A Database of Tonogenetic Events (DTE) and what it can tell us about tonogenesis

Lilja Sæbø (University of Oxford), Steven Moran (University of Neuchâtel),
Eitan Grossman (Hebrew University of Jerusalem)

Tonogenesis has become a topic of increasing interest, including numerous descriptions of tonogenetic events across a variety of language families (Haudricourt 1954, Arnold 2020, Hyslop 2009) with much recent work towards models (Dockum 2019, Gehrmann 2022) and typologies of tonogenesis (Hyman 2018, Hyslop submitted). To aid in these endeavors, we have constructed a preliminary database of tonogenetic events (DTE), which aims to organize the documentation that exists on tonogenesis into a format that is easy to filter, search, and compare.

The DTE currently describes 229 tonogenetic events from 90 language varieties in 26 language families across five macro-areas (North America, Africa, Asia, Europe and Papunesia). Along with language variety metadata, the database includes information about the triggering context of each tonogenetic event, the resultant tone, and a description of the effect that the tonogenetic event had on pitch. The triggering contexts have been divided into five main groups: onset, coda, nucleus (i.e. tonogenesis triggered by a feature of the vowel itself, such as for instance height (Ratliff 2015:253), stress and word type (i.e. tonogenesis triggered by the loss or gain of a syllable). Thus, the coding of the DTE provides a general classification of tonogenetic events reported in the literature, making it straightforward to see what kind of triggering contexts result in what kind of tones.

Our goal is to provide an open source and easily accessible resource that can be harnessed to ask, and answer, typological questions related to tonogenesis. Towards this aim, we have so far investigated two areas of tonogenesis that we will report on in this talk. The first is the relationship between the triggering context and the resultant tone, and the second is an areal classification of the distribution of different types of tonogenesis.

As to the first issue, our data largely concurs with trends already reported in the literature, such as the fact that voiced onsets tend to give rise to lower tones than voiceless onsets (Kingston 2011, Hombert, Ohala and Ewan 1979). While some triggers almost always have a given effect on the pitch, there were other triggers that could have various different effects. An example is voiceless aspirated and unaspirated stops, where there is no clear trend as to what trigger will give the highest tone. Regarding areality, the DTE for example shows that among the languages in the sample from Asia, it is very common to have undergone a two step tonogenetic process similar to that described for Vietnamese by Haudricourt (1954). That is, in the first step the coda consonants first create two or more different contour tones, and in a second step these tones are doubled by a second series of tonogenetic events based on the voicing/aspiration of the onset, generally creating a high and a low register. In our data, this kind of tonogenesis is not found outside of Asia. Other areal trends include the fact that the most common context for tonogenesis in North America is codas, while word-type tonogenesis is the most common in Europe. In this talk, we will discuss both topics in greater detail.

In accordance with principles for open cross-linguistic typological research put forth in the Cross-Linguistic Linked Data (CLLD) project (Forkel et al. 2018), we are making the DTE available as a CLDF data set. This makes its contents interoperable with Glottolog and numerous other typological databases (e.g., WALS, PHOIBLE, ASJP), so that new types of questions can be asked about tonogenesis. For example, one can investigate whether there are any correlations between the current phonological system of a language and the types of tonogenesis that it has undergone. Thus, in line with the goals of this workshop, we aim to bring together linguists and data sources to generate dialogue and collaboration to shed light on the diachronic study of tone and the processes of tonogenesis.

References

Arnold, L. 2020. Highs and lows: A previously unattested tone split from vowel height in Metnyo Ambel. *Transactions of the Philological Society*, 118, 141–158.

Dockum, R. 2019. The Tonal Comparative Method: Tai Tone in Historical Perspective. PhD thesis, Yale University.

Forkel, R., et al. 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data*, 5:180205.

Gehrmann, R. 2022. Desegmentalization: Towards a Common Framework for the Modeling of Tonogenesis and Registrogenesis in Mainland Southeast Asia with Case Studies from Austroasiatic. PhD thesis, The University of Edinburgh.

Haudricourt, A. G. 1954. De l'origine des tons en vietnamien. *Journal Asiatique*, 242, 69–82. English translation by Marc Brunelle: The origin of tones in Vietnamese.

Hombert, J. M, Ohala, J. J, and Ewan, W. G (1979). Phonetic explanation for the development of tones. *Language*, 55(1), 37–58.

Hyman, L. 2018. Towards a Typology of Tone System Changes. In: H. Kubozono and M. Giriko (Eds.), *Tonal Change and Neutralization*, 203–222.

Hyslop, G. 2009. Kurtöp Tone: A tonogenetic case study. *Lingua*, 119, 827–845.

Hyslop, G. Submitted. Toward a Typology of Tonogenesis: refining the model.

Kingston, J. (2011). Tonogenesis. In: Oostendorp, M., Hume, E. V.. and Ewen, C. J. eds. *The Blackwell Companion to Phonology.* John Wiley & Sons.

Ratliff, M. 2015. Tonoexodus, Tonogenesis, And Tone Change. In: Honeybone, P., Salmons, J eds., *The Oxford Handbook Of Historical Phonology*. Oxford University Press. 245–261.