

R YOU SURE THAT YOUR /R/ IS TRILLED? A METHODOLOGICAL CAVEAT

RÉMI ANSELME^{*1}, FRANÇOIS PELLEGRINO¹, and DAN DEDIU²

^{*}Corresponding Author: remi.anselme@univ-lyon2.fr

¹Laboratoire Dynamique du Langage, CNRS/University of Lyon, Lyon, France

²University of Barcelona & Catalan Institute for Research and Advanced Studies (ICREA),
Barcelona, Spain

Databases are great tools for typological studies and for understanding the evolution of language, but they can rarely be taken at face value. Many (if not all) databases have various limitations and usage constraints (see, for example, for WALS (Cysouw, Dediu, & Moran, 2012; Hunley, Bowern, & Healy, 2012) and UPSID (Simpson, 1999)), which can "trickle down" to (or even be amplified by) the studies based on them. We focus here on *phonemic databases*, which can be used, among others, to compare phonemic inventories across languages and to make claims about language evolution through perceptual or articulatory mechanisms for instance. However, such databases often suffer from an intrinsic ambiguity between *phoneme/allphone/phone*, which often percolates down to research that aims to study the *phonetic systems* of languages from (supposedly) *phoneme inventories* (for a recent example, see (Winter, Sóskuthy, Perlman, & Dingemanse, 2022)). One common hurdle is that sometimes the phoneme is seen as a monolithic object without variation in its phonetic realizations, and without contextual dependencies. Moreover, *written symbols* can be misleading because the same symbol may represent different phonetic realities (Anderson et al., 2018). To drive these points home, we will focus here on *r* (Barry, 1997).

Not all segments are directly comparable, and the *rhotics* class is a perfect example because of their great phonetic variability (Chabot, 2019; Lindau, 1985). In particular, by default the rhotic *r* is often considered as an alveolar trill: from a *phonemic* point of view, UPSID (Maddieson, 1984) reports that among the languages that have a rhotic, at least half have one alveolar trill or more. But to say that a language has an alveolar trill *phoneme* does not inform about its *potential phonetic realizations*. Several studies show that the trill alveolar phonemes have few of their allophones that are actually trilled (Sebregts, 2014; Blecua, 2002; Rafat, 2010), potentially leading to a distorted picture. As an example, we consider here a recent article (Winter et al., 2022) which correlates the presence of the trilled /r/ sound with "roughness" cross-modally. Their result is supported by several approaches, one being a cross-linguistic statistical study of 332 languages,

where, in order to determine whether a language has a trilled /r/ or not, the authors rely on (a) their phonetic judgments based on the literature, and (b) PHOIBLE (Moran & McCloy, 2019). Although their phonetic judgments took precedence over PHOIBLE, both methods tend to force the interpretation of the phoneme as having a single realization and may introduce some residual subjectivity.

As a sort of proof of concept, we decided to reproduce their process, by focusing on a random sub-sample of the languages they considered (34 of 332; $\approx 10\%$) for which they mention a trilled /r/. Our rerun is based on the authors' own comments in the data files used for data preparation, which we used as a starting point for searching grammars and reports on the languages of interest. When the authors based their judgment on PHOIBLE, we accessed the primary sources in there to assess whether /r/ was indeed trilled (however, this is still problematic because of the limitations mentioned in the previous paragraph). We systematically tried to use several resources and avoid making decisions based on only one source.

The findings on this sub-sample are that, first, in 16 of the languages (47% of the sub-sample), we fail to reach the same conclusions as the authors. There are three languages that should not have been included in the original analysis as they show a contrast between a trilled /r/ and a non-trilled /r/ (exclusion criterion in the original study). Eight other languages should not have been considered as having a trilled /r/ because the rhotic is described as not trilled in the primary sources even if the symbol *r* is used in the grammars (leading to confusion). And in another six languages, the articulation of the rhotic is not defined in the grammars, making it impossible to tell if it is a trilled /r/ (*sonorant*, *resonant*, *liquid*, or *rhotic*) or not, other than by implicitly assuming an interpretation of the symbol *r* as a trilled /r/ (which is not warranted). For the remaining 18 languages (53% of the sub-sample), we may consider them as having a trilled /r/, since we did find grammars containing phonetic information. Still, this is not always the case, with some decisions made solely on the basis of the mention "trill" in the phoneme inventory of the languages' reference grammars.

Should our preliminary results generalize to the whole sample of 332 languages, it may question the validity of the statistical cross-linguistic findings of a cross-modal association between the alveolar trill and "roughness". However, our point here is not to criticize this particular study, but to highlight the dangers of assuming phonetic meaning from written symbols allegedly representing phonological systems. In fact, we want to underline that, while large databases (and their associated statistical methods) are essential for the modern language sciences and the study of language change and evolution, they cannot be used without considerable care and expertise. The good news is that such expertise is available and achievable, as well as incorporable in large-scale databases (see, for example, the inclusion of multiple inventories and of allophones in PHOIBLE), and that new statistical approaches, capable of dealing with the remaining uncertainty, can be developed (probably in a Bayesian framework).

References

- Anderson, C., Tresoldi, T., Chacon, T., Fehn, A.-M., Walworth, M., Forkel, R., & List, J.-M. (2018). A cross-linguistic database of phonetic transcription systems. *Yearbook of the Poznan Linguistic Meeting*, 4(1), 21–53.
- Barry, W. J. (1997). Another r-tickle. *Journal of the International Phonetic Association*, 27(1-2), 35–45.
- Blecua, B. (2002). *Las vibrantes del español: manifestaciones acústicas y procesos fonéticos*. Unpublished doctoral dissertation, Universitat Autònoma de Barcelona.
- Chabot, A. (2019). What's wrong with being a rhotic? *Glossa: a journal of general linguistics*, 4(1), 38.
- Cysouw, M., Dediu, D., & Moran, S. (2012). Comment on “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa”. *Science*, 335(6069), 657–657.
- Hunley, K., Bowern, C., & Healy, M. (2012). Rejection of a serial founder effects model of genetic and linguistic coevolution. *Proceedings of the Royal Society B: Biological Sciences*, 279(1736), 2281–2288.
- Lindau, M. (1985). The Story of /r/. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (p. 260). Academic Press Inc.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.
- Moran, S., & McCloy, D. (Eds.). (2019). *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History.
- Rafat, Y. (2010). A Socio-Phonetic Investigation of Rhotics in Persian. *Iranian Studies*, 43(5), 667–682.
- Sebregts, K. D. C. J. (2014). *The Sociophonetics and Phonology of Dutch r*. Unpublished doctoral dissertation, LOT, Netherlands Graduate School of Linguistics, Utrecht.
- Simpson, A. P. (1999). Fundamental problems in comparative phonetics and phonology: does upsid help to solve them. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 1, pp. 349–352). San Francisco.
- Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, 12(1), 1035.