The role of pragmatics and politeness in explaining prosodic variability

Stella Gryllia[1], *Mary Baltazani*[2], *Amalia Arvaniti*[3]

[1]University of Leiden, Netherlands
[2]University of Oxford, UK
[3]University of Kent, UK

s.gryllia@hum.leidenuniv.nl, mary.baltazani@phon.ox.ac.uk, a.arvaniti@kent.ac.uk

Abstract

Twenty speakers (10F, 10M) took part in a discourse completion task (DCT) to examine effects of politeness and context on tunes used with wh-questions in Greek: they heard and saw on screen short scenarios ending in a wh-question. DCTs were controlled for power, solidarity, and context (with scenarios leading to the wh-questions being used either to request information or to make a statement). The results confirmed the role of context: the two context types led to the elicitation of distinct tunes, L*+H L-!H% for information-seeking questions, and L+H* L-L% for implicit statements, with lower scaling and later alignment of the accentual H in the former, and differences in final F0 consistent with a !H% and L% boundary tone respectively. In addition, questions after information contexts were shorter, but with a significantly longer final vowel. Politeness also affected duration, with conditions requiring a greater degree of politeness (the addressee being non-solidary and of different social status than the speaker) leading to lower speaking rate. The results indicate that tunes are associated with different durational profiles, which are also influenced by politeness. These results support recent studies showing that the study of intonation must include parameters beyond F0.

Index Terms: intonation, politeness, Greek, wh-questions

1. Introduction

Greek wh-questions are typically produced with one of two tunes, which in autosegmental terms are represented as L*+H L-!H% and L+H* L-L% [1], [2], [3]. Previous research has shown that straightforward questions used to request information are typically, *though not exclusively*, produced with the former tune. This tune is also rated more polite by native speakers of Greek, though only when produced by female talkers (while the use of either tune is considered equally polite for male talkers), [3].

The L+H* L-L% tune can also be used for straightforward questions, but it is also used by speakers to make a statement or assert an opinion, typically a negative one; we refer to questions used in this way as *implicit statements*. As an illustration, in the example shown in (1), the speaker uses a question in order to elicit information from the addressee; in (2), however, when the speaker finds out there is no milk, she uses the question *What will I make him cappuccino with now*? to assert that making cappuccino is impossible and thus to express displeasure at the addressee's lack of foresight which led them to not buying a sufficient quantity of milk. We note that the fact that this assertion takes the form of a question does not make it less confrontational than an overt statement, i.e. using a question here with the L+H* L-L% tune is not a politeness device. As

mentioned, questions with this tune are used by speakers and interpreted by listeners as straightforward questions, since they are formally questions; however, the interpretation of implicit statement is robust, considering that the only cue for this interpretation is the tune: the two interpretations can apply even when questions are string identical [3].

This brief description of the phonology and pragmatics of the two tunes is based on large samples of perceptual data [3], but the production studies on the realization of the tunes are based on small numbers of participants; e.g. [2] and [3] were both based on four speakers each. Here we use a large sample of questions elicited from 20 Greek speakers to examine in more detail the phonetics of the two tunes in relation to the possible contexts in which they appear; we also consider the role of politeness in the realization of the tunes. Specifically we examine if the differentiation between the two tunes applies in this large sample in order to determine whether (a) the two tunes are indeed distinct and used consistently in response to different pragmatic contexts; and (b) whether politeness affects the choice of tune, as the results of [3] indicate, or leads instead to gradient changes in realization, as discussed in the literature of politeness and prosody (cf. [4], [5]).

2. Methods

2.1. Participants

Twenty native speakers of Standard Athenian Greek (10 females, and 10 males) were recorded. They were between 18 and 24 years old (mean 21, s.d. 2). At the time of the recording they were all students at the University of Athens, the University of Patras or the TEI of Western Greece (also located in Patras). They had all been brought up in Athens and had lived there most of their lives.

2.2. Materials

The materials were a set of scenarios that ended with a wh-question. The scenarios were created so as to take into account two elements, pragmatic context and politeness, with the latter being operationalized as power and solidarity (ideally, *degree of imposition* should also be included in the exploration of politeness, but the use of wh-questions precluded use of this politeness-related factor). Two types of contexts were used, following [3]: context A provided participants with a situation which naturally ended with asking a question in order to receive information (see 1 below); context B provided participants with a situation in which the wh-question was unlikely to serve its primary purpose of seeking information and was used instead as an implicit statement (see 2 below). In addition, the contexts were created so as to manipulate power and solidarity: situations were either between solidary or between non-solidary

speakers; within each set, speakers were equal, or unequal with the speaker being either superior or inferior to the addressee. There were 10 scenarios per combination of context, power and solidarity for a total of 120 wh-questions (2 contexts * 2 power * 3 solidarity * 10 scenarios).

(1) Context A
You return home from school hungry. You want to eat but only if dinner will be ready within half an hour, as you are going to the pool for swim practice and you don't want to feel full. You see your mother boiling pasta and you ask:
[ˈpote θa ˈine etimi i makaɾoˈnaða]
When will the pasta be ready?

(2) Context B
When your dad does the grocery shopping, he gets stingy and does not buy enough of anything. One day you have many friends at home and you run out of milk although you had warned him about it. When a friend asks you for a double cappuccino, you tell your dad:
[meˈti na tu ˈftçakso ˈtoɾa toŋ gapuˈtsino]
What will I make him cappuccino with now?

The questions varied with each scenario but had similar structure. They varied in number of syllables from 9 to 16 (mean = 11.5, mode = 11, s.d. 1.53). The wh-words or expressions varied so that half began with a stressed syllable (e.g. [ˈti] "what", [ˈpote] "when"), while the other half begun with an unstressed syllable (e.g. [me ˈti] "with what"). In addition, in all questions (with a few exceptions), the stressed syllable of the wh-word and the next stressed syllable were separated by two unstressed syllables, which is optimal in Greek [6]. All questions ended in words with penultimate stress (the default in Greek).

In total, 2400 utterances were recorded. The data of one male participant (M3) were discarded as it turned out that he had failed to respond to many of the stimuli. This yielded a total of 2280 usable tokens from 19 speakers (10 females and 9 males).

2.3. Procedures

The recordings took place at the University of Patras. The participants were recorded in a quiet room, using a laptop and a Yeti microphone set to cardioid. The task was an amended version of the Discourse Completion Task [7], in that the participants were given both the description of a situation (the scenario) and the text with which to respond to it (the wh-question). Each scenario and question combination was presented on a Powerpoint slide, one combination at a time. The participants saw the scenarios on screen and heard a prerecorded version. They saw the questions only in writing and were asked to utter them in a way appropriate for the situation described in the scenario. The slides were presented in random order, with each speaker being presented with a different randomization.

2.4. Measurements and statistical analysis

The questions were annotated using the facilities of Praat [8]. Here we report on the following measurements, which are illustrated in Figure 1:

1. Accentual H (AH) scaling; the F0 value in ERB at the highest point in the contour; this was always located in the vicinity of the stressed vowel of the wh-word;

2. Accentual H alignment, i.e. the distance of AH from the onset of the stressed vowel of the wh-word in ms;

3. Scaling of the final boundary tone (FB), i.e. the F0 value in ERB at the end of the contour;

4. Speaking rate over the entire question, defined as the duration of the question divided by the number of syllables (as defined by each question's phonological representation);

5. Pitch range in ERB, defined as the difference between the maximum and minimum F0 of each question;

6. The duration of the accentual vowel and the last vowel in the question.

As we were interested in systematic differences between tunes, rather than individual differences among speakers, we analyzed the resulting data using linear mixed effects models with CONTEXT, POWER, and SOLIDARITY as fixed factors, and speakers and items as random intercepts [9].

Based on previous results [3], we hypothesized that there would be systematic differences between questions uttered after context A and context B. Specifically, we expected that context A would favor the use of the L*+H L-!H% tune and context B would favor the use of the L+H* L-L% tune. Consequently we expected that after context A questions would show later AH alignment, and higher FB scaling; observation suggested that the accentual and final vowel would also be longer than after context B. We further expected that politeness would affect production in a gradient manner: specifically we expected that when the addressee was non-solidary or non-equal in power to the speaker, questions would be produced with lower speaking rate and extended pitch range.
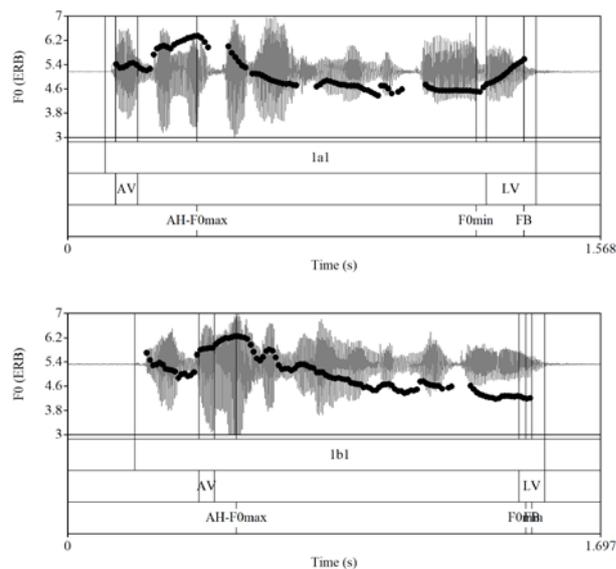

Figure 1. Illustration of measurements. Top panel: [ˈpote na peˈɾaso na to ˈpaɾo] "when should I come pick it up?" in response to a context of type A; bottom panel: [jaˈti na ti ˈblino ti bjaˈtela] "why should I wash the platter?" in response to a context of type B.

3. Results

3.1. Accentual High

We found a significant effect of CONTEXT on the scaling of the accentual H, AH [est. = −0.138, SE = 0.030, t = −4.558]: in context B (question as implicit statement), AH was significantly

higher ($\overline{x} = 6.83$ ERB) than in context A ($\overline{x} = 6.70$ ERB); see Figure 2. There was no effect of SOLIDARITY [est. = −0.009, SE = 0.030, t = −0.288], or POWER [for equal compared to inferior, est. = 0.001, SE 0.037, t = 0.035; for superior compared to inferior, est. = −0.052, SE = 0.037, t = −1.414].
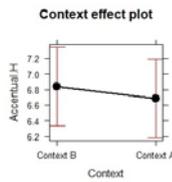

Figure 2. Effect of CONTEXT on the scaling of AH.

There was also a significant effect of CONTEXT on the alignment of AH [est. = 0.057, SE = 0.004, t = 14.234], with AH aligning significantly earlier in context B ($\overline{x}$ = 38 ms) than context A ($\overline{x}$ = 44 ms). SOLIDARITY also had a significant effect on the alignment of AH [est. = −0.009, SE = 0.004, t = −2.232]: AH was aligned significantly earlier when the speakers were solidary ($\overline{x}$ = 41 ms) than non-solidary ($\overline{x}$ = 42 ms), though the difference was minimal. There was a similarly small effect of POWER on AH alignment for the contrast inferior to superior [est. = 0.011, SE = 0.005, t = 2.158], such that AH was aligned earlier ($\overline{x}$ = 41 ms) when the speaker was inferior to the addressee than when the speaker was superior ($\overline{x}$ = 42 ms). On the other hand, there was no effect of power for the comparison between inferior and equal [est. = −0.005, SE = 0.005, t = −1.000]; see Figure 3.
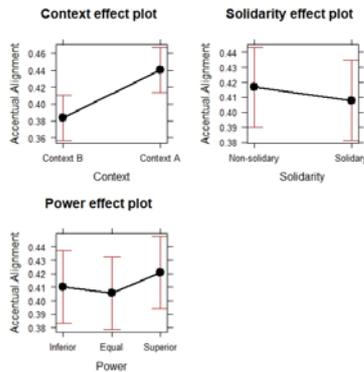

Figure 3. Effects of CONTEXT, SOLIDARITY and POWER on the alignment of AH.

3.2. Final Boundary Tone

CONTEXT had a significant effect on the scaling of the final boundary tone, FB [est. = 0.379, SE = 0.026, t = 14.834]; in context B, FB was significantly lower ($\overline{x}$ = 4.54 ERB) than in context A ($\overline{x}$ = 4.92 ERB). SOLIDARITY also had a significant effect on the scaling of FB [est. = −0.070, SE = 0.026, t = −2.745]; FB was higher when the speakers were non-solidary ($\overline{x}$ = 4.77 ERB) than when they were solidary ($\overline{x}$ = 4.69 ERB). POWER partially affected the scaling of FB; when the speaker was equal to the addressee, FB was significantly lower ($\overline{x}$ = 4.63 ERB) than when the speaker was inferior to the addressee ($\overline{x}$ = 4.76 ERB), [est. = −0.139, SE = 0.031, t = −4.439]. There was no effect of

POWER for the comparison between inferior and superior [est. = 0.031, SE = 0.031, t = 1.005]; see Figure 4.

3.3. Speaking rate

CONTEXT had no effect on speaking rate [est. = −0.869, SE = 0.711, t = −1.22]. SOLIDARITY, on the other hand, affected speaking rate [est. = −5.919, SE = 0.711, t = −8.32], such that speaking rate was significantly slower when the speakers were non-solidary ($\overline{x}$ = 147.06 ms/syllable) than when they were solidary ($\overline{x}$ = 141.15 ms/syllable). We also found an effect of POWER; in scenarios where the speaker was inferior to the addressee, the speaking rate was significantly slower ($\overline{x}$ = 146.27 ms/syllable) than when the two were equal ($\overline{x}$ = 141.44 ms/syllable) [est. = −4.907, SE = 0.869, t = −5.65]. There was no effect of POWER in the comparison of inferior with superior [est. = −1.701, SE = 0.871, t = −1.95]; see Figure 5.
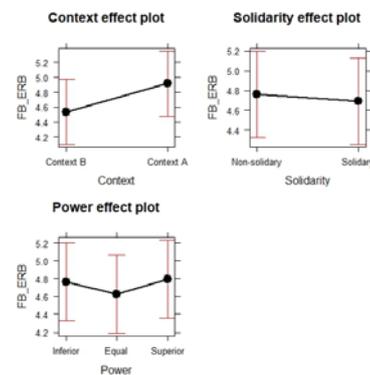

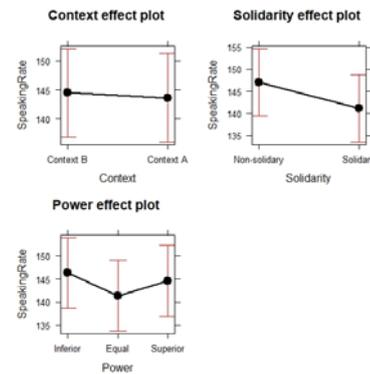Figure 4. Effects of CONTEXT, SOLIDARITY and POWER on the scaling of FB.

Figure 5. Effects of CONTEXT, SOLIDARITY and POWER on speaking rate.

3.4. Pitch range

CONTEXT had a significant effect on pitch range [est. = −0.106, SE = 0.0419, t = −2.530], which was larger in context B ($\overline{x}$ = 2.45 ERB) than in context A ($\overline{x}$ = 2.35 ERB); see Figure 6. We found no effect of SOLIDARITY [est. = −0.038, SE = 0.042, t = −0.914] or POWER [for equal compared to inferior, est. = 0.036, SE = 0.051, t = 0.702; for superior compared to inferior, est. = −0.030, SE = 0.051, t = −0.586].
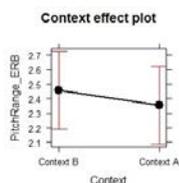
Figure 6. Effect of CONTEXT on pitch range.

3.5. Duration of the accented vowel and the last vowel

CONTEXT had a significant effect on the duration of the accented vowel [est. = −2.713, SE = 0.719, t = −3.775], which was longer in context B ($\overline{x}$ = 54.16 ms) than in context A ($\overline{x}$ = 51.43 ms), though the difference was small; see Figure 7. We found no effect of SOLIDARITY [est. = 0.465, SE = 0.719, t = 0.647] or POWER [for equal compared to inferior, est. = −1.591, SE = 0.878, t = −1.812; for superior compared to inferior, est. = 0.653, SE = 0.881, t = 0.741].
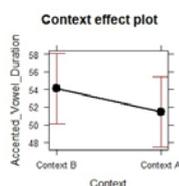

Figure 7. Effect of CONTEXT on the duration of the accented vowel.

There was also an effect of CONTEXT on the duration of the last vowel [est. = 12.905, SE = 1.102, t = 11.709], which was significantly longer in context A ($\overline{x}$ = 107.4 ms) than in context B ($\overline{x}$ = 94.52 ms); see Figure 8. There was no effect of SOLIDARITY [est. = 0.850, SE = 1.102, t = 0.771], or POWER on the duration of the last vowel [for equal compared to inferior, est. = −2.434, SE 1.347, t = −1.808; for superior compared to inferior, est. = 0.357, SE = 1.351, t = 0.264].
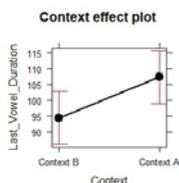

Figure 8. Effect of CONTEXT on the duration of the last vowel.

4. Discussion and Conclusions

The results confirmed previous descriptions of the two tunes used with wh-questions in Greek, reported in [1], [2], [3]: they showed that speakers respond differently to the two types of contexts used here, context A, in which questions were used in their prototypical function, namely to request information, and context B, in which questions were used as implicit statements. In response to context A, speakers produced accents with lower scaling and later alignment than in context B; this difference is consistent with what is known about the pitch accents of Greek represented as L*+H and L+H* [1], [10], and it indicates that L*+H was primarily used in response to context A, while L+H*

was primarily used in response to context B. In addition, the difference in the scaling of the final boundary tone (FB), is also consistent with earlier descriptions that prototypical questions in Greek are more likely to end in a !H% boundary tone, while those used as implicit statements are more likely to end in L%. In addition, the data show effects of the tune on segmentals, particularly the longer duration of the final vowel in context A. Although this difference could be dismissed as simply associated with the final rise (prevalent in context A), our results indicate that this explanation is insufficient, in that final rises were of small excursion and not always present.

In addition, the results indicate that intonation data show substantial variation when a large number of participants and utterances are examined, as in the present study. This is reflected in the fact that in many instances the differences found between the two tunes are not as large as reported in previous studies (e.g. [3]), an effect due to this variability. This is illustrated in Figure 1, which shows that the accentual peak is after the accented vowel of the wh-word in both questions, although the expectation from previous studies, such as [3], was that late peak alignment would be present only in the question ending in !H% (top panel). Examining the role and limits of such variability is critical for understanding how to connect phonological representations of intonation with their realization and is part of our planned research.

With respect to politeness, we find small but relatively consistent effects. First, our results do not show an overall increase in pitch range as an indication of politeness, as often expected (e.g. [4]). This is, however, consistent with other empirical studies, such as [5] on politeness in Catalan which also showed no changes in pitch range related to politeness. On the other hand, we did find that the FB was scaled higher when the speakers were non-solidarity and when there was a power difference between them (whether the speaker was inferior or superior to the addressee). Solidarity also affected speaking rate so that questions were produced with a slower rate when speakers were non-solidary. These results indicate that although some effects of politeness are global, such as the effect of solidarity on speaking rate, others are local; e.g. we do not see an overall increase of pitch range, but a local effect on the FB only. Moreover, the results on politeness are in line with previous research on politeness in Greek and on the use of indirect devices [11].

Overall the results uncover an intricate interplay between categorical and gradient effects both on F0 and on duration, relating to both politeness and the context-dependent choice of tune, which clearly affects segmental timing in addition to the scaling and alignment of tonal targets. This indicates, on the one hand, the need to pay close to attention to pragmatics in the study intonation (and prosody more generally), and on the other, the impossibility of incorporating this variability into representations of intonation in some systematic way. In turn, these results argue in favor of streamlined phonological representations of intonation coupled with rich phonetics.

5. Acknowledgements

6. References

[1] A. Arvaniti, and M. Baltazani, "Intonation analysis and prosodic annotation of Greek spoken corpora," in S. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, pp. 84–117, Oxford: Oxford University Press, 2005.

[2] A. Arvaniti, and D. R. Ladd, "Greek wh-questions and the phonology of intonation," *Phonology*, vol. 26, pp. 43-74, 2009.

[3] A. Arvaniti, M. Baltazani, and S. Gryllia, "The pragmatic interpretation of intonation in Greek wh-questions," *Proceedings of Speech Prosody 7*, 20-23 May 2014, Dublin, retrieved 04/01/2018 from http://www.speechprosody2014.org/, 2014.

[4] C. Gussenhoven, *The Phonology of Tone and Intonation*, Cambridge: Cambridge University Press, 2004.

[5] M. Nadeu, and P. Prieto, "Pitch range, gestural information, and perceived politeness in Catalan," *Journal of Pragmatics*, vol. 43, issue 3, pp. 841–854, 2011.

[6] A. Arvaniti, D. R. Ladd, and I. Mennen, "Stability of tonal alignment: the case of Greek prenuclear accents," *Journal of Phonetics,* vol. 26, pp. 3–25, 1998.

[7] S. Blum-Kulka, J. House, and G. Kasper. "Investigating cross-cultural pragmatics: An introductory overview," in S. Blum-Kulka, J. House, and G. Kasper (Eds.), *Cross-cultural Pragmatics: Requests and Apologies*, pp. 1–34, Norwood, NJ: Ablex, 1989.

[8] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 6.0.36, retrieved 11 November 2017 from http://www.praat.org/

[9] R. H. Baayen, D. J. Davidson and D. M. Bates, "Mixed-effects modelling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, pp. 390–412, 2008.

[10] A. Arvaniti, D. R. Ladd, and I. Mennen, "Tonal association and tonal alignment: Evidence from Greek polar questions and emphatic statements," *Language and Speech,* vol. 49, pp. 421–450, 2006.

[11] M. Sifianou. "Politeness and off-record indirectness," *International Journal of the Sociology of Language* 126, pp. 63-179, 1997.