

## Information structure in a spoken corpus of Cameroon Pidgin English

Ayafor, Miriam, Melanie Green, and Gabriel Ozon

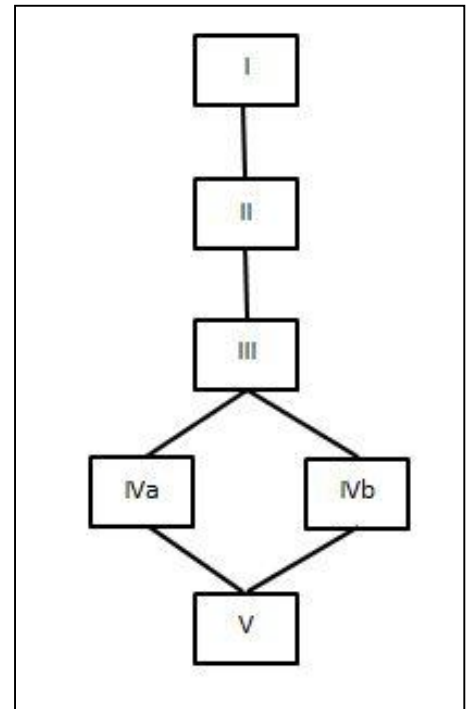
Cameroon Pidgin English (CPE) is an expanded pidgin/creole spoken in some form by an estimated 50% of Cameroon's 22,000,000 population (Lewis et al. 2014), primarily in the Anglophone west regions, but also in urban centres throughout the country. As a primarily spoken language, CPE has no standardised orthography, but enjoys a vigorous oral tradition, not least through its presence in the broadcast media. However, it resists close documentation due to its stigmatised status in the face of French and English, prestige languages of Cameroon, where it also co-exists with an estimated 280 indigenous languages (Lewis et al. 2014).

We are in the first year of a two-year, British Academy/Leverhulme-funded project aimed at constructing a 240,000-word pilot corpus of transcribed spoken CPE dialogues and monologues, with partial POS-tagging, glossing and translations.

As a testing ground for the pilot corpus methodology, we constructed a small, 120,000-word 'pre-pilot' corpus consisting of (i) spoken CPE (Ayafor, Green and Ozón, in prep.), (ii) existing published sources (Ayisi & Longinotto 2005; Bellama et al. 2006; Todd 1979), and (iii) elicited examples. This pre-pilot corpus was designed to test our approaches to recording, transcribing, coding and devising a POS-tagset (and tagging system).

Drawing on this pre-pilot corpus, we report on the methodological stages for investigating information structure in CPE, which we identify as the following:

- I. **Elicitation** of representative examples of topic and focus constructions. For example, question-answer pairs for the identification of new information focus (1), and assertion-correction pairs for the identification of contrastive focus (2), as well as cleft constructions (3). This allows the identification of function words and word orders associated with focus and topic constructions, such as copulas and relativisers.
- II. **Extracting** the set of utterances containing the relevant function words (or n-grams containing those function words) from a corpus of naturally-occurring language.
- III. **Coding** those utterances to separate topic/focus constructions from other grammatical constructions containing the same function words (4), (5).
- IV. (a) The identification of recurring **constructional patterns** involved then offers the potential for automatic retrieval (6). This also allows (b) the identification of overt markers of **contexts** in which e.g. focus constructions are likely to occur, e.g. question/answer pairs.
- V. The resulting dataset is then available for prosodic analysis.



We identify a number of methodological challenges, including the following:

- Orthography: the lack of a standardised spelling system conspires against automatic token retrieval. Cross-checking transcriptions becomes an expensive necessity in the absence of a pre-existing annotation scheme.
- Multifunctionality: the expression *na*, which serves both as a copula and as a focus marker, is extremely frequent, requiring the manual coding of over 3,000 tokens from our 120,000-word pre-pilot corpus.

- Multifunctionality: similarly, the expression *fo*, which functions as topic marker, preposition, and infinitival particle, is also extremely common, requiring the manual coding of over 5,000 tokens.

## Examples

- (1) Q: (na) fo hu i bi gif de chop?  
 COP PREP who(m) she ANT give DEF food  
 'Who did she give the food to?'  
 A: i bi gif de chop fo yi boi-pikin object focus (new information)  
 she ANT give DEF food PREP her boy-child  
 'S/he gave the food to his/her son.'
- (2) Q: i bi bai tomato, no bi so?  
 she ANT buy tomatoes, NEG be so?  
 'She bought tomatoes, didn't she?'  
 A: no-oh, na banana i bi bai object focus (contrastive)  
 no, COP bananas she ANT buy  
 'No, s/he bought bananas.'
- (3) na wit cane weh yu bit yi cleft  
 COP PREP cane REL you beat her  
 'It was with a cane that you beat her?'
- (4) yi nem na Atiqu copular clause  
 his name COP Atiqu  
 'His name was Atiqu.'
- (5) sumo ting weh a wan tok relative clause  
 small thing REL I want say  
 'something I want to say.'
- (6) (na) NP (weh) NP (TMA) V ex-situ object focus schema

## Abbreviations

ANT = anterior tense; COP = copula; DEF = definite determiner; NEG = negation; PREP = preposition; REL = relativiser.

## References

- Ayafor, Miriam, Melanie Green and Gabriel Ozón. In preparation. *A spoken corpus of Cameroon Pidgin English*. Ms, University of Sussex.
- Ayisi, Florence and Kim Longinotto (dir.) 2005. *Sisters in Law* [DVD]. London: Vixen Films.
- Bellama, David, Solomon Nkwelle and Joseph Yudom. 2006. *An introduction to Cameroonian Pidgin*. 3rd ed. Cameroon Peace Corps.
- Lewis, M. Paul, Gary F. Simons and Charles D. Fennig (eds.). 2014. *Ethnologue: language of the world*. 17<sup>th</sup> edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Todd, Loreto (editor and translator). 1979. *Some day been dey: West African Pidgin folktales*. London: Routledge and Kegan Paul.