

The corpus *Deutsch in Namibia (DNam)*: A resource for contact-linguistics, variation studies, and sociolinguistics

This paper describes the corpus *Deutsch in Namibia (DNam)*, ‘German in Namibia’), which is openly accessible via the *Datenbank für Gesprochenes Deutsch (DGD)*, ‘Database for Spoken German’). The corpus is a new digital resource that systematically documents language use and linguistic attitudes within the German-speaking minority in Namibia. The published corpus contains over 18 hours of recordings with full transcriptions encompassing 224,392 tokens, covering data from 110 speakers aged 14–75 (cf. Table 1). Audio files and transcriptions are complemented by detailed metadata on speakers, including biographical data (place and year of birth, gender, occupation, family's immigration history) and information on language competence and language use.

The corpus consists of three types of data:

- 1) *Group conversations*. Groups of two to five speakers were recorded in peer-group settings with absent researchers as they talked about a range of topics, yielding data on informal conversations.
- 2) *‘Language Situations’ elicitations*. Speakers described a fictional accident, presented as a photo story, for different communication partners in formal and informal situations. This set-up elicited a broader linguistic repertoire and supports investigations of formal versus informal usages of German in Namibia.
- 3) *Semi-structured interviews*. Speakers were interviewed about topics of language use, language competence, language acquisition, and language attitudes. Interviews were conducted with groups of speakers in urban and rural contexts. The interviewer guided the conversation using impulse questions. This set-up yielded further language-biographical information on speakers, and sociolinguistic data on attitudinal patterns, linguistic evaluations, and language ideologies in the context of Namibian German.

The data from the *conversation groups* and *language situations* include both adolescents and adults, the *interviews* included in the corpus are only with adults. Table 1 gives an overview of the different corpus components.

Method	Tokens	Duration	Speakers	Recordings
Group conversations	115.004	9:15:00 hours	65	21
“Language Situations”	51.509	4:41:30 hours	103	198
Semi-structured interviews	57.879	4:42:15 hours	15	7
Total	224.392	18:38:45 hours	110	226

Table 1: Corpus composition

The data were transcribed according to *cGAT-minimal* (Selting et al. 2009; Schmidt/Schütte/Winterscheid 2015) using the *EXMARaLDA-Partitur-Editor* (Schmidt 2016), and subsequently anonymised, tokenised, normalised, and POS-tagged with the *STTS 2.0* tagset (Westpfahl 2014; Westpfahl et al. 2017).

The corpus provides an empirical basis for a range of research questions, including grammatical and pragmatic analyses, investigations into register distinctions, and sociolinguistic studies. Since speakers of German in Namibia are typically also fluent in English and/or Afrikaans, the corpus is a valuable resource for contact-linguistic studies. In this respect, Namibian German provides a particularly interesting case due to its vitality and its use across formal and informal settings, which is unusual for German as a heritage language in post-migration settings (Wiese et al. 2017). In its entirety, the *DNam* corpus can fill a significant empirical research gap for German in Namibia and beyond, including comparative studies on German as a heritage language.

References

- Schmidt, Thomas (2016): EXMARaLDA Partitur-Editor. Manual. www.exmaralda.org/pdf/Partitur-Editor_Manual.pdf (last access: 14.7.2020).
- Schmidt, Thomas/Schütte, Wilfried/Winterscheid, Jenny (2015): cGat. Konventionen für das computergestützte Transkribieren in Anlehnung an das Gesprächsanalytische Transkriptionssystem 2 (GAT 2). http://agd.ids-mannheim.de/download/cgat_handbuch_version_1_0.pdf (last access: 14.7.2020).
- Selting, Margret/Auer, Peter/Bart-Weingarten, Dagmar/Bergmann, Jörg/Bergmann, Pia/Birkner, Karin/Couper-Kuhlen, Elizabeth/Deppermann, Arnulf/Gilles, Peter/Günthner, Susanne/Hartung, Martin/Kern, Friederike/Mertzluff, Christine/Meyer, Christian/Morek, Miriam/Oberzaucher, Frank/Peters, Jörg/Quasthoff, Uta/Schütte, Wilfried/Stukenbrock, Anja/Uhmann, Susanne (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, p. 353–402. www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf (last access: 14.7.2020).
- Westpfahl, Swantje (2014): STTS 2.0? Improving the Tagset for the Part-of-Speech-Tagging of German Spoken Data. In: Levin, Lori/Stede, Manfred (eds.): *LAW VIII – The 8th Linguistic Annotation Workshop*. Dublin: Association for Computational Linguistics and Dublin City University, p. 1–10. <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/3189> (last access: 14.7.2020).
- Westpfahl, Swantje/Schmidt, Thomas/Jonietz, Jasmin/Borlinghaus, Anton (2017): STTS 2.0. Guidelines für die Annotation von POS-Tags für Transkripte gesprochener Sprache in Anlehnung an das Stuttgart Tübingen Tagset (STTS). Version 1.1, March 2017. <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/6063> (last access: 14.7.2020).
- Wiese, Heike/Simon, Horst J./Zimmer, Christian/Schumann, Kathleen (2017): German in Namibia: A vital speech community and its multilingual dynamics. In: *Language & Linguistics in Melanesia*, p. 221–245. www.langlxmlanesia.com/LLM%20S%202017_Wiese%20et%20al.pdf (last access: 14.7.2020).