New Methods for Exploring Intonosyntax: Introducing an Intonosyntactic Treebank for Nigerian Pidgin

Emmett Strickland[1,2], Anne Lacheret-Dujour[1], Sylvain Kahane[1], Marc Evrard[2]
Perrine Quennehen[1], Bernard Caron[3], Francis Egbokhare[4], Bruno Guillaume[5]
[1]Paris Nanterre University, CNRS, MoDyCo, France
[2]Paris-Saclay University, CNRS, LISN, France
[3]CNRS, LLACAN, IFRA, France
[4]University of Ibadan, Nigeria
[5]Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
{emmett.strickland, anne.lacheret, sylvain.kahane, perrine.quennehen}@parisnanterre.fr
marc.evrard@lisn.upsaclay.fr, bruno.guillaume@inria.fr
bernard.l.caron@gmail.com, francis.egbokhare@gmail.com

Abstract

This paper presents a new phonetic resource for Nigerian Pidgin, a low-resource language of West Africa. Aiming to provide a new tool for research on intonosyntax, we have augmented an existing syntactic treebank of Nigerian Pidgin, associating each orthographically transcribed token with a series of syllable-level alignments and phonetizations. Syllables are further described using a set of continuous and discrete prosodic features. This new approach provides a simple tool for researchers to explore the prosodic characteristics of various syntactic phenomena. In this paper, we present the format of the corpus, the various features added, and several explorations that can be performed using an online interface. We also present a prosodically specified lexicon extracted using this resource. In it, each orthographic form is accompanied by the frequency of its phoneme-level variants, as well as the suprasegmental features that most frequently accompany each syllable. Finally, we present several additional case studies on how this corpus can be used in the study of the language's prosody.

Keywords: Nigerian Pidgin, Intonosyntax, Prosody, Treebank, Corpus Linguistics

1. Introduction

Nigerian Pidgin, or Naijá, is a large but poorly resourced contact language spoken by as many as 100 million people in West Africa. The majority of its vocabulary was inherited from English, though it has developed a distinct set of grammatical and prosodic features that distinguish it from its lexifier. One widely described feature is its apparent use of lexical tone, whose exact nature has been the subject of several analyses (Mafeni, 1971; Elugbe and Omamor, 1991; Faraclas, 1996).

We present NaijaSynCor-Prosody, a new resource that allows researchers to quantify both syntactic and prosodic structures in Naijá. This work is based on the NaijaSynCor corpus, a 30-hour treebank of transcribed Nigerian Pidgin. We first describe this corpus and how it was modified to facilitate the study of segmental, suprasegmental, and intonosyntactic features. We then showcase several concrete applications of this corpus, starting with an attempt at validating a tonal minimal pair. We then present a segmentally and prosodically specified lexicon in which each word is accompanied by its most common phonological and prosodic forms. We then describe an additional case study that combines the preexisting syntactic annotations with the new prosodic annotations in an exploration of the prosodic features of compounds. The studies presented in this paper are not intended to be exhaustive or conclusive, but rather to showcase the resource's potential as a tool for researchers.

2. Corpus description

2.1. Context: the NaijaSynCor treebank

The resource described in this paper is a continuation of the French National Research Agency (ANR) project NaijaSynCor (Bernard Caron, 2021), itself heavily influenced by the Rhapsodie project on spoken French (Anne Lacheret-Dujour, 2019). The NaijaSynCor project oversaw the development of a corpus of 500,000 tokens transcribed from 321 recordings of monologues and dialogues taking place in various social contexts. 88 recording sessions corresponding to roughly 150,000 tokens spread over eight dialogues and 80 monologues were manually annotated as dependency trees under the Surface Syntactic Universal Dependencies (SUD) scheme before being converted to Universal Dependencies (UD) (Gerdes et al., 2018). This annotation process was carried out by five Nigerian Pidgin speakers working full-time for 2 years in a shared office. All of them were masters and PhD-level students in linguistics recruited at the

LREC-COLING 2024, pages 12207–12216
20-25 May, 2024.

University of Ibadan. Annotations were reviewed by the project director, and contentious annotations were resolved through consensus in a workshop setting. This gold standard was then used to train a parser to automatically annotate the remaining files. Each transcribed recording session is segmented into illocutionary units (IUs) (Pietrandrea et al., 2014) represented as dependency trees encoded in the CoNLL-U tabular data format. IUs are also associated with various metadata including a numeric speaker ID, allowing researchers to access various sociolinguistic information such as the participant's age, sex, profession, and level of education. For a more detailed description of this corpus and its format, see (Kahane et al., 2021).

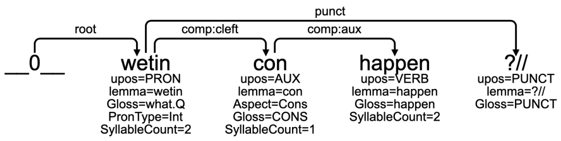

Figure 1: Example of Naijá dependency tree

In addition to the treebank and parser, the NaijaSynCor project led to the creation of several other tools, including a Wiktionary based on a lexicon extracted from the corpus, a phonetic alignment model, and a speech synthesis system for use in perceptual experiments (Manfredi et al., 2021; Bigi et al., 2020; Strickland et al., 2023a).

2.2. Building an intonosyntactic treebank

In building this resource, our goal was to preserve all morphosyntactic information in the original NaijaSynCor treebank while adding a detailed layer of segmental and suprasegmental information. Like a traditional treebank, this augmented corpus allows researchers to perform quantitative studies of various morphological and syntactic phenomena such as those presented in (Courtin et al., 2018). However, users will now have access to various segmental and suprasegmental information alongside the original annotations.

In building this resource, we used the alignment software SPPAS (Bigi et al., 2020) to produce a `.TextGrid` phonetic alignment for the 80 monologues in the gold standard corpus, whose orthographic transcriptions had been carefully verified over the course of the NaijaSynCor project. These alignments were then grouped into syllables, which are the primary unit of study in the field of prosody and broadly recognized as the tone-bearing unit (TBU) in tonal analyses of Nigerian Pidgin. The syllabic alignments and phonetic transcriptions were then manually verified and corrected by human annotators. Under our annotation scheme, speech sounds were represented using the SAMPA-X format. These syllabic transcriptions formed the basis of the segmental annotation in this corpus. The NaijaSynCor-Prosody corpus counts 112k syllables and 90k tokens, excluding the punctuation marks used to annotate pauses, utterance boundaries, and other phenomena.

To form our suprasegmental annotations, F0 information was extracted from the original audio recording of each monologue and encoded in the `.PitchTier` format. Pitch tracking errors were manually corrected using the software Analor (Lacheret and Victorri, 2002). The syllabified `.TextGrid` alignments and corresponding `.PitchTier` files were then used as input for the prosodic modeling software SLAM3 (Strickland et al., 2023b), the most recent version of the SLAM model (Obin et al., 2014; Liu et al., 2019). For each syllable, SLAM3 produced categorical textual labels describing the beginning and end pitch values of each syllable, as well as any internal F0 peaks. At each instance, two labels were produced: a global label expressing height values relative to the mean F0 of the utterance in which the syllable appears, and a local label expressing height relative to the mean F0 of the syllable's immediate context spanning from the beginning of the previous syllable to the end of the following syllable.

Several other prosodic features were also extracted from these labels, and are described below. In addition to the discrete SLAM labels and associated features, we also produced a set of mostly continuous labels using the `.PitchTier` and `.TextGrid` files directly. These include the duration in milliseconds and the average F0 in Hz. All of these features were assigned to each syllable in the corpus, with the exception of `UtteranceMeanF0`, which represents the mean F0 of the entire illocutionary unit. This feature was assigned to the syntactic root of each dependency tree. The full list of prosodic features assigned to each syllable is as follows:

1. **`Glo`**: SLAM3 contour with *m* tone centered on mean F0 of utterance.

2. **`Loc`**: SLAM3 contour with *m* tone centred on mean F0 of immediate context.

3. **`SlopeGlo`**: The slope extracted from the `Glo` label, with *Rise*, *Fall*, and *Flat* as possible values.

4. **`SlopeLoc`**: The slope extracted from the `Loc` label.

5. **`AvgHeightGlo`**: A rough categorical average of the pitch values represented in the `Glo` label, with *L*, *M*, and *H* as possible values.

6. **`AvgHeightLoc`**: A rough categorical average

of the pitch values represented in the `Loc` label.

7. **`Duration`**: The length of the syllable measured in milliseconds

8. **`MeanF0`**: The mean F0 of the syllable measured in Hz.

9. **`SemitonesFromUtteranceMean`**: number of semitones between `MeanF0` and `UtteranceMeanF0`. A negative value would denote a pitch lower than the utterance mean.

10. **`Slope`**: The slope of the contour extracted directly from a linear regression of pitch values in the `.PitchTier` file, with *Rise*, *Fall*, and *Flat* as possible values depending on whether the slope exceeds the glissando threshold described in (Hart et al., 1990).

Pitch-related features are occasionally annotated with the value `X` when no fundamental frequency was detected within the syllable. Additional information unrelated to pitch or duration was also encoded. An underlying `SylForm` feature allows for a graphical representation of the SAMPA-X phonetic transcription associated with each syllable. We also accounted for the many instances where two adjacent tokens are pronounced as a single syllable. Consider the `.TextGrid` alignment in Figure 2.

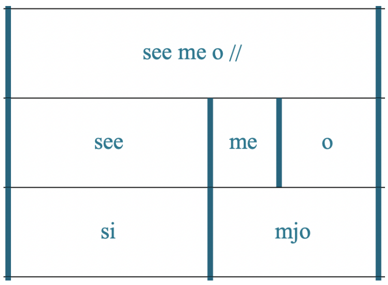

Figure 2: Two tokens pronounced as one syllable

From top to bottom, the three tiers in this example represent the IU-level alignment, the orthographic token-level alignment, and the phonetized syllable-level alignment. While there is a 1:1 alignment between the verb *see* and its syllabic representation, the pronoun *me* and the discourse marker *o* are pronounced as a single syllable even though they are represented as separate tokens in the treebank. In some rare instances, as many as three tokens can share a single syllable. The backend architecture we developed for this intonosyntactic treebank allows us to represent such cases.

We also account for a second type of incongruity between the token and syllable-level alignments. In our corpus, syllable-level alignments are liable to overstep word boundaries in cases where a word ending with a consonant is followed by one beginning with a vowel. A typical outcome in this scenario is for the consonant at the end of the first word to be pronounced as the onset of the second word's first syllable. The `.TextGrid` alignment in Figure 3 demonstrates this phenomenon.

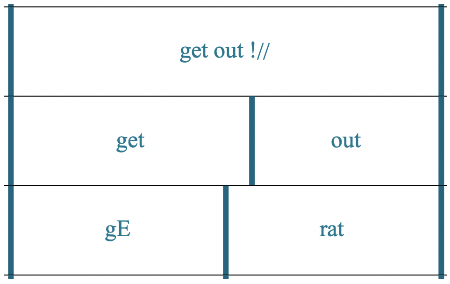

Figure 3: Final consonant of *get* realized as a rhotic onset on *out*

In our annotation scheme, the token *get* is annotated with the feature `SylForm=gE` while the token *out* is annotated with `SylForm=rat`. This second token is marked with the feature `ExternalOnset=Yes` to signal that the syllable's onset is a consonant from a previous word, in this case an underlying /t/ being realized as a rhotic. We find this system to be a useful and intuitive way to encode misalignments between syntactic and prosodic units, a problem also encountered during the production of the Rhapsodie corpus (Lacheret-Dujour et al., 2019). This example also highlights the treebank's encoding of highly noncanonical pronunciations which were obscured in the original NaijaSynCor treebank.

2.3. Data visualization

To visualize and explore the relationships between tokens and syllables, we use the GREW-Match tool (Guillaume, 2021) alongside a modified CONLL-U encoding which accounts for the various features described in the previous subsection. This allows us to observe the syntactic and syllabic data within the same interface to make observations or test linguistic hypotheses.

Figure 4 is an example of the encoding. Each utterance is represented as a graph with two kinds of nodes: word nodes (in black) and syllable nodes (in purple). Special (blue) edges are used to link words to the syllables of which they are comprised. These edges are labeled with a number corresponding to the position of that syllable within the word.

As discussed in section 2.2, syllable boundaries do not always align with word boundaries. Our system allows users to visually distinguish these cases. When a single syllable is shared by two tokens, the syllable node is placed between them

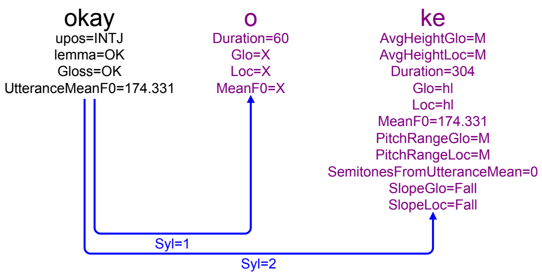

Figure 4: Example of data encoding

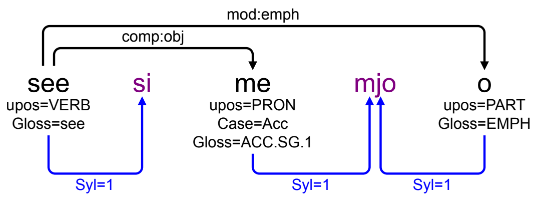

Figure 5: Example of a fused syllable (note that only a small subset of the features is shown)

on the graph and linked to each of them with a blue edge. Figure 5 provides the final representation of one such case, which also corresponds to the alignment represented in Figure 2.

Meanwhile, syllables with an external onset from a preceding word are distinguished with an orange edge connecting them to the final syllable of that word, as illustrated in Figure 6. This example again corresponds to the alignment in Figure 3.

2.4. Exploring the treebank with Grew-Match

Users can now request and quantify a wide range of syntactic and prosodic phenomena in the treebank using Grew syntax. As an illustration, we will use the case of go, which can function either as a future-marking auxiliary or as a verb of movement. Descriptions of Nigerian Pidgin typically reference these as a tonal minimal pair, with a low tone marking the auxiliary form and a high tone marking the verb. To see if this distinction is represented in the corpus, we wrote the following Grew request to

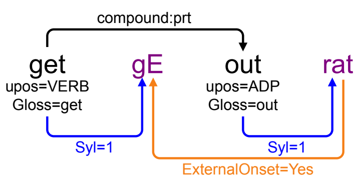

Figure 6: Example of an ExternalOnset annotation

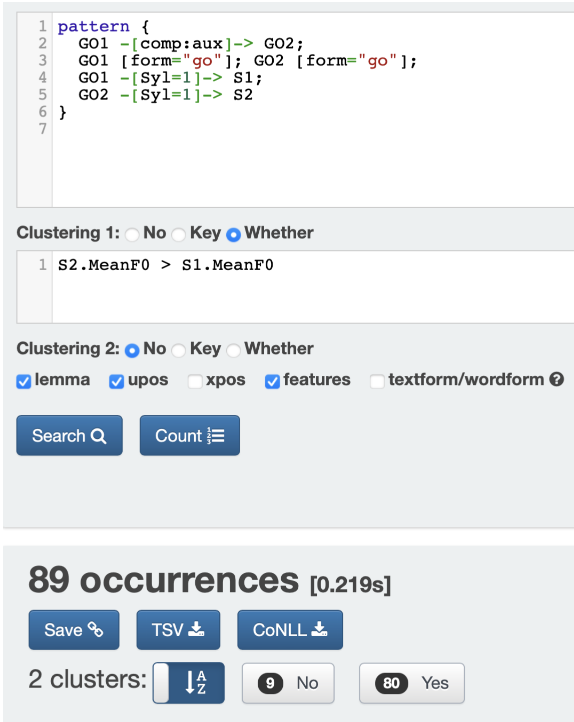

Figure 7: Grew-Match graphic interface

identify all cases in which a go annotated as an auxiliary was linked to a verbal go as its complement.

```
pattern {
  GO1 -[comp:aux]-> GO2;
  GO1 [form="go"]; GO2 [form="go"];
  GO1 -[Syl=1]-> S1; GO2 -[Syl=1]-> S2
}
```

The first two lines of this request locate two tokens (`GO1` and `GO2`) connected by a `comp:aux` (auxiliary complement) relation, both of which carry the form go. The remainder of the request ensures that both tokens are associated with separate syllable nodes (`S1` and `S2`), effectively excluding any instances in which both words constitute a single fused syllable.

The resulting constructions can then be grouped depending on whether they meet the condition `S2.MeanF0 > S1.MeanF0`, signifying a higher F0 on the syllable associated with the second go. Since each pair represents two words uttered by the same speaker in the same utterance, we can directly compare their mean F0 values in Hz.

The full request and corresponding results can be seen in Figure 7, which displays a screenshot of the Grew-Match graphic interface. Of the 89 occurrences of a verbal go linked to an auxiliary go, 80 matches feature a higher F0 on the verb. The data in the corpus is therefore compatible with the literature describing a difference in pitch between the two uses. A different query provides additional evidence in this view. To test whether this pitch

	Utterance Mean Exceeded	
	Yes	No
AUX	308	1549
VERB	386	457
p-val	p<0.000001	
odds radio	0.235	

Table 1: Occurrences of GO clustered by part of speech and mean pitch relative to utterance

S.AvgHeightLoc \ GO.upos	1854 AUX	843 VERB
1618 M	1013	605
781 L	759	22
298 H	82	216

Figure 8: Occurrences of GO clustered by part of speech and average height relative to immediate context

difference is a generalized feature of the language rather than something that is only observable when the two forms follow one another, we searched the corpus using the following query, which identifies any tokens that carry the form GO and are associated with a syllable bearing the feature `MeanF0`.

```
pattern {
  GO [form=go];
  GO -[Syl=1]-> S;
  S [MeanF0]
}
```

We then used the GREW-Match interface to cluster the results along two dimensions: the part of speech of each occurrence of GO, and whether the mean F0 of the associated syllable exceeded that of the utterance in which it appeared. The latter was defined with the condition `S.SemitonesFromUtteranceMean > 0`. Since this test involves comparing the prosodic qualities of GO from various speakers with different baseline F0s, using semitones from the utterance mean was the most appropriate metric in this instance.

The results in Table 1 show that the overwhelming majority of cases (83.4%), GO does not exceed the utterance mean when functioning as an auxiliary. However, the verbal equivalent exceeds this threshold in nearly half of cases (45.8%). We applied Fisher's Exact Test to this contingency table and achieved a p-value below 10^{-6}, indicating a high degree of statistical significance. Meanwhile, a calculated odds ratio of 0.235 demonstrates that auxiliaries are only 23.5% as likely as verbs to have a mean F0 exceeding the utterance mean.

At a glance, the results are even more striking when we modify the request to cluster along the `AvgHeightLoc` feature, which provides a rough categorical height for each occurrence relative to its local context, a more linguistically interesting metric than its height relative to the utterance mean. The output on GREW-Match in Figure 8 shows that low heights represent only 2.6% of verbs, but 40.9% of auxiliaries.

The tests showcased in this subsection are examples of quantitative research that can be performed directly using the GREW-Match interface. Our case study with GO shows that our resource provides multiple features that can be used to explore a given phenomenon, though some may be better-suited than others for a specific task. In our case, the use of discrete pitch values calculated relative to a syllable's immediate context yielded the most striking results. We encourage researchers using this resource to experiment with the different features outlined in 2.2 to decide which ones align best with their specific needs.

The corpus and a larger list of suggested queries are readily available on the Naijá page of the GREW-Match website[1]. The website also provides a more general tutorial[2] to those unfamiliar with the query syntax.

3. Further applications

The following section showcases several additional ways in which we have applied this intonosyntactic treebank to our own research on Nigerian Pidgin. Here, we intend not only to showcase some of the preliminary findings we have made using this resource, but also to provide inspiration for how such a resource might be used by other researchers exploring similar topics.

3.1. Extracting a prosodically specified lexicon

Over the course of the NaijaSynCor project, researchers used the most common transcribed tokens in the original treebank as the basis of a Wiktionary with a page for each of the most common words in the corpus (Manfredi et al., 2021). Speakers of Nigerian Pidgin were recruited to update each page with the corresponding word's definition, IPA pronunciation, and tone values on each syllable. While informative, this segmental and suprasegmental information corresponded to the words as they are pronounced in isolation. The IPA transcriptions were highly canonical, and typically decided by a single annotator. Meanwhile, final syllables were often assigned falling tones because

[1] https://naija.grew.fr/?corpus=SUD_Naija-NSC-prosody

[2] https://match.grew.fr/?tutorial=yes

pronouncing them in isolation introduced influence from utterance-final pitch lowering.

We leveraged the segmental and suprasegmental annotations introduced in this paper to produce a more data-driven alternative that describes each wordform as it is produced in spontaneous speech. For each token in the treebank, we extracted the part of speech and number of occurrences of each segmental variant; i.e., the SAMPA-X transcription reconstructed from the associated syllables. We also extracted various prosodic information, including the locations of syllables with the highest pitch and duration, and the slope of each syllable.

This allowed us to reconstruct an archetypical form for each entry based on the most frequent phonetic form and prosodic structure. Using speaker metadata, we also calculated the number of times each variant was uttered by a member of each sociolinguistic category related to sex, age, and birthplace. Finally, our lexicon also included the phonetic transcription extracted from the corresponding Wiktionary page when one was available, allowing for future comparisons between the canonical forms selected by annotators and the most common forms extracted from the spontaneous data in our corpus.

All segmental, prosodic, and morphological information in our lexicon was drawn directly from the data encoded in the intonosyntactic treebank. While inspecting the document, we noticed that the rarest segmental transcriptions typically corresponded to hapax variants, annotation inconsistencies, and dysfluent contexts such as false starts and hesitations. While identifying these unusual forms was particularly valuable for error mining, we decided to base most of our research on a version of the lexicon that includes only variants occurring at least five times in the corpus.

This lexicon comprises over 1423 different segmental realizations, with each of the 824 transcribed wordforms being associated with an average of 1.73 different segmental forms. This resource opens the door to new avenues of research regarding sociolinguistic influences on pronunciation, the tonal patterns of different parts of speech, and the diversity of tonal patterns on different words. Discussing all of these would fall beyond the scope of this paper, but we would like to highlight one focused study made possible by this lexicon. During our exploration of the data, we noticed that many of the canonically consonant-final entries were pronounced with the final phoneme elided. In some cases, the most common vowel-final variant was nearly as common as the dominant consonant-final form. Noticing that many vowel-final forms seemed predominantly uttered by men, we calculated the total number of occurrences of vowel-final and consonant-final variants uttered by each sex.

	Full lexicon		**Restricted Lexicon**	
	C-Fin.	**V-Fin.**	**C-Fin.**	**V-Fin.**
M	7.2k	36k	5.6k	14k
F	5.4k	21k	4.1k	7.5k
p-val	p<0.000001		p<0.000001	
O.R.	0.773		0.742	

Table 2: Occurrences of consonant and vowel-final variants by speaker sex

Two variants of this study were performed. A full lexicon variant counting 69,086 total word instances considered all phonetic forms present in the lexicon, including those with exclusively consonant or vowel-final variants. Meanwhile, a restricted lexicon variant of 30,160 instances counted only wordforms when they were associated with both consonant and vowel-final pronunciations.

The results in Table 2 show that in both cases, consonant-final forms constitute a far higher proportion of all productions among females than they do among males. This effect is observable regardless of whether the study is limited to the full lexicon. In both cases, we applied Fisher's Exact Test to the results and achieved a p-value below 10^{-6}, indicating a statistically significant correlation between sex and the eliding of word-final consonants. The tests yielded odds ratios of 0.742 and 0.773, meaning that words uttered by men are 74.2% to 77.3% as likely to end in a consonant as those uttered by women. This effect is naturally somewhat stronger when the study is limited to words with both consonant and vowel-final variants. We interpret these results to mean that males are somewhat more likely to elide consonants and consequently show greater variation between consonant and vowel-final forms, while females are somewhat more likely to preserve word-final consonants. This trend went unnoticed for the entirety of the NaijaSynCor project, but became readily apparent thanks to the exploratory techniques enabled by our intonosyntactic treebank. Further analysis will be required to prove if other factors may be playing a role, but this finding appears consistent with sociolinguistic literature showing that women are more likely to use prestigious varieties (Labov, 1990).

3.2. The prosody of compounds and modifiers

One additional linguistic question that arose during our exploration of the corpus was whether there exists a prosodic difference between words connected by a `mod` relation, the default dependency relation linking nouns to adjectives and other modifiers, and those connected by a `compound` relation. In our corpus, the latter is intended for use in cases

Form	Lemma	POS	Pronunciation	Occurrence	Highest_Syl	Highest_Syl_Consistency	Longest_Syl
fight	fight	VERB	fal	2	Syl1	100.00%	Syl1
fight	fight	VERB	falt	5	Syl1	100.00%	Syl1
fight	fight	VERB	falti	1	Syl1	100.00%	Syl1
fight	fight	NOUN	fal	1	Syl1	100.00%	Syl1
find	find	VERB	fal	9	Syl1	100.00%	Syl1
find	find	VERB	faln	7	Syl1	100.00%	Syl1
find	find	VERB	fal~	25	Syl1	100.00%	Syl1
find	find	VERB	fal~n	8	Syl1	100.00%	Syl1
fine	fine	ADJ	fal(n)	3	Syl1	100.00%	Syl1

Figure 9: Assorted features in lexicon

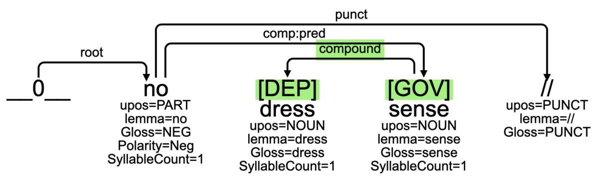

Figure 10: Example of dependency tree containing monosyllabic words linked by a `compound` relation

where a word pairing has become lexicalized, and carries a semantic value that cannot be adequately described as a simple combination of its components. According to the SUD annotation guidelines used in this corpus[3], the English examples *green cars* or *green shirts* would typically be annotated with a `mod` relation linking the noun to its modifier. However, *green beans* and *green screens* would be annotated with a `compound` relation because they refer to concepts more specific than beans and screens, which are green. The line between the two is sometimes ambiguous, and the `compound` relation was typically based on the intuitions of the annotators.

We applied the intonosyntactic corpus to a machine learning task designed to explore any prosodic differences between modifiers and compounds. We extracted from the corpus adjacent words linked either by a `compound` relation or a `mod` relation. We chose to limit this study to monosyllabic pairs in order to limit the number of prosodic factors at play, and facilitate direct comparisons between the height and length of each component word. We also included only constructions that have a noun as the syntactic head, like those in the above examples. For each pair in the dataset, we extracted various prosodic features from the treebank. Namely, which of the two token's syllables had a higher value for `duration` and `MeanF0`, as well as various non-relative characteristics of each

[3]https://surfacesyntacticud.github.io/guidelines/pcm/

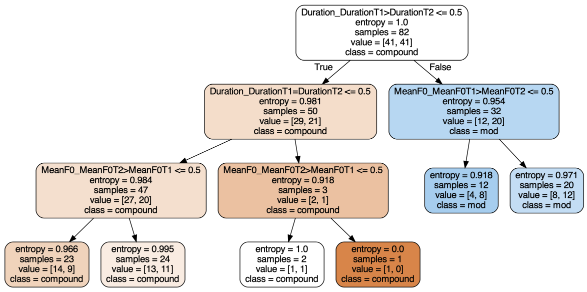

Figure 11: Example of decision tree based on relative `duration` and `MeanF0`

syllable such as `slope`. The dataset included 82 pairs linked by a `compound` relation and 549 pairs linked by a `mod` relation. To ensure parity between the sizes of the two categories, we randomly selected a subset of 82 `mod` pairs.

The 164 total instances were then divided into train and test sets of equal size before being fed into the `DecisionTreeClassifier` provided by `scikit-learn`. The various prosodic features were provided as input variables to predict the syntactic relation connecting each word pair. To ensure a balanced representation of `mod` and `compound` labels in the train and test sets, instances were also stratified by label. When repeating the experiment, results differed significantly from iteration to iteration. This is presumably the result of randomizing the composition of our train and test sets over such a small pool of data. However, decision trees based on the relative `duration` and `MeanF0` of the two tokens appeared to generate the most consistent and often highest accuracy scores.

The example in Figure 11 shows one such tree produced using these variables, which achieved an overall F-score of 0.60. In this particular instance, `duration` was sufficient to classify instances into the categories of `compound` (orange) and `mod` (blue). More specifically, pairs in which the first token was longer than the second were predicted to have a `mod` relation, while pairs in which the second

	MeanF0		Duration	
	Mod	Comp.	Mod	Comp.
T1>T2	304	43	213	22
T2>T1	245	39	332	58
p-val	p=0.636		p=0.048	
odds ratio	1.125		1.691	

Table 3: Relative `duration` and `MeanF0` of tokens in monosyllabic `compound` and `mod` pairs

	MeanF0		Duration	
	Male	Female	Male	Female
T1>T2	16	27	33	15
T2>T1	18	21	25	9
p-val	p=0.502		p=0.806	
odds ratio	0.691		0.792	

Table 4: Relative `duration` and `MeanF0` in monosyllabic `compound` pairs uttered by each sex

token was longer were predicted to have a `compound` relation. This pattern frequently emerged across different randomizations of the dataset, suggesting that this was the most reliable and universal feature distinguishing the two syntactic labels. This result was surprising given that most descriptions of the language's prosody treat duration as secondary to pitch. We decided to calculate the statistical significance of both of these binary features across the entire corpus using Fisher's Exact Test. The results in Table 3 confirm a statistically significant trend towards a longer second token in monosyllabic pairs annotated with a `compound` relation, with a p-value below the traditional threshold of 0.05. However, no significant correlation was found for pitch.

We suspected that certain sociolinguistic variables may have played a role in determining whether compounds were marked using pitch or duration. Using speaker metadata, we calculated contingency tables for compounds uttered by speakers belonging to each sociolinguistic category within the fields of sex, age, and birthplace. For each of these categories, separate tables were calculated for the variables `duration` and `MeanF0`, and p-values calculated using Fisher's Exact Test. For an example, see the contingency tables represented in Table 4. Note that sex was the only field originally treated as a binary in the metadata. When calculating the p-values for the different age categories and birthplaces, we compared compounds uttered by members of each category against those uttered by persons outside of that category, e.g., speakers from Abuja against those born outside of Abuja.

The results presented in Table 5 fail to show a statistically significant relationship between any sociolinguistic category and the use of pitch or duration to mark compounds. Because none of the p-values fell below 0.05, effect sizes were excluded for the sake of readability. A relatively low p-value of 0.083 was achieved when examining relative duration in speakers from Oyo State, but this was still above the traditional threshold of 0.05. One interpretation of these results is that the use of pitch or duration in the marking of compounds is uniform across all segments of society studied in this paper.

	MeanF0	Duration
Sex		
Male	0.502	0.806
Female	0.502	0.806
Age		
<16	1.0	1.0
16-30	0.827	1.0
31-45	0.825	0.629
46-60	1.0	0.315
Birthplace		
Abia	1.0	0.292
Abuja	0.703	0.667
Akwa Ibom	0.476	1.0
Anambra	0.476	1.0
Bayelsa	1.0	1.0
Cross Rivers	1.0	1.0
Delta	0.318	0.148
Edo	0.665	1.0
Enugu	1.0	0.552
FCT	1.0	1.0
Kaduna	0.342	0.577
Katsina	0.476	1.0
Kogi	1.0	1.0
Lagos	0.269	1.0
Niger	1.0	0.552
Ofa	1.0	1.0
Ogun	1.0	0.293
Ondo	0.223	0.502
Osun	0.602	1.0
Oyo	1.0	0.083
Plateau	0.602	0.204
Rivers	0.476	1.0
Sokoto	1.0	0.766

Table 5: P-values calculated from each sociolinguistic category

Given the complexities of this dataset, which involves multiple variables and potential interactions between them, a multivariate multivariate analyses such as multiple correspondence analysis or linear mixed-effects models would be more appropriate. Going forward, we will apply a more robust statistical analysis to this particular problem to draw more definitive conclusions. Dividing speakers into a large number of categories depending on their state of birth may also be a less than ideal approach to the geographic dimension of this research, and may be obfuscating meaningful trends. In future tests, we will try grouping these categories into larger regions and introducing other sociolinguistic variables.

4. Concluson

We have presented a new method for exploring intonosyntax through this new resource for Nigerian Pidgin, and have provided several concrete examples of its utility to researchers. While all of the case studies presented here are illustrative and demand more complete analyses, they have already demonstrated a likely relationship between speaker sex and final-consonant elision, and a possible role of duration in the marking of compounds. To our knowledge, neither of these was documented prior to the construction of this resource.

Carrying out the experiments presented in this paper also brought our attention to how this resource might be improved and expanded upon in the future. To revisit our experiments using prosodic variables in decision trees, we found that the various models we produced consistently predicted `compound` labels with far greater accuracy than `mod` labels. For example, the decision tree model represented in Figure 11 yielded an F-score of 0.65 for the `compound` label compared to only 0.52 for `mod`. One possible interpretation of this result is that compounds have a more rigid prosodic structure, or occur in a more restricted set of prosodic environments. While this possibility will be explored in future research, a more likely culprit is inconsistency in the syntactic annotation. It seems plausible that the syntactic annotators only attributed the rarer `compound` relation in cases where two tokens clearly and obviously represented a compound, and the more general `mod` relation in more ambiguous cases. The abstract nature of the `compound` relation, which has no obvious syntactic criteria, may also have been challenging to annotators. A preliminary exploration of the corpus has revealed various inconsistencies in the way these two labels are applied. We believe that revisiting these annotations would greatly improve the accuracy of our decision trees, and further validate the role of duration in distinguishing these constructions.

This opens the door to an additional potential application of the intonosyntactic data presented in this paper. If prosodic data can be used as a predictor for certain syntactic labels, it stands to reason that they might also be used as criteria when performing automatic corrections to improve future iterations of this corpus. A primary objective going forward will be to apply the methods presented in this paper to the remainder of the NaijaSynCor treebank, which was annotated using a parser. We believe that exploiting the sorts of prosodic information presented in this paper would be particularly useful for improving the automatic annotations of certain ambiguous constructions in this portion of the treebank, most notably compounds.

Expanding the intonosyntactic treebank to this portion of the NaijaSynCor corpus would also achieve a second goal, which is expanding the number of occurrences of certain rare constructions. One of the difficulties of studying compounds in this paper was the relative rarity of monosyllabic word pairs. While we plan to devise methods to study word pairs of various syllable structures, having a larger dataset could only be an asset to researchers working on Nigerian Pidgin going forward.

5. Bibliographical References

Brigitte Bigi, Oyelere S Abiola, and Bernard Caron. 2020. Resources and tools for automated speech segmentation of the african language Naija (Nigerian Pidgin). In *Human Language Technology. Challenges for Computer Science and Linguistics: 8th Language and Technology Conference, LTC 2017, Poznań, Poland, November 17–19, 2017, Revised Selected Papers 8*, pages 164–173. Springer.

Marine Courtin, Bernard Caron, Kim Gerdes, and Sylvain Kahane. 2018. Establishing a language by annotating a corpus. In *annDH 2018 Annotation in Digital Humanities*, volume 2155, pages 7–11. CEUR.

Ben Ohiọmamhẹ Elugbe and Augusta Phil Omamor. 1991. *Nigerian Pidgin: Background and Prospects*. Heinemann Educational Books Nigeria PLC.

Nicholas Faraclas. 1996. *Nigerian Pidgin*. Routledge.

Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. 2018. SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In *Universal dependencies workshop 2018*.

Bruno Guillaume. 2021. Graph matching and graph rewriting: Grew tools for corpus exploration, maintenance and conversion. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 168–175.

Johan't Hart, René Collier, and Antonie Cohen. 1990. *A Perceptual Study of Intonation*. Cambridge University Press.

Sylvain Kahane, Bernard Caron, Emmett Strickland, and Kim Gerdes. 2021. Annotation guidelines of UD and SUD treebanks for spoken corpora. In *Proceedings of the 20th International*

Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021), pages pp–35. Association for Computational Linguistics.

William Labov. 1990. The intersection of sex and social class in the course of linguistic change. *Language variation and change*, 2(2):205–254.

Anne Lacheret and Bernard Victorri. 2002. La période intonative comme unité d'analyse pour l'étude du français parlé: modélisation prosodique et enjeux linguistiques. *Verbum: Analecta Neolatina*, 1(24):55–72.

Anne Lacheret-Dujour, Sylvain Kahane, Rachel Bawden, Serge Fleury, and Ilaine Wang. 2019. Exploration of the rhapsodie corpus: Data structure, formats and query tools. In *Rhapsodie*, pages 271–283. John Benjamins.

Luigi Liu, Anne Lacheret-Dujour, and Nicolas Obin. 2019. Automatic Modelling and Labelling of Speech Prosody: What's New with SLAM+?? In *International Congress of Phonetic Sciences (ICPhS)*, Melbourne, Australia. International Phonetic Association and Australasian Speech Science and Technology Association, Australasian Speech Science and Technology Association Inc.

Bernard Mafeni. 1971. Nigerian pidgin. *The English Language in West Africa*, pages 95–112.

Stefano Manfredi, Bernard Caron, Kim Gerdes, and Marine Courtin. 2021. NaijaSynCor: A Syntactic Treebank, a Parser and a Wiktionary for Naija. In *Summer Conference of the Society of Pidgin and Creole Linguistics*, En ligne (Paris), France. SeDyL, INALCO.

Nicolas Obin, Julie Beliao, Christophe Veaux, and Anne Lacheret. 2014. SLAM: Automatic stylization and labelling of speech melody. In *Speech prosody*, page 246.

Paola Pietrandrea, Sylvain Kahane, Anne Lacheret-Dujour, and Frédéric Sabio. 2014. The notion of sentence and other discourse units in corpus annotation. *Spoken corpora and linguistic studies*, pages 331–364.

Emmett Strickland, Dana Aubakirova, Dorin Doncenco, Diego Torres, and Marc Evrard. 2023a. NaijaTTS: A pitch-controllable TTS model for Nigerian Pidgin. In *ISCA Speech Synthesis Workshop*, Grenoble, France.

Emmett Strickland, Marc Evrard, and Anne Lacheret-Dujour. 2023b. SLAM 3: An updated stylization model for speech melody. In *International Congress of Phonetic Sciences*, Prague, Czech Republic.

6. Language Resource References

Anne Lacheret-Dujour. 2019. *Rhapsodie Treebank*. MoDyCo. Available via http://rhapsodie.modyco.fr.

Bernard Caron. 2021. *NaijaSynCor Treebank*. LLACAN, MoDyCo, University of Ibadan. Available via https://naijasyncor.huma-num.fr.