Intonation as an Encoder of Speaker Certainty: Information and Confirmation
Yes-No Questions in Catalan

Maria del Mar Vanrell, Ignasi Mascaró, Francesc Torres-Tamarit and Pilar Prieto

Language and Speech published online 11 July 2012
DOI: 10.1177/0023830912443942

The online version of this article can be found at:
http://las.sagepub.com/content/early/2012/07/11/0023830912443942

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Language and Speech can be found at:

Email Alerts: http://las.sagepub.com/cgi/alerts

Subscriptions: http://las.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> OnlineFirst Version of Record - Jul 11, 2012

What is This?
Intonation as an Encoder of Speaker Certainty: Information and Confirmation Yes-No Questions in Catalan

Maria del Mar Vanrell
Universitat Pompeu Fabra – Universitat Autònoma de Barcelona, Barcelona – Bellaterra, Spain

Ignasi Mascaró
Institut Menorquí d’Estudis, Maó, Spain

Francesc Torres-Tamarit
Universitat Autònoma de Barcelona, Bellaterra, Spain

Pilar Prieto
Institució Catalana de Recerca i Estudis Avançats – Universitat Pompeu Fabra, Barcelona, Spain

Abstract
Recent studies in the field of intonational phonology have shown that information-seeking questions can be distinguished from confirmation-seeking questions by prosodic means in a variety of languages (Armstrong, 2010, for Puerto Rican Spanish; Grice & Savino, 1997, for Bari Italian; Kügler, 2003, for Leipzig German; Mata & Santos, 2010, for European Portuguese; Vanrell, Mascaró, Prieto, & Torres-Tamarit, 2010, for Catalan). However, all these studies have relied on production experiments and little is known about the perceptual relevance of these intonational cues. This paper explores whether Majorcan Catalan listeners distinguish information- and confirmation-seeking questions by means of two distinct nuclear falling pitch accents. Three behavioral tasks were conducted with 20 Majorcan Catalan subjects, namely a semantic congruity test, a rating test, and a classical categorical perception identification/discrimination test. The results show that a difference in pitch scaling on the leading H tone of the H+L* nuclear pitch accent is the main cue used by Majorcan Catalan listeners to distinguish confirmation questions from information-seeking questions. Thus, while a ¡H+L* pitch accent signals an information-seeking question (i.e., the speaker has no expectation about the nature of the answer), the H+L* pitch accent indicates that the speaker is asking about mutually shared information. We argue that these results have
implications in representing the distinctions of tonal height in Catalan. The results also support the claim that phonological contrasts in intonation, together with other linguistic strategies, can signal the speakers’ beliefs about the certainty of the proposition expressed.

**Keywords**
Catalan, confirmation-seeking questions, information-seeking questions, tonal perception, tonal scaling

### 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 lexical/affixal marking of epistemic modality), prosody (different choices of intonational contours) or facial and body gestures (Palmer, 2001). In the prosody literature, there is ample evidence that languages may use prosody to distinguish between information-seeking questions (questions for which the speaker has no particular bias with respect to the answer he/she expects) and confirmation questions (questions for which the speaker has some bias based on beliefs, 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 do-support (*Did Jim leave early?*), canonical confirmation-seeking questions tend to exhibit declarative syntax (*Jim left early?).*1 In Romance languages, the relationship between question type and surface syntactic structure is less clear because these languages have a freer word order and are pro-drop.2 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, 2003).

In some Romance languages, speakers can signal the distinction of information- and confirmation-seeking questions through prosodic means. 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 Italian variety, questions about new information are marked by means of a L+H* accent while questions about given information are signaled 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; in the latter it is located in the syllable preceding the nuclear one. 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 he/she does.

Similarly, Germanic languages other than English 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 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 present 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.

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 paper aims to further explore the extent to which intonational contrasts cue this distinction from a perceptual point of view. Recent results from a production experiment on different varieties of Eastern Catalan (Central and Balearic Catalan) revealed that while Ibizan/Formenteran Catalan speakers use boundary tones to mark the distinction between confirmation- and information-seeking yes-no questions, Central, Majorcan and Minorcan Catalan speakers use a different type of nuclear pitch accent (Vanrell et al., 2010; see Appendix 1 for the survey used to elicit the different types of questions and Appendix 2, Table A2.1 for a summary of the nuclear configurations found for information- and confirmation-seeking questions in each of the following varieties of Eastern Catalan: Central Catalan, Majorcan Catalan, Minorcan Catalan and Ibizan/Formenteran Catalan). 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, Aguilar, Mascaró, Torres-Tamarit, & Vanrell, 2009; Aguilar, De-la-Mota, & Prieto, 2009–2011; Prieto, in press).

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

![Figure 1](https://example.com/figure1.png)

**Figure 1.** Schematic representation of the nuclear accents found in Majorcan confirmation-(left) and information-seeking (right) yes-no questions with the pretonic and tonic 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 borders.
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”) 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 intonational contours thus lies in the tonal height of the leading tone H.4

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 the question, and confirmation-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. Thus, Escandell-Vidal (1996, 1998) characterizes interrogative sentences according to whether the speaker’s 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-Vidal, how much a speaker can choose from a variety of questions depends on his/her own available knowledge and how much knowledge he/she assumes the hearer to have. The diagram in Table 1 (from Escandell-Vidal, 1996, as adapted by Prieto, 2002, and the present study) schematically represents the different degrees of knowledge on the part of the speaker and presumed knowledge possessed by the hearer. The left part of the graph represents the speakers’
knowledge about the content of the question and the right part of the graph the hypothesis that the speaker formulates about the knowledge possessed by the hearer.

Neutral questions (those in which the speaker seeks to obtain information that he/she 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 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: he/she suspects that he/she already knows the answer to his/her question and thus merely seeks confirmation of his/her hypothesis. The resulting output is a confirmation-seeking question, which Escandell-Vidal (1996) 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. One of the aims of this study will be to test this hypothesis and one of the experiments included in this paper (the rating test) is thus intended to test whether Majorcan 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 providing empirical support for Escandell-Vidal’s (1996) model.

The pitch accent contrast investigated in this article involves a difference in pitch height (namely \( \dagger H + L^* \) 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. For example, Ladd and Morton (1997) and Chen (2003) applied the Categorical Perception (CP) paradigm to a contrast between two pitch accents in English, the normal high accent and the emphatic high accent. Evidence was found for a shift between these two categories in identification, even though a related discrimination peak was not found (Ladd & Morton, 1997). On the basis of these results, Chen (2003) argued that the absence of a discrimination peak might be related not to the nonexistence of categorical perception but rather to a hypothetical unsuitability of applying the CP paradigm to a pitch height contrast. In a production and perception experiment, Calhoun (2004) found that themes and rhemes are marked by distinctive pitch accents and that the most reliable cue to the theme and rheme accents is pitch height.

Table 1. Diagram from Escandell-Vidal (1996) as adapted by Prieto (2002) as well as the present study. The left-hand triangle represents the speaker’s knowledge and the right-hand triangle represents the hearer’s presumed knowledge. The center text shows the resulting types of interrogative output.

<table>
<thead>
<tr>
<th>Real knowledge on the part of the speaker</th>
<th>Presumption of the knowledge in possession of the hearer</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Diagram" /></td>
<td><img src="image" alt="Diagram" /></td>
</tr>
</tbody>
</table>

Neutral questions

Oriented questions

Exam-style questions

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. Vanrell (2006) investigated the distinction between wh- questions and yes-no questions in Majorcan Catalan. Results from an identification task showed that it is possible to switch the perceived category simply by manipulating the pitch height of the leading tone (i.e., from a H tone to a super-high tone and vice versa). Both reaction time (RT) and discrimination results revealed a peak that coincided with the shift in the identification function, thus confirming that the difference in pitch accent type is discrete and has a phonological character. Similarly, 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 variety 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 (also known as counter-expectational 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, Vanrell and Prieto (2010) conducted 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. counter-expectational questions). They concluded that counter-expectational questions differ from both narrow focus statements and contrastive focus statements in a discrete way, but the perceived difference between narrow focus statements and contrastive focus statements cannot be explained exclusively by scaling differences. In addition, it seems that this specific pattern (L+¡H*) found in counter-expectational questions for Central Catalan (Borràs-Comes et al., 2010) and Bari Italian (Savino & Grice, 2007, 2011) is also attested in the intonation of other Romance languages that have received lesser attention until recently such as Friulian and Logudorese Sardinian.

In the boundary tone domain, similar contrasts in H scaling have been reported. Thus, in languages like English, German, Greek, Korean, Portuguese, Spanish or Catalan a mid level boundary tone is needed in order to account for linguistic contrasts (Beckman & Ayers-Elam, 1997, for English; Beckman, Díaz-Campos, McGory, & Morgan, 2002, for Spanish; Lee, 2003, for Korean; Arvaniti & Baltazani, 2005, for Greek; Grice, Baumann, & Benzmüller, 2005, for German; Prieto, Torres-Tamarit, & Vanrell, 2008, for Catalan; Frota, in press, for European Portuguese). For example, Prieto et al. (2008) demonstrated by means of four experimental tasks (a semantic congruity task with the original stimuli, an identification task with stimuli containing two scaling continua, five types of discrimination tasks and an imitation task) that Catalan listeners perceive the contrast between the two levels of pitch height LH% and L!H% in a discrete fashion. In addition, from the methodological point of view, they claim that a combination of semantically motivated tasks like congruity and identification tasks (along with reaction time measures) can be profitably used to investigate intonational categories.

One of the overarching goals of this article will be to provide a test case for comparing the outcome of different perceptual tasks for the study of intonational phonology. The Categorical
Perception (CP) paradigm 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, 2009; Dilley, 2010) and differences in pitch height in both tonal languages (Francis & Ciocca, 2003; Francis, Ciocca, & Kei Chit Ng, 2003) and intonational languages, for boundary tones (Remijsen & van Heuven, 1999; Post, 2000; Schneider & Linフト, 2003; Cummins, Doherty, & Dilley, 2006; Falé & Hub Faria, 2006; Prieto et al., 2008; Schneider, Dogil, & Möbius, 2009) as well as for pitch accents (Ladd & Morton, 1997; Chen, 2003; Vanrell, 2006; 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 the CP paradigm to the perception of not only intonational languages (Kohler, 1987; Ladd & Morton, 1997; Remijsen & van Heuven, 1999; Schneider & Linフト, 2003; Cummins et al., 2006; Falé & Hub Faria, 2006) but also tonal languages (Francis & Ciocca, 2003). Studies that report asymmetries in the application of the CP paradigm to pitch height contrasts seem to agree that two different contours are more successfully discriminated when the second one – regardless of which two points in a contour are compared – has a higher pitch.

A general finding in the articles that have applied the CP paradigm to the study of intonational 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; Prieto et al., 2008, for Catalan) or pitch accents (Ladd & Morton, 1997, for English; Gili-Fivela, 2008, 2009, for Italian; Dilley, 2010, for American English). The explanations for this lack of a clear peak in discrimination functions are varied. Some researchers (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 favor a continuous rather than a categorical perception. Other researchers like Gili-Fivela (2009, pp. 370–371) 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”. 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”. In the study by Prieto et al. (2008) in which different discrimination tasks were applied, it is claimed, as in Chen (2003), Kohler (2006), Niebuhr and Kohler (2004) and Frota (in press), 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 of the CP paradigm to investigate 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; Vanrell, 2006; Savino & Grice, 2007, 2011), semantically motivated identification and discrimination tasks (Frota, in press), a combination of semantically motivated tasks like congruity and identification tasks (along with RT measures) (Prieto et al., 2008), or triple answer identification tasks together with semantic congruity tests (Borràs-Comes 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 paper, 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 paper 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 Prieto et al., 2008; Borràs-Comes et al., 2010; Crespo-Sendra, Vanrell, & Prieto, 2010; Rathcke & Harrington, 2010; 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 & Linfelt, 2003, for German; Cummins et al., 2006, for English; Falé & Hub Faria, 2006, for European Portuguese; Prieto et al., 2008, for Catalan; or the CP paradigm applied to pitch accents in Kohler, 1987, for German; Ladd & Morton, 1997, for English; Vanrell, 2006, for Catalan; Gili-Fivela, 2008, 2009, for Italian; Dilley, 2010, for American English).

This paper 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 states 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 intonational 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 a pitch height continuum extending between the two target intonational 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 RT measurements, listeners were asked 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 (Prieto et al., 2008; Borràs-Comes et al., 2010; Crespo-Sendra et al., 2010; Rathcke & Harrington, 2010). The advantage of using such a task is that it allows us to evaluate the degree of perceived appropriateness of target intonational 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 he/she had tangerines. They then heard audio recordings of the ensuing dialogue. The two dialogues were as follows:

(1) Només has de mester mandarines, però no saps si en tenen.
‘You only need tangerines but you do not know whether they have them or not.’
—Bon dia, teniu mandarines?
—‘Good morning. Do you have any tangerines?’
—Eh… sí, ara vénc des Mercapalma i n’he duites.
—‘Er… yeah, I’ve just come from the wholesale market and have brought some.’

(2) Només has de mester mandarines i saps que sempre en tenen.
‘You only need tangerines and you know that they always have them.’
—Bon dia, teniu mandarines?
—‘Good morning. Do you have any tangerines (I suppose so)?’
—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 (upper panel) and for the spliced stimulus inserted into the confirmation-seeking question dialogue (lower 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.’) were produced at a lower volume compared to the dialogues so that it resembled a voice-over.

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 of 40 trials (2 congruous dialogues + 2 incongruous dialogues × 5 repetitions × 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-Vidal, 1996, 1998). Escandell-Vidal’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 paper). In Catalan confirmation-seeking questions we also often find confirmation marks such as polarity adverbs (no, oï), the

Figure 3. Waveform, F0 contour, and Cat_ToBI transcription for the unmanipulated target confirmation-seeking question dialogue (upper panel) and for the spliced stimulus inserted into the confirmation-seeking question dialogue (lower panel).

Note: Owing to space limitations, the figures show only part of the dialogue that the listeners heard.
noun *veritat* or the particle *eh* (Cuenca, 1997; Hernanz & Rigau, 2006; Prieto & Rigau, 2007).\(^5\) 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 his/her interlocutor had tangerines or not. The materials consisted of 4 different audio stimuli recorded by the first author: 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 intonational contour), a tag question (*Teniu mandarines, no?* ‘You have tangerines, don’t you?’), and a broad focus statement (*Teniu mandarines ‘You have tangerines’). The test consisted of 40 trials (4 stimuli × 5 repetitions × 2 blocks). Subjects had to press one of four possible options, namely “1” for ‘no idea’, “2” for ‘maybe yes’, “3” 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 (Remijsen & van Heuven, 1999, for Dutch; Schneider & Linftert, 2003, for German; Cummins et al., 2006, for English; Falé & Hub Faria, 2006, for European Portuguese; Prieto et al., 2008, for Catalan) or between pitch accents (Kohler, 1987, for German; Ladd & Morton, 1997, for English; Vanrell, 2006, for Catalan; Gili-Fivela, 2008, 2009, for Italian; Dilley, 2010, for American English). 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 above-mentioned articles only three really offer unarguable evidence of categorical perception, with a clear discrimination peak in the expected position (Kohler, 1987; Schneider & Linftert, 2003; Vanrell, 2006). Even perception studies performed on tonal languages do not always show clear patterns of discrimination results (see Francis et al., 2003, for a review). The explanation given 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 first author, 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 (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 his/her intonation. They had to press the “N” key for ‘no idea’ (if the speaker had no clear expectation or bias about the answer) or the “M” key for ‘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 × 2 base tokens × 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 × 2 base stimuli × 4 blocks).

Figure 4. Diagram of the eleven stimuli created from a confirmation-seeking question base stimulus (left panel, solid line) and an information-seeking question base stimulus (right panel, solid line) with the pretonic and tonic syllables of the target word mandarines (‘tangerines’). White boxes represent the consonants and shaded boxes the vowels. Boxes representing the nuclear accent-bearing syllable are marked with thicker borders.
3 Results

3.1 Congruity test results

Figure 5 shows the rate of “congruous” responses to both congruous (black bars) and incongruous 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 “congruous” responses to congruous dialogues for both information-seeking and confirmation-seeking meanings. By contrast, the average rate of “congruous” responses in the incongruous 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 (congruous vs. incongruous) for both information-, $T = 1780, p < .001, r = -.57$ and confirmation-seeking, $T = 1053, p < .001, r = -.55$ question meanings were significant. These results indicate that listeners are extremely sensitive to the incongruous use of confirmation- and information-seeking questions.

3.2 Rating test results

Table 2 presents the ratings (in columns) for each stimulus (in rows) in the rating test with results for the 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

![Figure 5](image.png)

**Figure 5.** Rate of “congruous” responses to congruous (black bars) and incongruous dialogues (grey bars), separated by information-seeking question meaning (left) and confirmation-seeking question meaning (right).
that listeners were asked to rate on a 4-point scale the likelihood that the speaker would obtain an affirmative answer to his/her 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 admitted 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 database. 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, $\chi^2(6) = 984.482, p < .05$.

All in all, our results indicate that the available knowledge/presumed knowledge scale proposed by Escandell-Vidal (1996, 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 tokens.

![Identification rate for confirmation and information questions](image)

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

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Answer category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No idea</td>
</tr>
<tr>
<td>Information-seeking question</td>
<td>370</td>
</tr>
<tr>
<td>Confirmation-seeking question</td>
<td>48</td>
</tr>
<tr>
<td>Tag question</td>
<td>1</td>
</tr>
<tr>
<td>Statement</td>
<td>7</td>
</tr>
</tbody>
</table>
find statistical differences between the response rates for stimuli 5 and 6, information-seeking base stimulus, the results are nearly what we would expect. Observe that we and also for stimuli 4 and 5, and 6 are not significant, Table 3, since the differences between response rates for stimuli 5 and 6 for both continua, there should be significant differences between the response rates so on) for each continuum. According to the CP paradigm, since the boundary is located between stimuli 5 and 6 for both continua (see Figure 6).

In order to claim that a contrast is discrete, 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 favor of the categorical nature of the contrast tested. With the aim of determining 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 through the Curve Estimation procedure in SPSS. From the SPSS fitted logistic curves, we obtained the $b1$ and the $b0$ parameters. While the term $b1$ is related to the slope (with lower values reflecting steeper curves), $b0$ is a constant. The boundary is computed from these two terms using the following formulato solve for $x$ when $y = 0.5$, that is, when speakers identify the stimulus in question with the equal numbers of “confirmation-seeking” and “information-seeking” responses:

$$\text{boundary} = -\frac{\ln b0}{\ln b1}$$

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 3 shows the results of ten Wilcoxon matched pairs signed rank tests 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, 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 3, since the differences between response rates for stimuli 5 and 6 are not significant, $T = 30, p > .05, r = -.12$. 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, $T = 19, p < .05, r = -.25$ and also for stimuli 4 and 5, $T = 46, p < .05, r = -.22$ and 3 and 4, $T = 0, p < .01, r = -.26$.

<table>
<thead>
<tr>
<th></th>
<th>Confirmation-based stimuli</th>
<th>Information-based stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2</td>
<td>$T = 2, p &gt; .05, r = -.04$</td>
<td>$T = 2, p &gt; .05, r = -.04$</td>
</tr>
<tr>
<td>2–3</td>
<td>$T = 0, p &gt; .05, r = -.18$</td>
<td>$T = 6, p &gt; .05, r = -.03$</td>
</tr>
<tr>
<td>3–4</td>
<td>$T = 28, p &gt; .05, r = -.10$</td>
<td>$T = 0, p &lt; .01, r = -.26$</td>
</tr>
<tr>
<td>4–5</td>
<td>$T = 81, p &lt; .05, r = -.21$</td>
<td>$T = 46, p &lt; .05, r = -.22$</td>
</tr>
<tr>
<td>5–6</td>
<td>$T = 30, p &gt; .05, r = -.12$</td>
<td>$T = 19, p &lt; .05, r = -.25$</td>
</tr>
<tr>
<td>6–7</td>
<td>$T = 50, p &gt; .05, r = -.15$</td>
<td>$T = 15, p &gt; .05, r = -.20$</td>
</tr>
<tr>
<td>7–8</td>
<td>$T = 19, p &gt; .05, r = -.25$</td>
<td>$T = 28, p &gt; .05, r = -.10$</td>
</tr>
<tr>
<td>8–9</td>
<td>$T = 32.5, p &gt; .05, r = -.04$</td>
<td>$T = 20, p &gt; .05, r = -.02$</td>
</tr>
<tr>
<td>9–10</td>
<td>$T = 30, p &gt; .05, r = -.02$</td>
<td>$T = 13.5, p &gt; .05, r = -.05$</td>
</tr>
<tr>
<td>10–11</td>
<td>$T = 11, p &gt; .05, r = -.14$</td>
<td>$T = 10.5, p &gt; .05, r = 0$</td>
</tr>
</tbody>
</table>
When looking at standard error values of the mean for every stimulus (Table 4), 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 continuum display the highest standard error values. These results are by no means unexpected since they show that listeners agreed in their responses when listening to stimuli 1 and 11 because they represent the canonical categories, while this agreement decreases as the crossover point between the categories approaches.

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

<table>
<thead>
<tr>
<th>Table 4. Standard error of the mean of the identification rate for information- and confirmation-based stimuli.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Figure 7. Averaged reaction time (RT) responses (in ms) for all subjects.
the RTs for the confirmation-seeking question-based continuum and the grey line shows the RTs for the information-seeking question-based continuum.\(^8\)

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., 2006), we separated our database into two groups according to whether the subjects had or did not have musical training.\(^9\) The results were consistent with Vanrell (2006) in the sense that musicians were faster than 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- and confirmation-seeking question continua, \(F(1, 688) = 16.188, p < .001\), and confirmation-seeking, \(F(1, 701) = 24.878, p < .001\), question continua, but no interaction between musical training, stimulus, and RT for either of the two continua. This means that even though non-musicians are always slower than musicians, both musicians and non-musicians display the same behavior concerning 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 associated with the preaccentual syllable. Complementary evidence in favor 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 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 preaccentual syllable and before 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 panel) and information-seeking (right panel) question-based continua. \(d'\) 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 occurring 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; Prieto et al., 2008, for Catalan). As we noted above in the Introduction, the explanations for this lack of a clear peak in discrimination functions are varied (Remijsen & van Heuven, 1999; Chen, 2003; Niebuhr & Kohler, 2004; Kohler, 2006; Prieto et al., 2008; Gili-Fivela, 2009; Frota, in press). In short, 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 (see Introduction) of applying the CP paradigm to intonational contrasts.

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) Iconic memory is known to last approximately 250 milliseconds (Sáiz, Baqués, de la Fuente, Pousada, & Vera, 2008, p. 21). Echoic memory is thought to last a little longer (Sáiz et al., 2008, p. 23), but such factors as the time separating the sound to be discriminated and 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, p. 175). However, there seems to be agreement that the discrimination ability/competence disappears when the interstimulus interval is above three seconds (Crowder, 1981). The maximum duration of our discrimination stimuli was 2276 ms, still below the point at which the same–different discrimination performance stops being effective.

(b) The second argument arises from previous studies (Vanrell, 2006) in which a clear peak was obtained for discrimination results. In these earlier studies the mean duration of discrimination stimuli was 2265 ms with an ISI (interstimulus interval) of 500 ms. In the

Figure 8. Discrimination results presented as d' for each stimulus pair in each order of presentation for the confirmation- (left panel) and information-seeking (right panel) question-based continua.
present study, the mean total duration of the discrimination stimuli was 2276 ms but with an ISI of 300 ms. It seems that the difference in the duration 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 Vanrell (2006) and the one used in the present study. However, the results of Prieto et al. (2008), 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.

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 one (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- and information-seeking, $T = 10829, p < .001, r = -.107$, and information-seeking, $T = 1782, p < .001, r = -.11$, question-based continua. 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 & Hub Faria, 2006; Vanrell, 2006; Prieto et al., 2008), in the sense that it appears that subjects have trouble discriminating between stimuli when the direction of change in fundamental frequency is downwards. In previous research, these asymmetries have been related to the F0 declination or 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, and evidence for compensation of this effect has been shown for languages such as American English (Pierrehumbert, 1979), Dutch (Gussenhoven & Rietveld, 1988), and Cantonese (Wong, 1999). According to Francis and 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 a 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.

4 Discussion and conclusions

The patterns of results obtained from the three perceptual experiments 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 incongruous use of confirmation- and information-seeking questions. As we saw in Figure 5, incongruous dialogues display a low average rate of “congruous 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-Vidal (1996) 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 intonational patterns: *Teniu mandarines?* ‘Do you have any tangerines?’ (with the upstepped H tones 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 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 his/her 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 the speaker’s 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 his/her question and seeks confirmation of his/her 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-seeking 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-seeking 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 Curve Estimation procedure. 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 stimulus 11 since they represent the endpoints of a continuum which we may interpret as canonical categories tested in this study, while this agreement decreases as the proximity to the crossover point increases. In Table 4 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 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 fitted logistic curve. 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), Prieto et al. (2008) 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. According to Prieto et al. (2008), who applied different types of discrimination tasks to determine the viability of categorical perception for studying intonation contrasts, the problem with discrimination has to do with the fact that “the comparison of pairs of stimuli is too focused on the acoustic perceptive properties of the stimuli rather than on their phonological patterning”. 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 is at an ISI of about three seconds (Crowder, 1981), and the maximum duration of our discrimination stimuli was below this limit; and (b) previous results have shown that although an increased ISI can improve discrimination performance, it does not lead to the emergence of a discrimination peak (Prieto et al., 2008).

Hence, the patterns of results obtained from the congruency test as well as the identification test (together with the RT measurements) prove that pitch scaling on the H level has a phonological character in distinguishing information- from confirmation-seeking questions in Majorcan Catalan. We argue that the absence of discrimination peaks should not be blamed on a lack of discreteness in the pitch contrast under examination but rather on a hypothetical unsuitability of the CP paradigm as applied to intonational contrasts (Chen, 2003; Prieto et al., 2008; Savino & Grice, 2011). Our data thus add further support to the important role of perception tasks (and especially perception tasks more related to linguistic contexts, like congruity tasks) in determining the nature of the relationship between intonation and meaning.

Our results have implications for the tonal representation of pitch accent contrasts in Catalan couched within the Cat_TolBI transcription system (Prieto et al., 2009; Aguilar et al., 2009–2011; Prieto, in press). 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 additional tonal pitch levels to account for relevant linguistic contrasts. This has been shown to be the case, for example, for mid boundary tones (Beckman & Ayers-Elam, 1997, for English; Beckman et al., 2002, for Spanish; Lee, 2003, for Korean; Arvaniti & Baltazani, 2005, for Greek; Grice et al., 2005, for German; Prieto et al., 2008, for Catalan; Frota, in press, for European Portuguese) 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 Vanrell (2006) for the presence of an upstepped high leading tone in the Majorcan variety. These two sets of results 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.

In sum, the results reported in this article confirm that a difference in pitch scaling on the leading H tone of the H+L* nuclear pitch accent 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. 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 is by means of morphological marking. The following example is taken from Suena (De Haan, 2001), 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, p. 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 his/her proposition. All in all, our findings represent further evidence that intonation constitutes a robust linguistic strategy to mark the pragmatic category of epistemicity 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, Benus, Hirschberg, Sneed German and Ward (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 intonational 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.

Acknowledgements

Preliminary versions of this article were presented at the PaPI 2009 in Las Palmas, the Workshop on Prosody and Meaning in Barcelona, and at Speech Prosody 2010 in Chicago. We are grateful to Laura Dilley, Mariapaola D’Imperio, Gorka Elordieta, Martine Grice, José Ignacio Hualde, Greg Kochanski, Aditi Lahiri, Olivier Niebuhr and Elina Savino for their helpful comments and suggestions. The various Majorcans who participated unselfishly as listeners in our experiments deserve special mention. We would like to thank Meghan Armstrong, Joan Borràs, Verónica Crespo, Jaume Monreal, Marianna Nadeu and Gemma Rigau for their comments and advice on the first written version of this article and for technical support. Finally, we also thank Isabel Falé and Frank Kügler (reviewers) and Dr. John Coleman (as Associate Editor) for their constructive comments from which our paper has greatly benefited. This research has been funded by projects FFI2009-07648/FILO and CONSOLIDER-INGENIO 2010 CSD2007-00012 (awarded by the Ministerio de Ciencia e Innovación) and by project 2009 SGR 701 (awarded by the Generalitat de Catalunya).

Notes

1. However, according to Gunlogson (2001), who analyzes the meaning and use of rising and falling declaratives compared to rising interrogatives, the picture is not always so clear. It is true that declarative questions are more appropriate in situations where the questioner is assumed to be partial and informed, 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.

2. In Catalan, while information-seeking yes-no questions are characterized by having a dislocated subject in preverbal or postverbal position (when it is present), confirmation-seeking yes-no questions like echo-questions tend to present the subject in preverbal position and without dislocation (Rigau, 2002; Prieto & Rigau, 2007; Vanrell et al., 2010) (e.g., (S),VO or VO,(S) for information-seeking yes-no questions: Tens una barberia, tu? (have a barber’s shop, you ‘Do you have a barber’s shop?’) and (S)VO for confirmation-seeking yes-no questions: Bé, tu has passat per davant ca na Janera? (well, you have gone past Janera’s house ‘Well, have you gone past Janera’s house?’)).
3. The upstep diacritic (¡) here refers to an extra high tone for information-seeking questions with respect to the plain H tone that is found for confirmation-seeking questions.

4. 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 intonational patterns. Thus, while confirmation-seeking questions show a steady high tonal plateau which extends from the beginning of the sentence 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.

5. 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; Vanrell et al., 2010). According to Vanrell et al. (2010), 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.

6. This formula as well as the procedure to calculate the location of the category boundary are taken from Keating (2004).

7. In the case of multiple comparisons, the post-hoc Bonferroni correction was applied by adjusting the $p$-values.

8. The outliers with values greater than $1.5 \times$ interquartile range (IQR) (1834 ms and 1717 ms for the confirmation-seeking and information-seeking continua respectively) were eliminated following the standard assumption that extremely long RTs reflect a lapse in the subject’s attention.

9. Nine 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 eight had not.

10. 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.

References


Appendix 1: Materials of the intonation survey

Situations for eliciting information-seeking yes-no questions:

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

2. Vas amb un amic i et телефонen. 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.
   Target question: ‘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.
   Target question: ‘Do you have a pen?’

3. Entres a una botiga on no havies estat mai i demanes al botiguer a veure si tenen mandarines.
   Target question: ‘Teniu mandarines?’
   You have just entered a shop you have never been in before. Ask the shopkeeper whether he has any tangerines.
   Target question: ‘Do you have any tangerines?’

Situations for eliciting 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.
   Target question: ‘Dus els caramel·los?’
   A friend of yours has bought cough drops for you because you had requested it. Ask your friend whether he’s bringing the cough drops.
   Target question: ‘Are you bringing the cough drops (I suppose 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.
   Target question: ‘M’has tret el llibre?’
   A friend of yours has had to go to the library and you have asked him/her to take a book out of the library for you. When he/she arrives, ask him/her whether he/she has got the book.
   Target question: ‘Have you brought me the book (I suppose so)?’
Table A2.1. Schematic representation of the nuclear configuration of information- (upper panel) and confirmation-seeking (lower panel) yes-no questions for Central Catalan, Majorcan Catalan, Minorcan Catalan and Ibizan/Formenteran Catalan.

Appendix 2: Nuclear configurations for information- and confirmation-seeking yes-no questions in Catalan dialects

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 agafés la bossa amb les mandarines. Just abans de partir, li demanes si ha agafat la bossa amb les mandarines.

Target question: ‘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, and he/she is supposed to bring a bag of tangerines. Before leaving, ask him/her whether he/she has brought the bag of tangerines.

Target question: ‘Have you brought the bag of tangerines (I suppose so)?’