[Abstract submitted for presentation as a talk]

The dialexical set: a diagnostic tool for studying sound change

The study of sound change in comparative linguistics includes many familiar tools and outputs such as sound correspondences, cognate sets, and reconstructed phonemes. In this paper I propose a broadly generalizable new concept for sound change comparison called the ***dialexical set***, defined as a diagnostic set of specific etyma which pattern together regularly within a given individual lect, but vary across related lects.

The *dialexical set* differs from existing comparative concepts as follows: **(1)** *Cognate sets* focus on shared form-meaning correspondence; by contrast, *dialexical sets* are for studying sound change, and compile etyma and associated reflexes, regardless of semantic change. A given etymon may belong to one cognate set in a given language, but to several dialexical sets for each constituent sound. **(2)** *Reflex sets* are attestations associated with cognate sets, but do not necessarily entail any regularity across reflexes. The reflex set of some etymon includes all forms descended from it, whether regular or idiosyncratic, and is often used to reconstruct the proto-form of that etymon; whereas *dialexical sets* are clusters of etyma that pattern together, entailing regularity, and thus useful as diagnostics for studying sound change. (3) *Sound correspondences* are a form of analysis of the surface outcomes of sound change, but identifying them also often precedes the hypothesizing of etyma. They are derived from the examination of surface forms, but unlike *dialexical sets*, sound correspondences are not always tied back to specific etyma, especially not when published, and thus are not immediately extensible to additional related languages. (4) *Etysets* (Cooper 2014) are a type of distributable dataset, 'cognate sets, phylogenetic trees, and reconstructed proto-forms', and thus could include *dialexical sets* but are distinct from them.

The notion of the *dialexical set* is inspired by insights from two distinct comparative traditions: English ***lexical sets*** (Wells 1982) and Tai ***tone boxes*** (Gedney 1972). Wellsian lexical sets are 'large groups of words which tend to share the same vowel' and represented by a keyword. For example, the set KIT comprises 'those words whose form in the two standard accents [RP and General American English] has the stressed vowel /ɪ/' (1982:127). While not an explicitly diachronic concept, lexical sets neatly reflect and reveal sound changes in the many varieties of English across the globe. Tai tone boxes, on the other hand, group etyma based on shared segmental environments that conditioned modern tone categories (e.g. voiced vs. voiceless proto-onsets). Each cell represents some combination of proto-tone and natural class of proto-onsets; it comprises a set of etyma that pattern together tonally across the set of Tai lects. Reflexes of these etyma share a modern tone in a given Tai language, regardless of subsequent changes to onsets.

We can recast some existing tasks using this idea. Reconstructed phonemes can be treated as the labels of hypothesized *dialexical sets*. For example, a hypothesized proto-phoneme *k can also be taken as the label of a set of etyma that pattern together in the languages under comparison. This is true regardless of whether all reflexes of *k stayed as /k/, changed to another sound, or split into a variety of sounds. Reconstructed sounds may represent different etyma in different positional contexts, but each reconstructed sound can be viewed as a label representing some set of etyma. When viewed in this light, we can view the discovery and labeling of *dialexical sets* as one major task of language comparison.

The major benefit of *dialexical sets* is that they are **clusters of specific etyma**. Unlike, say, tables of reconstructed phonemes, or lists of sound correspondences, defining a dialexical set entails listing the set of etyma that it comprises (as Wells and Gedney both did). This enables immediate diagnostic usefulness to other researchers wishing to study additional related lects at any level of granularity (language, dialect, sociolect, idiolect, etc), without needing necessarily access to all of the original reflex data available to the linguist(s) who proposed the dialexical sets. This thus also addresses a shortcoming of much published comparative work.

[Abstract submitted for presentation as a talk]

References

Cooper, Doug. 2014. Data Warehouse, Bronze, Gold, STEC, Software. In *Proceedings of the Workshop on the Use of Computational Methods in the Study of Endangered Languages* (ComputEL-1), Baltimore, 91-99.

Gedney, William J. 1972. A checklist for determining the tones in Tai dialects. In *Studies in Linguistics in Honor of George L. Trager*, 423-437. Mouton.

Wells, J. C. 1982. *Accents of English 1: An Introduction*. Cambridge University Press.