A quantitative approach to Nigerian Pidgin prosody: revisiting tonal analyses

Emmett Strickland and Anne Lacheret-Dujour
MoDyCo Laboratory, Paris Nanterre University

Languages with lexically specified prosody can primarily be grouped into stress-accent languages, in which every lexical word carries one prominent syllable marked through some combination of pitch, duration, and intensity; and tone languages, in which pitch targets are intrinsically associated with certain morphemes. The prosodic systems documented for creole languages are diverse, covering both tonal and stress-accent configurations, mixed systems which incorporate both tone and stress, and etymologically split systems in which the rules governing the prosody of lexemes are determined by the language of origin. This submission describes an ongoing effort to apply novel computational approaches to exploring the prosody of Nigerian Pidgin, an Afro-Atlantic English Lexifier Creole.

Spoken by as many as 100 million speakers, Nigerian Pidgin is the world's most widely spoken creole language. Like many other creoles, Nigerian Pidgin has a stress-accent lexifier, but is spoken alongside predominantly tonal languages such as Igbo, Yoruba, and Hausa. Most previous descriptions of the language's lexical prosody have described it as a tone language or pitch accent language in which underlyingly high-toned syllables only occur once in a given word. The majority of supposed minimal pairs are monosyllabic and oppose lexical and grammatical functions, with the most frequent examples including gó 'go' and gò 'FUT', or déy 'to be' and dèy 'IMPF'. The few polysyllabic minimal pairs which have been described include fàdà 'father' and fàdá 'Catholic priest', though it is unclear whether this distinction has been maintained in modern varieties of the language. Such analyses can be said to be the result of stress-to-tone mapping, a process by which the pitch prominences in stressed syllables are reanalyzed as high tones by native speakers of tone languages. To our knowledge, no major description of Nigerian Pidgin has described it as a stress accent language. Other varieties of West African Pidgin English have also been described as tonal, perhaps most notably Pichi which exhibits a similar set of minimal pairs, as well as an apparent use of pitch the marking of grammatical case on pronouns.

In this submission, we revisit previous descriptions of Nigerian Pidgin prosody using an innovative, corpus-driven approach. By augmenting an existing syntactic treebank with syllable-level prosodic features, we are able to quantify the interactions between morphosyntactic variables, and the prosodic variables of pitch, duration, and intensity. Using roughly five hours of transcribed and annotated spontaneous language, we have automatically extracted occurrences of common minimal pairs. Thus far, this approach has successfully demonstrated significant differences in pitch between each pair, which at first glance would appear to validate a tonal analysis of the language's prosody. However, these differences in pitch also co-occur with significant differences in length, a feature of Nigerian Pidgin which to our knowledge has not been previously documented. We therefore believe that the traditional wisdom of placing Nigerian Pidgin in the same typological category as its tonal substrates deserves reexamining. Indeed, this initial result appears consistent with a stress-based analysis. If Nigerian Pidgin is indeed a stress language, the pitch differences we observe in these supposed minimal pairs are simply one of several correlates of lexical stress or the absence thereof. This would make the pitch difference observed between gó 'go' and gò 'FUT' equivalent to the pitch difference between the stressed lexical item *butt* and the unstressed function word *but* in English. We also present our ongoing exploration of the correlation between pitch, duration, and intensity in other areas. For example, we will present

findings from a prosodically specified lexicon extracted automatically from the treebank. Using this lexicon, we examine whether the highest pitch syllables of polysyllabic words also tend to co-occur with increased length and intensity, and explore whether this supports a tonal or stress-based analysis of the language.

This presentation will begin with an overview of Nigerian Pidgin and previous studies of its prosody. We will then present the syntactic treebank used in our research and how we developed it into an innovative intonosyntactic resource to facilitate our study of the language. We will then present results from our exploration of supposed minimal pairs present in the corpus. Finally, we will showcase other applications of corpus and how they may contribute to the study of Nigerian Pidgin prosody.