of the speaker and presumed knowledge possessed by the hearer. The triangle at left represents the speakers’ knowledge about the content of the question and the inverted triangle at right represents the hypothesis that the speaker formulates about the knowledge possessed by the hearer. The center column shows the resulting interrogative output of the combination of the speaker’s real knowledge together with the hearer's presumed knowledge.

![Diagram](image)

**Figure 1:** The lefthand triangle shows the speaker’s real knowledge and the righthand triangle shows the hearer’s presumed knowledge. (Diagram from Escandell (1993) as adapted by Prieto (2002a) as well as the present study.)

Neutral questions (those in which the speaker seeks to obtain information that s/he lacks) are located at the top of the scale because they imply minimal knowledge on the part of the speaker and at the same time a maximal presumption of the knowledge in possession of the addressee. By contrast, at the other opposite extreme of the scale one finds exam-style questions, which imply a maximal degree of knowledge on the part of the speaker and no presupposition of knowledge on the part of the interlocutor. However, there are times when the speaker’s ignorance is not total: s/he already knows the answer to her/his question and thus merely seeks confirmation of her/his hypothesis. The resulting output is a confirmation-seeking question, which Escandell calls an “oriented question” and which would be located at some intermediate point in the diagram. In this type of question, the speaker expects either a confirming or a disconfirming answer. In Catalan confirmation-seeking questions we also often find confirmation marks such as polarity adverbs (*no*, *oi*), the noun *veritat* or the particle *eh*
(Cuenca 1997, Hernanz & Rigau 2006, Prieto & Rigau 2007). Regarding the syntactic order of constituents, confirmation-seeking questions in Catalan can also have the subject in preverbal position and without dislocation (Rigau 2002, Prieto & Rigau 2007). It is also possible for confirmation-seeking questions not to present confirmation marks. In this case, the type of question will be triggered by either the syntactic order (again, only when the subject is expressed) or the intonation. Though it is well established at this point that languages can use different strategies to mark the epistemic disposition towards a proposition, that is, the speakers’ degree of certainty about it, such as lexical marking (the choice of modal adverbs), morphological marking, facial and body gestures, and prosody (Palmer 2001), much less work has been done thus far regarding the specific contribution of prosody in the expression of speaker certainty in questions.

1.5. Goals

In some Romance languages, speakers can signal the distinction between (or function of) information- and confirmation-seeking questions by means of prosody. Grice and Savino (1997) demonstrated that in Bari Italian the choice of a specific pitch accent reflects the degree of confidence with which the speaker believes the information to be shared with the interlocutor. In this variety of Italian, questions about new information are marked by means of a L+H* accent while questions about given information are signalled through a H+L* accent. Recent findings for Puerto Rican Spanish (Armstrong 2010) suggest that the difference between questions in which the speaker has no belief about the propositional content and questions where the speaker believes the propositional content to be true is expressed in this variety by a difference in the location of the high tone. Thus, while in the former the peak is associated with the nuclear syllable (H* L%), in the latter it is located in the syllable preceding the nuclear one (H+L* L%). Perceptively, the nuclear syllable is heard high in questions when the speaker does not have any specific belief about the propositional content and low when s/he does.

Similarly, Germanic languages such as Dutch and German are able to convey this contrast through intonation. Haan (2001) analyzed in a production experiment the intonation of yes-no questions with declarative syntax, yes-no questions with verb inversion and wh- questions in Dutch. A rising nuclear pattern was found in 100% of the yes-no questions with declarative syntax, 94% of the yes-no questions with inversion and 64% of the wh- questions. The author offers no explanation for the differing percentages in the appearance of the final rise. However, Kügler (2003) suggests that it could be due to a difference in the speakers’ expectation of an answer in the case of a yes-no question. Kügler’s (2003) study of Leipzig German revealed that the interaction between intonation and information structure was displayed through the use of different boundary tones. According to him, questions in which the speaker is asking for new information are produced with a high boundary tone but questions in which the speaker expects a particular answer based on given information in the previous context employ a low boundary tone.
Our main goal in the first part of this chapter is to analyze prosodically, but also from the semantic and pragmatic point of view (but without losing sight of syntax), information-seeking questions and confirmation-seeking questions in Central and Balearic Catalan (Majorcan, Minorcan and Ibizen/Formenteran). Since the latter are formally declaratives, presupposition differences can be expressed by means of prosodic features as the primary cues. The motivation to study these varieties of Catalan is that quite often dialectal variation can be seen in the intonation of interrogatives, and we felt that it would be of considerable interest to test how this dialectal variation interacts with speaker certainty. Both types of questions have been designated differently by different authors and in different languages. Thus while Droste (1962), Quirk et al. (1987) and Huddleston (1994) talk about yes-no questions/declarative questions, Escandell (1993) introduces the term neutral/oriented questions, Haan (2001) refers to (new) information/confirmation questions, Grice and Savino (1997, 2003a, 2003b) use the terms queries/checks, and these types have also been designated confirmation/information-seeking requests by Santos and Mata (2010). A second goal is to test the reliability of an intonation survey of the sort also known as a Discourse Completion Test (henceforth DCT) in the pragmatics literature (Kasper & Dahl 1991, Cohen 1996, Beebe & Cummings 1996, Nurani 2009).

The remainder of the third chapter is organized as follows. Section 2 contains the methodology used in this production experiment, section 3 presents the results for each of the four varieties under study, that is, Central and Balearic Catalan (Majorcan, Minorcan and Ibizen/Formenteran Catalan) and, finally, section 4 presents a summary of the major findings as well as its general conclusions.

2. Methodology

2.1. Participants

The participants in our production experiment were twenty-one women25 aged between 16 and 45 from the following locales in the two sub-regions of the Eastern Catalan linguistic domain: a) Central Catalan (Barcelona province: 1 subject from Barcelona, 2 subjects from Borredà, 1 subject from Sabadell; Girona province: 1 subject from la Bisbal d’Empordà and 1 subject from Figueres; Tarragona province: 1 subject from Reus), and b) Balearic Catalan (the island of Majorca: 1 subject from Alcúdia, 1 subject from Campos, 3 subjects from Llucmajor, 2 subjects from Ses Salines; the island of Minorca: 1 subject from Ciutadella, 1 subject from Alaior and 1 subject from Maó: the smaller adjacent islands of Ibiza and Formentera: 2 subjects from Ciutat d’Eivissa and 1 subject from Sant Josep; 1 subject from Sant Ferran). See Figure 1 for the location of the locales from which data analyzed in this study was collected.

---

25 We justify choosing only women because of their wider pitch range. It allows to create smoother intonation contours.
2.2. Materials

The corpus that has been analyzed in this chapter was obtained by means of an intonation survey. This survey was based on the DCT (Kasper & Dahl 1991, Cohen 1996, Beebe & Cummings 1996, Nurani 2009) and on the questionnaire used by Prieto (2001). It is an inductive method in which the researcher presents the subject with a series of situations (such as “You go into a shop you have never been in before and ask the shop assistant if they sell sugar”) and then asks him or her to respond accordingly. The full survey is made up of 17 situations that allow the researcher to obtain a wide range of interrogative contours such as information-

---

26 This map as well as the map appearing in page 10 were created by Joan Borràs-Comes. We would like to thank him for his technical help in this work.
seeking questions, confirmation-seeking questions, counterexpectational questions and echo-questions. However, only information- and confirmation-seeking questions have been analyzed in this chapter. Since we were interested in the prosodic differences that exist between information- and confirmation-seeking questions, we collected contours that were syntactically ambiguous, that is, they were VO sentences with an object NP (or PP in some cases). Thus, since the subject was not present, it was not possible to determine the question type on the basis of whether the subject was or was not dislocated (see section 1.3 of this chapter). The six situations used for the elicitation of the materials were as follows:

- Situations used to elicit information-seeking yes-no questions:

1. *Tens un poc de mal de coll. Li demane s al teu amic si té un caramel·lo.*
   Potential answer: ‘*Tens un caramel·lo?’*
   You have a sore throat. Ask your friend whether he has a cough drop.
   Potential answer: ‘Do you have a cough drop?’

2. *Vas amb un amic i et telefonen. Necessites apuntar un número de telèfon que t’han donat. No tens res per escriure i li demanes al teu amic si té un boli.*
   Potential answer: ‘*Tens un boli?’*
   You are walking with a friend and suddenly someone rings you. You need to write down a phone number but you do not have a pen. Ask your friend whether he has a pen.
   Potential answer: ‘Do you have a pen?’

3. *Entres a una botiga on ho havies estat mai i demanes al botiguer a veure si tenen mandarines.*
   Potential answer: ‘*Teniu mandarines?’*
   You have just entered a shop you have never been in before. Ask the shopkeeper whether he has any tangerines.
   Potential answer: ‘Do you have any tangerines?’

- Situations used to elicit confirmation-seeking yes-no questions:

1. *Un amic teu t’ha anat a comprar caramel·los pel mal de coll perquè tu li ho havies demanat. Demana-li si els du.*
   Potential answer: ‘*Dus els caramel·los?’*
   A friend of yours has bought cough drops for you because you had requested it. Ask your friend whether has got the cough drops.
   Potential answer: ‘Have you got the cough drops (I assume so)?’

2. *Un amic teu havia d’anar a la biblioteca. Li has demanat que aprofitant el viatge et tragués un llibre que tu havies de mester. Quan arriba li demanes si ha tret el llibre.*
Potential answer: ‘M’has tret el llibre?’

A friend of yours has had to go to the library and you have asked him to take a book out of the library for you. When s/he arrives, ask her/him whether s/he has got the book.
Potential answer: ‘Have you brought me the book (I assume so)?’

3. Tu i un amic teu estau a punt de partir de casa per anar a una excursió. Ho heu preparat tot junts i li havies demanat que ell agafàs la bossa amb les mandarines. Just abans de partir, li demanes si ha agafat la bossa amb les mandarines.

Potential answer: ‘Has agafat la bossa de mandarines?’

‘You and a friend of yours are about to leave to go on a trip. You have arranged everything together beforehand and agreed that s/he would bring a bag of tangerines. Before leaving, ask her/him whether s/he has brought the bag of tangerines.

Potential answer: ‘Have you brought the bag of tangerines (I assume so)?’

Note that the background situation specifies the kind of information that the speaker has about the answer to the question. For example, in the information-seeking questions there is no previous information about whether the interlocutor has or does not have what is requested by the speaker. By contrast, in the confirmation-seeking questions, it is assumed that the answer to the question will be affirmative because the speaker had previously required of the interlocutor a specific action and now s/he wants to check whether the hearer has carried out the task or not (believing that s/he has indeed done so).

We elicited a total of 6 contours x 21 subjects to yield a total of 126 contours.

2.3. Procedure

Descriptions of the prompt situations were read aloud randomly to the participants by native speakers of each variety. Speakers were then asked to respond appropriately to the situation as spontaneously as possible. The language register used by the interviewer was always that of a casual conversation between interlocutors sharing a high degree of familiarity with each other. Questions were recorded only once but when a speech disfluency such as “um” or “er”, breaks or irregularities occurred, the prompt situation was described a second time at the end of the full interview session and the subject’s response recorded again.

This type of intonation survey is also known in pragmatics as a DCT (Nurani 2009). The DCT has important advantages (Kasper & Dahl 1991, Cohen 1996, Beebe & Cummings 1996 among others): it allows researchers to collect a large

---

27 We are particularly indebted to Ignasi Mascaró, Pilar Prieto and Francesc Torres-Tamarit for help with the collection of and comments on the data for Minorcan, Northern Central and Ibizan/Formenteran Catalan.
amount of data in a short time, it is able to create model responses which are likely to occur in spontaneous speech and it provides stereotypical responses for a socially appropriate context.

Subjects were recorded on a TASCAM TEAC Professional Division digital audio tape recorder using a SHURE PG81 microphone. The recorded utterances were digitized at a 44100 Hz sample rate using Goldwave software.

2.4. Analysis

The intonational analysis was conducted using a notational method based on the AM framework (Beckman & Pierrehumbert 2008, Jun 2005, Ladd 1996, and Pierrehumbert 1980, among others) as applied to Catalan, that is, the Cat_ToBI (Prieto 2006, Prieto & Vanrell 2007, Prieto et al. 2009, Prieto in press, Aguilar et al. (coords.) 2009-2011) (see the general Introduction for an overview about the system). In addition to the general tonal units proposed by the Cat_ToBI, the upstep diacritic (¡) was used to mark a difference in tonal height that is systematic and relevant in marking a specific pragmatic meaning.

3. Results

Results obtained for each of the four dialectal areas (Central Catalan, Majorcan Catalan, Minorcan Catalan and Ibiza/Formenteran Catalan) and two types of questions (information- and confirmation-seeking questions) are described below. The percentages show a series of general tendencies that allow us to identify the preferred pattern—described schematically here using the Cat_ToBI system. The fundamental frequency contours shown for each dialectal variety, namely information- and confirmation-seeking questions, were produced by the same speaker so that these contours were directly comparable.

3.1. Central Catalan

Central Catalan speakers use both the rising pattern (L* HH%) and the falling pattern (H+L* L%) without distinction when uttering information- and confirmation-seeking questions. However, it was noted that in the case of information-seeking questions there is a clear preference for the rising pattern, with 92% rising contours versus 8% falling contours. By contrast, in confirmation-seeking questions, the rate of falling contours is 23% while that of rising contours is 75%. Thus, our interpretation is that while the rising pattern is more common in information-seeking questions, both rising and falling patterns can be found in confirmation-seeking questions. The rising pattern is characterized by a low pitch accent (L*), aligned with the nuclear syllable (last accented syllable of the sentence), followed by rising boundary tones (HH%). The falling pattern is produced
with a falling bitonal accent (H+L*), a high tone associated with the preaccentual syllable and a low tone aligned with the nuclear syllable, and a low boundary tone. We can observe the difference between these two different contours in Figure 3. Both contours were produced by the same speaker from Sabadell in Barcelona province.

![Waveforms and fundamental frequency contours](Chap3Fig3top.wav) and (Chap3Fig3bottom.wav), as produced by a Central Catalan speaker.

**Figure 3:** Waveforms and fundamental frequency contours of the information-seeking question (upper panel) *TENI mandarines?* ‘Do you have any tangerines?’ and the confirmation-seeking question (bottom panel) *QUE has agafat les mandarines?* ‘Have you taken the tangerines?’ as produced by a Central Catalan speaker.

The rising contour (upper panel) is characterized by the presence of a low tone, aligned with the syllable *-ri* (*mandarines*). The boundary tones located at the postnuclear region are rising. The falling contour (bottom panel) presents a falling pitch accent, associated with the syllable *-ri* (*mandarines*). The final boundary tones are falling.

Previous work such as Prieto (2002b) and Prieto and Rigau (2007) have already reported this kind of contrast in Central Catalan.
3.2. Majorcan Catalan

In Majorcan Catalan, 97% of both information- and confirmation-seeking questions were produced with a falling nuclear accent H+L* L%. Interestingly, it seems that in Majorcan Catalan the difference between information- and confirmation seeking questions is marked by means of a different tonal height associated with the leading tone H. Thus, 60% of the information-seeking questions were realized with an extra high tone associated with the preaccentual syllable, while 64% of the confirmation-seeking questions presented a falling accent H+L* whose H tone was not extra high but simply high. In other words, even though both types of questions are characterized by the same falling nuclear accent, what distinguishes one meaning from the other is that in the case of information-seeking questions the leading H tone is upstepped (¡H+L*) while in confirmation-seeking questions the leading H is non-upstepped (H+L*). Low boundary tones are found for both question types. Figure 4 shows this tonal scaling difference related to the leading tone H produced by a speaker from Algaida in Mallorca.

Note that although both contours have the same nuclear configuration, that is, a falling nuclear accent aligned to the syllable -ri- (mandarines) followed by a low boundary tone, in the specific case of the information-seeking question (upper panel), the preaccentual syllable -da- (mandarines) is approximately 100 Hz (5.4 semitones) higher than in the confirmation-seeking question. This difference in tonal height between information- and confirmation-seeking questions in Majorcan Catalan was previously observed in Payà and Vanrell (2005), in which yes-no questions collected following the Map Task methodology were prosodically and pragmatically analyzed.
Figure 4: Waveforms and fundamental frequency contours of the information-seeking question (upper panel) *Teniu mandarines?* ‘Do you have any tangerines?’ [Chap3Fig4top.wav] and the confirmation-seeking question (bottom panel) *Has agafat sa bossa de ses mandarines?* ‘Have you taken the bag of tangerines?’ [Chap3Fig4bottom.wav], as produced by a speaker of Majorcan Catalan.

3.3. **Minorcan Catalan**

Like in Majorcan Catalan, in Minorcan Catalan a difference in pitch scaling on the leading H tone of the H+L* nuclear pitch accent is the main cue to distinguish confirmation- from information-seeking questions. Thus, for information-seeking questions, 71% of the contours were realized with an extra high leading tone while 14% of the contours contain a non-upstepped high associated with the preaccentual syllable. Interestingly, in confirmation-seeking questions, the percentage of contours realized with a non-upstepped leading H tone increases and it is about 32% (while in information-seeking question it is just 14%) versus 66% of contours whose preaccentual syllable is extra high. The main difference between Minorcan and Majorcan Catalan is that while in Majorcan final boundary tones are always low, in Minorcan they are high for both information- and confirmation-seeking questions. Hence, although both varieties use the same strategy to mark the difference between the two question types, what distinguishes one variety from another is the use of different boundary tones. Figure 5 illustrates the contrast produced by a speaker from Maó in Minorca.
Figure 5: Waveforms and fundamental frequency contours of the information-seeking question (upper panel) *Que tens un caramel·lo?* ‘Do you have a cough drop?’ [Chap3Fig5top.wav] and the confirmation-seeking question (bottom panel) *Que m’has tret es llibre?* ‘Have you brought me the book?’ [Chap3Fig5bottom.wav], as produced by a speaker of Minorcan Catalan.

In both contours a falling nuclear accent is aligned either with the syllable *-mel-*(caramel-lo) (upper panel) or with the syllable *lli-* (llibre) (bottom panel). Also for both types of questions, high boundary tones are observed. The basic difference lies in the pitch height of the preaccentual syllables (*-ra-* of caramel-lo and *és*), which in the specific case of the information-seeking question is 120 Hz (5.5 semitones) higher.

3.4. Ibizan and Formenteran

Ibizan and Formenteran Catalan combine the strategy of using a different nuclear configuration, specifically different boundary tones like in Central Catalan (see 3.1.), with the strategy used by Majorcan and Minorcan Catalan in which the pitch height of the preaccentual syllable is the main cue to distinguishing the question types under study (see section 3.2. and 3.3. of this chapter). 68% of information-seeking contours were produced with high boundary tones while just 50% of the confirmation-seeking contours presented this pattern. Regarding low boundary tones, 11% of the information-seeking questions were produced with this pattern while the percentage increased to
22% in confirmation-seeking questions. All in all, it seems that Ibizan/Formenteran Catalan speakers have two strategies: the first one is based on the use of different boundary tones (H+L* HH% for information-seeking questions and H+L* L% for confirmation-seeking questions) and the second one involves a difference in pitch height related to the leading tone H (namely ¡H+L* HH% vs. H+L* HH%) while the boundary tones remain the same. The upstepped leading H tone rate for information-seeking questions was 18% while for confirmation-seeking questions it was 11%. The proportion of non-upstepped leading H tones for information- and confirmation-seeking questions is very similar (about 30% for each question type). Figure 6 illustrate the first strategy, that is, that in which different boundary tones are used (both contours were produced by a subject from Sant Ferran in Formentera) and Figure 7 shows the contrast in which a difference in tonal scaling of the leading tone H is at work (this subject is from the town of Ibiza on the island of the same name).

As can be observed, both contours have the same nuclear pitch accent, that is, a falling accent H+L* aligned with the syllable -ri- (mandarines). The difference between

![Waveforms and fundamental frequency contours of the information-seeking question (upper panel) Teniu mandarines? ‘Do you have any tangerines?’ [Chap3Fig6top.wav] and the confirmation-seeking question (bottom panel) Has agafat sa bossa de ses mandarines? ‘Have you taken the bag of tangerines?’ [Chap3Fig6bottom.wav], as produced by a speaker of Formenteran Catalan.](image_url)
the two contours is based on the F0 trajectory during the postnuclear syllables. In information-seeking questions a rising trajectory is observed, while in confirmation-seeking questions it falls. This is the same phenomenon that was also seen in Central Catalan.

We observe in Figure 7 that both contours apparently contain the same nuclear configuration, namely, a falling pitch accent aligned with the syllable -mel- (caramel·los) that precedes a final rising movement. Thus, the distinction between information- and confirmation-seeking questions is conveyed by a difference in tonal height of the preaccentual syllable, which in the specific case of Figure 7 is on the order of 50 Hz (2.3 semitones). Recall that both Majorcan and Minorcan Catalan make a similar use of the tonal scaling associated with the preaccentual syllable.

**Figure 7:** Waveforms and fundamental frequency contours of the information-seeking question (upper panel) *Que tens un caramel·lo? ‘Do you have a cough drop?’* [Chap3Fig7top.wav] and the confirmation-seeking question (bottom panel) *Que m’has dut es caramel·los? ‘Have you brought me the cough drops?’* [Chap3Fig7bottom.wav], as produced by a speaker of Ibizan Catalan.
4. Discussion and conclusions

The purpose of this production study was to analyze the interrogative intonation patterns of confirmation- and information-seeking question in several varieties of Central and Balearic Catalan. The data was collected by means of an intonation survey that allowed us to obtain a wide range of interrogative contours related to the different intentions that can underlie these interrogative contours. Regarding the syntactic structure, particular interest was placed on VO interrogative sentences. In such sentences, since the subject was not present, the question type could not be determined on the basis of whether the subject was dislocated or not (see section 1.3. of this chapter). Results from these analyses show that there is dialectal variation in the way in which speakers mark the difference between information- and confirmation-seeking questions. Thus, varieties of Central Catalan mark the distinction between these two question types by means of different nuclear configurations; other varieties such as Majorcan and Minorcan Catalan use the pitch height of the leading tone H for this purpose, whereas Ibizan and Formenteran Catalan use both strategies, that is, they have a contrast in which different boundary tones are essential but they can also use the tonal scaling of the preaccentual syllable as a main predictor. Table I summarizes the possible nuclear configurations that have been found for each dialectal variety (Central Catalan, Majorcan Catalan, Minorcan Catalan, Ibizan/Formenteran Catalan) and each question type (information- versus confirmation-seeking questions) under study. For each dialectal variety and type of question: a schematic representation of the tonal movements from the preaccentual syllable is shown, with the corresponding Cat_ToBI labeling below each diagram.

<table>
<thead>
<tr>
<th></th>
<th>Central Catalan</th>
<th>Majorcan Catalan</th>
<th>Minorcan Catalan</th>
<th>Ibizan/Formenteran Catalan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information-</td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
<td><img src="image4" alt="Diagram" /></td>
</tr>
<tr>
<td>seeking questions</td>
<td>L* HH%</td>
<td>H-L* L%</td>
<td>H-L* HH%</td>
<td>H-L* HH%</td>
</tr>
<tr>
<td>Confirmation-</td>
<td>H-L* L%</td>
<td>H-L* L%</td>
<td>H-L* HH%</td>
<td>H-L* HH%</td>
</tr>
<tr>
<td>seeking questions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table I:** Schematic representation of the nuclear configuration of information- (upper panel) and confirmation-seeking yes-no questions (bottom panel) for Central Catalan, Majorcan Catalan, Minorcan Catalan and Ibizan/Formenteran Catalan. Shaded boxes represent accented syllables.

Our results for Central Catalan (taking into account only boundary tones) and Ibizan/Formenteran Catalan (specifically the contrast that relies on the use of different boundary tones) agree with those of Kügler (2003), since like in Leipzig German the
variation we found in final boundary tones is related to the speakers’ epistemic disposition towards a proposition. Thus, low boundary tones signal that the speaker expects a particular answer based on belief, world knowledge or discourse context, while rising boundary tones indicate that s/he has no specific bias with respect to the answer.

The results from Majorcan and Minorcan Catalan correspond to the results of Grice and Savino (1997, 2003a, 2003b) and Armstrong (2010) since the difference between questions for which the speaker has no particular bias with respect to the answer s/he expects and those for which the speaker has some bias is conveyed through a different nuclear pitch accent choice.

With respect to the methodology used here, we claim that the DCT can be profitably used to investigate intonational categories as well as to study how these intonational categories interact with pragmatics. The evidence comes from cues other than tonal which have the potential for signaling speaker (un)certainty such as speech rate (information-seeking questions tend to be slower than confirmation-seeking questions), the use of right-dislocations in confirmation-seeking questions (i.e., L’has agafada, sa bossa de ses mandarines? ‘Have you taken it, the bag of tangerines?’) or the use of “uh”, “um” or pauses in information-seeking questions (Eh... tens un caramel·lo? ‘Uh... do you have a cough drop?’). We conclude, thus, that the DCT is an appropriate method to control for the information that a speaker has based on beliefs, expectations, world knowledge or information that is available discoursively.

Our results challenge the general assumption of the standard Autosegmental-Metrical (AM) model (Pierrehumbert 1980, Pierrehumbert & Beckman 1988), in the sense that pitch height variation has been assumed to have a paralinguistic function that corresponds to an increase or decrease in the expression of emphasis or prominence. However, this chapter presents positive evidence that pitch height variation can also convey linguistic distinctions. For that reason, our next step will be to test through perceptual tasks whether listeners, in the specific case of Majorcan Catalan, can vary in their expectation about the answer simply on the basis of a different tonal height associated with the preaccentual syllable.
CHAPTER 4

Intonation as an encoder of speaker certainty: perception data

This chapter focuses on one of the Balearic varieties analyzed in the previous chapter, Majorcan Catalan. In this specific variety, the differences between questions for which the speakers has no particular bias with respect to the answer s/he expects and those for which the speaker has some bias are conveyed through a difference in the pitch height of the leading tone H. By means of three behavioural tasks, we seek to determine whether this difference in pitch scaling on the leading H tone of the H+L* nuclear pitch accent is the main cue used by the listeners to distinguish confirmation-seeking yes-no questions from information-seeking yes-no questions. A second goal of this chapter is to assess and compare different experimental methods such as the congruity task, the rating task and the Categorical Perception paradigm which was described in Chapter 1 of this dissertation.

1. Introduction

As is well known, languages may rely on a variety of strategies for the expression of the speakers’ epistemic disposition towards a proposition, and specifically the speakers’ degree of certainty about it, namely lexical marking (for example, by the choice of modal adverbs such as surely), morphological marking (in languages with morphemic/affixal marking of epistemic modality), prosody (different choices of intonation contours) or facial and body gestures (Palmer 2001). In the prosody literature, there is ample evidence that languages may distinguish prosodically between information-seeking questions (questions for which the speaker has no particular bias with respect to the answer s/he expects) and confirmation questions (questions for which the speaker has some bias based on belief, expectations, world knowledge or information that has become available in the discourse context; see Bolinger 1989; Büring & Gunlogson 2000). Studies that have applied the Map Task technique (Carletta et al. 1995, Grice & Savino 1997) for collecting interrogative data have referred to information- and confirmation-seeking questions as queries and checks.

In languages like English, these two different types of questions can display different syntactic properties. While canonical information-seeking questions in English are characterized by subject-verb inversion and the presence of an auxiliary verb (Did Jim leave early?), canonical confirmation-seeking questions tend to exhibit declarative syntax (Jim left early?). In Romance languages, the relationship between question
type and surface syntactic form is less clear because these languages have a freer word order and offer the possibility of not expressing the syntactic subject. Hence, because syntactic aspects alone do not always allow a listener to differentiate between question types, “the decision about the informational-status of a certain question can rely heavily upon prosodic features as the primary cues” (Grice & Savino 2003a).

In some Romance languages, speakers can signal the distinction of (or function of) information- and confirmation-seeking questions through prosodic means by a specific pitch accent choice like in Bari Italian (Grice and Savino 1997) or Puerto Rican Spanish (Armstrong 2010). Similarly, Germanic languages other than English are able to convey this contrast through intonation. Thus, according to Kügler the differing percentages in the appearance of the final rise found in Haan (2001) for yes-no questions could be due to the speakers’ expectation of an answer. Likewise, in Leipzig German (Kügler 2003) the interaction between intonation and information structure is conveyed through the use of different boundary tones.

All of these studies have relied on production data to claim that the distinction between these two question-types is cued by an intonational contrast. This study aims to further explore the extent to which intonational contrasts cue this distinction from a perceptual point of view. Results in the previous chapter revealed that while Central and Ibiza/Formenteran Catalan speakers use boundary tones to mark the distinction between confirmation- and information-seeking yes-no questions, Majorcan and Minorcan Catalan speakers use a different type of nuclear pitch accent. The present investigation focuses on the contrast found in Majorcan Catalan, a variety in which the difference between information- and confirmation-seeking yes-no questions is marked through the use of two nuclear pitch accents, ¡H+L* and H+L* (see Figures 1 and 2). Both question types are characterized by a falling nuclear pitch accent H+L*, that is, a H leading tone aligned with the preaccentual syllable and a L* tone associated with the last stressed syllable. While the information-seeking yes-no question has a higher H leading tone (Figure 1, right panel), the confirmation-seeking yes-no question has a non-upstepped H leading tone (Figure 1, left panel). For the transcription of the examples, we use the most recent version of Cat_ToBI (Prieto et al. 2009, Prieto in press, Aguilar et al. (coords.) 2009-2011).

Figure 2 illustrates the difference between the fundamental frequency contours of the confirmation- and information-seeking versions of the yes-no question Teniu mandarines? ‘Do you have any tangerines?’ (upper and lower panels, respectively), as produced by a Majorcan Catalan speaker. While the prenuclear part is almost identical in the two pitch contours, the preaccentual syllable -da- (manda-rines) of the information-seeking question is realized with a significant increase in pitch range, compared to the confirmation-seeking question. The main difference between the two intonation contours thus lies in the tonal height of the leading tone H.29

but she argues that the questioning function of declaratives is a matter not only of syntax but also of the interaction between syntax, intonation and context.

29 Observe that the difference in tonal height of the H leading tone also triggers an intonational difference in the phonetic realization of the prenuclear part of the two intonation patterns. Thus, while confirmation-seeking questions show a steady high tonal movement which extends from the beginning of the sentence
Figure 1: Schematic representation of the nuclear accents found in Majorcan confirmation- (left) and information-seeking (right) yes-no questions with the two penultimate syllables of the target word mandarines (‘tangerines’). White boxes represent consonants and shaded boxes represent vowels. Boxes representing the nuclear accent-bearing syllable are marked with thicker edges.

Figure 2: Waveforms and fundamental frequency contours of the confirmation- (upper panel) [Chap4Fig2top.wav] and information-seeking (bottom panel) [Chap4Fig2bottom.wav] versions of the yes-no question Tenui mandarines? ‘Do you have any tangerines?’, as produced by a Majorcan Catalan speaker.

As we noted above, the intonation literature has mainly focused on the existing dichotomy between information-seeking yes-no questions, or questions in which the speaker has no particular expectations about the answer to question, and confirmation-to the end of the preaccentual syllable, information-seeking questions exhibit a well-defined rising slope that goes from the beginning of the sentence to the end of the preaccentual syllable.
seeking yes-no questions, or questions in which the speaker is asking for acknowledgement of mutually shared information. In line with this, recent studies in pragmatics have proposed that there exists a gradient scale based on the speaker’s knowledge on the one hand and the presumed knowledge in possession of the hearer on the other. As it was stated in the previous chapter, Escandell (1993, 1998) distinguishes between: neutral questions, located at the top of the scale, in which the speaker seeks to obtain information that s/he lacks, exam-style questions, which imply a maximal degree of knowledge on the part of the speaker and no presupposition (located at the other opposite extreme of the scale of knowledge on the part of the interlocutor) and confirmation-questions (located at some intermediate point in the scale) in which the speaker’s ignorance is not total since s/he already knows the answer to her/his question and thus merely seeks confirmation of her/his hypothesis. One of the aims of this chapter will be to test this hypothesis and one of the experiments included in this work (the rating test) has is thus intended to test whether Catalan listeners can distinguish between four levels of knowledge presupposition depending on a variety of acoustic and semantic cues in a speaker’s output, thus provinding empirical support for Escandell’s model.

The pitch accent contrast investigated involves a difference in pitch height (namely ¡H+L* vs. H+L*). Within the standard Autosegmental-Metrical (AM) model (Pierrehumbert 1980, Pierrehumbert & Beckman 1988), pitch height variation has been assumed to have a paralinguistic function that corresponds to an increase or decrease in the expression of emphasis or prominence. However, more recent studies have shown that differences in pitch height can also convey linguistic distinctions with respect to pitch accents (Ladd & Morton 1997, Chen 2003, Calhoun 2004 for English; Prieto 2004 for Spanish; Savino & Grice 2007, 2011 for Bari Italian; Borràs-Comes et al. 2010 for Central Catalan) but also in the boundary tone domain (Beckman & Ayers-Elam 1997, for English; Arvaniti & Baltazani 2005, for Greek; Grice et al. 2005, for German; Beckman et al. 2002, for Spanish; Lee 2003, for Korean; Frota in press-b for European Portuguese)

One of the overarching goals of this chapter will be to provide a test case for comparing the outcome of different perceptual tasks for the study of intonational phonology. The Categorical Perception paradigm (CP) has been one of the most commonly used methods to examine the nature of tonal contrasts and has been applied to differences in peak alignment (Kohler 1987, D’Imperio & House 1997, Chen 2003, Gili-Fivela 2008, 2009) and differences in pitch height in both tonal languages (Francis et al. 2003, Francis & Ciocca 2003) and intonational languages, for boundary tones (Remijsen & van Heuven 1999, Post 2000, Schneider & Linfert 2003, Cummins et al. 2006, Falé & Faria 2006, Schneider et al. 2009) as well as for pitch accents (Ladd & Morton 1997, Chen 2003, Gili-Fivela 2008, 2009, Dilley 2010). The application of the CP paradigm has generally uncovered the presence of asymmetries in discrimination results. Asymmetries in tonal perception occur when the discrimination of a tonal change presented in one direction is easier compared to the same change presented in

---

30 See the Introduction for a literature review about the nature of pitch height contrasts.
the reverse direction (see Chapter 1 for a review of the literature about asymmetries in tonal perception).

A general finding in the articles that have applied the CP to the study of intonation contours is the lack of a clear peak in discrimination functions in contrasts that involve boundary tones (Remijsen & van Heuven 1999, for Dutch; Cummins et al. 2006, for English; Falé & Hub Faria 2006, for European Portuguese) or pitch accents (Ladd & Morton 1997, for English; Dilley 2010, for American English; Gili-Fivela 2008, 2009, for Italian). The explanations for this lack of a clear peak in discrimination functions are varied. Some researchers (Remijsen and van Heuven 1999) argue that this absence of a discrimination peak could be due to the fact that intonational units, like vowel phonemes, are encoded over long time intervals, other researchers (Gili-Fivela 2009) argue that the mixed results obtained could be due to a difference in the pragmatic categories under study, while other scholars (Chen 2003, Kohler 2006, Niebuhr & Kohler 2004, and Frota in press-a) claim the unsuitability of discrimination tasks due to the fact that they are too focused on the acoustic perceptive properties rather than on tonal categories. All in all, it seems that the absence of a clear peak should be related not to the nonexistence of a phonological distinction but rather to the unsuitability for different reasons of applying the CP paradigm to intonational contrasts.

Given this methodological problem, other alternatives have been proposed such as the use of Reaction Time (RT) measures, together with the results of semantically motivated identification tasks (Chen 2003, Falé & Hub Faria 2006, Savino & Grice 2007, 2011), semantically motivated identification and discrimination tasks (Frota in press-a), semantically motivated tasks like congruity and identification tasks (along with Reaction Time measures) or triple answer identification tasks (Borrás-Comas et al. 2010). All in all, it seems that the general proposal would be to prevent the activation of the “auditory mode” using tasks that ensure contextual and semantic accessibility. In this chapter we will provide results from a variety of perception tasks that have been previously used in the intonation literature. These tasks will allow us to compare their outcome with the same test case and to evaluate their suitability for the investigation of intonational contrasts.

Our main goal in this chapter is to investigate whether Majorcan Catalan listeners use the height of the leading tone in the H+L* nuclear accent as the primary perceptual cue to distinguish the contrast between information- and confirmation-seeking questions. In order to do so — while remaining aware of the methodological issues that may arise when testing the nature of intonational contrasts — we analyze subjects’ responses to three types of perception experiments, namely a congruity test, a rating test and a test set based on the application of the classical CP paradigm. These techniques have been previously used to explore the perceptual processing of intonation contrasts (the congruity task in Rathcke & Harrington 2010, Borràs-Comes et al. 2010, Crespo-Sendra 2011, the rating test in Swerts & Krahmer 2008, Nadeu & Prieto 2011, the CP paradigm applied to boundary tones in Remijsen & van Heuven 1999, for Dutch; Schneider & Linftert 2003, for German; Cummins et al. 2006, for English; Falé & Hub Faria 2006, for European Portuguese; or the CP paradigm applied to pitch accents in Kohler 1987,
This fourth chapter is organized as follows. Section 2 contains the methodology used in these three perception experiments, section 3 presents the results for each of the three experiments and, finally, section 4 discusses the major findings of this work and presents its overall conclusions.

2. Methodology

2.1. Participants and general procedure

In order to investigate whether Majorcan Catalan listeners use the height of the leading tone in the H+L* nuclear accent as the primary perceptual cue to the contrast between information- vs. confirmation-seeking questions, three perception experiments were conducted: a congruity task in which participants were asked to evaluate the degree of appropriateness/congruity of the target utterances with different intonation patterns to different pragmatic contexts, a rating task in which the participants were asked to rate the perceived degree of certainty of utterances that contained both types of falling accents and, finally, an identification task based on the CP paradigm, which included Reaction Time measurements. In this last task, listeners were asked to identify the meaning of an utterance when its intonation was varied across an acoustic continuum extending between the two target intonation contours. This was followed by a single discrimination test to assess whether the shift from one category to the other found in the identification results corresponded to a clear peak in the discrimination results.

Twenty speakers of Majorcan Catalan aged between 16 and 35 participated in this experiment. None of them reported a history of hearing disability.

The presentation of the stimuli was prepared using E-prime version 1.2 (Psychology Software Tools Inc., 2009) and lasted a total of 50 minutes. Subjects were seated at a laptop in a quiet room and the stimuli were played back through headphones. All of them participated in the full sequence of experiments and performed the various tests in the same order: congruity test, rating test, identification test and discrimination test.

Since we were interested in Reaction Time measurements, listeners were instructed to maintain their hands near the keyboard and press the keys as fast as they could. There was no break between the different perceptual tests, except when time was needed to change the prompt script.

2.2. Congruity test methodology

Congruity tests have been applied successfully to the analysis of intonation contrasts (Rathcke & Harrington 2010, Borràs-Comes et al. 2010). The advantage of using such a task is that it allows us to evaluate the degree of perceived appropriateness of target intonation patterns to different pragmatic contexts.
Two similar everyday contexts were used for establishing the contextual appropriateness of the confirmation/non-confirmation meanings. In both contexts listeners had to imagine that they had just entered a store and wanted to ask the shopkeeper whether s/he had tangerines. They then heard audio recordings of the ensuing dialogue. The two dialogues were as follows:

Information-seeking question dialog
Només has mester mandarines, però no saps si en tenen (voice-over).
‘You only need tangerines but you do not know whether they have them or not.’
Speaker A: —Bon dia, teniu mandarines?
—‘Good morning. Do you have any tangerines?’
Speaker B: —Eh… sí, ara vènc des Mercapalma i n’he dutes.
—‘Er… yeah, I’ve just come from the wholesale market and have brought some.’

Confirmation-seeking question dialog
Només has mester mandarines i saps que sempre en tenen (voice-over).
‘You only need tangerines and you know that they always have them.’
Speaker A:—Bon dia, teniu mandarines?
—‘Good morning. Do you have any tangerines (I suppose so)?’
Speaker B: —Clar! Ves si en tenim! Com sempre!
—‘Of course we have tangerines, as always!’

In half of the recorded dialogues, the question Teniu mandarines? was consistent with the pragmatic context. In the other half, the information-seeking question intonation was inserted into the confirmation-seeking context and vice versa. The splicing was performed using Goldwave software. The two panels in Figure 3 show the waveform, F0 contour and Cat_ToBI transcriptions for the unmanipulated target confirmation-seeking question dialogue (top panel) and for the spliced stimulus inserted into the confirmation-seeking question dialogue (bottom panel). The two target auditory stimuli Bon dia, teniu mandarines? for this experiment as well as the ensuing dialogues were recorded by two native speakers of Majorcan Catalan. The initial sentences contextualizing the kind of information that the speaker has about the answer of the questions (e.g., ‘You only need tangerines but you do not know whether they have them or not.’) was produced at a lower volume compared to the dialogues so that it resembled a voice-over.
Figure 3: Waveform, F0 contour, and Cat_ToBI transcription for the unmanipulated target confirmation-seeking question dialogue (top panel) [Chap4Fig3top.wav] and for the spliced stimulus inserted in the confirmation-seeking question dialogue (bottom panel) [Chap4Fig3bottom.wav].

Thus, the test consisted of two yes-no questions whose intonation was coherent with the pragmatic context and two yes-no questions whose intonation was not coherent with the pragmatic context. Listeners had to answer whether they regarded the intonation of
the yes-no questions as “congruent” with the pragmatic context (by pressing the “C” key) or “incongruent” (by pressing the “I” key). The test consisted of a total 40 trials (2 congruent dialogues + 2 incongruent dialogues x 5 repetitions x 2 blocks).

2.3. Rating test methodology

The goal of this task was to test whether Majorcan Catalan listeners are able to identify different degrees of speaker presupposition in an utterance depending on prosodic and syntactic cues (Escandell 1993, 1998). Escandell’s hypothesis is that there is a continuum of varying degrees of presupposition that goes from neutral questions, at one extreme, which imply a minimal knowledge on the part of the speaker and a maximal presumption of knowledge on the part of the addressee, to exam-style questions at the other extreme in which it is the speaker rather than the addressee that has maximal knowledge about the truth value of the question (exam-style questions are not analyzed in this chapter). In Catalan confirmation-seeking questions we also often find confirmation marks such as polarity adverbs (no, oi), the noun veritat or the particle eh (Cuenca 1997, Hernanz & Rigau 2006, Prieto & Rigau 2007).31 Our prediction was that speakers would be able to distinguish among these different degrees of presupposition by relying on prosodic cues such as the difference in pitch height related to the leading tone in information- and confirmation-seeking questions or on morphosyntactic cues such as the presence of confirmation marks in tag questions. We also hypothesized that even though both tag questions and confirmation questions without confirmation marks seek to confirm a hypothesis, they differ in the degree of presupposition held by the speaker. For that reason, the materials used for the rating test were an information-seeking question, a confirmation-seeking question without confirmation marks, a tag question and a declarative. The declarative sentence was intended to act as “control stimulus” in the sense that it represents the maximal presupposition of an affirmative answer, given that it is an affirmative sentence.

In the rating test, listeners had to rate on a 4 point scale how certain the speaker was about whether her/his interlocutor had tangerines or not. The materials consisted of 4 different audio stimuli recorded by the author of this dissertation: an information-seeking question (Teniu mandarines? ‘Do you have any tangerines?’), a confirmation-seeking question (Teniu mandarines? ‘Do you have any tangerines (I suppose so)?’, with the appropriate intonation contour), a tag question (Teniu mandarines, no? ‘You have tangerines, don’t you?’), and a broad focus statement (Teniu mandarines ‘You have tangerines’). The broad focus statement acted as a control since the speaker is asserting that “the interlocutors have tangerines” and is therefore completely sure about this. The test consisted of 40 trials (4 stimuli x 5 repetitions x 2 blocks). Subjects had to press one of four possible options, namely “1” for ‘no idea’, “2” for ‘maybe yes’, “3”

31 Regarding the syntactic order of constituents, confirmation-seeking questions in Catalan can also have the subject in preverbal position and without dislocation (Rigau 2002, Prieto & Rigau 2007). As was stated in Chapter 3, it is also possible for confirmation-seeking questions not to present confirmation marks. In this case, the type of question will be triggered by either the syntactic order (only when the subject is expressed) or the intonation.
for ‘probably yes’ and “4” for ‘absolutely yes’, with number values reflecting the strength of certainty of a “yes” answer.

2.4. Tests based on the classical CP paradigm

Several studies have successfully applied the standard CP paradigm to the study of intonation contrasts, whether between boundary tones, for example, L% vs. H% contrasts that prototypically mark the difference between questions and statements or between pitch accents (see the general Introduction of this dissertation). In these investigations (whether they involved boundary tone or pitch accent contrasts), while identification curves show clear discrete effects, discrimination results are less clear-cut and generally show much weaker evidence for categorical perception. As a result, although claims are made of “categorical perception” for various contrasts, out of all the abovementioned articles only three really offer unarguable evidence of categorical perception, with a clear discrimination peak in the expected position (Kohler 1987, Remijsen & van Heuven 1999, Schneider & Linfert 2003). Even perception studies performed on tonal languages do not always show clear patterns of discrimination results (see Francis et al. 2003, for a review). One of the explanations given (see general Introduction) is that tonal contrasts, like vowel contrasts, may be perceived in a less categorical manner than consonantal contrasts.

One natural token of the information-seeking question version of Teniu mandarines? (‘Do you have any tangerines?’) and one token of the confirmation-seeking version of the same question were recorded again by the author of this dissertation, a native speaker of Majorcan Catalan, this time without the presence of the greeting bon dia (‘good morning’). The small acoustic differences between these two natural tokens that were not related to the nuclear region (e.g., initial pitch height and possible differences in speech rate) were neutralized. In the information-seeking question token the leading tone was 320 Hz while in the confirmation-seeking question token it was 210 Hz. A linear stylization of the rising-falling movement was carried out for each token. Three points were interpolated: a point L1 at the rising onset, a point H at the peak, and a point L2 at the falling offset. L1 was aligned in both tokens with the onset of the syllable -da- (mandarines ‘tangerines’), H with the offset of the vowel of the syllable -da- and L2 with the offset of the vowel of the syllable -ri- (mandarines ‘tangerines’). From these two base tokens, twenty stimuli were created by means of PSOLA synthesis in Praat (Boersma & Weenink 2009). Ten stimuli were created by shifting the peak downwards from the information-seeking question token in ten steps of 11.2 Hz each and conversely by shifting the peak upwards from the confirmation-seeking question token. Including the two base tokens, this yielded a total of 22 stimuli. The two graphs in Figure 4 schematically show the stimuli that were created. These stimuli were used for both the identification and discrimination tasks.
In the identification task subjects were asked to respond after each stimulus by indicating the answer the speaker expected as revealed by her/his intonation. They had to press the “N” key for Ni idea ‘no idea’ (if the speaker had no clear expectation or bias about the answer) or the “P” key for Per ventura ‘maybe yes’ (if the speaker seemed to be expressing some certainty or previous knowledge about the truth value of the answer). The materials for the identification task consisted of 110 trials (11 stimuli x 2 base tokens x 5 blocks).

The materials for the discrimination task consisted of pairs of the same stimuli that were used in the identification task. First 20 pairs of stimuli were created in low-high order, meaning that the peak of the second stimulus was always higher in pitch than that of the first stimulus (10 from the information-seeking question and 10 from the confirmation seeking-question). Then 20 high-low-ordered pairs of stimuli were created (again, 10 for each type of question). Finally, 22 pairs which contained identical stimuli were created. As they performed the task, listeners were asked to decide whether they heard each recorded pair of stimuli as “same” or “different”. The discrimination test consisted of 248 trials (10 low-high pairs + 10 high-low pairs + 11 identical stimuli pairs x 2 base stimuli x 4 blocks).

2.5. Statistical analyses

In this chapter non-parametric tests were used when the dependent variable in our data set was measured in a nominal scale (congruity test, rating test and identification/discrimination tests responses) and parametric tests when the dependent variable was continuous (RT measurements). Results were obtained by SPSS statistics software (SPSS for Windows). With respect to non-parametric tests, the Wilcoxon matched pairs signed rank test was used to: (1) compare the responses to two conditions (congruent vs. incongruent) for both information- and confirmation-seeking questions in the congruity test, (2) compare the identification rate between two conditions (e.g., stimulus 1 on the continuum vs. stimulus 2 on the continuum) in determining the location of the boundary shift for both question types, and (3) compare the two functions (low-high-ordered vs. high-low-ordered stimuli) for both question types in the...
discrimination test, and the Friedman test to test the differences between the scores obtained for each stimulus in the rating test. For the Wilcoxon matched pairs signed rank test, we report the test statistic, its significance and the effect size (see Chapter 1, section 2.4. for a detailed explanation about these values). For the Friedman test, the test statistic (denoted by $\chi^2$), its degrees of freedom and its significance are reported. The post-hoc Bonferroni correction was applied in cases in which multiple comparisons were performed. The univariate ANOVA run with the GLM procedure of SPSS was performed as parametric test for each type of question (information- and confirmation-seeking question) with RT measurement as the dependent variable, and with one between-subjects factor, namely Stimulus (11 steps in the pitch range continuum). The results from the univariate ANOVA test include the $F$-ratio, the degrees of freedom from which it was calculated and the significance value.

3. Results

3.1. Congruity test results

Figure 5 shows the rate of “congruent” responses to both congruent (black bars) and incongruent dialogues (grey bars), separated into information-seeking (left) and confirmation-seeking meanings (right). Recall that the key difference between the two interrogative types is in the height of the leading tone $H$, which in the case of an information-seeking question is upstepped (see Figure 2). The results revealed an average rate of 0.91 for “congruent” responses to congruent dialogues for both information-seeking and confirmation-seeking meanings. By contrast, the average rate of “congruent” responses in the incongruent dialogue was 0.12 and 0.09 for the information-seeking and confirmation-seeking meanings respectively. Results from the Wilcoxon matched pairs signed rank test revealed that the differences between the two conditions (congruent vs. incongruent) for both information- ($T = 1780, p < .001, r = -.57$) and confirmation-seeking question ($T = 1053, p < .001, r = -.55$) meanings were significant. These results indicate that listeners are extremely sensitive to the incongruent use of confirmation- and information-seeking questions.

3.2. Rating test results

Table I presents the ratings (in columns) for each stimulus (in rows) in the rating test with results for information-seeking question (Teniu mandarines? ‘Do you have any tangerines?’), the confirmation-seeking version of the same question, the tag question (Teniu mandarines, no? ‘You have tangerines, don’t you?’) and the statement (Teniu mandarines ‘You have tangerines’). Recall that listeners were asked to rate on a 4 point scale the likelihood that the speaker would obtain an affirmative answer to her/his question. The possible answers varied from less certainty (answer “1”) to more certainty (answer “4”) of getting a “yes” answer to the utterance. Since there was a listener who
admitting to having trouble remembering which number was related to the ‘no idea’
meaning and which one to the ‘absolutely yes’ meaning, the responses of this particular
speaker were systematically removed from the final data base. The final results
demonstrate that listeners do succeed in perceiving the prosodic cues related to the
contrast between information- and confirmation-seeking questions. This provides clear
evidence that listeners base their decisions about the truth value of the sentences on not
only morphosyntactic but also prosodic cues. Results of a Friedman test revealed that
the differences between the scores obtained for each stimulus were significant ($\chi^2(6) =
984.482, p < .05$).

![Figure 5: Rate of “congruent” responses to congruent (black bars) and incongruent dialogues (grey bars), separated by information-seeking question meaning (left) and confirmation-seeking question meaning (right).](image)

<table>
<thead>
<tr>
<th>ANSWER CATEGORY</th>
<th>No idea</th>
<th>Maybe</th>
<th>Probably yes</th>
<th>Absolutely yes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STIMULUS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information-seeking question</td>
<td>370</td>
<td>23</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Confirmation-seeking question</td>
<td>48</td>
<td>311</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td>Tag Question</td>
<td>1</td>
<td>10</td>
<td>373</td>
<td>17</td>
</tr>
<tr>
<td>Statement</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>380</td>
</tr>
</tbody>
</table>

Table I: Table presenting the ratings (in columns) for each stimulus (in rows).
All in all, our results indicate that the available knowledge/presumed knowledge scale proposed by Escandell (1993, 1998) has clear validity at the perceptual level.

3.3 Classical CP test results

3.3.1 Identification results

Figure 6 shows the identification rate for the continuum created from the confirmation-seeking question base token (black bars) and the information-seeking question base token (grey bars). The “identification rate” is defined as the number of “confirmation-seeking question” responses (in confirmation-seeking-question-based stimuli) or “information-seeking question” responses (in information-seeking-question-based stimuli). In Figure 6 we can see that the functions show the expected S-shape and hence that it is possible to change the perceived category simply by shifting the leading tone upwards or downwards (depending on the base category).

![Figure 6](image)

**Figure 6**: Identification rate for the continuum created from the confirmation-seeking question (black line) and information-seeking question base tokens (grey line).

In order to claim the discretness of a contrast, it is very important to determine the exact location of the boundary between the categories. If the location of the category boundary corresponds to the RT/discrimination peak, we will have evidence in favour of the categorical nature of the contrast tested. With the aim of getting the boundary between the two categories for the two curves obtained for the confirmation- and information-seeking question continua respectively, the set of data points was fitted to a logistic function in SPSS software (SPSS for Windows) (see Chapter 2, section 2.2.1. for an explanation about the formula and the procedure to calculate the location of the category boundary).

Thus, for the confirmation-seeking-question-based continuum, when $y$ equals 0.5 $x$ is 5.31, and for the information-seeking-question-based continuum, when $y$ equals 0.5 $x$ is 5.8. Thus the boundary is located between stimuli 5 and 6 for both continua (see Figure 6).
Table II shows the results of ten Wilcoxon matched pairs signed rank test comparing the identification score between adjacent stimuli (e.g., stimulus 1 vs. stimulus 2, stimulus 2 vs. stimulus 3, and so on) for each continuum. According to the CP paradigm, since the boundary is located between stimuli 5 and 6 for both continua, there should be significant differences between the response rates for these two stimuli. This is not the case for the continuum created from the confirmation-seeking base stimulus, as can be seen in Table II, since the differences between response rates for stimuli 5 and 6 are not significant. However, for the continuum created from the information-seeking base stimulus, the results are nearly what we would expect. Observe that we find statistical differences between the response rates for stimuli 5 and 6, and also for stimulus 4-5 and 3-4 (shaded values).

<table>
<thead>
<tr>
<th></th>
<th>Stim. 1-2</th>
<th>Stim. 2-3</th>
<th>Stim. 3-4</th>
<th>Stim. 4-5</th>
<th>Stim. 5-6</th>
<th>Stim. 6-7</th>
<th>Stim. 7-8</th>
<th>Stim. 8-9</th>
<th>Stim. 9-10</th>
<th>Stim. 10-11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confirmation yes-no questions</strong></td>
<td>( T = 2 )</td>
<td>( T = 0 )</td>
<td>( T = 28 )</td>
<td>( T = 81 )</td>
<td>( T = 50 )</td>
<td>( T = 19 )</td>
<td>( T = 32.5 )</td>
<td>( T = 30 )</td>
<td>( T = 11 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( p &gt; .05 )</td>
<td>( p &gt; .05 )</td>
<td>( p &gt; .05 )</td>
<td>( p &lt; .05 )</td>
<td>( p &gt; .05 )</td>
<td>( p &gt; .05 )</td>
<td>( p &gt; .05 )</td>
<td>( p &gt; .05 )</td>
<td>( p &gt; .05 )</td>
<td></td>
</tr>
<tr>
<td><strong>Information yes-no questions</strong></td>
<td>( T = 2 )</td>
<td>( T = 6 )</td>
<td>( T = 0 )</td>
<td>( T = 46 )</td>
<td>( T = 19 )</td>
<td>( T = 15 )</td>
<td>( T = 28 )</td>
<td>( T = 20 )</td>
<td>( T = 13.5 )</td>
<td>( T = 10.5 )</td>
</tr>
<tr>
<td></td>
<td>( r = -.04 )</td>
<td>( r = -.18 )</td>
<td>( r = -.21 )</td>
<td>( r = -.12 )</td>
<td>( r = -.15 )</td>
<td>( r = -.25 )</td>
<td>( r = -.04 )</td>
<td>( r = -.02 )</td>
<td>( r = -.14 )</td>
<td></td>
</tr>
</tbody>
</table>

**Table II:** Results of ten Wilcoxon matched pairs signed rank tests comparing the identification rate between adjacent stimuli for information- and confirmation-based stimuli.

When looking at standard error values of the mean for every stimulus (Table III), we observe that identification rates yield a higher standard error as they get closer to the most ambiguous rate, 0.5, corresponding to 5.31 for the confirmation-seeking-question-based continuum and 5.8 for the information-seeking-question-based continuum. Thus, stimuli 5 and 6 for each continua display the highest standard error values (shaded values). These results are by no means unexpected since they show that listeners agreed in their responses when listening to stimulus 1 and 11 because they represent the canonical categories, while this agreement decreases as the crossover point between the categories approaches.

<table>
<thead>
<tr>
<th></th>
<th>Stim. 1</th>
<th>Stim. 2</th>
<th>Stim. 3</th>
<th>Stim. 4</th>
<th>Stim. 5</th>
<th>Stim. 6</th>
<th>Stim. 7</th>
<th>Stim. 8</th>
<th>Stim. 9</th>
<th>Stim. 10</th>
<th>Stim. 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confirmation yes-no questions</strong></td>
<td>.032</td>
<td>.030</td>
<td>.040</td>
<td>.049</td>
<td>.057</td>
<td>.056</td>
<td>.054</td>
<td>.046</td>
<td>.043</td>
<td>.043</td>
<td>.033</td>
</tr>
<tr>
<td><strong>Information yes-no questions</strong></td>
<td>.027</td>
<td>.035</td>
<td>.028</td>
<td>.051</td>
<td>.057</td>
<td>.054</td>
<td>.048</td>
<td>.038</td>
<td>.039</td>
<td>.034</td>
<td>.032</td>
</tr>
</tbody>
</table>

**Table III:** Standard error of the mean of the identification rate for information- and confirmation-based stimuli.

### 3.3.2. Reaction Time results

Reaction Time (RT) measurements have been proposed to be a good alternative to the discrimination task in testing the hypothetical discreteness of a contrast (Pisoni & Tash 1974, Chen 2003). Chen (2003, p. 100) claims that “short RTs for within-category identification and long mean RTs for across-category identification are essential properties of linguistically real identification categories”. Figure 7 plots averaged RT
responses (in ms) for all subjects. The black line shows the reaction times for the confirmation-seeking-question-based continuum and the grey line shows the RTs for the information-seeking-question-based continuum (RT measurements did not include the duration of the stimulus).

![Graph showing averaged reaction time (RT) responses for all subjects.](image)

**Figure 7:** Averaged Reaction Time (RT) responses (in ms) for all subjects.

A clear peak in RT measurements for the information-seeking question continuum is obtained at stimulus 6, but not for the confirmation-seeking question continuum, since a RT plateau extends from stimulus 4 to stimulus 6. The location of the peak (in the case of the information-seeking-question continuum) and the plateau (in the case of the confirmation-seeking-question continuum) coincides with the boundaries calculated from the fitted logistic curves. The results of univariate ANOVAs, run with the GLM procedure of SPSS, indicated that there were statistically significant differences between stimuli with respect to RT for the confirmation-seeking-question continuum ($F(10, 700) = 1.918, p < .05$) but not for the information-seeking question continuum ($F(10, 687) = 1.439, p > .05$). We hypothesize that the absence of statistical differences in the case of the information-seeking-question continuum can be explained by subject variability. Since several studies seem to show differences in tonal perception accuracy depending on the musical training of listeners (Schellenberg 2002, Cummins et al. 2005), we separated our database into two groups according to whether the subjects had or did not have musical training.$^{32}$ The results were consistent with the results obtained in the study presented in Vanrell (2006) in the sense that musicians were faster that non-musicians in reacting to all stimuli. The results of the univariate ANOVAs showed that there was a significant interaction between musical training and RT for the information- ($F(1, 688) = 16.188, p < .001$) and confirmation-seeking-question continua ($F(1, 701) = 24.878, p < .001$) but no interaction between musical training, stimulus and RT for

$^{32}$ Of the 17 listeners whose responses were analyzed in this study had musical training (i.e., had taken music lessons for more than ten years) and 8 had not.
either of the two continua. This means that even though non-musicians are slower than musicians, both musicians and non-musicians display the same behavior with respect to the continuum, that is, they all tend to be slower at the frontier region between the categories than within the same category.

In sum, the RT measurements corroborate the finding that the main cue to the distinction between information- and confirmation-seeking questions is the pitch height of the tone associated with the preaccentual syllable. Complementary evidence in favour of this comes from the time alignment between the subjects’ responses and the stimuli. Though the listeners were instructed to always press the keys after the end of the stimulus, a large percentage of responses were given immediately after the onset of the preaccentual syllable and before the end of the stimulus. For the confirmation-seeking question continuum, 17% of the responses were given between the onset of the preaccented syllable and the end of the stimulus and 19% of responses followed this pattern in the case of the information-seeking question continuum. This raises the question as to whether there were subjects that were basing their judgments on cues other than those present in the preaccentual syllable. The answer is that only 4 responses of the total number of responses were given before the preaccentual syllable, which represents the insignificant percentage of 0.02% out of the total number of responses. This shows that the neutralization of cues not related to the nuclear region such as the initial pitch height was indeed effective.

3.3.3. Discrimination results

Figure 8 shows the discrimination results presented as d' for each stimulus pair in each order of presentation (low-high-ordered and high-low-ordered stimuli) for the confirmation (left-hand graph) and information-seeking-question-based continua (right-hand graph). d' scores were calculated on the basis of “different” responses to the pairs that were truly different (hits) and “different” responses to the pairs that were actually the same (false alarms). Following Macmillan and Creelman (1991), d' was calculated using roving methods (see Macmillan & Creelman 1991, Table A5.4, pp. 338-354). As can be seen, no clear peak is present in the frontier region between the categories; rather, we find two unexpected discrimination peaks occurred at pairs 2_3 and 7_8 for the confirmation-seeking-question-based continuum and at pairs 2_3 and 5_6 for the information-seeking-question-based continuum. No match was found between this function and the identification results.

The lack of a clear peak in discrimination functions is not new in the literature and in fact seems to be a constant in studies in which the CP paradigm is applied to intonational contrasts (Ladd & Morton 1997, for English; Remijsen & van Heuven 1999, for Dutch; Cummins et al. 2006, for English; Falé & Hub Faria 2006, for European Portuguese; Dilley 2010, for American English; Gili-Fivela 2008, 2009, for Italian) As we noted above in the introduction, the explanations for this lack of a clear peak in discrimination functions are varied (Remijsen and van Heuven 1999, Chen 2003, Kohler 2006, Niebuhr & Kohler 2004, Gili-Fivela 2009). In short, according to these studies, the absence of a clear peak should be related not to the existence of a
phonological distinction but rather to the unsuitability of applying the CP paradigm to intonational contrasts for different reasons (see the general Introduction of this dissertation).

One could also argue that the absence of a clear discrimination peak could be due in the present study to the length/duration of stimuli and the fact that short-memory effects may be interfering with task decisions. We rule out this explanation based on the following two arguments:

a) The iconic memory is known to last 250 milliseconds. Echoic memory\(^{33}\) is thought to last a little longer (Sáiz, Baqués, de la Fuente, Pousada & Vera 2008) but factors as the time separating the to-be-discriminated sounds but also the nature and the duration of the two items that should be discriminated can affect the discrimination performance (Crowder 1978, 1981, 1982). This can be summarized as follows “if two activations are close enough together in time and similar or identical in channel of arrival, they will mutually inhibit one another” (Crowder 1981). However, it seems that there is agreement that a contrast between two stimuli disappears at all, that is, labeling is not different in the “same” and the “different” conditions, at about three seconds of interstimulus interval (Crowder 1981). The maximum duration of our discrimination stimuli was 2,276 ms, still below the point in which the same-different discrimination performance stops being effective.

(b) The second argument has to do with previous studies (Chapter 1) in which a clear peak was obtained for discrimination results. In these previous studies the mean total duration of discrimination stimuli was 2,265 ms with an ISI (interstimulus interval) of 500 ms. In the present study, the mean total duration of the discrimination stimuli was 2,276 but with an ISI of 300 ms. It seems that the difference in the length of the stimuli between the two studies cannot have caused the different discrimination results, since there is a difference of only 11 ms between the two stimuli pairs. By contrast, we have a difference of 200 ms between the ISI used in the discrimination task of Chapter 1 of this dissertation and the one used in the present study. However, the results of the investigation presented in Chapter 5 of this dissertation in which different types of discrimination tasks were used, showed that an increased ISI can improve the discrimination performance but does not lead to the emergence of a discrimination peak. For that reason, we conclude that the lack of a discrimination peak in the present study is probably not due to short-memory or ISI effects. Our hypothesis is that the absence of a discrimination peak could depend on the categories tested meaning that a linguistic distinction such as the one between questions and statements or yes-no and wh-questions (recall that in Chapter 1 a clear discrimination peak was found) would be more easily categorized than a linguistic distinction related to focus (Gili-Fivela 2009) (see the general Conclusions for a more detailed explanation).

---

33 Echoic memory is the auditory version of sensory memory, that is, the ability to retain impressions of sensory information after the original stimulus has ceased and before they are processed by working memory.
In both functions a clear order-of-presentation effect was found: higher d' scores are obtained for curves in which the stimuli are presented in low-high order, that is, pairs in which the second stimulus has a higher peak than the first (curve with circles). A Wilcoxon matched pairs signed rank test revealed a significant difference between the two functions (low-high-ordered vs. high-low-ordered stimuli) for both confirmation- ($T = 10829.00, p < .001, r = -.107$) and information-seeking-question-based continua ($T = 1782, p < .001, r = -.11$). These results confirm the findings of previous studies (Kohler 1987, Ladd & Morton 1997, Remijsen & van Heuven 1999, Schneider & Linftert 2003, Cummins et al. 2006, Falé & Faria 2006, Prieto et al. 2010), in the sense that it appears that subjects have trouble discriminating between stimuli when the direction of change in fundamental frequency is downwards (see Chapter 1, section 4 for a possible explanation).

4. Discussion and conclusions

The patterns of results obtained from the three perceptual experiments clearly show that Majorcan Catalan listeners use the height of the leading tone H as the main cue to distinguish between information- and confirmation-seeking questions. First, the congruity test results show that listeners are extremely sensitive to the incongruent use of confirmation and information-seeking questions. As we saw in Figure 5, incongruent dialogues display a low average rate of “congruent responses”: 0.12 and 0.09 for the information-seeking and confirmation-seeking meanings respectively.

Another important result of our data is that it allows us to conclude that the knowledge/presumed presupposition model proposed by Escandell (1993) has a clear perceptual correlate. In the rating test, listeners had to interpret prosodic and morphosyntactic cues and relate them to a presupposition scale about the potential response to the sentence. They listened to three sentences that presented the same overt syntactic order but different intonation patterns: Teniu mandarines? ‘Do you have any mandarines?’
tangerines?’ (with the upstepped H tone aligned with the preaccentual syllable; with the plain H tone associated with the preaccentual syllable; and with declarative intonation). They also heard a sentence that contained a confirmation tag Teniu mandarines, no? ‘You have tangerines, don’t you?’. Our results show that listeners can clearly recognize morphosyntactic and prosodic cues and relate these cues to a specific degree of presupposition about the likelihood that the speaker will get a “yes” answer to her/his utterance. Thus, listeners can perceptually establish a hierarchy of presupposition on the part of a speaker that ranges from the broad-focus statement, in which the speaker has maximal knowledge about the utterance, to the information-seeking yes-no question in which her/his knowledge is much lower. Tag questions and confirmation questions occupy more central positions within this hierarchy. First, confirmation questions containing a plain H tone associated with the preaccentual syllable are interpreted as indicating that the speaker has a certain idea about the answer to her/his question and seeks confirmation of her/his hypothesis. Second, tag questions, by contrast, indicate that speakers have nearly all the information related to the truth value of the sentence but that there is still a little space for uncertainty. These results, then, confirm the existence of a gradient scale based on knowledge of the speaker and presupposed knowledge on the part of the hearer which is syntactically but also prosodically expressed.

Moreover, the identification results provide clear evidence about the discrete nature of this contrast. In Figure 6 we observed that the original stimulus —the confirmation-seeking question in the case of the confirmation-question-based continuum and the information-seeking question in the case of the information-seeking-question-based continuum— did not exert any effect on the categorical perception of this contrast. Thus, the functions obtained are undoubtedly S-shaped with an identification rate that goes from 0.86 to 0.19 (in the case of the confirmation-question-based continuum) and from 0.06 to 0.86 (in the case of the information-seeking-question-based continuum) within 5 steps of the 11-step-continuum. Statistical analyses show that the sharpest differences in the identification rate for adjacent stimuli in the identification task are located around stimulus 5, which corresponds to the boundary calculated from the logistic regression. Evidence of the linguistic nature of this contrast comes also from standard error values. As expected, listeners show more agreement in their responses when listening to stimulus 1 and 11 since they represent the endpoints of a continuum which we may interpret as canonical categories tested in this study, while this agreement decreases with proximity to the crossover point increases. In Table IV it can be observed that identification rates yield higher standard errors as they approach the most ambiguous rate, 0.5. A mean RT peak/plateau can be observed in Figure 7 which coincides roughly with the boundaries calculated from the fitted logistic curves. As was predicted, listeners are faster at within-category identification than across-category identification. However, the evidence from RT results is not supported statistically. We therefore asked whether this absence of significant differences could be due to dispersion. After breaking down our data into two groups, musicians and non-musicians, we confirmed that the variability could be explained by varying listener performance according to whether they had or did not have musical training. Musicians
were faster in responding to the stimuli than non-musicians, thus causing the variability in RT results. In spite of the significant effect of musical training on the RT, however, we verified that both musicians and non-musicians showed the basic behavior expected for a linguistic contrast, that is, they were all slower at across-category identification and faster at within-category identification.

Regarding discrimination results, no clear peak was found that coincided with the boundary calculated from the logistic curve regression. We claim that the nonexistence of a discrimination peak does not necessarily point to the absence of categorical perception, but might instead reflect the unsuitability of discrimination tasks as applied to intonational contrasts, as has been shown in Chen (2003) and Savino and Grice (2011). Hence, we are facing a problem related not to the nature of the contrast itself but rather to the methodology. We also argue that the absence of a clear discrimination peak cannot be attributed to the length/duration and ISI (interstimulus interval) values, which might be interfering with task decision. There are two arguments that back up this conclusion, namely that (a) there is agreement that the temporal threshold for which a contrast between two stimuli disappears at all is at about three seconds of ISI (Crowder 1981) and the maximum duration of our discrimination stimuli was still below this point; and (b) results presented in Chapter 5 of this dissertation have shown that although an increased ISI can improve discrimination performance, it does not lead to the emergence of a discrimination peak. Since a clear discrimination peak that corresponded to the shift in the identification function was obtained in Chapter 1, we hypothesize that this difference between these two sets of results could be due to a difference in the pragmatic categories under study (see the general Conclusions for a more detailed explanation about this hypothesis).

The results in this chapter have implications for the tonal representation of pitch accent contrasts in Catalan by means of Cat_ToBI (Prieto et al. 2009, Prieto in press, Aguilar et al. (coords.) 2009-2011). Though the standard Autosegmental-Metrical approach claims that only two tones (L and H) are sufficient to capture all the categorical differences in English, it is becoming increasingly clear that some languages require other tonal pitch levels to account for relevant linguistic contrasts. This has been shown to be the case, for example, with respect to the mid boundary tones (Beckman & Ayers-Elam 1997, for English; Arvaniti & Baltazani 2005, for Greek; Grice et al. 2005, for German; Beckman et al. 2002, for Spanish; Lee 2003, for Korean) or the extra high pitch accents in several languages (Savino & Grice 2007, 2011, for Bari Italian; Borràs-Comes et al. 2010, for Central Catalan). The present study confirms earlier evidence provided by the results presented in the first chapter of this dissertation for the presence of an upstepped high leading tone in the Majorcan variety. These two sets of results (Chapters 1 and 4) point to the existence of a three-way pitch scaling contrast in this variety of Catalan: an upstepped ¡H leading tone for information-seeking yes-no questions, a plain H leading tone for confirmation-seeking yes-no questions and a downstepped !H leading tone for wh-questions.

As it was observed in the previous paragraph, an upstepped leading H tone signals that the speaker has no particular expectation about the answer, while a non-upstepped leading H signals that the speaker is expressing her or his hypothesis about the state of
events while seeking confirmation. Our proposal is that interrogative intonation in Majorcan Catalan serves as a kind of epistemic marker in the sense that it indicates the degree to which a speaker is confident about the proposition expressed in a particular context. Typologically, one of the most common ways in which languages mark epistemic modality or evidentiality (referring to the marking of the commitment to the truth of the knowledge and the source of the knowledge, respectively) is by means of morphological marking. The following example is taken from De Haan (2001) from Suena, a New Guinean language (Wilson 1974): the sentence *ma-n-a sia* means ‘It’s true, I’ve really come’. In this sentence, the morpheme *sia* is used as a marker of certainty (Wilson 1974: 113). In a similar way, the absence of the upstep feature (¡) in Majorcan Catalan is expressing the speaker’s certainty about the truth value of her/his proposition. On the other hand, several Romance languages show a reportative/evidential marker that derives from the verb ‘say’ plus the complementiser *que* that expresses a meaning similar to English ‘apparently’ or ‘allegedly’. For example, in Sardinian the particle *nachi* can be found in questions and statements to indicate that the speaker has received auditory input, namely a description of an event reported by a third person. The following example: the question *Nachi benis a mandigare?* (‘You’re coming to eat, right?’) means that even though the speaker is asking whether the interlocutor is going to eat, s/he has an expectation that the answer will be affirmative since s/he has been previously informed of such possibility. This particle can be also found in declaratives, as in *Nachi chi haias coladu s’esame* (‘It seems that you’ve passed the exam’) with the same meaning described for interrogatives.

All in all, our findings represent further evidence that intonation constitutes a well-established linguistic strategy to mark the grammatical category of evidentiality across languages, and that its role in marking certainty, together with the interactions with other linguistic strategies, deserves to be further investigated. In a recent study of Gravano et al. (2008) the effect of contour type and epistemic modality on the perceived degree of certainty was assessed. Thirty native speakers of American English were asked to rate the degree of certainty of utterances that contained either the modal *would* or the verb *be* (e.g., *That would be me* vs. *That’s me*), which were also produced with different intonation contours (downstepped, declaratives and yes-no questions). They concluded that both the downstepped contour and the epistemic *would* are employed to convey speaker certainty, while the yes-no question contour is perceived to be highly uncertain (meaning that not only morphology but also intonation can act as an epistemic marker). Future studies will need to elucidate the potential interaction between these various linguistic strategies in the expression of a speaker’s certainty.

34 *Na* corresponds to the truncated form of *narat*, third person singular of the verb *narai* (‘to say’), while *chi* is a complementiser (‘that’).
35 Two recent studies (Prieto, Borràs-Comes, Roseano & Vanrell 2011 and Vanrell, Borràs-Comes, Roseano & Prieto 2011) aiming at determining the cues involved in the perception of confidence and uncertainty in Catalan conclude that even though lexical marking (the choice of modal adverbs such as *probably, perhaps*) is important for the conveyance of a pragmatic meaning such as uncertainty, it can be easily overridden by prosodic and gestural patterns.
Part III
Counterexpectational questions and statements of the obvious
CHAPTER 5

Evidence for mid boundary tones in Catalan

This chapter presents a perception experiment on the role of intonation of Catalan in distinguishing between statements of the obvious and counterexpectational questions. This distinction is based on a difference in pitch height of the final boundary tones. Thus, while counterexpectational questions are characterized by low-high boundary tones, statements of the obvious are produced with low-mid boundary tones. This is the only chapter of the dissertation dealing with a scaling contrast located at the boundary domain. The goal of this chapter is twofold. First, we want to investigate the functional role of F0 scaling differences at the boundary tone level in Catalan. The second goal of this chapter is to use evidence coming from several experimental techniques in order to investigate intonational categories.

1. Introduction

One of the controversial issues in the field of intonational phonology is whether variation in H pitch height can lead to the establishment of categorical contrasts. As mentioned before, the general approach to this topic by the Autosegmental-Metrical (AM) model is the original claim by Pierrehumbert (1980), borrowed from Bolinger (1988), that the primary phonological opposition between two tones, high (H) and low (L), are sufficient to distinguish pitch accent and boundary tone categories in English, while all remaining pitch range variation is attributed to phonetic variation, the application of downstep and upstep phonological rules, and the paralinguistic expression of emphasis and speaker involvement (Pierrehumbert 1980, Beckman & Pierrehumbert 1986, among many others). Studies of languages such as Spanish, English, French, Greek or German have documented a mid (and level) tone in utterance-final position (see Beckman & Ayers-Elam 1997 for American English, Grabe 1998 for British English and German, Post 2000 for French, Beckman et al. 2002 for Spanish, Lee 2003 for Korean, Arvaniti & Baltazani 2005 for Greek, Grice et al. 2005 and Peters 2006 for German, Frota in press-b for European Portuguese). Different representational proposals are available in the literature to represent this three-way contrast in height at the final intonational phrase boundary. While some authors have adopted the use of the M% level tone (e.g., Beckman et al. 2002 for Spanish, Lee 2003 for Korean), other authors have analyzed this boundary tone as a downstepped boundary tone !H% (e.g., Grice et al. 2005 for German, Arvaniti & Baltazani 2005 for Greek, Frota in press-b for European Portuguese), and others have used the 0% tone symbol to indicate that the pitch stays level after the nuclear accented syllable (see Grabe’s 1998 analysis of British English and German, Post 2000 analysis of French, and Peters’ 2006 analysis of German). In the last decade, a number of studies have investigated the role of pitch range in establishing categorical contrasts in English (Ladd & Morton 1997, Chen 2003, Calhoun 2004) as well as in other languages (Prieto 2004 for Spanish; Savino & Grice
2007, 2011 for Bari Italian; Borràs-Comes et al. 2010 for Central Catalan), and they have met with mixed results (see the general Introduction of the dissertation for a review of the results obtained in these studies).

In contrast with H scaling, the timing or alignment of fundamental frequency (F0) peaks or valleys with respect to segments has been consistently associated with intonational meaning distinctions in a number of languages (e.g., Purcell 1976 for Serbo-Croatian; Bruce 1977 for Swedish; Kohler 1987 for German; D’Imperio & House 1997, D’Imperio 2000 for Neapolitan Italian; Pierrehumbert & Steele 1989, Redi 2003, Dilley 2005, 2007, 2010 for English; Frota in press-b for European Portuguese; see Frota in press-b for a review). Even though all in all these articles provide evidence for claiming that changes in F0 alignment of peaks and valleys are especially salient and cue phonological distinctions across languages, some of them have provided contradictory (see for example Chen’s 2003 results on the English contrast between incredulity vs. uncertainty question meanings conveyed by the L+H* vs. L*+H pitch accents, in contrast with the original claim by Pierrehumbert 1980 and experiments that indicated the discreteness of this contrast in Pierrehumbert & Steele 1989, Redi 2003, and Dilley 2010).

Currently it is still unclear whether pitch range variation can give rise to distinctive linguistic categories across languages and researchers do not have a consensus view on this topic. In fact, the abovementioned evidence coming from different languages (see the general Introduction) seems to be only partially compatible with the general AM standard assumption that two tones are the basic units needed for representing intonational contrasts. In a recent experiment, Dilley (2005, 2007) constructed a continuum of pitch range differences, while controlling for F0 peak alignment across stimuli. Imitation data showed that speakers reproduced the continuous variation in the stimuli, suggesting that pitch range differences may be perceived as less categorical than F0 alignment differences. All in all, the original AM theoretical assumption that pitch scaling differences in the H domain are gradient, in contrast with alignment differences, is in need of crosslinguistic examination.

In Catalan, we find a number of boundary pitch movements that convey different discourse meanings, namely, L%, HH%, LH%, HL%, and LHL% (for an overview, see the Cat_ToBI proposal in Prieto et al. 2009; Prieto in press; Aguilar et al. (coords.) 2009-2011). First, Catalan has several instances of mid (or not extra high) boundary tones that are produced as rising or falling movements from the nuclear accent to a final mid tone, or as a complex falling and rising movement to a final mid tone (see schematic representations in Figure 1). The left panels illustrate the contrast between a statement of the obvious (e.g., (Home), la Bàrbara!, ‘Barbara, (obviously)!’) and a counterexpectational question (e.g., (Has dit) la Bàrbara?, ‘Did you say Barbara?’). The central panels illustrate the contrast between a disapproval statement (e.g., (No estic d’acord amb) la Bàrbara ‘(I disapprove of) Barbara’) and a yes-no question (e.g., (És) la Bàrbara? ‘(Is it) Barbara?’). Finally, the panels on the right illustrate the contrast
between an uncertainty statement\textsuperscript{36} (e.g., \textit{(Potser la) Bàrbara!} ‘(Perhaps) Barbara!’) and an emphatic statement (e.g., \textit{Bàrbara!} ‘(My God), Barbara!’). The basic difference between the three minimal pairs of nuclear configurations is the height of the boundary tone. It is also clear that the mid tone is not predictable from the preceding pitch accent.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1}
\caption{Schematic representations of three pairs of nuclear configurations in Catalan. Left panels: statement of the obvious (leftmost panel) vs. counterexpectational question (second from the left). Central panels: disapproval statement (left) vs. information-seeking yes-no question (right). Right panels: uncertainty statement (second from the right) vs. emphatic statement (rightmost panel).}
\end{figure}

The empirical basis for this investigation will be the contrast between counterexpectational questions and statements of the obvious exemplified above. Counterexpectational questions are used to express a denial of a discourse-activated assumption while statements of the obvious are used to express a statement on a piece of information that should be shared by the listener. Prosodically, whereas counterexpectational questions are produced with a sentence-final high boundary tone, statements of the obvious are produced with a mid boundary tone. The two graphs in Figure 2 illustrate the contrast between counterexpectational questions and statements of the obvious. The figures show the waveform, spectrogram, F0 contour, and Cat_ToBI labelling of the utterance \textit{La Bàrbara} ‘Barbara’ spoken with statement of the obvious intonation (top panel) and with a counterexpectational question intonation (bottom panel).

A number of experimental methods have been used to investigate what is categorical and thus linguistic in intonation and what is paralinguistic and gradient (see a review in Gussenhoven 2004, 2006; see also Dilley 2010). One method that comes from the study of consonantal contrasts and which has also been applied to intonational contrasts is the Categorical Perception paradigm, or CP (Liberman et al. 1957; Fry, Abramson, Eimas & Liberman 1962; Repp 1981; Delattre, Liberman, Cooper & Gerstman 1952; Tartter 1982; Sawusch, Nusbaum & Schwab 1980) but its suitability for investigating intonational contrasts is being questioned since while identification curves show clear discrete effects, discrimination results are less clear-cut. Researchers provide different explanations for this lack of a clear peak in discrimination functions (see Chapter 4, section 1). Several factors have been mentioned as possible sources to explain the null results of discrimination tasks both in intonation and in vowel studies. First, step size matters, as a very small step size in F0 between stimulus pairs can render the contrasts undetectable (Newport 1982, Repp 1984). Auditory storage time (measured in InterStimulus Interval, or ISI) has been shown to greatly affect discrimination performance: in essence, discrimination performance improves as ISI increases.

\textsuperscript{36} The results from Vanrell et al. (2011), in which the prosodic cues of confidence and uncertainty in Catalan are analyzed, show that not only pitch patterns are important in the conveyance of uncertainty but also temporal patterns such as speech rate or duration of the final syllable.
(Massaro 1974). Finally, the type of task has also been claimed to be important, namely, the fact that discrimination task is basically a low-level auditory task that can be answered in an auditory mode, without resorting to the phonological module (Newman 1984, Repp 1984, Gussenhoven 1986, Niebuhr & Kohler 2004, Gerrits & Schouten 2004). Gerrits and Schouten (2004) claim that the classic AX discrimination task (the type most often used) generally activates an ‘auditory mode’ which leads to common psychoacoustic patterns like order-of-presentation asymmetry effects and great differences in performance between subjects. In an attempt to address this issue, Gerrits and Schouten (2004) designed a discrimination task that prevents direct comparison between successive stimuli and forces the use a phonetic labeling strategy for discrimination results that are highly categorical for vowels. Following this, Frota (in press-a) designed a semantically motivated discrimination task to analyze an alignment intonational contrast in European Portuguese. She claimed that discrimination tasks, as well as identification tasks, should be semantically motivated in order to assess intonational categories.
With this in mind, we performed several variants of the AX discrimination task that took into account increased step-size, increased ISI, and variations in the type of task that would involve semantic access to the materials (see the Methodology section). Another method which has also been applied with success in empirical studies of intonation has been the imitation task, in which participants attempt to reproduce artificially generated continuous acoustic variation in stimuli (Pierrehumbert & Steele 1989; Redi 2003; Dilley 2005, 2007, 2010; Dilley & Brown 2007; Gili-Fivela 2009).

Based on a review of the literature, Gussenhoven (2004, 2006) has suggested that imitation tasks were one of the best available means to assess intonational categories. In their seminal paper, Pierrehumbert and Steele (1989) investigated the contrast between L+H* and L*+H in English. The authors argued that if the subjects were able to reproduce the continuum in their imitation, peak alignment differences must be gradient. However, if subjects’ imitations were to fall into two categories, peak alignment differences must be categorical. They found that by and large the distribution of peak alignments was bimodal in the imitation and therefore concluded that the distinction between early peak alignment and late peak alignment was discrete. The discreteness of the peak alignment contrast revealed by Pierrehumbert and Steele’s production data was borne out by the perception data (and also imitation data) reported in Dilley (2010), but with some caveats: while some of the sentences were produced with a discrete shift in the timing of produced peaks and valleys, other appear to be somewhat gradual. Similarly, the same English contrast was investigated by Redi (2003) with contradictory results: in the imitation task, participants failed to produce bimodal timing in F0 valleys in response to the stimulus continua.

Only a handful of studies so far have worked with imitation tasks on a pitch range continuum (Dilley 2007, Dilley & Brown 2007, Gili-Fivela 2009, Rathcke & Harrington 2010). Dilley (2005, 2007) constructed a continuum of pitch range
differences, while controlling for F0 peak alignment across stimuli. Imitation data showed that speakers reproduced the continuous variation in the stimuli, thus suggesting that pitch range is a phonetic dimension which is gradient in English. On the other hand, Gili-Fivela (2009) manipulated both the alignment and scaling patterns of a rising pitch accent in narrow focus and a rising-falling pitch accent in contrastive narrow focus. In the H*+L to H* L- continuum, two speakers produced imitated productions that belonged to two acoustically different populations on the basis of alignment differences; yet a third speaker produced patterns depending on both alignment and scaling differences. In sum, previous studies of intonation using this technique have been inconclusive with respect to its suitability. Finally, Rathcke and Harrington (2010) performed an imitation task with a continuum of a falling F0 contour over the final syllable in German. The results provided evidence for a two-way distinction between an early and a later (H*) peak accent, but not for a three-way distinction nor to a categorical distinction between H+!H* and H+L*.

In order to assess the performance of these experimental techniques to the study of intonational phonology, we performed a variety of tasks on the same linguistic contrast, paying special attention to the relevance of the semantic dimension in these tasks.

The goals of this chapter are twofold. First, we investigate whether Catalan listeners perceive two types of distinct tonal rising movements at the end of the sentence and whether they relate them systematically to different discourse meanings. Do differences in F0 scaling trigger a perceptual change from one meaning to the other? The results will provide a wider crosslinguistic perspective on whether scaling differences can encode discrete categories in the boundary area as opposed to simply within-category phonetic variation.

Second, we compare the results from different experimental techniques to the analysis of the same tonal contrast. By comparing the results from different tasks we will be in a position to assess the suitability of these methods for the study of intonational categories. With this aim in mind, we conducted four sets of experiments. Experiment 1 consisted of a congruity task in which listeners were asked to evaluate the appropriateness of the intonational contours to a discourse context. Experiment 2 consisted of an identification task. Experiment 3 was comprised of different types of discrimination tasks to assess differential sensitivity to F0 scaling differences. We implemented several types of discrimination schemes in order to seek converging evidence regarding the viability of the categorical perception paradigm for the study of intonational categories. Finally, Experiment 4 involved an imitation task in which participants attempted to reproduce F0 scaling differences as closely as possible.

The fifth chapter is organized as follows. In section 2 we discuss the general methodology of the experiments that will be conducted, together with a detailed description of each of the experiments performed. Section 3 presents the results of the four experiments. Finally, section 4 discusses the implications of these results for understanding the nature of intonational contrasts and presents concluding remarks.
2. General methodology

2.1. Experimental tasks

In order to test whether the change from a mid boundary tone to a high boundary tone (and conversely) is categorically perceived by Catalan listeners, we undertook a variety of experimental tasks.

2.1.1. Experiment 1: Congruity task

The congruity task was designed to investigate whether listeners are aware of the semantic appropriateness of a particular intonation contour to a given discourse context and are able to detect an incongruent use of a target intonation pattern. Participants in this experiment were asked to classify appropriateness/congruity of the stimulus to a given discourse context in a semantic task. This task has been successfully used by other researchers investigating intonational contrasts (see Kohler 1987; Niebuhr 2007; Rathcke & Harrington 2006, 2010; Frota in press-a). In our task, listeners were asked to judge contextual appropriateness of the target stimulus to a given discourse context through a dual response task (“congruent” vs. “incongruent”). The materials consisted of two target dialog contexts (question context vs. statement context) that were combined with our two target contour types (see the Materials section).

2.1.2. Experiment 2: Identification task

Identification tasks have been used to examine the extent to which listeners assign different meanings to different intonation contours. In our case, subjects were asked to respond after each stimulus according to how they would answer the question in a real situation. In other words, if they perceived the statement of the obvious (Home), la Bàrbara!, ‘Barbara, (obviously)!’, they were to press the “A” key on the keyboard (for the target answer Ah! = “Of course”), whereas if they perceived the counterexpectational question (Has dit) la Bàrbara? ‘(Did you say) Barbara?’ , they were to press the “S” key (for the target answer Sí = “Yes”). The materials for the identification task consisted of 22 stimuli coming from two continua created with the two base stimuli (see the Materials section).

2.1.3. Experiment 3: Discrimination tasks

As noted above, in Experiment 3 we performed several variants of the classic AX (same/different) discrimination task in order to assess sensitivity to F0 scaling differences. The idea behind this task is that hearers will only be able to discriminate pairs of items that are located across different perceptual categories, as informed by the results of identification tests (Repp 1984). The materials for all discrimination experiments consisted of pairs of stimuli taken from the continua created with two base stimuli (see the Materials section).
(a) Experiment 3a. AX Discrimination Task

In this task, listeners were asked to compare two target stimuli in the synthesized continuum and say whether they are the same or different.

(b) Experiment 3b. AX Discrimination Task with increased step size

Our goal was to evaluate whether an increase in the step size in the target materials would make any difference in the results. This experiment replicates the AX discrimination task but with an increased step size. Once more, subjects had to judge whether two stimuli were the same or different. Yet in this case the distance between the stimuli was two steps along the stimulus continuum instead of one.

(c) Experiment 3c. AX Discrimination Task with increased ISI

Our goal was to test if discrimination performance improves as the ISI increases, as demonstrated by Massaro (1974), among others. We implemented the classic AX discrimination task but with an increased InterStimulus Interval (ISI) between stimuli. The interval between the two stimuli in each pair was expanded from 200 to 500 ms.

(d) Experiment 3d. Four Interval Discrimination Task

Following Gerrits and Schouten (2004), we adapted the discrimination task in such a way that subjects were prevented from performing it only in an exclusive auditory mode. In what is known as a “four-interval discrimination task”, subjects are expected to be free to both use auditory processing and access the linguistic labeling component. The stimuli were presented randomly in two possible forms AABA and ABAA, and the subjects had to decide the position of the “B” stimulus. They were told that three of the four stimuli in each combination were going to be identical and that the oddball would be either the second or the third one. The interstimulus interval was kept at 200 ms.

(e) Experiment 3e. Semantic Discrimination Task

In this task subjects were asked to perform an AX task, but this time they had to judge the stimuli both on semantic and acoustic grounds. First of all, they had to determine whether the second stimulus was the same as or different from the first stimulus. If they heard the second stimulus to be the same as the first one, subjects had to press the “0” key for “no difference”, but if they heard the second stimulus as different from the first, subjects had to label this stimulus as either a declarative (“D” key) or interrogative (“I” key). We designed this task (acoustic judgment + labeling of the second item) on the assumption that the labeling module would be ‘activated’ during the acoustic discrimination phase too, inducing listeners to access the linguistic labeling component.
2.1.4. Experiment 4: Imitation task

For this experiment, participants had to hear a sentence over the headphones, and they were told to imitate it as closely as possible using a comfortable pitch range, and not trying to imitate the exact range of the stimuli. The same 22 stimuli making up the two continua were used (see the Materials section).

2.2. Speech material

The target speech materials for all the experiments were produced by two native speakers of Central Catalan. They were recorded in a quiet room of the Universitat Autònoma de Barcelona using a Marantz Professional PMD660 digital recorder and two Rode NTG-2 microphones. The recorded utterances were digitized at a 44100 sample rate using the Goldwave software.

For all the experiments, the presentation of the stimuli was prepared using E-prime version 1.2 (Psychology Software Tools Inc. 2009). Subjects were seated at a laptop in a quiet room and the stimuli were played back through headphones. All the experimental tasks were conducted at the Universitat Autònoma de Barcelona. Since we were interested in RT measurements, for experiments 1-3 listeners were instructed to maintain their hands near the keyboard and press the keys as fast as they could. In all the tasks, subjects were given written instructions about how they were to respond.

2.2.1. Stimuli for Experiment 1

First, two dialogues were recorded that were considered to be appropriate contexts for the production of a counterexpectational question (produced with a final high boundary tone) and of a statement of the obvious (produced with a final mid boundary tone). The Catalan version (and the English translation) of the target dialogues is reproduced as follows, with the target words underlined:

Target question dialog
Speaker A: —Al final vindrà a sopar qui tu volies.
"You know, the person you wanted will be coming for dinner."
Speaker B: —Qui? La Bàrbara?
"Who? Barbara?"
Speaker A: —Això mateix.
"That’s right."
Speaker B: —Ah!
"Ah!"

Target statement dialog
Speaker A: —Sé que m’ho has dit moltes vegades, però se m’ha oblidat un altre cop. Qui vindrà al sopar?
"I know you have said it many times, but he forgot it another time. Who will come for dinner?"
—‘I know you’ve told me a million times, but I’ve forgotten. Who’ll be coming for dinner?’

Speaker B: —Sí home, la Bàrbara!
—‘Barbara!’

Speaker A: —Ah, és cert!
—‘Oh yeah, that’s right!’

We then created an alternative “incongruent” pair of dialogs by inserting the target utterance with statement of the obvious intonation into the question paragraph, and vice versa. The splicing was performed using Goldwave software. The two panels in Figure 3 show the waveform, F0 contour, and Cat_ToBI transcriptions for the non-manipulated target question dialogue (top panel) and for the spliced stimuli inserted in the question dialogue (bottom panel).

The four target dialogs (the statement and the question dialogs with the two intonation contours) were the basic stimuli in this experiment. Stimuli were repeated 5 times in 2 blocks. We thus obtained a total of 800 responses (2 contours x 2 contextual situations x 5 repetitions x 2 blocks x 20 listeners). Thus, each participant heard a total of 40 tokens (2 contours x 2 contextual situations x 5 repetitions x 2 blocks).
Stimuli for Experiments 2-4

For Experiments 2-4 (identification task, discrimination tasks, and imitation task), two sets of stimuli were created from two original tokens produced by a native speaker of Central Catalan, e.g., one token of the counterexpectational question *(Has dit) la Bàrbara?* ‘(Did you say) Barbara?’*, and one token of the statement of the obvious *(Home), la Bàrbara!,* ‘Barbara, (obviously)!’. Note that the target word contains exclusively voiced and sonorant consonants (recall that orthographic “b” in this context is pronounced as bilabial approximant in Central Catalan). The stimuli were selected from several pronunciations of these sentences inserted in two different contextual paragraphs. For the final selection of the base stimuli, we checked that the F0 peak and valley of the pitch accent L+H* was approximately the same in the two examples. The basic values of F0-final targets originally produced were 172.5 Hz for the statement and 278.4 Hz for the question.

To create the stimuli, the F0 value at the end of the sentence was manipulated using the resynthesis script in Praat (Boersma & Weenink 2009). A linear stylization of the final rising movement was carried out. The following five points were interpolated: a point at the beginning of the contour, a point at the onset of the rising L+H* accent, a point at the F0 maximum of the L+H* accent, a point at the onset of the final rising boundary movement, and a point at the offset of the final rising boundary movement, that is...
!H% or H% depending on the base stimulus (see Figure 4). From these two base tokens, eleven stimuli were created by means of PSOLA synthesis. Eleven stimuli were created by shifting the peak downwards from the yes-no question token in 11 steps of 11.53 Hz each, and conversely by shifting the peak upwards from the statement of the obvious token. The original tokens lasted 780 ms for the statement and 795 ms for the question (including a silence sequence before and after the token). The two graphs in Figure 4 schematically show the 22 stimuli that were created and that were used for Experiments 2-4. The stimuli numbers correspond to the low-high F0 values for the stimuli created from the !H% base stimulus and to the high-low F0 values for the stimuli created from the H% base stimulus.

Figure 4: Idealized schemas of the stimuli used for the Experiments 2-4. The left panel shows the stimuli constructed from the statement base stimulus (Home), la Bàrbara!, ‘Barbara, (obviously)! ‘ and the right panel the stimuli constructed from the question base stimulus (Has dit) la Bàrbara? ‘(Did you say) Barbara?’.

For the identification task (Experiment 2), subjects heard a total of 110 sequences of la Bàrbara (11 stimuli x 2 base stimuli x 5 blocks).

For the discrimination tasks, the materials consisted of pairs of stimuli from the two continua. Twenty pairs were created in low-high order, meaning that the second stimulus is always higher in frequency than the first (ten from the statement continuum and ten from the question continuum). Twenty pairs were created in high-low order, meaning that the second stimulus is always lower in frequency than the first (ten from the statement continuum and ten from the question continuum). Additionally, 22 control pairs which contained two identical stimuli (11 from the statement continuum and 11 from the question continuum) were created. Thus for the AX discrimination task, participants heard a total of 310 tokens (31 stimuli pairs -10 low-high pairs, 10 high-low pairs, 11 control pairs x 2 contours x 5 repetitions). Listeners were asked to decide whether pairs of stimuli differing in F0 scaling sounded the same or different. If the two stimuli sounded the same, they were to press the “I” key on the keyboard for Catalan igual (‘same’) and if the stimuli sounded different, they were to press the “D” key for Catalan diferent (‘different’). In this experiment, we used one-step size and 200 millisecond interval between the stimuli.

For the increased step size discrimination experiment, participants heard a total of 290 tokens (2 contours x 29 stimuli pairs -9 low-high pairs, 9 high-low pairs and 11
control pairs- x 5 blocks). The interstimulus interval was kept at 200 ms. The materials for the four-interval discrimination task was made up of 10 ABAA combinations + 10 AABA combinations x 2 contours x 5 blocks for a total of 200 tokens per speaker. The interstimulus interval was kept at 200 ms. The materials for the semantic discrimination task consisted of the same stimulus pairs presented in the increased step size discrimination task (a total of 290 tokens per participant with an ISI of 200 ms). Finally, for the AX discrimination task with increased ISI, we used the same stimulus pairs than in the AX discrimination task (310 tokens per participant), however, the interval between the two stimuli in each pair was expanded from 200 to 500 ms.

The materials used for Experiment 4 (the imitation experiment) were the same 22 randomized stimuli coming from the two continua in two blocks (see Figure 2). Each block was preceded by a test phase where listeners could hear 5 random examples. No time limit was imposed for obtaining the responses and the possibility of repetition was allowed. A total of 858 imitations were obtained (11 stimuli x 2 continua x 3 repetitions x 13 subjects). However, results from two of the subjects had to be eliminated due to the fact that the final part of the utterance was uttered in creaky voice and thus the F0 values could not be easily extracted. Hence a total of 660 imitations were analyzed.

2.3. Participants

The full set of tasks was performed in three different sessions. In the first session participants performed the semantic congruity task, the identification task, and the AX discrimination task. In the second session, they performed the AX discrimination task with increased step size, the four interval discrimination task, the semantic discrimination task, and the increased ISI discrimination task. Finally, a separate session was devoted to the imitation task.

Twenty speakers (thirteen females and seven males) of Central Catalan aged between 18 and 30 participated in the first session. They were all undergraduate students at the Universitat Autònoma de Barcelona and they had no specific background in phonetics. None of them reported a history of hearing disability. All the tasks were performed in the same order: semantic congruity task, identification task, and AX discrimination task. There was a break between the tasks, and practice blocks before all of them. The whole set of tasks lasted approximately 60 minutes.

Twenty speakers (fourteen females and six males) of Central Catalan aged between 18 and 30 participated in the second session (approximately a 60% of them had participated in the first session five months before). They were all undergraduate students at the Universitat Autònoma de Barcelona and they had no specific or little background in phonetics. None of them reported a history of hearing disability. This session lasted a total of 45 minutes. All the listeners participated in the full sequence of tasks and performed the various tests in the same order: increased step size, the four interval discrimination task and the semantic discrimination task. There was a break of 30 minutes between the second and third task. Finally, half of the participants of this session also performed the increased ISI discrimination task that lasted 15 minutes.
Thirteen speakers (11 females and 2 males) of Central Catalan aged between 18 and 30 participated in the imitation task. All of them had participated in the first or second session and, for that reason, were familiarized with the materials. The task lasted 15 minutes.

2.4. Statistical analyses

In this chapter non-parametric tests were used when the dependent variable in our data set was measured in a nominal scale (congruity test and identification/discrimination tests responses) and parametric tests when the dependent variable was continuous (RT measurements and imitation task data). Results were obtained by SPSS statistics software (SPSS for Windows). The Wilcoxon matched pairs signed rank test was used to: (1) compare the responses to two conditions (congruent vs. incongruent) for both counterexpectational questions and statements of the obvious in the congruity test, (2) compare the difference between hit rate and false alarm rate (d prime) between two conditions (e.g., pair 1_2 in the continuum vs. pair 2_3 in the continuum) in each order of presentation for both meanings in determining where the discrimination peak was located in the discrimination results, and (3) compare the difference between hit rate and false alarm rate (d prime) between low-high and high-low orders of presentation for each pair of stimuli in the discrimination results. The Friedman test was used to test the differences between identification scores for each stimulus of the continua created for each meaning in the identification task. Regarding non-parametric tests, we reported the same values than in the previous chapter. The post-hoc Bonferroni correction was applied in cases in which multiple comparisons were performed. The univariate ANOVA run with the GLM procedure of SPSS was performed as parametric test for each meaning (counterexpectational question and statement of the obvious) with RT measurement as the dependent variable, and with one between-subjects factor, namely Stimulus (11 steps in the pitch range continuum). This test was also used on the imitation data with the dependent variable Scaling of the F0 value (in ERB) and with the independent factor Stimulus Category (i.e., the three types of stimuli, namely, Question, Statement, and Ambiguous) for each speaker. The results from the univariate ANOVA test include the F-ratio, the degrees of freedom from which it was calculated and the significance value.

3. Results

3.1. Experiment 1: Congruity task

Figure 5 shows the rate of “congruent” responses to both congruent (black bars) and incongruent dialogues (grey bars), separated into question (left) and statement (right) meanings. The results reveal an average congruity rate of 0.89 and 0.92 of “congruent” responses to congruent dialogues for question and statement meanings respectively. By contrast, the average rate of congruent responses in the incongruent dialogue was 0.29
and 0.38 for question and statement meanings respectively. That is, even though there is a sharp contrast between congruity conditions, subjects accept non exemplary intonation roughly 30% of the time. Moreover, there is an asymmetry between the congruity rate for question and statement meanings in incongruent dialogues: the statement of the obvious meaning yields a higher congruity rate, meaning that LH% tends to be better tolerated for the statement of the obvious context (L!H%) than the reverse. This suggests that intonational contours that fall into the phonetic space of counterexpectational questions can be pragmatically adequate exemplars in a statement context. Results from the Wilcoxon matched pairs signed rank test revealed that the differences between the two conditions (congruent vs. incongruent) for both counterexpectational question ($T = 0, p < .001, r = -.78$) and statement of the obvious ($T = 0, p < .001, r = -.78$) meanings were significant.

![Figure 5](image)

**Figure 5**: Rate of “congruent” responses to congruent (black bars) and incongruent dialogues (grey bars), separated by question meaning (left) and statement meaning (right).

### 3.2. Experiment 2: Identification task

Figure 6 shows the average rate of identification (and standard error values) as a function of stimulus step number for all 20 listeners. The two curves present a clear S-shape and the shift from one category to another occurs in the range of stimulus 4 to 7 in the two curves. Importantly, a full crossover from 0.2 to more than 0.8 is achieved in two steps in both curves. The graph also shows that mean standard error values are higher in the central stimuli (namely stimuli 5-7), that is, when the stimulus is clearly ambiguous, and lower for the less ambiguous stimuli.
Two Friedman’s ANOVA were applied to identification data. The analyses revealed a significant main effect of stimulus number for the question based continuum ($\chi^2(10) = 358.029, p < .001$) and for the statement of the statement based continuum ($\chi^2(10) = 279.664, p < .001$). In order to compare the two different curves obtained respectively for the question and statement continua, the set of data points was fitted with a logistic function in the SPSS statistical analysis program (SPSS for Windows) (see Chapter 2, section 2.2.1. for an explanation about the formula and the procedure to calculate the location of the category boundary).

For the question-based continuum, when “y” equals 0.5 “x” is 5.31, and for the statement-based continuum, when “y” equals 0.5 “x” is 6.11. Thus the boundary is located between stimuli 5 and 6 for the question continuum and around stimulus 6 for the statement continuum (see graph).

Figure 7 plots averaged RT responses (in ms) for all subjects. Black bars show the reaction times for question-based stimuli and grey bars show the reaction times for statement-based stimuli. A mean RT peak is obtained at stimulus 5 for the question continuum and at stimulus 6 for the statement continuum. The RT values for these stimuli are 994.72 and 840.81 ms respectively (RT values include stimulus duration). At the other end of the spectrum, stimuli 11 for the question continuum and stimuli 1-2 for the statement continuum obtain the lowest RT measurements (513.10, 577.34, and 576.20 ms respectively).

An important effect of stimulus number was found on the RT measurements for both continua (univariate ANOVAs ($F(10, 525) = 5.251, p < .001$) for the question based continuum and ($F(10, 529) = 3.059, p < .01$) for the statement based continuum). Yet post-hoc analyses revealed no significant differences for adjacent stimuli for any of the continua. We hypothesize that the absence of a statistically significant ‘peak’ in RTs for the central continua could be explained by subject variability.
3.3.  Experiment 3: Discrimination tasks

3.3.1.  AX Discrimination Task

The two graphs in Figure 8 show the discrimination results presented as $d'$ (d prime) for each adjacent stimulus pair, as a function of order of presentation (high-low vs. low-high sequences), for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph). The percentage of correct discriminations between different pairs is highly susceptible to subjects who tend to give only “different” or “same” responses all the time, and it should be interpreted in terms of the listener’s response bias, that is, his or her tendency to qualify stimuli pairs as “same” or “different”. In that sense, Signal Detection Theory attributes responses to a combination of sensitivity and bias. Sensitivity (d prime) is the variable that is being investigated and bias is what we must take into account so that sensitivity measure is meaningful. A higher $d'$ means that the signal is more readily detected. D’ scores were calculated on the basis of “different” responses to the pairs that were truly different (hits) and “different” responses to the pairs that were actually the same (false alarms). Following Macmillan & Creelman (1991: Table A5.4, pp. 338-354), $d'$ was calculated using roving methods. The graphs show that discrimination rates are higher for low-high than for high-low pairs, for both continua. Yet no discrimination peaks appear at the expected category boundary region, with unexpected peaks appearing at the 2-3 and the 9-10 comparisons for the question- and statement-based continua respectively, in both orders of presentation.
Figure 8: Sensitivity (d’) to difference between stimulus pairs, calculated for both orders of presentation (low-high and high-low pairs), for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph).

Nine Wilcoxon matched pairs rank tests were carried out on hit rate minus false alarm rate comparing the responses to adjacent pairs for low-high and high-low functions for each continuum (see Table I). Results revealed a significant difference between pairs 2_3 and 3_4 for the statement-based continuum and between pairs 1_2 vs. 2_3 and 2_3 vs. 3_4 for the question-based continuum (both in the low-high functions) (shaded values). However, no match is found between these results and the results of the identification task. In both graphs, a clear order-of-presentation effect was found: another Wilcoxon matched pairs rank test revealed a significant difference between the two functions (at $p < .001$ for both the statement-based and the question-based continua).

A very consistent order-of-presentation effect was found in the data, in both continua. That is, pairs of stimuli presented in the low-high order were better discriminated than those with high-low order. This result is in agreement with previous discrimination experiments dealing with tonal scaling differences (Kohler 1987, Ladd & Morton 1997, Remijsen & van Heuven 1999, Schneider & Linftert 2003, Schneider et
al. 2009, Falé & Faria 2006, Cummins et al. 2006, Gili Fivela 2009, Chapter 1 and Chapter 4 of this dissertation). In these experiments, pairs of stimuli were more successfully discriminated when the second one had a higher pitch than the other way around. This order-of-presentation asymmetry has also been found in tonal languages (Francis & Ciocca 2003).

In sum, no discrimination peak was found in this first task. Several factors have been mentioned as possible sources of null results in discrimination tasks, such as step size, auditory storage time (measured in InterStimulus Interval, or ISI), and type of task. As mentioned before, some of the problems reported regarding discrimination tasks relate to the fact that it is an auditory task that can be answered in an auditory mode (Newman 1984, Repp 1984, Gussenhoven 1986, Niebuhr & Kohler 2004, Gerrits & Schouten 2004). Taking this into consideration, we planned a set of variants of the discrimination task to find out whether these factors were interfering with our results.

### 3.3.2. AX Discrimination Task with increased step size

The two graphs in Figure 9 represent the averages of the 20 listeners’ individual d’ scores for each stimulus pair and order of presentation, for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph). The results of Wilcoxon matched pairs rank tests reveal no statistically significant differences between adjacent pair comparisons (see Table II), confirming that there is no discrimination peak at the expected category boundary. That is, discrimination results are the same regardless of the fact that the step size in between stimuli was doubled. Even though order of presentation effects are clear in both graphs, differences are only significant for the question-based continuum ($p < .001$).

![Figure 9: Sensitivity (d’) to difference between stimulus pairs, calculated for both orders of presentation (low-high and high-low pairs), for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph).](image-url)
Table II: Results of eight Wilcoxon matched pairs signed rank tests comparing differentiation rates between adjacent pairs of stimuli for low-high and high-low functions for each continuum.

<table>
<thead>
<tr>
<th></th>
<th>1_3 vs. 2_4</th>
<th>2_4 vs. 3_5</th>
<th>3_5 vs. 4_6</th>
<th>4_6 vs. 5_7</th>
<th>5_7 vs. 6_8</th>
<th>6_8 vs. 7_9</th>
<th>7_9 vs. 8_10</th>
<th>8_10 vs. 9_11</th>
</tr>
</thead>
<tbody>
<tr>
<td>!H low-high</td>
<td>T = 74</td>
<td>T = 67.5</td>
<td>T = 99.5</td>
<td>T = 74</td>
<td>T = 90</td>
<td>T = 33</td>
<td>T = 86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.17</td>
<td>r = -.005</td>
<td>r = -.24</td>
<td>r = -.23</td>
<td>r = -.06</td>
<td>r = -.40</td>
<td>r = -.10</td>
<td></td>
</tr>
<tr>
<td>!H high-low</td>
<td>T = 70.5</td>
<td>T = 45</td>
<td>T = 70.5</td>
<td>T = 86.5</td>
<td>T = 30</td>
<td>T = 95.5</td>
<td>T = 98.5</td>
<td>T = 60</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.25</td>
<td>r = -.10</td>
<td>r = -.25</td>
<td>r = -.21</td>
<td>r = -.31</td>
<td>r = -.20</td>
<td>r = -.16</td>
<td>r = -.25</td>
</tr>
<tr>
<td>H low-high</td>
<td>T = 58.5</td>
<td>T = 41</td>
<td>T = 53.5</td>
<td>T = 70</td>
<td>T = 47</td>
<td>T = 74</td>
<td>T = 52</td>
<td>T = 53</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.20</td>
<td>r = -.24</td>
<td>r = -.16</td>
<td>r = -.12</td>
<td>r = -.07</td>
<td>r = -.02</td>
<td>r = -.10</td>
<td>r = -.17</td>
</tr>
<tr>
<td>H high-low</td>
<td>T = 51</td>
<td>T = 55.5</td>
<td>T = 56.5</td>
<td>T = 39</td>
<td>T = 64</td>
<td>T = 50</td>
<td>T = 42</td>
<td>T = 68</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.11</td>
<td>r = -.05</td>
<td>r = -.13</td>
<td>r = -.18</td>
<td>r = -.16</td>
<td>r = -.16</td>
<td>r = -.22</td>
<td>r = 0</td>
</tr>
</tbody>
</table>

3.3.3. AX Discrimination Task with increased ISI

The data plotted in Figure 10 represent the averages of the 10 subjects’ individual d’ scores for each stimulus pair and order of presentation. Again, we find two unexpected discrimination peaks at pairs 4-5 and 7-8 for both continua, confirming that increasing the ISI in the presentation of the stimuli does not significantly change the state of affairs.

Wilcoxon matched pairs rank tests (Table III) show that only the peak located at the 4-5 pair is significantly different from adjacent pairs and only for the low-high order (shaded values). Isolated significant effects appear between pairs 2-3 and 3-4, and between pairs 6-7 and 7-8, but only in the H continuum and in only one of the orders of presentation. However, the important result is that this discrimination peak does not match with the results of the identification task (recall that we found the shift between...
categories at the 5.31 for the question continuum, and at 6.11 for the statement continuum).

<table>
<thead>
<tr>
<th></th>
<th>1_2 vs. 2_3</th>
<th>2_3 vs. 3_4</th>
<th>3_4 vs. 4_5</th>
<th>4_5 vs. 5_6</th>
<th>5_6 vs. 6_7</th>
<th>6_7 vs. 7_8</th>
<th>7_8 vs. 8_9</th>
<th>8_9 vs. 9_10</th>
<th>9_10 vs. 10_11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>!H low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>T = 21</td>
<td>T = 21</td>
<td>T = 0</td>
<td>T = 28.5</td>
<td>T = 15</td>
<td>T = 14</td>
<td>T = 10</td>
<td>T = 10</td>
<td>T = 23</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.31</td>
<td>r = -.69</td>
<td>r = -.79</td>
<td>r = -.46</td>
<td>r = -.05</td>
<td>r = -.17</td>
<td>r = -.21</td>
<td>r = -.18</td>
<td></td>
</tr>
<tr>
<td><strong>!H high</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>T = 7</td>
<td>T = 16</td>
<td>T = 39</td>
<td>T = 7</td>
<td>T = 37.5</td>
<td>T = 7.5</td>
<td>T = 15</td>
<td>T = 12</td>
<td>T = 11</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.04</td>
<td>r = -.37</td>
<td>r = -.04</td>
<td>r = -.05</td>
<td>r = -.05</td>
<td>r = -.05</td>
<td>r = -.03</td>
<td>r = -.16</td>
<td></td>
</tr>
<tr>
<td><strong>H low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high</td>
<td>T = 18.5</td>
<td>T = 1</td>
<td>T = 36</td>
<td>T = 1</td>
<td>T = 15</td>
<td>T = 34</td>
<td>T = 7</td>
<td>T = 13</td>
<td>T = 24</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &lt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.24</td>
<td>r = -.69</td>
<td>r = -.79</td>
<td>r = -.63</td>
<td>r = -.05</td>
<td>r = -.04</td>
<td>r = -.05</td>
<td>r = -.05</td>
<td>r = -.26</td>
</tr>
<tr>
<td><strong>H high</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>T = 9</td>
<td>T = 6</td>
<td>T = 9</td>
<td>T = 3</td>
<td>T = 6</td>
<td>T = 26</td>
<td>T = 10</td>
<td>T = 22</td>
<td>T = 13</td>
</tr>
<tr>
<td></td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>r = -.01</td>
<td>r = -.29</td>
<td>r = -.47</td>
<td>r = -.23</td>
<td>r = -.12</td>
<td>r = -.64</td>
<td>r = -.35</td>
<td>r = -.17</td>
<td>r = -.05</td>
</tr>
</tbody>
</table>

Table III: Results of nine Wilcoxon matched pairs signed rank tests comparing differentiation rates between adjacent pairs of stimuli for low-high and high-low functions for each continuum.

Wilcoxon matched pairs rank tests comparing low-high and high-low functions for question- and statement-based continua revealed significant order of presentation effects only for the question-based continua (p < .001).

In sum, the results from the five types of discrimination experiments are very consistent and point to the same conclusion. None of the discrimination tasks performed was able to obtain a discrimination peak in the expected region. On the contrary, the five tasks obtained consistent order-of-presentation effects, indicating that these types of tasks trigger a set of responses that are based on purely acoustic grounds.

### 3.3.4. Four-Interval Discrimination Task

The two graphs in Figure 11 represent the averages of the 20 listeners’ individual d’ scores for each stimulus pair, for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph). Order-of-presentation was not calculated in this data because of inherent ambiguity –recall that four stimuli were presented in each trial, either ABAA or AABA. The graphs show no clear peaks of discrimination in the expected region. Wilcoxon matched pairs rank tests (see Table IV) confirm the significance of three pair comparisons for the question-based continuum, namely, 1_2 vs. 2_3, 2_3 vs. 3_4, and 8_9 vs. 9_10 (shaded values). Yet neither of those corresponds with the category boundary.
130

**Figure 11:** Sensitivity (d') to difference between adjacent stimuli calculated for the four-interval discrimination task, for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph).

<table>
<thead>
<tr>
<th>Stimulus Pairs</th>
<th>1_2 vs 2_3</th>
<th>2_3 vs 3_4</th>
<th>3_4 vs 4_5</th>
<th>4_5 vs 5_6</th>
<th>5_6 vs 6_7</th>
<th>6_7 vs 7_8</th>
<th>7_8 vs 8_9</th>
<th>8_9 vs 9_10</th>
<th>9_10 vs 10_11</th>
</tr>
</thead>
<tbody>
<tr>
<td>!H based stimuli</td>
<td>T = 56</td>
<td>T = 46.5</td>
<td>T = 59.5</td>
<td>T = 45.5</td>
<td>T = 72</td>
<td>T = 76</td>
<td>T = 53</td>
<td>T = 49</td>
<td>T = 108</td>
</tr>
<tr>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td>r = -.05</td>
<td>r = -.18</td>
<td>r = -.006</td>
<td>r = -.10</td>
<td>r = -.16</td>
<td>r = -.35</td>
<td>r = -.007</td>
<td>r = -.23</td>
<td>r = -.23</td>
<td></td>
</tr>
<tr>
<td>H based stimuli</td>
<td>T = 84.5</td>
<td>T = 25.5</td>
<td>T = 72</td>
<td>T = 51.5</td>
<td>T = 69.5</td>
<td>T = 84</td>
<td>T = 55.5</td>
<td>T = 16</td>
<td>T = 112.5</td>
</tr>
<tr>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &lt; .05</td>
<td>p &gt; .05</td>
<td>p &gt; .05</td>
<td></td>
</tr>
<tr>
<td>r = -.48</td>
<td>r = -.04</td>
<td>r = -.04</td>
<td>r = -.11</td>
<td>r = -.01</td>
<td>r = -.32</td>
<td>r = -.06</td>
<td>r = -.64</td>
<td>r = -.04</td>
<td></td>
</tr>
</tbody>
</table>

**Table IV:** Results of nine Wilcoxon matched pairs signed rank tests comparing differentiation rates between adjacent pairs of stimuli for each continuum.

### 3.3.5. Semantic Discrimination Task

Recall that in this task listeners were asked to perform two decisions. First, they were to determine whether the second stimulus was the same or different from the first stimulus and if the second stimulus was different they had to classify it as either being a question or a statement. The rationale behind the design of this labeling decision was that the semantic module would be activated during the whole discrimination task, and even during the first purely acoustic discrimination task. The two graphs in Figure 12 represent the averages of the 20 listeners’ individual d' scores as a function of each stimulus pair and of order of presentation (low-high vs. high-low pairs), for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph). Again we find no clear discrimination peaks aligned with the category boundaries in any of the two graphs, Yet the Wilcoxon matched pairs rank tests (see Table V) show statistically significant differences between the 5_7 and 6_8 pair comparisons continua (only for the low-high comparison) (shaded values), revealing a greater sensitivity to F0 differences in the pairs located at the boundary predicted by the logistic function. Significant order of presentation effects were found for both continua (p < .001).
Figure 12: Sensitivity (d’) to difference between adjacent stimuli calculated for both orders of presentation (low-high and high-low), for H-based or question stimuli (left graph) and for !H-based or statement stimuli (right graph).

Table V: Results of eight Wilcoxon matched pairs signed rank tests comparing differentiation rates between adjacent pairs of stimuli for low-high and high-low functions for each continuum.

<table>
<thead>
<tr>
<th></th>
<th>1.3 vs 2.4</th>
<th>2.4 vs 3.5</th>
<th>3.5 vs 4.6</th>
<th>4.6 vs 5.7</th>
<th>5.7 vs 6.8</th>
<th>6.8 vs 7.9</th>
<th>7.9 vs 8.10</th>
<th>8.10 vs 9.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>!H low-</td>
<td>(T = 53)</td>
<td>(T = 64.5)</td>
<td>(T = 32)</td>
<td>(T = 0)</td>
<td>(T = 66.5)</td>
<td>(T = 40)</td>
<td>(T = 21.5)</td>
<td>(T = 17)</td>
</tr>
<tr>
<td>high</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &lt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
</tr>
<tr>
<td></td>
<td>(r = -.09)</td>
<td>(r = -.06)</td>
<td>(r = -.37)</td>
<td>(r = 0)</td>
<td>(r = -.51)</td>
<td>(r = -.30)</td>
<td>(r = -.14)</td>
<td>(r = -.40)</td>
</tr>
<tr>
<td>!H high-</td>
<td>(T = 45.5)</td>
<td>(T = 46.5)</td>
<td>(T = 66.5)</td>
<td>(T = 0)</td>
<td>(T = 63)</td>
<td>(T = 28.5)</td>
<td>(T = 35)</td>
<td>(T = 30.5)</td>
</tr>
<tr>
<td>low</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
</tr>
<tr>
<td></td>
<td>(r = -.10)</td>
<td>(r = -.08)</td>
<td>(r = .34)</td>
<td>(r = 0)</td>
<td>(r = -.28)</td>
<td>(r = -.09)</td>
<td>(r = -.04)</td>
<td>(r = -.15)</td>
</tr>
<tr>
<td>H low-</td>
<td>(T = 40.5)</td>
<td>(T = 45)</td>
<td>(T = 68)</td>
<td>(T = 63.5)</td>
<td>(T = 11)</td>
<td>(T = 64)</td>
<td>(T = 28.5)</td>
<td>(T = 61.5)</td>
</tr>
<tr>
<td>high</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &lt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
</tr>
<tr>
<td></td>
<td>(r = -.02)</td>
<td>(r = -.11)</td>
<td>(r = 0)</td>
<td>(r = -.30)</td>
<td>(r = -.04)</td>
<td>(r = -.03)</td>
<td>(r = -.36)</td>
<td>(r = -.13)</td>
</tr>
<tr>
<td>H high-</td>
<td>(T = 22.5)</td>
<td>(T = 57)</td>
<td>(T = 84)</td>
<td>(T = 45.5)</td>
<td>(T = 62.5)</td>
<td>(T = 32.5)</td>
<td>(T = 25.5)</td>
<td>(T = 52.5)</td>
</tr>
<tr>
<td>low</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
<td>(p &gt; .05)</td>
</tr>
<tr>
<td></td>
<td>(r = -.02)</td>
<td>(r = -.19)</td>
<td>(r = -.32)</td>
<td>(r = -.27)</td>
<td>(r = -.03)</td>
<td>(r = -.12)</td>
<td>(r = -.33)</td>
<td>(r = -.02)</td>
</tr>
</tbody>
</table>

3.4. Experiment 4: Imitation task

The dependent measure in this imitation study is the scaling of the F0 value of the utterance-final boundary tone. First, two subjects had to be eliminated because their voice properties did not allow Praat to register accurately the final H% boundary tones in their imitations. The results from the productions by 11 subjects are shown below. The 11 graphs in Figure 13 show the mean F0 value (in Hz) of the boundary tone in the imitated versions of the stimuli, for each speaker. At first glance, while some subjects’ F0 data show clear evidence of a bimodal distribution (specifically subjects 3, 6, 8, 10, 13), this is not the case for some other subjects, whose data appear to be somewhat more gradual.
In order to make comparisons across subjects, utterance-final F0 values were normalized into ERBs. In addition, the different stimuli were pooled according to three categories as suggested by the results of Experiment 2, that is, “statement of the obvious” category (from stimulus 1 to stimulus 4), frontier region (stimuli 5-6 since the boundaries between the two categories were located between stimuli 5 and 6 for the question continuum and around stimulus 6 for the statement continuum) and “counterexpectational question” category (from stimulus 7 to stimulus 11). The results of the univariate ANOVA revealed significant effects of Category on Scaling of the boundary tone for all the speakers with the exception of speakers 9 and 12 (see Table VI).

Taken together, the imitation data show evidence of statistically significant differences in the tonal scaling patterns produced when imitating stimuli 1-4 in comparison with the patterns obtained when imitating stimuli 7-11 (which correspond to the non-ambiguous stimuli; see Experiment 2), and also in comparison with the patterns obtained when imitating ambiguous stimuli 5-6. Thus, it seems that listeners’ imitations are consistent with their perception of two distinct tonal categories. Yet the fact that we did not find an effect of stimulus number for two of the speakers is a warning sign about using this task: it seems likely that some speakers are poor imitators or are simply not
able to perform the task successfully. In general, though, the data support the presence of discrete categories for at least eight out of eleven speakers.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>F(2, 63) = 43.691, p &lt; .0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker 3</td>
<td>F(2, 60) = 36.406, p &lt; .0001</td>
</tr>
<tr>
<td>Speaker 5</td>
<td>F(2, 44) = 5.341, p &lt; .01</td>
</tr>
<tr>
<td>Speaker 6</td>
<td>F(2, 62) = 57.282, p &lt; .0001</td>
</tr>
<tr>
<td>Speaker 7</td>
<td>F(2, 59) = 49.285, p &lt; .0001</td>
</tr>
<tr>
<td>Speaker 8</td>
<td>F(2, 63) = 21.863, p &lt; .0001</td>
</tr>
<tr>
<td>Speaker 9</td>
<td>F(2, 59) = 2.106, p &gt; .05</td>
</tr>
<tr>
<td>Speaker 10</td>
<td>F(2, 48) = 33.714, p &lt; .0001</td>
</tr>
<tr>
<td>Speaker 11</td>
<td>F(2, 61) = 80.677, p &lt; .0001</td>
</tr>
<tr>
<td>Speaker 12</td>
<td>F(2, 46) = .874, p &gt; .05</td>
</tr>
<tr>
<td>Speaker 13</td>
<td>F(2, 62) = 37.448, p &lt; .0001</td>
</tr>
</tbody>
</table>

**Table VI:** Summary of the results of the univariate ANOVAs for each subject in the imitation experiment.

4. Discussion and conclusions

One of the goals of the present investigation was to analyze the performance of several experimental methodologies that have been used to study intonational contrasts. The results of the experiments presented in this chapter can be summarized as follows. First, the results of Experiment 1 (semantic congruity task) show that the two target intonation contours (question vs. statement contours) are mainly appropriate in their respective semantic contexts and that incongruent situations are heavily penalized by the listener.

As expected, evidence of two categories was obtained in the identification experiment (Experiment 2); importantly, a mean RT peak was observed at the identification boundary for both continua. However, the results from the five types of discrimination task (Experiment 3) failed to show evidence of the presence of a clear discrimination peak in between categories. Finally, evidence from the imitation study (Experiment 4) showed that the majority of speakers produced two groups of productions when imitating the different range of stimuli.

Given previous results in the literature, the conflicting evidence from the discrimination tasks was an expected outcome. Previous studies had shown that even in cases where identification results show a dual contrast, discrimination results are very weak or null. For example, no discrimination peaks were found by Falé and Hub Faria (2006) or Gili-Fivela (2008), plateaux rather than peaks were found by Schneider and Linfert (2006), and two peaks were found by Dilley (2007). As noted above, the problems reported with the discrimination tasks are of various kinds. One of the conflicting factors is the choice of step size, as well as the time of the interstimulus interval (Newport 1982, Repp 1984). Many researchers have pointed out that the discrimination task may be more properly considered a purely perceptual or auditory
task that can be answered in an auditory mode (Newman 1984, Chen 2003, Gussenhoven 2004, 2006, Niebuhr and Kohler 2004, Kohler 2006, Gerrits & Schouten 2004). While in the identification task subjects are asked to map their perception patterns onto cognitive and linguistic categories, this is not the case in the classical acoustic-oriented discrimination tasks. That is why Gerrits and Schouten (2004), Schneider et al. (2009) and Frota (in press-a) tested other types of discrimination tasks intended to trigger semantic access. We performed several variants of the discrimination task, including Gerrits and Schouten’s (2004) proposal of a four-interval discrimination task and a semantic discrimination task that was intended to activate the listeners’ labeling/semantic component, with no success in any of the tasks.

At this point, together with Chen (2003), Niebuhr and Kohler (2004) and Kohler (2006), we might be tempted to claim that discrimination tasks (and thus the CP paradigm) are not suitable for investigating discreteness in intonational contrasts. The fact is that standard discrimination tasks are auditory in nature and systematically obtain asymmetrical effects in the perception of intonational contrasts (order of presentation effects) rather than tonal categories. In such acoustic-oriented tasks, the comparison of pairs of stimuli is too focused on acoustic grounds rather than on their phonological patterning. Thus these tasks are not successful in tapping into the linguistic nature of the contrasts. In this chapter, even in the two tasks that could be claimed to activate the speakers’ semantic/labeling component, we obtained null results.

Yet recent evidence has shown that a context-matching discrimination task has been shown to provide successful results, both for alignment (Schneider et al. 2009), and for scaling (Frota in press-a). The procedure followed by these two studies differed from our semantic discrimination task. In the case of Frota (in press-a), a context appeared before the presentation of each pair of stimuli, and listeners had to decide with of the stimuli in the pair was appropriate to the context. In the case of Schneider et al. (2009), each stimulus consisted of two phrases, namely, the context plus the target phrase, and speakers had to decide whether the two stimuli were the same or different. In our semantic task, subjects had to perform a standard discrimination task, followed by a labeling task if the second item was perceived to be different. Thus, the importance of the semantic dimension in the design of phonetic discrimination tasks is still an open question, and it is clearly an area for further research.

In light of the results reported in this article, we claim that the converging evidence from three types of tasks (congruity, identification, and possibly imitation) represents a clear argument in favor of a discrete mid boundary tone category in Catalan. Perhaps one of the most powerful tests of discreteness in intonation is the identification task. Against this idea, one may argue that when listeners are faced with a continuum and must make a binary decision they are forced to automatically categorize the exemplars into two categories (e.g., Ladd and Morton 1997, see also recent evidence from Borràs-Comes et al. 2010 who show that identification tasks can induce an automatic and artificial separation of the data into two categories). To test this hypothesis, we conducted a pilot identification experiment identical to Experiment 2 but using only half of the phonetic space. In each one of the blocks, listeners heard stimuli from only one half of the two continua (i.e., only stimuli 1 to 5 for each continuum) and had to respond
either ‘yes’ (corresponding to the question interpretation) or ‘of course’ (statement interpretation). Five listeners participated in the experiment and we obtained a total of 150 responses (5 stimuli from the question continuum + 5 stimuli from the statement continuum x 3 blocks x 5 subjects). Figure 14 shows the rate of identification responses (and standard error values) as a function of stimulus step number (averaged over subjects). The responses clearly show that when listeners hear only half of the continuum in each one of the blocks, they do not automatically classify them in a binary way (e.g., stimulus 5 of both continua appears at chance level).

Figure 14: Proportion of identification responses (and standard error values) as a function of stimulus step number (averaged over subjects).

With respect to the methods to be employed for the investigation of intonation categories, special attention has to be paid to the semantic relevance of the task, and discourse meanings have to be available through semantically motivated tasks or tasks where contextual information is clearly accessible. In this sense, semantic congruity tasks are especially sensitive for the analysis of phonological categorization in natural discourse settings. Our congruity study revealed a partial overlap between the two intonational categories and a higher degree of acceptability of non-exemplary intonation in incongruent situations. Even though this is not among the issues being investigated in this chapter, it is an important question in the field and we believe that this task can represent an adequate tool to analyze the status of intonational categories within a discourse context. Thus intonational categories could be reliably investigated with a combination of semantic congruity tasks and identification tasks (and possibly with context-matching discrimination tasks; see Schneider et al. 2009, Frota in press-a). Results from imitation tasks, on the other hand, have to be taken with caution and more research is needed to evaluate the imitation capabilities.

The idea of enriching the traditional Pierrehumbert H/L dichotomy with a finer differentiation of pitch range is not new and has long been advocated by researchers
such as Ladd (1993, 1994). He states that “The central point of the descriptive proposals I have made here and elsewhere is that the Bruce-Pierrehumbert approach to intonational phonology must be enriched with a notion of categorical distinction of pitch range. We need to get rid of the idea that any distinction that is orthogonal to the basic opposition between High and Low tones is *ipso facto* gradient; both gradient factors and categorical ones play a role in the vertical scale of any given tone” (Ladd 1994:60).

Even though the AM framework assumes a primary phonological opposition of pitch between high (H) and low (L), a good number of languages have been documented to have at least a three-way distinction in height at the boundary tone level (see Beckman & Ayers-Elam 1997, for American English; Grabe 1998, for British English and German; Post 2000, for French; Beckman et al. 2002, for Spanish; Lee 2003, for Korean; Arvaniti & Baltazani 2005, for Greek; Grice et al. 2005 and Peters 2006, for German; Frota in press-b, for European Portuguese). Even though there is a need for more experimental work to further describe the scaling contrasts found in boundary position in these languages and to determine whether those contrasts are categorical or gradient in nature, evidence is accumulating that edges in many languages can be marked by (at least) three levels of tonal height. Crosslinguistic evidence reveals that at the edge of a prosodic unit, languages may use many tonal contrasts to represent different pragmatic meanings. This is supported by Korean, which has more than 9 boundary tone combinations with their own pragmatic meaning (all describable in terms of L and H but differing by a number of tonal changes: L, H, LH, HL, LHL, HLH, LHLH, HLHL, LHLHL (Jun p.c.). Just as there are tonal languages with two or three distinct level tones 37, there is accumulating crosslinguistic evidence that intonation languages may need more than two level tones to satisfactorily represent intonation contrasts.

Different authors have proposed different representational schemes to express phonological distinctions in height at the boundary level. Some systems have encoded directly a mid M% boundary tone, such as Beckman et al. (2002) for Spanish or Lee (2003) for Korean. On the other hand, other authors have used the 0% tone symbol to indicate that the pitch stays level after the nuclear accented syllable (see Grabe’s 1998 analysis of British English and German, Post 2000 analysis of French, and Peter’s 2006 analysis of German). Finally, another solution has been to analyze non-high boundary tones as !H% (e.g., Grice et al. 2005 for German, Arvaniti & Baltazani 2005 for Greek, Frota in press-b for European Portuguese). This symbol of downstepped !H% has been used by some authors to indicate that the resulting tone was the result of downstep rules. For example, Grice et al. (2005) analyze an utterance-final mid-level tone in German as a H% that is downstepped by a preceding L-phrase accent, while Beckman and Ayers-Elam (1997) analyze a similarly scaled tone in English as a L% boundary tone that is upstepped by a preceding H-phrase accent. Following Ladd (1983, 1984), other authors have used the downstepped !H% tone to express not merely an automatic process of

37 For example, in Thai or African languages there is a distinction between lexical tones that are High and Super-high (McHugh 1999, Rose 1997).
phonetic realization, but a separate phonological feature that may be selected in its own right. For Gr_ToBI, Arvaniti and Baltazani claim that there is a phonological distinction in Greek—that is, a meaningful intonational choice—between downstepped (H%) and non-downstepped (H%) final high boundary tone. Similarly, Frota (in press-b) analyzes the greeting call in European Portuguese as (L+)H* !H%, with spreading of !H% in the post-tonic stretch. The extended duration of the boundary syllable is taken to go hand in hand with the special nature of !H% in this contour. Currently there is ambiguity in the way ToBI systems use the !H% unit either as a phonetic or as a phonological category, that is, either as a result of a downstep rule, or to represent a paradigmatic intonational contrast at the phonological level.38

In Catalan, a series of non-high boundary tones, namely those produced at the end of vocatives, obvious statements, and contradiction statements, share the phonetic space occupied in between low and high boundary tones. For Cat_ToBI, we initially proposed the presence of a mid tone in the system M% (see Prieto et al. 2009), as Beckman et al. (2002) did for Spanish ToBI. There were two main reasons to use this phonological unit. First, this unit represented a transparent choice that directly encodes the three-way phonological distinction in height. Moreover, in Catalan the non-high boundary tone analyzed in this study is not predictable from preceding L and H edge tones or nuclear configurations, and is independent from the phonological process of downstep, or successive lowering of H tones (see the minimal contrasts in Catalan intonational phonology in Figure 1). In essence, the mid tone category (M%) was allowing us to represent a non-high boundary level, reserving the tone !H to indicate the original stepping mechanism among H tone targets. Yet even though our initial representational choice for mid boundary tones in Catalan was M% over the other two options, the most recent version of this system uses the !H% symbol, because of two important reasons: first, in order to be able to offer a consistent analysis of pitch height contrasts both at the pitch accent level and at the boundary tone level; in Catalan, as well as in other Romance languages, we have a two-way distinction at the H level in rising pitch accents which we express as L+H* vs L+¡H* (see Borràs-Comes et al. 2010); second, in order to be able to offer a more general transcription analysis of the same contour in other European languages (e.g., German, Greek, and Portuguese).

Thus the downstep !H feature represents two different kinds of sources in Catalan, as well as in other languages, either an independent choice (arguably with meaning consequences), or appear as a result of a contextual process (such as when H follows another H tone or when H precedes X within a certain configuration). As Frota (p.c.) points out, this situation would be no different from what happens rather frequently in segmental phonology, where a segment (or feature) may be distinctive in the sound system of a language and also generated by rule in specific contexts (as in the case of

---

38 As has been recently pointed out by Rathcke and Harrington (2006, 2010), “if we allow the system to contrast H+L* and H+¡H* pitch accents, then this means that the frequency scaling of the (starred) tones that are associated with the accented syllable is no longer phonetic as argued in e.g., Pierrehumbert (1980), but is implicitly phonological because there is now a paradigmatic contrast between three tonal levels: high-star in (L+)H*, low-star in H+L* and effectively a mid-star in H+¡H*.
voiceless plosives in Dutch, which are phonemically contrastive and also can be contextually motivated in word-final position.

To sum up, this chapter has presented a set of experiments that aimed at investigating the phonological status of two rising boundary tones in Catalan and their use for different discourse functions. Two relevant areas of current intonation research have been addressed, namely, (a) the nature of intonational contrasts within AM theory of intonation; and (b) methodological issues which arise when the nature of intonational variation is explored experimentally.

In general, the results from a set of four experiments provided evidence that Catalan listeners perceive a contrast between two different types of rises in phrase-final position that are used for two different discourse functions, namely counterexpectational questions and statements of the obvious. Results of a semantic congruity task (Experiment 1) showed that Catalan listeners are very sensitive to the pragmatic appropriateness of the two target intonation contours with their discourse context. Evidence for having two categories was also obtained in an identification experiment (Experiment 2). Yet, the results from the five types of discrimination tasks (Experiment 3) failed to show evidence of the presence of a clear discrimination peak in between categories, even in tasks that were designed to activate a semantic labeling component. Evidence from the imitation study (Experiment 4) showed that the majority of speakers produced two groups of productions when imitating the different range of stimuli. In sum, results from three of our four experiments, namely the congruity task, the identification task, and the imitation task, support the presence of a discrete intonational contrast between two types of boundary tones in Catalan.

One of the valuable contributions of this investigation is the comparable data obtained using several experimental techniques which investigate the same phenomenon, something which allows to perform a comparison of their sensitivity. The null results from the five types of discrimination tasks were a predictable result and provide support for the conclusions drawn by earlier studies in trying to evaluate the application of discrimination tasks and the CP Paradigm to the investigation of intonational categories. In these five experiments, we controlled for the factors that had been previously identified as potential problems with the discrimination task, such as the choice of step size (Newport 1982, Repp 1984), the duration of the ISI (interstimulus interval; Massaro 1974), as well as the type of task (Gerrits & Schouten 2004). Yet null results were obtained notwithstanding. Even though we might be tempted to join Chen (2003), Kohler (2004), and Niebuhr and Kohler (2004) in claiming that discrimination tasks are not suitable for investigating discreteness in intonational contrasts (recall that perceptual asymmetries are obtained consistently in all types of discrimination tasks), we surely have to restrict this property to acoustic oriented discrimination tasks. As for semantic oriented discrimination tasks, two recent investigations have performed context-matching discrimination tasks with successful results, both in alignment (Frota in press-a) and in scaling (Schneider et al. 2009). We believe that the issue of the semantic relevance of the discrimination tasks and their adequacy in the investigation of intonation categories is very important and deserves further investigation.
In methodological terms, we suggest that a combination of semantic congruity tasks and identification tasks (involving RT measures) can be profitably used to investigate intonational categories. On the other hand, it should be stressed that imitation data must be analyzed with caution, as some speakers were not able to successfully perform this kind of task and we seem to have two groups of results. We believe that the tasks employed for the investigation of intonation categories have to pay special attention to the semantic relevance of these units and ensure contextual and semantic accessibility. In this respect, the congruity tasks are a very promising tool to make perceptual testing more ‘ecologically valid’, as listeners performing this task have been shown to be especially sensitive to the appropriateness of intonation contours in given discourse meanings.
Part IV
Narrow contrastive focus and broad focus declaratives
CHAPTER 6

The contribution of alignment, duration and scaling to the perception of contrastive focus in Catalan, Italian and Spanish

This chapter investigates the relevance of three prosodic parameters (alignment, duration and scaling) in the conveyance of narrow contrastive focus. In this study, Catalan is compared with two typologically related languages, Italian and Spanish. In particular, we seek to determine how the Effort Code is instantiated in the expression of narrow contrastive focus both in the production and in the perception of speech. According to the Effort Code putting more effort on speech production will lead to greater articulatory precision (de Jong 1995). The Effort Code is related to the expression of focus in the sense that wider pitch excursions will be used to signal meanings that are relevant from an informational point of view. A dual production and perception experiment based on a identification task (together with RT measurements) will be conducted. This is the only chapter in which Catalan is compared to other Romance languages.

1. Introduction

It is well known that languages differ in the ways that they express narrow contrastive or corrective focus, that is, in the way they mark a constituent that is the direct rejection of an alternative (i.e., “it’s B, not A”). According to Elordieta (2007), a cross-linguistic typological distinction can be established depending on how contrastive focus is marked: (a) by means of intonation, (b) via syntactic movement optionally combined with prosodic mechanisms to ensure prominence or (c) by using specific morphemes also optionally combined with syntactic displacement, prosodic marking or a combination of both. In languages such as English or German the intonational prominence may be shifted to different positions in the clause while the syntactic structure remains constant. On the other hand, languages such as Catalan, Italian or Spanish use word order for focus marking (Vallduví 1991, Ladd 199640), although recent studies show that in these languages intonational strategies are as acceptable as syntactic ones (Estebas-Vilaplana 2009 for Catalan; Face & D’Imperio 2005 for Italian and Spanish), especially when word order is kept invariant. With respect to prosody,

39 We would like to thank Barbara Gili-Fivela and Antonio Stella for their help in the general design of the experiment and for the collection and analysis of the Italian data.
40 Vallduví (1991: 295) states that “contrary to English, in Catalan intonational prominence is fixed on clause-final position and syntactic operations must be used to make the focus (or a subset of it) fall under prominence”. Likewise, Ladd (1996: 191) asserts that “word order modifications in languages like Spanish and Italian may indirectly achieve the accentual effects that English accomplishes directly by manipulating the location of the nuclear accent”.

various prosodic cues have been shown in a number of languages to distinguish narrow contrastive focus from broad focus (henceforth NCF and BF) sentences: alignment of the peak of the focal accent with respect to segmental landmarks such as the onset or offset of the accented syllable, pitch range (also pitch scaling of the peak) of the focal accent and duration of the syllable bearing the focal accent. As far as alignment is concerned, crosslinguistic evidence has shown that NCF is characterized by the use of retracted pitch peaks (De-la-Mota 1995, 1997 and Beckman et al. 2002 for Spanish; Estebas-Vilaplana 2009 for Central Catalan; D’Imperio 2002 and Gili-Fivela 2002, 2006, 2008 for Neapolitan and Pisa Italian respectively; Smiljanic 2004 for Serbo-Croatian; Manolescu et al. 2009 for Romanian). On the other hand, pitch range (or tonal scaling) as a prosodic marker of focus across languages is a controversial issue. In languages such as Romanian (Manolescu et al. 2009) and Serbo-Croatian (Smiljanic 2004), NCF leads to greater pitch excursion on the accented syllable, while for Italian and Spanish it is a debated question among researchers whether F0 pitch range increase is or is not an acoustic correlate of NCF (D’Imperio 2002 and Gili-Fivela 2005, 2006 for Italian; De-la-Mota 1995, 1997 and Face 2001, 2002 for Spanish). The importance of duration in the conveyance of NCF has been claimed in studies such as those of Eady et al. (1986) for English, Jun and Lee (1998) for Korean, and Baumann et al. (2007) and Kügler (2008) for German. Although it is apparent from the abovementioned studies that languages employ different prosodic strategies to convey the pragmatic contrast between NCF and BF, very few studies have analyzed the interplay between these cues in a systematic way.

One related issue that is still relatively under-explored concerns the nature of these prosodic strategies that are used as focus marking. According to a theory regarding the biological foundations of intonation such as Gussenhoven’s (2002, 2004) biological codes and specifically according to the so-called Effort Code, there is a correlation between a wider excursion of the pitch movement and an increase in the importance of the message from an informational point of view. In other words, the interpretation is that the speaker is being forceful because s/he believes the contents of her/his message are relevant. Thus, the grammaticalisation of the information interpretation is typically found in the expression of focus in which the intonational structure will favor a situation whereby focused information will be characterized by relatively wide pitch excursions (Gussenhoven 2004).

The present study aims to explore the specific contribution of tonal alignment, duration and pitch height as well as the interplay between these prosodic cues in the expression of NCF in Catalan, Italian and Spanish. Previous research conducted on the contrast between BF and NCF initial accents for Catalan (Prieto et al. 2005) and Castilian Spanish (De-la-Mota 1995; Nibert 2000; Face 2001; Hualde 2002, 2003) conclude that although the main difference is based on a different alignment of the peak (BF accents have late peaks, \( L+\rightarrow H^* \), while NCF accents have earlier F0 peaks, \( L+H^* \), see Figure 1, left panel), a post-focal pitch reduction or a higher F0 peak height can be found in NCF utterances. Likewise, in the varieties of Italian spoken in both Pisa (Gili-Fivela 2006, 2008) and Lecce (Stella & Gili-Fivela 2009) BF initial accents have late peaks (\( L+H^* \)), while NCF accents have earlier F0 peaks (\( H^*+L \)) (see Figure 1, right
Interestingly, in both varieties the nuclear L+H* accent appears to be higher in F0 (possibly against the prediction of the Effort Code) and duration is found to be greater for syllables bearing the NCF accent (for Pisa Italian, Gili-Fivela 2005, 2006).

Figure 1: Schematic representation of the alignment difference between NCF and BF initial accents for Catalan and Spanish (left panels) and Italian (right panels). Shaded boxes represent accented syllables.

The motivation behind choosing these three languages is that they provide a good test case for the crosslinguistic predictions of the Effort Code in expressing the NCF through higher and retracted peaks. While Italian would contradict the general trend proposed by this code (as accents found in NCF have earlier but lower peaks), Catalan and Spanish would agree with its predictions.

The present study will investigate the prosodic patterns of BF and NCF initial accents in Catalan, Italian and Spanish declarative sentences from both a production and a perception perspective. We seek to determine how the informational interpretation of the Effort Code is found in the expression of NCF both in the production and in the perception of speech. Our hypothesis is that, regardless of whether the languages under study follow or not the general trend proposed by the Effort Code in production, the marking of BF and NCF found in production will be mirrored at the perceptual level. This hypothesis is consistent with Exemplar Theory (Bybee 2006), for which language input is stored in memory in rich detail as exemplars and that targets for subsequent productions are derived from them, or with theories about language acquisition for which the conventional assumption is that perception precedes production and there should therefore be a correspondence between production and perception.

In order to test this hypothesis, we conducted a production experiment in which the data obtained for each of two types of focus and for the three languages respectively were analyzed acoustically. In the second experiment, we tested the specific contribution of tonal alignment, duration and tonal scaling through a perception experiment based on an identification task (together with RT measurements). This study is innovative because it seeks to shed light not only on the specific prosodic cues used in the conveyance of NCF but also on the interaction that seems to exist among them. From the methodological point of view, it will also add to the current body of research on specific varieties of Catalan (Majorcan) and Italian (Lecce) that have not been studied previously in the field of focus while allowing us to compare different languages following the same controlled methodology.

The chapter is organized as follows. Section 2 presents the methodology (participants, materials and procedure) and the results for the production experiment. Section 3 presents the methodology and results for the perception experiment, that is, the identification task together with an analysis of the Reaction Time measurements and the logistic regressions. Finally, section 4 discusses the major findings of this research and states its overall conclusions.
2. Experiment 1: Production experiment

The aim of this experiment was to investigate the prosodic features of alignment, pitch range and pitch scaling, and duration of the initial accents in BF and NCF utterances in Catalan, Italian and Spanish. To this end a corpus was designed which contained declarative sentences with BF and NCF initial accents with two different stress positions. Five native speakers for each language participated in this production experiment, reading a total of 300 sentences per language. It was expected that the different pragmatic meanings might trigger differences in alignment, pitch range or scaling of the focus accents as well as differences in duration of the syllable bearing the focus accent. Additionally, this production study was expected to prove the existence of crosslinguistic differences in the expression of the Effort Code in NCF utterances.

2.1. Methodology

2.1.1. Participants

Five native speakers of Majorcan Catalan (3 women, 2 men) aged between 20 and 36 (mean = 25.8), five native speakers of Lecce Italian (3 women, 2 men) aged between 22 and 29 (mean = 26.4) and five native speakers of Madrid Spanish (3 women, 2 men) aged between 35 and 45 (mean = 38.4) participated in this experiment. All participants had been born and raised in their respective geographic regions. They reported having little or no prior training in phonology/phonetics and no history of hearing disability.

2.1.2. Experimental materials and elicitation method

The corpus that has been used in this study contains declarative sentences with BF or NCF in initial position. Our definition of broad focus is a carrier of new information in which there is no particular constituent that is focalized, which is equivalent to saying that the whole expression is the focus constituent. On the other hand, contrastive focus “marks a constituent that is a direct rejection of an alternative,” which implies the exclusion of an alternative (Gussenhoven 2007). For instance, when the sentence Mary is coming is pronounced as an answer to the question What’s happening? the entire sentence is new information with no specific element emphasized. However, when the same sentence is an answer to the question Is Peter coming?, Mary is highlighted and is in contrastive focus (Manolescu et al. 2009).

The discourse setting materials in our production experiment had a simple syntactic structure, namely SVO sentences with a subject NP and object NP (or PP in some cases) and a verbal VP, all consisting of one single constituent (see Table I for examples for each of the languages under study).

Whenever possible, the target words were composed of voiced consonants (to avoid segmentally-induced effects on the F0 curve). The target words always contained the stress on the penultimate or antepenultimate syllable to provide more room for tonal
accent realization and all consisted of three syllables (see Table II for some examples of the target words used for each language). Words were chosen on the basis of the following three criteria: a) they should be high frequency words that did not present problems of pronunciation; b) at least the consonant and the vowel of the accented syllable should share place of articulation across the three languages; and c) the syllable structure of the syllable bearing the focus accent should be the same across the three languages.

<table>
<thead>
<tr>
<th>Broad focus</th>
<th>Narrow Contrastive focus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paroxytone target words</strong></td>
<td><strong>Paroxytone target words</strong></td>
</tr>
<tr>
<td>Catalan</td>
<td>Catalan</td>
</tr>
<tr>
<td>Q: <em>Què t’han dit?</em> (<em>What did they tell you?</em>)</td>
<td>Q: <em>Na Tina?</em> (<em>pers. article Tina [proper name]</em>)</td>
</tr>
<tr>
<td>A: <em>Que na Marina vendrà demà</em> (<em>That Marina is coming tomorrow</em>)</td>
<td>A: <em>No, na MARINA vendrà demà</em> (<em>No, MARINA is coming tomorrow</em>)</td>
</tr>
<tr>
<td><strong>Italian</strong></td>
<td><strong>Italian</strong></td>
</tr>
<tr>
<td>Q: <em>Hai saputo qualcosa?</em> (<em>Have you found out anything?</em>)</td>
<td>Q: <em>La Sabina?</em> (<em>pers. article Sabina [proper name]</em>)</td>
</tr>
<tr>
<td>A: <em>Sì, la Marina verrà domani</em> (<em>Yes, Marina is coming tomorrow</em>)</td>
<td>A: <em>No, la MARINA verrà domani</em> (<em>No, MARINA is coming tomorrow</em>)</td>
</tr>
<tr>
<td><strong>Spanish</strong></td>
<td><strong>Spanish</strong></td>
</tr>
<tr>
<td>Q: <em>¿Qué te han dicho?</em> (<em>What did they tell you?</em>)</td>
<td>Q: <em>¿Tina?</em> (<em>pers. article Tina [proper name]</em>)</td>
</tr>
<tr>
<td>A: <em>Que Marina vendrá mañana</em> (<em>That Marina is coming tomorrow</em>)</td>
<td>A: <em>No, MARINA vendrá mañana</em> (<em>MARINA is coming tomorrow</em>)</td>
</tr>
<tr>
<td><strong>Proparoxytone target words</strong></td>
<td><strong>Proparoxytone target words</strong></td>
</tr>
<tr>
<td>Catalan</td>
<td>Catalan</td>
</tr>
<tr>
<td>Q: <em>Què t’han dit?</em> (<em>What did they tell you?</em>)</td>
<td>Q: <em>Sa màquina?</em> (<em>The machine?</em>)</td>
</tr>
<tr>
<td>A: <em>Que sa làmina no està acabada</em> (<em>That the sheet is not finished</em>)</td>
<td>A: <em>Sa LÀMINA no està acabada</em> (<em>The sheet is not finished</em>)</td>
</tr>
<tr>
<td><strong>Italian</strong></td>
<td><strong>Italian</strong></td>
</tr>
<tr>
<td>Q: <em>Hai saputo qualcosa?</em> (<em>Have you found out anything?</em>)</td>
<td>Q: <em>La pagina?</em> (<em>The page?</em>)</td>
</tr>
<tr>
<td>A: <em>Sì, la lamina è piegata</em> (<em>Yes, the sheet is folded</em>)</td>
<td>A: <em>No, la LAMINA è piegata</em> (<em>No, the SHEET is folded</em>)</td>
</tr>
<tr>
<td><strong>Spanish</strong></td>
<td><strong>Spanish</strong></td>
</tr>
<tr>
<td>Q: <em>¿Qué te han dicho?</em> (<em>What did they tell you?</em>)</td>
<td>Q: <em>¿La màquina?</em> (<em>The machine?</em>)</td>
</tr>
<tr>
<td>A: <em>Que la lámina no está acabada</em> (<em>That the sheet is not finished</em>)</td>
<td>A: <em>La LÁMINA no está acabada</em> (<em>The SHEET is not finished</em>)</td>
</tr>
</tbody>
</table>

**Table I:** Examples of the series of question-answer pairs used in this experiment. Target words are marked in boldface.
Table II: Examples of target words used for each language grouped according to stress pattern (paroxytone and proparoxytone).

<table>
<thead>
<tr>
<th></th>
<th>Catalan</th>
<th>Italian</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paroxytone</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>target words</td>
<td>Marina</td>
<td>Marina</td>
<td>Marina</td>
</tr>
<tr>
<td>(proper name)</td>
<td>[məˈɾina]</td>
<td>[maˈrina]</td>
<td>[maˈrina]</td>
</tr>
<tr>
<td>bombona</td>
<td>[bomˈbona]</td>
<td>remone</td>
<td>bombona</td>
</tr>
<tr>
<td>(gas cylinder)</td>
<td>(big oar)</td>
<td>[reˈmone]</td>
<td>(gas cylinder)</td>
</tr>
<tr>
<td><strong>Proparoxytone</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>target words</td>
<td>làmina</td>
<td>làmina</td>
<td>làmina</td>
</tr>
<tr>
<td>(sheet)</td>
<td>[ˈlaˈmina]</td>
<td>[ˈlaˈmina]</td>
<td>[ˈlaˈmina]</td>
</tr>
<tr>
<td>Mòdena</td>
<td>[ˈmoðena]</td>
<td>Monaco</td>
<td>Mòdena</td>
</tr>
<tr>
<td>(Modena, Italian city)</td>
<td>(Modena, Italian city)</td>
<td>[ˈmonako]</td>
<td>[ˈmoðena]</td>
</tr>
</tbody>
</table>

Five speakers per language read three repetitions of the two blocks (paroxytone and proparoxytone target words) for a total of 300 sentences per language (5 sentences x 2 focus types x 2 stress positions x 3 repetitions x 5 speakers). In order to facilitate the reading and interpretation of the sentences by the speakers, sentences were elicited by means of questions, which were read by the experimenter (see Table I). The series of question-answer pairs were presented through a PowerPoint presentation at random irrespective of the stress pattern. Speakers were asked to read the sentences at a normal speech rate and avoid marked readings and pauses. The questions to elicit the NCF sentences were designed so that they rhymed or were very similar to the BF target words (see Table I).

Catalan speakers were recorded in a quiet room at the Universitat Pompeu Fabra and Spanish speakers were recorded under similar conditions at the Universidad Autónoma de Madrid using a Marantz Professional PMD660 digital recorder and Rode NTG-2 microphone. Italian speakers were recorded in an anechoic room at the Centro di Ricerca Interdisciplinare sul Linguaggio in Lecce using a Shure SM-86 condenser microphone and Goldwave 5.25 software installed on a desktop computer equipped with an Edirol UA-5 external sound card. The sentences were digitized at a 44100 Hz sample rate.

2.1.3. Analysis

The recordings were digitized and processed using Praat (Boersma & Weenink 2009). The intonational patterns of the utterances were analyzed by means of auditory analysis and inspection of the F0 traces. Spanish speaker 4 had a low F0 with creaky voice and this made it difficult for Praat to collect L points. As a consequence, five utterances by this subject were eliminated from the database. The final database for Spanish thus contained a total of 295 utterances. The other two languages contained 300 examples each.

Six segmental and two tonal targets were labelled by hand, using the following segmental and tonal labels:
**Segmental labels**
- Onset of the target utterance (o)
- Onset of the first syllable of the target word (osyl)
- Onset of the syllable with which the accent is associated (oasy)
- Offset of the syllable with which the accent is associated (ofasy)
- Offset of the target word (ofw)
- End of the utterance (e)

**Tonal labels**
- Initial F0 valley value in focus accents (L)
- F0 peak value in focus accents (H)

Measurements corresponding to these labels were collected by Praat software automatically through a script into a file in .txt format. With respect to pitch timing, two measures of H and L alignment were used for purposes of subsequent statistical exploration: (i) distance from point L to the onset of the syllable bearing the focus accent (see Figure 2, left panel), and (ii) distance from point H to end of the syllable bearing the focus accent (see Figure 2, central panel). Pitch range was measured as the distance between the peak (H*) and the preceding low target (see Figure 2, right panel).

![Figure 2: Illustration of the measures of alignment and pitch range performed for the data.](image)

Figures 3, 4 and 5 illustrate the orthographic, segmental and prosodic transcription of the utterance ‘Marina will come tomorrow’ produced with BF and NCF meaning for each of the three languages under study. The first horizontal tier contains the orthographic transcription, while the segmental and tonal codes appear in the other tiers. The second tier marks the segmental labels explained above, whose number can vary depending on the number of syllables of the target word. Finally, the third tier contains the tonal information regarding the focus accents.

These are the criteria that were followed in order to determine the tonal landmarks:

- *Initial F0 valley value in focus accents.* In the minima, the L was located just before the onset of the rise at the point in which the F0 rising velocity increases. This roughly corresponded in Catalan and Spanish with the onset of the syllable bearing the focus accent while in the case of Italian the L target could occur before the accented syllable.
-F0 peak value in focus accents. The peak was generally easy to pinpoint and was situated just before the fall. Sometimes a plateau was found. When this happened, the H was placed at the end of the plateau, immediately before the contour started to fall.

**Figure 3:** Waveform, spectrogram, F0 contour and intonational labelling used for the Catalan BF utterance *Que na Marina vendrà demà* ‘That Marina is coming tomorrow’ (left) and the Catalan NCF utterance *Na MARINA vendrà demà* ‘MARINA is coming tomorrow’ (right) (Catalan speaker 5) [Chap6Fig3.wav].

**Figure 4:** Waveform, spectrogram, F0 contour and intonational labelling used for the Italian BF utterance *BF Sì, la Marina verrà domani* ‘Yes, Marina is coming tomorrow’ (left) and the Italian NCF utterance *No, la MARINA verrà domani* ‘No, MARINA is coming tomorrow’ (right) (Italian speaker 3) [Chap6Fig4.wav].
2.1.4. Statistical analyses

In this experiment repeated measures ANOVA tests were performed because we had a situation in which the same people take part in different experimental conditions (and not different people taking part in different experimental conditions). Different measures were used as dependent variables for statistical exploration, namely (a) H Alignment, (b) L Alignment, (c) Pitch Range of the focus accent, (d) H Scaling of the focus accent, (e) L Scaling of the focus accent, and (f) Duration of the accented syllable. For each one of the dependent variables, we ran a repeated measures ANOVA with Focus Type (two levels) as the within-subject factor and Language (three levels) as the between-subject factor across Subjects (F1) and Items (F2). The results from the repeated measures ANOVA tests include the $F$-ratio, the degrees of freedom from which it was calculated and the significance value.

2.2. Results

2.2.1. Alignment results

Figure 6 shows the box plots of the distance from H to the end of the accented syllable in milliseconds for BF (dark grey boxes) and NCF (light grey boxes) for all speakers of Catalan (left), Italian (centre) and Spanish (right). Observe that there is a substantial difference between median points of BF and NCF for Catalan and Spanish on one hand and Italian on the other. Thus, in the case of Catalan and Spanish the peak for NCF is
realized at the end of the accented syllable (with the exception of Spanish speaker 5) while for BF the peak is systematically aligned after the end of the accented syllable. For Italian, the situation is a little bit different since both BF and NCF peaks are aligned before the end of the accented syllable. However, the NCF peaks are aligned always earlier than BF peaks. The repeated measures ANOVA analysis with H Alignment as the dependent variable revealed statistically significant effects of Focus Type ($F_1(1, 4) = 109.562, p < .0001, F_2(1, 8) = 145.856, p < .0001$) and Language ($F_1(2, 8) = 10.242, p < .01, F_2(2, 16) = 31.663, p < .0001$). The interaction Focus Type * Language was also significant ($F_1(2, 8) = 22.456, p = .001, F_2(1.208, 9.662) = 7.633, p < .0001$). Post-hoc tests with the Bonferroni Adjustment comparing means of H Alignment for BF and NC across Languages showed that Italian is significantly different from Catalan ($p < .05$) but not from Spanish ($p > .05$).

![Figure 6: Box plots of the distance from H to the end of the accented syllable in milliseconds for BF (dark grey boxes) and NCF (light grey boxes) for all speakers of Catalan (graph on the left), Italian (centre) and Spanish (graph on the right).](image)

With respect to the distance from L to the start of the accented syllable, the L point is located for the three languages near the start of the accented syllable for both BF and NCF accents. These results confirm previous observations in the sense that the L point is little or not at all affected by focus type (Estebas-Vilaplana 2009 for Central Catalan; Gili-Fivela 2005, 2006 for Pisa Italian; De-la-Mota 1995, 1997 for Spanish). The repeated measures ANOVA analysis with L Alignment as the dependent variable revealed no statistically significant effects of Focus Type across subjects ($F_1(1, 4) = 1.725, p > .05$) but significant differences in Focus Type across items ($F_2(2, 16) = 20.077, p < .01$) and significant effects of Language ($F_1(2, 8) = 8.689, p < .05, F_2(2, 16) = 35.489, p < .0001$). The interaction Focus Type * Language was not significant ($F_1(2, 8) = .664, p > .05, F_2(2, 16) = 1.777, p > .05$). Post-hoc tests with the Bonferroni Adjustment comparing means of L Alignment for BF and NC across Languages show that Italian is significantly different from Catalan ($p < .0001$) and Spanish ($p < .001$).

### 2.2.2. Pitch range and pitch scaling results

Figure 7 shows the box plots of pitch range of the focus accent in semitones within the initial BF (dark grey boxes) and NCF (light grey boxes) accents for all speakers of
Catalan (graph on the left), Italian (centre) and Spanish (graph on the right). As can be observed, the pitch range of the focus accent seems to be a very stable cue for Italian speakers since all the speakers with the exception of speaker 1 mark NCF through narrower pitch excursions. By contrast, we find no consistent tendency either for Catalan or for Spanish. Thus, Catalan speakers 3, 4, 5 and Spanish speakers 3 and 5 realize the NCF pitch accents with wider pitch excursions but there are speakers such as speaker 2 for Catalan and speakers 2 and 4 for Spanish that present the opposite tendency, that is, they realize the NCF pitch accent with a narrower pitch excursion. Furthermore, some Catalan and Spanish speakers seem not to use pitch range to mark the difference between BF and NCF accents (speakers 1 for Catalan and Spanish). The repeated measures ANOVA tests with Pitch Range of the focus accent as the dependent variable revealed no significant effects of Focus Type ($F(1, 4) = 0.195, p > .05$) or Language ($F(1, 8) = 3.131, p > .05$) across Subjects. However, significant differences were obtained when the analysis was performed across items for Focus Type ($F(2, 8) = 11.166, p < .05$) and Language ($F(2, 16) = 47.952, p < .0001$). The interaction Focus Type * Language was always significant ($F(2, 8) = 5.173, p < .05, F(2, 16) = 32.404, p < .0001$). Post-hoc tests with the Bonferroni Adjustment comparing means of Pitch Range of the focus accent for BF and NC across items showed that Spanish is significantly different from both Catalan ($p < .0001$) and Italian ($p < .0001$). Our interpretation is that this difference between Spanish and the other two languages is due to the variability found in this specific language. Thus, while Italian exhibits a very consistent behavior (NCFs are always narrower in pitch excursion) and Catalan shows a tendency to have broader pitch range for NCF (three out of five speakers follow this pattern), Spanish is a mixture: two out of five speakers produce broader pitch range for NCF but also two out of five speakers exhibit the inverse tendency and one speaker does not mark focus type by means of pitch amplitude at all.

Additionally, we looked at the tonal scaling of L and H tones of the focus accents. Figure 8 shows the box plots of H F0 in Hz within the initial BF (dark grey boxes) and NCF (light grey boxes) accents for all speakers of Catalan (graph on the left), Italian (centre) and Spanish (graph on the right). We observe that the behavior of the H tones is very similar to the results presented with respect to pitch range of the focus accents.
Thus, lower NCF peaks are very systematic in Italian. However, in Catalan and Spanish no specific tendency is found. Some speakers tend to realize the NCF through higher peaks (speakers 4 and 5 for Catalan and speakers 3 and 5 for Spanish) but other speakers such as speaker 2 for Catalan and speakers 2 and 4 for Spanish produce lower NCF peaks. In addition, other Catalan and Spanish speakers do not use H scaling at all to mark the difference between BF and NCF (speakers 1 and 3 for Catalan and speaker 1 for Spanish). Across speakers, the repeated measures ANOVA analyses on the data with H Scaling of the focus accent as the dependent variable revealed no effect of Focus Type on H scaling ($F(1, 4) = 1.058, p > .05$) nor any effect of Language ($F(1, 8) = .149, p > .05$). However, the Focus Type by Language interaction was significant ($F(1, 8) = 5.644, p < .05$). By contrast, significant effects across items were found with respect to Focus Type ($F(2, 8) = 43.495, p < .0001$) and Language ($F(2, 16) = 20.497, p < .0001$). The interaction between Focus Type and Language was again significant ($F(2, 16) = 41.518, p < .0001$). Post-hoc tests with the Bonferroni Adjustment comparing means of H Scaling of the focus accent for BF and NC across items showed that Italian is significantly different from Catalan ($p < .01$) and Spanish ($p < .01$).

![Figure 8: Box plots of H F0 in Hz within the initial BF (dark grey boxes) and NCF (light grey boxes) accents for all speakers of Catalan (graph on the left), Italian (centre) and Spanish (graph on the right).](image)

Regarding the L Scaling, generally speaking no differences depending on focus type were found. However, some speakers (speaker 3 for Catalan and speakers 3 and 4 for Italian) tend to realize NCF accents by means of lower F0 values but at first sight it seems to be a very isolated tendency since the other speakers tend not to realize the focus distinction on the basis L targets. The results of a repeated measures ANOVA with L Scaling of the focus accent as the dependent variable revealed no effects of Focus Type ($F(1, 4) = 3.384, p > .05$) nor effects of Language ($F(1, 8) = 0.571, p > .05$), and no Focus Type by Language interaction was found either ($F(1, 8) = 3.851, p > .05$). When the test was performed across items, significant effects of Focus Type ($F(2, 8) = 17.701, p < .01$), Language ($F(2, 16) = 20.497, p < .0001$) and also a significant interaction Focus Type * Language ($F(2, 16) = 10.820, p = .001$) was found.

All in all, these results show that the difference between NCF and BF is marked through a difference in tonal scaling of the H tone and not a difference in pitch span,
since H tones of NCF accents are raised in Catalan and Spanish and lowered in Italian, while L tones are not affected by focus type.

2.2.3. Duration results

Figure 9 shows the box plots of duration of the initial accented syllable in milliseconds for BF (dark grey boxes) and NCF (light grey boxes) accents for all speakers of Catalan (graph on the left), Italian (centre) and Spanish (graph on the right). The graphs reveal that there is a significant and systematic effect of duration only for Italian, where we see that syllables bearing NCF accents become longer. For Catalan and Spanish, there are also some speakers (speakers 1, 4, and 5 for Catalan and speaker 2 for Spanish) that show a slight tendency to have the same pattern found in Italian. The repeated measures ANOVA tests performed on the data with Duration of the accented syllable as the dependent variable revealed significant effects of Focus Type ($F_{1}(1, 4) = 60.079, p < .01, F_{2}(1, 8) = 90.203, p < .0001$) and Language ($F_{1}(2, 8) = 4664.740, p < .0001, F_{2}(1, 8.001) = 536.782, p < .0001$). The interaction Focus Type * Language was also significant ($F_{1}(2, 8) = 59.421, p < .01, F_{2}(1.001, 8.004) = 87.846, p < .0001$). Post-hoc tests with the Bonferroni Adjustment comparing means of Duration for BF and NCF across Languages showed that Italian is significantly different from both Catalan ($p < .0001$) and Spanish ($p < .0001$).

![Figure 9](image-url)

**Figure 9**: Box plots of duration of the accented syllable in milliseconds for BF (dark grey boxes) and NCF (light grey boxes) accents for all speakers of Catalan (graph on the left), Italian (centre) and Spanish (graph on the right).

2.3. Discussion

This production experiment shows interesting results relating to BF and NCF in initial position in Catalan, Italian and Spanish. In order to compare the two different accents, measures of alignment of the H and L targets, pitch range and F0 of the peak and valley and duration of the syllables bearing the focus accents were taken. The results reveal an important effect of focus type on the alignment of the H target for the three languages under study. These results thus confirm previous observations that both Catalan and Spanish speakers align the NCF peak near the end of the accented syllable while the
peak for BF accents is aligned much later (De-la-Mota 1995, 1997 and Beckman et al. 2002 for Spanish; Estebas-Vilaplana 2009 for Central Catalan). In Italian the difference between NCF and BF accents is also conveyed by differences in alignment with the syllable. NCF accents are always retracted in comparison with BF accents and they are aligned roughly halfway through the syllable. By contrast, we do find later peaks for BF accents, which tend to be aligned near the end of the syllable bearing the focus accent. Interestingly, the alignment realization of the BF in Italian coincides with the realization of the NCF in Catalan and Spanish, which perceptively can cause Catalan/Spanish speakers to perceive Italian speakers as focalizing excessively.

Differences were found with respect to pitch range of the focus accent depending on the focus type across languages, this difference being caused by a different pitch height solely of the H tone. Importantly, for Catalan and Spanish, tonal scaling of the H target seems not to play a key role in marking the different types of focus (as suggested by Face 2001, 2002) but, by contrast, it is a very stable cue in Italian. Our results support previous findings (Gili-Fivela 2005, 2006) in characterizing NCF accents as having lower peaks (see the Introduction for an overview about scaling results found for Pisa Italian). This is particularly interesting because it would imply that Lecce Italian goes against the prediction of the Effort Code. In an attempt to explain these striking results, our proposal would be that the complexity of NCF accents in Italian, that is, the presence of a rising-falling movement, can act as a substitute for a salience marker in this particular language. We should also add that there exist other languages such as Akan (Kügler & Genzel 2011) in which the expression of prominence does not necessarily go hand in hand with an increase in pitch height.

Finally, our results reveal substantial differences in duration only for Italian, in the sense that syllables bearing NCF accents are longer. Also in the case of Catalan and Spanish, there were some speakers that displayed a similar behavior to what was seen in Italian. This allows us to conclude that for Catalan and Spanish duration is an optional cue while for Italian it seems to be used systematically.

3. Experiment 2: Perception experiment, identification task

As we saw in the previous experiment, speakers use different cues to different extents depending on the language to distinguish BF from NCF initial accents. Our next goal was therefore to test whether there is a correspondence between the cues used in

41 Though we found an effect of focus type on H tonal scaling whose magnitude depended on the language, the L point was less affected by it. However, there were three speakers (one Catalan speaker and two Italian speakers) that tended to realize lower L points for NCF. Interestingly, the Catalan speaker in question did not produce different tonal scaling of H at all. Thus, in some way, s/he was using a strategy based on scaling of L target to mark the difference. With respect to Italian speakers, they both mark the difference between the two accents through lower peaks for NCF accents. It seems that their strategy is to use a lower register for NCF.

42 Akan is a tone language that belongs to the Kwa branch of the Niger-Kongo family spoken in Ghana (Kügler & Genzel 2011).
production for each specific language and those used in perception, that is, whether the specific use of the Effort Code by each language in production is also reflected at the perceptual level. Our hypothesis was that irrespective of whether languages follow the Effort Code prediction or not in terms of production, perception will reflect the specific production found for each language. In addition we aim to determine the specific contribution of each prosodic parameter (alignment, duration and scaling) for each language by means of an identification task.

3.1. Methodology

3.1.1. Participants

Twenty native speakers of Majorcan Catalan (5 women, 15 men) aged between 16 and 45 (mean = 28.65), twenty native speakers of Lecce Italian (17 women, 3 men) aged between 19 and 45 (mean = 27.75) and twenty native speakers of Madrid Spanish (14 women, 6 men) aged between 23 and 53 (mean = 31.75) participated in this experiment. They reported having little or no prior training in phonology/phonetics and no history of hearing disability.

3.1.2. Materials

The prompt stimuli used in this perception experiment were created from a set of recordings of the sentence Na Marina vendrà (Cat.) / La Melania verrà (It.) / Marina vendrá (Sp.) (‘Marina/Melania [proper name] will come’) with NCF meaning/prosody, as spoken by three female native speakers of each variety who had participated in the production experiment (but did not participate otherwise in the perception experiment). Catalan and Spanish materials were recorded in a quiet room at the Universitat Pompeu Fabra in Barcelona using a Marantz Professional PMD660 digital recorder and Rode NTG-2 microphone. Italian materials were recorded in an anechoic room at the Centro di Ricerca Interdisciplinare sul Linguaggio in Lecce using a Shure SM-86 condenser microphone and Goldwave 5.25 software installed on a desktop computer equipped with an Edirol UA-5 external sound card. All sentences were digitized at a 44100 Hz sample rate. The criteria for selecting the materials were based on quality and clarity; we also guaranteed that the original alignment, duration and scaling values corresponded to the mean values obtained for that speaker in the production experiment.

Four 7-step continua per language were created in which alignment, duration and scaling were manipulated from a NCF to a BF interpretation. The manipulation was done as follows: in the first continuum only alignment was manipulated while duration and scaling were neutralized; for the second continuum we manipulated alignment and duration but scaling was maintained; in the third continuum only alignment and scaling were at work while duration was kept constant and, finally, in the fourth continuum the three prosodic features were manipulated (see Table III for the step values used for each parameter and language). The manipulations were done again using Praat. Keeping in mind that the manipulation of duration might interact with the manipulation of
alignment (especially in the case of Italian, where the alignment differences take place on the same syllable), the following precautions were taken: (a) the order of manipulation was always duration, alignment and then scaling; (b) once a new step for duration was created, we carefully checked that the peak occupied the same position (in terms of ratio) that it had occupied before we manipulated duration; and (c) given that the manipulations in duration could cause a little irregularity in the F0 rising movement, this irregularity was repaired by means of a linear stylization of the rising-falling movement. Figure 10 illustrates schematically the creation of the continua.

<table>
<thead>
<tr>
<th>Language</th>
<th>alignment</th>
<th>duration</th>
<th>scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>22 ms</td>
<td>10 ms</td>
<td>8 Hz (from 316 to 286 Hz)</td>
</tr>
<tr>
<td>Italian</td>
<td>15 ms</td>
<td>12 ms</td>
<td>12 Hz (from 239 to 309 Hz)</td>
</tr>
<tr>
<td>Spanish</td>
<td>22 ms</td>
<td>10 ms</td>
<td>8 Hz (from 322 to 274 Hz)</td>
</tr>
</tbody>
</table>

Table III: Specific values of the steps of alignment, duration and scaling for each language.

Figure 10: Schematic representation of the creation of the alignment continuum (top left), alignment + duration continuum (top right), alignment + scaling continuum (bottom left) and alignment + duration + scaling continuum (bottom right) for Catalan. White boxes represent the consonants and shaded boxes the vowels. Boxes representing the syllable bearing the nuclear accent are marked with thicker borders.

3.1.3. Procedure

Subjects were seated before a laptop in a quiet room and the stimuli were played back through headphones. Since we were interested in not only utterance type identification choices but also RT measurements, listeners were instructed to maintain their hands near the keyboard and press the keys as fast as they could.

Subjects received written instructions about how they were supposed to respond after each stimulus. In addition, the prompt stimuli appeared in a dialogue so that listeners would understand clearly the difference in meaning between the two stimuli. Thus, if Catalan listeners perceived the BF _Na Marina vendrà_ (‘Marina [proper name] will come’), they were to press the ‘N’ key on the keyboard (for _Normal_, indicating a
‘normal’ declarative), whereas if they perceived the NCF Na MARINA vendrà, they were to press the ‘C’ key (for Correcció = ‘Correction’, since the speaker is correcting the information that s/he has just received from the interlocutor). The experiment lasted approximately 10 minutes and there was a break of 10 seconds between each of the five blocks. The final identification experiment thus consisted of 140 tokens per language (7 steps x 4 continua x 5 blocks).

Stimuli were presented using E-prime version 1.2 (Psychology Software Tools Inc., 2009) software running either on a MSI U100 Wind Notebook laptop equipped with a Realtek HD sound card and Sennheiser HD 202 headphones (Catalan and Spanish) or on a HP desktop computer equipped with a Creative Sound Blaster sound card and Sennheiser HMD 280 headphones (Italian).

3.1.4. Statistical analyses

Although our data did not follow the normal distribution, repeated measures ANOVA tests were used because of the same reasons explained in Chapter 2, section 2.1.4. Two different measures were used as dependent variables for statistical exploration, namely (a) identification responses and (b) RT measurements, with Stimulus (seven levels) and Parameter (four levels) as within-subject factors and Language as the between-subject factor. The results from the repeated measures ANOVA tests include the $F$-ratio, the degrees of freedom from which it was calculated and the significance value.

3.2. Results

3.2.1. Identification results

The four graphs in Figure 11 show the identification rate for the alignment continuum (top left), alignment + duration continuum (top right), alignment + scaling continuum (bottom left) and alignment + duration + scaling (bottom right). The ‘identification rate’ is defined as the number of ‘normal’ responses over the total number of responses. We observe that for all three languages speakers are able to change from a NCF to a BF interpretation merely on the basis of alignment (Figure 11, top left); however, for Catalan and Spanish the identification rate as ‘normal’ for stimulus 1 is higher than expected. That would indicate that the parameter ‘alignment’ is more helpful for Italian speakers than for Catalan or Spanish speakers. The results are a little better when ‘duration’ is added to ‘alignment’ (Figure 11, bottom right). Here the three languages have a very similar behavior. However, when ‘scaling’ is added to ‘alignment’ a different picture emerges. It seems that while Catalan and Spanish speakers prefer scaling, Italian speakers are more sensitive to duration, judging by the high and low identification rate of ‘normal’ responses obtained for the Italian stimuli 1 and 7 respectively (Figure 11, bottom left). For Catalan and Spanish, the results improve substantially when all three prosodic features are at work but for Italian they are a little poorer than when only alignment and duration are exerting an effect. The repeated measures ANOVA analysis revealed statistically significant effects of Stimulus
The effect of Language was also significant \( (F(2, 59) = 4.686, \ p = .013) \).

\( (F(2.469, 145.695) = 328.924, \ p < .0001) \) and Parameter \( F(2.642, 155.875) = 111.766, \ p < .0001 \). The effect of Language was also significant \( (F(2, 59) = 4.686, \ p = .013) \).

3.2.2. Reaction Time results

The four graphs in Figure 12 show the averaged reaction time (RT) responses (in ms) for the alignment continuum (top left), alignment+duration continuum (top right), alignment + scaling continuum (bottom left) and alignment + duration + scaling continuum (bottom right) for Catalan (black bars), Spanish (dark grey bars) and Italian (light grey bars). RTs did not include the duration of the stimulus. We find clear RT peaks (marked with the dashed line) for the three languages only in the case of the ‘alignment + duration’ continuum (Figure 12, top on the right) and the ‘alignment + duration + scaling’ continuum (Figure 12, bottom on the right). Interestingly, Italian speakers are always slower than Catalan and Spanish speakers but especially when the
parameter ‘scaling’ is at work (these results agree with the general identification results).

From the previous results, it is observed that speakers are faster or slower depending on the stimulus (i.e., depending on whether or not it is a stimulus corresponding to the frontier region between the category NCF and BF) but also depending on the parameters that are at work. The three graphs in Figure 13 show averaged RT responses (in ms) for Catalan (top on the left), Italian (top on the left) and Spanish (bottom) with results separated according to the parameter or combination of parameters that is at work: alignment (black line), alignment + duration (dark grey line), alignment + duration + scaling (light grey line) and alignment + scaling (broken line). Assuming that speakers are generally slower when the parameter or combination of parameters is not helpful, we observe that there are important differences across languages. Thus, for Catalan the less helpful parameters are alignment and alignment + duration (Fig. 13, top left), for Italian they are alignment + duration + scaling and alignment + scaling (Fig. 13, top right) and, finally, for Spanish the less effective combinations are alignment + duration

![Figure 17: Averaged RT responses (in ms) for the alignment continuum (top left), alignment + duration continuum (top right), alignment + scaling continuum (bottom left) and alignment + duration + scaling continuum (bottom right) for Catalan (black bars), Spanish (dark grey bars) and Italian (light grey bars).](image)
and alignment + scaling (Fig. 13, bottom). These results agree with the general results of the identification task in the sense that again it seems that the parameter ‘alignment’ alone is more helpful for Italian than for Catalan and Spanish listeners and Catalan and Spanish listeners are more sensitive to ‘scaling’ than Italians. The repeated measures ANOVA analysis performed on the data with RT as the dependent variable revealed statistically significant effects of Stimulus ($F(5.011, 285.650) = 18.208, p < .0001$, Parameter ($F(2.831, 161.372) = 3.430, p = .021$) and Language ($F(2, 57) = 8.917, p < .0001$). Our next objective was therefore to check whether the peaks found for RT measurements in these specific continua broadly coincided with the boundary between the two categories as calculated through logistic regression (see 3.2.3. below).

![Figure 13: Averaged RT responses (in ms) for Catalan (top left), Italian (top left) and Spanish (bottom) with results separated according to the parameter or combination of parameters that is at work: alignment (black line), alignment + duration (dark grey line), alignment + duration + scaling (light grey line) and alignment + scaling (broken line).](image)

### 3.2.3. Logistic regression results

In order to compare mathematically the four identification curves obtained according to the parameters that were manipulated (alignment, alignment + duration, alignment +
scaling, and alignment + duration + scaling) for each language, the set of data points making up each of the four curves was fitted to a logistic function using SPSS (SPSS for Windows) (see Chapter 2, section 2.2.1. for an explanation about the formula and the procedure to calculate the location of the category boundary). Table IV shows the values of ‘x’ for each language and continuum when ‘y’ equals 0.5 (left-hand value in each cell) but also the stimulus in which the RT peak/plateau was located (right-hand value). Shaded boxes indicate when there is coincidence between the boundary located through the logistic regression and the RT peak/plateau obtained for each continuum and language.

<table>
<thead>
<tr>
<th>Language</th>
<th>alignment</th>
<th>alignment+duration</th>
<th>alignment+scaling</th>
<th>alignment+duration+scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>2.67</td>
<td>3.24</td>
<td>3.86</td>
<td>4.20</td>
</tr>
<tr>
<td>Italian</td>
<td>3.86</td>
<td>4.06</td>
<td>4.18</td>
<td>4.02</td>
</tr>
<tr>
<td>Spanish</td>
<td>1.61</td>
<td>3.51</td>
<td>3.47</td>
<td>4.22</td>
</tr>
</tbody>
</table>

Table IV: Boundary values for each continuum and language calculated through a logistic regression (left-hand value in each cell) and RT peak/plateau (right-hand value). Boxes are shaded when the two values are similar.

These results suggest that Italian speakers are more sensitive to changes in ‘alignment + duration’ while Spanish speakers seem to react better to ‘alignment + scaling’ modifications. Interestingly, there is agreement between the three languages with respect to the importance of the effect of the combination of the three prosodic features on subjects’ responses.

Table V shows the slope values for the identification curve obtained for each continuum and language. The slope value is not returned directly by the function, but since the term ‘b1’ is related to the slope, we will use this value as a measure of slope. Lower values reflect steeper curves (shaded boxes in table V), i.e., conditions under which listeners are more successful at distinguishing between the two different meanings. The results confirm what was suggested above, that is, that Catalan and Spanish speakers need the three prosodic parameters to be at work in order to clearly distinguish the two different focus conditions, while Italian speakers have a clear preference for the combination of alignment and duration. Furthermore, Catalan speakers prefer the combination of alignment and scaling, Italian speakers consider alignment alone sufficient and, finally, Spanish speakers opt for the combination of alignment and duration.

<table>
<thead>
<tr>
<th>Language</th>
<th>alignment</th>
<th>alignment+duration</th>
<th>alignment+scaling</th>
<th>alignment+duration+scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>0.445</td>
<td>0.320</td>
<td>0.237</td>
<td>0.208</td>
</tr>
<tr>
<td>Italian</td>
<td>0.4</td>
<td>0.364</td>
<td>0.557</td>
<td>0.460</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.570</td>
<td>0.394</td>
<td>0.439</td>
<td>0.373</td>
</tr>
</tbody>
</table>

Table V: Slope values for each continuum and language calculated by means of a logistic regression. Shaded boxes indicate lower values for each language.
4. Discussion and conclusions

In this study, we explored the specific contribution of tonal alignment, duration and tonal scaling to the expression of NCF as opposed to BF in Catalan, Italian and Spanish. To this end, a production experiment and a perception experiment were carried out. In the production experiment, measures of alignment of the H and L targets, pitch range, F0 and valley of the focal accent, and duration of the syllables bearing them were taken. An important effect of focus type on the alignment of the H target for the three languages was found. Consistent with other studies, NCF peaks occurred earlier than BF peaks in all three languages (see Estebas-Vilaplana 2009 for Central Catalan; D’Imperio 2002 and Gili-Fivela 2002, 2006, 2008 for Neapolitan and Pisa Italian respectively; De-la-Mota 1995, 1997 and Beckman et al. 2002 for Spanish). By contrast, no focus effect on the alignment of the L target was found for any of the three languages (Estebas-Vilaplana 2009 for Central Catalan; Gili-Fivela 2005, 2006 for Pisa Italian; De-la-Mota 1995, 1997 for Spanish). With respect to scaling features, Italian speakers are the only ones to use the tonal scaling of the H target in a systematic fashion. Our results agree with previous work about Pisa Italian (Gili-Fivela 2005, 2006) in which NCF peaks are lower than BF peaks. For Catalan and Spanish, though results were not so clear cut, there was a tendency for some speakers to mark the NFC accents through higher peaks. Finally, NCF accents were systematically longer syllables only for Italian (similar to Gili-Fivela 2005, 2006) but a tendency in the same direction was found for Catalan and Spanish (consistent with Estebas-Vilaplana 2009 for Central Catalan and De-la-Mota 1995, 1997).

The Italian scaling results are interesting in terms of Biological Codes (Gussenhoven 2002, 2004), specifically the Effort Code, since they contradict the general predictions described by this code. Interestingly, another study has shown that the Akan language does not follow the general prediction of the Effort Code either (Kügler and Genzel 2011). This study undertook a production experiment to investigate the tonal and durational means of encoding focus and givenness in Akan. To this end the researchers recorded eleven speakers uttering sentences with target words occurring in sentence-medial and sentence-initial position in different information structure constructions. The sentences were elicited through context questions that put target words either in wide, narrow or contrastive focus in pre-focal or post-focal position in order to study givenness. Prosodic parameters such as duration and F0 were measured on the target words. Contrary to the predictions of the Effort Code (which predicts a correlation between the importance associated with the element and the effort, resulting in higher pitch targets), they found a significantly lower realization of both high and low tones in contrastive focus. Hence, they propose an alternative view on the prosodic expression of focus in the sense that what seems to be important is not the direction of change to mark the focus but rather the deviation from the neutral register. These language-specific strategies to mark NCF are hinted at by Gussenhoven (2002) when he states that “an important aspect of the present conception of intonational meaning is that while the nature of the meanings is related to the way our speech organs produce pitch variation, there is no implication that the physical condition that lies at the basis of these meanings
need be present in order to create the forms”. Turning to our own results for Italian, our interpretation is that the substitute variable in pitch height in the specific case of Lecce Italian is the complexity of the tonal movement, which is to say that it is the rising-falling movement aligned with the accented syllable that is being used as an NCF marker. We speculate that the Italian impossibility of going higher in the NCF accent is due in origin to mechanical reasons, that is, since the F0 has to go up and down within the minimum space of one syllable, speakers have found it easiest to simply avoid pitch rises. Once speakers became aware of this connection between mechanical procedure and meaning they applied it to conveying the NCF pragmatic meaning. We could say, then, that the scaling strategy remains subordinate to the alignment strategy due to articulatory or mechanical factors.

The different sets of results yielded by the identification task (i.e., the identification responses, the RT measurements and the logistic regression results) present converging evidence for the following two assertions: (a) listeners make active perceptual use of the three prosodic parameters found in production in the three languages (thus confirming our hypothesis about a direct correspondence between production and perception regardless of whether in terms of production the languages under study agree with the assumption derived from the Effort Code); and (b) the three languages have their own specific preferences for particular prosodic parameters. Thus, though only alignment was a stable cue in production for Catalan and Spanish speakers while all three parameters were robust for Italian speakers, the three languages used a combination of the three prosodic features whenever they were available in perception. However, our results also point to preferences for each language with regard to perception. First, when possible, Catalan listeners select the three cues together, but if they are not all available, they have a special predilection for the combination of alignment and scaling. Second, Italian speakers do have a preference for the combination of alignment and duration or, in lieu of that, for alignment alone. Finally, Spanish speakers, like Catalan speakers, give priority to the stimuli combining the three prosodic cues: yet when they are not available, alignment and duration can also be sufficient. One might ask why while Catalan and Spanish speakers primarily select the three cues together, Italian speakers are satisfied with only two parameters. Our interpretation is that they select the salience cue par excellence, that is, the rising-falling movement (expressed by means of alignment) and then a possible consequence of this complex movement, an increase in the duration of the syllable bearing the focus accent. It is understandable, then, that these cues are preferred over scaling, which contradicts universal assumptions about language.

In sum, the results reported in this chapter represent further evidence that typologically related languages such as Catalan, Italian and Spanish can use different strategies as prosodic NCF markers. These results contribute to a better comprehension of the principles involved in both production and perception in the expression of NCF across Romance languages. Finally, we would also like to highlight the importance of testing cross-linguistically the perceptual salience of the cues found in production in order to fully understand the interplay that exists between production and perception mechanisms.
Conclusions

The purpose of this dissertation was to investigate whether pitch range variation gives rise to distinctive linguistic categories in Catalan in terms of both production and perception. The study has also addressed the issue of the appropriateness of different perceptual and production experimental methods such as the Discourse Completion Test and the imitation task for production, and the congruity task, the identification task and different types of discrimination tasks for perception.

Chapter 1 reported the application of the Categorical Perception paradigm to a pitch height contrast in the nuclear accent that distinguishes yes-no and wh- questions in Majorcan Catalan. Using two natural tokens produced by a female speaker, two intonational continua were created, from the yes-no question to the exemplar wh-question contour and vice versa. Forty-two Majorcan Catalan listeners participated in a two-part experiment consisting of an identification and a discrimination task. The results from the identification task showed that it is possible to switch the perceived category by manipulating the pitch height of the leading tone. Also, RTs were shorter within categories and longer between categories. Discrimination results revealed that the shift in the identification function corresponded to the peak in the discrimination function. The comparisons between obtained and predicted discrimination results indicated that discrimination can be predicted from identification results on the basis of phonetic categorization. These results confirmed that the difference in pitch height of the leading tone in nuclear pitch accents for yes-no and wh- questions in Majorcan Catalan is discrete and has a phonological character. Interestingly, although the scaling difference of the leading tone of the falling nuclear accent was shown to be the main perceptual cue that Majorcan listeners used to distinguish yes-no questions from wh-questions, identification results also revealed a small effect of the base stimulus. This seemed to suggest that the acoustic information located at the prenuclear region might play a role in the identification task. For that reason, a second perception experiment was carried out in which the role of nuclear and prenuclear accentual/tonal cues in interrogative interpretation was assessed (Chapter 2). In this perception experiment, identification tasks and the gating paradigm were used to investigate whether Catalan listeners use the pitch accent cues in the prenuclear part of the contour in online recognition of yes-no questions and wh-questions. Results indicated that although the most important perceptual cue to identify interrogative sentences lies in the nuclear part of the contour, Catalan listeners are able to exploit the accentual properties of the interrogative particles to a greater extent when the nuclear intonational properties are neutralized. The implications of these results for the AM theory of compositional meaning are discussed at the end of the chapter.

Chapter 3 presented an empirical investigation of the relationship between intonational phonology and pragmatics using empirical data from Eastern Catalan (specifically from the varieties of Central and Balearic Catalan), a language which can signal the distinction between (or function of) information- and confirmation-seeking question by means of prosody. This chapter was concerned with the issue of the
interplay between dialectal variation in intonation and the expression of epistemic modality through the use of an intonation survey based on a Discourse Completion Test strategy. The results showed that the use of specific nuclear pitch accents or boundary tones (depending on the variety of Catalan) is related to the speakers’ epistemic disposition towards a proposition. In addition, with respect to the methodology, it is claimed that the Discourse Completion Test can be profitably used to investigate intonational categories as well as to study how these intonational categories interact with pragmatics. The goal of Chapter 4 was to determine whether the difference in pitch scaling on the leading H tone of the H+L* nuclear pitch accent is used as an epistemic marker by Majorcan Catalan speakers (i.e., whether it is the main cue used by the listeners to distinguish confirmation questions from information-seeking questions). To this end, three behavioral tasks were conducted with twenty Majorcan Catalan subjects, namely a semantic congruity test, a rating test and a classic categorical perception identification/discrimination test. The results confirmed that a difference in pitch scaling on the leading H tone is the main cue used by Majorcan Catalan listeners in distinguishing between a confirmation-seeking and an information-seeking request. Thus, an upstepped leading H tone signals that the speaker has no particular expectation about the answer, while a non-upstepped leading H signals that the speaker is expressing his or her hypothesis about the state of events while seeking confirmation.

Experiments reported in Chapter 5 dealt with the distinction between statements of the obvious and the counterexpectational questions. The aim of this chapter was to investigate the functional role of F0 scaling differences at the boundary tone level in Catalan (in this language the distinction between counterexpectational questions and statements of the obvious lies in the scaling properties of the boundary tone) and to use evidence coming from several experimental techniques to investigate intonational categories. Listeners participated in four experimental tasks: a semantic congruity task with the original stimuli, an identification task with stimuli containing two scaling continua, five types of discrimination task and an imitation task. The results provided converging evidence that Catalan listeners perceive the contrast between the two levels of pitch height in a discrete fashion. Methodologically, this experiment provided evidence that a combination of semantically motivated tasks like congruity and identification tasks (along with RT measures) can be profitably used to investigate intonational categories.

Finally, Chapter 6 investigated the relevance of three prosodic parameters (alignment, duration and scaling) in conveying narrow contrastive focus. In this study, Catalan was compared to two typologically related languages, Italian and Spanish. In particular, we sought to determine how the Effort Code was instantiated in the expression of narrow contrastive focus both in the production and in the perception of speech. To this end a dual production and perception experiment based on an identification task was conducted. Results for the production part of the experiment showed that narrow contrastive focus accents have earlier peaks for all three languages but F0 peaks are systematically lower only in Italian (while in Catalan and Spanish, though scaling results are not so clear cut, there is a tendency for some speakers to mark the narrow contrastive focus accents through higher peaks). Syllables bearing the
contrastive focus accents are also longer in the three languages. Regarding the results for the perception part of the experiment, converging evidence was found not only for an active perceptual use of the three prosodic parameters present in production but also for language-specific preferences for particular prosodic parameters, which are explained by the particular Italian marker of salience, that is, the rising-falling movement aligned with the accented syllable.

On the basis of the results summarized above, we can conclude that pitch range differences in the falling nuclear pitch accent H+L* (Chapters 1-4) and in the final rising boundary tones LH% (Chapter 5) in Catalan are crucial in making linguistic distinctions between information-seeking questions versus wh- questions/confirmation-seeking questions, and statement of the obvious versus counterexpectational questions. The fact that pitch range can be encoded phonologically and that levels other than High are Low are needed to account for linguistic contrasts is not an isolated result but adds to the findings for other languages in the domain of pitch accents (Hirschberg & Ward 1992, Ward & Hirschberg 1985, Ladd 1994, Ladd 1996, Ladd & Morton 1997, Chen 2003, for English; Savino & Grice 2007, 2011 for Bari Italian; Borrás-Comes et al. 2010 for Catalan; Roseano et al. 2011 for Friulian; Vanrell et al. 2011 for Logudorese Sardinian) and boundary tones (Beckman & Ayers-Elam 1997 for American English; Grabe 1998 for British English and German; Post 2000 for French; Beckman et al. 2002 for Spanish; Lee 2003 for Korean; Arvaniti & Baltazani 2005 for Greek; Grice et al. 2005 and Peters 2006 for German; Frota in press-b for European Portuguese).

An important issue concerns how the Catalan prosodic annotation system (the Cat_ToBI system: Prieto et al. 2009, Prieto in press, Aguilar et al. (coords.) 2009-2011) can capture the scaling differences that are shown to be of a categorical nature as well as how this system can be accommodated in a common transcription system for Romance languages within the Tones and Break Indices (ToBI) framework. Our proposal would be to include the upstep (‘¡’) and the downstep (‘!’) features to indicate either a paradigmatic choice (i.e., with consequences for meaning) or a syntagmatic choice (i.e., between sequences of the same phonemic tone as the result of a contextual and thus predictable process). Thus, given the evidence offered by Chapters 1-4 of this dissertation, we would propose to include the category ¡H+L* in the Cat_ToBI inventory. At the same time, these two sets of point to the existence of a three-way contrast in the specific variety of Majorcan Catalan: an upstepped high for information-seeking questions (¡H+L*), a plain high for confirmation-seeking questions (H+L*), and a downstepped high for wh- questions (!H+L*). In similar fashion, our proposal would be to add the !H% boundary tone to the intonational grammar of Catalan.43 We are aware that including the categories ¡H+L* and !H% in the Cat_ToBI phonological analysis of tones renders the concept of upstep ambiguous since the upstep/downstep features will represent either the result of an upstep/downstep rule or a paradigmatic

43 Recall that as we explained in Part III, for Cat_ToBI, we initially proposed the presence of a mid tone in the system (M%). One of the main reasons for replacing the original M% with the present !H% is to be able to offer a consistent analysis of pitch height contrasts both at the pitch accent level and at the boundary tone level. Another important reason is to be able to offer a more general transcription analysis of the same contour in other European languages (e.g., German, Greek and Portuguese) (see Part III of the dissertation).
intonational contrast at the phonological level. Our arguments to maintain these categories in the Cat_ToBI system regardless of the ambiguity generated are as follows: (1) the use of \( \text{H}+\text{L*} \) \( \text{L}\% \) and \( \text{H}+\text{L*} \) \( \text{L}\% \) to distinguish yes-no questions from wh-questions/confirmation-seeking questions/declaratives is productive in other Romance languages (Fernández-Rey 2005, for Galician; Vanrell et al. 2011 for Logudorese Sardinian), as is the existence of the non-high boundary tone \( \text{H}\% \) for not only Romance but also other languages (Beckman & Ayers-Elam 1997 for American English; Grabe 1998 for British English and German; Post 2000 for French; Beckman et al. 2002 for Spanish; Lee 2003 for Korean; Arvaniti & Baltazani 2005 for Greek; Grice et al. 2005 and Peters 2006 for German; Frota in press-b for European Portuguese) and (2) as Frota (p.c.) points out, this situation of ambiguity would be no different from what happens rather frequently in segmental phonology, where a segment (or feature) may be distinctive in the sound system of a language and also generated by rule in specific contexts (as in the case of voiceless plosives in Dutch, which word-finally are contextually motivated). There seem to be solid reasons to reassess the concept of upstep and downstep within the AM framework and also to arrive at a consensus between tonal and intonational languages about the use of labels standing for pitch height levels others than \( \text{L} \) and \( \text{H} \).

Regarding methodology, our suggestion is that a combination of methods can be used to study how intonational categories interact with pragmatics, as clearly observed in the case of the Discourse Completion Test (in production). Similarly, in the perception domain, semantic congruity tasks and identification tasks (involving RT measures and the application of regression as a method of analysis) can be profitably used to investigate intonational categories. With respect to production, we would like to highlight that imitation data must be analyzed with caution, as some speakers were not able to successfully perform this kind of task. We seemed to have two groups of results, one produced by “good imitators” and the other by “poor imitators” (Chapter 5). In perception, puzzling results were obtained in the application of the Categorical Perception paradigm. Thus, a clear discrimination peak was obtained in Chapter 1, as predicted by the identification results. By contrast, no peak was found either in Chapter 4 or in Chapter 5. The possible explanations for this lack of a clear peak in discrimination functions are varied, as we saw in the general Introduction of the dissertation. To summarize, some of these reasons are that (1) intonational units, like vowel phonemes, are encoded over long time intervals (which could favour a continuous rather than a categorical perception) (Remijen & van Heuven 1999), (2) short-term memory effects may be interfering with task decision, (3) the tasks are too focused on the acoustic perceptive properties of the stimuli rather than on the semantic perception of tonal categories (Chen 2003, Kohler 2006, Niebuhr & Kohler 2004, Frota in press-a), and (4) the results would depend on the categories tested, that is, a linguistic distinction such as the one between questions and statements or yes-no and wh-questions would be more easily categorized than a linguistic distinction related to focus since “focus may easily become part of the paralinguistic domain, involving degrees of emphasis rather than categorical changes of meaning/interpretation” (Gili-Fivela 2009: 12-13). We rule out explanations (1) and (3) because they are not able to justify the fact
that a clear discrimination peak is obtained in Chapter 1 but not in Chapters 4 or 5. Explanation (2) should be also excluded because in the cases in which a discrimination peak was not found, the maximum duration of our discrimination stimuli was still below the point at which the same-different discrimination performance ceased to be effective (see Chapter 4). Hence, the explanation that best fits the results obtained in Parts I-III of this dissertation is (4). The results obtained in Chapter 1 are consistent with the results of Remijsen and van Heuven (1999) and Schneider and Linert (2003) in which the Categorical Perception paradigm has been applied successfully to a pitch range contrast since a clear discrimination peak is found at the expected position. Interestingly—and coming back to the explanation (4)—one point in common between these three studies is that they all deal with a tonal scaling contrast related to the prototypical categories of illocutionary force (Quirk 1985) such as declaratives, yes-no questions or wh-questions. By contrast, the perception experiments presented in Chapters 4 and 5 of this dissertation involve categories that refer to beliefs such as confirmation-seeking questions, counterexpectational questions or statements of the obvious. This interpretation is supported by studies in which asymmetries (where discrimination of a change presented in one direction is easier compared to the same change presented in the reverse direction) are found in the perception of vowels, colors, linear direction or geometric figures (Rosch 1975, Tversky & Gati 1978, Polka & Bohn 2003). These studies demonstrate the existence of prototypes that contain the ideal characteristics of a given category (contradicting the assumption of symmetry that characterizes a Euclidean space). In the vowel space, peripheral vowels act as reference points (Polka & Bohn 2003), which causes them to be easily recognizable to speakers, while central vowels are less salient. Our interpretation would be that pragmatic categories such as declaratives, yes-no questions or wh-questions would act as reference points and this would facilitate their discrimination. Hence, since discrimination results can depend on the pragmatic categories involved, we conclude that categorical perception in intonation may be a function of the prototypicality of the original tokens and that therefore this paradigm is not sufficiently robust as a method. Our additional claim is that the tasks employed for the investigation of intonational categories must pay special attention to the semantic relevance of these units and ensure contextual and semantic accessibility. In this regard, congruity tasks (Chapters 4 and 5) are a very promising tool to make perceptual testing more ‘ecologically valid’, as listeners performing this task have been shown to be especially sensitive to the appropriateness of intonation contours in given discourse meanings. Similarly, it is very important to bear in mind the polifunctionality nature of intonational contours, which highlights once again the importance of the context in which a particular specific contour is produced.

Our results underscore the importance of perception experiments to determine what aspects of intonation have linguistic value. That does not mean that production data are not important—on the contrary, they are very important since they bring out the research questions that will guide perceptual studies. Yet production data can be very variable, as we have different intonational strategies. With respect to the phonological analysis of intonation, the relevance of a particular part of the contour (see Chapter 2) or the role of a given parameter or other prosodic features (Chapters 2 and 6) can be
resolved only with the performance of production studies. Yet perception experiments in which tokens can be gated or manipulated freely can be used to verify our claims. Ongoing technological advances now allow us to compare different languages following the same controlled methodology, and perception results are essential if we wish to propose well grounded phonological analyses (Chapter 6).

The findings of Chapter 6 allow us to acknowledge the fact that Catalan does not differ so much intonationally from other typologically related languages such as Italian and Spanish at least from the point view of broad and narrow contrastive focus sentences. It is also important to highlight with respect to the main purpose of this dissertation that the contrast between broad and narrow contrastive focus presented in this chapter is realized phonologically through a difference in peak alignment and consequently the role of scaling is secondary and phonetic in nature.

In conclusion, this dissertation has contributed to enhance our understanding of the principles involved in the expression of different pragmatic meanings through prosodic mechanisms across Romance languages. Our findings are encouraging, specifically those that allow us to see that they are not isolated results but add to the findings for other typologically related languages since they point to the possibility of a consensus-based analysis of Romance intonation in the not so distant future.
Appendix

Tonal inventory of the Cat_ToBI system

Pitch accents

![Graph showing pitch accents H*, L*, H+L*, L+>H*, L+H*, L+¡H*, L*+H]

Boundary tones

Final boundary tones at a break indice of level 4

![Graph showing boundary tones L%, !H%, H%, HH%, LH%, L!H%, HL%, LHL%]

Final boundary tones at a break indice of level 3

![Graph showing boundary tones L-, !H-, H-, HH-, LH-, LHL-]

Initial boundary tone

![Graph showing initial boundary tone %H]
Maria del Mar Vanrell was born in Llucmajor, Mallorca, on 23 July 1980. She studied Catalan Philology at the Universitat de les Illes Balears, finishing her degree in 2003. Since then, she has taught Catalan to both high school students and adult learners (both native and non-native speakers) and also worked for two years as an associate lecturer at the Universitat Autònoma de Barcelona. In September 2006 she defended her Master’s thesis entitled “The phonological role of tonal scaling in Majorcan Catalan interrogatives”, supervised by Dr. Pilar Prieto, and was awarded a DEA (Advanced Studies Diploma) in Cognitive Science in Linguistics by the Universitat Autònoma de Barcelona. She has received various research grants: one from the Universitat Autònoma de Barcelona allowed her to carry out research at the Università degli Studi di Cagliari (February 2009), another from the Spanish Ministerio de Ciencia e Innovación covered a stay at the Centro di Ricerca Interdisciplinare sul Linguaggio (January-April 2010), and two grants from the Universitat Pompeu Fabra funded visits to the Università degli Studi di Sassari (June 2010) and the Centro di Ricerca Interdisciplinare sul Linguaggio (February 2011) respectively. Her research is in the field of Catalan intonational phonology, with interests in the prosody-pragmatics interface, acquisition of intonation and intonational phonology of other Romance languages such as Italian, Sardinian and Spanish. She has participated in various international conferences related to her specializations such as the International Congress of Phonetic Sciences (August 2007), the International Conference on Speech Prosody (May 2006 and 2010), Phonetics and Phonology in Iberia (June 2005, 2007, 2009 and 2011) and Tone and Intonation in Europe (September 2008 and 2010). Currently, she is a doctoral fellow and member of the Group of Prosodic Studies (GrEP) at the Universitat Pompeu Fabra, and she is also a member of the Center for Theoretical Linguistics (CLT) at the Universitat Autònoma de Barcelona.
References


SPSS for Windows, Rel. 15.0.1. 2006. Chicago, USA: SPSS Inc.


