Evidentiality as an areal feature

an East Caucasian perspective on typology vs. reality

Evidentiality (i.e. the marking of information source) as a grammatical category is considered an areal feature in different parts of the world. Similar systems are found in contiguous areas, and the category seems highly borrowable in contact situations. Ideas about particular evidential areas, however, often remain impressionistic and detached from geographical data. Large-scale typological studies do take into account geography, but tend to simplify linguistic reality by reducing clusters of heterogeneous speech communities to a single dot representing a language and a discrete value.

 The present paper explores how much detail areal typology requires to paint an accurate picture, by comparing three types of data and the linguistic maps they produce: 1) a subset of a large-scale typological study – two maps from WALS (de Haan 2013a, b); 2) a more detailed inquiry into evidentiality in the languages of the eastern Caucasus; and 3) a dataset of all East Caucasian villages and the languages spoken there.

 The WALS maps show a general division between the northwestern part of the area (where evidentiality is attested as part of tense) and the southeastern part (where evidentiality is expressed with special affixes or clitics). The second dataset provides a clearer picture: evidentiality as part of tense is widespread in the northwest but almost completely absent in the southeast, while affixes and clitics are spread throughout the area. The results do not contradict the general pattern revealed by the WALS sample but importantly, the pattern is not limited to the East Caucasian family. It also extends to local Turkic languages. The villages dataset revealed a lack of data for villages in the transitional zone, so the exact range of both areas remains to be ascertained.[[1]](#footnote-2)

 Moroz (2019) suggested that taking samples is not a viable option for linguistic data, due to gaps in knowledge and other biases that prevent balanced sampling. At least in the case of evidentiality in the East Caucasus, a sample of 10 languages proved sufficient to reveal the main patterns in the area. However, that is all it does. I argue that sample-based inquiries should be updated with more data and specialist knowledge to corroborate and refine patterns, and ultimately explain where the observed patterns come from.

**References**

de Haan, Ferdinand. 2013a. ‘Coding of evidentiality’. In: Matthew S. Dryer & Martin Haspelmath (eds.) *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/78, accessed on 2019-09-27)

de Haan, Ferdinand. 2013b. ‘Semantic distinctions of evidentiality’. In: Matthew S. Dryer & Martin Haspelmath (eds.) *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/78, accessed on 2019-09- 27)

Moroz, George. 2019. *Analysing phonological systems: on Bayesian typological research*. SLE, 21- 24 August 2019, University of Leipzig. (Presentation)

1. This is relevant because the border might coincide with a border between two distinct languages, or it could divide a single language into idioms that have the feature and those that lack it. [↑](#footnote-ref-2)