Laboratory Phonology
Journal of the Association for Laboratory Phonology

Arvaniti, A 2016 Analytical Decisions in Intonation Research and the Role of Representations: Lessons from Romani. *Laboratory Phonology: Journal of the Association for Laboratory Phonology,* 7(1): 6, pp. 1–43, DOI: http://dx.doi.org/10.5334/labphon.14

JOURNAL ARTICLE

Analytical Decisions in Intonation Research and the Role of Representations: Lessons from Romani

Amalia Arvaniti
English Language and Linguistics, SECL, University of Kent, Cornwallis NW, Canterbury CT2 7NF, UK
a.arvaniti@kent.ac.uk

This paper presents an analysis of the intonational system of Greek Thrace Romani. The analysis serves to highlight the difficulties that spontaneous fieldwork data pose for traditional methods of intonational research largely developed for use with controlled speech elicited in the laboratory or under laboratory-like conditions from educated speakers of standardized languages. It leads to proposing a set of principles and procedures which can help deal with the variability inherent in spontaneous data; these principles and procedures apply particularly to data from less homogeneous speech communities but are relevant for the intonation analysis of any linguistic system. This approach relies on the understanding that autosegmental-metrical representations of intonation are phonological representations, not means of faithfully depicting pitch contours per se. It follows that representations should capture what is contrastive in the intonational system under analysis. In turn, this entails that new categories are posited, taking the meaning of tonal events into account and after due consideration of all legitimate sources of phonetic variation. It is argued that following this procedure allows for more robust analyses and is particularly advantageous when data are highly variable. This view is discussed in light of the analysis of Greek Thrace Romani, and in combination with recent proposals for greater uniformity and phonetic transparency in intonational representations, traits which are said to lead to greater insights in typological and cross-varietal research. It is shown that these goals are not better served by a level of broad phonetic transcription which encodes an arbitrary selection of phonetic variants.

1 Introduction

Much of the research in intonation has been laboratory-based. Paradigms for data collection and techniques that are widely accepted in intonational research were originally developed for use with controlled speech elicited from educated speakers of standardized languages. These paradigms and analytical techniques, concisely described in Jun and Fletcher (2014), have been successfully adopted in the documentation and analysis of a variety of intonational systems that go well beyond the languages for which they were originally developed (see Gussenhoven, 2004; Jun 2005a, 2014, for a number of languages analyzed along these lines). These methods have been very useful in determining a number of properties of the systems studied, including prosodic type, levels of phrasing, and tonal inventory.

The data collection paradigms in particular are well suited for the study of mainstream standardized languages as they largely involve the reading aloud of multiple repetitions of specially prepared sentences, short dialogues, or passages. Care is typically taken to use words mostly composed of sonorants in order to minimize microprosodic perturbations, a practice that results in largely smooth pitch tracks (cf. Jun & Fletcher, 2014, for advice on this point). Cooperative games and tasks are also used, like the map task (Anderson et al., 1991),

various forms of the discourse completion task (DCT; see, e.g., Borràs-Comes et al., 2014, and references therein), and specially designed games (e.g., Swerts et al., 2002). An example of the widespread use of these data collection paradigms is the Interactive Atlas of Spanish Intonation which includes data from 10 varieties of Spanish elicited using most of the tasks mentioned above (Prieto & Roseano, 2010). Using such tasks allows for the collection of semi-spontaneous data that still contain a number of controlled parameters. For instance, the words used are comprised mostly of sonorants and may be controlled for other variables like the position of stress or the type of structure elicited (e.g., the original HCRC map task maps contain single nouns, compounds, and noun phrases). In short, these practices result in largely smooth pitch tracks that are mostly uniform both within and across study participants and allow researchers to test specific hypotheses about the role of metrical structure, information structure, or any other parameter that is of interest.

The uniformity of data collected in the laboratory is so extensive that it is often considered a natural feature of human speech (see, e.g., Ladd, 1999) or at least highly desirable (Xu, 2010). For this reason it is worth enumerating the facets of similarity researchers have come to expect from intonation data based on characteristics of speech elicited in the laboratory or under laboratory-like conditions, particularly from educated speakers of standardized languages. First, such speakers can read aloud fluently and can do so for multiple repetitions while maintaining a consistent style that is similar across speakers and familiar to all from school.[1] This in turn means that balanced experimental designs with data that are comparable across speakers are the norm. Even semi-spontaneous tasks such as the map task or the DCT are based on skills that participants are likely to be familiar with, such as map reading and role-playing. Thus even in these less controlled tasks, participants are expected to maintain a consistent speaking style, speech rate, and volume and to follow turn-taking (Sachs et al., 1974). Further, speakers in the laboratory are likely to be young, educated, and middle class, characteristics that facilitate research in practical ways well: for example, such participants are likely to have healthy voices and use modal phonation (unless a different phonation mode is sociolinguistically appropriate for the community, such as creaky voice in California; Podesva, 2007; Yuasa, 2010). The importance of these elements cannot be underestimated, but it becomes apparent only when these conditions are not met (see, e.g., Henrich et al., 2010, on the expectations arising from research based on samples from Western, Educated, Industrialized, Rich, and Democratic [WEIRD] societies).

As a result of the above, many researchers have come to expect quite uniform data when it comes to intonation, and this expectation is by and large fulfilled, as even in studies that involve varied samples, intonational norms are shared among participants. As an illustration, Ritchart and Arvaniti (2014) investigated uptalk in California using the map task and eliciting data from a large number of speakers who varied in terms of socioeconomic class, gender, ethnicity, linguistic background, and geographical origin. They found mostly gender-related differences in the frequency and discourse function of uptalk, but few differences relating to form. Chung and Arvaniti (2013) report data from 15 Seoul Korean speakers, all of whom conformed fully to the intonational patterns of Korean as described in Jun (2005b), even when performing cycling, a relatively artificial task (Cummins & Port, 1998).

[1] This statement is not meant to imply that there are no exceptions, such as participants who cannot read a corpus of isolated sentences in a consistent style or speaking rate, or using a constant melody. But typically such speakers are the exception and not included in analysis.

The limited variability found in such data has led not only to expectations of uniformity in the realization of intonation but has also shaped the field's views about the perceived importance of such uniformity. In part the problem relates to the focus on form in much intonational research. This has been so both because attempts at codifying intonational meaning proved too complicated (as in the British School; e.g., Halliday, 1967; O'Connor & Arnold, 1973), and because the consideration of meaning has been limited to basic distinctions such as question vs. statement (for a discussion, see Arvaniti, 2011; Beckman & Venditti, 2011). The focus on form coupled with the ability to easily extract pitch tracks and treat them as faithful depictions of intonation appears to have strengthened the view that differences in form, particularly in the alignment of tones with respect to segmental landmarks, are sufficient to establish distinct tonal categories (see Kochanski, 2010, and Beckman & Venditti, 2011, for discussion of these practices from different perspectives). As a result of this view, small differences in alignment (and, to a much lesser extent, scaling) can be considered crucial in determining a tonal inventory and are incorporated into phonological analyses (see, e.g., Prieto et al., 2005, and relevant discussion in Arvaniti et al., 2006a). In turn, the focus on such differences has led to proposals for a level of intonational representation akin to that of a broad phonetic transcription (Hualde & Prieto, 2016).

While phonetic detail has taken such an important role in intonation research, concerns have also been voiced that similar intonational phenomena are not analyzed in the same way across languages. This is discussed at some length in Ladd (2008a, pp. 107–119), who argues that "if transcriptions are language-specific, we are left with no theoretically meaningful way to pursue cross-language comparison" (p. 115). This argument could be interpreted as a plea for more abstract phonological presentations, since abstractions are more likely to converge cross-linguistically, thereby facilitating comparisons. Ladd, however, seems to take the opposite stance: phonetic detail should be faithfully and uniformly represented in intonation research; doing so leads to phonetic transparency which in turn means that crosslinguistic similarities will not be obscured by language-specific representations.

I contend here that Ladd's arguments, which focus on the importance of intonational form, implicitly question the legitimacy of abstract phonological representations for intonation and by extension the very legitimacy of intonation as a fully-fledged part of phonological structure (doubts about a fully-fledged phonology of intonation have a long pedigree; see Crystal, 1969). The fact that intonation is treated differently from other aspects of phonological structure becomes evident if one compares Ladd's arguments on intonation with his arguments about segmentals. With respect to segmentals, Ladd (2011) argues persuasively against the use of a systematic phonetic level and in favour of abstract representations, on the one hand, and of measurable phonetic detail on the other (for similar arguments, see also Pierrehumbert et al., 2000). However, the systematic phonetic level he argues against when it comes to segmentals is precisely the type of level that Hualde & Prieto (2016) argue in favour of with respect to intonation; they do so using Ladd's own arguments about intonational representations (Ladd, 2008a, pp. 107–130, 2008b).

The focus on phonetic detail has led to neglecting intonational meaning to an extent that is rather striking when juxtaposed to standard practice in segmental analysis. One cannot imagine a fieldworker deciding that a particular vowel is phonemic in language *x* simply because it sounds similar to a vowel that is phonemic in language *y*. Surely, our hypothetical fieldworker would first wish to consult with native speakers of language *x*, use standard tests such as the presence of (near) minimal pairs, study the role of context in

observed variation and *consider the entire phonological system* before establishing the status of that vowel. In other words, she would rely on meaning differences, and context and system-internal observations to reach a decision, not on the precise value of the vowel's formants or their similarity to values used in another language.

Of the above criteria, meaning in particular has not featured prominently in descriptions of intonation (but see Jun & Fletcher, 2014, for good advice on this point). The importance of meaning becomes evident when data that do not conform to the uniformity assumptions discussed earlier are examined: when faced with variable data, it is difficult if not impossible to rely on similarity of form during analysis. Here, a corpus of Greek Thrace Romani is used to illustrate how analytical decisions can be made in the face of such variable data. Section 2 presents in more detail the reasons for the extensive variability in this corpus; section 3 presents the corpus and the principles used for analysis; section 4 illustrates the use of these principles with respect to stress, tonal inventory, and phrasing in Romani; finally, section 5 discusses the analysis in light of recent calls for surface phonetic representations of intonation (Hualde & Prieto, 2016), more typological research (Ladd, 2008a, pp. 107–130, 2008b), and the assumed superiority of laboratory data (Xu, 2010).

2 Sources of variability in Greek Thrace Romani

Greek Thrace Romani (henceforth *Romani*) is a Vlax variety of Romani spoken by Muslim Roma in Greek Thrace (Adamou, 2010; Adamou & Arvaniti, 2014; Arvaniti & Adamou, 2011). The Roma are a non-sedentary people who arrived in Europe from North East India approximately 600 years ago. As a people they have long suffered persecution; e.g. an estimated 220,000 Roma died at the hands of the Nazis and their collaborators during WWII (among many, Martins-Heub, 1989; Tyalglyy, 2009, and references therein). Possibly as a result of hostile attitudes toward them, the Roma form relatively closed communities that do not easily admit strangers, especially non-Roma.

The above apply to the Greek Roma communities as well. The exact number of Roma in Greece is not certain, as the Greek census of 2011 did not include questions about ethnicity. Estimates range from a minimum of 180,000 to a maximum of 350,000 Roma (or approximately 2.5% of Greece's population). The community is not homogeneous: some Greek Roma remain non-sedentary while others are settled in well-known neighbourhoods (e.g., Aghia Varvara in the outskirts of Athens). Communities also differ in terms of religion, with some groups being Muslim and others Christian. In addition, Greek Roma speak a number of Romani dialects; of these, Balkan Romani and Vlax Romani, originating in the Black Sea and Transylvania respectively, are the main ones (Matras, 2002).

The Romani variety in focus here has been referred to in previous work as *Greek Thrace Xoraxane Romane* (i.e., Turkish Romani) and is a mixture of Turkish and Romani as the name implies (Adamou, 2010; Adamou & Arvaniti, 2014). It is recognized as such by the speakers themselves who consider it to be distinct from Romani proper (Adamou, 2010; Adamou & Arvaniti, 2014). The data were collected from two communities, Anahoma, close to Komotini, and Drosero, close to Xanthi, both towns in Greek Thrace (see **Figure 1**); each community counts approximately 300 members. Although the distance between Xanthi and Komotini is only 55 km, there are dialectal differences between Anahoma and Drosero: in Anahoma, the Romani variety has mostly Vlax features, while Vlax and Balkan Romani are more mixed in Drosero (Adamou & Arvaniti, 2014). The differences are partly due to patterns of intermarriage in the two communities (the community in Drosero having closer ties with Roma in Bulgaria than the community in Anahoma). The ambient languages, Turkish and Greek, also exert an influence that adds to variability as is shown in more detail below (Adamou, 2010).

Figure 1: Map of Greece, showing Xanthi and Komotini, the Greek Thrace towns in the outskirts of which the Roma communities discussed here are based. Source: By Lencer [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons.

The speakers of the communities discussed here are trilingual in Romani, Turkish, and Greek. They tend to use Romani at home and within their community; they use Turkish and Greek for trade and other business, transactions with authorities, etc. However, Turkish is rapidly replacing both Romani proper and Turkish Romani (Adamou, 2010). This is in part due to proximity and close business ties with Turkey, but the trend is also strengthened by the fact that the speakers are Muslim and thus officially considered part of the Greek Muslim minority (which is strongly associated with Turkey). The classification of the Roma as part of the Muslim minority is based on the 1923 Treaty of Lausanne which ended WWI between Turkey and neighbouring states including Greece. The delineation of minorities in the treaty was based on religious rather than ethnic divisions following the practice of the Ottoman Empire.

A corollary of the above is that the Muslim Roma of Greece can be educated either in minority schools, which are Turkish-medium, or in mainstream Greek-medium schools. Although education is compulsory in Greece up to age 15, most of the Roma have at best elementary education. This applies to the speakers in the communities under discussion as well, most of whom have little or no schooling.[2] Further, because of the minority arrangements, there is no provision in Greek schools for Roma children to learn to read and write in Romani; thus the standard Romani variety used for transnational communication and education purposes in some European countries (Matras, 1999, 2005) is not known among the Roma in the communities under discussion. As a result of this situation, the reading of controlled sentences in Romani is out of the question, while the translation of sentences from Greek or Turkish is fraught with difficulties as speakers freely mix their three languages. Semi-controlled tasks, though possible as the present corpus demonstrates (see Section 3), must be chosen with care: unschooled speakers are not always comfortable with tasks such as map reading or the role-playing required by DCT, and can be weary of describing images depicting non-naturalistic situations (as happens, for instance, in the Questionnaire for Information Structure or QUIS; Skopeteas et al., 2006).

The linguistic situation of the communities discussed here has additional consequences for data-gathering. Frequent code-switching and mixing using three languages means that

[2] According to FRA and UNDP, in 2011 less than 10% of Roma children in Greece attended pre-school or kindergarten, while just over 35% of Roma children aged 7–15 attended school; less than 1% of the Roma population overall had completed upper secondary education.

semi-controlled data exhibit more variation than that found in monolingual communities. For example, in elicitation with materials from QUIS, the same speaker would use the Greek word [ˈkokini] for 'red.F' and the Romani word [loˈli] in response to prompts immediately following one another. In another QUIS game, one participant would consistently stress the word for 'gorilla' on the penult, as in Greek, producing [ɣoˈrila], while her interlocutor would vacillate between penultimate and default final stress producing both [ɣoˈrila] and [ɣoriˈla]. Such alternations mean that attempts to control prompts so as to avoid obstruents or elicit specific stress patterns can be easily thwarted.

Real-world and cultural sources of variability may also interfere with the quality of recordings, making certain types of measurements difficult to obtain. As in many Roma communities, living conditions do not allow for quiet recordings. According to FRA and UNDP (2012), the average number of persons per room is 2.7 for the Greek Roma (as opposed to just over 1 for the non-Roma population), while 35% of Greek Roma live in households without basic amenities, such as electricity, an indoor kitchen, bathroom, or toilet. Such conditions mean that quiet indoor recordings are rarely possible; recordings are likely to take place outdoors or involve bystanders. As the Roma communities discussed here are "high involvement" (Tannen, 1987), overlaps in conversation and multiple conversations taking place at the same time are also common. Finally, relying on spontaneous conversations also means that recordings tend to be uneven in speaking rate, volume, and pitch level and span, especially when the conversations become animated. As an indication, pitch in the speech of several female participants well exceeded 500 Hz when they were engaged in spontaneous conversation; the same women had substantially lower maxima in the semi-controlled data from QUIS, rarely exceeding 300 Hz. Similarly, the male speaker who participated in several tasks, reached a maximum of 280 Hz and rarely fell below 100 Hz when telling a story, but kept to a low level and small span of between 80 Hz and 180 Hz when taking part in various QUIS tasks.

In terms of analysis, the difficulties presented by the variability in the data are compounded by the fact that there is little research on Romani prosody on which to build an analysis: the bibliography of Romani linguistics by Bakker and Matras (2003) has more than 2,500 entries but just six publications that touch on intonation. They all treat Eastern European varieties but none that is dialectally close to Greek Thrace Romani. Thus, building on previous descriptions, as suggested by Jun & Fletcher (2014), is not possible in this instance. As with other previously undescribed languages, an analysis can only be based on (i) general assumptions related to typology, (ii) existing knowledge of realizational variability, (iii) a finite dataset, and (iv) knowledge of neighbouring systems (though, as Jun & Fletcher, 2014, point out, neighbouring systems may not necessarily share typological similarities with the system under analysis).

Taken all together, the elements discussed above mean that any corpus of Romani is likely to include multiple sources of variability and noise both literally and figuratively: the speakers do not speak a uniform variety and code-mix using three languages, the main one of which, Romani, shows extensive dialectal variation even among small groups like those examined here. Further, the speakers are unlikely to be educated to a degree that would allow them to read aloud with ease scripted materials, certainly not in Romani, and may approach some tasks (e.g., those involving role-play) with misgivings due to their unfamiliarity. Recordings are likely to be noisy because privacy and quiet spaces are hard to come by, while conversations are animated and involve multiple participants; at the same time, community members cannot afford to travel to studios and may even be skeptical of such endeavours.

3 Data and principles of analysis

3.1 The Romani data

The extensive variability created by the sources discussed above means that standard paradigms for data collection, even adapted to a fieldwork situation with an unschooled population, would be unlikely to be successful or would lead to a small sample in terms of number of speakers, a highly inadvisable outcome given the interspeaker variability present in the language. The Romani corpus discussed here contains instead mostly spontaneous speech from 10 speakers and a variety of speaking styles.[3] Specifically, the data include the following: story-telling from three speakers, two male and one female; spontaneous conversations, involving a total of nine speakers (eight female); semi-controlled data elicited from two female and three male speakers using QUIS (Skopeteas et al., 2006); elicitation of words and short phrases based on the Intercontinental Dictionary Series (Ritchie Key & Comrie; http://lingweb.eva.mpg.de/ids/) produced by one male speaker (who also contributed one story, and took part in the QUIS tasks and in spontaneous conversations). The majority of the speakers had little or no schooling. The two youngest participants were 16 years old and the oldest was in her 50s, but most participants were in their 20s or 30s. All gave oral informed consent.

3.2 Basic principles of analysis

In order to analyze the Romani data the following principles were adhered to. First, the aim was to arrive at a *phonological analysis*, not to develop a *phonetic transcription* of the Romani intonational system. This aim is not specific to Romani, to this particular project, or to non-standardized linguistic varieties. Rather, it is in line with the principles underlying the development of AM: what are often referred to as *AM transcriptions* are in fact meant to be phonological representations characterized by underspecification (Arvaniti, 2011; Arvaniti & Ladd, 2009; Beckman et al., 2005). This understanding of AM is in line with a more general understanding of the organization of sound systems which recognizes both the need of abstraction and the need for phonetic detail (Beckman et al., 2007; Ladd, 2011; Pierrehumbert, 2002).

This understanding of the nature and purpose of phonological representations has several consequences for analysis. First, it means that the aim of the presentations was not to more or less faithfully depict the course of F_0. As will be argued in more detail in Section 5.2., the course of F_0 can be represented much more accurately by the pitch tracks themselves. Instead, the aim of the analysis here was to determine the intonational elements that are contrastive in the Romani system. Adopting this view entails that meaning cannot be dismissed and phonetic form cannot be considered without reference to meaning. Rather, the analysis follows similar lines to those used to establish the segmental contrasts in a sound system: intonational events are identified and examined in terms of their pragmatic meaning to determine whether they are contrastive in the system under analysis; meaning in this instance would involve the role played by different intonational elements in discourse (e.g., highlighting, showing finality; cf. Pierrehumbert & Hirschberg, 1990). Decisions about the representation of the intonational elements deemed to be contrastive are based on (i) standard practice, (ii) system internal considerations (cf. Gussenhoven, 2007), and (iii) acceptance of what Arvaniti and Ladd (2009, p. 63) have called "lawful variability" (cf. Cangemi & Grice, 2016; Cole & Shattuck-Hufnagel, 2016; Frota, 2016). Each of these elements is discussed in detail below.

[3] Though the majority of the data comes from spontaneous speech, many illustrations here are from QUIS, as the data were less noisy and produced with clearer speech.

Standard practice was taken into consideration in determining the appropriate representation of tonal events. Thus, H was used to represent tones deemed to be high in a melody with respect to the speaker's range and other tones in the same contour; L was used for tones deemed to be low by the same criteria (cf. Pierrehumbert, 1980, pp. 68–75). Conventions that are gradually becoming established in the field were also followed, such as Jun & Fletcher's (2014) recommendation to dispense with the + sign in bitonal pitch accents unless there is evidence that the two tones align independently of each other; thus here LH* is used instead of L+H*.

System-internal considerations mean that phonetic detail was not part of the representations unless there was evidence it was contrastive. Decisions on contrastiveness were guided by meaning in combination with form: differences in form were considered contrastive after taking into account focus and information structure and the pragmatic function of utterances in discourse (cf. Pierrehumbert, 1980, pp. 59–63). The analysis involved several iterations, leading to both bottom-up and top-down decisions: a first set of data determined the original analysis, which was then used to annotate more data; analytical decisions were further tested with semi-controlled QUIS data which in turn made it clear that additional refinements and revisions were necessary.

In addition, the analysis was kept as simple as possible until this proved untenable. In other words, rather than annotating phonetic detail and determining at a later stage if doing so was justified (as Jun & Fletcher, 2014, recommend), the analysis started with the simplest possible annotation labels. For instance, instead of marking rising pitch accents with both a L and a H tone (i.e., as L*+H, L+H*, L*H, LH*, etc.), only H* was originally used. Once additional data indicated that narrow focus is signalled by the use of a rising accent with consistently different realization, a distinction between H* and LH* was adopted (see Section 4.2.1.).

The fact that simple representations were adopted means that not all tonal events are represented in the most phonetically exhaustive way possible. For instance, in Romani H* may show a rise to a peak. This rise is an optional element determined by context and thus considered part of the accent's phonetic realization — more specifically, of the scope of the accent's variability — but is not seen here as essential for its representation. Although others have argued against the loss of phonetic transparency in such cases (e.g., Ladd & Schepman, 2003), what is advocated here is standard practice in segmental phonology. For instance, in all accounts of English phonology, voiceless stops are represented as /p/, /t/ and /k/, i.e., with the IPA symbols for voiceless unaspirated plosives, even though /p^h/, /t^h/ and /k^h/, the symbols for voiceless aspirated plosives, would provide more faithful representations. This is in line with IPA guidelines (IPA, 1999; for a discussion see Ladd, 2011); it reflects the understanding that aspiration need not be part of the symbolic representation of these phonemes since they do not contrast for aspiration with any other phonemes of English. On the other hand, in Romani, which has a three-way contrast between prevoiced, short-lag, and long-lag VOT (Adamou & Arvaniti 2014), incorporating VOT into the phonological representation is essential. As a result of these widely accepted practices, a sound phonetically similar to English /p/ is represented as /p^h/ in Romani phonology, since in that system it contrasts with unaspirated /p/. System-internal considerations comparable to those pertaining to English stops led to the decision to represent the most frequent rising accent of Romani as H* (see Section 4.2.1.).

As noted above, the decision to adopt representations that are as simple as possible was also based on the understanding that intonational elements exhibit lawful variability, and that such variability in intonation should be considered at least as normal as it

is considered for segments (cf. Cole & Shattuck-Hufnagel, 2016). Some of the variation observed is related to speaker and style. In the present data it was immediately evident that some participants used clearer speech than others, but also that speech clarity depended on the task: spontaneous, animated conversations showed extensive coarticulatory effects as compared to the QUIS data; these differences were evident in intonation as well.

Variability may also relate to dialect. This particular point could not be explored in detail here due to the extensive code-mixing of Romani (but see Adamou & Arvaniti, 2014, for some dialectal differences in stress). Nevertheless, it is a type of variability worth discussing as it has often been neglected in intonation research. A good case in point is the contrast between H* and L+H* in English. This contrast is posited by Pierrehumbert (1980, ch. 4) and a pragmatic analysis of the difference between the two accents is presented in Pierrehumbert and Hirschberg (1990). The existence of the contrast, however, has been strongly disputed by others (see Ladd, 2008a, pp. 96–97, for a discussion). Indeed Ladd and Schepman (2003) propose that the representation (L+H)* replace H* and L+H*, on the grounds that all "sagging transitions" between high accents in English involve an F_0 dip consistently aligned with the onset of the accented syllable. As Arvaniti and Garding (2007) show, however, this argument, though valid for Ladd and Schepman's production data (which are based on one RP and one Scottish speaker), does not apply to all dialects of English. In Arvaniti and Garding's study, speakers from Minnesota clearly followed a pattern similar to that described by Ladd and Schepman (2003), i.e., always used an accent which started with a clear and consistent dip and relied on pitch range to distinguish new information from contrastive focus. Southern California speakers, on the other hand, maintained an equally clear distinction between H* and L+H*, using an accent with a shallow and inconsistently present dip (H*) for new information, and an accent with a consistently present and prominent dip with stable alignment (L+H*) to indicate contrastive focus. Data like these clearly show that dialectal differences must be given due consideration in intonation research; no researcher would discuss the vowels of "English" without specifying the variety being examined, or argue about the vowel contrasts in U.S. varieties based on data from RP. The same principle should consistently apply to intonation research as well.

In addition to the above sources of variability, context-related lawful variation should also be considered. Some contextual factors affecting the realization of tones are discussed below. They all largely reflect aspects of tonal crowding and undershoot (Arvaniti & Ladd, 2009; Fougeron & Jun, 1998; Grabe, 1998, ch. 5; Ladd, 2008a, pp. 180–184; Arvaniti et al., 2006b).

- Tonal context. Tonal events are affected by proximity to other events, with tonal crowding often resulting in elision or undershoot so that pitch modulations evident in some contexts are eliminated in others; e.g., Arvaniti et al. (2000) show that the L tone of L*+H pitch accents in Greek can be severely undershot or eliminated altogether if L*+H accents are on adjacent syllables. Undershoot and changes in alignment are also reported with respect to the tones of the Greek *wh*-question melody (Arvaniti & Ladd, 2009). The present data show undershooting of H* accents when they appear on consecutive syllables. For edge tones in particular, differences in realization depend on the location of pitch accents. As shown in detail in Section 4.2.3., L% boundary tones are manifested as low F_0 points when the nuclear accent is in absolute phrase-final position but as low F_0 stretches if it appears earlier. Tonal context may affect the realization of tonal

events even when crowding is not an issue; cf. Venditti et al. (2008), who illustrate significant variation in the scaling and alignment of the Japanese accentual H* + L depending on the nature of the boundary tones that follow.

- Location of the tone within the utterance. The prosodic position of a tonal event can also result in different realizations. Romani, for instance, shows positional variants of the H* pitch accent which is realized as a rise with peak delay in utterance-initial position but as a high fall in utterance-final position (see Section 4.2.1).
- Interactions of stress with phrasing. Variation often depends not only on the position of a tonal event within an utterance (e.g., initial, medial, or final) but also on its precise location. In the case of pitch accents, this is determined by the position of the stressed syllables. Thus in MAE-ToBI the contrast between H* and L + H* is considered to be neutralized in absolute utterance-initial position, as the L tone of L + H* is not realized in this context (Brugos et al., 2006, ch. 2.5). A similar situation is observed in Greek *wh*-questions, which show a rise from a low point when the stressed syllable of the *wh*-word is not utterance-initial; the rise is truncated if the *wh*-word starts with the stressed syllable (Arvaniti & Ladd, 2009; Arvaniti et al., 2014). In Romani, the first H* accent in an utterance is likely to show truncation in absolutely initial position as compared to its realization on a later syllable (see 4.2.1.).
- Segmental context. Segmental context also affects the realization of tones. The presence of voiceless obstruents may obscure glissandos, a phenomenon interpreted as truncation (but see Niebuhr, 2008, 2012). In the present corpus, this is evident in the realization of the H* accent which rarely shows a rise if the accented syllable starts with a voiceless obstruent. Languages may differ in how such environments are treated: Grabe (1998, ch. 5) reports that German shows truncation in these circumstances, while English prefers compression. Further, segmental context effects may overlap with location effects. In the Romani corpus all *wh*-questions started with a *wh*-word with initial stress and a voiceless initial consonant, such as /so/ 'what', /kon/ 'who', and /ˈkaste/ 'to whom'. Until additional evidence is available, it is assumed here that in Romani the contrast between H* and LH* is neutralized in this context. In such instances, positing the simpler representation (here H*) was preferred.
- Speaking rate effects. Changes in speaking rate can lead to the reorganization of speech, and intonation is no exception. Fougeron and Jun (1998) show that in French changes in speaking rate can affect pitch range and lead to the deletion or undershoot of underlying tones. Arvaniti and Garding (2007) and Mixdorff et al. (2014) report similar patterns for English and German, respectively. In the present corpus fast, less careful speech was characterized by a greater degree of tonal undershoot and anticipatory coarticulation of tones than more careful, deliberate styles. Since spontaneous speech tends to be fast, especially when speakers are animated, effects of speaking rate must be carefully considered when determining the tonal inventory.
- Language (and melody) specificity in choice of strategies. Examples of compression and truncation like those discussed above have led to suggestions that languages either compress or truncate (Grabe, 1998, ch. 5). The situation, however, is clearly more complicated than an *either/or* choice suggests (Ladd, 2008a, p. 182). In some languages at least, preferences in realization may differ depending on context. In Greek, *wh*-questions and consecutive L* + H accents show truncation of the L tone, as noted, but in polar questions compression is preferred for the L + H-L% edge tone configuration (Arvaniti et al., 2006b). Similarly, the routine calling

melody of Polish shows compression of the initial rise, while the melody used for urgently calling someone shows truncation of a similar rise (Arvaniti et al., 2016). Given the above, it is important during analysis to keep in mind that both options, truncation and compression, may be available to speakers. This is illustrated in the Romani data as well; while many tones are eliminated, the L* accent of polar questions shows evidence of compression instead (see Section 4.2.2).

- The nature of tones. Differences between L and H tones were discussed in Pierrehumbert (1980, pp. 68-75) and have been observed in several studies since (e.g., Prieto, 1998, 2006). Ladd (2008a: 182) mentions that L tones tend to be undershot or truncated more often than H tones. Arvaniti and Garding (2007, p. 569) also note that L tones show more consistent alignment than H tones, "the alignment of which appears to be affected by various parameters, such as emphasis [. . .], metrical factors, and speaking rate." Taken together these observations suggest that H tones may show more variable alignment, while L tones show more variable scaling. The Romani data support both observations indicating that it is important to consider whether one is dealing with a L or H tone when assessing variability.

In addition to the above, recent evidence indicates that tonal events may be manifested by a variety of means, including not just F_0 changes but also differences in the duration, amplitude, or quality of the segments involved (e.g., Arvaniti et al., 2016; Niebuhr, 2008, 2012; see also Cole & Shattuck-Hufnagel, 2016, and references therein). This in turn suggests that some cues to tonal events may be redundant and thus not present at all times. For example, the LH* accent of Romani used to mark narrow focus is typically realized with a rise from a low F_0 point and a peak within the accented vowel (see Section 4.2.1). At the same time, however, syllables associated with a LH* accent are typically longer and louder, cues that in context can be sufficient for the correct identification of the accent even if it lacks the rise from a low point or shows peak alignment later than expected. Though this is a topic that requires much work, it is worth bearing in mind when considering variability that not all instances of every tonal event will exhibit all possible traits associated with that event and that sometimes non-F_0 cues may be the only ones present.

To sum up, the positing of phonological contrasts in the Romani intonational system was based on basic principles of phonological analysis, the consideration of both meaning and form, and the expectation that the realization of the posited contrastive elements would show lawful variability. Linguistic sources of variability taken into account included the phrasal position of tonal events, their interaction with the segmental, tonal, and metrical context, and language-specific preferences in resolving tonal crowding. Differences between L and H tones, the possibility of redundant cues, and speaker- and style-specific differences were also considered. Crucially, the weight attributed to these factors hinged on the role of meaning. Tonal elements were posited as contrastive if differences in form were shown to operate in discourse in a way that reflected pragmatic differences, such as the presence of focus, distinctions between given and new information, and the pragmatic function of utterances in discourse (cf. Arvaniti, 2011; Pierrehumbert & Hirschberg, 1990). Given that the data were either part of QUIS, which was specifically designed to probe matters of information structure, or came from natural conversations and story-telling, in which it was possible to establish pragmatic meaning from the context and interlocutors' reactions, this practice served analysis well. Analytical decisions were revised in light of new data, particularly the semi-controlled data from QUIS, and were verified again by examining whether they remained adequate when additional spontaneous data were considered.

4 Illustrations

4.1 Stress

A first step to any analysis is to determine the prosodic type of the system under examination. Existing analyses of the same or related varieties can be of help. However, previous analyses should not be the only source of information, as even varieties of the same language may belong to different prosodic types (cf. Jun & Fletcher, 2014): e.g., Gussenhoven (2004, pp. 228–252) discusses varieties of Central Franconian which encode a tonal contrast absent in mainstream varieties of the German-Dutch dialectal continuum; Hualde et al. (2002) and Kim and Jun (2009) also describe varieties of Basque and Korean, respectively, that are tonal unlike the standard varieties of these languages.

Existing analyses report that Romani is a language with fixed stress on the ultima (Matras, 2002). Auditorily, this appears to apply to most of the Thrace Romani vocabulary as well. The present corpus further suggests that stressed vowels are longer and louder than unstressed vowels though quality differences are small (Adamou & Arvaniti, 2014). Loans, however, have introduced variation in stress location. For instance, when words are borrowed from Turkish, they often acquire Romani morphology; the addition of suffixes in particular leads to stress shifting to the penult; e.g., a Turkish word like *pembe* 'pink' when used with a feminine noun acquires a feminine suffix *–a* yielding /pemˈbea/ 'pink.F' with penultimate stress. Such examples are not uncommon (Adamou & Arvaniti, 2014).

Crucially, in Romani declaratives with broad focus all words show a pitch rise or high pitch on the syllable perceived as stressed; this applies whether the utterance presents new or given information. This is illustrated in **Figure 2**: as can be seen, even function words like the preposition [kaj] and the classifier [taˈneja] show a pitch rise; such pitch rises on function words are a common occurrence in the corpus. Data like these could lead

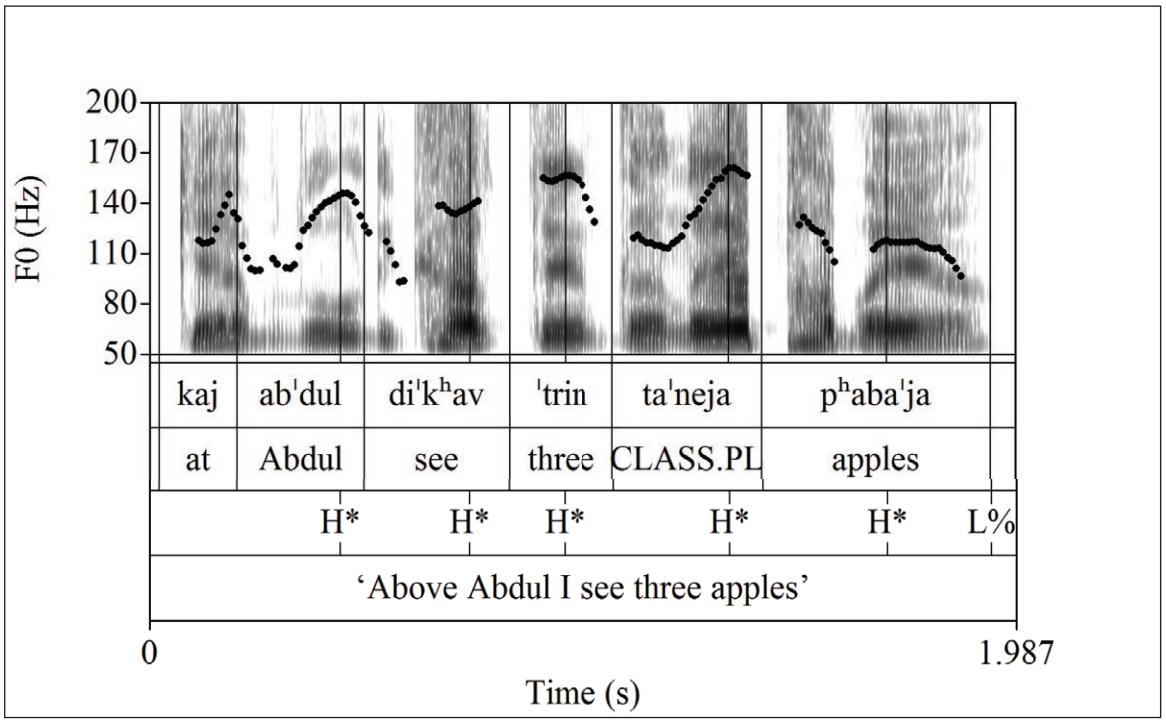

Figure 2: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of "all new" utterance from QUIS, illustrating the use of accentuation even on function words such as [kaj] 'at' and [taˈneja], a classifier. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav2.

to the conclusion that this variety of Romani has a lexical pitch accent system in which one syllable per word carries a rising melody, or that high or rising F_0 is a feature of stress.

The connection between stress and high or rising pitch is still well accepted thanks to early work on the topic (Fry, 1958) and despite plenty of subsequent research clarifying the relationship between stress and intonation (e.g., Beckman, 1986; Beckman & Edwards, 1994). Here the standard view of the autosegmental-metrical (AM) framework of intonational phonology is adopted, namely that stress is independent of changes in pitch related to intonation (Arvaniti, 2011; Ladd, 2008a, pp. 49–55). The connection between the two is indirect: stress is determined by metrical structure; in turn, stressed (metrically prominent) syllables are licensed for association with a pitch accent but need not always be accented.

Data like those in **Figure 2** demonstrate why it is crucial to examine not only declaratives that present new information, as is customarily the case, but other types of utterances as well, including questions and declaratives with early narrow focus (cf. Jun & Fletcher, 2014). Doing so allows us to separate the effects of stress from those of intonation. Evidence from questions and narrow focus utterances makes it clear that pitch rises in Romani are not an exponent of stress or a lexical property of words, but an independent phenomenon. Sentences with early narrow focus show that pitch rises are not present postfocally. This is seen in **Figure 3** in which only the negative particle [naj] is accented, while content words [maj'muna] 'monkey' and [a'ia] 'bear' show falling and flat F_0, respectively. The same applies to *wh*-questions, like that in **Figure 4**: there is only one marked pitch movement, that on the *wh*-word [so] 'what', after which F_0 drops until the end of the utterance. Utterances like these clearly show that typologically Thrace Romani is a linguistic variety which has stress and uses pitch primarily to encode intonational differences.

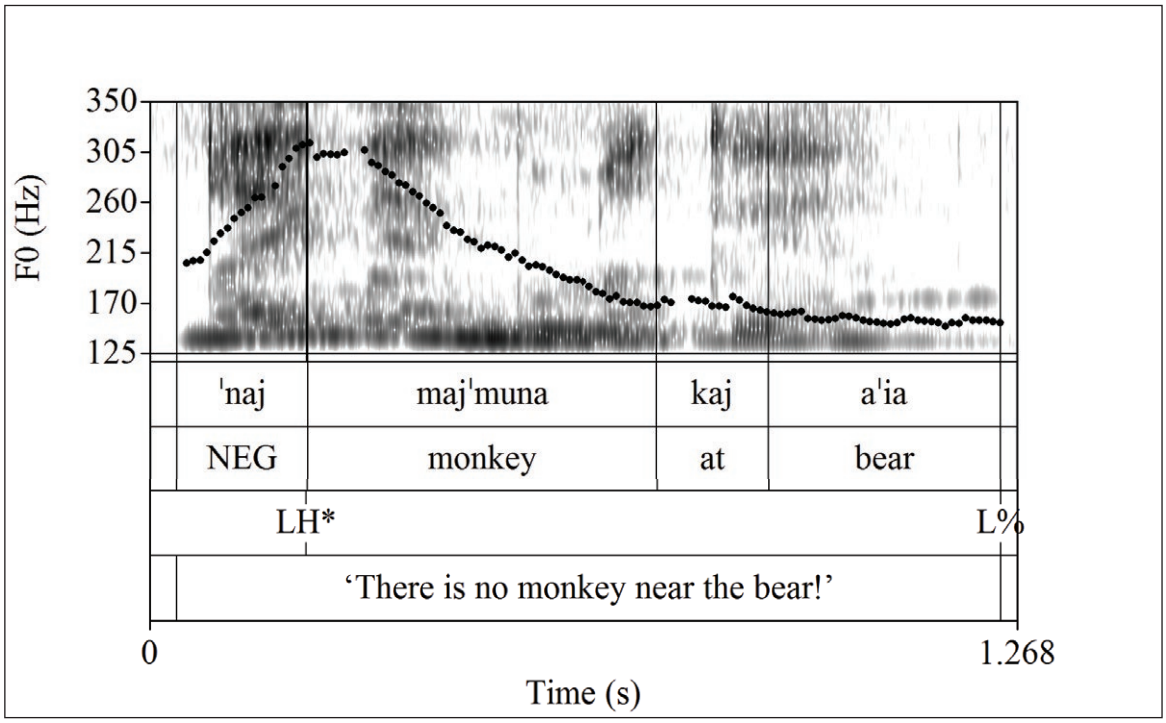

Figure 3: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from a QUIS game, illustrating the use of low (flat or falling) F_0 on stressed syllables. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav3.

4.2 Tonal inventory

The above discussion of stress indicates that pitch modulation in Romani should be treated as postlexical, i.e., as intonation. The next step then is to determine the number and nature of tonal events — pitch accents and edge tones — and their use in the system.

4.2.1 High pitch accents

The discussion of stress clearly showed that stressed syllables are often, though not always, realized with rising or high pitch. Rising and high pitch are interpreted here as reflexes of a H* pitch accent. The corpus indicates that the H* accent can take several forms: sometimes it shows a rise from a low point, while at other times it is manifested as high F_0, a plateau, or a fall. These different realizations can be seen in **Figures 2, 4–8, 10–12**, **14, 16,** and **17**. Most of the observed variation in the realization of H* can be explained by context. The first accent in an utterance is usually realized with a substantial rise and delayed peak (see **Figures 2, 5,** and **8**). Most subsequent accents do not exhibit either of these characteristics except in careful speech: compare **Figure 2** from QUIS and **Figure 5** from a spontaneous and rather animated conversation. On the other hand, the accentual rise may be barely present if the utterance starts with a voiceless consonant; this is shown in **Figure 4** in which the H* accent is on [so] 'what'.

Unlike prenuclear H* accents, which often show a rise, utterance-final words typically show a fall that starts on the stressed syllable. This is illustrated in **Figure 6** (a similar final accent can be seen in **Figure 10** on [merdeˈfea] 'ladder'). **Figure 6** includes three accents: the first is realized as a rise that spans the entire accented vowel, the second as high F_0, and the last one as a fall throughout the stressed vowel of [loˈle] 'red'. The difference appears to be context-related with the final accent being realized as a fall under

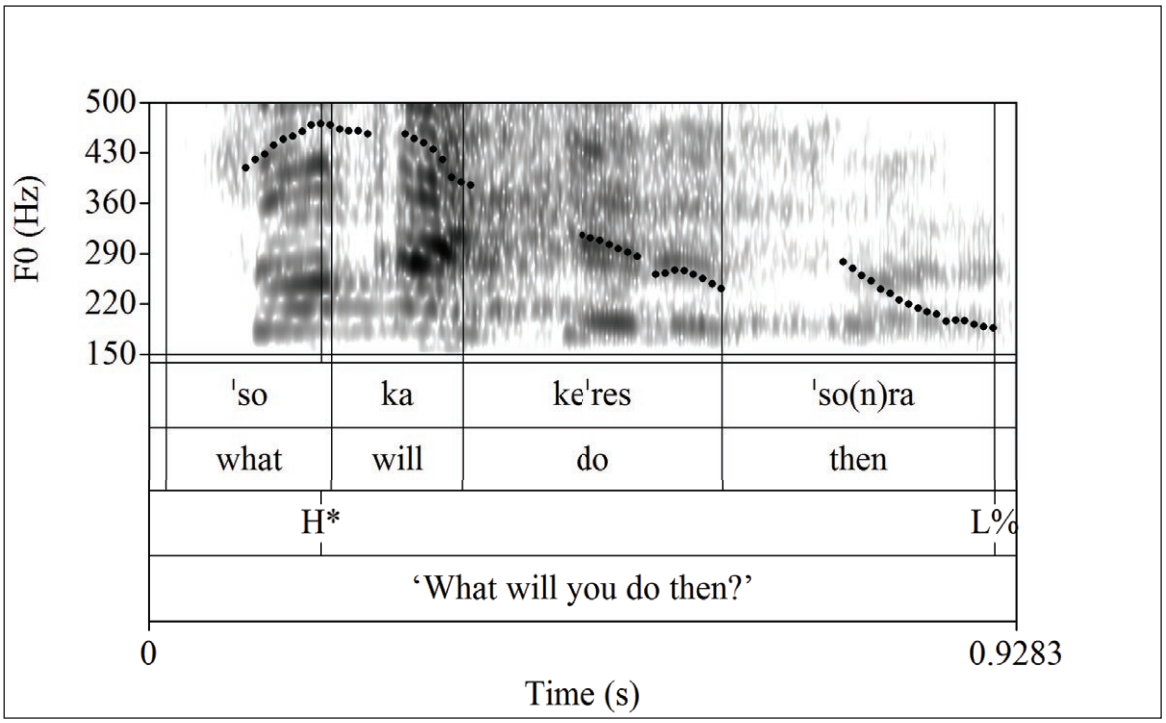

Figure 4: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of a *wh*-question from spontaneous conversation, illustrating the lack of F_0 rises on content words after the *wh*-word [so] 'what'. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav4.

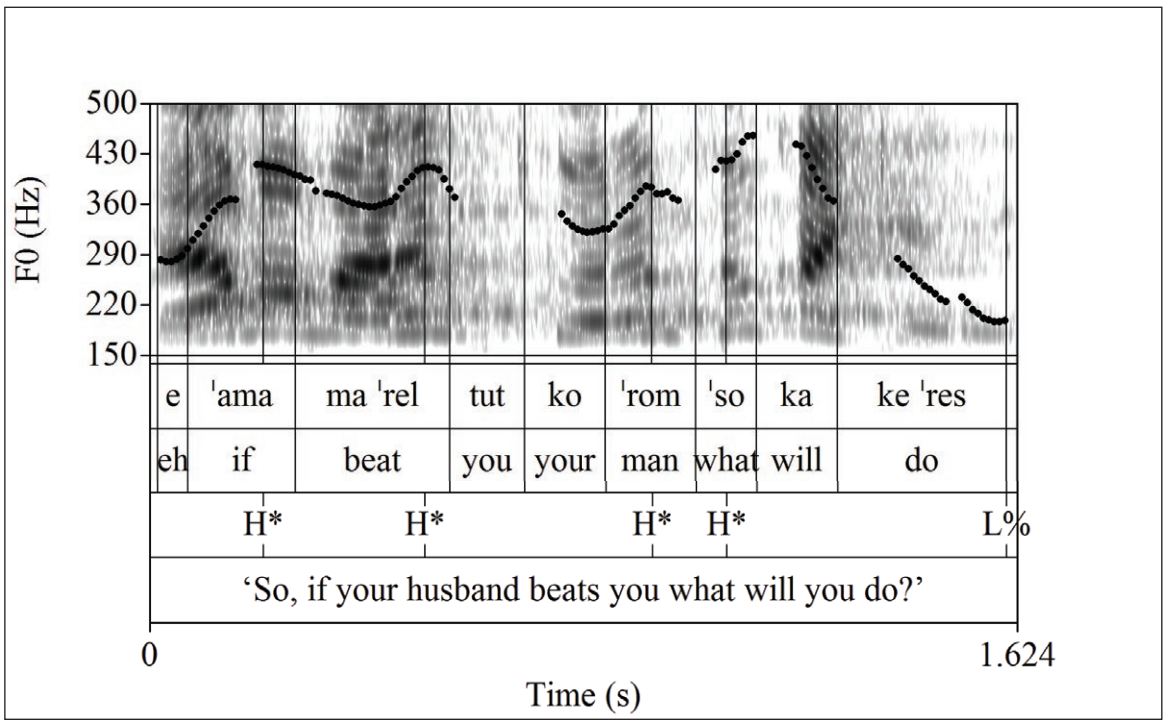

Figure 5: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of a hypothetical followed by a *wh*-question; data from spontaneous conversation, illustrating the variable realization of the H* pitch accent. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav5.

pressure from the upcoming L% (on edge tones; see Section 4.2.3.). On the other hand, when tonal crowding is reduced, as in **Figure 7** where the last word has antepenultimate stress, the H* accent may be realized as high F_0 instead. Given that the differences in realization can be explained by context and the location of stress, they do not warrant a phonological distinction: there is no evidence that final accents in sentences like those in **Figures 5, 6, 7,** or **10** serve any different purpose than prenuclear accents. As noted earlier, in sentences encoding "all new" or given information, all words are accented, suggesting that the main function of the accents is to highlight stressed syllables (Arvaniti & Adamou, 2011; cf. Calhoun, 2010). A non-exhaustive presentation of the variation of H* is given in **Table 1**.

There are, however, realizations of high accents in Romani which indicate that not all can be represented as H*. In utterance-final position, one can observe a difference between accents in broad focus utterances, which show the flat or falling F_0 discussed above, and accents with a marked dip and a rise-fall contour as on the word [aˈraχni] 'spider' in **Figure 8**. This accent is represented as LH*. The two accent types serve distinct purposes: LH* is used to mark narrow focus in declaratives (cf. Pierrehumbert & Hirschberg, 1990). The same accent is also found in early narrow focus, as in **Figure 9**: here F_0 lowers from the onset of the second phrase to the onset of the stressed syllable of [tʃalaˈvel] before rising to a peak within this syllable; following words are unaccented (Arvaniti & Adamou, 2011). The realization of this accent can be juxtaposed to the H* accent on [fuˈlel] 'descend' in **Figure 10** which encodes new information: though this accent also shows a significant pitch rise (being phrase-initial), its range is reduced relative to LH* (the speaker is the same in **Figures 9** and **10**); the peak is aligned with the end of the accented syllable, while the following noun [merdeˈfea] 'ladder' is also accented. The difference between LH* and H* in focal position can also be seen in **Figure 11** which

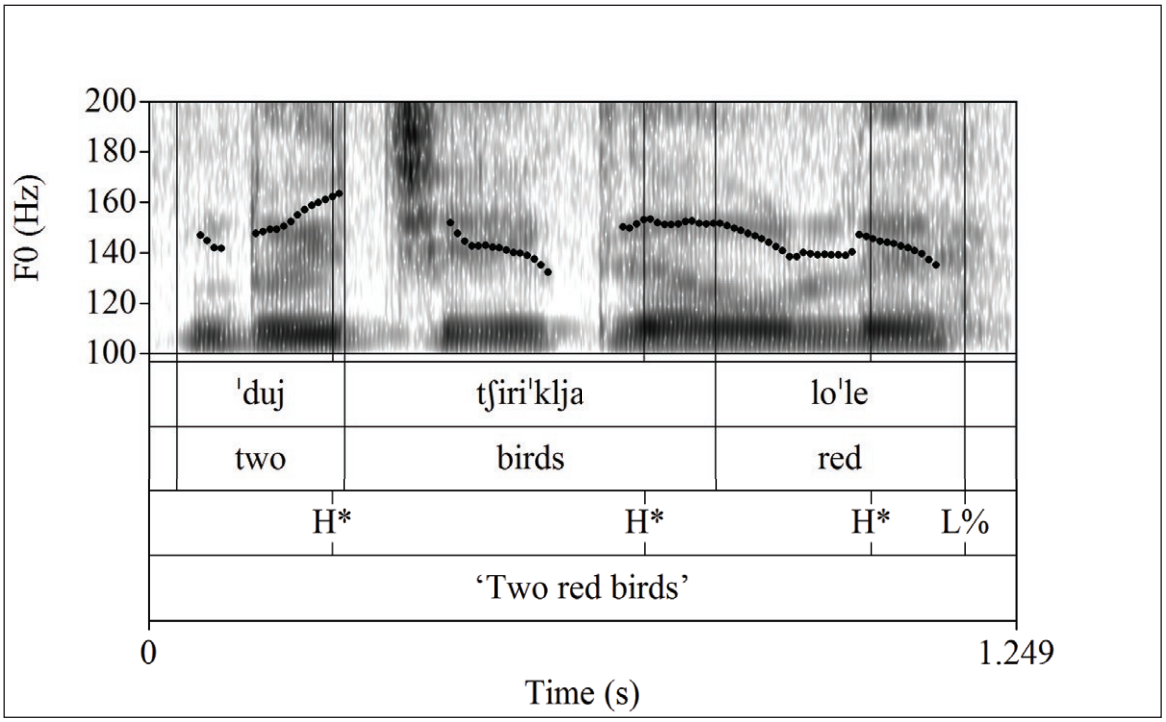

Figure 6: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of a broad focus utterance from QUIS, illustrating the realization of H* in different contexts, including in absolutely utterance-final position. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav6.

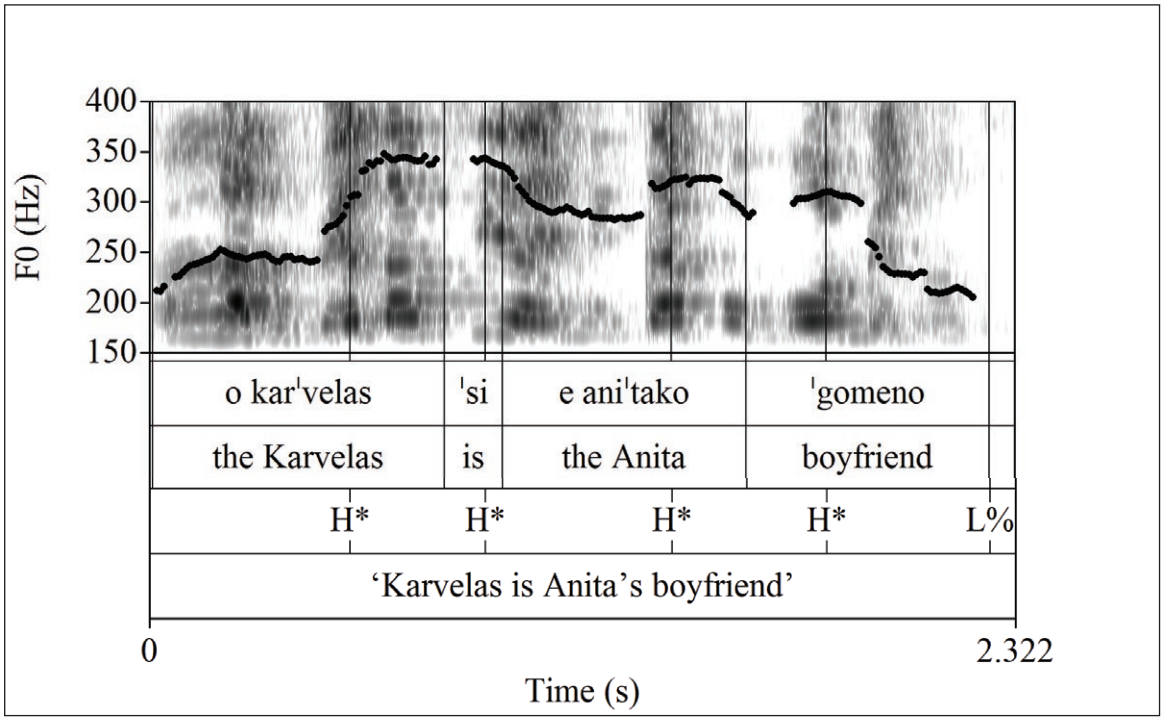

Figure 7: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of a broad focus utterance from spontaneous conversation, illustrating the realization of H* in different contexts, including in nuclear but not absolutely utterance-final position. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav7.

Context	Realization	Illustration
Utterance initially		
On first syllable		Figures 4, 6
On non-initial syllable		Figures 2, 5, 7
Utterance medially		
Adjacent to other accent		Figure 2
Non-adjacent to other accent		Figures 5, 6, 7
Utterance finally		
Final stress		Figures 2, 6
Non-final stress		Figures 7, 10

Table 1: Schematized F_0 contour (continuous line) depicting context-dependent realizations of the H* pitch accent on the target syllable (grey box) and, where applicable, on neighboring syllables (white boxes). This is not an exhaustive list of possible H* realizations in Romani.

includes a short exchange during a QUIS game revolving around the word [aˈia] 'bear': the first token is contrastive and bears a LH* accent, while the second is the interlocutor's confirmation and thus given information and bearing H* instead; a difference in overall scaling between the two accents, in addition to shape, is obvious here too. Given the above, the difference between H* and LH* cannot be attributed to a simple expansion of pitch range as has been advocated by Ladd for English (e.g., Ladd & Morton, 1997); pitch level and span are both involved, but there are also differences in alignment. Further, the low F_0 at the onset of the accented syllable is systematic for LH*, and can be the outcome of a drop in F_0 (**Figure 8**), a low stretch (**Figure 9**), or a combination of the two depending on context. The rise, especially in final position in a long utterance, can be small: it is just sufficient for the accent to sound high rather than falling and in this position it gives the accent its characteristic rise-fall shape (**Figures 8** and **10**). In short, an analysis with only one accent, whether this is represented as H* or LH*, does not appear to be satisfactory for Romani.

The above data beg the question: why not posit instead that Romani has a H*L accent in utterance final position and a LH* accent elsewhere, realized with an optional L component that is present particularly when the accent is used in corrective or contrastive contexts? This analysis would be phonetically transparent and faithful to the most frequent realizations of the two accents. The reason why such a solution is not adopted has to do with the function of the accents within the Romani intonational system. If the above analysis were adopted, Romani would be said to have a LH* accent that has a variety of functions: it is used both to mark new information and for metrical purposes but can also mark narrow focus when needed. The H*L is also used to mark new information (but only at the end of utterances) and can also serve metrical purposes. Relying on the role of the accents in the system makes it clear that this analysis is not optimal as it posits two

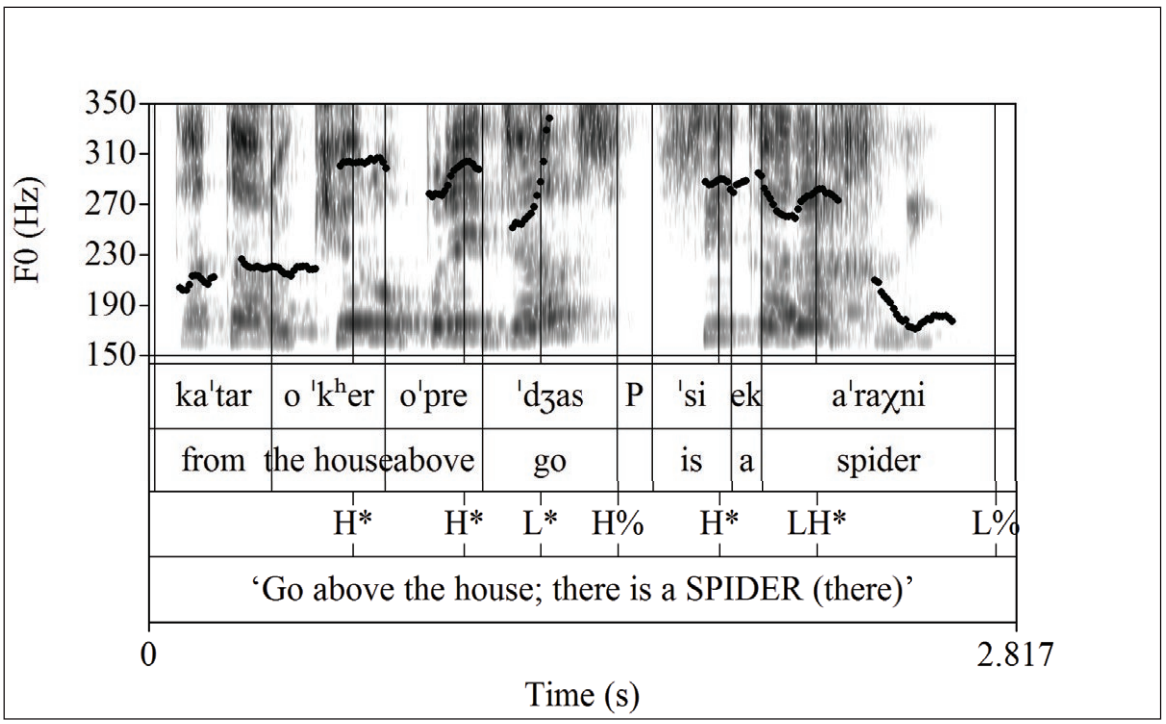

Figure 8: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from a QUIS map-task with narrow contrastive focus on the final word. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav8.

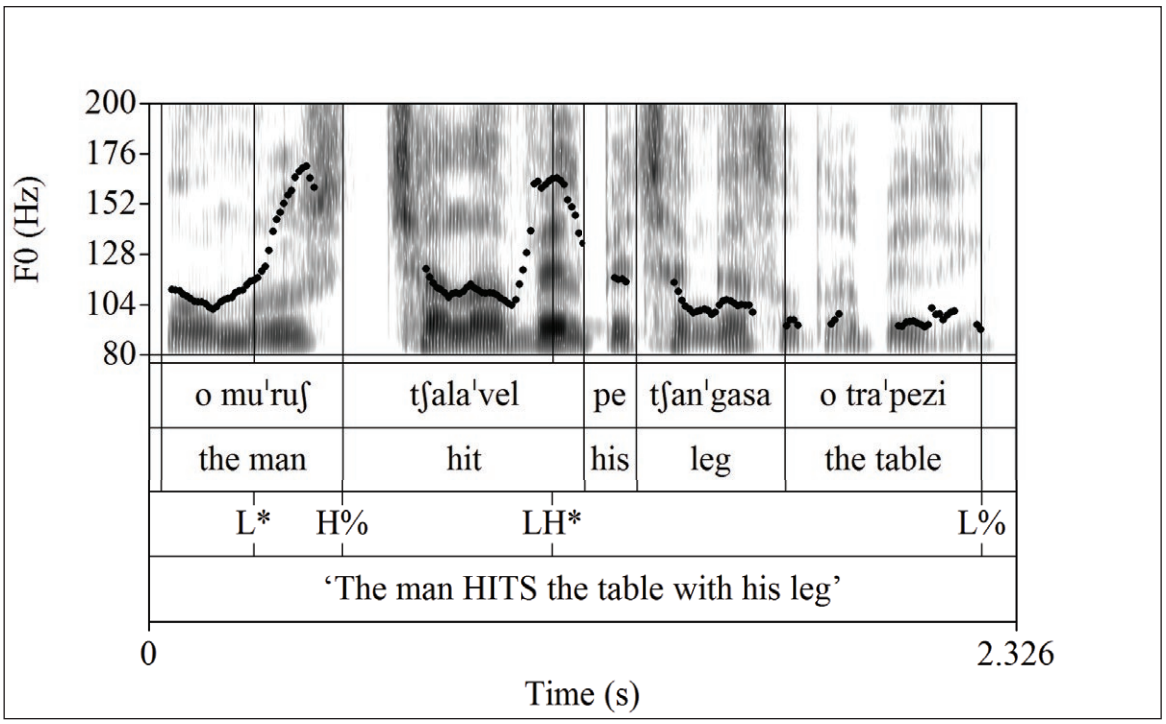

Figure 9: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from QUIS elicitation with narrow contrastive focus on the verb. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav9.

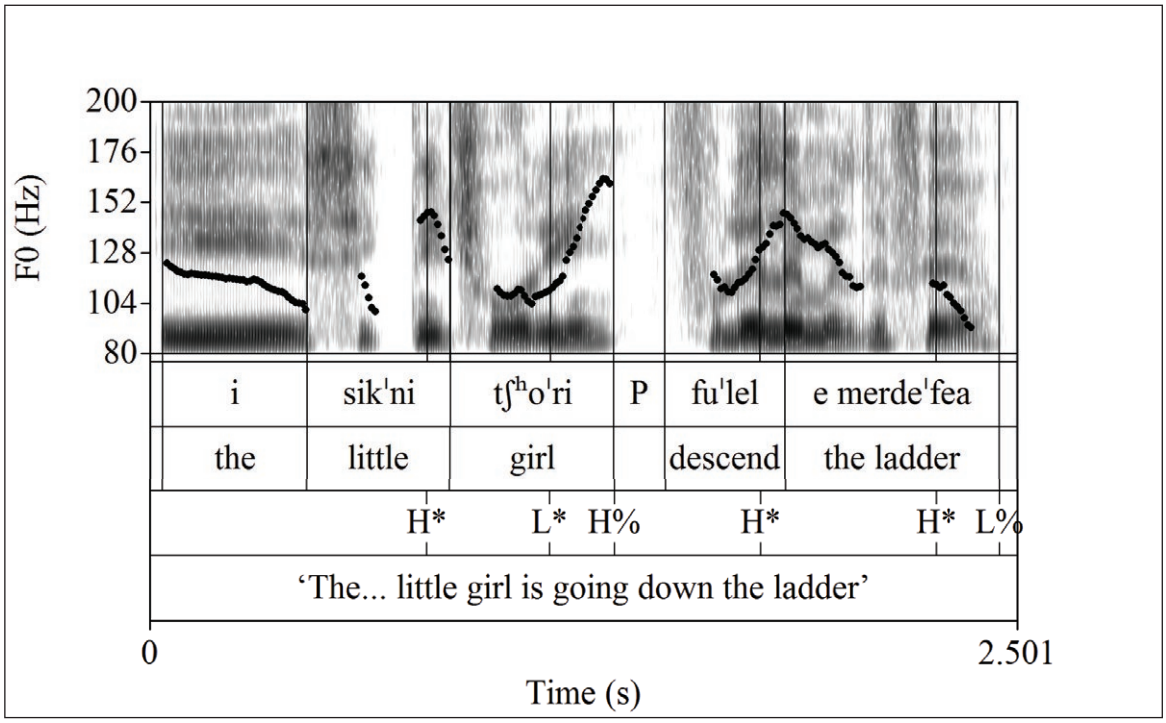

Figure 10: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of a broad focus utterance from QUIS elicitation encoding “all new” information. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav10.

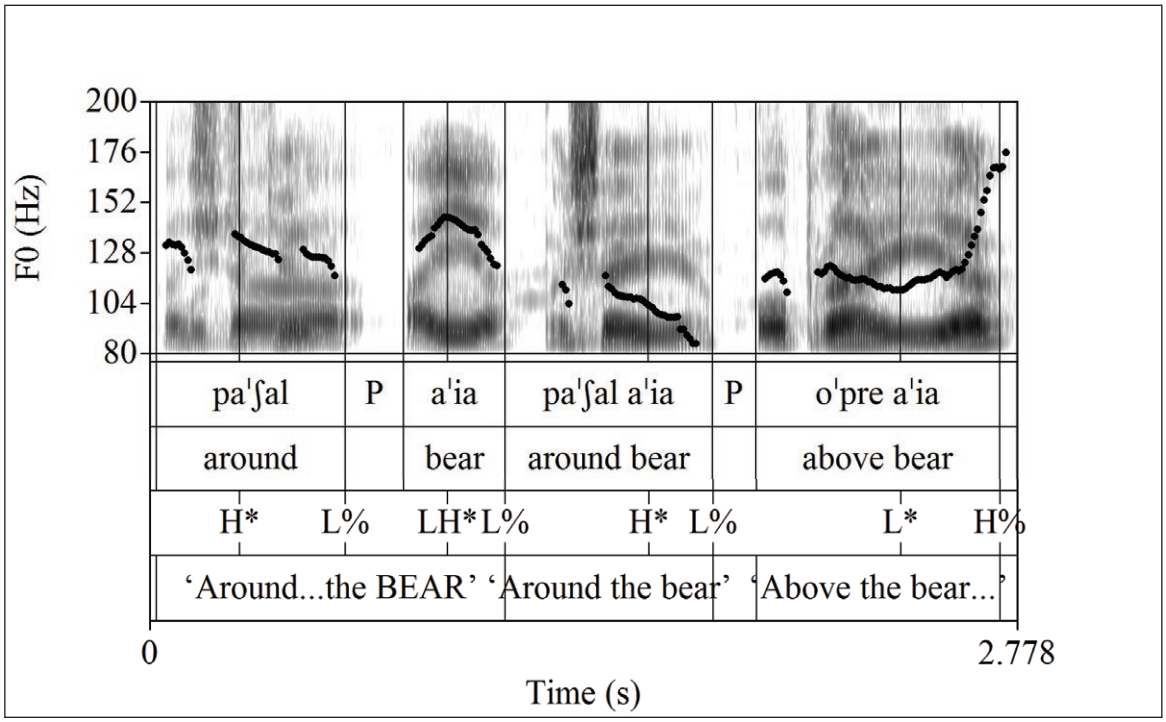

Figure 11: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from a QUIS game illustrating the differences between LH*, H*, and L* accents on the same word [aˈia] ‘bear’. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav11.

accents on the basis of form but with mixed functions. Further, this analysis assumes that the ubiquitous F_0 fall at the end of declaratives is due sometimes to a L edge tone (after LH*) and sometimes to the accent itself (after H*L). The consequences of particular pitch accent choices for the analysis of edge tones are discussed in more detail in Section 4.2.3.

4.2.2 Low pitch accents

Romani also shows low pitch accents. These appear in two main environments in the corpus: before a continuation rise and in polar questions. Low accents in both cases are represented as L* and serve to highlight the word in focus in environments indicating non-completion.

A canonical instantiation of L* is seen in **Figure 12** on [ˈmatʃka] 'cat' which has low, flat F_0 on its stressed syllable. **Figure 13** exemplifies the realization of L* in a polar question. Like the H* accent, L* exhibits realization variability. In general, L* is realized as a dip or low-F_0 stretch that is more pronounced and longer compared to the dip of rising accents discussed in Section 4.2.1.; as a result, the accent sounds low not high. The difference in the extent of the low F_0 stretch is illustrated in **Figure 11** which includes a LH*, a H* and a L* accent on the word [aˈia] 'bear'. Low F_0, however, is often realized on the syllable *preceding* the one with stress, while the stressed syllable itself is low but rising. This happens particularly if the stressed syllable is phrase-final; this can be observed in the first phrase in **Figures 9** and **10**, where [muˈruʃ] 'man' and [tʃʰoˈri] 'girl', respectively, show a deliberate dip on their first (unstressed) syllable. It is only when the stressed syllable is further from the boundary tone that the L* is fully realized, as in **Figures 11** and **12**. However, the preponderance of final stress means that such realizations of L* are not very frequent.

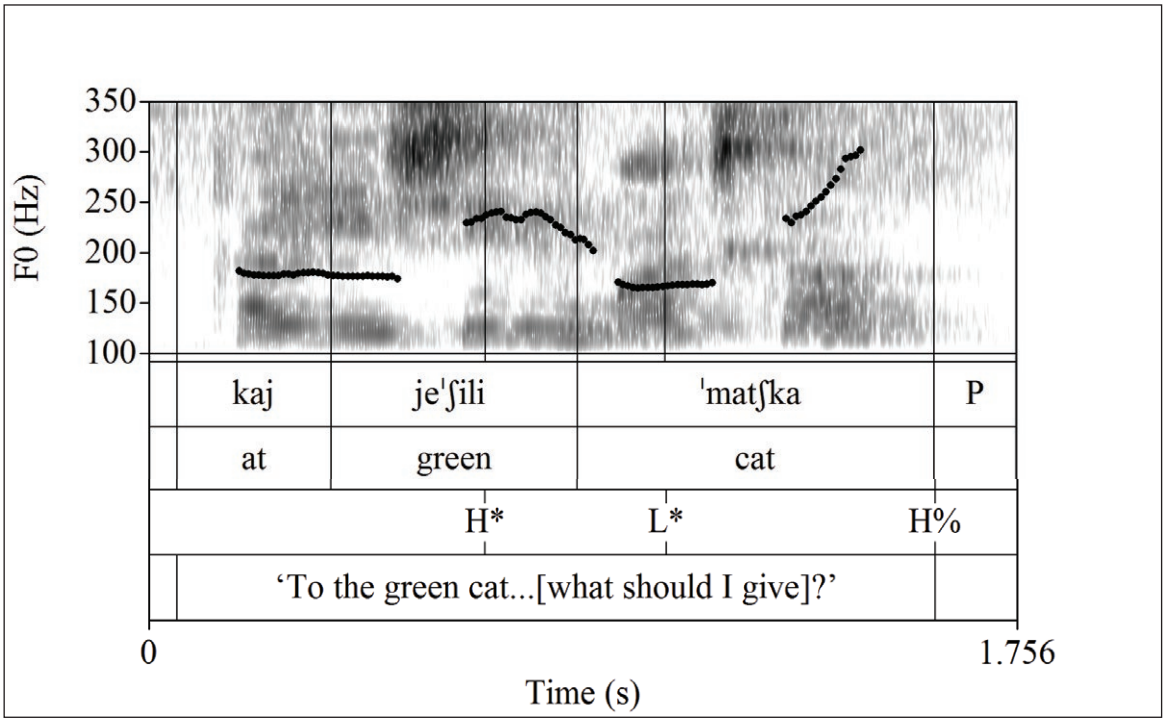

Figure 12: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from a QUIS game illustrating the realization of L* in the absence of tonal crowding. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav12.

The dip in F_0 reflecting a L* pitch accent can be less pronounced in the case of polar questions; e.g., in **Figure 13**, [iˈklan] starts low but F_0 rises smoothly afterwards. The fact that the dip in questions like that in **Figure 13** is the reflex of a L* is supported by utterances like that in **Figure 14**: the melody here shows a low F_0 stretch on the last vowel of [laˈtʃʰo] 'nice' which is the focus of the question and is followed by the HL% boundary tone typical of polar questions (see Section 4.2.3.). While the boundary L is undershot in this instance, due to tonal crowding, it is clear that the F_0 dip associated with the L* focal accent is considered essential for the melody and thus fully realized by elongating the last vowel of [laˈtʃʰo] 'nice'.

The decision to analyze these accents as L* when most instances are characterized by undershoot due to extensive coarticulation with upcoming H tones may be met with skepticism. Could it be, for example, that questions use the same LH* accent as statements to indicate narrow focus? Why not use H* for the accent in continuation rises, since F_0 is often rising on accented syllables? The answers to these questions lie in basic principles for distinguishing L and H accents, system internal considerations, and analytical coherence (Gussenhoven, 2007).

First, L* accents in both continuation rises and polar questions show scaling that is low relative to the speaker's range and other accents, as shown in **Figures 12–14**. Second, polar questions in particular always end in a L edge tone (see **Figures 13** and **14**); if so, then analysing their melody as LH* L% would make them identical to narrow-focused statements. This is patently false, however, and this stands to reason: speakers should wish to differentiate statements from questions. The difference has primarily to do with the shape of the pitch rise and the location of the peak. In narrow focus statements, the rise and fall are symmetrical; the rise is convex in shape and the peak is typically reached on the accented syllable, after which F_0 begins to fall. In polar-questions, the contour starts with a low F_0 stretch, while the rise is concave and followed by a fall of relatively short duration. This difference in shape is illustrated in **Figure 15** which shows the contour of the word [aˈia] 'bear' with LH* (from **Figure 11**) and as a polar question (from a different speaker with different pitch range but almost identical duration). The differences support the observation above that the H tone of the LHL sequence in polar questions occurs close to the end of the question (see **Figure 13**). Due to the limited variation in stress location in Romani, it is not clear if the phrase-final, phrase-penultimate, or last *stressed* syllable is the docking site of this H tone (though **Figure 13** and other similar examples suggest it is the last stressed syllable, as in Greek; Arvaniti et al., 2006b; Grice et al., 2000). Despite this uncertainty, it is clear that the H is not aligned with the accented syllable of the word in focus unless this word is phrase-final. This suggests that the H tone is less likely to be part of the pitch accent itself and more likely to be part of an edge tone. These considerations lead to the overall analysis of the polar question melody as L* HL%. The only alternative analysis involving a LH* accent would be to represent the polar question melody as LH* HL%. But this would entail the presence of a plateau between the two H tones; this is not attested, however, although plateaus are frequent in Romani (see Section 4.2.4.)

A viable alternative would be to represent the accent of polar questions as L*H and the entire melody as L*H L%, accepting that the accentual H is aligned independently of the L* tone (see Gussenhoven, 2007, for arguments that pitch accent tones need not be bound to each other). If so, then the posited L* used in continuation rises could be seen as the flip side of H*, an accent used primarily for metrical purposes; as such, this accent can be elided or severely undershot. Its presence is required simply to create a perceptual contrast with the upcoming H% boundary tone (cf. Gussenhoven, 2007). A similar reversal of polarity is reported for Greek by Baltazani and Jun (1999). At present it is not possible to determine which analysis is optimal. This is due both to the fact that

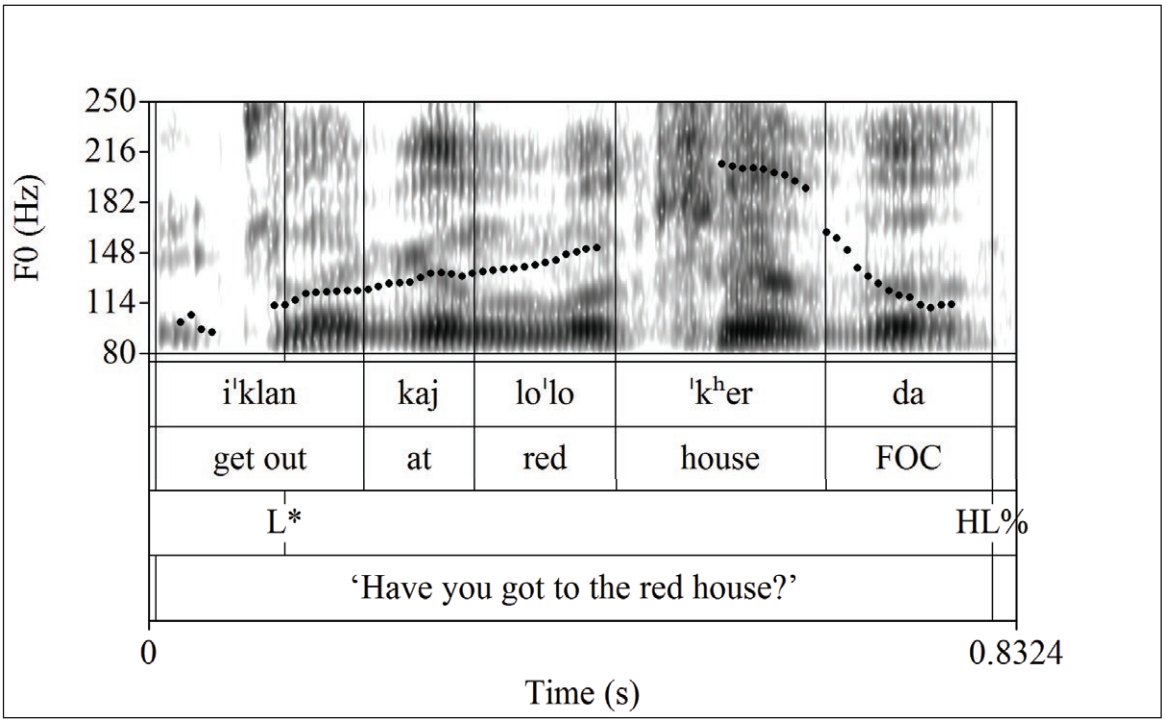

Figure 13: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from a QUIS game illustrating the polar question melody of Romani when no tonal crowding is involved. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav13.

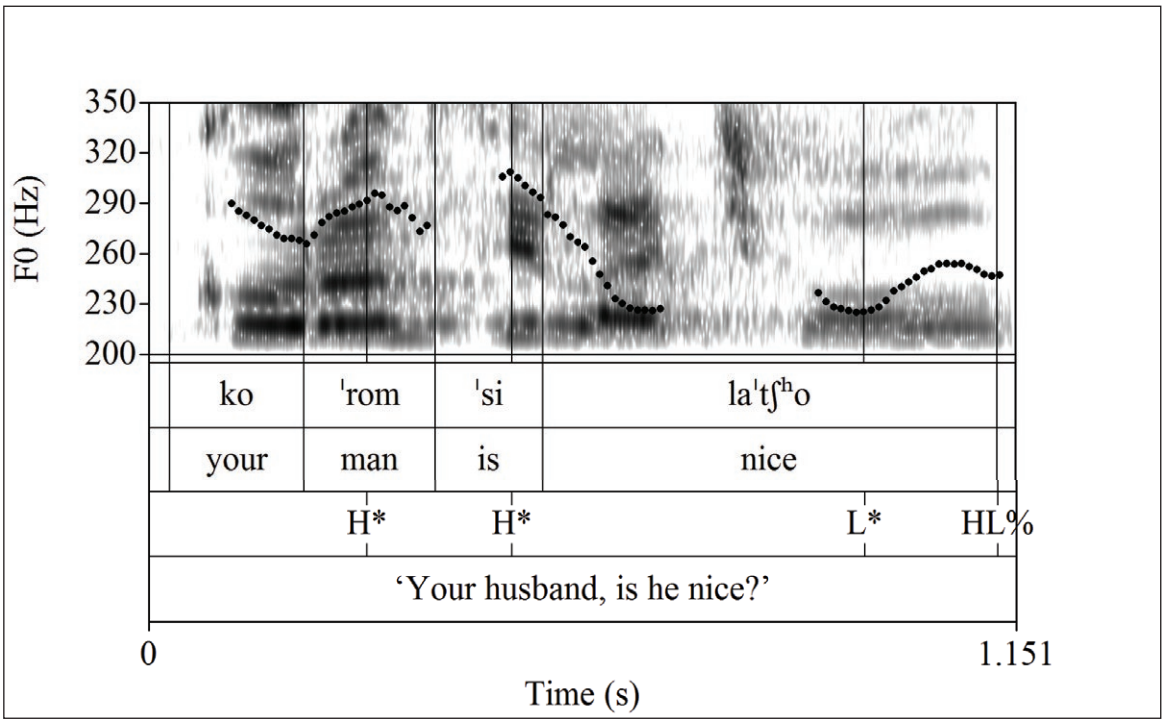

Figure 14: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from spontaneous conversation illustrating the polar question melody of Romani in the presence of tonal crowding. Click on the figure to listen to the sound file. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav14.

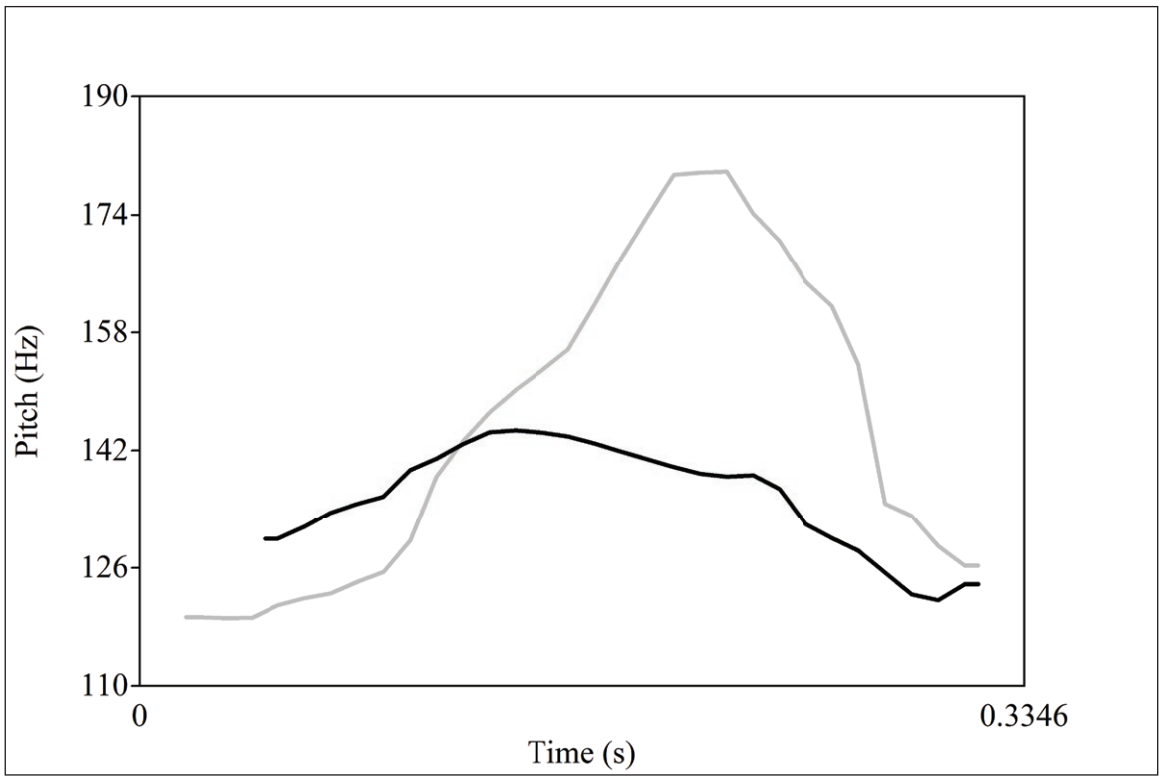

Figure 15: F_0 contours (in Hz) of the word [a'ia] 'bear' with L* HL% (gray line) and LH* L% (black line). This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav15a (gray line) and http://dx.doi.org/10.5334/labphon.14.wav15b (black line).

the corpus contains relatively few instances of polar questions and continuation rises and because we do not as yet have strict criteria in intonational research to assess alternative analyses (but see Ritter and Grice, 2015, and Gussenhoven, 2016). I return to this point in Section 5.3.

4.2.3 Edge tones and phrasing

Edge tones are often discussed together with phrasing. Following Pierrehumbert (1980), many analyses have adopted two types of edge tones, phrase accents and boundary tones (e.g., L- and L% respectively). Since Beckman and Pierrehumbert (1986), these two types of edge tones have been linked to distinct levels of phrasing, the intermediate phrase (ip) for phrase accents and the Intonational Phrase (IP) for boundary tones. Evidence in favour of the independence of phrase accents and boundary tones in such configurations has been reported *inter alia* in Arvaniti et al. (2006b) for the Greek L+H-, Barnes et al. (2006) for the English L-, and Arvaniti and Ladd (2009) for the Greek L-. On the other hand, the need for two levels of phrasing has been hotly disputed by some (e.g., Gussenhoven, 2004, pp. 316–319; Ladd, 1983; see Ladd, 2008a, pp. 142–147 for a discussion). Nevertheless, combinations of tones are clearly needed for observational adequacy, independently of whether one adopts the notion of a phrase accent and relates it to the presence of two levels of phrasing. For instance, the presence of two H tones each with its own target accounts for final rises in English which show a step-up from one high pitch level to the next (Brugos et al., 2006; Pierrehumbert, 1980). Similarly, Ritchart and Arvaniti (2014) analyze Southern California uptalk as L* L-H%; the L-H% edge tone configuration accounts for the late onset and low scaling of the uptalk rise as compared to rises in questions, which are analyzed as L* H-H%. Independently of whether one assumes that one of

these tones is a phrase accent and the other a boundary tone each demarcating a different phrasal constituent, it is clear that both are needed to adequately represent this difference between questions and statements with uptalk.

In order to determine whether a language has one or two levels of phrasing, Jun and Fletcher (2014) propose that one uses disambiguation (of the *Mary is not drinking because she is unhappy* type) or increasingly longer utterances in which "weight" is added to specific constituents. The assumption is that these manipulations will break down long utterances into shorter phrases. One can then examine if these shorter phrases are comparable to longer ones or present their own characteristics.[4] A somewhat different approach is adopted by Arvaniti and Baltazani (2005) in the GRToBI analysis of the Greek intonational system. The authors annotated ips and IPs based on (impressionistic) degree of juncture, then compared the two types: they found that phrases annotated as ips had less complex tonal movements (simple rise or fall) and less extreme scaling that those annotated as IPs; e.g., while at the end of IPs Greek speakers reached the bottom of their range, they did not do so at the end of ips ending in L-.

The procedure of Arvaniti and Baltazani (2005) is not easy to use with a diverse corpus like that of Romani. Establishing a speaker's pitch range in the laboratory is much easier than in natural speech, particularly when speakers touch upon sensitive topics, become excited, etc. (see Section 2). The approach of Jun and Fletcher (2014) is more appropriate in such circumstances, if suitable data are available. In the present corpus, however, attempts to elicit such longer utterances resulted in short phrases separated by prolonged pauses; **Figures 8, 9, 10,** and **16** illustrate the types of substantial breaks speakers of Romani used when at most a minor break would be expected. This can be juxtaposed to spontaneous animated speech in which expected breaks are missing; e.g., in **Figure 5** there is no break between the subordinate and main clauses. Thus, although there are perceived differences in strength between some phrasal boundaries, it is not possible to discern systematic differences between them in terms of function, scaling, or tonal configuration. In turn this suggests that one level of phrasing is sufficient for Romani (barring new data). The edge tones include L%, H%, HL%, and LH%. HL% is found at the end of polar questions, as in **Figures 13** and **14**. LH% is attested in *wh*-questions (not illustrated).

A reason why researchers posit two types of edge tones is that phrase accents often fill the gap between the last pitch accent and the end of the utterance or show secondary association (Arvaniti & Ladd, 2009; Barnes et al., 2006; Grice et al., 2000). In Romani there is no evidence for secondary association.[5] Spreading appears to apply only to the L% boundary tone which spreads to the left when focus is early: in such instances, F_0 starts dropping towards the end of the stressed syllable of the accented word and remains low for the remainder of the utterance, though no consistent pattern for the extent of the spread can be discerned (cf. **Figures 3, 4,** and **9**). **Figure 7** shows a different instance of L% spreading: here, F_0 falls immediately after the stressed antepenult of [ˈgomeno] 'boyfriend' so that the last two syllables in the utterance are both low in pitch.

Finally, it is worth noting that the present analysis of edge tones follows the established practice of separating final (nuclear) pitch movements into a pitch accent and following

[4] A similar result can be achieved using Rapid Prosody Transcription, a system in which boundaries are marked by lay participants; the results can be used to examine the acoustic parameters associated with high boundary scores (Cole & Shattuck-Hufnagel, 2016).

[5] It is possible that secondary association is needed to account for the realization of the LH% boundary tone attested with *wh*-questions. The corpus, however, contains too few questions with this pattern to allow for further analysis.

edge tones. Thus "nuclear falls" in Romani declaratives are analyzed as a sequence of a H* pitch accent and a L% boundary tone. This type of analysis goes back to Pierrehumbert (1980, ch. 1) who analyzed English nuclear falls as consisting of a H* pitch accent followed by a L-L% edge tone configuration. This is not the only possibility, however. Gussenhoven (2004, pp. 296–299) analyses the same English nuclear fall as consisting of a H*L pitch accent followed by a L% boundary tone or L_t in his notation (for additional arguments for "off ramp" analyses of English melodies, see Gussenhoven, 2016; see also Peters, Hanssen & Gussenhoven (2015) for "off ramp" analyses of a number of Germanic varieties). A discussion of the two views is beyond the scope of the paper, but it is worth keeping in mind when determining how best to analyze a particular language that any analysis of edge tones hinges on decisions about the accent inventory and vice versa.

4.2.4 The use of plateaux

In the Romani corpus, plateaux are quite frequent (see, e.g., **Figures 2, 3, 6, 7, 8,** and **12**). An utterance in which plateaux are used almost exclusively is shown in **Figure 16**. In some languages differences between peaks and plateaux are meaningful; this applies, e.g., to Neapolitan Italian (D'Imperio, 2000; D'Imperio et al., 2000). In others, like British English, it is clear that plateaux affect estimates of pitch accent scaling but not necessarily accent identity (Knight, 2008). In Romani, however, peaks and plateaux appear to be realizational variants of L and H tones both phrasal and accentual, so that plateaux and glissandos are interchangeable. Compare **Figures 16** and **17**, both showing utterances elicited from the same speaker during a QUIS task. In **Figure 16** the F_0 of almost every syllable is flat, independently of stress and association with a tone (cf. unaccented [ka] in [kaˈfe] and accented [ˈpa] in [ˈpasta]). In **Figure 17**, on the other hand, plateaux and glissandos coexist: the

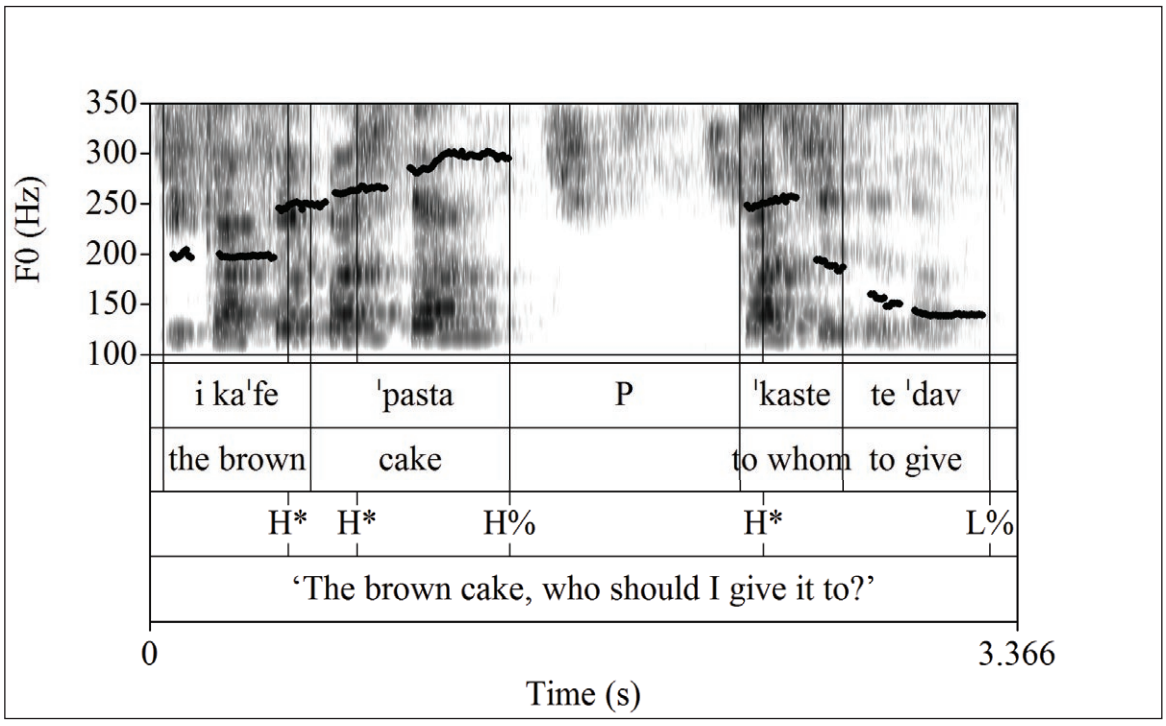

Figure 16: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from QUIS, illustrating the use of plateaux. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav16.

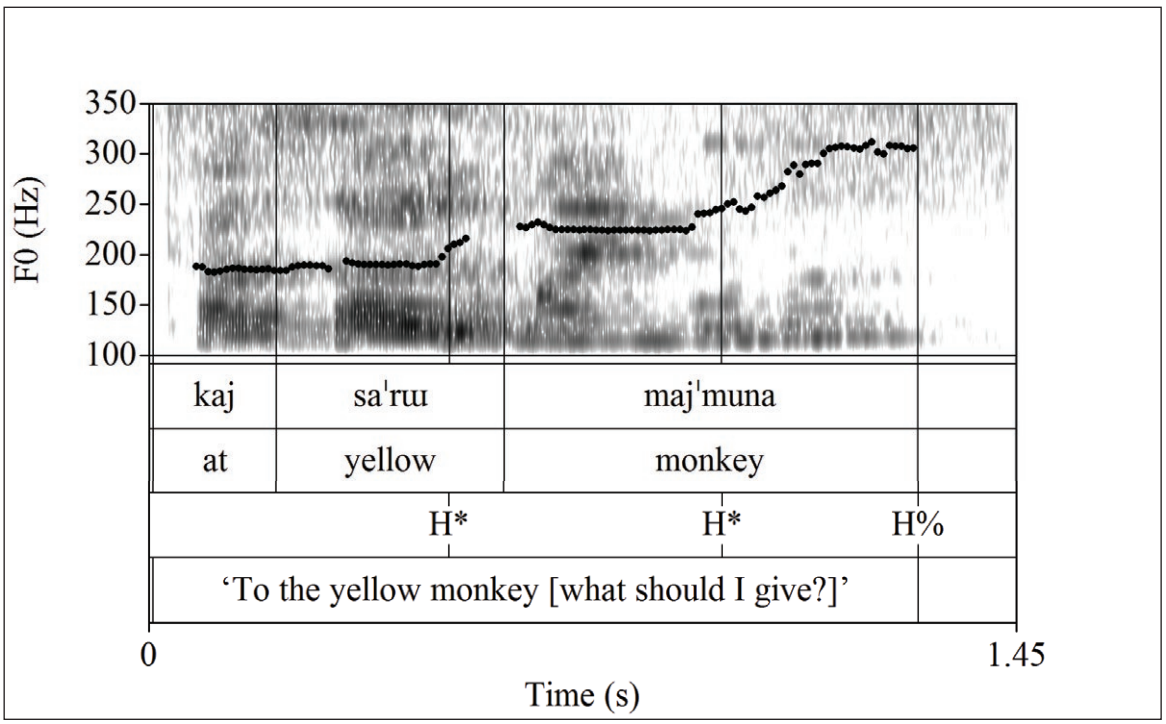

Figure 17: Spectrogram, F_0 contour (in Hz), AM annotation, and gloss of an utterance from QUIS, illustrating the mixing of plateaux and glissandos in the same utterance; the speaker is the same as in Figure 16. This audio content is available at: http://dx.doi.org/10.5334/labphon.14.wav17.

H* accents are realized as rises but the two initial (unstressed) syllables and the final H% are realized as plateaux. The fact that glissandos and plateaux can be used interchangeably and mixed in the same utterance indicates that there is little difference between them in Romani. Plateaux appear to be more frequent in ritualistic and formal speech, such as story-telling and QUIS games, respectively. Realizations of H%, plateaux are frequent as a floor-holding device, particularly when pauses mid-utterance are involved, as in **Figure 16** (this use is akin to plateaux attested in Greek and some varieties of English; cf. Arvaniti & Baltazani, 2005, on Standard Greek; Clopper & Smiljanic, 2011, and Ritchart & Arvaniti, 2014, on U.S. English varieties). Overall, these observations hint at a stylistic rather than a pragmatic difference between plateaux and glissandos in Romani, indicating that the difference need not be part of the phonological representation.

5 Discussion

5.1 Laboratory and spontaneous data

The above presentation of some elements of the Romani prosodic system shows that the principles used here allow for the development of a phonological analysis even when the data present multiple sources of variation. The variety of speech styles included in the corpus allowed for a more robust analysis: variability was present and had to be taken into consideration, while decisions were not based on a uniform (and, for that reason, possibly unrepresentative) dataset as is typical of laboratory studies.

This does not mean that laboratory data are not useful or should be dispreferred in research. Indeed, the analysis presented here serves to show that it is counterproductive to pit laboratory and spontaneous data against each other, considering one or the other

inherently superior or better suited for research (cf. Xu, 2010). Rather, what is proposed and illustrated here is a back and forth between the two: spontaneous data allow one to establish a set of hypotheses about the system under analysis; these can then be tested by means of controlled or semi-controlled data; any changes should be subjected to new scrutiny using spontaneous data and, if necessary, to further revision. Thus, the present work shows that it is possible to use spontaneous and (semi-)controlled data synergistically and that each type can provide answers to particular problems during analysis. Given the importance of meaning advocated here, however, approaching a previously undescribed intonational system using primarily spontaneous data was advantageous, as such data include a wealth of information in terms of both linguistic and pragmatic context that can serve as analytical tools; e.g., new, given, and contrastive information could be tracked from discourse, and linguistic context could provide clues as to the reasons for variation.

5.2 Problems with a level of broad phonetic transcription

The present corpus illustrates issues that can arise when variability clashes with established notions of uniformity in intonational realization. Some of the features discussed here may be more prevalent in speech communities with an oral tradition and no established standard, but once spontaneous data become more common in research, the overall variability observed here is likely to prove comparable to that found in other speech communities. Thus the present corpus can be treated as an extreme example of variability which allows us to sharpen the intonational analysis toolkit. The lessons learned apply to the analysis of all languages, not exclusively to the present data, to Romani in particular, or to non-standardized languages like Romani.

I argue that the analysis presented here was easier to arrive at by not using a level of broad phonetic transcription as is often advocated (e.g., Hualde & Prieto, 2016; Jun, 2005; Jun & Fletcher, 2014). As discussed amply in Ladd (2008a, ch. 3), this approach represents one of the two main views about intonational analysis that played a part in the development of annotation systems, beginning with the original ToBI system for the prosodic annotation of English (Silverman et al., 1992). Specifically, the approach taken here is that advocated by some of the ToBI developers who took the position that the analysis of intonation using autosegmental-metrical representations is phonological in nature and therefore it need not faithfully represent every detail of the pitch contours (Beckman et al., 2005; Ladd, 2008a, p. 111; see also Gussenhoven, 2007, for similar views).

Adopting this position is not meant to denigrate the importance of phonetic detail. The value of phonetic detail in understanding speech has been noted for at least the past 30 years and is constantly affirmed by new evidence (see, *inter alia,* Browman & Goldstein, 1992, on the repercussions of ignoring fine-grained phonetic detail in understanding allophonic variation; Scobbie et al., 2000, on covert contrast; Edwards et al., 2015, on the problems of ignoring phonetic detail in language acquisition). Intonation is no exception to this understanding. Almost a quarter of a century after the original ToBI system, it is undeniable that fine-grained phonetic detail is present in production and crucial for the processing of intonational categories (Barnes et al., 2012; Cangemi & Grice, 2016; Cole & Shattuck-Hufnagel, 2016; D'Imperio, 2000; D'Imperio et al., 2000; Knight, 2008; Knight & Nolan, 2006). Thus the need for more research in this area is indisputable.

However, as advocated by Ladd (2011) for segmentals, investigating the details of phonetic realization neither necessitates recourse to broad phonetic transcriptions nor does it obviate the need for an abstract phonological analysis of intonation. Specifically, in his (2011) paper, Ladd argues in favour of such an abstract level of analysis and against a *systematic phonetic level*, the equivalent of a broad phonetic transcription. As Ladd shows,

a systematic phonetic level is problematic as it converts one symbolic representation into another. Such more fine-grained symbolic representations may capture some details about realization but cannot capture all phonetic detail as shown by a large body of research of the past 30 years. To give but one example, timing is an important aspect of phonetic realization that no symbolic representation can capture, by definition (Port & Leary, 2005).

The problem can be illustrated by first using a segmental example. At the phonological level, it is generally agreed that English has a phoneme /k/. At a systematic phonetic level, several allophones may be recognized, depending on a researcher's emphasis on a particular aspect of realization; e.g., the detailed descriptions of Cruttenden (1994, pp. 138–157) and Ladefoged and Johnson (2011, pp. 57–65) focus in turn on aspiration, place of articulation, and type of release. Based on such descriptions, an aspirated (long-lag VOT) and an unaspirated (short-lag VOT) allophone are typically recognized for /k/, [kʰ] and [k] respectively (cf. Section 3.2.). If emphasis is placed instead on place of articulation, [k], [k̟], and [k̠] allophones may be postulated (cf. Cruttenden, 1994, p. 153). Together VOT length and place of articulation would yield six /k/ allophones (possibly more if aspiration and place of articulation are combined with compatible types of release). However, these allophones (or any other for that matter) would not do justice to the attested variation in the realization of English /k/: VOT varies gradiently based on stress, quality of the following vowel, position in the foot, word, and phrase, and even on dialect (e.g., Cruttenden, 1994, pp. 140–142; Keating, 1984; Stuart-Smith et al., 2015); the exact place of articulation of /k/ is also different for each following vowel (Cruttenden, 1994, p. 153). This means that what is represented in a broad phonetic transcription — the realizations typically referred to as *allophones* — will be incomplete and arbitrary. As Browman and Goldstein (1992, p. 164) note: "many allophonic differences are just quantitative differences that are large enough that phoneticians/phonologists have been able to notice them, and to relate them to distinctive differences in other languages."

By extension, a systematic, broad phonetic representation of intonation can only amount to an arbitrary collection of allotones without capturing the full gamut of variation. This is in fact explicitly noted by Hualde and Prieto (2016) who define broad phonetic transcription as "a form of transcription that includes *a certain amount of redundant, phonologically non-contrastive detail* that is nevertheless a systematic aspect of the language [emphasis added]." *A certain amount* is precisely the problem with such a system, as it is not clear how this amount can be determined (cf. Cangemi & Grice, 2016, for similar arguments). For instance, Hualde and Prieto (2016: Figure 5) use !H% to phonetically transcribe an underlying L% boundary tone which, being undershot due to tonal crowding, is scaled higher than typical (by approximately 20 Hz). However, in that same figure, the H* of the L+H* pitch accent is also undershot, being scaled lower by approximately 20 Hz as well, but this change is not transcribed. Similarly arbitrary decisions could have been made for the Romani data presented here had a level of broad phonetic transcription been used. As Browman and Goldstein (1992) note, attention might have been paid to variants that have been used as distinctive tonal elements in other languages; !H% used by Hualde and Prieto to indicate an undershot L% is such an example.

It is this abritrariness of broad phonetic transcriptions that drives the position that abstract representations are more successfully combined with exemplars, detailed traces of phonetic realization (Pierrehumbert, 2002). As Beckman et al. (2007) have shown, both these levels — which can be loosely equated to phonological and phonetic — play a part in speech production and perception. What is doubtful, however, is that an intermediate systematic phonetic level plays a useful role either in linguistic behaviour or linguistic

analysis (Ladd, 2011; Pierrehumbert et al., 2000; for similar arguments, see also Cole & Shattuck-Hufnagel, 2016). If this applies to segmentals, then it is unclear why something different is advocated for intonation. Based on the above, it is clear that the issue is not whether a broad or a narrow transcription of intonation is to be preferred, while discussion cannot be fruitfully focused on the level of detail to be transcribed.[6] It is not possible for any type of *transcription* to capture the full gamut of possible variability, while at the same time, using an intermediate systematic phonetic level can stop researchers from capturing essential generalizations (Arvaniti & Ladd, 2009; Browman & Goldstein, 1992).

5.3 The typology of intonation

The need for typological comparisons is an argument that has been put forward in favour of more surface faithful and detailed representations of intonation. As noted earlier, typological research is said to be hindered when similar phenomena are represented in different ways across languages (Ladd, 2008a, pp. 107–119, 2008b; Prieto & Hualde, 2016). At first glance, this seems like a legitimate concern. There are several elements of this argument, however, that warrant further scrutiny. First, it is not clear what kind of typology would require such consensus among representations. The typology envisaged either by Hyman (2006) or Beckman and Venditti (2011), to take two very different views, is not concerned with whether a system has a LH* accent and another a L+<H*, but rather with the origin and function of tones. As Hyman (2006) notes, any *phonological* typology must deal *not* with surface phonetic details but rather with the analytical categories used to make sense of these details in a given linguistic system. Thus Romani would be classed as a language that uses only postlexical tones (intonation) in combination with stress. For a typology of this sort, more generic categories would work better to bring a cross-linguistic understanding about; but generic categories are unlikely to be phonetically transparent.

If, on the other hand, a phonetic typology is envisaged, then details are better captured in terms of algorithms or patterns of realization rather than by a detailed but still symbolic notation which, as shown in Section 5.2., is unlikely to adequately capture all variability in realization. Peak delay is a good example of the inadequacy of a symbolic system in capturing commonalities that would be of use in constructing a phonetic typology: if peak delay is a parameter to encode, how far from the onset of a stressed vowel should a peak be before an accent is annotated as having a delayed peak? In answering this question one needs to consider the fact that peak location is only the outcome of an algorithm and thus only an approximation to begin with (Beckman & Venditti, 2011; Kochanski, 2010). Further, as shown in more detail below, the answer is clearly related to the system to which the accent belongs: if all peaks are systematically delayed, is it worth annotating delay at all? What if, like in Romani or Neapolitan Italian (Cangemi & Grice, 2016), peaks show substantial variability in alignment? Similar arguments apply to the transcription of undershoot: how far from typical must a given tone's scaling be before it is annotated? Can general criteria be established or should undershoot be defined for each speaker separately based on their pitch range and if so, how? If questions like these cannot be answered in a straightforward way — both because of logistical issues to do with

[6] It is sometimes argued that broad phonetic transcriptions would be useful for applications, such as speech synthesis. It is not possible to address this point here in detail, but typically speech synthesis systems rely on mark-up languages and implementation rules that incorporate quantitative measures rather than symbolic representations alone (cf. Hirschberg, 2006; Sproat et al., 1998; van Santen et al., 1997; Venditti et al., 2008).

how we measure turning points and define scaling relations, and because the answers to these questions cannot possibly be the same for all languages — we need to question the usefulness of such a level of transcription.

The issue of how to analyze linguistic systems and do typological comparisons is of concern to typologists in general. Some argue, like Ladd (2008a, pp. 107–119) or Hualde and Prieto (2016), that we need a predetermined set of categories into which to fit the elements of different systems. Others like Haspelmath (2010, 2015) argue that a typology which relies on a limited set of categories from which all languages choose is unsatisfactory for many reasons. An obvious one is that such categories can be unnecessarily restrictive and may fail to capture essential generalizations (cf. Haspelmath, 2015, on clitics). This is particularly likely to be true in the field of intonational phonology, as only a fraction of languages have been adequately described and thus the whole gamut of possibilities in terms of the organization of prosodic features and their realization is simply unknown. The proposal by Hualde and Prieto (2016) illustrates this point. The authors provide a series of five labels for bitonal accents (H + L*, H* + L, L + H*, L + < H*, L* + H) and propose canonical realizations for them. However, it is not certain that these five labels are sufficient to adequately capture all possible pitch accents researchers are likely to encounter as more languages are analyzed. Hualde and Prieto acknowledge that these labels should be broad enough to cover differences in realization, but this statement in itself implies that the essential categories are determined. This carries precisely the risk discussed by Haspelmath (2010, 2015).

To avoid such problems, typologists argue that one can have recourse to *comparative concepts*, which can be used for cross-linguistic comparison, while recognizing that language-specific categories are needed to account for phenomena specific to each language (Haspelmath 2010, 2015). What Ladd (2008a p. 110) calls "sustained level phrase-final pitch" could be such a concept when it comes to intonation. Pierrehumbert (1980) analyzed sustained level phrase-final pitch in English as a H-L% sequence of edge tones. In other analyses of English, however, it is argued to reflect the absence of a specific boundary tone (Ladd, 1983; Grabe, 1998, ch. 4, following Gussenhoven, 1984). In turn, the absence of a boundary tone is notated in some analyses as 0% (Grabe, 1998), or by not positing a tone at all, as in the German ToBI system, GToBI, in which sustained level phrase-final pitch is annotated as H-% (Grice et al., 2005). Arvaniti and Baltazani (2005), on the other hand, analyze sustained level phrase-final pitch in Greek as !H-!H%. Differences like these are seen by Ladd (2008a, pp. 107–119) as a problem. Ladd argues that sustained level phrase-final pitch is "on the face of it, a similar intonational phenomenon in different languages" (2008a, p. 110) and thus it should be presented in a similar way in all of them, because different representations can lead to the conclusion that languages differ more than they really do. Ladd's point is well taken; his discussion, however, glosses over differences that relate to system-internal relationships between tonal elements in the languages he considers. Yet, the representation of sustained level phrase-final pitch (or any other intonational phenomenon, for that matter) does depend on the overall system of the language under analysis; by glossing over this critical point, Ladd makes the different representations appear utterly arbitrary, though they are motivated by system-internal consistency. This can be clearly seen if one compares English and Greek.

In Pierrehumbert's (1980) analysis, sustained level phrase-final pitch comes about in the following manner. The H- of the H-L% configuration is downstepped because it is preceded by a H L sequence of tones (a H* + L accent to be exact); this is based on a more general tenet according to which all HLH tonal sequences trigger downstep of the second H tone (Pierrehumbert, 1980, p. 139). The downstepping of H- is explained

somewhat differently in the revised analysis of Beckman and Pierrehumbert (1986), in which all bitonal accents are said to trigger downstep independently of the sequence of tones involved. Finally L% is upstepped because it follows a H- phrase accent; this solution is possible because in English H-L% sequences in which L% is fully scaled are not attested (but see Gussenhoven, 2016). Now in Greek, sustained level phrase-final pitch follows either a L* + H or L + H* pitch accent, depending on the melody. Crucially, there is no evidence that L* + H or L + H* triggers downstep in Greek (Arvaniti, 2003; Arvaniti & Baltazani, 2005). In addition there is no HLH sequence on which downstep would apply.[7] Thus, neither of the two explanations of downstep used for English is possible in Greek, nor is there any other context-related reason for the downstep. This leads to the conclusion that downstep has to be treated as an independent feature in Greek (as also argued for English in Ladd, 1983, and for Dutch in Gussenhoven, 2005). Further, unlike English, sequences of H-L% without L% upstep are attested in Greek (Arvaniti et al., 2006b), making a L%-upstep rule like that of English equally unsuitable for Greek. In short, sustained level phrase-final pitch in Greek cannot be analyzed in the same way as sustained level phrase-final pitch in English. One can of course question whether H-L% is the only possible or optimal way of analysing sustained level phrase-final pitch in English; e.g., Gussenhoven (2016) presents cogent arguments against this analysis. This, however, remains an analytical decision *about English* and as such it should have little bearing on how sustained level phrase-final pitch is analyzed in Greek or any other language. As this example demonstrates, different decisions are the outcome of different system requirements. Among the differences appears to be the fact that the tonal space is carved up in ways that make it impossible to use just L and H tones for the analysis of all linguistic systems. Indeed any explicit use of the downstep feature argues in essence for a system with three levels (Brugos et al., 2006; Ladd, 1983; cf. Liberman, 1975).

At best then, one could argue that Arvaniti and Baltazani (2005) could have followed the convention established by GToBI and used !H-% instead of !H-!H% to indicate the lack of change in pitch (cf. Grice et al., 2005).[8] This however, is a simple question of notation, not a question of analysis or typology. On the other hand, and this is a crucial difference, an analysis whereby sustained level phrase-final pitch is represented either by a 0% boundary tone (as in Grabe, 1998) or no boundary tone at all (as in Gussenhoven, 2004, pp. 313–315) would require altogether different analytical decisions. As indicated in Section 4.2.3., such an analysis would require that the pitch accent of the melody includes the downstep which in the GRToBI analysis is represented as a sequence of two distinct tonal events, !H- and !H%, both independent of the pitch accent. Whether one or the other theoretical position is superior is beyond the scope of this paper, though it is likely that each is better suited for some languages than others.

The discussion above should serve to highlight the fact that differences among analyses are not all qualitatively the same. The distinctions between them should be acknowledged, as some are genuine problems with straightforward solutions and others are part of the nature of research itself. The differences are of three types which are discussed below primarily in relation to AM analyses of the vocative chant in a variety of languages (see **Table 2**). The vocative chant is used here because it is the most characteristic use of sustained level phrase-final pitch which, as noted above, has been a matter of some debate.

[7] I do not consider here the fact that alternative representations of Greek accents could create a HLH sequence; rather, I assume that the representations of the accents are correct.

[8] Even this representation may not be optimal for Greek, as !H% is needed independently of sustained level phrase-final pitch to represent the final tone in Greek *wh*-questions and related melodies (Arvaniti & Ladd, 2009; Arvaniti et al., 2014; Baltazani, 2006).

(i) Differences between intonational systems. Differences between systems arise for two reasons: first, melodies that are similar in form and function may still show differences substantial enough to warrant distinct representations; second, distinct representations may be required for the sake of analytical consistency. The former type is illustrated by Frota (2016), regarding the rise-fall contour associated with narrow focus in both Portuguese and Catalan. Frota shows that, despite superficial similarities, both production and perception data indicate that the former is an off-ramp H* + L and the latter an on-ramp L + H* pitch accent. On the other hand, the representation of sustained level phrase-final pitch in Greek discussed above exemplifies the system-internal considerations that force a particular analysis.

(ii) Distinct analytical positions. As noted in Section 4.2.3. and above, decisions about how to carve up a melody into distinct tonal events have consequences for their representation. This is the reason why the Dutch vocative chant is analyzed without recourse to an edge tone in Gussenhoven (2005): in his analyses, the drop from a high to mid-level pitch (which is then sustained) is analyzed as part of the H*!H pitch accent. H*!H % may indeed be best for Dutch as it reflects the fact that the melody applies to successive feet, when available, a behaviour typical of pitch accents (Gussenhoven, 2005; see Grice et al., 2000, for an alternative analysis). This type of analysis is not suitable for Greek or Polish, however, since both languages have only one level of stress, a metrical difference that makes iteration of the melody impossible (Arvaniti, 2007a; Arvaniti & Baltazani, 2005; Arvaniti et al., 2016).

(iii) Notational differences. Notational differences are evident in the representations of the vocative chant in **Table 2**; e.g. L + H* and LH* represent pitch accents with very similar characteristics; !H-0% and H-% arguably represent the same thing, sustained mid-level pitch as a reflex of phrasal tones.

The three types of differences discussed above cannot be approached in the same way. Differences between systems should be accepted as inevitable. Languages cannot be expected to have the same tonal inventory, use the same melodies, carve up the tonal space in the same manner, exhibit the same interactions between tones, or otherwise realize the same phonological entities in the same manner in all contexts (cf. the differences between Portuguese and Catalan reported in Frota, 2016). The prosodic type of the language in question and the interaction between metrical and tonal structure are additional sources of cross-linguistic variation. Such differences, as argued above, may lead by necessity to very different analyses, if those analyses are to be internally consistent.

On the other hand, disagreements in notation can be resolved relatively easily by agreeing on a set of consistent conventions. Such agreement could be reached on how to annotate a sequence of two identical edge tones: L-L%, L-% or L-0% etc. (but see Sections 5.4 and 5.5. below). Similarly, agreement should be possible on whether multi-tonal accents are best represented with the plus sign between tones or not (e.g., L + H* or LH*), or whether Jun and Fletcher's (2014) proposal to distinguish the two in a principled manner is to be preferred. It is important to keep in mind, however, that such differences are trivial (however intimidating they may be to non-initiates).

Distinguishing between notational and analytical disagreements is crucial for any attempt to standardize AM representations, especially as it appears that the two types of disagreement are sometimes overlooked: Hualde & Prieto (2016) treat the difference between !H% in the analysis of the Portuguese vocative chant (Frota et al., 2015) and

Language	Pitch accent	Phrasal tones
Catalan (Borràs-Comes et al., 2015)	L+H*	!H%
Dutch (Gussenhoven, 2005)	H*!H*	%
English (Brugos et al., 2006)	H*	!H-L%
German (Grice et al., 2005)	L+H*	H-%
Greek (Arvaniti & Baltazani, 2005)	L*+H	!H-!H%
Hungarian (Varga, 2008)	H*	!H-0%
Polish (Arvaniti et al., 2016)	LH*	!H-%
Portuguese (Frota et al., 2015)	(L+)H*	!H%

Table 2: AM representations of sustained phase-final pitch as used in the vocative chant.

!H-% in the German equivalent (Grice et al., 2005) as being on a par with the difference between the German !H-% and the !H-!H% used in Greek (Arvaniti & Baltazani, 2005). However, the difference between the Greek and German analyses is one of convention (a notational difference) while that between Portuguese and German reflects different analytical decisions about edge tones: Frota et al.'s analysis of Portuguese relies on boundary tones, while Grice et al.'s analysis of German posits both phrase accents and boundary tones. Similarly, the difference between Ladd and Schepman's (2003) analysis of English rising accents as (L + H)* and the use of H* for Romani is not a difference in notation; rather, it is a different analytical approach to the role and significance of the initial rise in such accents.

Analytical differences are not easy to resolve as they reflect different approaches to phenomena, often coupled with different requirements of the systems under analysis. Nevertheless, agreement in analytical decisions appears to be a desideratum for some; e.g., Hualde & Prieto (2016) talk of "the potential use of a generally accepted set of intonational labels and phonetic implementation rules that can be common across languages". Such a goal, however, would not only force all languages onto a phonetic Procrustean bed, but would also require that all researchers espouse the exact same principles and solutions to problems of analysis. Such homogeneity of opinion would be detrimental to scientific inquiry, and very unlikely to be achieved.

Though analytical differences among researchers will and should persist, useful progress could be made by working towards a generally agreed set of criteria and diagnostic tests that would allow researchers to evaluate alternative analyses for the same linguistic system on a consistent basis. Examples of recent research along these lines include Peters et al. (2015), Ritter and Grice (2015), and Gussenhoven (2016). As these studies indicate, criteria could relate, on the one hand, to levels of adequacy that analyses must meet, and on the other, to the empirical evidence that must support an analysis. Such criteria could include types of empirical evidence required to determine whether one or two types of edge tones are needed for the analysis of a given language, whether to posit one or more types of rising accents and what their tonal composition might be. Focusing on the development of such diagnostic tests, on the one hand, and on standardizing notation where appropriate, on the other, should help resolve many points of disagreement among analyses.

5.4 Phonetic transparency in intonation

Another reason put forward for more similarity in cross-linguistic representations of intonation is the need for phonetic transparency (Ladd, 2008a, p. 112). As Ladd concedes, however, there are some problems with this argument, in that phonetic transparency can complicate rather than facilitate comparisons across linguistic varieties. Ladd uses the vowel system of Scottish English to illustrate this difficulty: Scottish English does not have a contrast between /ʊ/ and /u/, and the vowel used in place of both is best transcribed as [ʉ]. Thus, neither /ʊ/ nor /u/ used in standard analyses of English is a good representation for the high mid central Scottish vowel; however, if, in the name of phonetic transparency, both /ʊ/ and /u/were to be replaced by /ʉ/ in the analysis of Scottish English, it would be difficult to compare the Scottish English vowel system with that of Southern Standard British English.[9] Ladd notes, however, that, while neither /ʊ/ nor /u/ is an ideal representation for the high central Scottish English vowel, no one would consider representing the vowel of *brick* or *break* using /ʊ/. In other words, there is some largely agreed upon phonetic substance related to these symbolic representations.

To my knowledge at least, the same applies to analyses of intonation. There are no analyses in which high pitch is represented by a L tone and low pitch by a H tone, a counterintuitive analytical decision equivalent to Ladd's *brick* transcribed with /ʊ/. The main point of disagreement across intonational analyses concerns sustained level phrase-final pitch (essentially mid-level pitch). This is due partly to differences among languages, as noted in Section 5.3., and partly to historical reasons which resulted in the adoption to a two-tone system forcing some rather cumbersome representations of pitch that is neither low nor high but is contrastive (Arvaniti, 2011). It is no coincidence that this is a main topic of scrutiny for four out of six papers in 2016 (Arvaniti, 2016; Frota, 2016; Gussenhoven, 2016; Prieto and Hualde, 2016).

The fact that phonological representations of intonation are not phonetically arbitrary is illustrated in **Table 2** which lists the representations of the vocative chant, a melody that according to Ladd shows "striking similarity across the languages of Europe" (Ladd, 2008a, p. 119). As can be seen in **Table 2**, all analyses involve a rising or high accent followed by sustained mid-level pitch. Differences in representation may seem overwhelming at first glance, but do not really obscure similarities and are not any more arbitrary than any phonological analysis of segments. Granted, one has to know that in the English system !H-L% involves an upstep of L%, but this is no different from having to learn that /p/ in English is aspirated in most contexts or that /b/ is rarely fully voiced (Keating, 1984) and thus that the symbols /p/ and /b/ do not represent quite the same sounds in English and French. What is noteworthy is that these types of discrepancies have long been accepted in segmental phonology, but are still treated as highly undesirable in the analysis of intonation. As I have argued elsewhere, one possible explanation is that intonation is not seen as being on a par with the rest of phonological structure even by those who study it (Arvaniti, 2007b). This tendency is probably reinforced by the relative phonetic transparency of L and H which forces a phonetic interpretation of phonological representations of intonation, impossible for abstract symbols like /p/ or /b/.

[9] Ladd does not address distributional, weight-related, and metrical criteria that could help determine the optimal (or least objectionable) phonological representation of Scottish [ʉ].

5.5 Dialectology, cross-linguistic comparisons, and the choice of categories

Dialectological research is another argument that has been used in support of a broad phonetic level of intonation transcription (Hualde & Prieto, 2016). Yet such transcriptions are now largely abandoned by dialectologists for the reasons discussed by Ladd (2008a, pp. 110–115) and briefly in Section 5.3. Following Wells (1982), instead of talking about /ʊ/ or /ɒ/, sociolinguists working on English talk about the FOOT and the LOT vowel respectively, a practice indicating a level of abstraction similar to that advocated here for intonation; talking about the FOOT vowel or the LOT vowel obviates the need to label the phonetic substance of these vowels but does allow for fruitful comparisons.

One thing to notice about this practice, however, is the cultural hegemony it reflects. The list of words used is based on categories that come from the system of Standard Southern British English. As luck would have it, it is the English vowel system with the largest number of vowel contrasts and thus it serves English dialectology well, but one wonders what that list of words would have been had it first been proposed by a speaker from Los Angeles or Newcastle; in the former case, there would be no separate entries for THOUGHT and LOT, while in the latter STRUT would be missing instead. Such biases, which are inevitable, add another layer of arbitrariness. Problems of this sort are inevitably compounded when a system of broad phonetic transcription is used precisely because phonetic substance cannot be left unspecified.

Indeed, problems do arise in dialectological comparisons when researchers opt for overly transparent phonological presentations. One such case is the proposal of Ladd and Schepman (2003) to collapse H* and L+H* into (L+H)*. As noted in Section 3.2., Arvaniti and Garding (2007) have shown that there is dialectal variation in the realization of these accents: in their study, speakers from California made a consistent distinction between H* and L+H*, using H* for new information and L+H* for contrastive focus, as suggested by Pierrehumbert (1980) and Pierrehumbert & Hirschberg (1990). Speakers from Minnesota, on the other hand, clearly had one pitch accent, L+H*, and relied on scaling to indicate the difference between new and contrastive information, as Ladd and Schepman (2003) would predict. Given these dialectal differences, collapsing the two categories in all descriptions of English intonation would be counterproductive as it would obscure a more important difference across dialects of English: the presence (or absence) of the L+H* vs. H* contrast. Doing so would be equivalent to using /ʉ/ for the analysis of all English varieties because Scottish has this vowel, or using only /ɔ/ for CAUGHT and LOT because Western varieties of US English have merged these vowels into one.

There are some additional concerns with respect to dialectology that go beyond cultural hegemony. A common transcription system implies that varieties of a language share some common core. In languages with well accepted and known standardized forms, this may be desirable and realistic and may have some psychological reality as well in that non-standard speakers are likely to be familiar with the standard. Experience, however, suggests that this does not apply to all speech communities, even those with highly codified standards: thus, British speakers are far more aware of a UK-wide English standard and are familiar with terms such as *RP*, *Queen's English*, and *BBC English*; for U.S. speakers, on the other hand, concepts like *Mainstream American English* or *General American English* hold little reality. This makes the enterprise of a common system possibly useful for linguists but of little psychological validity. This is all the more so for speakers like the Roma in the present study who are not familiar with a standard form of their language. In such circumstances, it would be highly unrealistic to posit that a common system for all Romani varieties must be used either for segmental or prosodic analysis, as such a system has no bearing on specific varieties and speakers. This state of affairs is likely to hold for speakers of other languages without a standard and without a written and schooling

tradition. This in turn means that while we discuss phonetic transparency, we make analytical decisions that fit one variety better than others, as would happen if the present analysis were to be made the base of intonation analysis in other Romani varieties.

The tendency for some linguistic varieties to take priority is implicit in cross-linguistic work as well; e.g., Hualde et al. (2002) argue that Lekeitio Basque is like Japanese; if Basque had been analyzed first it would be Japanese that would have to fit the Basque type. Although the similarities in these particular systems may render this difference trivial, issues of precedence can have consequences for other analyses: it is undeniable that many analytical decisions in intonation have been the way they are because of the influence of English.

6 Conclusion

In conclusion, the corpus presented here shows variability on a scale rarely encountered in data from educated monolingual speakers of standardized languages, though presumably common in many non-standardized linguistic varieties, particularly those showing extensive contact. Variability on this scale poses challenges for intonational analysis and highlights the importance of distinguishing between phonetic realization and phonological representation during analysis and determining intonational phonology on the basis of *meaningful* contrasts as in the rest of a language's phonological system. Though the need to adhere to these principles may be more obvious under conditions of variability like those discussed here, the argument made is that the principles would be useful in the intonational analysis of all languages. Such an analysis can be usefully and fruitfully compared with analyses of other languages leading to successful typological comparisons. This can be achieved without recourse to an intermediate level of broad phonetic transcription which cannot do justice to the full gamut of variability in any data, but will inevitably focus on some variable elements whether they are especially significant or not. Doing so may well obscure real cross-linguistic similarities and lead to researchers missing both significant generalizations and the opportunity to explore the full gamut of prosodic variation present in the world's languages.

Supplementary Files

For accompanying TextGrid, Pitch, and wav files, go to http://dx.doi.org/10.5334/labphon.14.smo.

Acknowledgements

Many thanks to my collaborator on this project, Evangelia Adamou, who introduced me to Romani, and to the Roma consultants without whose generous participation this work would not have been possible. The recordings, glosses, and translations of the examples cited in this paper are by Evangelia Adamou. Thanks are also due to the editors and two anonymous reviewers for their detailed and insightful comments. The financial support of CNRS through the Fédération de Typologie et Universaux Linguistiques, and of an Illinois-WUN International Development Grant is hereby gratefully acknowledged.

Competing Interests

The author declares that they have no competing interests.

References

Adamou, E. 2010. Bilingual speech and language ecology in Greek Thrace: Romani and Pomak in contact with Turkish. *Language in Society, 39*, 147–171. DOI: http://dx.doi.org/10.1017/S0047404510000035

Adamou, E., & Arvaniti, A. 2014. Illustrations of the IPA: Greek Thrace Xoraxane Romane. *Journal of the International Phonetic Association, 44*(2), 223–231. DOI: http://dx.doi.org/10.1017/S0025100313000376

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. 1991. The HCRC Map Task Corpus. *Language and Speech, 34*, 351–366.

Arvaniti, A. 2003. Peak scaling in Greek and the role of declination. *Proceedings of XVth ICPhS*, Barcelona, 4–9 August 2003, pp. 2269–2272.

Arvaniti, A. 2007a. Greek phonetics: The state of the art. *Journal of Greek Linguistics, 8*, 97–208. DOI: http://dx.doi.org/10.1075/jgl.8.08arv

Arvaniti, A. 2007b. On the relationship between phonology and phonetics (or why phonetics is not phonology). Special Session: Between Meaning and Speech: On the Role of Communicative Functions, Representations and Articulations. *Proceedings of ICPhS XVI*, pp. 19–24.

Arvaniti, A. 2011. The representation of intonation. In van Oostendorp, M., Ewen, C., Hume, E., & Rice, K. (Eds.), *Companion to Phonology* (pp. 757–780). Malden, MA: Wiley-Blackwell.

Arvaniti, A., & Adamou, E. 2011. Focus expression in Romani. *28th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings Project.

Arvaniti, A., & Baltazani, M. 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In Jun, S.-A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press, pp. 84–117. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199249633.003.0004

Arvaniti, A., Baltazani, M., & Gryllia, S. 2014. The pragmatic interpretation of intonation in Greek wh-questions. *Proceedings of Speech Prosody 7*, Dublin, May 20–23, 2011. Retrieved from http://www.speechprosody2014.org.

Arvaniti, A., & Garding, G. 2007. Dialectal variation in the rising accents of American English. In Cole, J., & Hualde, J. I. (Eds.), *Papers in Laboratory Phonology*. Berlin, New York: Mouton de Gruyter, pp. 547–576.

Arvaniti, A., & Ladd, D. R. 2009. Greek wh-questions and the phonology of intonation. *Phonology, 26*, 43–74. DOI: http://dx.doi.org/10.1017/S0952675709001717

Arvaniti, A., Ladd, D. R., & Mennen, I. 2000. What is a starred tone? Evidence from Greek. In Broe, M. & Pierrehumbert, J. (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, pp. 119–131.

Arvaniti, A., Ladd, D. R., & Mennen, I. 2006a. Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech, 49*, 421–450. DOI: http://dx.doi.org/10.1177/00238309060490040101

Arvaniti, A., Ladd, D. R., & Mennen, I. 2006b. Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication, 48,* 667–696. DOI: http://dx.doi.org/10.1016/j.specom.2005.09.012

Arvaniti, A., Zygis, M., & Jaskula, M. 2016. The phonetics and phonology of the Polish calling contours. In Żygis, M., & Malisz, Z. (Eds.), *Phonetica, 73,* Special Issue "Slavic perspectives on prosody".

Bakker, P., & Yaron, M. 2003. *Bibliography of modern Romani linguistics*. Amsterdam: Benjamins. DOI: http://dx.doi.org/10.1075/lisl.28

Baltazani, M. 2006. Intonation and pragmatic interpretation of negation in Greek. *Journal of Pragmatics, 38*(10), 1658–1676. DOI: http://dx.doi.org/10.1016/j.pragma.2005.03.016

Baltazani, M., & Jun, S.-A. 1999. Topic and focus intonation in Greek. *Proceedings of the XIVth International Congress of Phonetic Sciences, 2*, 1305–1308.

Barnes, J., Shattuck-Hufnagel, S., Brugos, A., & Veilleux, N. 2006. The domain of realization of the L- phrase tone in American English. *Proceedings of Speech Prosody* 2006, Dresden.

Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. 2012. Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology, 3*(2), 337–383. DOI: http://dx.doi.org/10.1515/lp-2012-0017

Beckman, M. E. 1986. *Stress and non-stress Accent*. Dordrecht: Foris. DOI: http://dx.doi.org/10.1515/9783110874020

Beckman, M. E., & Edwards, J. 1994. Articulatory evidence for differentiating stress categories. In Keating, P. A. (Ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III*. Cambridge: Cambridge University Press, pp. 7–33.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun, S.-A. (Ed.), *Prosodic typology :The phonology of intonation and phrasing*. Oxford: Oxford University Press, pp. 9–54. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199249633.003.0002

Beckman, M. E., Munson, B., & Edwards, J. 2007. The influence of vocabulary growth on developmental changes in types of phonological knowledge. In Cole, J., & Hualde, J. I. (Eds.), *Laboratory Phonology 9*. Berlin, New York: Mouton de Gruyter, pp. 241–264.

Beckman, M. E., & Pierrehumbert, J. 1986. Intonational structure in Japanese and English. *Phonology, 3*, 15–70.

Beckman, M. E., & Venditti, J. J. 2011. Intonation. In Goldsmith, J., Riggle, J., & Yu, A. C. L. (Eds.), *The handbook of phonological theory*. Malden, MA: Blackwell, pp. 485–532. DOI: http://dx.doi.org/10.1002/9781444343069.ch15

Borràs-Comes, J., Sichel-Bazin, R., & Prieto, P. 2015. Vocative intonation preferences are sensitive to politeness factors. *Language and Speech, 58*(1), 68–83. DOI: http://dx.doi.org/10.1177/0023830914565441

Borràs-Comes, J., Vanrell, M. M., & Prieto, P. 2014. The role of pitch range in establishing intonational contrasts. *Journal of the International Phonetics Association, 44*(1), 1–20. DOI: http://dx.doi.org/10.1017/S0025100313000303

Browman, C. P., & Goldstein, L. 1992. Articulatory phonology: An overview. *Phonetica, 49*, 155–180. DOI: http://dx.doi.org/10.1159/000261913

Brugos, A., Shattuck-Hufnagel, S., & Veilleux, N. 2006. *Transcribing prosodic structure of spoken utterances with ToBI* (MIT Open Courseware). Retrieved from http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm.

Calhoun, S. 2010. The centrality of metrical structure in signaling Information Structure: A probabilistic perspective. *Language, 86*(1), 1–42. DOI: http://dx.doi.org/10.1353/lan.0.0197

Cangemi, F., & Grice, M. 2016. The Importance of a Distributional Approach to Categoriality in Autosegmental-Metrical Accounts of Intonation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology,* 7(1): 9, pp. 1–20, DOI: http://dx.doi.org/10.5334/labphon.28

Chung, Y., & Arvaniti, A. 2013. Speech rhythm in Korean: Experiments in speech cycling. *Proceedings of Meetings on Acoustics (POMA)* 19.060216. Retrieved from http://scitation.aip.org/content/asa/journal/poma/19/1/10.1121/1.4801062.

Clopper, C. G., & Smiljanic, R. 2011. Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 39*(2), 237–245. DOI: http://dx.doi.org/10.1016/j.wocn.2011.02.006

Cole, J., & Shattuck-Hufnagel, S. 2016. New Methods for Prosodic Transcription: Capturing Variability as a Source of Information. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1): 8, pp. 1–29, DOI: http://dx.doi.org/10.5334/labphon.29

Cruttenden, A. 1994. *Gimson's pronunciation of English*. London: Edward Arnold.

Crystal, D. 1969. Review of M. A. K. Halliday, *Intonation and Grammar in British English*. The Hague: Mouton, 1967. *Language*, *45*(2), 378–393. DOI: http://dx.doi.org/10.2307/411669

Cummins, F., & Port, R. F. 1998. Rhythmic constraints on stress-timing in English. *Journal of Phonetics, 31*, 139–148. DOI: http://dx.doi.org/10.1016/S0095-4470(02)00082-7

D'Imperio, M. 2000. *The role of perception in defining tonal targets and their alignment* (Unpublished doctoral dissertation). The Ohio State University.

D'Imperio, M., Terken, J., & Piterman, M. 2000. Perceived tone "targets" and pitch accent identification in Italian. *Proceedings of Australian International Conference on Speech Science and Technology* (SST), *8*, 201–211.

Edwards, J. R., Beckman, M. E., & Munson, B. 2015. Cross-language differences in acquisition. In Redford, M. (Ed.), *The handbook of speech production*. Malden, MA: Wiley-Blackwell, pp. 530–554.

Fougeron, C., & Jun, S.-A. 1998. Rate effects on French intonation: Phonetic realization and prosodic organization. *Journal of Phonetics, 26*(1), 45–70. DOI: http://dx.doi.org/10.1006/jpho.1997.0062

FRA & UNDP: European Union Agency for Fundamental Rights and United Nations Development Program. 2012. The situation of Roma in 11 EU states: Survey results at a glance. Luxembourg: Publications Office of the European Union. Retrieved from http://www.scribd.com/doc/153872420/The-situation-of-Roma-in-11-EU-Member-States.

Frota, S. 2016. Surface and Structure: Transcribing Intonation within and across Languages. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1): 7, pp. 1–19. DOI: http://dx.doi.org/10.5334/labphon.10

Frota, S., Cruz, M., Svartman, F., Vigário, M., Collischonn, G., Fonseca, A., & Serra, C. 2015. Intonational variation in Portuguese: European and Brazilian varieties. In Frota, S. & Prieto, P. (Eds.), *Intonation variation in Romance*. Oxford: Oxford University Press, pp. 235–283. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199685332.003.0007; http://dx.doi.org/10.1093/acprof:oso/9780199685332.001.0001

Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech, 1*, 126–152.

Grabe, E. 1998. *Intonational phonology: English and German* (Unpublished doctoral dissertation). Max-Planck-Institute for Psycholinguistics and University of Nijmegen. Retrieved from http://www.phon.ox.ac.uk/files/people/grabe/thesis.html.

Grice, M., Baumann, S., & Benzmüller, R. 2005. German intonation in autosegmental-metrical phonology. In Jun, S.-A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press, pp. 55–83. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199249633.003.0003

Grice, M., Ladd, D. R., & Arvaniti, A. 2000. On the place of "phrase accents" in intonational phonology. *Phonology, 17*, 143–185. DOI: http://dx.doi.org/10.1017/S0952675700003924

Gussenhoven, C. 1984. *On the grammar and semantics of sentence accents*. Dordrecht: Foris. DOI: http://dx.doi.org/10.1515/9783110859263

Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511616983

Gussenhoven, C. 2005. Transcription of Dutch intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press, pp. 118–145. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199249633.003.0005

Gussenhoven, C. 2007. The phonology of intonation. In de Lacy, P. (Ed.), *The Cambridge handbook of phonology*. Cambridge: Cambridge University Press, pp. 253–280.

Gussenhoven, C. 2016. Analysis of Intonation: the Case of MAE_ToBI. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1): 10, pp. 1–35, DOI: http://dx.doi.org/10.5334/labphon.30

Halliday, M. A. K. 1967. *Intonation and grammar in British English*. The Hague: Mouton. DOI: http://dx.doi.org/10.1515/9783111357447

Haspelmath, M. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language, 86*(3), 663–687. DOI: http://dx.doi.org/10.1353/lan.2010.0021

Haspelmath, M. 2015. Defining vs. diagnosing linguistic categories: a case study of clitic phenomena. In Błaszczak, J., Klimek-Jankowska, D., & Migdalski, K. (Eds.), *How categorical are categories?* Berlin: De Gruyter Mouton, pp. 273–304. DOI: http://dx.doi.org/10.1515/9781614514510-009

Henrich, J., Heine, S. J., & Norenzayan, A. 2010. The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–135. DOI: http://dx.doi.org/10.1017/S0140525X0999152X

Hirschberg, J. 2006. Speech synthesis: prosody. In Brown, K. (Ed.), *Encyclopedia of language and linguistics*. Amsterdam: Elsevier, 2nd ed., pp. 49–55. DOI: http://dx.doi.org/10.1016/B0-08-044854-2/00914-7

Hualde, J. I., Elordieta, G., Gaminde, I., & Smiljanić, R. 2002. From pitch accent to stress accent in Basque. In Gussenhoven, C., & Warner, N. (Eds.), *Laboratory Phonology* 7. Berlin, New York: Mouton de Gruyter, pp. 547–584. DOI: http://dx.doi.org/10.1515/9783110197105.547

Hualde, J. I. and Prieto, P. 2016. Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1): 5, pp. 1–25. DOI: http://dx.doi.org/10.5334/labphon.11

Hyman, L. M. 2006. Word-prosodic typology. *Phonology, 23*, 225–257. DOI: http://dx.doi.org/10.1017/S0952675706000893

IPA. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Jun, S.-A. (Ed.). 2005a. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199249633.001.0001

Jun, S.-A. 2005b. Korean intonational phonology and prosodic transcription. In Jun, S.-A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press, pp. 201–229. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199249633.003.0008; http://dx.doi.org/10.1093/acprof:oso/9780199249633.001.0001

Jun, S.-A. (Ed.). 2014. Prosodic typology II: The phonology of intonation and phrasing. Oxford: Oxford University Press. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199567300.001.0001

Jun, S.-A., & Fletcher, J. 2014. Methodology of studying intonation: from data collection to data analysis. In Jun, S.-A. (Ed.), *Prosodic typology II: The phonology of intonation and phrasing*. Oxford: Oxford University Press, pp. 493–519. DOI: http://dx.doi.org/10.1093/acprof:oso/9780199567300.003.0016; http://dx.doi.org/10.1093/acprof:oso/9780199567300.001.0001

Keating, P. A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language, 60*(2), 285–319. DOI: http://dx.doi.org/10.2307/413642

Kim, J., & S.-A. Jun. 2009. Prosodic structure and focus prosody of South Kyungsang Korean. *Language Research, 45*(1), 43–66.

Knight, R.-A. 2008. The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux. *Language and Speech, 51*(3), 223–244. DOI: http://dx.doi.org/10.1177/0023830908098541

Knight, R.-A., & Nolan, F. 2006. The effect of pitch span on intonational plateaux. *Journal of the International Phonetic Association, 36*(1), 21–38. DOI: http://dx.doi.org/10.1017/S0025100306002349

Kochanski, G. J. 2010. Prosodic peak estimation under segmental perturbations. *Journal of the Acoustical Society of America, 127*(2), 862–873. DOI: http://dx.doi.org/10.1121/1.3268511

Ladd, D. R. 1983. Phonological features of intonational peaks. *Language, 59*, 721–759. DOI: http://dx.doi.org/10.2307/413371

Ladd, D. R. 1999. Perfect pitch. *New Scientist, 164*(2218), 87.

Ladd, D. R. 2008a. *Intonational phonology*. Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511808814

Ladd, D. R. 2008b. Review of Sun-Ah Jun (ed.) (2005). *Prosodic typology: the phonology of intonation and phrasing*. Oxford: Oxford University Press. *Phonology, 25*, 372–376.

Ladd, D. R. 2011. Phonetics in phonology. In Goldsmith, J., Riggle, J., & Yu, A. C. L. (Eds.), *Handbook of phonological theory*. Malden, MA: Blackwell, pp. 348–373. DOI: http://dx.doi.org/10.1002/9781444343069.ch11

Ladd, D. R., & Morton, R. 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics, 25*, 313–342. DOI: http://dx.doi.org/10.1006/jpho.1997.0046

Ladd, D. R., & Schepman, A. 2003. "Sagging transitions" between high accent peaks in English: "experimental evidence". *Journal of Phonetics, 31*, 81–112. DOI: http://dx.doi.org/10.1016/S0095-4470(02)00073-6

Ladefoged, P., & Johnson, K. 2011. *A course in phonetics*. Boston: Wadsworth Cengage Learning.

Liberman, M. 1975. *The intonational system of English* (Unpublished doctoral thesis). MIT.

Martins-Heub, K. 1989. Genocide in the 20th century: Reflections on the collective identify of German Roma and Sinti (Gypsies) after national socialism. *Holocaust Genocide Studies, 4*(2), 193–211. DOI: http://dx.doi.org/10.1093/hgs/4.2.193

Matras, Y. 1999. Writing Romani: The pragmatics of codification in a stateless language. *Applied Linguistics, 20*(4), 481–502. DOI: http://dx.doi.org/10.1093/applin/20.4.481

Matras, Y. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511486791

Matras, Y. 2005. The future of Romani: Toward a policy of linguistic pluralism. *Roma Rights Quarterly, 1*, 31–44.

Mixdorff, H., Leemann, A., & Dellwo, V. 2014. The influence of speech rate on Fujisaki model parameters. *EURASIP Journal on Audio, Speech, and Music Processing, 33*. Retrieved from http://asmp.eurasipjournals.com/content/2014/1/33. DOI: http://dx.doi.org/10.1186/s13636-014-0033-6

Niebuhr, O. 2008. Coding of intonational meanings beyond F0: Evidence from utterance-final /t/ aspiration in German. *Journal of the Acoustical Society of America, 124*(2), 1252–1263. DOI: http://dx.doi.org/10.1121/1.2940588

Niebuhr, O. 2012. At the edge of intonation – The interplay of utterance-final F0 movements and voiceless fricative sounds in German. *Phonetica,* 69(1–2), 7–21. DOI: http://dx.doi.org/10.1159/000343171

O'Connor, J. D., & Arnold, G. F. 1973. *Intonation of colloquial English: A practical handbook*. London: Longman.

Peters, J., Hanssen, J., & Gussenhoven, C. 2015. The timing of nuclear falls: Evidence from Dutch, West Frisian, Dutch Low Saxon, German Low Saxon, and High German. *Laboratory Phonology, 6*(1), 1–52. DOI: http://dx.doi.org/10.1515/lp-2015-0004

Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation* (Unpublished doctoral thesis). MIT.

Pierrehumbert, J. 2002. Word-specific phonetics. In Gussenhoven, C., & Warner, N. (Eds.), *Laboratory Phonology VII*. Berlin, New York: Mouton de Gruyter, pp. 101–139. DOI: http://dx.doi.org/10.1515/9783110197105.101

Pierrehumbert, J., Beckman, M., & Ladd, D. R. 2000. Conceptual foundations of phonology as a laboratory science. In Burton-Roberts, N., Carr, P., & Docherty, G. J. (Eds.), *Conceptual and empirical foundations of phonology*. Oxford: Oxford University Press, pp. 273–303. Reprinted 2012 in Cohn, A., Fougeron, C., & Huffman, M. (Eds.), *The Oxford handbook of laboratory phonology*. Oxford: Oxford University Press, pp. 17–39.

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

Podesva, R. J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics, 11*, 478–504. DOI: http://dx.doi.org/10.1111/j.1467-9841.2007.00334.x

Port, R. F., & Leary, A. P. 2005. Against formal phonology. *Language, 81*(4), 927–964. DOI: http://dx.doi.org/10.1353/lan.2005.0195

Prieto, P. 1998. The scaling of the L tone line in Spanish downstepping contours. *Journal of Phonetics, 26*, 261–282. DOI: http://dx.doi.org/10.1006/jpho.1998.0074

Prieto, P. 2006. Word-edge tones in Catalan. *Italian Journal of Linguistics, 18*(1), 39–71.

Prieto, P., D'Imperio, M., & Gili-Fivela, B. (2005). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech, 48*(4), 359–396. DOI: http://dx.doi.org/10.1177/00238309050480040301

Prieto, P., & Roseano, P. (eds.). 2010. *Transcription of intonation of the Spanish language*. Lincom Europa: München.

Ritchart, A., & Arvaniti, A. 2014. The form and use of uptalk in Southern California English. *Proceedings of Speech Prosody 7*, Dublin, May 20–23, 2014. Retrieved from http://www.speechprosody2014.org.

Ritter, S., & Grice, M. 2015. The role of tonal onglides in German nuclear pitch accents. *Language and Speech, 58*(1), 114–128. DOI: http://dx.doi.org/10.1177/0023830914565688

Sachs, H., Schegloff, E. A., & Jefferson, G. 1974. A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696–735. DOI: http://dx.doi.org/10.2307/412243; http://dx.doi.org/10.1353/lan.1974.0010

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. 2000. Covert contrast as a stage in the acquisition of phonetics and phonology. In Broe, M., & Pierrehumbert, J. (Eds.), *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*. Cambridge: Cambridge University Press, pp. 194–207.

Silverman, K. E. A., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P., Pierrehumbert, J. B., & Hirschberg, J. 1992. ToBI: A standard for labeling English prosody. *Proceedings of ICSLP 1992*.

Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C., & Krifka, M. 2006. *Questionnaire on information structure*. Potsdam: Universität Potsdam, Institut für Linguistik/Allgemeine Sprachwissenschaft.

Sproat, R., Hunt, A., Ostendorf, M., Taylor, P., Black, A., Lenzo, K., & Edgington, M. 1998. SABLE: A standard for TTS markup. *International Conference on Spoken Language Processing*, 1998.

Stuart-Smith, J., Sonderegger, M., Rathcke, T., & Macdonald, R. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology 6*(3–4), 505–549. DOI: http://dx.doi.org/10.1515/lp-2015-0015

Swerts, M., Krahmer, E., & Avesani, C. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics, 30*, 629–654. DOI: http://dx.doi.org/10.1006/jpho.2002.0178

Tannen, D. 1987. Conversational style. In Dechert, H.-W., & Raupach, M. (Eds.), *Psycholinguistic models of production*. Norwood, NJ: Ablex, pp. 251–267.

Tyaglyy, M. 2009. Were the "Chingené" victims of the Holocaust? Nazi policy toward the Crimean Roma, 1941–1944. *Holocaust Genocide Studies, 23*(1), 26–53. DOI: http://dx.doi.org/10.1093/hgs/dcp015

van Santen, J., Shih, C., & Möbius, B. 1997. Intonation. In Sproat, R. (Ed.), *Multilingual text-to-speech synthesis: The Bell Labs approach*. Dordrecht, Boston, London: Kluwer Academic Publishers, pp. 141–189.

Varga, L. 2008. The calling contour in Hungarian and English. *Phonology, 25*, 469–497. DOI: http://dx.doi.org/10.1017/S0952675708001607

Venditti, J., Maekawa, K., & Beckman, M. E. 2008. Prominence marking in the Japanese intonation system. In Miyagawa, S., & Saito, M. (Eds.), *Handbook of Japanese linguistics*. Oxford: Oxford University Press, pp. 456–512. DOI: http://dx.doi.org/10.1093/oxfordhb/9780195307344.013.0017

Wells, J. C. 1982. *Accents of English*. Cambridge: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511611759

Xu, Y. 2010. In defense of lab speech. *Journal of Phonetics, 38*, 329–336. DOI: http://dx.doi.org/10.1016/j.wocn.2010.04.003

Yuasa, I. P. 2010. Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech, 8*(3), 315–337. DOI: http://dx.doi.org/10.1215/00031283-2010-018

How to cite this article: Arvaniti, A 2016 Analytical Decisions in Intonation Research and the Role of Representations: Lessons from Romani. *Laboratory Phonology: Journal of the Association for Laboratory Phonology,* 7(1): 6, pp. 1–43, DOI: http://dx.doi.org/10.5334/labphon.14

Published: 30 June 2016

Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS