10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

# The contribution of pitch accents and boundary tones to intonation meaning

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

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

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

## Abstract

Greek wh-questions are typically uttered with one of two tunes. Questions with the L*H L-H% tune are interpreted as information-seeking; questions with the LH* L-L% tune may be interpreted as non-information seeking, carrying implicatures of a negative type. As the tunes are nuclear and differ in both pitch accent and boundary tone, we conducted two experiments to test the contribution of each to these pragmatic interpretations: participants (N = 66 in Exp1; N = 88 in Exp2) heard questions in which the stretch that included either the boundary tone (Exp1) or the pitch accent (Exp2) was high-pass filtered to remove F0 information. The participants were asked to bet €0-100 on the most likely of two utterances following the question, choosing between follow-ups that indicated the pragmatic intent of the question to be information- or non-information-seeking. Bets on follow-ups indicating information-seeking intent were expected to be higher after questions with a L*H pitch accent (Exp1), or a H% boundary tone (Exp2). The results show that both pitch accents and boundary tones made pragmatic contributions and affected responses (i.e. strength of bets), thereby supporting the view that intonation meaning is compositional and nuclear tunes are not processed as a whole.

**Index Terms**: intonation, pitch accent, boundary tone, Greek meaning compositionality, intonation meaning, questions

## 1. Introduction

A long-standing issue in intonation research is how to model and understand intonational meaning. In the existing literature, disagreements arise both regarding the *nature* of intonational meaning, i.e. what information intonation may encode, and its *structure*. A recurring question on structure relates to whether intonational meaning is compositional or can be treated as a gestalt, e.g. [1, ch.15], [2, ch.8].

Regarding the nature of intonation, research in the autosegmental-metrical model of intonational phonology (AM) [2] recognizes that the main function of intonation is pragmatic [3]. Intonation may indicate utterance modality [4], information structure ([5], [6], [7], [8]), epistemic modality ([9]), and implicatures ([10], [11]).

Despite implicit agreement on the nature and functions that intonation encodes, AM lacks a generally accepted and experimentally validated theory of intonational meaning. As a result, debates regarding the structure of intonational meaning persist. According to some accounts, intonational meaning is *compositional*, determined by independent contributions of each of a tune's constituents, e.g. [10]. In this view, each element of the tune is a morpheme; the (deliberately abstract) meanings of all elements contribute equally to the overall meaning of the tune, which is interpreted in context, [10], [7].

Others have argued, however, that tonal meaning arises from sequences of tones grouped into morphemes contributing as a whole to pragmatic interpretation, and thus viewing tunes as idioms, [1], [12], or abstractions, [2], [13]. Such accounts often assign priority to the meaning of the *nucleus*, the part of the tune that consists of the last pitch accent and following edge tones, [1], [2], [13]. In these accounts, intonational meaning is determined by the nuclear tune as a whole; further, the nuclear tune takes priority over the contribution of other tonal elements, such as prenuclear accents.

In the experiments reported here we address the issue of compositionality of intonational meaning by examining the role of pitch accents and boundary tones in the interpretation of wh-questions in Greek uttered with two well attested tunes, [11], [14], [15], [16]. Since these tunes are nuclear, as detailed in 1.1., they are an ideal testing ground for examining both compositionality and views regarding the role of the nuclear tune in shaping intonational meaning.

### 1.1. Background on Greek wh-questions

A number of studies of Greek wh-questions have shown that they are typically (though not exclusively) realized with one of two tunes, autosegmentally represented as L*+H L-!H% and L+H* L-L% (henceforth represented in simplified form as L*H L-H% and LH* L-L% respectively), [11], [16]. The tunes differ in pitch accent and boundary tone. In the L*H L-H% tune, the L*H accent is realized as a rise from a low point to a late peak occurring on the first postaccentual vowel or later; if the wh-word does not include pre-accentual syllables, the rise is curtailed but present. The L- phrase accent is realized as a low F0 stretch, and the !H% as a small rise relative to the speaker's range, starting after the last stressed vowel of the utterance, or approximately in the middle of this vowel if it is utterance-final [15]. In the LH* L-L% tune, the LH* accent is realized as a rise from a mid-value to an early peak synchronized with the accented vowel itself; if the wh-word does not include pre-accentual syllables, the rise may be elided and the tune starts with high F0. After the fall from the accentual peak, F0 remains low and flat or slightly declining until the end of the utterance. Because of the way the tunes end, for convenience, we refer to the L*H L-% tune as rising, and to the LH* L-L% as flat. Figure 1 illustrates these two tunes with utterances used as stimuli in the experiments.

The two tunes lead to different interpretations of the questions [11]. Questions with the rising tune are treated as *primarily* information-seeking. Although such questions may also be interpreted as ironic, e.g. as a way of feigning ignorance, overall, straightforward wh-questions used to request

 10.21437/SpeechProsody.2020-73

information are typically, though not exclusively, produced with this tune. Wh-questions with the flat tune can also be interpreted as information-seeking, since they are syntactically wh-questions, but they may also be treated as carrying implicatures of a negative type ([11]); e.g. a question such as *why should I hire Paulina?* with the flat tune, shown in Figure 1b, can indicate that the speaker doubts Paulina's credentials. Note that these negative interpretations are compatible with the main questioning function of these utterances; thus, the question in Figure 1b could be interpreted as simultaneously asking for reasons to hire Paulina and indicating that the evidence for doing so is weak. In other words, questions with the flat tune may have a dual function in conversation (see [11] for a discussion).

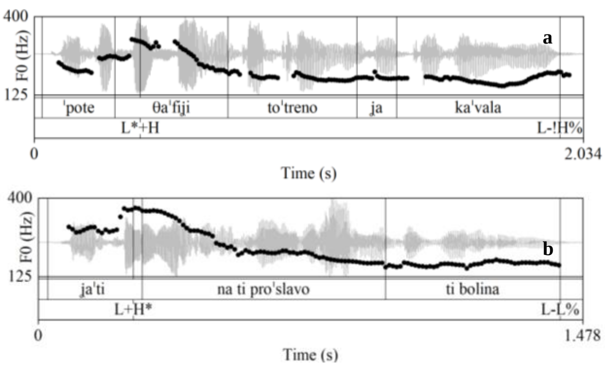


Figure 1: *Waveforms, F0 contours, transcriptions and AM annotations of two wh-questions used as stimuli in the present experiments; (a) "when will the train for Kavala leave?"; (b) "why [should I] hire Paulina?".*

### 1.2. Experimental design and hypotheses

Although this much is established based on [11], issues with this earlier analysis remain. First, [11] used a forced choice task, which required that participants treat the questions either as information or non-information-seeking but could not choose both interpretations, though both are available for the flat tune, as noted above. Second, [11] only tested the prototypical rising and flat tunes (L*H L-H% and LH* L-L%). Therefore, it is not possible to ascertain based on their results whether responses reflected separate effects of the pitch accent and boundary tone.

To address the question of compositionality we ran two experiments using the betting paradigm presented in [17]: rather than choosing an interpretation, participants had to bet on alternative interpretations of the utterances used as stimuli. We chose betting because it reflects the probabilistic nature of intonational meaning and its processing [18]. Further, betting does not force participants to pick only one interpretation, thereby acknowledging the lack of one-to-one correspondence between tune and meaning. In both experiments, participants heard wh-questions and were asked to bet between €0 and €100 on the most likely utterance that the speaker could use as a follow-up to their own question. The follow-ups provided an explanation of each question's *primary* intention as information-seeking or not. Examples are given in (1) and (2) with glosses of the two questions in Figure 1 and the two follow-ups for each one of them, one information-seeking (IS) and one non-information-seeking (NIS) follow-up.

We anticipated that if meaning is compositional, participants would place different bets depending on the pitch accent or boundary tone they heard (since only one was audible per experiment). If meaning is configurational, however, both experiments should lead to uncertainty (i.e. bets close to €50) except with controls, i.e. when the entire tune was audible.

| | | |
|---|---|---|
| (1) | Question | *When will the train for Kavala leave?* |
| | IS | *Do you happen to know?* |
| | NIS | *We've been waiting for too long!* |
| (2) | Question | *Why should I hire Paulina?* |
| | IS | *Is she looking for job?* |
| | NIS | *She is not qualified.* |

## 2. Experiment 1

### 2.1. Experiment 1: Method

#### 2.1.1. Experiment 1: Participants

Participants were recruited using social media and prolific (profilic.co; [19]). Participants recruited through prolific were modestly remunerated; those recruited through social media had the chance to win one of several gift cards. The results are based on 66 monolingual native speakers of Greek with no reported speech or hearing problems. Participants were randomly divided into two groups with each group hearing half of the stimuli and controls; 34 participants (14 female and 20 male, $\bar{x} = 29.4$, SD = 6.9) were assigned to version 1, while 32 participants (10 female and 22 male, $\bar{x} = 29.9$, SD = 6.9) were assigned to version 2. This was done to ensure the experiment was short and uncomplicated, as previous experience, [11], suggested this is critical for experiments on intonation meaning.

#### 2.1.2. Experiment 1: Stimuli & Procedure

The stimuli were based on 32 wh-questions elicited from 8 speakers of Standard Greek (4 female, 4 male) using a discourse completion task; each speaker contributed 2 questions with the rising and 2 questions with the flat tune. These 32 questions were used as controls and as the bases for the synthesized stimuli. To create the stimuli, the F0 of the controls was manipulated in Praat using the *Stylize pitch (2st)* function [20]. First, the initial F0 was set in the middle of the speaker's range for that question to render it ambiguous between LH* and L*H with respect to this cue (see section 1.1.). In addition, the stimuli were manipulated to have (i) an early peak (*EP*) in the middle of the wh-word's stressed vowel, typical of LH*, or (ii) a late peak (*LP*) placed 20 ms into the postaccentual vowel, typical of L*H. In other words, each initial question – referred to below as the *base* – was used to create both a stimulus with an early and a stimulus with a late peak. Since in the flat tune the accented vowel is longer than in the rising tune [16], this set up led to both matches and mismatches between F0 and segmental effects; e.g. a flat base manipulated to have an early peak presented a *match*, since the flat tune is LH* L-L% and early peak represents LH*; the same base manipulated to have a late peak presented a *mismatch*, since a late peak references L*H. The peak location manipulation yielded 64 stimuli (32 questions × 2 peak alignments). These were used together with the 32 controls to create 2 versions of the experiment, each based on 16 controls and the 2 stimuli derived from each one of them, one with L*H and one with LH*.

To minimize the contribution of the boundary tone, after F0 manipulation, the second half of each stimulus was high-pass filtered at 5 KHz with 12 dB attenuation using Audacity (but it was not otherwise altered). Filtering extended from the end of the first content word after the wh-word to the end of the question. For example, in (1), [ˈpote θaˈfiʝi toˈtreno ʝa kaˈvala] filtering started at the end of [θaˈfiʝi] *will leave*; in (2) [ʝaˈti na ti proˈslavo ti boˈlina], it started after [proˈslavo] *hire*. The overall effect gave the impression of a poor recording where some information was still available but not clearly so.

Each version of the experiment started with four practice trials, full questions from four of the speakers (these were not used in the main experimental session). In the main session, the stimuli were presented first, while controls (i.e. the original questions) were presented in one block of 16 trials at the end. This was done to avoid priming, i.e. so that participants did not hear the entire utterances with full F0 before they heard the stimuli, as this could influence their responses. Each version of the experiment included both stimuli with early and late peak.

The stimuli were presented using Psytoolkit in survey mode [https://www.psytoolkit.org/, [21]). Participants could hear each question as many times as they wished. They were instructed to bet €0 to €100 on one of two possible follow-ups of each question, like those presented in (1) and (2). A written version of the question appeared on the same screen as the follow-ups, because the manipulated questions were not fully audible. This is illustrated in Figure 2.

We anticipated that participants would find one follow-up more likely than the other based on the pitch accent: stimuli with a late peak (L*H) should lead to higher bets for information-seeking follow-ups than stimuli with an early peak (LH*). We expected controls to lead to the strongest bets, since the entire melody was available, followed by bets on *match stimuli*, i.e. those with matching pitch accent and base (such as L*H stimuli synthesized from a rising base), since segmental and F0 cues would work in them synergistically [22]. We anticipated that *mismatch stimuli*, i.e. those with mismatches between pitch accent and base (e.g. L*H stimuli synthesized from a flat base) would lead to bets closest to €50.

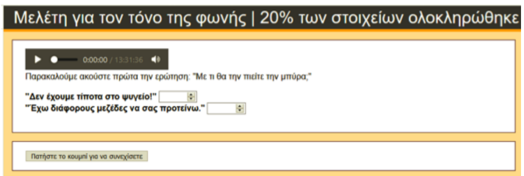


Figure 2: *Illustration of the screen viewed by participants in Experiments 1 and 2.*

#### 2.1.3. Experiment 1: Statistical analysis

We ran a number of linear mixed-effects models in R [23] using the *lmer* function of the *lme4* package [24], starting with a null model that included bets for information-seeking follow-ups as the dependent variable, and participants and items as random factors. The best fit model included bets as the dependent variable, pitch accent (early peak, late peak), stimulus type (control, matched, mismatched) and their interaction as fixed factors, and participants and items as random factors [Likelihood ratio test: $\chi^2 = 52.264$, $df = 2$, $p < 0.001$].

### 2.2. Experiment 1: Results

The results are illustrated in Figure 3. With respect to the controls, L*H led to higher bets for information seeking follow-ups than LH* [Est. = 39.493, S.E. = 3.098, $p < 0.001$]. This difference also applied to matches [Est. = 27.681, S.E. = 3.091, $p < 0.001$]. In mismatches, on the other hand, pitch accent did not differentiate bets [Est. = − 5.190, S.E. = 3.089, $p > 0.05$], with both LH* and L*H mismatches leading to bets close to €50 on average. Within LH* (early peak) stimuli, there was no difference between controls and matches, [Est. = − 1.028, S.E. = 1.764, $p > 0.05$], and both controls and matches lead to significantly lower bets than mismatches [controls vs. mismatches, Est. = − 17.697, S.E. = 3.093, $p < 0.001$; matches vs. mismatches, Est. = 16.669, S.E. = 3.090, $p < 0.001$]. For L*H (late peak) stimuli, bets were significantly higher for controls relative to both matches [Est. = − 10.784, S.E. = 1.768, $p < 0.001$], and mismatches [Est. = − 36.986, S.E. = 3.095, $p < 0.001$]. Matches also led to significantly higher bets than mismatches [Est. = − 16.202, S.E. = 3.091, $p < 0.001$].

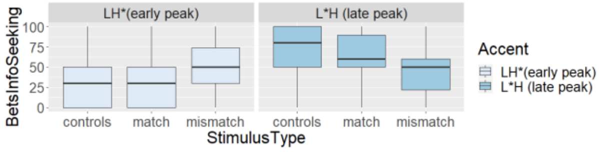


Figure 3: *Results of Experiment 1, separately for pitch accent and stimulus type.*

### 2.3. Experiment 1: interim discussion

The above results by and large support the hypotheses behind the experiment. First, they confirmed that the pitch accent alone can change the betting behavior of participants. We interpret this as an indication that the pragmatics of the questions is strongly influenced by the type of pitch accent used. Further, the results show that the base, which represents the segmental aspects of the tunes, such as differences in the duration of segments [16], also plays a part in how the questions are interpreted. This effect is evident with mismatches, i.e. when the base and type of accent are not matched (flat base with L*H; rising base with LH*). As predicted, with these stimuli, bets were closer to €50, indicating that the segmental and tonal cues conflict, leading to participant uncertainty. There is, in addition, a difference between LH* and L*H with respect to controls and matches: for the former, controls and stimuli yielded the same results, but for the latter they did not. A possible reason for this is that the peak in the L*H stimuli was not placed as late as it could be [16]. This is a point that requires further investigation.

## 3. Experiment 2

The methods were largely the same as in Experiment 1.

### 3.1. Experiment 2: Methods

#### 3.1.1. Experiment 2: Participants

The results are based on data from 88 monolingual native speakers of Greek with no reported speech or hearing problems; they were recruited using social media and were given the chance to win gift vouchers. They were divided into two groups of 41 and 47 participants each (version 1: 35 female, 6 male, $\bar{x}$ = 32, SD= 8.1; version 2: 34 female, 13 male, $\bar{x}$ = 33.7, SD= 7.1). As in Experiment 1, each group heard half of the stimuli and controls.

#### 3.1.2. Experiment 2: Stimuli, procedures, and statistical analysis

In Experiment 2 high-pass filtering affected the beginning of the stimuli, specifically the stretch from the onset of the question to the end of the first content word after the wh-word;

e.g., in (1), filtering ended at the offset of [θaˈfiji] *will leave*; in (2), it ended at the offset of [proˈslavo] *hire*. This minimized the contribution of the pitch accent. The F0 of the stimuli was manipulated to end flat, i.e. L%, or in a rise to the middle of the question's pitch range, i.e. H%. This yielded 64 stimuli (32 questions × 2 boundary tones), which were used together with the 32 controls to create 2 experiment versions, each including 16 controls and the two stimuli derived from each control (so the two versions included both L% and H% stimuli). As in Experiment 1, controls appeared in one final block.

We anticipated that participants would find one follow-up more likely than the other based on the boundary tone: rising stimuli (H%) should lead to higher bets for information-seeking follow-ups than flat stimuli (L%). We expected controls to lead to the strongest bets, followed by bets on stimuli with matching boundary tone and base. We anticipated that stimuli with mismatches would lead to bets closest to €50, since final question vowels are longer in the rising than the flat tune [16], and this should lead to uncertainty.

The results are based on the best fit model that included bets for information-seeking follow-ups as a dependent variable, pitch accent, stimulus type and their interactions as fixed factors, and participants and items as random factors, [Likelihood ratio test: $\chi^2 = 22.537$, $df = 2$, $p < 0.001$].

### 3.2. Experiment 2: Results

The results are illustrated in Figure 4. H% controls led to higher bets for information-seeking follow-ups than L% controls [Est. = 49.459, S.E. = 3.496, $p < 0.001$]. This difference also applied to matches [Est. = 44.085, S.E. = 3.522, $p < 0.001$] and mismatches [Est. = 16.013, S.E. = 3.523, $p < 0.001$]. Within L% stimuli, controls led to significantly lower bets than both matches [Est. = 3.974, S.E. = 1.660, $p < 0.05$] and mismatches [Est. = 13.432, S.E. = 3.523, $p < 0.001$] and matches lead to significantly lower bets than mismatches [Est. = − 9.457, S.E. = 3.523, $p < 0.001$]. For H%, there was no difference between controls and matches [Est. = − 1.400, S.E. = 1.704, $p > 0.05$], and both led to significantly higher bets than mismatches [Est. = − 20.014, S.E. = 3.496, $p < 0.001$, for controls vs. mismatches; Est. = − 18.614, S.E. = 3.523, $p < 0.001$, for matches vs. mismatches]. Finally, although L% mismatches led to higher bets than L% controls and matches, and H% mismatches led to lower bets than H% controls and matches, the two sets of mismatches remained distinct: L% mismatches led to significantly lower bets than H% mismatches [Est. = − 16.013, S.E. = 3.523, p < 0.001].

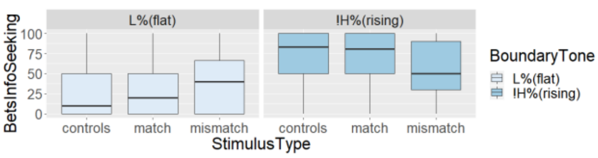


Figure 4: *Results of Experiment 2, separately for pitch accent and stimulus type.*

### 3.3. Experiment 2: Interim discussion

The results of Experiment 2 confirmed the hypotheses. Bets on information-seeking follow-ups were much lower for flat than rising stimuli. Controls showed the strongest bets (for or against information-seeking follow-ups), matches yielded overall similar results to controls, while mismatches led to weaker bets, though still differentiated between L% and H%.

## 4. General discussion

The main aim of these experiments was to examine compositionality in intonation meaning. The experiments confirmed the prediction that the pitch accent and the boundary tone make independent contributions to meaning: bets changed both in response to the pitch accent and the boundary tone in the absence of the other. This provides prima facie evidence for compositionality: if the tunes were configurations, participants should be more severely affected by the absence of tonal elements in the stimuli (boundary tone in Experiment 1, pitch accent in Experiment 2), while bet differences between controls and stimuli should have been more dramatic. Further, the fact that the responses to controls were stronger than responses to stimuli indicates that participants did not simply match the tune fragment they heard to one of the two tunes. These conclusions are in line with the results of [25] who tested entire tunes and combined LH* with H% and L*H with L% (in addition to testing the prototypical LH* L-L% and H*H L-H% tunes).

The results are also of relevance to debates regarding whether the nuclear tune is qualitatively different from other tune elements, such as prenuclear accents [1], [2], [13]. The present results indicate that Greek speakers do decompose the nuclear tunes tested here and consider separately the contribution of the pitch accent and boundary tone. The fact that they do not simply extrapolate from each component to the entire tune is evident from the unequal contributions of the pitch accent and boundary tone: as Figures 3 and 4 indicate, bets were stronger based on the boundary tone relative to the pitch accent. This is most evident with mismatches: pitch accent mismatches in Experiment 1 lead to bets close to €50, i.e. to uncertainty, while boundary tone mismatches lead to weak preferences but still differentiated L% from H%. These patterns agree with intuitions about the importance of the boundary tone in the interpretation of these tunes, e.g. [15].

Finally, the results support recent research (e.g. [16], [22]) regarding the role of intonation cues beyond F0. This is amply demonstrated here by the differences in bets between matches – where tonal and segmental cues were compatible – and mismatches – where such cues were contradictory. It is clear that segmental cues are taken into consideration by listeners, leading to greater uncertainty when cues do not match.

## 5. Conclusion

In conclusion, the results of these experiments provide evidence that intonational meaning is compositional with each tonal entity in the tune contributing independently to the tune's interpretation. This applied here to nuclear tunes of Greek wh-questions, suggesting that such tunes are not privileged and are not interpreted as a unit. This does not of course mean that the same applies to all languages, but it indicates that possible cross-linguistic differences must be considered. Finally, the present results confirm that segmental cues can aid the interpretation of intonation, strongly suggesting that their presence and contribution should be further investigated.

## 6. Acknowledgements

We thank the study participants and Nikos Vergis for setting the experiments in Psytoolkit. The financial support of the British Academy through grant SG160538 is hereby gratefully acknowledged.

# 7. References


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

[2] D. R. Ladd, *Intonational Phonology*. Cambridge: Cambridge University Press, 2008.

[3] J. Hirschberg, "Pragmatics and intonation," in L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics*, pp. 515–537. London: Blackwell, 2004.

[4] C. Bartels, *The Intonation of English Statements and Questions: A Compositional Interpretation*. New York: Routledge, 1999.

[5] D. Büring, *Intonation and Meaning*. Oxford: Oxford University Press, 2016.

[6] M. Steedman, "Information-structural semantics for English intonation," in C. Lee, M. K. Gordon, & D. Büring (Eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*, pp. 254–264. Dordrecht: Springer, 2007.

[7] M. Steedman, "The surface-compositional semantics of English intonation," *Language*, vol. 90, pp. 2–57, 2014.

[8] C. Féry, *Intonation and Prosodic Structure*. Cambridge University Press, 2017.

[9] P. Prieto and J. Borràs-Comes, "Question intonation contours as dynamic epistemic operators," *Natural Language and Linguistic Theory*, vol. 36, no 2, pp. 563–586, 2018.

[10] J. B. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication*, pp. 271–311. Cambridge, MA: The MIT Press, 1990.

[11] M. Baltazani, S. Gryllia and A. Arvaniti, "The intonation and pragmatics of Greek wh-questions," *Language and Speech*. https://doi.org/10.1177/0023830918823236, 2019.

[12] J. Peters, "Phonological and semantic aspects of German intonation," *Linguistik online*, vol. 88, 1/18, pp. 85–107. http://dx.doi.org/10.13092/lo.88.4191, 2018.

[13] F. Torreira and M. Grice, "Melodic constructions in Spanish: Metrical structure determines the association properties of intonational tone," *Journal of the International Phonetic Association*, vol. 48, no 1, pp. 9–32, 2018.

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

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

[16] S. Gryllia, M. Baltazani and A. Arvaniti, "The role of pragmatics and politeness in explaining prosodic variability," *Proceedings of Speech Prosody 2018*, https://www.isca-speech.org/archive/SpeechProsody_2018/, 2018.

[17] M. C. Frank and N. D. Goodman, "Predicting pragmatic reasoning in language games," *Science*, vol. 336, no. 6084, pp. 998. DOI: 10.1126/science.1218633, 2012.

[18] S. Calhoun, "The centrality of metrical structure in signaling information structure: A probabilistic perspective," *Language*, vol. 86, no. 1, pp. 1–42, 2010.

[19] S. Palan and C. Schitter, "Prolific.ac – A subject pool for online experiments," *Journal of Behavioral and Experimental Finance*, vol. 17, pp. 22–27, 2018.

[20] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[21] S. Stoet, "PsyToolkit - A software package for programming psychological experiments using Linux," *Behavior Research Methods*, vol. 42, no. 4, pp. 1096–1104, 2010.

[22] A. Arvaniti, "Crosslinguistic variation, phonetic variability, and the formation of categories in intonation," *Proceedings of ICPhS 2019*, https://assta.org/proceedings/ICPhS2019/, 2019.

[23] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2019.

[24] D. Bates, M. Maechler. B. Bolker and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4", *Journal of Statistical Software*, 67(1), pp. 1–48. Doi: 10.18637/jss.v067.i01, 2015.

[25] S. Gryllia, M. Baltazani and A. Arvaniti, "Evidence for the compositionality of tunes and intonational meaning," *Proceedings of ICPhS 2019*. https://assta.org/proceedings/ICPhS2019/, 2019.