

# Linguistic Distance and Mutual Intelligibility among South Ethiosemitic Languages: A Combined Approach

---

## Abstract

Ethiosemitic languages are variants of the Semitic language family which are spoken in Ethiopia and in Eritrea. They are classified into North and South Ethiosemitic. The North branch consists of Ge'ez, Tigre and Tigrigna while the South Ethiosemitic includes Amharic, Argoba, Harari and several Gurage varieties. Many of the South Ethiosemitic languages are closely related, and the speakers of one variety can sometimes communicate with the speakers of other varieties. The relative distance (Bender, 1971; Hudson, 2013) and mutual intelligibility (Ahland, 2003; Gutt 1980) among these languages received some attentions. There are also proposals of the classification of the languages based on shared morphological and grammatical features (*e.g.* Demeke, 2001; Hetzron, 1972, 1977; Laslau, 1969). However, the previous studies have several shortcomings. First, they are limited in terms of scope- especially the Gurage varieties received a marginal attention. Second, some of the studies are not supported by sufficient data (Demeke, 2001; Hetzron, 1972). There are also inconsistencies among mutual intelligibility studies (*cf.* Ahland, 2003 and Gutt, 1980) which possibly emanate from the inherent methodological limitations and complex intermingling among the languages.

The present study employed a combination of measures: structural distance (lexicostatistics and Levenshtien distance), mutual intelligibility, perceived distance (judged similarity and judged intelligibility) and geographical distance to determine the distance and mutual intelligibility among 10 South Ethiosemitic languages: *Chaha, Ezha, Harari, Silt'e, Gora, Gumer Soddo, Muher, Ennomer and Mesqan*. The study intended to (1) determine the relationship among the distance measures and (2) re-examine the previous classification of the languages. The lexical and Levenshtien distances were computed based on words in the translations of a fable-*the North Wind and the Sun*. The lexical distance was determined by computing the average of the percentage of non-cognate words within pairs of languages. The Levenshtien distance was determined by computing the cost- *insertion, substitution and deletion* required to transform the pronunciation of one cognate to another, using GabMap (see Nerbonne et al., 2011). Word categorization test was used for mutual intelligibility measure (Tang and Heuven, 2009). The test requires categorizing words under their Semantic categories; for example apple under FRUIT. The perceived distances were determined using the recordings of the translations of *the North Wind and the Sun*. After listening to each recording, selected participants determined (a) to what extent each recording is similar with their native language (*judged similarity*) and (b) to what extent they understand each of the recording (*judged intelligibility*). GabMap was employed for the clustering and cluster validation (multidimensional scaling and bootstrapping). The geographical distance among the language areas was obtained from Google Earth using GabMap.

The regression results show strong correlations among the distance measures (structural, mutual intelligibility and perceived distance). The clusters obtained from the measures are also largely similar. These results are consistent with previous correlation reports (Gooskens, 2013; Gooskens and Heeringa, 2004; Tang and Heuven, 2009). The clusters obtained from the measures are also very similar with the genetic classifications proposed by Demeke (2001) and Hetzron (1972). However, Harari and Soddo have shown deviations from the previous classifications. This divergence is associated with the influence of Cushitic languages. Strong association was also found between language distance and geographical distance which indicates the influence of language contact.

## References

- Ahland, C. (2003, April). Interlectal intelligibility between Gurage speech varieties. In *North American Conference on Afroasiatic Linguistics, Nashville, April*.
- Bender, M. L. (1971). The languages of Ethiopia: A new lexicostatistic classification and some problems of diffusion. *Anthropological Linguistics*, 165-288.
- Demeke, G. A. (2001). The Ethio-Semitic Languages (Re-examining the Classification). *Journal of Ethiopian Studies*, 57-93.
- Gooskens, C. (2013). Methods for measuring intelligibility of closely related language varieties. *Handbook of sociolinguistics*, 195-213.
- Gooskens, C., & Heeringa, W. (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language variation and Change*, 16(03), 189-207
- Gutt, E. A. (1980). Intelligibility and interlingual comprehension among selected Gurage speech varieties. *Journal of Ethiopian Studies*, 14, 57-85.
- Hetzron, R. (1972). *Ethiopian Semitic: studies in classification* (No. 2). Manchester University Press.
- Hetzron, R. (1977). *The Gunnän-Gurage Languages* (Vol. 12). Istituto orientale di Napoli.
- Hudson, G. (2013). *Northeast African Semitic: Lexical Comparisons and Analysis*. Harrassowitz.
- Leslau, W. (1969). Toward a classification of the Gurage dialects. *Journal of Semitic Studies*, 14(1), 96-109.
- Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., & Leinonen, T. (2011). Gabmap-a web application for dialectology. *Dialectologia: revista electrònica*, 65-89.
- Tang, C., & Van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, 119(5), 709-732