

Audiovisual Correlates of Interrogativity: A Comparative Analysis of Catalan and Dutch

Joan Borràs-Comes · Constantijn Kaland · Pilar Prieto ·
Marc Swerts

© Springer Science+Business Media New York 2013

Abstract Languages employ different strategies to mark an utterance as a polar (yes–no) question, including syntax, intonation and gestures. This study analyzes the production and perception of information-seeking questions and broad focus statements in Dutch and Catalan. These languages use intonation for marking questionhood, but Dutch also exploits syntactic variation for this purpose. A production task revealed the expected language-specific auditory differences, but also showed that gaze and eyebrow-raising are used in this distinction. A follow-up perception experiment revealed that perceivers relied greatly on auditory information in determining whether an utterance is a question or a statement, but accuracy was further enhanced when visual information was added. Finally, the study demonstrates that the concentration of several response-mobilizing cues in a sentence is positively correlated with the perceivers' ratings of these utterances as interrogatives.

Keywords Information-seeking questions · Speech perception · Gaze · Dutch · Catalan

J. Borràs-Comes (✉) · P. Prieto
Departament de Traducció i Ciències del Llenguatge, Campus de la Comunicació, Poble Nou,
Universitat Pompeu Fabra, C/Roc Boronat, 138, 08018 Barcelona, Spain
e-mail: joan.borras@upf.edu

P. Prieto
e-mail: pilar.prieto@upf.edu

C. Kaland · M. Swerts
Tilburg University, Tilburg, The Netherlands
e-mail: c.c.l.kaland@uvt.nl

M. Swerts
e-mail: m.g.j.swerts@uvt.nl

P. Prieto
Institució Catalana de Recerca i Estudis Avançats (ICREA), Universitat Pompeu Fabra, Barcelona,
Spain

Introduction

The world's languages have different grammatical means to mark an utterance as a polar (yes–no) question (e.g., *Are you hungry? Does the shop open on Saturday?*), including the use of different lexical items or morphemes, changes in the syntactic structure, or prosodic and gestural marking. While declaratives are considered to be the unmarked sentence type, primarily used to convey information with no special illocutionary force markers (Levinson 2010, p. 2742), questions are primarily used to seek information.

Cross-linguistically, morphosyntactic features have been shown to constitute a common way to identify polar questions. Among these strategies, we find the presence of question particles (*est-ce que* in French, *ne* in Latin), a specific interrogative word order (as in most Germanic languages), or a combination of such strategies. As Dryer (2008) states, most languages using these morphosyntactic strategies also employ a distinct intonation pattern, though some do not (e.g., Imbabura Quechua, spoken in Ecuador).

Prosody is also a very common resource to signal polar questions across languages. It can be used to assign question status to a declarative-formatted sentence (Stivers and Rossano 2010, for Italian), even in those languages that use morphosyntactic strategies (e.g., declarative questions; Englert 2010, for Dutch). Bolinger (1989) argued that the presence of high pitch in questions may even be considered a linguistic universal (i.e., the fact that the average pitch in questions tends to be higher than the average pitch in non-questions). Moreover, Cruttenden (1981) suggested that the universal dichotomy between falling and rising tunes may be associated with the abstract notions of *closed* (for falls) vs. *open* status (for rises).¹ However, some recent descriptive studies like Englert's (2010) have pointed out that this prosodic feature is not exclusively tied to interrogativity, but is also a common device for signaling continuation in statements or, at the level of discourse, both turn-giving and turn-keeping. In contrast with Bolinger's claim mentioned above, Rialland's (2007) analysis of 78 Central African languages showed that question prosodies without any high-pitched correlates are widespread and include falling intonations or low tones, lengthening, breathy termination, and open vowels.

Though the analysis of morphosyntactic and prosodic markers of polar questions has received considerable attention in the linguistics literature, less is known about the relevance of nonverbal cues. Nonetheless, various studies in the last three decades have taken into account the potential importance of eye gaze and certain facial and manual gestures. In fact, backchannel signals like facial expressions, head movements, and gaze, seem to be critically linked to listeners' attention, perception, and comprehension (Peters et al. 2005; Lysander and Horton 2012). Argyle and Cook (1976) argued that gaze serves three main purposes during face-to-face communication: seeking information, receiving signals that accompany the speech, and controlling the flow of the conversation. Cosnier's (1991) study of French spontaneous speech revealed that the gestural traits that characterize information-seeking questions are those that normally accompany informative verbal expressions, namely, eye gaze to the interlocutor, head elevation, an optional suspended hand gesture facing the interlocutor, and a variety of facial expressions which are then frozen while the speaker awaits a response. Cosnier in fact argued that gaze is as important as intonation and pauses for question marking and turn-taking. As Vilhjalmsson pointed out (1997, pp. 21–22), since the primary function of the eyes is to gather sensory input, the most

¹ As Snow and Balog (2002, p. 1031) suggest, *closed* would encompass “specific meanings like final, complete, or the intention of statements”, whereas *open* may have “the grammatical meaning of nonfinal in utterance-internal contexts and the pragmatic meaning of a yes/no question in utterance-final contexts”.

obvious function of gaze is perhaps information-seeking, since the speaker will at least look at the listener when feedback is expected.

Eyeblink movements have also been associated with questioning, though the results appear to be somewhat inconclusive. For instance, Srinivasan and Massaro (2003) made use of “talking heads” (synthetic representations of a human face) in which they varied specific auditory and visual characteristics to investigate whether these could differentiate statements from declarative questions in English. They found that both eyebrow raising and head tilting could increase the perceivers’ detection of a question, though participants tended to rely more on auditory cues. However, Flecha-García’s (2010) analysis of English spontaneous speech materials found that speakers do not use eyebrow raises in questions more often than in other types of utterances. Yet, incidentally, she also suggested that eyebrow raises may add a questioning meaning to any utterance—somewhat like adding a tag question (e.g., *isn’t it?*) at the end—even if the utterance does not express a question or request, whether verbally or prosodically (Flecha-García 2010, p. 553).

In line with this cross-linguistic variation, recent studies prefer to look at question marking as the set of features that contribute to response mobilization (Stivers and Rossano 2010, p. 29; Haan 2002). Stivers and Rossano (2010) found for both English and Italian that no single feature is present in all cases and thus conclude that no such feature appears to be intrinsic to the action of requesting information (p. 8). They state that if an assessment is accompanied by several response-mobilizing features, this increases the response relevance of the action (p. 28). From a cross-linguistic point of view, even though speakers of different languages rely on different question marking correlates, the same response-mobilizing resources—gaze, lexico-morphosyntax, prosody, as well as contextual epistemic asymmetry—seem to be available across languages, ethnicities, and cultures (Stivers and Rossano 2010, p. 29). In general, Rossano (2010) observed a trade-off relationship between mobilizing cues, and observed that Italian speakers tend to look more often at recipients when those utterances do not have a clear intonational marking. In addition, he found that speakers looked more at recipients during polar questions and alternative questions than during *wh*-questions, which can also be linked to the fact that the latter show a greater use of interrogative verbal cues than the other two types of questions (i.e., *wh*-words). Moreover, Levinson (2010; see also Stivers 2010) has shown that pragmatic inference is a cross-linguistic cue for interrogativity detection and can even represent the main question marker in a language (Levinson 2010, for Yélf Dnye). If the speaker makes a statement about anything of which the recipient has greater knowledge, this routinely attracts the recipient’s response (Labov and Fanshel 1977; Pomerantz 1980).

To our knowledge, no controlled experimental studies have been undertaken to explore what role verbal and nonverbal cues play in the production and perception of questions and whether there exists a trade-off relationship between different mobilizing correlates. To date, the majority of descriptions have been based on the analysis of controlled or natural corpora, and some perception studies have assessed the audiovisual identification of ‘biased’ questions (i.e., those conveying, for instance, counter-expectation, incredulity, or surprise), most of them by means of synthetic materials (Borràs-Comes and Prieto 2011; Borràs-Comes et al. 2011; Crespo-Sendra et al. 2013; House 2002; Srinivasan and Massaro 2003). There are still a number of open questions that have not received a complete answer, such as: Can we differentiate an information-seeking polar question from a broad focus statement by means of visual information alone? How does visual information contribute to question identification when added to auditory information? Does the simultaneous use of several questioning cues increase the perceiver’s identification of an utterance as a

question? Do nonverbal cues have a major role in those languages in which intonation and syntactic cues do not play a defining role?

The present study aims to compare interrogativity-marking strategies in Dutch and Catalan, two European languages that have been argued to rely on different resources for this distinction. One the one hand, Dutch polar interrogatives are characterized by subject/verb inversion, without making use of an auxiliary verb as is the case in English (Dut. *Heeft hij een snor?*, lit. ‘Has he a moustache?’, ‘Does he have a moustache?’; Englert 2010, p. 2668). By contrast, grammatical subjects are generally not produced in Catalan interrogatives (Cat. *Té bigoti?*, lit. ‘Has moustache?’, ‘Does he have a moustache?’). In terms of prosody, speakers of Dutch appear to draw on the overall set of phonological devices of their language for question-marking, though certain configurations are more likely to occur in questions than in statements, as happens with rising intonational contours (Haan 2002, p. 214). By contrast, in order to convey information-seeking polar questions most Catalan dialects have been claimed to use a specific intonational contour which consists of a low pitch accent followed by a rising boundary tone (Prieto et al. 2013).² Drawing on Rossano’s (2010) hypothesis, we expect that Catalan speakers will use a greater number of prosodic and gestural cues than Dutch speakers, since the latter use an additional syntactic strategy to mark questions (see also Geluykens 1988).

The present study has two related goals. First, we aim to describe the combination of syntactic, prosodic, and gestural cues used by Dutch and Catalan speakers for the marking of broad focus statements and information-seeking polar questions. In order to collect a series of broad focus statements and information-seeking polar questions for our perception experiment, we conducted a production task using two variants of the *Guess Who* game. As Ahmad et al. (2011) point out, the dynamic nature of games make them a good tool for investigating human communication in different experimental setups, especially if the outcome of a game is controllable in a systematic manner.

The second goal of the study is to test whether and how listeners of the two languages differentiate questions from statements, as well as to evaluate the relative importance of the different cues used in production and perception. A set of the stimuli obtained by means of the production task was therefore used as stimulus materials for a test in which participants had to guess whether an utterance was a statement or a question. Participants were presented with materials in three perceptual conditions: one in which only the auditory information was available, another one in which only the visual information was available, and a third one which presented simultaneously the full auditory and visual information of the actual recordings. This identification test allowed us to assess the relevance of the various features and their potential interaction effects.

Method for Recording Stimuli

Participants

Eighteen Dutch speakers (11 male, 7 female) and sixteen Central Catalan speakers (1 male, 15 female) participated in the production task. Participants played the game in pairs, taking turns in adopting the different available roles. Participants only played the game with other

² Catalan can also mark interrogativity through the expletive particle *que* (cf. *est-ce que* in French and *é que* in Portuguese), which is especially found for Central, Balearic, and North-western Catalan in confirmation-seeking questions (i.e., not in information-seeking questions; Prieto and Rigau 2007).

native speakers of their own language. They showed clear signs of engagement in the task and some of them even wanted to play more rounds than those indicated by the experimenter. All subjects were undergraduates at either the Tilburg University, The Netherlands, or the Universitat Pompeu Fabra in Barcelona, Spain. All participants played both variants of the game.

Procedure

In order to elicit statements and questions in a natural manner, we used two digital variants of the “Guess Who” board game as created by Suleman Shahid, from Tilburg University, and some colleagues (see Ahmad et al. 2011). In this game, participants were presented with a board containing 24 colored drawings of human faces. These faces differed regarding various parameters, such as gender or color of their skin, hair, and eyes. Some faces were bald, some had beards or moustaches, and some were wearing hats, glasses, or earrings. In the traditional version of “Guess Who”, the purpose of the game is to try to guess the opponent’s mystery person before he or she guesses yours.³

Given our need to elicit either statements or questions, we asked participants to play one of two different variations of the game. In the *question-elicitation* variation, participant A had to ask Participant B questions to try to determine the mystery person on B’s face card. Players took turns asking questions about the physical features of their respective “mystery persons” in an effort to eliminate the wrong candidates. Subjects were instructed to be truthful when answering questions about the mystery person. The winner was the player who guessed his/her mystery person first. In the *statement-elicitation* variation of the game, participants took turns making statements about their mystery person, while the other player listened and eliminated all characters that did not exhibit a particular feature. Again, it was the player who guessed the identity of their “mystery person” first that won.⁴ Note that both participants within a pair took turns in the course of both variations of the game and therefore both provided examples of questions and statements.

Setup

Participants sat in the same room, facing each other across a table and in front of two laptop computers arranged so that they could not see each other’s screen. Two camcorders were placed in such a way that they could record the upper part of each participant’s body. Before the start of each experiment, the camera was raised or lowered according to the participant’s height. Once the participants were seated, the experimenter gave spoken instructions, telling the participants about the game and procedure to be followed for each

³ This experimental setup provides a clear advantage over real situations. As Richardson et al. (2009) state, a question typically implies turn transition, and several studies have shown that gaze is related with turn-giving (Argyle and Cook 1976; Duncan and Fiske 1977; Kendon 1967, 1990). Moreover, Englert (2010) has shown for Dutch that questioners rely overwhelmingly on speaker gaze (90 %) for next speaker selection. Thus, in order to describe the nonverbal patterns that characterize questions one has to focus on those cases in which gaze plays no addressee-selection role, and this is controlled in our study since participants are engaged in dyadic situations.

⁴ In order to increase the number of interactions and communication flow between participants—and to avoid continuation rises in the intonation patterns they produced—we added an additional rule to the game: at the end of each turn, players had to try to guess the mystery person’s name. This additional set of questions was not subjected to analysis. As one of the reviewers suggests, this set of guesses might be used in a follow-up study on the auditory and visual aspects of (un)certainly in questions.

variation. Each game lasted approximately 20 min, the time it took for both variants of the game to be played and won (4–6 times each).

Method for General Perception Experiment

Participants

In the perception experiment, twenty Dutch listeners and twenty Catalan listeners (10 males, 30 females) rated the selection of 70 stimuli in their own language as being statements or questions. Though the stimuli were excerpts from recordings made during the first experiment, none of the participants in the first experiment took part in the second one. The stimuli were presented in three different conditions in a within-subjects design: Auditory-Only (AO), Visual-Only (VO), and AudioVisual (AV). In order to control for a possible learning effect, the AV condition was always the last to be presented to the participants, and the order of the two unimodal conditions was counterbalanced among subjects.

Materials

From the production recordings, 35 statements and 35 questions related to gender (e.g., *It is a man* vs. *Is it a man?*) were selected for each language in order to be included in the subsequent rating task. The selection was limited to this type of utterances in order to constrain their semantics, and the gender issue was selected because these utterances were found to be more frequent overall in the recordings. The final set of materials came from 17 Dutch speakers and 15 Central Catalan speakers. Whenever possible, we guaranteed that each speaker provided a similar number of statements and questions. Since some of the participants of the production task did not provide gender-related utterances, they had to be excluded when selecting the materials for the perception task.

Perception Procedure

The target 70 stimuli were presented to each group of same-language participants in a randomized order. Stimuli were presented to subjects using a desktop computer equipped with headphones. Subjects were instructed to pay attention to the stimuli and decide which interpretation was more likely for each stimulus by pressing the corresponding computer key for *statement* and *question*: ‘A’/‘P’ (*afirmació, pregunta*) for Catalan, and ‘S’/‘V’ (*stelling, vraag*) for Dutch. No feedback was given on the “correctness” of their responses. Participants could take as much time as they wanted to make a decision, but could not return to an earlier stimulus once they had made a decision on it.

The experiment was set up by means of E-Prime version 2.0 (Psychology Software Tools Inc., 2009), which allowed us to record responses automatically. A new stimulus was presented only after a response to the previous one had been given. The experiment was set up in a quiet research room at either Tilburg University and or the Universitat Pompeu Fabra, respectively. It lasted approximately 17 min. The total number of responses obtained was 8,400 (70 stimuli × 20 subjects × 3 conditions × 2 languages).

General Perception Results

Figure 1 shows the mean correct identification rates of the perception experiment broken down by Language (Dutch, Catalan), Condition (AO, VO, AV), and Intention (statement, question). The results in the graph show that participants in both languages were able to identify the two categories above chance level in all three presentation conditions. However, materials that included auditory information (i.e., AO and AV) were consistently more reliable conveyors of question identification.

A Generalized Linear Mixed Model (GLMM) analysis was run with the correct identification of the utterance category as the dependent variable, with language, condition, intention, and all the possible interactions as fixed factors and Subject and Item (Speaker) as random factors. Main effects for Language ($F(1, 155) = 6.58, p = .01, r = .044$) and Condition ($F(2, 8388) = 417.40, p < .001$) were found, but not for Intention ($F(1, 152) = 0.46, p = .50, r = .037$). Two interactions were also found to be significant: Language \times Condition ($F(2, 8388) = 21.50, p < .001$) and Condition \times Intention ($F(2, 8388) = 33.48, p < .001$).

Bonferroni post hoc tests were extracted in order to know the direction of the significant main effects and interactions. They show an effect of condition such that $AV > AO > VO$ (all paired comparisons, $p < .001$), that is, even though the difference between AO and AV conditions might not be clear in the graph, there is a significant difference between the two in the number of correct identifications. This pattern of results suggests that the visual channel provides additional cues that are useful to solve momentary ambiguities of the utterances coming from the presence of falling intonation or syntactic unspecification. Concerning the interaction Language \times Condition, Dutch participants were more accurate than Catalan participants only when auditory information was available: AO ($p = .002$)

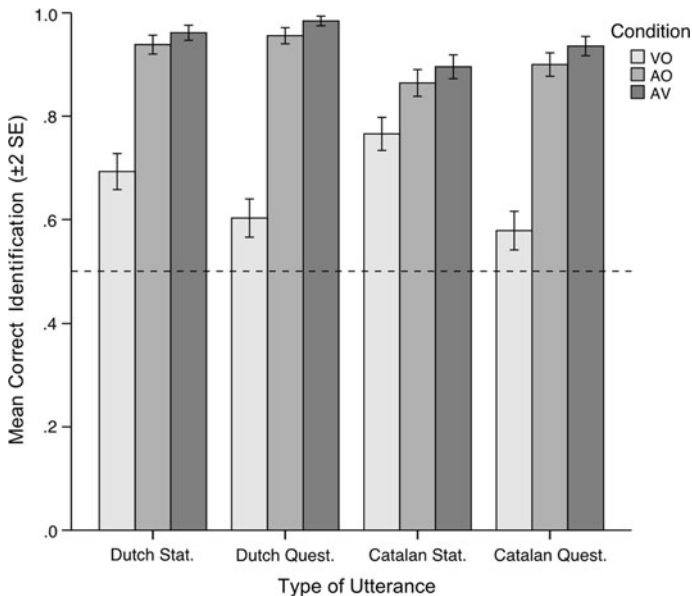


Fig. 1 Mean correct identification rate as a function of Language, Condition, and Type of utterance. VO video only, AO audio only, and AV audio + video

and AV ($p < .001$), and not in VO ($p = .53$). Concerning the interaction Condition \times Intention, statements were more accurately identified than questions in VO condition ($p = .001$), but questions were more accurately identified than statements in AV condition ($p = .006$), and no differences were found in the AO condition ($p = .19$).

In sum, the perception results shown here reveal that participants could identify questions and statements above chance level in all conditions. Specifically, participants' responses were better when auditory information was present, but a beneficial effect of visual cues was also shown when they were added to the auditory ones. In addition, Dutch participants' perception of auditory materials was found to be better than that of Catalan participants, with less of a difference between language groups when they were presented with VO materials, which allows us to hypothesize that language differences were most pronounced when the auditory components of the experiment materials were involved. Importantly, our results show that when only visual information is present, statements are better identified than questions (though the converse pattern holds in the AV condition). These questions are further investigated in the next section, where we analyze the perceived materials in terms of their specific auditory and visual features.

In the perception experiment, twenty Dutch listeners and twenty Catalan listeners (10 males, 30 females) rated the selection of 70 stimuli in their own language as being statements or questions. Though the stimuli were excerpts from recordings made during the first experiment, none of the participants in the first experiment took part in the second one. The stimuli were presented in three different conditions in a within-subjects design: Auditory-Only (AO), Visual-Only (VO), and AudioVisual (AV). In order to control for a possible learning effect, the AV condition was always the last to be presented to the participants, and the order of the two unimodal conditions was counterbalanced among subjects.

Method for Auditory and Visual Cues Experiment

Labeling Procedure

With the aim of assessing the discrimination power of prosodic and gestural cues, the first two authors of the article—native speakers of Catalan and Dutch, respectively, but with some knowledge of each other's language—independently coded the selected audiovisual materials (a total of 70 utterances) in terms of the following cues (based on Cosnier 1991):

- order of the sentence constituents (SV, VS, V)
- intonation (falling or rising boundary tone; i.e., L % versus H %)⁵
- gaze to interlocutor (presence, absence)
- eyebrow raising movement (presence, absence)

A featural analysis of the selected materials was performed in order to ensure that participants were faced with different kinds of syntactic forms, intonation patterns and the

⁵ Please note that the rising intonation category includes both Cat_ToBI continuation rises of the type M % and H %, as well as the HH % type, more typical of questions. These three types of rising boundary tones can be easily mistaken for one another when transcribing short and isolated sentences.

presence of gaze to interlocutor and eyebrow raising, which would allow us to assess whether those parameters played any role in subject judgments.⁶

The inter-transcriber agreement between the two labelers' coding was quantified by means of the Cohen's kappa coefficient (Cohen 1960), which gave an overall coefficient of .84, which means that the strength of the agreement was very good (Landis and Koch 1977). The coefficient was .86 for Dutch and .82 for Catalan. Concerning the different cues, it was .72 for the boundary contour, .91 for gaze, and .70 for eyebrow raising. Disagreements were resolved by consensus.

Audiovisual Properties of the Selected Materials

Table 1 presents the results of the presence of cues found in the database. Regarding Syntax, the subject was omitted in all Catalan sentences, which only displayed the verb and predicate (Cat. *És una dona*, lit. 'Is a woman', 'It is a woman'). In turn, all Dutch statements presented a SV order (Dut. *'t is een vrouw*, 'It is a woman') and all Dutch questions presented a VS order (Dut. *is 't een vrouw?*, 'Is it a woman?'). In terms of Intonation, the same pattern of results was attained for statements in the two languages, showing a great number of falling tones (mostly L* L % and some H* L %). Rising tones (L* H %) were found more often in Dutch questions than in Dutch statements (though Dutch questions exhibited a larger number of falling tones than rising tones; see Geluykens 1988). In turn, Catalan showed a clear majority of questions produced with a rising tone (L* H %, as in the case of Dutch).

Concerning the two visual cues labeled (presence of gaze, eyebrow-raising), the two languages showed similar distributions of their uses in statements and questions. Crucially, the presence of gaze and eyebrow-raising were found to be more present in questions. Overall, Catalan speakers also seem to use more non-syntactic cues than Dutch speakers.⁷

Auditory and Visual Cues Perception Results

Unimodal Perception of Auditory and Visual Features

The lack of syntactic marking in Catalan (i.e., zero degrees of freedom) makes it impossible for us to compute the interactions in which Language and Syntax are implied (see Table 2 for means). In order to know the effect of both syntax and intonation within Dutch, a language-specific GLMM analysis of the AO task was performed, with identification as the dependent variable, Syntax, Contour, and their interaction as fixed effects, and Subject and Speaker as random factors. All factors were significant: Syntax ($F(1, 107) = 331.192$, $p < .001$, $r = .024$), Contour ($F(1, 32) = 16.989$, $p < .001$, $r = .093$), and their

⁶ As one of our reviewers suggests, it would be interesting to correlate the presence of listener gaze (and also mutual gaze between speaker and listener) with the presence of speaker eyebrow raisings and other gestures. Probably these phenomena could be better explored with a different experimental set up (e.g., without screens in between).

⁷ The parameter "eye gaze" refers to speaker's gaze to interlocutor. One may argue that the use of a paradigm like the 'Guess Who' boardgame would elicit less mutual gaze than normal face-to-face conversations, because of the presence of the board. In fact, there is nothing in the set-up that forces participants to look at the board or at the other participant in neither the declarative nor the question conditions. They can freely choose when to look at the other. Interestingly, despite these conditions, we found differences in the presence of gaze to interlocutor from one intention to the other.

Table 1 Number of utterances containing the four labeled cues, for each sentence meaning, in Dutch and Catalan

	Dutch		Catalan	
	Statements	Questions	Statements	Questions
VS order	0	35	0	0
Rising intonation	4	13	4	33
Eye gaze	9	21	12	24
Eyebrow raising	5	9	6	16

Total number of utterances per cell = 35

Table 2 Proportion of sentences identified as questions by presence or absence of each audiovisual cue

	Dutch		Catalan	
	Absence	Presence	Absence	Presence
VS order	.14 (.34)	.85 (.36)	.48 (.50)	–
Rising intonation	.42 (.49)	.70 (.46)	.13 (.33)	.80 (.40)
Eye gaze	.33 (.47)	.71 (.45)	.35 (.48)	.60 (.49)
Eyebrow raising	.46 (.50)	.63 (.48)	.38 (.49)	.70 (.48)

interaction ($F(1, 59) = 6.087, p = .02$). Bonferroni paired contrasts crucially showed that the interaction Syntax \times Contour was related to the fact that a rising contour caused more question identifications when applied to a SV structure ($p < .001$), but not when applied to a VS structure ($p = .18$).

As for the perception of intonation differences, a GLMM analysis was conducted on the results of the AO task, with identification as the dependent variable, Language, Contour, and their interaction as fixed effects, and Subject and Speaker as random factors. There were main effects for Language ($F(1, 26) = 11.67, p = .002, r = .116$), Contour ($F(1, 2796) = 601.41, p < .001, r = .136$), and their interaction ($F(1, 2796) = 79.25, p < .001$). The significant interaction is due to the fact that Catalan listeners rated more falling contours as statements than Dutch listeners ($p < .001$), but this difference does not hold for rising contours ($p = .33$), suggesting that rising contours are perceived equally often as question-conveyors for both language groups. This is consistent with the patterns found in production.

Another GLMM analysis was conducted on the results of the VO task, with identification as the dependent variable, and Subject and Speaker as random factors. The fixed effects were Language, Gaze, Eyebrow, and all their possible interactions. Main effects were found for Gaze ($F(1, 2080) = 283.04, p < .001, r = .068$), Eyebrow ($F(1, 2792) = 21.04, p = .004, r = .059$) and Language ($F(1, 37) = 8.88, p = .005, r = .026$). Two interactions were also found to be significant: Gaze \times Eyebrow ($F(1, 2792) = 16.09, p < .001$), and the triple interaction Gaze \times Eyebrow \times Language ($F(1, 2792) = 4.43, p = .04$). The main effects of Gaze and Eyebrow are related to the patterns observed in production, i.e., that the presence of these cues increased 'question' responses. The main effect of Language suggests that Dutch participants gave overall more 'question' responses than Catalan participants. As for the Gaze \times Eyebrow interaction, Eyebrow had a significant effect on 'question' identification when in the presence of gaze ($p < .001$), but

not in its absence ($p = .68$). Regarding the triple interaction, a Language difference is found, such that Dutch participants provided more 'question' responses than Catalan participants when Gaze ($p = .003$) and Eyebrow ($p = .006$) appeared alone in the perceived materials, but not when these features co-appeared ($p = .33$) or were both absent ($p = .06$).

In sum, the presence of specific syntactic marking in questions had an effect on the subjects' decisions, as well as the presence of rising intonational contours. Moreover, an interaction between syntactic form and intonation was found such that rising contours triggered more question identifications when applied to utterances with an unmarked syntactic structure (i.e., SV sentences, not showing an explicit question structure). As for nonverbal parameters, we found that subjects rated significantly as questions those utterances in which the speaker looked to his/her interlocutor and produced eyebrow raising, with an interaction between both factors such that eyebrow raising had a significant effect on 'question' identification when in the presence of gaze.

Auditory and Visual Features Combined

A main question related to cue interaction is whether the presence of different cues related to questioning can significantly increase the detection of questions. To this end we created a new variable that contained the sum of the different cues to questioning found in both languages (i.e., VS syntax, rising intonation contour, presence of gaze, and eyebrow raising). A Pearson correlation (2-tailed) was conducted between the number of interrogative cues and the identification responses. The test identified a positive correlation of .736 in the case of Dutch and a correlation of .709 in the case of Catalan (in both cases, $p < .001$), which means that there is a high correlation for both languages between the incremental presence of cues to questioning and the participants' 'question' responses.

General Discussion and Conclusions

The first goal of the present study was to describe the syntactic, prosodic, and gestural strategies used by Dutch and Catalan speakers for marking information-seeking polar questions and broad focus statements. These two languages have been argued to mark interrogativity in two different ways. Whereas Dutch polar questions are characterized by the use of a syntactic verb fronting strategy and optional intonational marks (e.g., *Hij heeft een baard* vs. *Heeft hij een baard?*, lit. 'He has a beard' vs. 'Has he a beard?'), the main strategy to mark polar questions in Catalan is through the use of specific intonational patterns on the same syntactic structure (e.g., *Té barba* vs. *Té barba?*, lit. 'Has beard' vs. 'Has beard?'). On the one hand, the fact that Dutch indeed has a systematic syntactic strategy as described in the literature was confirmed by the results of our production task. As for prosody, both languages showed a great number of rising tones in questions, though Catalan (because of the lack of any lexico-morphosyntactic distinction) showed a stronger effect of intonation for interrogativity marking. Concerning gestures, both languages showed similar distributions of the use of gaze and eyebrow raisings, which were mainly found in questions.

The second and main goal of this investigation was to test whether listeners of the two languages could differentiate questions from statements in the different presentation conditions (AO, VO, AV), as well as to evaluate the relevance of the different cues used in perception. The results of our perception experiment with 20 Dutch listeners and 20

Catalan listeners confirmed that participants can identify questions and statements above chance level in all conditions. Importantly, perceivers showed a great reliance on auditory information, but also showed that (a) visual-only utterances were classified above chance; and (b) better accuracy in responses was exhibited when visual information was added to auditory information. This result confirms the importance of nonverbal cues in speakers' identification of pragmatic intentions but also suggests a higher importance of auditory cues in the perception of interrogativity.

Focusing on the auditory-only perception, Dutch participants were found to be more accurate than Catalan participants, which can be linked to the fact that Dutch uses an unambiguous syntactic strategy. With respect to the perceptual importance of syntax and intonation in Dutch, an analysis of the Dutch listeners' perception of AO information revealed that both factors were significant. Moreover, there was an interaction between the two, in the sense that rising contours led to more 'question' identification responses only when applied to an unmarked (SV) syntactic structure. This demonstrates that when both are available syntax has greater importance relative to intonation.

When focusing on the visual-only perception, gaze played an especially strong role in 'question' identification responses in both languages. This is in line with Rossano's (2010) production results for Italian, which showed that the occurrence of speaker gaze towards the recipient in dyadic interactions increases the likelihood of obtaining a response. As for eyebrow raising, a secondary role was found such that it powered 'question' responses only when in the presence of gaze.

More crucially, in the AV presentation, we found a positive correlation between the concentration of mobilizing cues in a sentence and its rating as an interrogative utterance, for both languages. This result is especially relevant for the theory of response relevance put forward by Stivers and Rossano (2010) and is consistent with previous findings from analyses of end-of-turn cues (see Duncan and Fiske 1977). While suggesting four main response-mobilizing features—namely interrogative lexico-morphosyntax, interrogative prosody, recipient-directed speaker gaze, and recipient-tilted epistemic asymmetry—they argue that the inclusion of multiple response-mobilizing features leads to higher response relevance than the inclusion of fewer or no features. In their own words, “a request (or an offer or information request) is high in response relevance, but a request designed ‘directly’ (e.g., with interrogative morphosyntax and/or prosody) would be still higher. Similarly, an assessment (or a noticing or announcement) would be low in response relevance. However, if it were designed with multiple response-mobilizing features, this would increase the response relevance of the action” (Stivers and Rossano 2010, pp. 27–28). In our data, a higher concentration of lexico-morphosyntactic, prosodic, and gestural cues increases the chances that utterances will be perceived as questions.

To our knowledge, the present study provides the first results of a controlled investigation on the perception of information-seeking polar questions compared with broad focus statements in two typologically different languages. First, we have found that auditory information has a greater effect in question identification (auditory cues > visual cues). As for visual cues, we have empirically shown that both auditory and visual cues play a role in this distinction in both Catalan and Dutch. Specifically, the addition of nonverbal cues to auditory cues enhances the perception of information-seeking questions. Also, a visual-only presentation of the materials led to successful interrogativity detection. In terms of its perceptual relevance, a greater effect was found for gaze compared to eyebrow raising. This pattern of results suggests, at least when taking into account Dutch and Catalan data, a cue value scale for interrogativity marking such that Syntax > Intonation > Gaze > Eyebrow.

In conclusion, this study shows how several verbal and nonverbal cues are systematically used in the production of interrogativity and how they crucially interact in its perception.

Acknowledgments We thank Suleman Shahid for his help in setting up the recording sessions, and Igor Jauk for his help in labeling part of the Catalan corpus. We also thank the audiences of The 13th Conference on Laboratory Phonology and The 5th European Conference on Tone and Intonation. This research has been funded by two research grants awarded by the Spanish Ministerio de Ciencia e Innovación, namely HUM2006-13295-C02-01/FILO, FFI2009-07648/FILO and Consolider-Ingenio 2010 Program CSD2007-00012, by a grant awarded by the Generalitat de Catalunya to the *Grup d'Estudis de Prosòdia* (2009SGR-701), and by a “Study abroad scholarship for research outside of Catalunya” 2010 BE1 00207, awarded by the Generalitat de Catalunya.

References

- Ahmad, M. I., Tariq, H., Saeed, M., Shahid, S., & Krahmer, E. (2011). Guess who? An interactive and entertaining game-like platform for investigating human emotions. In Jacko, J. A. (Ed.), *Human-computer interaction. Towards mobile and intelligent interaction environments* (Vol. 3, pp. 543–551). Lecture Notes in Computer Science, 6763. Berlin: Springer.
- Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.
- Bolinger, D. L. (1989). *Intonation and its uses: Melody in grammar and discourse*. London: Edward Arnold.
- Borràs-Comes, J., & Prieto, P. (2011). ‘Seeing tunes’. The role of visual gestures in tune interpretation. *Journal of Laboratory Phonology*, 2(2), 355–380.
- Borràs-Comes, J., Puglesi, C., & Prieto, P. (2011). Audiovisual competition in the perception of counter-expectational questions. In Salvi, G., Beskow, J., Engwall, O., & Al Moubayed, S. (Eds.), *Proceedings of the 11th international conference on auditory-visual speech processing* (pp. 43–46). Volterra, Italy: KTH Royal Institute of Technology.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cosnier, J. (1991). Les gestes de la question. In C. Kerbrat-Orecchioni (Ed.), *la question* (pp. 163–171). Lyon: Presses Universitaires de Lyon.
- Crespo-Sendra, V., Kaland, C., Swerts, M., & Prieto, P. (2013). Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics*, 47, 1–13.
- Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics*, 17(1), 77–91.
- Dryer, M. S. (2008). Polar questions. In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Eds.), *The world atlas of language structures online* (Chapter 116). Munich: Max Planck Digital Library. Retrieved from: <http://wals.info/chapter/116>.
- Duncan, S., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory*. New York: Wiley.
- Englert, C. (2010). Questions and responses in Dutch conversations. *Journal of Pragmatics*, 42(10), 2666–2684.
- Flecha-García, M. L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech Communication*, 52, 542–554.
- Geluykens, R. (1988). On the myth of rising intonation in polar questions. *Journal of Pragmatics*, 12, 467–485.
- Haan, J. (2002). *Speaking of questions. An exploration of Dutch question intonation*. LOT Dissertation Series, 52. Utrecht: LOT.
- House, D. (2002). Perception of question intonation and facial gestures. *Fonetik*, 44(1), 41–44.
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, 26, 22–63.
- Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. New York: Cambridge University Press.
- Labov, W., & Fanshel, D. (1977). *Therapeutic discourse*. New York: Academic Press.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Levinson, S. C. (2010). Questions and responses in Yéfi Dnye, the Papuan language of Rossel Island. *Journal of Pragmatics*, 42(10), 2741–2755.

- Lysander, K., & Horton, W. S. (2012). Conversational grounding in younger and older adults: The effect of partner visibility and referent abstractness in task-oriented dialogue. *Discourse Processes, 49*(1), 29–60.
- Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., & Poggi, I. (2005). A model of attention and interest using gaze behavior. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Intelligent virtual agents* (Vol. 3661, pp. 229–240). Lecture Notes in Computer Science. London: Springer.
- Pomerantz, A. M. (1980). Telling my side: “Limited access” as a “fishing” device. *Sociological Inquiry, 50*, 186–198.
- Prieto, P., Borràs-Comes, J., Crespo-Sendra, V., Roseano, P., Sichel-Bazin, R., & Vanrell, M. M. (2013). Intonational phonology of Catalan and its dialectal varieties. In S. Frota & P. Prieto (Eds.), *Intonational variation in Romance*. Oxford: Oxford University Press.
- Prieto, P., & Rigau, G. (2007). The syntax-prosody interface: Catalan interrogative sentences headed by *que*. *Journal of Portuguese Linguistics, 6*(2), 29–59.
- Psychology Software Tools Inc. (2009). E-Prime (version 2.0). Computer Program.
- Rialland, A. (2007). Question prosody: An African perspective. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes* (Vol. 2, pp. 35–62). Berlin: Mouton.
- Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science, 33*(8), 1468–1482.
- Rossano, F. (2010). Questioning and responding in Italian. *Journal of Pragmatics, 42*(10), 2756–2771.
- Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua, 112*, 1025–1058.
- Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice. Distinguishing statements from echoic questions in English. *Language and Speech, 46*, 1–22.
- Stivers, T. (2010). An overview of the question-response system in American English conversation. *Journal of Pragmatics, 42*(10), 2772–2781.
- Stivers, T., & Rossano, F. (2010). Mobilizing response. *Research on Language and Social Interaction, 43*(1), 1–31.
- Vilhjalmsson, H. H. (1997). *Autonomous communicative behaviors in avatars*. Unpublished Master’s thesis. Massachusetts Institute of Technology, Cambridge, MA.