

Word break conflicts and elision in Gbari nominalizations

In developing a writing system for an unwritten language, two major issues frequently arise: deciding where to mark word breaks, and how to handle vowel elision. Numerous semantic, grammatical, phonological, sociolinguistic and pedagogical criteria have been proposed to help provide guidance in such matters (e.g. Haspelmath 2011, Kutsch Lojenga 2014), but such criteria are often in conflict with each other and orthography developers are forced to prioritise certain factors over others (Eaton & Schroeder 2012). Complex nominalizations such as noun-incorporation constitute a continuum between one word and two words, with elision more likely the more frequently the compound is used. Current orthographic solutions typically involve either hyphens or apostrophes or both.

This paper looks at word breaks in Gbari [gby], a Nupoid language of Central Nigeria, in which the vast majority of words are monosyllabic or disyllabic, even though much longer words are possible through compounding. Word-initial vowels often get dropped in fluent speech, with current orthographic practice representing the elided vowel with an apostrophe. Both these factors have resulted in a great deal of orthographic variation in the transcription of phrasal nominalizations. The situation gets even more complex when the incorporated phrases are themselves complex, as illustrated by the following example:

- (1) *a-tⁿa-be-dyegoyi-sũ-yi*
PL-be.without-matter-good-do-AG
'wicked people'

Here, the initial plural prefix /a-/ conceptually applies to the agentive suffix /yi/ on the the verb /sũ/ 'do' at the end, suggesting, by some criteria at least, that they could be considered part of the same morphosyntactic word, even though words of such length are extremely rare.

This paper applies ten word break criteria to Gbari nominalizations alongside a comparison of nominalizations in both natural texts and translated texts, which tend to have a higher proportion of longer words. It proposes some guidelines towards an optimal orthography for Gbari, attempting to minimise word-break conflicts, whilst at the same time maintaining internal consistency, naturalness, and pedagogical clarity.

References:

- Eaton, Helen and Schroeder, Leila (2012) 'Word break conflicts in Bantu languages: Skirmishes on many fronts', *Writing Systems Research*, 4(2), 229-241.
- Haspelmath, Martin. (2011). 'The indeterminacy of word segmentation and the nature of morphology and syntax', *Folia Linguistica*, 45(1), 31-80.
- Kutsch Lojenga, Constance (2014). 'Basic principles for establishing word boundaries'. In Michael Cahill & Keren Rice (Eds), *Developing Orthographies for Unwritten Languages*, 73–106. Dallas: SIL International.