

Reflections on the role of language documentations in linguistic research

Stefan Schnell

University of Bamberg

Centre of Excellence for the Dynamics of Language

I reflect the role of language documentations in linguistic research beyond its most common linguistic use as a high-quality database for descriptive work. I show that the original Himmelmann-ian conception of documentations, as multi-varied and multi-purpose, and to some extent community-driven, enable a range of research outcomes that would not have been foreseeable within the traditional descriptive, typological and theoretical agendas. I argue that it is overall more fruitful for innovative linguistic research to invest into the processing of haphazard language documentation data rather than attempting to collect precisely the kind of data demanded by specific analytic goals.

1. Introduction¹ According to Himmelmann (1998) language documentations are ideally prepared not in service of any specific analytic agenda, but as broad multi-varied collections open to a variety of purposes and uses by different user groups, including speech communities and academic linguists. In this contribution I show in what ways data that are not controlled for any research purpose can play an important role in research outside traditional descriptive grammar writing. I show that documentation-based research always involves considerable efforts in additional or alternative processing of different types of data (Himmelmann 2012), but that it often does not need to involve further collection of more *specifically useful* data in order to play its increasingly important role in linguistics.

In what follows, I outline first how language documentations are seen to be empirically valuable (Section 2). I then summarize some prominent research agendas that traditionally

¹The ideas developed in this contribution have been developed while I was a postdoc researcher in the *ARC Centre of Excellence for the Dynamics of Language*. I am grateful for the ongoing support by the Centre. I would also like to thank Rebecca Defina, Cris Emmonds-Wathern, Jenifer Green, Yukinori Kimoto, as well as Lauren Gawne, Bradley McDonnell and one anonymous reviewer for helpful comments on an earlier draft of this paper. All remaining errors and shortcomings are my own responsibility.

build on rich performance data (Section 3). In Section 4 I turn to more recent research based on entire language documentations that has resulted in significant insights into language use across languages. In Section 5 I discuss the role of different text varieties in typologically oriented documentation-based research. I conclude my contribution in Section 6 with some reflections on further developments in corpus-based typology.

2. Empirical value of language documentations The most obvious and central value of language documentations for academic linguistics is that it provides an empirical basis on which linguistic analyses are accountable by way of giving access to the recorded data and their annotation (Himmelman 2006; also Gawne & Berez-Kroeker 2018 in this volume; Berez-Kroeker et al. 2018 for recent discussion). Aside from this more global value of accountability, different components of a documentation, "...a comprehensive record of the linguistic practices of a given speech community" (Himmelman 1998:166), are more or less useful in reflecting different aspects of these practices, in particular observable behavior and metalinguistic knowledge. Hence, Himmelman (1998) advocates aiming at a broad collection of data resembling different degrees of naturalness and spontaneity. Thus, a casual conversation recorded with little awareness of participants will be the best representation of naturally, and most frequently, occurring observable linguistic behavior in a speech community. Metalinguistic knowledge, on the other hand, is often not reflected in such recordings and specialized elicitations of, for instance, morphological paradigms together with comments on similarities of forms, etc would capture this instead, but not represent any naturally occurring speech event. Elicitation sessions, as well as other data that do not resemble any established communicative routines, like stimuli-based elicited texts, are on the other hand characterized by a high degree of spontaneity, which may reveal interesting aspects of a language systems otherwise not represented in more naturalistic data. A further dimension concerns the relevance of documentary activities for a given speech community: elicitations of a morphological paradigm are obviously not of any major concern for communities, but neither are more natural casual conversations. Instead, collection of different forms of verbal art, indigenous oral literature as well as encyclopedic knowledge of flora, fauna, material culture and so forth, and respective vocabulary, is often among the major desiderata of a given community (Himmelman 2006; Mosel 2014a).

Similarly, different types of data play different roles in linguistic research. Casual conversations resemble most accurately how a language is used at a given point in time in a community, and this data is ultimately crucial for a thorough understanding of language change and possible developments of evolutionary models thereof (e.g. Baxter & Croft 2016; Blythe & Croft 2012). It would be much less useful for a first descriptive account of a language which requires examples of complete, well-formed constructions which can be hard to come by in conversational data. Narrative texts from oral literature may be a much better data source for this purpose. Elicitations of specific structures can provide data most relevant for descriptions in the most immediate way, including data often not attested in any less controlled data type (see Evans 2008; Rhodes et al. 2006), but they are hardly ever really useful for studies in language variation and diversification. To what extent different types of data are restricted to very specific linguistic purposes or open to a variety thereof is discussed in detail in McDonnell (2018, in this volume). My main concern here is to show how documentations that are mainly concerned with the coverage of linguistic practices and the desires of speech communities can play and have played an important role in linguistic research beyond the traditional descriptive paradigm.

3. Traditional research on language use: in search for the right data Performance data have long been a focus of dedicated research traditions in linguistics, for instance variationist sociolinguistics (Meyerhoff 2010), general corpus linguistics (Biber & Conrad 2009), or conversation analysis (Seedhouse 2013), among many others. Of particular relevance for typological linguistics has been a line of research that DuBois (2017) calls “discourse and grammar”: established by Wallace Chafe and Talmy Givón in the 1970ies, it is concerned with patterns of reference and information packaging in discourse and seeks to explain these with reference to cognitive factors of language processing (e.g. Givón 1976; Chafe 1976; DuBois 1987 among many others). On the other hand, identified patterns in discourse are considered the seedbed of grammatical structures which *emerge* through frequent deployment of discourse patterns during communication, hence the emergentist credo that “grammars do best what speakers do most” (Du Bois 1985). Grammar and discourse is closely related to the tradition of *language variation and change*, where language-internal and -external (i.e. social, cultural, etc) factors are related to regularities of language use and resulting diachronic developments (Labov 1994; Croft 2000).

A major challenge for these research traditions has been to determine what kind of performance data is required, in line with their respective goals, for instance sociolinguistic interviews in sociolinguistics, etc. Finding appropriate performance data has been, and continues to be, a particular challenge in more typologically oriented research, like that in grammar and discourse: these research agendas require records of connected discourse from as many languages as possible, comparable, at least to a certain degree. A common response to this challenge is to use stimuli-based elicited narrative texts, most notably Chafe’s (1980) Pear Film, or Mayer’s (1969) Frog stories (see Slobin 2004), which ensure a minimal degree of comparability of different texts on each occasion of their elicitation. While such elicited texts can yield interesting observations on possible structures of a language system due to their high degree of spontaneity, they do not capture natural routines of linguistic performance (see Foley 2003 for critical discussion of the use of Frog stories in Watam). In extreme cases, they hardly resemble any kind of coherent discourse at all, as reported by DuBois (1980) for the elicitation of Pear stories in Sakapultek. It seems to me that for typologically oriented studies of language use, we are still exploring what the ideal dataset looks like, and in the following I will show that language documentations have a great deal to contribute to this quest.

4. A found treasure: language documentations in usage-based linguistic research

Although Himmelmann (1998) mentions potential uses of language documentations outside the standard grammaticographic line, specific research projects of this kind drawing extensively on documentation data started to take off not before about ten years ago or so. It is worth mentioning that the DoBeS program dedicated its final fully-fledged round of funding almost entirely to projects utilizing existing collections in broader research projects. Examples are Frank Seifart’s project² on the ratio of nouns, pronouns and verbs in spoken-language discourse, Anna Margetts’ project³ on three-participant

²DoBeS research project *The relative frequencies of nouns, pronouns, and verbs cross-linguistically* (PI Frank Seifart, 2012-2015),

³DoBeS research project *Cross-linguistic patterns in the encoding of three-participant events* (PI Anna Margetts, 2012-2016) and *Cross-linguistic patterns in the encoding of three-participant events—investigating BRING and TAKE* (PI Anna Margetts, 2017-2018)

constructions across languages, and Claudia Wegener's project⁴ on prosodic patterns in discourse structure in contact situations between languages from two different families. Not specifically funded by DoBeS, but developed in its context is Geoffrey Haig's and my own project on referential choice and argument realization in discourse (Haig & Schnell 2014, 2016a,b).

The great potential of documentation-based research lies in its focus on aspects of language production that are not typically part of structuralist descriptive and typological work and in the embeddedness of performance data in the cultural context of speech communities. Both of these aspects have to teach us a lot about how languages are used, and how this may influence their evolution. For instance, Himmelmann (2014) draws upon extensive spoken language data to bear on the long-standing challenge of explaining the suffixing preference in the languages of the world. His study identifies a systematic distribution of dysfluencies and pausing in spoken discourse that corroborates specifically constrained structural contexts for the development of affixal exponents of grammatical categories, hence explaining the typological preference. Seifart et al. (2018) show that across languages, the production of noun phrases affords more planning effort, the latter being determined by proxy measurements of pause probabilities and speech rate. The authors attribute this higher effort to the particular referential choices associated with noun phrase production, a conclusion of major relevance to questions of referential choice and language processing in general. Further prominently published documentation-based studies are Margetts (2015) and Haig & Schnell (2016b).

These examples bear witness of the fact that the role of documentation-based research is gaining ground in academic linguistics. They also seem to yield some methodological insights that are important for future developments: for one thing, documentation-based research of this kind involves considerable efforts of further processing of existing data. Seifart et al. (2018), for instance, draw on data with word-level time-aligned transcriptions and further annotations that required the development of forced (time-)alignment methods (Strunk et al. 2014). Haig & Schnell's (2016b) extensive corpus study on argument realization draws to a large extent on a multilingual corpus annotated for specific morphosyntactic and semantic features of syntactic arguments (Haig & Schnell 2014). This required the development and monitored implementation of annotation guidelines that are applicable to diverse languages. Similar kinds of annotation guidelines have been implemented in Margett's three-participant project (Margetts et al. 2017). These observations counteract occasional ideas that linguistic analysis could in some way just fall out of documentations, as long as these are well-structured and well-curated. Instead, analytical documentation-based work seems to always come with research-specific additional efforts of data processing.

For another thing, it seems obviously worth pursuing research on data that has not been collected for specific research goals, and has thus not been controlled for in relevant ways. The relevant aspect is that the data resemble real, usually spoken, communication between speakers of diverse communities, and this can be of any kind in order to yield relevant research findings.

5. The usefulness and utilization of original and introduced text varieties From a more corpus-linguistic perspective though, paying attention to the characteristics of

⁴DoBeS research project *Discourse and prosody across language family boundaries: two corpus-based case studies on contact-induced syntactic and prosodic convergence in the encoding of information structure* (PI Claudia Wegener, 2011–2013)

specific text varieties is vital. I will first discuss the example of a variationist study in Vera'a that draws on different types of documentation data, and then turn to recent developments in the field of corpus-based typology, where the use of stimuli-based versus original text data is a major concern.

5.1 Utilization of minimally varied corpora in variationist studies In a study of object realization in Vera'a (Oceanic, North Vanuatu) (Barth & Schnell 2018), we drew on a sub-corpus of the overall language documentation which resulted from a fairly typical documentation project within the DoBeS program.⁵ The alternation we were interested in was that between a pronoun versus zero as a form of realization for those objects that are not a full noun phrase. We investigated spoken narratives as well as descriptions of both floral and faunal species, so that the texts in our corpus resemble two different registers (narration, description) with three different ontological classes of global discourse topic (humans, fish, plants). We find that the best predictor for the use of a pronoun is the global discourse topicality of the referent in question, being either the human protagonists in stories, or the fish and plant species under discussion, thus refining Schnell's (2012) treatment of the alternation as an animacy effect. Only a cross-register analysis of this kind, together with the implementation of sophisticated statistical methodology, made it possible to disentangle the notoriously converging dimensions of animacy and global topicality. Again, this study involved a considerable amount of meticulous corpus annotation work with GRAID (Haig & Schnell 2011) and subsequent further coding of data. This annotated Vera'a corpus is being archived as part of Multi-CAST (Haig & Schnell 2014) with the *Language Archive Cologne* (LAC), thus ensuring reproducibility of this study in the sense of Gawne & Berez-Kroeker (2018, in this volume).

Our inclusion of descriptive texts in our corpus investigation was motivated by my fairly random observations of pronoun use during data processing. However, to collect such data in the first instance was not motivated by our study at all, but followed from the design of the preceding documentation project where a team of researchers from various disciplines and local language workers aimed to document a large range of communicative events and various cultural aspects of two speech communities, including encyclopedic knowledge and associated vocabulary (and folk taxonomies) of flora and fauna, material culture, social organization etc. In accordance with the interests of all participants, we collected not only oral literature (and produced written editions thereof, Vorës & Schnell 2012), but also descriptions of flora and fauna, and their names and taxonomy information. These collections served as a basis for dedicated community materials, akin to the materials for the Teop language of North Bougainville produced by Ulrike Mosel and collaborators (e.g. Mosel et al. 2010; Mosel 2014b,c). It is important to note that descriptive (as well as procedural) texts are not an established genre in Vera'a linguistic culture, and are in this sense not natural; relevant information is traditionally conveyed only by means of demonstration. Their collection was motivated entirely by the aim of documenting the ethnobiological knowledge contained therein. In conclusion, it is possible to arrive at typologically highly significant results by exploring those kinds of data that come up for different reasons during a documentation project.

⁵DoBeS documentation project *Documentation of Vurës and Vera'a, the two surviving endangered languages of Vanua Lava, Vanuatu* (PI Catriona Hyslop) and *Documenting biocultural diversity in the languages of Vurës and Vera'a* (PI Catriona Malau, 2009-2011).

5.2 Original and introduced text varieties in corpus-based typology The typological field most clearly concerned with language use is *corpus-based typology*, a relatively recently emerging field that seeks to determine cross-linguistic commonalities in language use as well as respective diversity. Some of the pioneering work in corpus-based typology continues (and considerably improves on) the Chafe-Givón tradition introduced above. For instance, Bickel (2003) and Stoll & Bickel (2009) take up the long-standing question as to whether speakers of diverse languages overtly realize all syntactic arguments (as in English), or tend to leave them zero (as in Japanese). Rather than considering specific grammatical rules that may constrain the occurrence of zero arguments, they employ corpus measurements that they call *referential density* or *lexical referential density*, respectively. Obviously, in order to arrive at a useful comparison of speakers' argument realization behavior one needs to compare texts of roughly the same content, since content will be a major factor determining whether a particular referent is familiar at a given point in discourse or not. To achieve this goal, Pear stories are used since here the stimulus ensures that different speakers, including those of different languages, will recount roughly the same content, having the same number of opportunities to verbalize specific referents.

Similar considerations motivate the use of the so-called *Family Problem Task* (San Roque et al. 2012) in the *Social Cognition* project (Barth & Evans 2017) which seeks to determine cross-linguistic differences in the realization of certain communicative tasks, for instance the expression of thoughts of others or reference to human beings. It enables comparison of relevant lexical and constructional choices by different speakers from different languages in precisely the same contexts, as determined by the structure of the stimulus. To compare such choices across speakers, languages, and different types of context will obviously not bear any useful insights.

A line of research where the use of stimuli-based data has proven to be problematic though is that of DuBois' (1987) famous hypothesis of *preferred argument structure* (PAS): based on a small corpus of Pear stories from the Mayan language Sakapultek (Guatemala), PAS has until recently been widely accepted as a usage-based account for ergative grammar in the world's languages, see Evans & Levinson (2009). Adducing a range of corpus data from different languages, Haig & Schnell (2016b) demonstrate that PAS does not seem to extend beyond this single corpus from Sakapultek, whose containing texts seem to be characterized by what Haig & Schnell (2016b) call a "telegraphic style", presumably due to the immense discomfort speakers experienced during the respective experiment, as reported by DuBois (1980). Moreover, Schnell (under revision) finds that patterns of referent introduction are much better explained by reference to the way characters are presented in the movie stimulus rather than universal cognitive constraints on information flow. Hence, the use of a single stimulus may bear analytical risks. The latter two studies draw to a large extent on corpus data from language documentations that are not controlled for content, but instead have the advantage of resembling much more closely the kinds of routines in language use and that are variable to some degree, so that respective findings are not entirely dependent on a single type of text data.

What kinds of discourse data should underlie corpus-based typological studies is thus not determined by general methodological principles but by specific requirements related to research design and goals. An undebatable general requirement that modern developments in corpus-based typology have made considerable progress in is the accountability of findings: Haig & Schnell (2016b) draw largely on an archived and web-accessible multilingual corpus, called Multi-CAST (Haig & Schnell 2016a), that enables scrutiny and replicability of their findings by other researchers. Likewise, the Social

Cognition corpus is going to be accessible via PARADISEC. This is a great improvement of earlier work where for instance Pear story corpora have almost never been made available. This is of course not to deny that the accessibility of language documentation as well could be improved considerably, see Gawne & Berez-Kroeker (2018, in this volume).

Haig & Schnell's (2016a) MultiCAST initiative as well as the Social Cognition project (and likewise Margett's 3-participant project) again involve tremendous efforts of project-specific data processing, adding further layers of specialized annotations triggering certain constructional variants and some semantic features. In all cases, the annotations are comparable in nature, being applicable to diverse languages and enabling cross-linguistic comparison of corpus analyses. These annotations are clearly an improvement over traditional variationist procedures where relevant information is typically added in separate spreadsheets. Combining various layers of data annotation, all time-aligned to the recorded signal, opens up unprecedented possibilities for further studies, as has partly been done where GRAID annotations (Haig & Schnell 2014) have been combined with other annotation, like Schiborr et al.'s (2018) *referent indexes* (Schnell et al. 2018).


6. Conclusions I hope to have shown here that documentation-based linguistic research has enormous potential to yield insights into language use and language systems that would not have been foreseeable from the perspective of established descriptive, typological or theoretical traditions. In this connection, the haphazard and often not academically driven nature of documentations can often be an advantage, since it may bring up data that would not have been planned for from an academic research point of view, but that nonetheless provides the most relevant insights, as in the case of Vera's plant and fish descriptions. Neither would this data have been collected if researchers had followed a purist ideal of naturalistic data. Although some research questions clearly demand the collection of very specific data (for instance directly comparable text data), it seems to me that time and effort is probably better invested into further data processing of what is there rather than collection of data in service of specific analytical goals. In this way, language documentations have an important role to play the scientific research into human language.

References

- Barth, Danielle & Nicholas Evans. 2017. SCOPIC Design and Overview. In Danielle Barth & Nicholas Evans (eds), *The Social Cognition Parallax Interview Corpus (SCOPIC): A Cross -linguistic Resource* (Language Documentation & Conservation Special Publication 12), 1–23.
- Baxter, Gareth & William Croft. 2016. Modeling language change across the lifespan: individual trajectories in community change. *Language Variation and Change* 28. 129–173.
- Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David L. Beaver, Shobhana Chelliah, Stanley Dubinsky, Richard P. Meyer, Nick Thieberger, Keren Rice, and Anthony C. Woodbury. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56(1). 1–18. <https://doi.org/10.1515/ling-2017-0032>
- Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.
- Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 79(4). 708–736.
- Blythe, Richard A. & Willam Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88(2), 269–304.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.
- Chafe, Wallace (ed.). 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: ALEX Publishing Company.
- Croft, William. 2000. *Explaining language change*. Harlow: Pearson Education.
- DuBois, John W. 1980. Introduction – The search for a cultural niche: Showing the Pear Film in a Mayan community. In Wallace Chafe (ed.). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, 1–8. Norwood, NJ: ALEX Publishing Company.
- DuBois, John W. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in syntax*, 343–366. Amsterdam & Philadelphia: John Benjamins.
- DuBois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.
- DuBois, John W. 2006. The Pear story in Sakapultek Maya: A case study of information flow and Preferred Argument Structure. In Mercedes Sedano, Adriana Boliva & Martha Shiro (eds.), *Haciendo Lingüística: Homenaje a Paola Bentivoglio*, 191–221. Caracas: Universidad Central de Venezuela.
- DuBois, John W. 2017. Ergativity in discourse and grammar. In Jessica Coon, Diane Massam & Lisa D. Travis (eds.), *The Oxford handbook of ergativity*, 23–58. Oxford: Oxford University Press.
- Evans, Nicholas. 2008. Review of Essentials of language documentation. *Language Documentation and Conservation* 2(2). 340–350.
- Evans, Nicholas & Stephen Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–448.

- Gawne, Lauren & Andrea Berez-Kroeker. 2018. Reflections on reproducible research. In Bradley McDonnell, Andrea Berez-Kroeker & Gary Holton (eds.), *Reflections on language documentation on the 20 year anniversary of Himmelmann 1998*, 22–32. (Special issue of Language Documentation & Conservation 15.) Honolulu: University of Hawaii Press. <http://hdl.handle.net/10125/24804>
- Givón, Talmy. 1976. Topic, pronoun, and grammatical agreement. In Charles N. Li (ed.), *Subject and topic*, 149–188. New York: Academic Press.
- Haig, Geoffrey & Stefan Schnell. 2014. Annotations using GRAID (Grammatical Relations and Animacy in Discourse). Manual version 7.0. ms. (Available at: https://www.academia.edu/10328418/Haig_Geoff_and_Stefan_Schnell._2014._Annotations_using_GRAID_Grammatical_Relations_and_Animacy_in_Discourse_.Manual_Version_7.0)
- Haig, Geoffrey & Stefan Schnell. 2016a. *Multi-CAST: multilingual corpus of spoken annotated texts*. Cologne: Language Archive Cologne. (<https://lac2.uni-koeln.de/de/multicast/>)
- Haig, Geoffrey & Stefan Schnell. 2016b. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36. 161–195.
- Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Gippert, Jost, Nikolaus P. Himmelmann & Ulrike Mosel (eds.), *Essentials of language documentation*, 1–30. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus P. 2012. Linguistic data types and the interface between language documentation and description. *Language Documentation & Conservation* 6, 187–207.
- Himmelmann, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words. *Language* 90.4, 927–960.
- Labov, William. 1994. *Principles of linguistic change*. Oxford: Blackwell.
- Li, Charles N. (ed.). 1976. *Subject and topic*. New York: Academic Press.
- Mayer, Mercer. 1969. *Frog where are you?* New York: Dial Books.
- Margetts, Anna, Katharina Haude, Nikolaus P. Himmelmann, Dagmar Jung, Sonja Riesberg, Stefan Schnell, Claudia Wegener, John Hajek, Andrew Margetts. 2017. Investigating three-participant events across text corpora. Paper presented at the workshop *Advances in corpus-based typology: exploring corpora of semi-parallel and indigenous texts*, convened by Geoffrey Haig, Stefan Schnell, Nicholas Evans, Danielle Barth, at the ALT12, Canberra, A.N.U., 15 December 2017.
- McDonnell, Bradley. 2018. Reflections on linguistic analysis. In Bradley McDonnell, Andrea Berez-Kroeker, and Gary Holton (eds.), *Reflections on language documentation on the 20 year anniversary of Himmelmann 1998*, 191–200. (Special issue of Language Documentation & Conservation 15.) <http://hdl.handle.net/10125/24820>
- Meyerhoff, Miriam 2010. *Introducing sociolinguistics*, 2nd edn. London/New York: Routledge.
- Mosel, Ulrike. 2014a. Corpus linguistic and documentary approaches in writing a grammar of a previously undescribed language. *Language Documentation and Conservation* 8, 135–157
- Mosel, Ulrike (ed.). 2014b. *Kehaa*. Shellfish. Kiel: ISFAS Allgemeine Sprachwissenschaft, Universität Kiel.
- Mosel, Ulrike. (ed.). 2014c. *Amaa moon bara otei vaa Teapu*. The life and work of Teop women and men. Kiel: ISFAS Allgemeine Sprachwissenschaft, Universität Kiel.

- Mosel, Ulrike, Mahaka, Mark, Enoch Horai Magum, Joyce Maion, Naphtali Maion, Ruth Siimaa Rigamu, Ruth Saovana Spriggs, and Jeremiah Vaabero, Marcia Schwartz and Yvonne Thiesen. 2010. *Ainu. The Teop-English dictionary of house building*. Kiel: Seminar für Allgemeine und Vergleichende Sprachwissenschaft, Universität Kiel.
- Rhodes, Richard A., Leanore C. Grenoble & Anna Berge. 2006. *Adequacy of documentation: A preliminary report to the CELP*. ms.
- Schiborr, Nils N., Stefan Schnell & Hanna Thiele. 2018. *RefIND – Referent Indexing in Natural-language Discourse: Annotation guidelines (v1.1)*. Bamberg / Melbourne: University of Bamberg / University of Melbourne. (<https://www.uni-bamberg.de/fileadmin/aspra/misc/RefIND-guidelines-v1.1.pdf>).
- Schnell, Stefan. Under revision. Revisiting information management in intransitive subjects. ms.
- Schnell, Stefan & Danielle Barth. 2018. Discourse motivations for pronominal and zero objects in Vera'a. *Language Variation & Change* 30(1). 51–81. doi: 10.1017/S0954394518000054.
- Schnell, Stefan, Nils N. Schiborr, and Geoffrey Haig. 2018. Is intransitive subject the preferred role for introducing new referents? Evidence from corpus-based typology. Paper presented at the workshop Comparative corpus linguistics: New perspectives and applications, convened by Natalia Levshina & Steve Moran, SLE31, Tallinn, September 1, 2018.
- Seddhouse, Paul. 2013. Conversation analysis. In Robert Bayley, Richard Cameron & Ceil Lucas (eds.). *The Oxford handbook of sociolinguistics*, 91–110. Oxford: Oxford University Press.
- Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja H. de Jong & Balthasar Bickel. 2018. Nouns slow down speech: evidence from structurally and culturally diverse languages. *PNAS* 115(22). 5720–5725. <https://doi.org/10.1073/pnas.1800708115>
- Slobin, Dan. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Strömquist & Ludo Verhoeven (eds.), *Relating events in narrative: Typological and contextual perspectives*, vol. 2, 219–257. Mahwah, NJ: Lawrence Erlbaum.
- Strunk, Jan, Florian Schiel & Frank Seifart. 2014. Untrained forced alignment of transcriptions and audio for language documentation corpora using WebMAUS. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. Reykjavik, Iceland: European Language Resources Association (ELRA), 3940–3947. <http://www.lrec-conf.org/proceedings/lrec2014/summaries/1176.html>

Stefan Schnell
 stefan.schnell@uni-bamberg.de
 orcid.org/0000-0003-2036-2263