# Toward Building a More Comprehensive Sound Change Database


Lauren Spradlin

The Graduate Center, The City University of New York

spradlin.lauren@gmail.com



4th Workshop on Sound Change, 19-22 April, 2017


THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

## Why We Need a Sound Change Database

- Historical phonologists have used the comparative method for over 150 years to reliably document sound correspondences and reconstruct sound changes in many of the world's languages and language families.
- Despite the breadth of work done in the field, there currently exists no database (or even paper volume) of the world's recurrent regular sound changes to consult should one want determine something as basic as the overall cross-linguistic frequency a particular sound change or whether a given sound change is more common than another.
  - How cross-linguistically common is word-final (de)voicing?
  - Which is more common: h > s or s > h?
- Historical linguistics texts often provide examples of 'common' and 'rare' types of sound change, but do not elaborate on the cross-linguistic applicability of such claims, or how they arrived at these conclusions.

> "All standard segmental historical phonologists need to develop a feeling for which kinds of changes are 'common' (or 'likely', 'natural' or 'possible') and which are 'unnatural' (or 'impossible'). This feeling is typically developed through a combination of experience, working through the histories of languages, and inherited wisdom, gained through reading textbooks or discussion with others. It's a rather haphazard basis for science: it would be really helpful if there were a systematic collection of types of changes, as found in the history of the languages of the world, ordered on a phonological basis, from which we could extract generalizations on a firmer basis about which changes happen commonly and what kinds of things are not found at all. This book aims to fill this need, and, as a historical phonologist, I'm grateful for it."
>
> -Honeybone in his 2009 review of Martin Kümmel's *Konsonantenwandel* (p. 280)

- Most historical reconstructions document sound changes within the context of a given language, subgroup, or language family, often without considering possible phonetic bases for such changes.
- Typologies of individual sound patterns can be found in the phonological literature, and are often informed by phonetics, but they are rarely discussed in terms of other similar sound changes.

A database could also aid in the investigation of the notable similarities between common types of historical sound change and common synchronic sound patterns, recurrent synchronic sound patterns observed in unrelated languages that are argued to share regular phonetically based sound change as a common source, and their potential parallel evolution constrained by the perceptual and articulatory properties of human speech claimed to be the force underlying their independent development (Evolutionary Phonology; Blevins 2004).

Based on the groundwork laid for the Handbook of Phonological Change, the current project's goals include:

- collecting as much empirical data as possible on the phonetic basis of sound change
- creating an electronic, publicly accessible database of regular sound change
- making the database flexible in structure, to accommodate questions asked with different starting assumptions (e.g., genetic relationships, feature systems, phonetic quality of segments, definitions of natural classes)

## Existing Databases

### Sound Change Databases

UNIDIA (Ben Hamed & Flavier 2009)

- A database for deriving diachronic universals
- "a data-based typology of sound change, universals, tendencies and sound change distributions"
- 10,349 sound changes, 302 languages

CBOLD (Comparative Bantu Online Dictionary; UC Berkeley)

- a lexicographic database to support and enhance the theoretical, descriptive, and historical linguistic study of the languages in the important Bantu family
- Contains a list of reconstructed Proto-Bantu roots and reflexes in 500+ daughter languages

### Sound Pattern Databases

PBase (Mielke 2008)

- A database of (synchronic) phonological patterns
- 7318 patterns, 629 languages, 8 feature systems, 398 features

## Proposed Database

### Data Entry: Quality Over Quantity

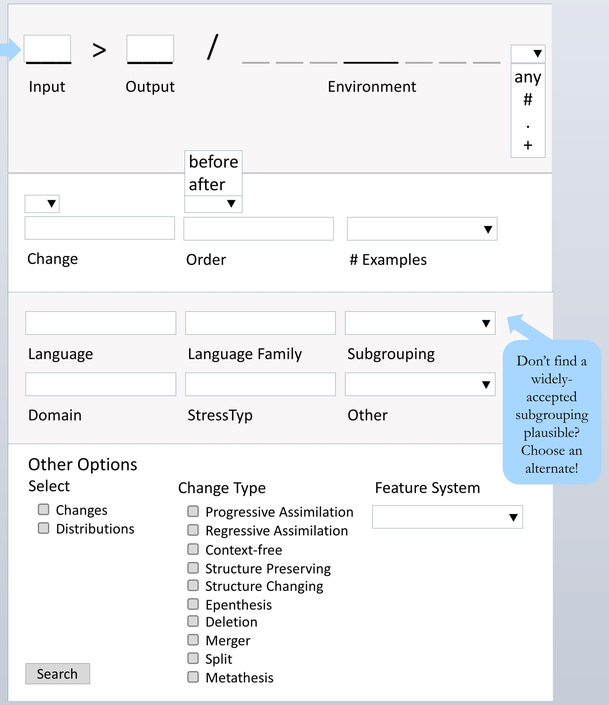
Many databases include as much information as possible, and leave vetting the quality of the search results to the user. We plan to take a different approach by:

- Entering only regular, exceptionless sound changes
- Including the best-documented language families and subgroups
- Relying on language experts when determining the relevant regular sound changes
- Noting how many examples of each sound change are attested

### Searchability

Users can investigate sound changes and inventories by general type, phonetic basis, target segment or structure, output segment or structure, language, language family, and other more specific attributes.

Specifiable by IPA character, phonetic content, orthographic character consistent with the language's tradition, natural classes definable using the feature systems available in PBase




### Search Results

A search will return a list of sound changes meeting the specified criteria from which the user can further investigate by clicking on the language, language family, specific sound change, or other search results.

Show 10 ▼ entries — Showing 1 to 1 of 1 entries — ◄ Previous Next ►

| Language | Lang Fam | Type | Description | Inputs | Outputs | Environments | # ex. | Source |
|---|---|---|---|---|---|---|---|---|
| Germanic | PIE | Change | m → n / __ # | m | n | __ # | 20+ | Fortson 2004 |

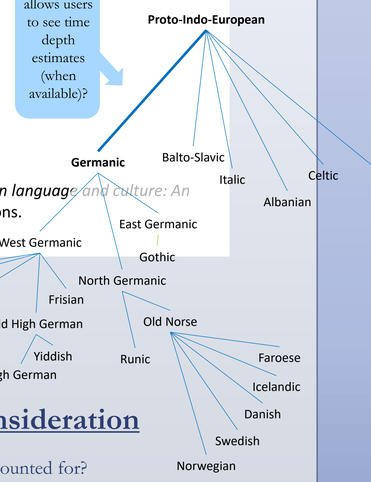
### Individual Sound Changes

| | bilabial | dental | alveolar | palatal | velar | labial-velar |
|---|---|---|---|---|---|---|
| nasal | m | | n | | | |
| stop | p b | t d | | | k g | $k^w$ $g^w$ |
| fricative | ɸ | θ | s z | | x | $x^w$ |
| approximant | | | | j | | w |
| lateral | | | l | | | |
| trill | | | r | | | |

| | |
|---|---|
| Language | Germanic |
| Language Family | Proto-Indo-European |
| Type | Sound Change |
| Description | *m > n / __ # |
| Input – Output | *m [+nas +lab -syll] – n [+nas -lab -syll] |
| Changes | -lab (or +alv) |
| Environment | 0 m – 1 # |
| Ordered? | No |
| Source | Fortson IV, B. W. (2011). *Indo-European language and culture: An introduction* (Vol. 30). John Wiley & Sons. |
| Number of examples | 20+ |

Including a phonemic inventory with each change could be helpful of lull users into a false sense of security.

Or with a tree that allows users to see time depth estimates (when available)?




Alerting users to when one change is part of a series of ordered changes is an unresolved seemingly complicated issue…

## Details Still Under Consideration

How should language / language family-specific quirks be accounted for?

- The Proto-Indo-European laryngeals
- Coding Proto-Malayo-Polynesian *N for phonetic qualities
- The same orthographic symbol is used for distinct (proto-)phonemes in different language families

What is the best way to represent ordered sound changes? Changes to inventories over time?

If it's possible, would visual relative chronologies of changes be helpful?

## Contact Information & Questionnaire

I welcome any and all input WSC4 attendees may have regarding this project. If you have expertise or experiences you'd like to share, please feel free to email me at spradlin.lauren@gmail.com.

Your opinions matter (a lot)! Please take the survey!
https://goo.gl/forms/AuN8B7gNWvSdBy3u1

Additionally, please take a moment to answer a questionnaire about what you'd like to see in a sound change database as a user. It's available at https://goo.gl/forms/pGSW3BcLgKCkPYGo2.

## Selected References


**Ben Hamed, M., and S. Flavier. 2009**. UNIDIA: A database for deriving diachronic universals. Historical linguistics 2007: Selected papers from the 18th International Conference on Historical Linguistics, Montreal, 6–11 August 2007, ed. by Monique Dufresne, Fernande Dupuis, and Etleva Vocaj, 259–68. Amsterdam: John Benjamins. Bickel, Balthasar. 2002. The AUTOTYP research program. Presented at the annual meeting of the Linguistic Typology Research Center, Utrecht, September 26-28. **Blevins, J. 2004**. Evolutionary phonology: The emergence of sound patterns. Cambridge: Cambridge University Press. **CBOLD** (http://www.cbold.ish-lyon.cnrs.fr/) **Dryer, M. S. & M. Haspelmath. (eds.) 2013.** The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info, accessed 2016-10-15.) **Kümmel, M. 2007**. Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion. Reichert Verlag. ISBN 978-3-89500-590-9. **Maddieson, Ian and Karen Precoda. 1989**. Updating UPSID. UCLA Working Papers in Phonetics 74:104-111. **Maddieson, Ian. 2015**. Constructing a global cross-linguistic database of basic phonological properties: Principles and challenges. Proceedings of the 18th International Congress of Phonetic Sciences. Martinet, André. 1955. Économie des changements phonétiques. Berne: A. Francke. **Maddieson I., S. Flavier, E. Marsico & F. Pellegrino. 2015**. LAPSyD: Lyon-Albuquerque Phonological Systems Database. (Available online at http://www.lapsyd.ddl.ish-lyon.cnrs.fr/lapsyd, accessed 2016-10-15.) **Michael, L., T. Stark, and W. Chang (compilers). 2012**. South American Phonological Inventory Database v1.1.3. Survey of California and Other Indian Languages Digital Resource. Berkeley: University of California. (Available online at http://linguistics.berkeley.edu/~saphon/en, accessed 2016-10-15.) **Mielke, J. 2008**. PBase. (Available online at http://pbase.phon.chass.ncsu.edu, accessed 2016-10-15.) **Mielke, J. 2010**. The frequency of segmental alternations: implications for sociolinguistic variation. Poster presented at NWAV, 39. **Moran, S., McCloy, D. & R. Wright (eds.) 2014**. PHOIBLE Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://phoible.org, accessed 2016-10-15.)