Encoding interrogativity intonationally in a second language

Aoju Chen and Ineke Mennen

Max Planck Institute for Psycholinguistics, Nijmegen
Queen Margaret University, Edinburgh
`aoju.chen@mpi.nl; imennen@qmuc.ac.uk`

Abstract

This study investigated how untutored learners encode interrogativity intonationaly in a second language. Questions produced in free conversation were selected from longitudinal data of four untutored Italian learners of English. The questions were mostly wh-questions (WQs) and declarative questions (DQs). We examined the use of three cross-linguistically attested question cues: final rise, high peak and late peak. It was found that across learners the final rise occurred more frequently in DQs than in WQs. This is in line with the Functional Hypothesis whereby less syntactically-marked questions are more intonationally marked. However, the use of peak height and alignment is less consistent. The peak of the nuclear pitch accent was not necessarily higher and later in DQs than in WQs. The difference in learners' exploitation of these cues can be explained by the relative importance of a question cue in the target language.

1. Introduction

Questioning is a major communicative function [9]. It serves the speaker's cognitive needs and is an effective means of starting and keeping up social interactions. Interrogativity can be expressed by means of syntax (i.e. the inversion of the subject and the finite verb), lexical cues (e.g. wh-words, question particles), and intonation. As inversion is found to occur also in non-questions in languages like Dutch and German and an utterance can function as a question without inversion (e.g. declarative questions), Haan [9] argues that inversion is not an essential formal characteristic of interrogativity. In the same vein, the presence of a question word is not an exclusive property of questions either, as question words are also frequently used in exclamatory utterances. In contrast, the distinctive power of question intonation characterised with high pitch is widely attested across languages [3]. The exact manifestation of high question pitch however differs from language to language. It can be realised both locally (e.g. a final rise, a higher pitch peak) and globally (e.g. a higher overall pitch).

The language-specificity in the realisation of high question pitch indicates that second language (L2) learners have to learn how to encode interrogativity intonationally in their L2. While some canonical patterns may be covered in some second language classrooms, the use of intonation in questions is by and large poorly taught or not specifically treated. The situation is even worse for untutored learners, who struggle to learn a language with little explicit information available on the structural properties of the language. Given the significance of questioning in everyday life, every L2 learner will be in a situation where he/she will need to ask (and answer) questions. But how exactly do learners ask questions intonationally in L2? This paper reports on a study investigating how untutored Italian learners of English use intonation to encode questions in spontaneous conversation in English at a given longitudinal point.

1.1. The Frequency Code and language-specificity in question intonation

The commonality of high pitch in questions, according to [7], results from grammaticalisation of the Frequency Code [10]. This code is based on the fact that large vocalisers tend to produce low-pitched sounds but small vocalisers high-pitched sounds. It associates low pitch with 'big' meanings (e.g. 'authoritative', 'confident') and high pitch with 'small' meanings (e.g. 'submissive', 'friendly'). When asking questions, the speaker is dependent on the other person's good will for information and therefore needs to present him/herself as being friendly and submissive. As a result, questions are often spoken with high pitch.

However, different languages may choose to realise high pitch in questions differently. Take for example the London variety of British English, our L2 learners' target language (TL), and Italian, our L2 learners' source language (SL). In London English, high question pitch manifests itself mainly in the rise at the end of the utterance, transcribed as H% [6]. The rise at the end is preceded by various nuclear accent types, including L*H, L*, H* and H*L. In standard Italian (i.e. the Tuscan variety), as in London English, declarative questions are spoken with H%, preceded by L* H- and H+L* L- [1]. The same is true for most Northern and Central varieties of Italian [5]. In the Southern varieties of Italian such as Neapolitan Italian, declarative questions are distinguished from statements in the alignment of the pitch peak in the rising nuclear accent LH. The peak was reached 40 ms later in declarative questions than in statements [5]. In line with the claim that a late peak can serve as a substitute for a high peak [7] because a higher peak takes a longer time to reach, it may be argued that high question pitch is realised as a late peak in Neapolitan Italian.

In this study, we are primarily interested in the 'how' part of expressing interrogativity in L2. We assume that learner question intonation emerges as a result of at least three factors, the Frequency Code, transfer of L1 realisation of interrogativity, and L2 realisation of interrogativity. The first question we addressed is how learners use the following cross-linguistically attested cues to encode interrogativity: (1) a final rise; (2) a high nuclear peak; (3) a late nuclear peak alignment.

1.2. Interaction between syntax and intonation in encoding questions

In her study on Dutch question intonation, Haan predicted that there is a functional trade-off between the syntactic and/or lexical markers of interrogativity and high question pitch (the Functional Hypothesis) [9]. Specifically, the less a question is syntactically and/or lexically marked, the more present high question pitch is. It follows that high pitch is maximally present in declarative questions but minimally present in wh-

questions. The Functional Hypothesis was borne out by Haan's data (read speech): final high pitch occurred most frequently in declarative questions and least frequently in wh-questions. The same pattern has also been reported for a number of varieties of British English including London English [6]. The second question we addressed in this study is thus whether the Functional Hypothesis is applicable to learners' use of peak height and peak alignment as well as their use of the final rise.

2. Method

2.1. The learner question corpus

This study is based on longitudinal data of four untutored Italian learners of English, Andrea, Lavinia, Santo and Vito, taken from the European Science Foundation (ESF) Second Language Database [11]. Table 1 shows the background information on the four learners, including gender, age of arrival (in the leftmost column under the name of the learner), region of origin (region), length (number of months) of stay in TL (stay), hours of occasional TL class (TLcl), rudimentary command of a third language (L3), and contact with TL speakers (contact).

Table 1. Background information on the learners

	Region	Stay	TLcl	L3	Contact
Andrea (M, 36)	Como (North)	5	30	French	low
Lavinia (F, 20)	Trieste (Northeast)	5	+600	French	some
Santo (M, 25)	Naples (Southwest)	7	0	none	rare
Vito (M, 22)	Caserta (Southwest)	12	150	none	high

These learners were newly arrived immigrants with secondary schooling and in their twenties or thirties at the start of the data collection of the ESF database. They were followed over a period of approximately 30 months, and recorded at 4 to 6 week intervals. During each visit they were asked to perform a number of tasks (e.g. free conversation, role play), and care was taken to ensure that after each 10 month cycle all tasks were performed at least once. In this study, we focused on questions produced in free conversation between the learners and their interlocutors (native speakers of English) at the end of the first cycle. Questions were operationalised as utterances that were uttered to seek information (e.g. time, place, reason, missing argument, truth value of an assertion etc.) from the interlocutor and were responded to by the interlocutor with the information required. A question was excluded from further analysis if it was unintelligible (due to noise) or parts of the utterance overlapped with the interlocutor's speech. When there were too few questions produced in the last recording session of the first cycle, we also included questions produced in slightly earlier or later sessions, as in the case of Andrea and Lavinia. As for Vito, questions produced during the first recording of the third cycle were included because he spoke very little English, and reverted to Italian during large parts of his conversations during the first two cycles. Three types of questions are covered in our corpus of learner questions:

- WH-questions (WQs): *Where you sleep?*
- Declarative questions (DQs): *You English?*
- Yes-no questions (Y/NQs) (with inversion and finiteness marking): *Are you English?*

The distribution of questions across learners is shown in Table 2. As can be seen, the learners produced very few Y/NQs. This may be related to two factors. First, Italian does not have the equivalent for the English Y/NQs. Second, the learners had not acquired finiteness marking in English at the time of recording and consequently seldom used inversion which involves the finite verb (Christine Dimroth, personal communication).

Table 2. Distribution of question types across learners

	WQ	DQ	Y/NQ	total	Sessions
Andrea	9 (9)	7 (7)	4 (3)	20 (19)	1-7, 1-8, 2-1
Lavinia	6 (5)	12 (7)	2 (0)	20 (17)	1-2, 1-5, 1-7
Santo	13 (13)	23 (23)	2 (2)	38 (38)	1-9
Vito	13 (13)	6 (6)	1 (0)	20 (19)	3-1
total	41	48	9	98	

* The number of questions included in the acoustic analyses are indicated in brackets.

2.2. Analyses

Analyses were conducted on the WQs and DQs. Y/NQs were not included for analysis because of the small of number of occurrences in our question corpus.

2.2.1. Intonational transcription

The intonation contour of each question was transcribed in terms of pitch accent and boundary tones following the ToDI (Transcription of Dutch Intonation) notation [8] and the transcription was done using Praat. ToDI was originally designed to represent the intonational structure of adult Dutch. Different from ToBI [2], ToDI intends to capture the intonation starting from the stressed syllable of a word (hence only pitch accents with trailing tones such as L*H and H*L) and employs only one phrase-type, i.e., intonational phrase. ToDI has been adopted to transcribe the intonation of English [6]. Recent work on intonational development [4] has shown that with minor adjustments, ToDI is also suitable for the transcription of intonation in child speech. Similar to transcribing child speech, the transcriber dealing with leaner speech cannot be always sure which TL intonational contrasts the learner intended to produce. It is nevertheless doable and productive to follow the principles of a certain notation, e.g. ToDI, to transcribe the intonation of learners by its 'face value'. For example, if a word (e.g. *emperor*) is spoken with a high tone on the stressed syllable but a low tone on the following unstressed syllables, the intonation of the word is transcribed as H*L. Each author transcribed the intonation of two learners' questions but both authors checked all labels and resolved cases of disagreement together.

2.2.2. Acoustic annotation

Three segmental landmarks were marked in each question, as illustrated in Figure 1:

- S0: the begin of the stressed syllable bearing the nuclear accent (hereafter the accented syllable)
- H: the location of the highest f0 before the fall in words spoken with non-rising intonation; words spoken with rising accents (e.g. L*H) were not marked for H because of lack of f0 peak in the accented syllable
- S1: the end of the accented syllable
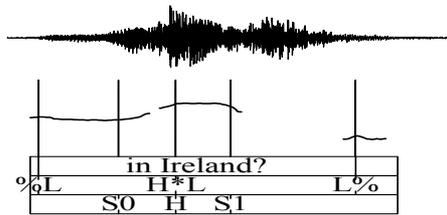
Figure 1. Segmental landmarks in the nuclear-accented syllable.

The alignment ratio was calculated on the basis of these landmarks: (H-S0)/(S1-S0). Further, peak height was measured by taking the f0 value at time H.

3.   Results and discussion

The length of the questions varied between one and eight words. 66% of the questions were spoken with a single pitch accent; the rest were spoken with two or more pitch accents. In the nuclear position, H*L, H*LH, !H*L, L*H and H* were used by all learners. In the prenuclear position, H*L and H* were used by all learners. These results suggest that the learners had a variety of pitch accent types at their disposal. Further, both the high and low boundary tones were used.

3.1. The final rise vs. the final fall

Figure 2 shows the distribution of final rises and final falls over question types for each learner. As can be seen, all the four learners used the final rise more frequently in DQs than in WQs. The occurrence of the final rise in DQs is rather high in comparison to the occurrence of the final rise in WQs in the cases of Andrea (86% vs 33%) and Lavinia (75% vs. 17%). These results show that our learners' use of final rise is in line with the Functional hypothesis, whereby the final rise is expected to occur more frequently in DQs than in WQs.

There is, however, a noticeable difference between Andrea and Lavinia on the one hand and Santo and Vito on the other hand. Specifically, the final rise was the predominant pattern in Andrea's and Lavinia's DQs and the final fall in their WQs whereas the final fall was largely the predominant pattern in Santo's and Vito's DQs and WQs. From the acquisitional point of view, this suggests that Andrea and Lavina were able to use the typical intonation of DQs and WQs in Southern British English by the beginning of the second cycle whereas Santo and Vito were not.

What has caused such a difference among the learners? Factors that have been shown to contribute to differences in learners' proficiency include age of arrival, length of residence in the TL country, length of schooling in TL (including occasional TL classes). However, as shown in Table 1, none of these factors distinguishes Andrea and Lavinia from Santo and Vito. The only factor that sets Andrea and Lavinea apart from

Santo and Vito turns out to be the region of origin. Both Andrea and Lavinea came from the North of Italy. As DQs are spoken with a final rise in most Northern varieties of Italian, we may consider transfer from the learners' native dialect as a possible explanation for the inter-learner difference. But further research on question intonation in these learners' dialects is needed to verify this explanation.
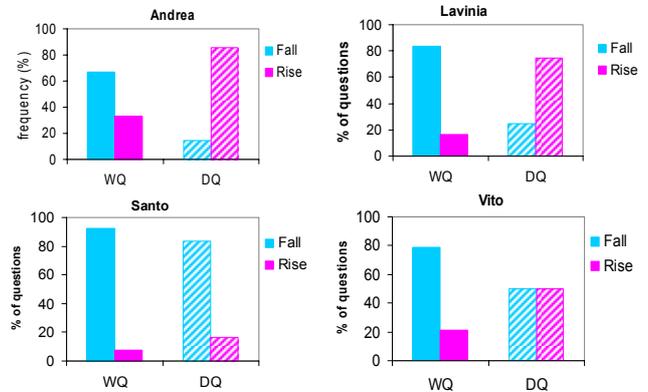
Figure 2. Distribution of final rise and final fall over question types.

3.2. Peak height and peak alignment

The alignment ratios appeared to be either in the initial one third of the stressed syllable or in the later half in a large number of questions irrespective of the number of peaks in the questions. We therefore considered 0.4 as the cut-off point between relatively early and relatively late peaks. Regarding peak height, as the learners differed in their pitch range, a cut-off point between relatively low peaks and relatively high peaks was set for each learner separately, taking into account of the heights of the peaks present in the learner' questions. The cut-off peak heights were: Andrea-125Hz, Lavinia-300Hz, Santo-200Hz for single-peaked questions/220Hz for multi-peaked questions, Vito-230Hz. On the basis of these cut-off points, we compared the proportion of questions with early or late alignments and high or low peaks between WQs and DQs for each learner.

The height and alignment of the nuclear peaks are plotted for Andrea and Santo. In each scatter plot, the reference line (the vertical line) for the X-axis separates the peaks that were reached before and after the 0.4 cut-off point; the reference line (the horizontal line) for the Y-axis separates the relatively higher peaks from the relatively lower peaks.

3.2.1.   Learner Andrea

As can be seen in Figure 3, across questions, all but 3 peaks (2 in WQ_single peaks, 1 in DQ_single peaks) were lower than 125 Hz. This suggests that peak height was not used to distinguish DQs from WQs. As regards peak alignment, in questions spoken with a single peak, the peak was reached relatively early in all DQs and in five out of the seven WQs. In questions spoken with multi-peaks, the peak was reached relatively early in the only WQ and in two out of four DQs. Given the fact that one of the two relative late peaks in DQs spoken with multi-peaks peaked shortly after the cut-off point, we do not consider this as a representative difference between WQs and DQs spoken with multi-peaks in Andrea's English. These results indicate that neither peak height nor peak alignment differed in different question types, lending no support to the Functional Hypothesis.

Andrea_WQ_single peaks

Andrea_DQ_single peaks

Andrea_WQ_multi-peaks

Andrea_DQ_multi-peaks
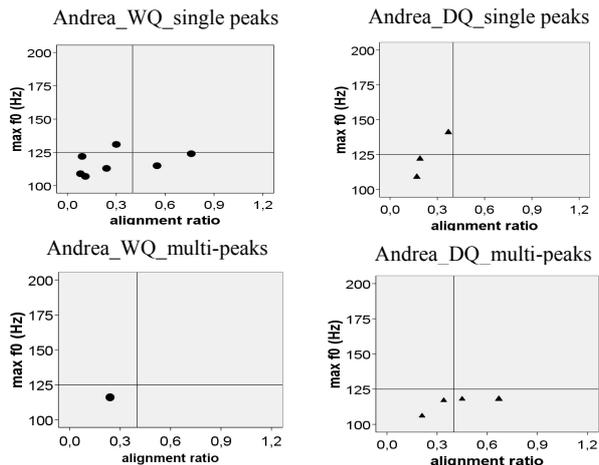
Figure 3. Height and alignment of nuclear peaks (Andrea)

3.2.2.   Learner Lavinia

Similar to Andrea, Lavinia made use of neither peak height nor peak alignment to distinguish DQs from WQs.

3.2.3.   Learner Santo

In questions spoken with a single peak, the peak was reached after the cut-off point in over one third of the time in both WQs (3 out of 8) and DQs (6 out of 17) (see Figure 4). In questions spoken with multi-peaks, the peak was reached relatively early in all the 4 WQs and in most of the DQs. Interestingly, in questions spoken with a single peak, the peak was lower than 200Hz in most of the WQs but was higher than 200Hz in over one third of the DQs. Further, in questions spoken with multi-peaks, the peak was equal to or lower than 220 Hz in all the WQs but was higher than 220 Hz in over half of the DQs. Clearly, Santo used higher peaks more frequently in the DQs than in the WQs. Thus, his use of peak height is in line with the Functional Hypothesis but not his use of peak alignment.

Santo_WQ_single peaks

Santo_DQ_single peaks

Santo_WQ_multipeaks

Santo_DQ_multipeaks
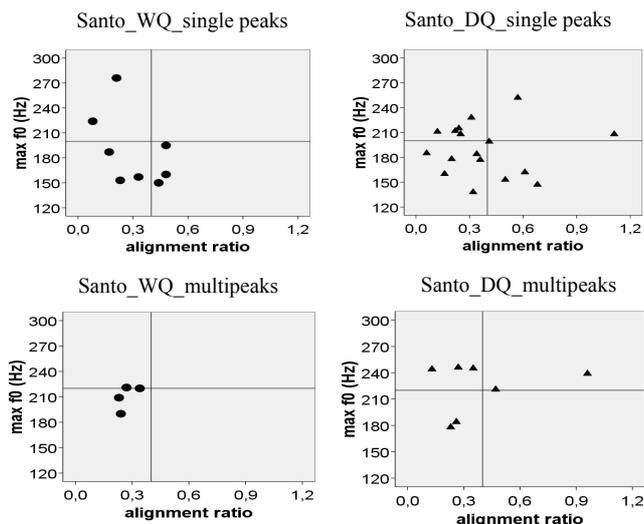
Figure 4. Height and alignment of nuclear peaks (Santo)

3.2.4.   Learner Vito

In Vito's data, there was no DQ spoken with multi-peaks. A comparison between WQs and DQs is only possible in the questions spoken with a single peak. We noted that the peak was reached later and realised with a higher pitch more often in WQs than in DQs, opposite to what is predicted by the Functional Hypothesis. It is worth mentioning that the peak was realised relatively late in five out of seven WQs spoken with multi-peaks. This is similar to WQs spoken with a single peak**.**

4.   Conclusions

We examined the use of intonation in encoding interrogativity in four untutored Italian learners of English. Importantly, we found that across learners the final rise was used more frequently in DQs than in WQs, in line with the Functional Hypothesis. We also found a clear intra-learner difference. That is, Andrea and Lavinia used the final rise as the predominant pattern of DQs whereas Santo and Vito used more final falls in both DQs and WQs. This difference between learners may be explained by transfer from the learners' native dialect. The use of high and late peaks is, however, less consistent. The peaks in the nuclear non-rising pitch accents were not necessarily lower in DQs than in WQs. This holds for all the four learners. As for peak height, only Santo used more high peaks in DQs than in WQs, in accordance with the Functional Hypothesis.

The difference in the use of the three question cues may be explained by differences in the relative importance of the cues in the target language. Final rising pitch is obviously most important as it is the cue used in London English. High and late peaks are probably less important as they are only accessible as question cues via the operation of the Frequency Code. Further, as high and late peaks can also be used to signal emphasis in English, the final rising pitch may be considered a more efficient cue to encode questions.

5.   References

[1]   Avesani, C., 1990. A contribution to the synthesis of Italian intonation. *Proceedings of the ICSLP*. 833-836.
[2]   Beckman, M.E.; Ayers, G.M., 1994. Guidelines for ToBI transcription (version 2.0, February 1994).
[3]   Bolinger, D., 1978. Intonation across languages. In Universals of human language Vol. II: Phonology, 471-524. Palo Alto: Stanford University Press.
[4]   Chen, A., 2007. Intonational realisation of topic and focus by Dutch-acquiring 4- to 5-year-old. *Proceedings of ICPhS. XVI*, 1553-1556.
[5]   D'Imperio, M., 2002. Italian intonation: An overview and some questions. *Probus* 14, 37-69.
[6]   Grabe, E., 2004. Intonational variation in urban dialects of English spoken in the British Isles, In *Regional Variation in Intonation,* P. Gilles, P. and J. Peters, J. (eds.). Linguistische Arbeiten, Tuebingen, Niemeyer, pp. 9-31.
[7]   Gussenhoven C., 2004. *The Phonolgy of Tone and Intonation*. Cambridge: Cambridge University Press.
[8]   Gussenhoven, C., 2005. Transcription of Dutch Intonation. In *Prosodic typology and transcription: A unified approach*, S.A. Jun (ed.) Oxford: Oxford University Press, pp. 118-145.
[9]   Haan, J., 2001. *Speaking of questions: an exploration of Dutch question intonation*. Utrecht: LOT.
[10] Ohala, J., 1983. Cross-language use of pitch: an ethological view. *Phonetica* 40, 1-18.
[11] Perdue, C., 1993. *Adult language acquisition: cross-linguistic perspectives (vol.2)*. Cambridge: CUP.