

A cross-linguistic comparison of content interrogatives

Siyu Liu

The Philipps University of Marburg

Liusi@students.uni-marburg.de

Almost every human language uses a universal class of function words, content interrogatives, to ask for missing information. However, it is broadly attested that content interrogatives differ across languages at nearly all linguistic levels (cf. Dixon 2012:400-420). Compared to previous studies that normally put the emphasis on the syntax, e.g., *wh*-movement or word order in interrogative clauses, the semantic properties of content interrogatives have much less been the focus of typologists.

This study aims at investigating the cross-linguistic diversity of semantic features of content interrogatives and identifying typical instantiations of each interrogative category. The investigation employs the approach of a *Massively Parallel Text*, which particularly refers to a text that is available in translations in a large number of languages (Cysouw & Wälchli 2007:95). Based on a massively parallel text, linguistic structures can be surveyed in a consistent context across languages with ample contextually-embedded instantiations throughout the text. The data of this research is collected from translational equivalents of the New Testament in the *Parallel Bible Corpus*. Instead of the traditional way to manually conduct a comparison, instantiations of content interrogatives are automatically classified into different groups on the basis of the (dis-)similarity via the quantitative method of *Partitioning Around Medoids* (Kaufman & Rousseeuw 2009:69-72). The contextual distribution of content interrogatives is then visualized in the graphical form by the technique of *Multidimensional Scaling* based on which patterns within the dataset can be interpreted (van der Klis & Tellings 2022).

It is expected that the underlying semantic distinctions will be uncovered by observing the contextual realization of content interrogatives. Moreover, by showing which contexts are prone to be identically denoted in most different languages, it is possible to generalize representative exemplifications of content interrogative categories.

Selected references:

- Cysouw, M. (2004). Interrogative words: an exercise in lexical typology. *Bantu grammar: description and theory*, 3, ZAS Berlin. URL: http://cysouw.de/home/manuscripts_files/cysouwQUESTION_handout.pdf.
- Cysouw, M. & Wälchli B. (2007). Parallel texts: using translational equivalents in linguistic typology. In *Language typology and universals* 60(2), 95-99.
- Dixon, R. M. (2012). *Basic linguistic theory volume 3: Further grammatical topics*. Oxford University Press on Demand.
- Haspelmath, M. (2010). Comparative concepts and descriptive categories in cross-linguistic studies. In *Language* 86 (3), 663-687.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.
- Wälchli, B. (2010). Similarity semantics and building probabilistic semantic maps from parallel texts. In *Linguistic Discovery*, 8(1), 331-371.
- van der Klis, M., & Tellings, J. (2022). Generating semantic maps through multidimensional scaling: linguistic applications and theory. In *Corpus Linguistics and Linguistic Theory*.