Comparing cross-language phonological profiles

George Moroz

Linguistic Convergence Laboratory (HSE University)

November 9, 2021

presentation is available here: tinyurl.com/yj2tacek

How I decided to give this talk?

- During the talk in our Lab with Misha and Ezequiel

Jeff Good: How you came up with the idea of calculating phonological distances? Is it some established procedure?

Me: No, we thought that it is the most obvious step…

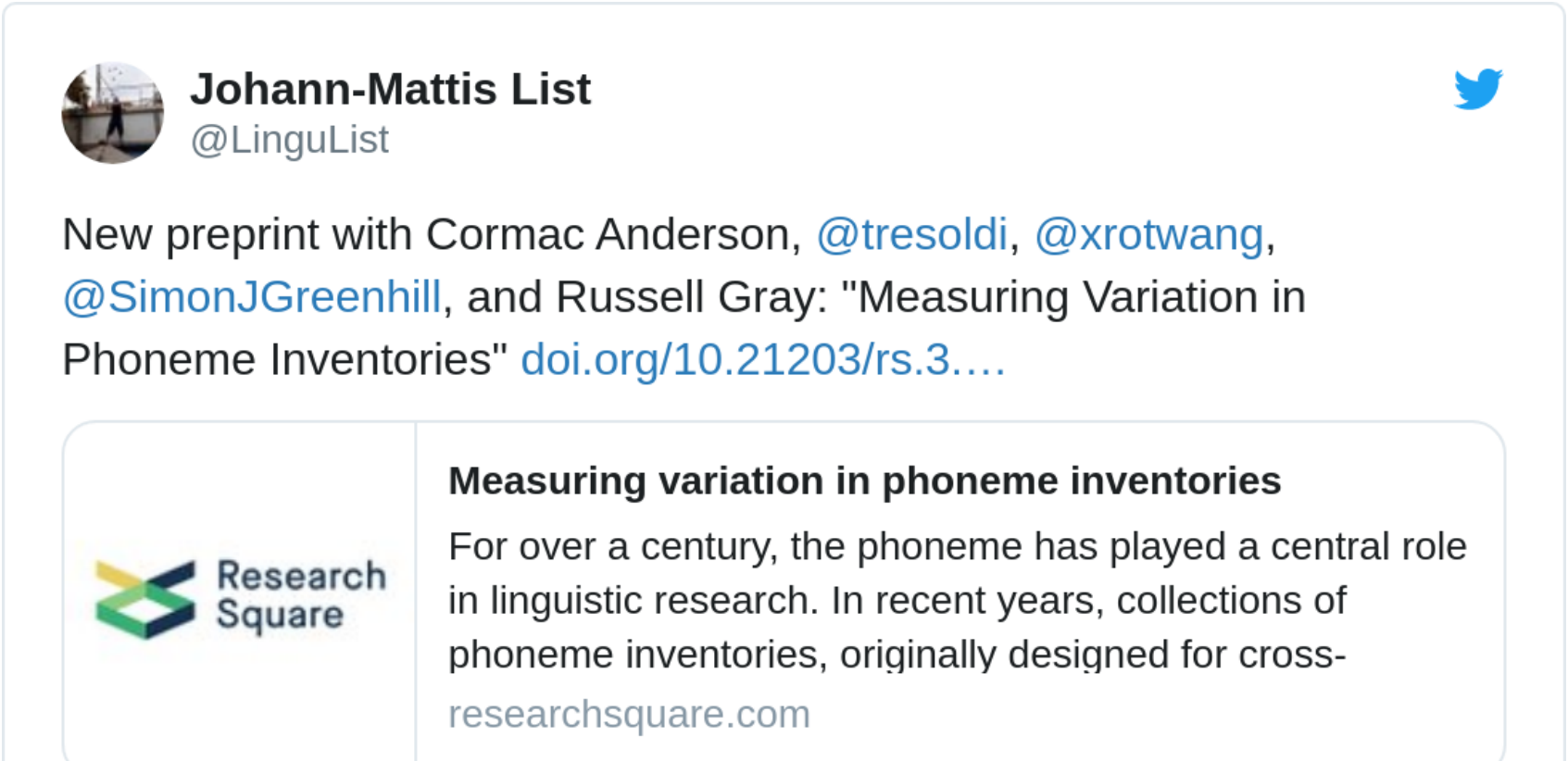
How I decided to give this talk?

- During the talk in our Lab with Misha and Ezequiel

Jeff Good: How you came up with the idea of calculating phonological distances? Is it some established procedure?

Me: No, we thought that it is the most obvious step…

- The second reason:



Overview

Materials for the analysis

Criticism by [Simpson 1999]

Complexity based approaches

Distance based approaches

Materials for the analysis

Materials for the analysis can be different:

- segment[1] inventory (and grammar, if you are lucky);
- dictionaries;
- parallel corpora;
- unparalleled corpora.

[1] Lets leave the phonology vs. phonetics debate aside.

Overview

Materials for the analysis

Criticism by [Simpson 1999]

Complexity based approaches

Distance based approaches

Criticism by [Simpson 1999]

[Simpson 1999] attacks UPSID[1]-like researches:

- phoneme masks allophones
  - Standard High German /ç/ stands for [ç], [x] and [χ];
  - "The allophone no longer represents the phoneme, it *replaces* it";
- phonological relations between segments is lost
  - comparing just vowel inventories it is impossible to get information about e. g. vowel harmony;
- there is no non-arbitrary way of assign phonological features (e. g. SPE [Chomsky and Halle 1968]) to segments.

[1]UPSID stands for UCLA Phonological Segment Inventory Database [Maddieson and Abramson 1987] which consists of the phonemic systems of a representative sample of 451 (this number changes from publication to publication) of the world's languages in machine-readable form. Now UPSID can be accessed via PHOIBLE database [Moran and McCloy 2019].

Criticism by [Simpson 1999]

[Simpson 1999] attacks UPSID[1]-like researches:

- phoneme masks allophones
  - Standard High German /ç/ stands for [ç], [x] and [χ];
  - "The allophone no longer represents the phoneme, it *replaces* it";
- phonological relations between segments is lost
  - comparing just vowel inventories it is impossible to get information about e. g. vowel harmony;
- there is no non-arbitrary way of assign phonological features (e. g. SPE [Chomsky and Halle 1968]) to segments.

My metaphor: omelet and pancakes share all ingredients, but they are significantly different meals.

[1]UPSID stands for UCLA Phonological Segment Inventory Database [Maddieson and Abramson 1987] which consists of the phonemic systems of a representative sample of 451 (this number changes from publication to publication) of the world's languages in machine-readable form. Now UPSID can be accessed via PHOIBLE database [Moran and McCloy 2019].

Overview

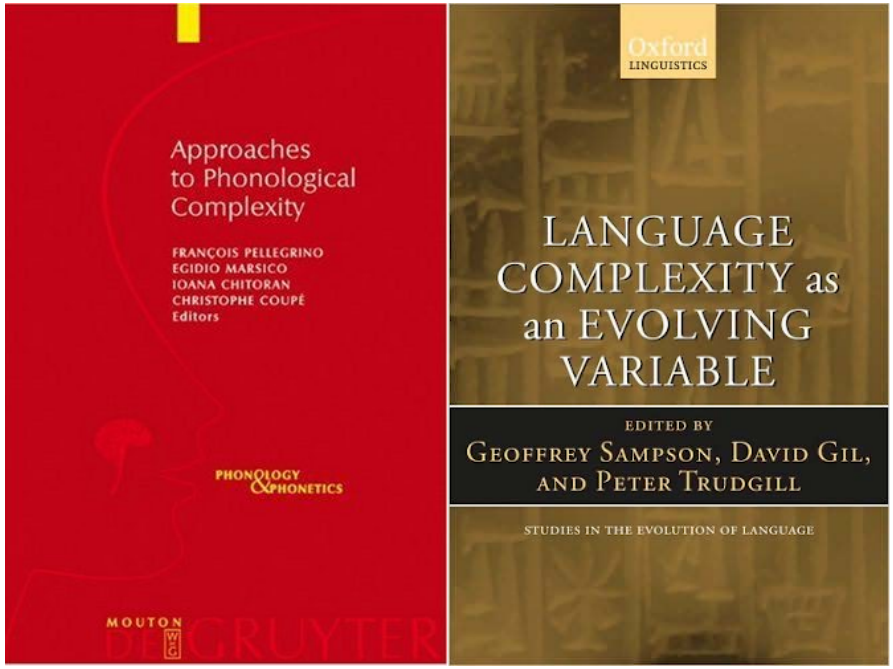
Materials for the analysis

Criticism by [Simpson 1999]

Complexity based approaches

Distance based approaches

[Pellegrino et al. 2009] and [Sampson et al. 2009]



[Pellegrino et al. 2009] and [Sampson et al. 2009]

- [Pellegrino et al. 2009]
  - [Chitoran and Cohn 2009]
  - [Ohala 2009]
  - [Maddieson 2009]
  - [Coupé et al. 2009]
- [Sampson et al. 2009]
  - [Nichols 2009]
  - [Deutscher 2009]

Overview

Materials for the analysis

Criticism by [Simpson 1999]

Complexity based approaches

Distance based approaches

Distance based approaches

- [Hoppenbrouwers and Hoppenbrouwers 2001] (after [Heeringa 2004])
- [Heeringa 2004]
- [Eden 2018] — Hamming distance
- [Anderson et al. 2021] — Jaccard similarity

Thank you for your attention!

References

Anderson, C., Tresoldi, T., Greenhill, S. J., Forkel, R., Gray, R. D., and List, J.-M. (2021). Measuring variation in phoneme inventories (preprint v1). *Research Square*.

Chitoran, I. and Cohn, A. C. (2009). Complexity in phonetics and phonology: gradience, categoriality, and naturalness. In *Approaches to phonological complexity*, pages 19–46. De Gruyter Mouton.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper and Row.

Coupé, C., Marsico, E., and Pellegrino, F. (2009). Structural complexity of phonological systems. In *Approaches to phonological complexity*, pages 141–170. De Gruyter Mouton.

References

Deutscher, G. (2009). "overall complexity": a wild goose chase? In *Language complexity as an evolving variable*, pages 243–252. Oxford University Press.

Eden, S. E. (2018). *Measuring phonological distance between languages*. PhD thesis, University College London.

Heeringa, W. J. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. PhD thesis, University Library Groningen][Host].

Hoppenbrouwers, C. A. J. and Hoppenbrouwers, G. A. (2001). *De indeling van de Nederlandse streektalen: dialecten van 156 steden en dorpen geklasseerd volgens de FFM*. Uitgeverij Van Gorcum.

Maddieson, I. (2009). Calculating phonological complexity. In *Approaches to phonological complexity*, pages 83–110. De Gruyter Mouton.

References

Maddieson, I. and Abramson, A. S. (1987). Patterns of Sounds. *The Journal of the Acoustical Society of America*, 82(S1):720–721.

Moran, S. and McCloy, D., editors (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.

Nichols, J. (2009). Linguistic complexity: a comprehensive definition and survey. In *Language complexity as an evolving variable*, pages 110–125. Oxford University Press.

Ohala, J. J. (2009). Languages' sound inventories: the devil in the details. In *Approaches to phonological complexity*, pages 47–58. De Gruyter Mouton.

Pellegrino, F., Marsico, E., Chitoran, I., and Coupé, C. (2009). *Approaches to phonological complexity*, volume 16. Walter de Gruyter.

References

Sampson, G., Gil, D., and Trudgill, P. (2009). *Language complexity as an evolving variable*. Oxford University Press.

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*, volume 1, pages 349–352.