Focus Expression in Romani Amalia Arvaniti and Evangelia Adamou

Near-final version of paper to appear in the proceedings of WCCFL 2010 PLEASE DO NOT QUOTE IN THIS FORM

Focus Expression in Romani

Amalia Arvaniti and Evangelia Adamou

University of California, San Diego and French National Centre for Scientific Research (CNRS)

1. Introduction

Cross-linguistically, focus is marked by syntactic, morphological and prosodic means. Languages like Italian or Catalan utilize syntactic changes to mark narrow focus (Ladd 2008 and references therein), while Wolof (Rialland & Robert 2001) and Chickasaw (Gordon 2007) rely on focus particles. Perhaps the most common strategy is prosodic focus marking, which takes many forms, including changes in prosodic phrasing as in Korean (Jun 2005) and Japanese (Venditti, Maekawa & Beckman 2008), the use of accentuation as in English (see Ladd 2008 for a review), and the selection of a particular accent to indicate narrow focus as in Greek (Arvaniti, Ladd & Mennen 2006).

Despite the large number of strategies available cross-linguistically for focus marking, it is widely held that most languages make use of a small number of strategies and typically do not use them concurrently (e.g. Büring 2009). For example, the use of focus particles tends to obviate the need for prosodic or syntactic changes as shown for Wolof (Rialland & Robert 2001) and Chickasaw (Gordon 2007). In addition, it is widely held that prosody is a lighter or less costly strategy than syntactic changes and thus to be preferred when possible. This widely accepted understanding of focus marking has recently been formalized as a "minimality condition" by Skopeteas & Fanselow (in press). According to Skopeteas & Fanselow, less complex strategies are preferred to more complex ones (if both available in a given language) following a markedness scale from lightest to most structurally complex: *in situ* (prosody) < reordering < cleft. This scale also assumes that prosodic and syntactic focus marking means are relatively independent of each other. Nevertheless, it is well known that many languages show consistent interactions between prosody and syntax: e.g., languages like Italian or Catalan use word order changes so that particular items that must be in focus can phonologically receive accent; in such languages, word order changes must accompany accentuation (Ladd 2008).

Our evidence from the variety of Romani spoken in Komotini (see section 2) is relevant to this debate since in this variety prosodic and morphosyntactic focus strategies are combined in the same utterance. Furthermore, the Komotini Romani data provide evidence for a novel strategy of focus marking, the use of stress-shift under focus discussed in more detail in section 6.4.

2. Komotini Romani

Romani is an Indo-Aryan language spoken in Europe, Australia, and the Americas. The variety of Romani on which we report is spoken by a small Muslim community settled in the suburbs of the city of Komotini, in the area of Greek Thrace (see Map). Komotini Romani belongs to the Vlach Romani branch and is heavily influenced by contact with Turkish since the Ottoman times (Adamou 2010). Its speakers are typically trilingual in Romani, Turkish and Greek with different degrees of competence in the three languages. They use Turkish and Greek for trade and other professional activities, and Romani mainly at home and as a community language. The majority of the Komotini Roma have received practically no formal education in any of their languages and are not literate in Romani.

3. Corpus and analysis

Our corpus includes mainly (though not exclusively) two types of data: (a) natural dialogs between one female and one male speaker and (b) story telling by these speakers and an additional

male speaker, all in their thirties (20 minutes of story telling with synchronized sound and annotation are available at http://lacito.vjf.cnrs.fr/archivage/languages/Romani_fr.htm via Lacito's Oral Tradition Language Documentation Program; http://idiom.ucsd.edu/~arvaniti/WCCFL2010 provides audio files and PRAAT annotation files of the examples and figures in this paper). Some additional data from female speakers from Xanthi (see Map), whose variety is very similar to Komotini Romani, are also included in our corpus. These are data from conversations between consultants and the second author and from more formal elicitation sessions.

The data were prosodically analyzed following the principles of the autosegmental-metrical framework of intonational phonology. The analysis was based on the simultaneous inspection of waveforms, spectrograms and pitch contours using PRAAT. Given that virtually nothing is known about the intonational system of Komotini Romani, we stress that our analysis is tentative at this point and requires further (and controlled) verification. For this reason, we do not provide autosegmental representations of the F0 contours in the figures, though we do briefly discuss our current analysis.



Map: Thrace, Greece. The recordings took place in the city of Komotini (circled).

4. Prosodic and focus-related features of Romani

As mentioned, Komotini Romani is influenced by Turkish. Among other features, it has borrowed the particle *da*, which is used for both focus marking and topicalization, as in Turkish (Göksel & Özsoy 2003), but also as a coordinator. This is exemplified in (1) below (all examples are given in broad phonological transcription).

(1) te na gaˈrados da getʃiˈnesin ˈmansa
to NEG hide.2SG and live.2SG me.INSTR
"...so you won't hide and live with me."
(following "My mother told me to burn it so you won't turn into a snake ...")

In Komotini Romani, arguments are normally postverbal. In turn, OV appears to be reserved for topicalization and narrow focus marking as in other Romani dialects (Matras 2002, 2006). Compare, for instance, examples (2) and (3): (2) is used during story telling at a point where $d \mathcal{J}uv$ "louse" has already been mentioned, while (3) is used when the speaker juxtaposes her previously unimpeded filling of prescriptions with the current situation where she is required to show her identity card in order to have her prescription filled. SV and VS are both possible in Komotini Romani but differences in their function are not fully understood. Tentatively we can say that VS is the canonical order, as shown in (4), while, SV indicates narrow focus (see e.g., Fig. 17) or topicalization (see e.g., Fig. 15). This largely agrees with observations from other Romani dialects (see e.g., Matras 1995, 2002, who argues that SV is reserved for contrastive thematic roles).

(2) *lel kaˈja dʒw* take.3SG this louse "He takes this louse."

- (3) aka na mi taf tota man gen now my identity card ask.3PL "Now they ask for my IDENTITY CARD."
- (4) ki da-pes by tyn o gavu tno gather.3SG-REFL all the village people "All the village people gather."

Similarly to other conservative Romani dialects (Matras 2002), Komotini Romani has final stress in the native parts of its vocabulary. Stress can also fall on the penult or antepenult in the numerous borrowings and when certain suffixes and enclitics (such as case and TMA markers) are involved, as in examples (5) and (6a-b) respectively. However, minimal pairs like that in (7a-b) are rare.

- (5) aste hava "hospital"
- (6a) 'apo "pill" (6b) 'apora "pills"
- (7a) *kana* "when" (7b) *ka ha* "ears"

5. Intonation and the marking of broad focus

Our data show that in Komotini Romani pitch modulation is not lexically determined but is a phrasal property. That is, pitch is used to mark prosodic boundaries, indicate the pragmatic function of an utterance and highlight particular words in discourse. This is illustrated in Figs. 1-4 which show melodies used with declaratives (Figs. 1 and 2), a negative declarative (Fig. 3) and a wh-question (Fig. 4). As can be seen, the overall melodies differ, but in each some pitch movements co-occur with stressed syllables and others with phrasal boundaries. We analyze the former as *pitch accents* and the latter as *boundary tones*. We tentatively recognize at least three types of pitch accent in the Komotini Romani system, L+H*, H+L* and L* and at least one level of phrasing that is delimited by the presence of a H% or L% boundary tone: compare, e.g., the pitch rise on *ma'mi* "grandma" in Fig. 1 (black dotted line), with the fall on *ga'va* "villages" in the same figure.

As can be seen in Fig. 1, content words are typically accented. In prenuclear accents, F0 rises from a low point to a peak reached during the accented vowel suggesting a L+H* pitch accent. Such accents can be seen in prenuclear position on 'dzalas' "went" and 'papo "grandpa" in Fig. 1, but they are also used in nuclear position, as on ga 'va "villages" in the same figure. L+H* accents may also be used for narrow focus marking, as on naj "is not" and so "what" in Figs. 3 and 4 respectively. On the other hand, we tentatively analyze the nuclear pitch accent in broad focus utterances as H+L* since it starts with high pitch (often creating a plateau with the preceding accent, if one is present) and falls rapidly throughout the duration of the accented syllable. Accents of this sort can be seen on toma fila "cars" in Fig. 2 and on naj "is not" in Fig. 13 (which also illustrates the above-mentioned plateau).

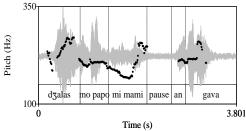


Fig. 1: *dʒalas mo ˈpapo mi ma mi | an ga va* went.3PL my grandpa my grandma to villages "My grandpa and grandma went to villages [to work]."

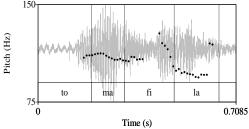


Fig. 2. toma fila car.PL "Cars."

6. Narrow focus

As mentioned, Romani displays an interesting and unusual array of focus marking strategies, possibly because of the influence of Greek and Turkish. These strategies are discussed below.

6.1. Narrow focus marked by prosody alone

As Fig. 3 shows, Komotini Romani can mark narrow focus by prosody alone. In Fig. 3, the negation verb *naj* "is not" is accented (with the accent tentatively analyzed as L+H*), while the rest of the utterance is deaccented. In our corpus, negative statements are the only unambiguous instance of *in situ* use of prosody to mark narrow focus.

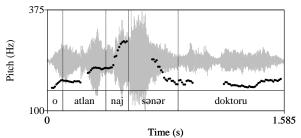


Fig. 3: o ad han [atlan] naj se her dokto hu the Adnan is.NEG nerve doctor "Adnan IS NOT a neurologist!"

6.2. Narrow focus marked by prosody and word order

The most common strategy for marking narrow focus in Komotini Romani combines prosody with syntactic marking. For instance, in the wh-question in Fig. 4, the wh-word is in utterance-initial position. It is accented while the rest of the utterance is deaccented. In Fig. 5, a change in word-order from canonical VO to OV is combined with deaccenting of the verb to mark narrow focus on 'apora "pills" when the speaker uses this utterance to introduce a new topic. In Fig. 6 non-canonical OV order is also used; in this case, the object is a noun phrase which carries narrow focus only on but "many"; the noun of each NP is deaccented as is the following verb. The entire pattern is repeated three times. Likewise in Fig. 7, the subjects tfank "leg", kuj "arm" and dan da "teeth" precede the verb of their phrase and are accented while the verbs are deaccented, a combination that puts the subjects in narrow focus in all three phrases.

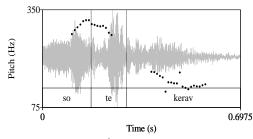


Fig. 4: so te ke rav what to do.1SG "WHAT can I do?"

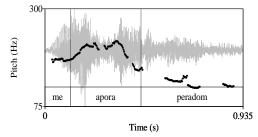


Fig. 5: me 'apora pera'dom my pills lost.1SG "I lost my PILLS."

¹ This pattern which requires that the wh-word be placed in initial position and carry the only accent in the utterance is also found in Greek (Arvaniti & Ladd 2009), Romanian and Hungarian (Ladd 2008: 224 ff.), as well as in other varieties of Romani, such as Erli and Kalderaš spoken in Bulgaria (Grigorova 1998).

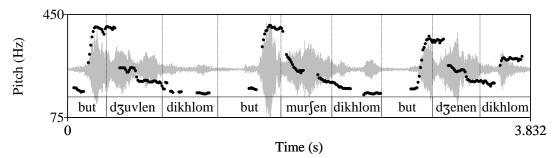


Fig. 6 but dɔuˈvlen di khlom | but muˈrʃen di khlom | 'but dɔeˈnen di khlom ...
many women saw.1SG many men saw.1SG many people saw.1SG
"Many women I've seen, many men I've seen, many people I've seen ..."

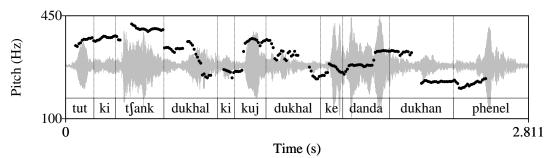


Fig. 7 tut | ki tʃank du khal | ki kuj du khal | ke dan da du khan phe hel you.ACC your leg hurt.3SG your arm hurt.3SG your.PL teeth hurt.3PL say.3SG "You, your leg hurts, your arm hurts, your teeth hurt, he says..."

6.3. Narrow focus marked by the focus marker da, prosody and word order

Accentuation and word order changes can be accompanied by the use of the clitic *da* to mark focus. In such cases, the word to which *da* cliticizes is accented, and the remainder of the utterance is deaccented, as illustrated in Figs. 8, 9 and 10. As can be seen in these figures, *da* when used as a focus marker attaches to nouns and pronouns the position of which is always preverbal.

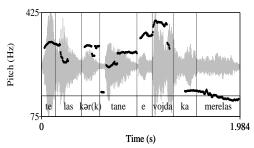


Fig. 8: te las kər(k) ta he
to took.3SG forty pieces
e voj da ka me relas
eh she.NOM FOC will died.3SG
"If she had taken forty pieces, eh, SHE
[too] would have died."

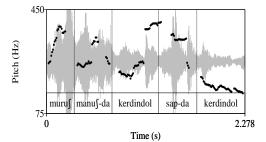


Fig. 9: mu'ruf | ma'nuf da ker'dindol | sap da boy man FOC became snake FOC ker'dindol became "Into a boy, a MAN he turned, into a SNAKE he turned."

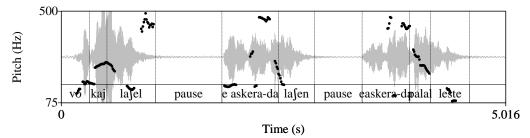
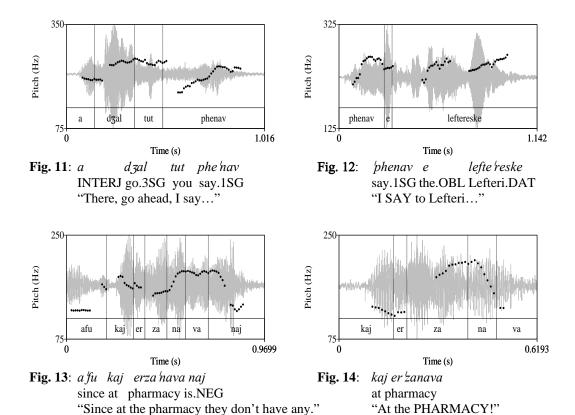


Fig. 10: vo kaj la fel e a skera da la fen e a skera da pa la leste he where leave.3SG the soldiers FOC leave.3PL the soldiers FOC behind him.LOC "Wherever he goes, the SOLDIERS go, the SOLDIERS follow him."

6.4. Narrow focus marked by accentuation and stress-shift

Finally, our data show that speakers of Komotini Romani employ an unusual strategy in combination with accentuation to mark focus, namely a non-metrically motivated stress-shift to an earlier syllable from the one canonically stressed. Stress-shift is used for narrow focus marking as in Fig. 12, though we have also observed it in cases in which focus scope is ambiguous, as in Fig. 14. On the other hand, Figs. 11 and 13 show canonical stress placement for the words *phe hav* "I say" and *erza hava* "pharmacy" respectively, when these are accented but not in focus. The changes in stress placement are evident in two ways. Between Fig. 11 and Fig. 12 there is a clear difference in the articulation of *phenav* "I say" in that the first vowel [e] is hardly visible in Fig. 11 but clearly articulated in Fig. 12; the opposite obtains for the second vowel [a]. In Fig. 13 vs. Fig. 14 there are differences both in tonal alignment and in the relative duration of [za] and [na]. In the case with canonical stress (Fig. 13) [za] is 1.5 times as long as [na]; it is also low in pitch and the pitch rise starts on the following syllable. The same pitch pattern is found in Fig. 14 but starting one syllable earlier, so that [za] is rising in pitch; in addition, in this case [za] is 2.2 times as long as [na]. To our knowledge, stress shift for focus marking is not explicitly examined in previous studies of Romani but it is mentioned in passing in Lee (2005: 7). Our data confirm that it is used at least in the Komotini variety.

The Romani stress shift shares similarities with a number of phenomena found in other languages, such as the English Rhythm Rule and accent shift (Ladd 2008: 234ff.), the French "accent d'insistence" (e.g., Dahan & Bernard 1996 among many) and a phenomenon comparable to the French one found in Greek (see Arvaniti 2007 for a review). However, unlike French in which the phenomenon is phrasal, Romani has lexical stress which is shifted here. Similarly, the Greek "accent d'insistence" (which is manifested as word-initial high pitch) is additional to the accent on the lexically stressed syllable and likely functions delimitatively rather than culminatively. In English, on the other hand, the accent shift is used to contrast morphologically similar words within an utterance, as in Bolinger's oft-cited example "This whisky wasn't EXported, it was DEported" (Boligner 1961: 83). This does not apply to our Komotini Romani instances of stress shift under focus. Finally, the English Rhythm Rule is motivated by metrical considerations that do not apply in our data. Indeed, in some instances of stress-shift, the use of this strategy may lead to a stress clash, as in (8a) below (in which the final vowel of $su \, stei(a)$ "pants" is not produced and a short pause is introduced between the object and verb, possibly to remedy the clash; 8b represents the canonical stress pattern for ki hav "I sell"). In short then, the Romani stress-shift differs in crucial respects from all of the above phenomena, and thus appears to be a novel strategy associated solely with focus marking.

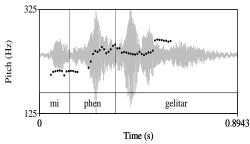


7. Focus, topicalization and the role of deaccenting

Our data also provide several instances of topicalization which show that both narrow focus marking and topicalization are achieved by left dislocation and accentuation of the dislocated constituent. What distinguishes narrow focus marking from topicalization is that when the dislocated constituent is focused, the remainder of the utterance is deaccented and phrased with it. In topicalization, the remainder of the utterance is not deaccented and the topicalized constituent can form its own phrase (though this is not necessary). For instance, in cases like that in Fig. 15, non canonical SV order is used and the subject is accented, but the verb is also accented with a non-downstepped L+H* accent. The effect is that of topicalizing the subject rather than putting it in narrow focus. Fig. 16 illustrates the phrasal break (after *gomeno* "boyfriend") that is often used in topicalization; it also shows accents on *si* "is" and *kar velas* (name) in the following phrase. Finally, Fig. 17 shows that topicalization and narrow focus may be combined: the object *man* "me" is topicalized and forms its own phrase, while *o le fteri li as* "Lefteri hired" forms a distinct phrase in which non-canonical SV is used with narrow focus on the subject and deaccenting of the verb.

8. Discussion and conclusions

The Romani data show that focus marking strategies additional to those already known may be available cross-linguistically. One such strategy is the stress-shift of Komotini Romani. In addition, Romani can be added to the small number of languages, such as Serbian (Godjevac 2004), which have a large repertoire of focus marking means and tend to use them concurrently. It is equally important to note that in Komotini Romani the concurrent use of syntactic and prosodic means for focus marking does not appear to be motivated by constraints in one of these domains, as, e.g., in Catalan or Italian (see section 1). In other words, our data suggest that the use of several focus marking strategies in Komotini Romani is a preference rather than a necessity.



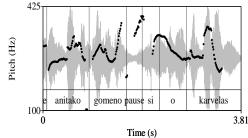


Fig. 15: mi phen ge li-tar my sister went.3SG-away "My sister's GONE."

Fig. 16: e ani tako gomeno | si o kar velas the Anita.GEN boyfriend is the Karvelas "Anita's boyfriend is Karvelas."

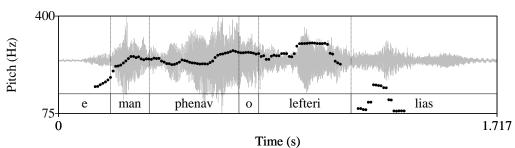


Fig. 17: e man | phe nav | o lefteri li as eh me.ACC say.1SG the.NOM Lefteri took.3SG "Eh, as for me, I say, it was LEFTERI who hired me."

The reasons for the rich focus marking strategies of Komani Romani are not entirely clear. It is possible that this is the result of borrowing and fusion. In other words, it is possible that in addition to its "native" strategies, Komotini Romani has adopted strategies from the languages it has been mostly influenced by, namely Turkish (the use of *da*) and Greek (specific uses of accentuation not common in other Indo-Aryan languages, such as Hindi and Bengali; Féry 2009). Evidence for such influences has been documented for Kalderaš and Erli Romani which have borrowed the Bulgarian question particle *li* and show intonational similarities with Bulgarian, their language of contact (Grigorova 1998).

Another possible reason could be that the use of several strategies to mark focus has a cumulative effect akin to the use of expletives in English. Thus, while a speaker of English may say "now they ask me for freaking id," a Romani speaker can instead change word order and use accentuation to achieve the same effect. This explanation seems plausible, given that the data examined here were not typical elicitation utterances which tend to be dispassionate and least marked from a discourse perspective. Rather, in the corpus of this study, speakers mostly discussed personal topics (altercations at work, medical problems) or "performed" in front of a live audience (story telling) and thus were animated and emotional. Therefore, it is possible that the relative rarity of languages like Romani may be simply an epiphenomenon of the data that have been collected from other languages rather than evidence of the typological rarity of Romani. Obviously, the two explanations offered here are not mutually exclusive: it is possible that the reason why Komotini Romani speakers do not have to resort to expletives is the fact that they can avail themselves instead of a rich focus marking system which came about thanks to borrowing (similarly to the French heritage speakers in Bullock 2009).

Under either of these explanations, however, it remains clear that the concurrent use of various focus marking strategies in Romani runs counter to the proposal of Skopeteas & Fanselow (in press) discussed in section 1 and to similar views expressed in the literature (Büring 2009, among others) which do not anticipate rich focus marking systems like that of Romani. As we have shown, Romani rarely uses prosody independently of morphosyntactic marking and typically combines two or three strategies in the same utterance.

A related point is that clefting were not present in our data either for focus or topic marking, though clefting is used in other Romani dialects (Matras 2002). The absence of clefts suggests that Komotini Romani exploits its rich options for accentuation, phrasing and word order to mark both topic and focus. The similarity between focus marking and topicalization in Komotini Romani clearly shows that concentrating on the accentuation (or lack thereof) of a particular word or constituent and using that as the sole criterion of whether focus has been prosodically marked can be misleading, and that taking into account the phrasing and intonational structure of the entire utterance is necessary for understanding how focus is marked in a particular linguistic variety.

In conclusion, the Romani data show both universally noted tendencies in focus marking as well as language-specific patterns, confirming the view that focus and accentuation patterns cannot be easily reduced to a limited set of universal principles. It thus remains clear that in order to fully understand the role of intonation and its interaction with syntax and pragmatics in focus marking more cross-linguistic work using both spontaneous speech and controlled data is necessary.

Acknowledgements

We thank the speakers, and gratefully acknowledge the financial support of Lacito-CNRS for the collection of the data (2005-2009), and a seed grant from the Worldwide Universities Network through the University of Illinois, Urbana-Champaign that facilitated our collaboration.

References

Adamou, E. 2010. Bilingual speech and language ecology in Greek Thrace: Romani and Pomak in contact with Turkish. *Language in Society* 39/2: 147-171.

Arvaniti, A. 2007. Greek phonetics: The state of the art. Journal of Greek Linguistics 8: 97-208.

Arvaniti, A., D. R. Ladd. 2009. Greek wh-questions and the phonology of intonation. Phonology 26: 43-74.

Arvaniti, A., D. R. Ladd & I. Mennen. 2006. Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech* 49: 421-450.

Bolinger, D. 1961. Contrastive accent and contrastive stress. Language 37: 83-96.

Bullock, B. E. 2009. Prosody in contact in French: A case study from a heritage variety in the US. *International Journal of Bilingualism* 13: 165-194.

Büring, D. 2009. Towards a typology of focus realization. In M. Zimmermann & C. Féry (Eds.), *Information Structure: Theoretical, Typological, and Experimental Perspectives*, pp. 177-205.

Dahan, D. & J-M. Bernard. 1996. Interspeaker variability in emphatic accent production in French. *Language and Speech* 39: 341-374.

Féry. C. 2009. Indian languages as intonational 'phrase languages'. Universität Potsdam manuscript. Available online at http://www.sfb632.uni-potsdam.de/~fery/aktuell.html. Accessed on March 18, 2010.

Godjevac, S. 2004. Focus Projection in Serbo-Croatian. Chicago, IL: The University of Chicago Press.

Gordon, M. 2007. The intonational realization of contrastive focus in Chickasaw. In C. Lee, M. Gordon & D. Büring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, pp. 69-82. Dordrecht: Springer.

Göksel, A. & A. S. Özsoy. 2003. dA: a focus/topic associated clitic in Turkish. Lingua 113: 1143-1167.

Grigorova, E. 1998. Interrogative intonation of two Bulgarian Romani dialects: Sofia Erli and Kalderaš. *Grazer Linguistische Studien* 50: 45-63.

Jun, S. 2005. Korean intonational phonology and prosodic transcription. In S. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, pp. 201-229. Oxford University Press.

Ladd, D. R. 2008. Intonational Phonology. Cambridge: Cambridge University Press.

Lee, R. 2005. Learn Romani: Das-duma Rromanes. University of Hertfordshire Press.

Matras, Y. 1995. Connective (VS) word order in Romani. *Sprachtypologie und Universalienforschung* 48(1/2): 189-203.

Matras, Y. 2002. Romani: A Linguistic Introduction. Cambridge: Cambridge University Press.

Rialland, A. & S. Robert. 2001. The intonational system of Wolof. Linguistics 39: 893-939.

Skopeteas, S. & G. Fanselow (in press). Focus types and argument assymetries. A cross-linguistic study in language production. In C. Breul (Ed.), *Contrastive information structure*. Benjamins.

Venditti, J. V., Maekawa, K., & Beckman, M. E. 2008. Prominence marking in the Japanese intonation system. In S. Miyagawa & M. Saito (Eds.), *Handbook of Japanese Linguistics*, pp. 456-512. Oxford University Press.