Phonology **32** (2015) 537–541. © Cambridge University Press 2016
doi:10.1017/S0952675715000305

*Underspecification in intonation revisited: a reply to Xu, Lee, Prom-on and Liu**

Amalia Arvaniti
University of Kent

D. Robert Ladd
University of Edinburgh

We are naturally pleased that Xu *et al*. (2015) have taken the trouble to address our critique of the PENTA model, and it is useful to have a concise restatement of PENTA's aims and assumptions. However, we believe that their reply does not address the key point of our earlier paper (Arvaniti & Ladd 2009), which was that syllable-by-syllable specification of F0 does not makes theoretical sense in a language where F0 functions at the phrase or utterance level, and does not permit adequate quantitative modelling of complex intonation contours in short utterances.

To begin with the theoretical issue, Arvaniti & Ladd focused on a central problem which arises in describing intonation, the fact that contours with similar functions and globally similar shapes can apply to utterances of very different lengths. An abstract representation in terms of phonological landmarks such as local peaks provides a way of expressing the systemic equivalence of such contours, irrespective of the length of the utterance to which they are applied. Defining contours in terms of such landmarks entails the existence of what we termed SPARSE TONAL SPECIFICATION: there need not be an intonational target for every syllable, and the F0 on any given syllable may reflect nothing more than a transition between an earlier target and a later one. Conversely, in short utterances, a syllable may bear two or more intonational specifications. This idea does not, of course, originate with Arvaniti & Ladd; it is implicit in Bruce's pioneering analysis of the Swedish accent distinction (1977), and sparse tonal specification as a general principle was explicitly discussed with respect to Japanese by Pierrehumbert & Beckman (1988). The purpose of Arvaniti & Ladd's paper was simply to show how this principle, in addition to making phonological sense, provides insight into various phonetic details of the contours on Greek wh-questions, and to show that the same phonetic details are difficult to account for under PENTA's assumption of

* E-mail: A.ARVANITI@KENT.AC.UK, BOB.LADD@ED.AC.UK.

'syllable-sized pitch targets'. To avoid misunderstanding, we emphasise that what we mean by this phrase is simply that each syllable has an underlying pitch specification, a PITCH TARGET in PENTA. The details of the F0 are determined by context in combination with these targets; the issue is whether every syllable needs an underlying pitch specification at all.

In their reply, Xu *et al*. do not address this fundamental challenge. They simply restate the assumption (p. 515):

> PENTA's imperative for pitch target for each syllable comes from its core assumption about speech articulation, as represented by the TA model shown in Fig. 2. That is, the F0 contour of every syllable comes from a single mechanism: articulatory approximation of an underlying pitch target in synchrony with the syllable. Thus there is no other way of generating an F0 contour for a syllable than assigning it an underlying pitch target.

They justify their unwillingness to abandon this core assumption in two ways. First, they believe that they have a superior conception of intonational function, and second, they claim that the qTA component of PENTA successfully models and predicts the phonetic detail of a wide variety of contours, based on this function-centred view. We briefly address these two points in turn.

With regard to function, Xu *et al*. state that the autosegmental-metrical approach to intonation is concerned purely with form. This statement betrays a fundamental misunderstanding. Autosegmental-metrical phonology, like any phonological analysis, examines form *together with meaning,* attempting to determine which phonetic differences signal meaning distinctions. Unlike PENTA, that is, it does not assume that certain very specific communicative functions like 'focus' are easily definable and identifiable across languages. Rather, the autosegmental-metrical literature includes several accounts of intonational meaning (e.g. Gussenhoven 1984, Pierrehumbert & Hirschberg 1990, Steedman 2014) which are based on the assumption that intonation can be used to encode a variety of often very broad or general pragmatic meanings, and that specific intonational nuances are determined by intonational form and context operating in tandem (see Ladd 2008: ch. 1 for further discussion). These researchers do not agree on one single analysis of intonational meaning, because, instead of defining a limited set of communicative functions *a priori*, autosegmental-metrical theory considers intonational meaning to be subject to empirical investigation, with ordinary assumptions about the relation between meaning and form.

As for the argument based on modelling, it has two clear weaknesses. First, *any* argument based on quantitative modelling needs to acknowledge that models and quantitative predictions can be reasonably successful even in the absence of sound theoretical understanding. To take an extreme example, the ancient Babylonians were able to predict eclipses with remarkable accuracy based solely on empirically observed periodicities and without any clear idea of the earth's position relative to the sun and the

moon (Steele 1997); closer to the topic at hand, Lindblom (e.g. 2004) has often cautioned against confusing phonetic 'curve-fitting' with genuine understanding. There is no doubt that Xu's early work on tonal coarticulation in Mandarin, based as it is on serious attempts to understand the physical basis of speech F0 control (e.g. Xu 1999, Xu & Wang 2001, Xu & Sun 2002), makes an important contribution to our knowledge, but the fact that it yields a fairly accurate model of spoken F0 contours in Chinese is no guarantee that its theoretical insights into speech production are either correct or more widely applicable.

Second and more important, Xu *et al*. have not answered our specific points about the ways in which PENTA is in principle unable to describe certain features of the Greek wh-question contours discussed in Arvaniti & Ladd (2009). In their §4 they present qTA simulations of two medium-length illustrative contours, focusing primarily on the problem of stress clash. They avoid the more general problem of comparing very short and long contours, which was our central point, and ignore some of our relevant findings. Space does not permit a detailed discussion, but we would note at least the following.

(i) Xu *et al*. account for our finding that the nuclear high peak is aligned earlier in stress clash contexts by invoking the 'target strength' of the immediately following stressed syllable. They note (2015: §4.2.1) that 'because there is no anticipatory mechanism in qTA', more distant stressed syllables would not be expected to have any such effect, which is consistent with Arvaniti & Ladd's paper. However, they do not mention our finding (2009: 58) that the effect of stress clash is significantly greater in short sentences than in long ones, which does seem to require look-ahead.

(ii) Moreover, although they invoke the 'target strength' of the post-nuclear syllable to explain the effects of stress clash on the alignment of the nuclear accent peak, they go on to explain the *absence* of effects of stress clash on the *scaling* of the same nuclear accent peak by saying (§4.2.4) that 'there is no real leftward push from the first post-focus syllable'. They do not comment on the apparent contradiction between this explanation and the previous point.

(iii) They suggest that greater 'target strength' on a final stressed syllable will account for the differences we report in the alignment of the sentence-final rise. They do not make clear why the *contour* target on a sentence-final post-focus stressed syllable should yield a lower F0 (their Fig. 8b) while the *level* target on a non-final post-focus stressed syllable should have a higher F0 (Fig. 8a), though this stipulation may help them more closely approximate our empirical data for medium-length utterances. They also say nothing about the fact that stressed syllables that are neither sentence-final nor immediately post-focus have no effect on F0 whatever, as clearly shown in Arvaniti & Ladd (2009: Figs 1c, 2).

(iv) More generally, they make no attempt to model the stretches of low level F0 between the postnuclear F0 fall and the sentence-final rise. Their simulation of the contour in Fig. 8b shows a simple slope from the nuclear peak to the onset of the final syllable, and they even speculate (§4.1) that

Greek wh-questions may show 'a progressive rise throughout the sentence', which flatly contradicts the available literature on Greek wh-questions (e.g. Botinis 1998, Grice *et al*. 2000, Arvaniti & Baltazani 2005, Alexopoulou & Baltazani 2012, Arvaniti *et al*. 2014 – and Arvaniti & Ladd 2009).

We conclude by noting a more general problem with PENTA, which is that Xu *et al*. talk about 'prosody', but really mean F0. We suggest that a narrow conception of prosody as F0 is an important motivation for a model in which F0 is specified syllable-by-syllable. In Mandarin, F0 does need to be lexically specified for every syllable if it is to be properly modelled phonetically, and PENTA provides an elegant and accurate model of Mandarin F0 contours. However, because they believe that PENTA captures something fundamental about how F0 functions in all languages, Xu *et al*. assume that F0 in any language must therefore be controlled by syllable-by-syllable specifications. But the same assumption can just as plausibly lead us to the conclusion that voice quality must also be specified syllable-by-syllable in all languages. In some Nilotic languages, every syllable has one of two distinctive voice qualities, in addition to distinctive tone and quantity; in Vietnamese and some Chinese languages, the syllable tones typically involve both voice quality and F0 specifications. Models of speech production *in any of these languages* will therefore necessarily involve a voice-quality specification for every syllable. But since in all languages every syllable *has* voice quality, and since this is created by the mechanisms of speech production, PENTA's logic suggests that any model of voice quality in any language will also necessarily involve specifications for each syllable. As voice quality in most European languages is often a matter of long-term 'settings' (Laver 1980), any such syllable-by-syllable specification, no matter how successfully it modelled phonetic detail, would necessarily miss something fundamental about how voice quality is used. We believe that the same is true of PENTA's approach to F0 in languages with utterance-level F0 patterns. Xu *et al*.'s reply does not address this issue.

REFERENCES

Alexopoulou Theodora & Mary Baltazani (2012). Focus in Greek *wh*-questions. In Ivona Kučerová & Ad Neeleman (eds.) *Contrasts and positions in information structure*. Cambridge: Cambridge University Press. 206–246.

Arvaniti, Amalia & Mary Baltazani (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In Sun-Ah Jun (ed.) *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press. 84–117.

Arvaniti, Amalia, Mary Baltazani & Stella Gryllia (2014). The pragmatic interpretation of intonation in Greek wh-questions. In Nick Campbell, Dafydd Gibbon & Daniel Hirst (eds.) *Social and linguistic speech prosody: proceedings of the 7th international conference on speech prosody*. 1144–1148. Available (October 2015) at http://fastnet.netsoc.ie/sp7/sp7book.pdf.

Arvaniti, Amalia & D. Robert Ladd (2009). Greek wh-questions and the phonology of intonation. *Phonology* **26**. 43–74.

Botinis, Antonis (1998). Intonation in Greek. In Daniel Hirst & Albert Di Cristo (eds.) *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press. 288–310.
Bruce, Gösta (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.
Grice, Martine, D. Robert Ladd & Amalia Arvaniti (2000). On the place of phrase accents in intonational phonology. *Phonology* **17**. 143–185.
Gussenhoven, Carlos (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
Ladd, D. Robert (2008). *Intonational phonology*. 2nd edn. Cambridge: Cambridge University Press.
Laver, John (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
Lindblom, Björn (2004). Emergent phonology. Course given at the Graduate School of Language Technology in Finland, Helsinki University. Available (October 2015) at http://www.ling.helsinki.fi/kit/tutkijakoulu/courses/lindblom.shtml.
Pierrehumbert, Janet B. & Mary E. Beckman (1988). *Japanese tone structure*. Cambridge, Mass.: MIT Press.
Pierrehumbert, Janet B. & Julia Hirschberg (1990). The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.) *Intentions in communication*. Cambridge, Mass.: MIT Press. 271–311.
Steedman, Mark (2014). The surface-compositional semantics of English intonation. *Lg* **90**. 2–57.
Steele, J. M. (1997). Solar eclipse times predicted by the Babylonians. *Journal for the History of Astronomy* **28**. 133–139.
Xu, Yi (1999). Effects of tone and focus on the formation and alignment of f_0 contours. *JPh* **27**. 55–105.
Xu, Yi, Albert Lee, Santitham Prom-on & Fang Liu (2015). Explaining the PENTA model: a reply to Arvaniti and Ladd. *Phonology* **32**. 505–535 (this issue).
Xu, Yi & Xuejing Sun (2002). Maximum speed of pitch change and how it may relate to speech. *JASA* **111**. 1399–1413.
Xu, Yi & Q. Emily Wang (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication* **33**. 319–337.