

A Comparative Analysis of Nigerian Linguist Native Speakers and Untrained Native Speakers Categorising Four Accents of Nigerian English

Umar Gombe Muhammad¹, Peter French^{1,2}, and Eleanor Chodroff¹

¹University of York, ²JP French Associates

In the field of LAAP (Language Analysis in the Asylum Process), there has been a great deal of debate over who should undertake the task of inferring a speaker's country and region of socialisation based on their language and dialect: academic/professional linguists with detailed knowledge of the languages/varieties that may be at issue (see LNOG, 2004; Fraser, 2009; 2011), by naïve native speakers of those languages/varieties, or by a combination of both (see Cambier-Langeveld, 2010; 2012; Fraser, 2011; Foulkes, French and Wilson, 2019; Wilson, 2009). Opposing positions within this debate have largely been argued on principle alone, without support from empirical studies. Contributing to this debate, this presentation concerns research which forms part of a larger study designed to determine which of the following four methods performs most accurately and reliably in determining the first language (L1) of Nigerian speakers of English. It also seeks to establish a basis for combining selected methods.

- (1) Review of speech samples by educated but linguistically naïve native speakers of the L1s concerned;
- (2) Examinations of the samples by native speaker of the L1s who are also academic linguists holding staff positions in Nigerian universities;
- (3) Examinations of the samples by general phoneticians and forensic phoneticians, with L1 British English, working either in UK universities or forensic speech science practices. (These participants are provided with priming material in the form of lists of phonetic features that distinguish the four Nigerian English varieties from one another);
- (4) Analysis of the samples by an automatic accent recognition system (YACCDIST-Brown, 2016).

The material presented in the present abstract concerns only the relative performance of methods 1 and 2. It is anticipated that no method will perform with 100% accuracy. Assuming this, one might ask whether the errors of the highest performing methods are in complementary distribution such that, for example, the samples misidentified by naïve native speakers and those misidentified by UK phoneticians or by the automatic system do not (substantially) overlap. If that is indeed the case, one then has a principled and empirically-grounded basis for selecting which methods to combine in achieving optimum performance in LAAP casework.

Methods

Sixteen recordings made during a fieldwork visit to Nigeria of L1 of speakers of Hausa, Igbo, Kanuri and Yoruba speaking in English were selected for the experiment (categorisation task). These included four speakers from each of the four language groups; additionally, 2 foil recordings (Ghanaian and Guinean English speakers) were added. For Method 1, 80 linguistically naïve educated native speakers, mainly in the form of university students and administrative staff, were recruited in the universities and cities of Kano, Nsukka, Maiduguri and Ibadan. For Method 2, 25 academic linguists (with various specialisations in e.g., phonetics/phonology, syntax, semantics and sociolinguistics of their Nigerian L1) were recruited from four Nigerian universities: Bayero University, Kano (Hausa linguist); University of Nigeria, Nsukka (Igbo linguists); University of Maiduguri (Kanuri linguists); University of Ibadan (Yoruba linguists). Using Qualtrics Survey Software, the two groups of participants were asked to listen to the recordings (which have equal duration of 30 seconds, comprising readings of part of a phonetically-balanced test — The Rainbow Passage — and spontaneous speech narrating their life experiences) under equivalent conditions and assign each of the recordings to an L1 (Hausa, Igbo, Kanuri, Yoruba or non-Nigerian).

Results

Overall, native speakers from all 4 L1 backgrounds, irrespective of whether they were linguists, performed well above chance level; a mixed-effects logistic regression was used to assess the effect of L1 match and expertise (linguist or non-linguist) on language identification accuracy. L1 match was significant, indicating that listeners exposed to stimuli of their L1 were approximately 2.26 times as likely to be accurate than when exposed to stimuli of other L1s ($\beta_{match} = 0.81, p < 0.001$). Native speaker linguists, however, were only slightly better than naïve native speakers on a numeric basis; the difference was not significant and applied to only 3 of the 4 L1 linguist groups ($p > 0.05$). Igbo linguists, who were the best performing group at identifying their L1 Igbo stimuli, were the poorest group at identifying other L1s (Hausa, Kanuri & Yoruba), while Kanuri linguists, who were the poorest at identifying their L1 Kanuri, were the best performing group at identifying other L1s (Hausa, Igbo & Yoruba). Hausa samples were easiest to identify by the linguists and non-linguists (well above chance level); Kanuri samples were the most challenging to identify by both groups (only at chance level). In a separate analysis, a set of salient L1 features were identified such as 'the' in Kanuri or 'h' in Yoruba. Samples which were richer in L1 cues (in terms of the number of pre-defined salient L1 features present in the recordings) were more correctly assigned to their L1s than others with fewer L1 features.

Discussion

Since the linguists were only slightly more accurate than the non-linguists in identifying the L1s, these findings offer empirical support for having educated native speakers involved in LAAP casework, even without linguistic training. Performance of both groups in identifying other Nigerian L1s (well above chance level) is facilitated by the listeners' familiarity with speakers of other L1s, owing to the movement of the participants in Nigeria and the use of English as lingua franca, as confirmed from them. Both groups were only around the chance level in identifying the two foils. The next stages of the research (testing the abilities of the UK phoneticians and an automatic accent recognition system-YACCDIST) are currently underway. In contrast to the previous methods, the UK phoneticians will also have the option of using auditory-perceptual or acoustic analysis help guide their categorisation of the speech samples.

References

- Brown, G (2016). Exploring forensic accent recognition using the Y-ACCDIST system. *Proceedings of the 16th Speech Science and Technology Conference*, Sydney, Australia, 6-9 December 2016, 305-308.
- Cambier-Langeveld, T. (2010). The role of linguists and native speakers in language analysis for the determination of speaker origin. *International Journal of Speech, Language and the Law*, 17 (1), 67-93.
- Cambier-Langeveld, T. (2012). Clarification of the issues in language analysis: a rejoinder to Fraser and Verrips. *International Journal of Speech, Language and the Law*, 19 (1), 95-108.
- Foulkes, P., French, P. and Wilson, K. (2019). LADO as forensic speaker profiling. In Patrick, P., Schmid, M. and Zwaan, K, (Eds). *Language Analysis for the Determination of Origin*. Cham: Springer, pp. 91-116.
- Fraser, H. (2009). The role of 'educated native speakers' in providing language analysis for the determination of the origin of asylum seekers. *International Journal of Speech, Language and the Law*, 16 (1), 113-138.
- Fraser, H. (2011). The role of linguists and native speakers in language analysis for the determination of speaker origin: a response to Tina Cambier-Langeveld. *International Journal of Speech, Language and the Law*, 18 (1), 121-130. <https://doi.org/10.1558/ijsll.v18i1.121>.
- Language and National Origin Group. (2004). Guidelines for the use of language analysis in relation to questions of national origin in refugee cases. *International Journal of Speech, Language and the Law*, 11(2), 261-266.
- Wilson, K. (2009) Language analysis for the determination of origin: native speakers vs. trained linguists. MSc dissertation, University of York.