

A spoken corpus of Cameroon Pidgin English: Compilation, applications and next steps

Melanie Green, Sussex (UK)

Gabriel Ozón, Sheffield (UK)

Cameroon Pidgin English (CPE) is an expanded pidgin/creole spoken in some form by an estimated 50% of Cameroon's 22,000,000 population (Simons & Fennig 2017). CPE is spoken primarily in the Anglophone west regions, but also in urban centres throughout Cameroon. As a predominantly spoken language, CPE has no standardised orthography, but enjoys a vigorous oral tradition, not least through its presence in the broadcast media. The language has stigmatised status in the face of French and English, prestige languages of Cameroon, where it also co-exists with an estimated 280 indigenous languages (Simons & Fennig 2017).

We describe the spoken corpus of CPE, a British Academy/Leverhulme-funded pilot study (Green et al. 2016, Ozón et al. 2017). The corpus consists of 30 hours of recordings made in five locations, resulting in a total of 240,000 words (80 texts of 15 minutes/3,000 words). Proportions of text types are guided by the International Corpus of English project (Nelson 1996), and the texts contain mark-up and part-of-speech-tagging. The corpus files, which are freely available from the Oxford Text Archive (<http://ota.ox.ac.uk/desc/2563>), include sound files (mp3 and wav), raw and annotated text files, participant metadata, a field manual, a tagging manual and a spelling list.

We then briefly describe some case studies of linguistic phenomena that the pilot corpus allows us to investigate, focusing on grammatical and lexical phenomena, as well as codeswitching, demonstrating that while a small corpus provides a robust test-bed for the investigation of grammatical phenomena, a larger dataset is required for the full investigation of lexical and sociolinguistic phenomena.

Finally, we outline our plans for a 1-million-word corpus, a project for which a funding application is in preparation.

References

- Green, Melanie, Miriam Ayafor and Gabriel Ozón. 2016. *A spoken corpus of Cameroon Pidgin English: pilot study*. British Academy/Leverhulme funded digital database (ref. SG140663) Available at: <http://ota.ox.ac.uk/desc/2563>.
- Nelson, Gerald. 1996. The design of the corpus. In Sidney Greenbaum (ed.). *Comparing English worldwide. The International Corpus of English*. Oxford: Clarendon Press, 27–35.
- Ozón, Gabriel, Miriam Ayafor, Melanie Green and Sarah Fitzgerald. 2017. A spoken corpus of Cameroon Pidgin English. *World Englishes* 36: 427–447.
- Simons, Gary F. and Charles D. Fennig (eds.). 2018. *Ethnologue: Languages of the World, Twenty-first edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.