

Grammar writing, documentation, data collection

Convenor: Aimée Lahaussais

This section focuses on recent advances in practices underlying grammar writing, documentation, and data collection, and the interconnections between them.

Grammar writing is a well-theorized discipline (e.g. Ameka, Dench & Evans 2006; Noonan 2007; Payne & Weber 2007; Nakayama & Rice 2014) but the subfield of **metagrammaticography** (Lehmann & Maslova 2004) is still very young, with questions of how grammar-writing decisions affect the use made of resulting grammars (for language comparison, for typology) beginning to come to the fore. Questions of how grammars are organized and the impact on how the data is interpreted by readers (Cristofaro 2006; Kelly & Lahaussais 2021), of the place of traditionally more marginal word classes, such as interjections and ideophones (Dingemanse 2018; Heine 2023), of the place of diachronic data or commentary in a descriptive (and traditionally synchronically-oriented) grammar (e.g. Rankin 2006; Post 2009), of what constitutes a representative dataset for grammar writing (e.g. Good 2012; Mosel 2012), of what statements about productivity actually mean, all represent interesting avenues for research. A corollary question concerns the reasons for the inherent challenges in using descriptive grammars when carrying out areal or typological research.

The **push towards open access** has shaped grammar-writing, documentation and data collection in very concrete ways. There is a growing expectation that examples in grammars and other types of descriptive documents be accessible through and linked to oral archives, with time-aligned sound and annotation files; publicly funded institutions and grant agencies are increasingly insistent that materials be made available, both in the form of primary data and the associated analyses, the latter typically through open access grammars (such as those published by Language Science Press) and open access journals (such as those focusing on descriptive and methodological questions, like *Language Documentation and Conservation*, and typology, like *Linguistic Typology at the Crossroads*).

Developments in the tools that accompany the descriptive and documentation process are also undergoing advances: to the traditional toolkit of **word lists and questionnaires** we can add **stimuli** carefully informed by typological and psycholinguistic advances, as well as **video recordings** of field sessions, making it possible to carry out multimodal studies featuring gesture. **Annotation** has also been affected: Despite widespread adoption of the Leipzig Glossing Rules, which greatly facilitates the accessibility of interlinear glossed texts, increased consistency of glossed material is needed to make it usable for language comparison and machine readability (List, Sims & Forkel 2021; Chelliah, Burke & Heaton 2021). Efforts towards the automatization of annotation, through methods making it possible to automatically produce phonemic transcriptions of audio files in the field (e.g. Wisniewski, Michaud & Guillaume 2020),

and the use of interlinear glosses to generate grammars (Bender et al. 2013), represent remarkable steps forward. And *electronic grammaticography* (cf. Nordhoff 2012) and its associated methods (which circle back to changes in expectations about open access of both analyses and data) will continue to lead to major headway in the production of grammars.

Data collection has been impacted and changed by recent public health and political crises, which have made it increasingly difficult to reach some field sites. This has resulted in the development of new techniques for *hybrid fieldwork*, and has also increased attention paid to using *legacy data sources* on minority languages, often necessitating collaboration with historians of linguistics for contextualization.

The section welcomes papers in any of these areas as well as papers which explore the interconnections between them, particularly those dealing with these questions from the point of view of under-resourced languages.

References

- Ameka, Felix, Alan Dench & Nicholas Evans (eds.). 2006. *Catching Language: The Standing Challenge of Grammar Writing* (Trends in Linguistics 167). Berlin & New York: Mouton de Gruyter.
- Bender, Emily M., Michael Wayne Goodman, Joshua Crowgy & Fei Xia. 2013. Towards Creating Precision Grammars from Interlinear Glossed Text: Inferring Large-Scale Typological Properties. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 74–83. Sofia, Bulgaria: Association for Computational Linguistics.
- Chelliah, Shobhana, Mary Burke & Marty Heaton. 2021. Using interlinear gloss texts to facilitate cross-language comparison and improve language description. *Indian Linguistics* 82(1–2). 1–24.
- Cristofaro, Sonia. 2006. The organization of reference grammars: a typologist user's point of view. In Felix Ameka, Alan Dench & Nick Evans (eds.), *Catching Grammar: The Standing Challenge of Grammar Writing*, 137–170. Berlin & New York: Mouton de Gruyter.
- Dingemanse, Mark. 2018. Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics* 3(1). 1–30. <https://doi.org/10.5334/gjgl.444>.
- Good, Jeff. 2012. Deconstructing descriptive grammars. In Sebastian Nordhoff (ed.), *Electronic Grammaticography* (Language Documentation & Conservation Special Publications 4), 2–32. Honolulu: University of Hawai'i Press.
- Heine, Bernd. 2023. *The grammar of interactives*. Oxford: Oxford University Press.
- Kelly, Barbara & Aimée Lahaussais. 2021. Chains of influence in Himalayan grammars: models and interrelations shaping descriptions of Tibeto-Burman languages of Nepal. *Linguistics* 59(1). 207–245. <https://doi.org/10.1515/ling-2020-0255>.

- Lehmann, Christian & Elena Maslova. 2004. Grammaticography. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie. Ein Handbuch zur Flexion un Wortbildung* (Handbücher Des Sprach- Und Kommunikationswissenschaft 17/2), 1857–1882. Berlin & New York: W. de Gruyter.
- List, Johann-Mattis, Nathaniel A. Sims & Robert Forkel. 2021. Towards a sustainable handling of interlinear-glossed text in language documentation. *ACM Transactions of Asian Low-Resource Language Information Processing* 20(2). 15 pages. <https://doi.org/10.1145/3389010>.
- Mosel, Ulrike. 2012. Advances in the accountability of grammatical analysis and description by using regular expressions. In Sebastian Nordhoff (ed.), *Electronic Grammaticography* (Language Documentation & Conservation Special Publications 4), 235–250. Honolulu: University of Hawai'i Press.
- Nakayama, Toshihide & Keren Rice (eds.). 2014. *The Art and Practice of Grammar Writing* (Language Documentation and Conservation Special Publication 08). Honolulu: University of Hawai'i Press.
- Noonan, Michael. 2007. Grammar Writing for a grammar-reading audience. In Thomas Payne & David Weber (eds.), *Perspectives on grammar writing* (Benjamins Current Topics 11), 113–126. Amsterdam/Philadelphia: John Benjamins.
- Payne, Thomas & David Weber (eds.). 2007. *Perspectives on grammar writing* (Benjamins Current Topics 11). Amsterdam/Philadelphia: John Benjamins.
- Post, Mark W. 2009. The phonology and grammar of Galo “words”: A case study in benign disunity. *Studies in Language* 33(4). 934–974.
- Rankin, Robert L. 2006. The interplay of synchronic and diachronic discovery in Siouan grammar-writing. In Felix Ameka, Alan Dench & Nick Evans (eds.), *Catching Grammar: The Standing Challenge of Grammar Writing*, 527–547. Berlin & New York: Mouton de Gruyter.
- Wisniewski, Guillaume, Alexis Michaud & Séverine Guillaume. 2020. Phonemic transcription of low-resource languages: To what extent can preprocessing be automated? In *Proceedings of the 1st Joint SLTU and CCURL Workshop (SLTU-CCURL 2020)*, 306–315. European Language Resources Association (ELRA).