

InTraSAL: an intonational model for South Asian languages

Sameer ud Dowla Khan, Reed College ICPHS Satellite Meeting on the Intonational Phonology of Typologically Rare or Understudied Languages, University of Melbourne, 4 August 2019

It is a well-known fact that despite their classification into several unrelated families, the languages of South Asia share many features, spanning syntax, morphology, and (segmental) phonology (Emeneau 1956). It has also been documented that in all of these features, we find regional variation that can also cross family lines (Subbārāo 2012). For example, while South Asian languages (SALs) generally use non-nominative case marking for subjects of experiencer verbs, the choice of case varies by region, with most languages (spanning Indo-European, Dravidian, and Austroasiatic) using the dative, and languages in the Eastern Zone (spanning Indo-European and Tibeto-Burman) using the genitive, as shown in Figure 1. The current study expands the range of features examined to include intonational phonology, documenting both shared properties and regional variations within those features, based on close annotations of recordings of the North Wind and the Sun fable in 11 languages from both Indo-European and Dravidian families. These also form the basis of a proposed prosodic annotation system designed to capture this range of variation within and across SALs: Intonational Transcription of SALs (InTraSAL).

Figure 1. Experiencer subject case in SALs. Circle = DAT; square = GEN; triangle = NOM

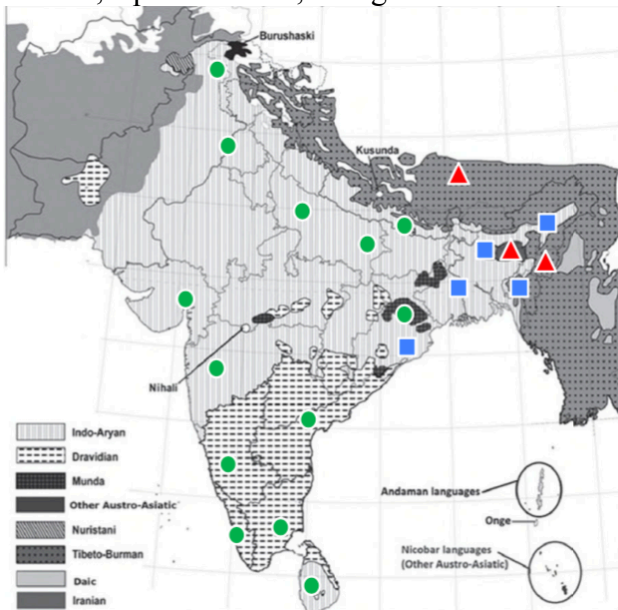
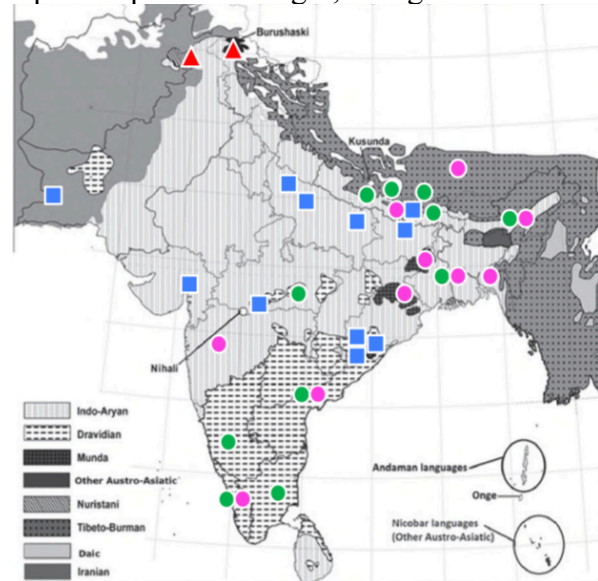


Figure 2. Patterns of stress assignment in SALs. Green circle = initial; pink circle = initial + weight; square = penult + weight; triangle = contrastive



Among the prosodic features shared across most SALs are (1) non-contrastive tone or stress (accent) placement, (2) an affinity between the location of stress and the presence of a low intonational tone, and (3) at least one high intonational tone affiliated before or after this low tone. The most canonical case of this, in which a word-sized rising contour is built from a low tone on the stressed initial syllable and a high tone at the end of the word/phrase, has been described in countless descriptions of Bengali, Hindi, Tamil, and other SALs from both Indo-European and Dravidian families (e.g. Féry 2010). In most cases, this Repeating Rising Contour (RRC) can be described as involving a low pitch accent L^* and a high boundary tone H_a projected by the accentual phrase (AP). Some analyses, however, propose (Féry 2010) or consider (Keane 2014) other interpretations of this contour, using left- and right-edge boundary tones and no pitch accents, suggesting little to no role of stress placement in determining this shape.

Even with the most basic finding of the RRC across SALs, it is clear that there are several significant areas of variation that have not been fully understood. These include (1) default stress assignment, (2) the timing of the H tone, and (3) RRCs with more complex contours (not covered here due to space limitations). The current study examines these to see how one prosodic model can capture this variation.

The first of these variations is the fact that stress assignment is not uniform across SALs: while “core Indic” languages (e.g. Gujarati, Hindi, Sindhi, Urdu) is described as having a weight-sensitive right-aligned trochaic system of stress assignment (i.e. stress the penult unless another syllable is heavier), Dravidian (e.g. Kannada, Malayalam, Tamil, Telugu) and “peripheral Indic” languages (e.g. Assamese, Bengali, Marathi) are described as having left-aligned trochaic stress assignment with an unclear relation to weight (i.e. stress the first syllable, possibly with a shift to the second syllable when heavy); see Fig. 2. Thus, syllable count and weight can determine whether these proposed stress systems predict the presence of a L tone (predicted if it is a pitch accent L*), or if the L tone is invariably phrase-initial (predicted if it is a boundary tone). From examining languages like Hindi, Sindhi, and Urdu (Fig. 3), it is clear that in situations where the stressed syllable is non-initial (e.g. *šimāli* ‘northern’), the L target aims for that stressed syllable rather than the left edge of the word, suggesting its identity as a pitch accent (Genzel 2007, Patil et al. 2008) rather than a boundary tone (contra Féry 2010).

Figure 3. L tone on non-initial stress in Urdu.

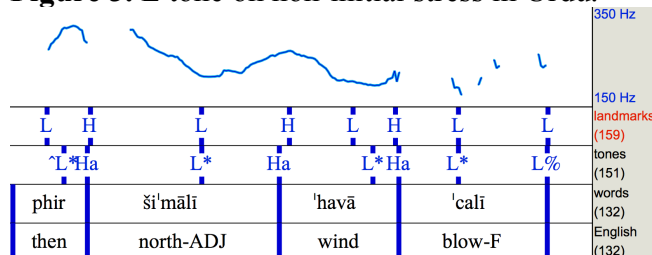
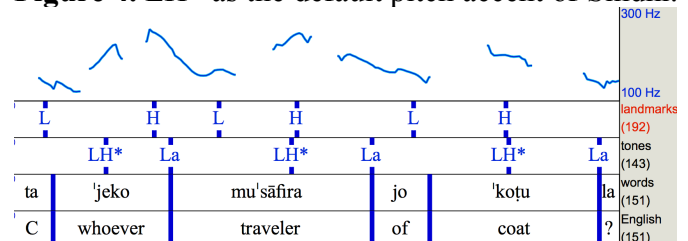


Figure 4. LH* as the default pitch accent of Sindhi.



The second area of RRC variation concerns the location(s) of the H target in each RRC. While this H tone is canonically on the AP-final syllable in most SALs, suggesting its identity as a right-boundary tone, there are situations in which the H appears immediately after (L*H), on (LH*), or before (HL*) the stressed syllable, suggesting its identity as part of the pitch accent. With the exception of Sindhi, where LH* appears to be the default pattern (Fig. 4), these stress-aligned H tones mostly appear in longer words, or are simply an alternative pattern to the usual L*...Ha rise. This further questions the entirely boundary-tone-driven analysis of Féry (2010), as these tones follow the stress rather than the edge.

Finally, one major contribution of this work is the InTraSAL prosodic annotation system, which not only provides labels for pitch accents (i.e. L*, H*, LH*, L*H, HL*), boundary tones (e.g. Ha, La, HLa, L%, HL%...), break indices, and words (with conventions appropriate for the phoneme inventories of SALs), but also has a built-in system for how to approach unfamiliar SALs. As seen in Figs. 3–4, InTraSAL includes a Landmarks tier, in which transcriber-perceived tonal targets are transcribed only by relative height (L, H) and phonetic timing, without further interpretation as to their phonological identity, alignment, or association. Once these are labeled, the transcriber can then begin a further analysis on the Tones tier, as far as they feel confident in the underlying system. This allows for both (1) less transcriber-imposed structure before a language has been sufficiently explored, and (2) a better check on existing prosodic models of SALs, as the landmarks should be able to correspond roughly to intonational phonological tones.

Altogether, with a broader and more systematic collection of data from across SALs, and a more flexible prosodic annotation system to describe it, InTraSAL will hopefully help shed light on the typology of SAL intonation both as a holistic region and as a collection of diverse languages.

References

- Féry, C. 2010. Indian languages as intonational ‘phrase languages’. In: Hasnain, I.; Chaudhury, S. (eds). *Problematizing Language Studies. Festschrift for Rama Agnihotri*. Delhi: Aakar Books, 288-312.
- Genzel, S. 2007. *Phonetische Realisierung von Fokus auf Adjektiven im Hindi*. MA thesis, Potsdam Univ.
- Keane, E. 2014. *Intonational phonology of Tamil*. In Jun, S.-A. (ed.). *Prosodic Typology II: Phonology of Intonation & Phrasing*. OUP.
- Patil, U.; Kentner, G.; Gollrad, A.; Kügler, F.; Féry, C.; Vasisht, S. 2008. Focus, word order, and intonation in Hindi. *JSAL* 1(1).
- Subbārāo, K. V. 2012. *South Asian languages: a syntactic typology*. Cambridge University Press.