Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Prosody and cognitive accessibility in left-detached topics: lessons from Nigerian Pidgin

Emmett Strickland[1], *Anne Lacheret-Dujour*[1], *Candide Simard*[2]

[1]MoDyCo, Paris Nanterre University, France
[2]School of Pacific Arts, Communication and Education, University of the South Pacific, Fiji
emmett.strickland@parisnanterre.fr, anne.lacheret@parisnanterre.fr, candide.simard@usp.ac.fj

Abstract

In Nigerian Pidgin, or Naija, a language spoken by some 100 million people in Nigeria, the role of prosody in topic-marking remains an understudied question, particularly with regards to the way in which speakers distinguish between topics which are readily present in the minds of their interlocutors, and less cognitively accessible ones such as those being introduced for the first time. Advances in the field of natural language processing have yielded effective methods for the automatic extraction of stylized prosodic contours from aligned sound files, facilitating the wide-scale study of syntactic units and their melodic properties.

Through a combination of manual pragmatic annotation, automatic prosodic stylization, and exploratory statistical analysis, we aim to better understand the prosodic differences between categories of left-detached topics. We hypothesize that there exists a clear relationship between the cognitive accessibility of topics and their associated prosodic contours. More precisely, the most cognitively accessible topics will be characterized by falling F0 contours with low fundamental frequency ranges. Conversely, we consider that less accessible topics will be associated with rising contours and ones covering wide frequency ranges. We also postulate that contours ending in a very high final pitch will be almost exclusive to less accessible topics.

Index Terms: Nigerian Pidgin, Naija, creoles, prosody, information structure

1. Introduction

Nigerian Pidgin, or Naija, is an English-lexifier pidgincreole spoken by roughly 100 million people in West Africa and within the Nigerian diaspora. A product of Afro-European contact in the 18th, 19th and 20th centuries, Naija has played an increasingly prominent role in Nigerian society since the country's independence in 1960 [1]. Today, it is used as the primary vehicular language spoken between Nigeria's various ethnolinguistic groups, and the preferred vernacular language of most urban centers. Despite a strong lexical resemblance to English, the language has developed a distinct set of syntactic and prosodic characteristics.

Since 2017, the French National Research Agency (ANR) project NaijaSynCor has sought to study these characteristics though the development of an extensive corpus of spoken Naija representing 319 speakers and roughly 30 hours of recorded speech. The corpus takes the form of a syntactic treebank of transcribed speech in which each token is assigned a pair of numeric values representing its estimated start and end times in the original recording. This data structure allows researchers to automatically align syntactic structures with their locations in a corresponding audio file.

The treebank was developed using a two-tiered syntactic model based on that of the Rhapsodie corpus of spoken French [2]. A microsyntactic annotation scheme describes marked relations of dependency such as those between verbs and their arguments and modifiers, while a system of macrosyntactic annotation divides utterances into three major types of internal units defined by their level of illocutionary autonomy. All completed utterances contain at least one nucleus, a fully autonomous unit capable of forming an acceptable utterance in isolation and in the same context. There also exist two optional types of units whose contextual function and acceptability depends on the existence of a nucleus. Prenuclei, also known as the left periphery, are located before a nucleus and are delimited by the symbol <. Postnuclei, located to the right, are delimited by the symbol >. Figure 1 provides an example of a syntactic tree containing a prenucleus, as well as a corresponding utterance-level alignment.

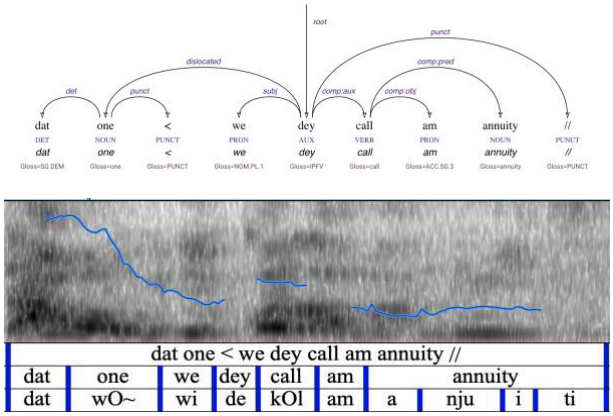

Figure 1: *Syntactic tree with phonetic alignment.*

The left periphery is known to play an important role in information structure, being used in Naija and many other languages to encode two basic elements: 1) framesetting topics that delimit the applicability of information, such as ***yesterday***, *it rained* and 2) aboutness topics, *i.e.* topics that an utterance is intended to provide information about, such as the one seen in the Naija utterance ***dat one*** *<* *we de call am annuity* 'that one, we call it annuity' represented in Figure 1.

These left-detached topics can be categorized according to their cognitive accessibility, *i.e.,* the degree to which they are present in the mental representation of the interlocutor at the

10.21437/SpeechProsody.2022-4

time of their utterance [3], [4], [5], [6], [7]. Prior research shows that new and therefore less cognitively accessible elements can be highlighted by a melodic prominence [8], [9], leading us to speculate that different topic categories are characterized by varying intonational features. While a substantial body of literature exists on the role of prosody in topic-marking [10], [11], no work has been done regarding Naija.

In this study, we have used a combination of manual pragmatic annotation, automatic melodic stylization, and exploratory statistical analysis to better understand the intonational characteristics of these left-detached topics. Specifically, we sought to determine whether it is possible to highlight the preferred intonational contours of each type of aboutness topic, hypothesizing that new and therefore less cognitively accessible topics are more likely to be marked with rising contours and very high final pitches, whereas more accessible ones are characterized by falling contours with lower F0 ranges. These hypotheses were partly supported by our results, though further research is required. The approach described in this paper is language-independent, and could therefore serve as useful basis for the study of intonation in other languages, provided the existence of a time-aligned corpus of recorded speech.

2. Data and methodology

2.1. Corpus and data preparation

This study draws from monologues of 23 speakers from the NaijaSynCor corpus, corresponding to about two hours of recorded speech, 2000 utterances, and 700 prenuclei. The sample covers several discourse types, including narratives, instructions, and sermons. Each monologue was represented by an audio file (.wav), a time-stamped pitch contour (.PitchTier) manually verified and corrected using the prosodic annotation software ANALOR [12], a raw text file representing each utterance in the form of a syntactic tree (.conllu) created under the Surface-Syntactic Universal Dependencies (SUD) annotation scheme [13], and a .TextGrid alignment containing a word and syllable-level segmentation of each utterance.

In a second step, we developed a Python script to construct a second set of .TextGrid alignments specifying the estimated positions of each prenucleus in the 23 monologues. A second tier allowed us to distinguish between simple prenuclei, which contain a single pragmatic element, and complex prenuclei like *and my parents*, which contains both a connector and a topic (see Figure 2). Only aboutness topics appearing in simple prenuclei were considered in this study. Each alignment was manually inspected in Praat.

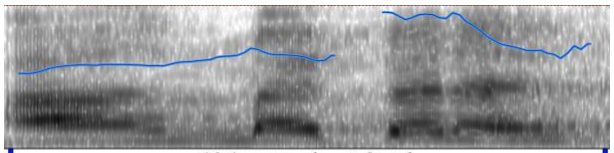

Figure 2: *Segmentation of a complex prenucleus.*

2.2. SLAM+ and contour stylization

For each prenucleus, a stylized melodic contour was generated using the software SLAM+[1], described more fully in [14]. This software intends to reduce the F0 variations within any user-defined unit of speech (in this work, left-detached topics) into a streamlined modelization which represents only those which are truly perceptible. Using a cleaned .PitchTier as well as the corresponding .TextGrid alignment as input, SLAM+ models the melodic characteristics of each interval of a specified tier using an annotation scheme which takes into account the following variables:

- The F0 value of the first voiced section of the segment.
- The F0 value of the final voiced section of the segment.
- If applicable, the value and relative position of the most salient F0 peak within the segment.

In our study, F0 values are described using five elementary tones calculated relative to the speaker's mean F0 (Table 1). Meanwhile, the temporal position of the internal saliency is expressed as one of three possible numeric values: 1, for saliencies located at the beginning of the contour; 2, for those located in the middle; and 3, for those at the end.

Table 1: *Pitch levels in SLAM+.*

Elementary tones	Description
H	extreme-high
h	high
m	medium
l	low
L	extreme-low

Contours with no internal saliency are expressed as a simple pair of elementary tones. The label Hm therefore represents a simple falling tone beginning with an extreme-high pitch and ending with a medium pitch. If an internal saliency is present, it is described using a letter and a number located at end of the code. The contour Hmh2 therefore represents a falling contour with a high saliency at the middle of the segment.

Figure 3 shows contours generated for the continuing topic *most of di young people wey dey America dat time* 'most of the young people in America at the time' and the contrastive topic *di kind cloth wey I dey sew* 'the kind of cloth that I'm sewing'. Note that only the global contours, represented in green, were considered in our analysis. The local contours, represented in red, are generated through an alternative means of calculating the elementary tone values, and are not considered in this study.

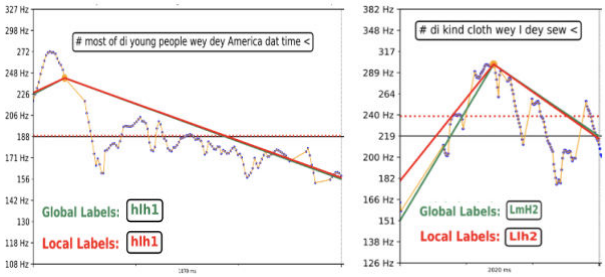

Figure 3: *Contours from two categories of topic.*

[1] https://github.com/vieenrose/SLAMplus

2.3. Pragmatic annotation

A spreadsheet was produced containing the temporal values, textual content, and global contour labels of each prenucleus. In a separate column, we manually assigned each aboutness topic a label according to its pragmatic function (see Figure 4). We limit the scope of this study to four categories of aboutness topic adapted from [15]. These are listed here in decreasing order of cognitive accessibility:

- TOP_A CONTINUING for continuing topics, i.e., those which are already at the center of discussion at the time of their utterance. We consider these to be the most cognitively accessible.
- TOP_A RESUMED for resumed topics, i.e., those previously introduced in the monologue, but which are no longer at the center of discussion at the time of their utterance.
- TOP_A NEW for new topics, i.e., those mentioned in the monologue for the first time.
- TOP_A CONTRAST for contrastive topics. The two people referenced in the utterance ***Jack*** *walks to work, but* ***Jill*** *takes the bus* would both be annotated as contrastive.

Pragmatic type	Contour	Begin	End	Text
TOP_A NEW	lmh2	19.374	20.042	my parents <
TOP_CONTRAST	mlh2	47.58	48.607	# my dad <
TOP_A RESUMED	mmh1	82.107	83.337	# even me sef <
TOP_A CONTINUING	hlH1	96.79	97.83	to do my registration <

Figure 4: *Annotated prenuclei in spreadsheet: topic types and corresponding contours*

2.4. Statistical output

Using the .tsv spreadsheet containing the automatically extracted global contours and the manual topic annotations, we generated a Multiple Correspondence Analysis (MCA) plot [16] to visualize correspondences between topic labels and the associated global contour labels generated by SLAM+.

All instances of the four topic categories occurring in simple prenuclei were imported into RStudio. We then declared the column corresponding to the pragmatic category as the active variable, while the column corresponding to the global contour was treated as illustrative.

An MCA plot was produced by rendering both variables visible, resulting in a mapping of the aboutness topic categories with their associated contours. We found a plot representing the first two dimensions sufficient for the purposes of this study.

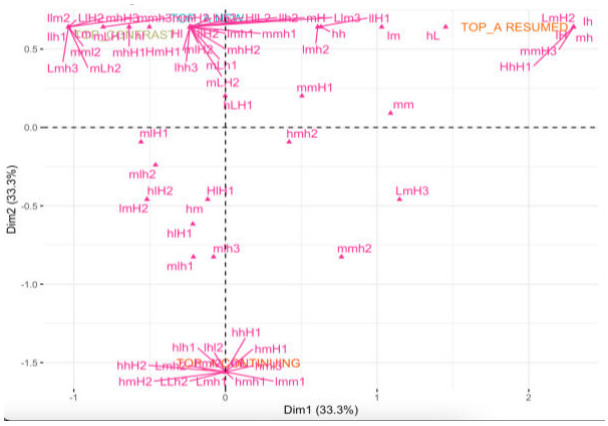

Figure 5: *MCA plot of topics and associated contours.*

Figure 5 shows the MCA plot based on topics located in simple prenuclei. The global contour labels are distributed according to their association with four pragmatic types corresponding to contrastive topics (top left), new topics (top center), resumed topics (top right) and continuing topics (bottom center).

3. Analysis of results

The MCA plot generated clearly demonstrates that each class of aboutness topic is associated with a distinct cluster of melodic contours. Two primary groupings of topics are observed: continuing topics, the most cognitively accessible of the four, are situated alone at the bottom of the vertical axis, whereas the three less accessible topics fall relatively close together at the top of the plot. This would suggest that the category of continuing topics has a distinct set of melodic characteristics which distinguish it from the three less cognitively accessible categories, supporting the hypothesis of a relationship between cognitive accessibility and intonational features.

Within the cluster at the top of the plot, one can also observe that new topics and contrastive topics pattern particularly closely to one another, and that several contour labels are located directly between the two. This suggests that these two topic categories share certain melodic characteristics in common, and therefore co-occur with a similar set of intonational patterns. One can also observe a range of contour labels located directly between the points corresponding to new and resumed topics, which therefore exclusively share a set of associated contour labels in common. The contours located near the middle of the plot appear to lack any strong association with any of the topic labels.

A central question that arises when interpreting this MCA plot is whether the discrete contour labels that make up each cluster are united by certain shared characteristics. Indeed, one might expect each type of topic to pattern with contours representing a certain melodic feature or set of features such as direction (rising, falling, or plateau), height or position of the internal saliency, or final F0 value.

Of the six contour labels directly associated with resumed topics, four feature rising contours, and all complex contours have a very high internal saliency. It is somewhat more difficult to identify a common set of features that unite the contours associated with the new and contrastive topics. One can observe a mix of rising and falling contours, as well as a range of positions and F0 values for the internal saliencies. One distinguishing feature of this group is a set of falling contours ending with a very low F0 value. While these do not make up a particularly large portion of the contours in this group, they are virtually absent outside of new and contrastive topics. There is also a significant set of contour labels such as llm2, LlH2, or llh1 which represent convex contours with low F0 values at the edges.

It is also difficult to identify more general intonational features typical of continuing topics. These are associated with both rising and falling contours, some of which contain very high internal saliencies. This partly contradicts our hypothesis that these topics would be characterized by falling contours. This suggests that, unlike less accessible topics, continuing topics do not need to be marked by any specific melodic patterns. The intonational variation observed within this

category therefore appears to be independent of information structure.

Based on these observations, one can draw several inferences regarding the melodic patterns associated with each topic category. Speakers of Naija tend to use rising contours or contours with a very high internal saliency when reintroducing topics. New and contrastive topics can be distinguished with falling contours which traverse a wide range of pitch values and end with a very low F0, or convex contours with low values at the edges. Finally, continuing topics are associated with the widest range of contour types.

More broadly, our initial hypotheses are generally supported. Continuing topics, the most cognitively accessible pragmatic type, do pattern frequently with falling contours, though to a lesser degree than initially predicted. These topics also tend to pattern with a relatively narrow fundamental frequency range, unlike their new and contrastive counterparts. Rising contours co-occur with all pragmatic types, but are better-represented among the three less cognitively accessible categories, also supporting our hypothesis. Rising contours appear to be a defining feature of resumed topics in particular. One can finally note that contours ending in a very high final tone are completely absent among the continuing topics, confirming our hypothesis that such contours are exclusive to less accessible topics.

4. Discussion and conclusion

This paper aimed to better understand the relationship between intonation and four categories of left-detached aboutness topics in Naija, with the primary hypothesis that they are characterized by distinguishing intonational features, according to their cognitive accessibility. Using Multiple Correspondence Analysis to explore our data, we observed that continuing topics and resumed topics are associated with differing sets of prosodic contours. These are also distinct from the categories of new and contrastive topics, which appear to share similar prosodic characteristics. Other initial hypotheses were partly validated. Namely, less accessible topics appear to pattern with higher fundamental frequency ranges and rising contours. However, our research also allowed us make several additional observations which merit further study, such as the apparent lack of a strong relationship between intonation and information structure in continuing topics, as well as the complexity posed by morphosyntactic factors.

Our approach suffered shortcomings often encountered with spontaneous speech. This study is based on a dataset of about 700 simple and complex prenuclei, yielding a limited number of tokens of the four subtypes of aboutness topics. Furthermore, the number of tokens of each of these subtypes is uneven, with resumed topics being relatively rare in our dataset. A larger dataset would have provided more tokens, allowing us to better consolidate the observations made in this paper, and make more meaningful generalizations about the role of intonation in left-detached topics.

Secondly, though the stylized contours generated by SLAM+ encode the positions of internal saliencies, we were unable to detect any link between these positions and the four topic categories. It would be reasonable to conclude that the position of internal saliencies has no relation to information structure. However, a later inspection of our data suggests that this issue is complicated by other factors. Notably, internal saliencies occurring at position 1 often correspond to the first lexical morpheme of the utterance. It is therefore possible that some contours are strongly influenced by morphosyntactic features such as whether the prenucleus begins with a functional or lexical word. Finally, when listening to several prenuclei with an initial saliency, we typically perceived the saliency itself as corresponding to the initial F0 value of the utterance. We therefore suspect that a melodic contour labeled mlH1 is perceptively equivalent to the simple falling contour Hl.

We will conclude by presenting a methodological approach we plan to implement in future studies in order to generate less cluttered and more interpretable MCA plots without losing information. Instead of using the global contour labels as statistical variables, we will introduce a set of labels which group them into categories and sub-categories based on shared intonational features. For example, a family of simple rising contours might cover all contours which lack an internal saliency, and which end with a higher F0 value than they begin. This category might be subdivided based on whether contours end with a high to extreme-high F0 value. These subcategories could also be subdivided based on exact final F0 value. This hypothetical classification is visualized in Figure 6.

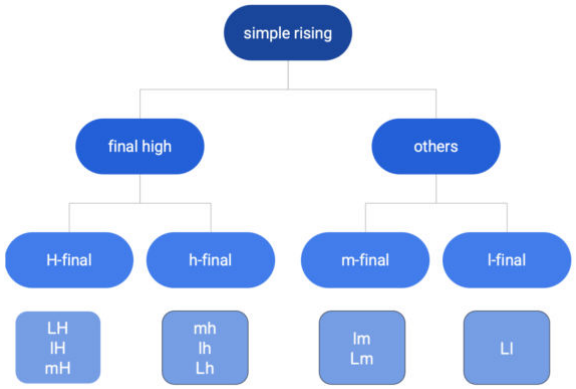

Figure 6: *Proposed subclassification of contour labels*

The main challenge going forward will be to determine which features to consider when forming these categories.

So far, we have managed to draw promising results from our data, and have identified several avenues to improve our methodology. We believe that the methods outlined in this paper could also serve as a basis for prosodic analysis in other languages, given the language-independent nature of the tools used.

5. Acknowledgements

We would like to recognize the French National Research Agency for financing the NaijaSynCor project[1] which made this this research possible. We also thank the members of the NaijaSynCor team for their commitment to the accuracy of the data used in our work.

[1] http://naijasyncor.huma-num.fr/index.html

6. Bibliography

[1] B. Bigi, B. Caron and A. Oyelere, "Developing Resources for the Automated Speech Processing of the African Language Naija (Nigerian Pidgin)," in *8th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznan, Poland, 2017.

[2] A. Lacheret, S. Kahane, J. Beliao, A. Dister, K. Gerdes, J.-P. Goldman, N. Obin, P. Pietrandrea and A. Tchobanov, "Rhapsodie: a Prosodic-Syntactic Treebank for Spoken French," in *Language Resources and Evaluation Conference*, Reykyavik, Iceland., 2014.

[3] E. Prince, "Toward a taxonomy of given-new information," in *Radical Pragmatics*, New York, Academic Press, 1981, pp. 223-256.

[4] K. Lambrecht, *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*, Cambridge University Press, 1994.

[5] A. Lacheret and J. François, "De la notion de détachement topical à celle de constituant thématique extrapropositionnel," *Cahiers de praxématique,* no. 40, pp. 167-198., 2003.

[6] S. Baumann and M. Grice, "The Intonation of Accessibility," *Journal of Pragmatics,* vol. 38, no. 10, pp. 1636-1657, 2006.

[7] T. Givon, *Topic Continuity in Discourse: A Quantitative Cross-Linguistic Study*, Philadelphia: John Benjamins, 1983.

[8] D. Bolinger, *Intonation and Its Parts*, Palo Alto: Stanford University Press, 1986.

[9] W. Chafe, "Topics and Point of View," in *Subject and Topic*, New York, Academic Press, 1976, pp. 25-55.

[10] C. Simard, *The Prosodic Contours of Jaminjung, a Language of Northern Australia,* Manchester: The University of Manchester, 2010.

[11] E. Adamou, K. Haude and M. Vanhove, *Information Structure in Lesser-described Languages*, Amsterdam: John Benjamins, 2018.

[12] M. Avanzi, A. Lacheret-Dujour and B. Victorri, "ANALOR. A Tool for Semi-Automatic Annotation of French Prosodic Structure," in *Speech Prosody*, Campinas, Brazil, 2008.

[13] K. Gerdes, B. Guillaume, S. Kahane and G. Perrier, "SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD," in *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, Brussels, Belgium, 2018.

[14] L. Liu, A. Lacheret-Dujour and N. Obin, "Automatic Modelling and Labelling of Speech Prosody: What's New with SLAM+," in *International Congress of Phonetic Sciences (ICPhS)*, Melbourne, Australia, 2019.

[15] M. Götze, C. Endriss, S. Hinterwimmer, I. Fiedler, S. Petrova, A. Schwarz, S. Skopeteas, R. Stoel and T. Weskott, "Information structure," in *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure.*, Potsdam, Universitätsverlag Postsdam, pp. 147-187.

[16] G. Desagulier, "Multiple Correspondence Analysis," in *Corpus Linguistics and Statistic with R: Introduction to Quantitative Methods in Linguistics*, Springer, 2017, pp. 268-276.