

De nouvelles méthodes pour l'exploration de l'interface syntaxe-prosodie : un treebank intonosyntaxique et un système de synthèse pour le pidgin nigérian

Emmett Strickland^{1,2} Anne Lacheret-Dujour¹ Marc Evrard² Sylvain Kahane¹
Dana Aubakirova² Dorin Doncenco² Diego Torres² Perrine Quennehen¹
Bruno Guillaume³

(1) MoDyCo, 200 Avenue de la République, 92001 Nanterre, France

(2) LISN, Campus Universitaire, bât.507 Rue du Belvédère, 91405 Orsay, France

(3) LORIA, Campus Scientifique, 615 Rue du Jardin-Botanique, 54506 Vandœuvre-lès-Nancy, France

{emmett.strickland, anne.lacheret, sylvain.kahane,
perrine.quennehen}@parisnanterre.fr, marc.evrard@lisn.paris-saclay.fr,
bruno.guillaume@inria.fr, {dana.aubakirova, dorin.doncenco,
diego-andres.torres-guarin}@universite-paris-saclay.fr

RÉSUMÉ

Cet article présente deux ressources récemment développées pour explorer l'interface prosodie-syntaxe en pidgin nigérian, une langue à faibles ressources d'Afrique de l'Ouest. La première est un treebank intonosyntaxique dans laquelle chaque token est associé à une série de traits prosodiques au niveau de la syllabe, ce qui permet d'analyser diverses structures syntaxiques et prosodiques en utilisant une même interface. La seconde est un système de synthèse de la parole entraîné sur le même ensemble de données, conçu pour permettre un contrôle direct sur les contours intonatifs de la parole générée. Cet outil a été développé pour nous permettre de tester les hypothèses formulées à partir de l'exploration du treebank. Cet article est largement une adaptation de deux publications récentes présentant chaque outil, avec un accent sur leur interconnexion dans notre recherche en cours.

ABSTRACT

New methods for exploring the prosodic-syntactic interface : an intonosyntactic treebank and speech synthesis system for Nigerian Pidgin.

This paper presents two newly-developed resources for exploring the prosody-syntax interface in Nigerian Pidgin, a low-resource language of West Africa. The first is an intonosyntactic treebank in which tokens are associated with a series of syllable-level prosodic features, allowing for analyses of various syntactic and prosodic structures within the same interface. The second is a speech synthesis system trained on the same dataset which is designed to allow for direct control over the prosodic contours of generated speech. This tool was developed to allow us to test hypotheses made from the exploration of this treebank. This paper is largely an adaptation of two recent publications presenting these tools with an emphasis on their interconnectedness in our ongoing research.

MOTS-CLÉS : pidgin nigérian, linguistique de corpus, prosodie, treebank, synthèse de la parole.

KEYWORDS: Nigerian Pidgin, corpus linguistics, prosody, treebanks, speech synthesis.

1 Introduction et contexte

Le pidgin nigérian, ou naijá, est une langue créole peu dotée pourtant parlée par quelque 100 millions de locuteurs en Afrique de l’Ouest. Si l’essentiel de son lexique a été hérité de l’anglais, il a développé un ensemble de caractéristiques prosodiques et grammaticales qui le distingue de son lexifieur. Notamment, plusieurs analyses de cette langue postulent l’existence d’un système de ton lexical, même si ses caractéristiques exactes diffèrent selon l’analyse (Mafeni, 1971; Elugbe & Omamor, 1991; Faraclas, 1996).

Cette publication s’inscrit dans un projet de recherche en cours visant à mieux comprendre la prosodie du pidgin nigérian en utilisant une méthode novatrice qui combine la linguistique de corpus et la linguistique expérimentale. Concrètement, notre objectif est d’exploiter deux outils qui ont été produits parallèlement et qui permettent, respectivement, d’explorer des structures intonosyntaxiques dans un corpus de pidgin nigérian parlé, et de tester des hypothèses extraites de ce corpus dans un contexte expérimental. Cette approche permettra donc de découvrir de nouvelles phénomènes dans un corpus de parole spontanée, et de les valider grâce à des expériences perceptives contrôlées. Ces deux outils, le corpus NaijaSynCor-Prosody (Strickland *et al.*, 2024), et le système de synthèse NaijaTTS (Strickland *et al.*, 2023a), ont été développés simultanément comme des ressources complémentaires produites à partir des mêmes données. Les deux ressources ont déjà fait l’objet de deux publications, que cet article prolonge pour mettre en évidence leurs rôles complémentaires dans une méthodologie partagée.

2 NaijaSynCor-Prosody : Un treebank intonosyntaxique

Cette section présente le corpus NaijaSynCor-Prosody, une extension récente du corpus NaijaSynCor. Elle adapte une publication antérieure (Strickland *et al.*, 2024) qui a introduit ce corpus, et explique son rôle dans notre flux de travail expérimental plus large.

Le corpus NaijaSynCor-Prosody s’inscrit dans la continuité du projet ANR NaijaSynCor (Manfredi *et al.*, 2021), ainsi que du projet Rhapsodie sur le français parlé (Lacheret-Dujour *et al.*, 2019). Le projet NaijaSynCor a piloté le développement d’un corpus de 500 000 tokens transcrits à partir de 321 enregistrements de monologues et de dialogues se déroulant dans divers contextes sociaux. Un ensemble de 88 sessions d’enregistrement correspondant à environ 150 000 tokens répartis sur huit dialogues et 80 monologues ont été annotées manuellement en arbres de dépendance selon le schéma des *Surface Syntactic Universal Dependencies* (SUD) avant d’être converties en *Universal Dependencies* (UD, Gerdes *et al.*, 2018). Ce *gold standard* a ensuite été utilisé pour entraîner un analyseur syntaxique afin d’annoter automatiquement les fichiers restants (Guiller, 2020). Chaque session d’enregistrement transcrite est segmentée en unités illocutoires (UI) (Pietrandrea *et al.*, 2014) représentées sous forme d’arbres de dépendance encodés sous le format de données tabulaires CoNLL-U. Les UI sont également associées à différentes métadonnées, dont un identifiant numérique du locuteur, ce qui permet aux utilisateurs d’accéder à diverses informations sociolinguistiques telles que l’âge, le sexe, la profession et le niveau d’éducation. Se référer à Kahane *et al.* (2021) pour description plus détaillée de ce corpus et de son format.

En construisant le corpus NaijaSynCor-Prosody, notre objectif était de préserver les informations morphosyntaxiques du treebank NaijaSynCor original et d’ajouter une couche détaillée d’informations segmentales et suprasegmentales. Comme un treebank traditionnel, ce corpus augmenté permet aux

utilisateurs d'effectuer des études quantitatives de divers phénomènes morphosyntaxiques tels que ceux présentés par [Courtin et al. \(2018\)](#). Cependant, les utilisateurs auront désormais accès à diverses informations segmentales et suprasegmentales en plus des annotations originales.

Pour construire cette ressource, nous avons utilisé le logiciel d'alignement SPPAS ([Bigi et al., 2020](#)) afin de produire un alignement phonétique `.TextGrid` des 80 monologues du corpus de référence, dont les transcriptions orthographiques avaient été soigneusement vérifiées au cours du projet NaijaSynCor. Ces alignements ont ensuite fait l'objet d'une segmentation en syllabes, l'unité porteuse de ton dans plusieurs analyses du pidgin nigérian. Les alignements syllabiques et les transcriptions phonétiques ont ensuite été vérifiées et corrigées manuellement par des annotateurs. Dans le cadre de notre système d'annotation, les phonèmes sont représentés à l'aide du format X-SAMPA ([Wells, 1995](#)). Ces transcriptions syllabiques constituent la base de l'annotation segmentale de ce corpus.

Nos annotations suprasegmentales consistent en une mesure de la fréquence fondamentale (F0) appliquée aux enregistrements audio des monologues et conservée au format `.PitchTier`. Les erreurs de pitchtracking ont été corrigées manuellement à l'aide du logiciel Anamor ([Lacheret & Victorri, 2002](#)). Les alignements de chaque syllabe ainsi que leurs transcriptions phonétiques sont conservées au format `.TextGrid`. Ces deux formats de fichier ont ensuite été utilisés comme données d'entrée pour le logiciel de modélisation prosodique SLAM3 ([Strickland et al., 2023b](#)), la version la plus récente du modèle SLAM ([Obin et al., 2014](#); [Liu et al., 2019](#)). Pour chaque syllabe, SLAM3 a produit des étiquettes catégorielles décrivant les valeurs de hauteur de début et de fin de chaque syllabe, ainsi que les maxima de F0.

Plusieurs autres traits prosodiques ont également été extraits des étiquettes SLAM, notamment sa hauteur moyenne catégorielle et sa pente. Nous avons également produit un ensemble d'étiquettes continues en utilisant les fichiers `.PitchTier` et `.TextGrid`, notamment la durée (en ms) de chaque syllabe ainsi que sa F0 moyenne (en Hz)

Pour visualiser les relations entre les tokens et les syllabes, nous utilisons l'outil GREW-Match ([Guillaume, 2021](#)) avec un encodage CoNLL-U modifié pour prendre en compte les différents traits décrivant les syllabes. Cela permet d'observer les données syntaxiques et syllabiques via la même interface pour faire des observations ou tester des hypothèses linguistiques.

La figure 1 est un exemple de l'encodage utilisé. Chaque énoncé est représenté sous la forme d'un graphe avec deux types de nœuds : les nœuds de mots (en noir) et les nœuds de syllabes (en violet). Des arêtes spécifiques (bleues) sont utilisées pour relier les mots aux syllabes qui les composent. Ces arêtes sont étiquetées avec un numéro correspondant à la position de la syllabe correspondante dans le mot.

Les utilisateurs peuvent désormais quantifier un large éventail de phénomènes à l'interface de la syntaxe et de la prosodie à partir du treebank en utilisant la syntaxe GREW. À titre d'illustration, nous présenterons le cas de GO, qui peut fonctionner soit comme un auxiliaire marquant le futur, soit comme un verbe de mouvement. Les descriptions du pidgin nigérian les présentent généralement comme une paire minimale tonale, avec un ton bas marquant la forme auxiliaire et un ton haut marquant le verbe. Pour vérifier si cette distinction est représentée dans le corpus, nous utilisons la requête GREW suivante.¹

```
pattern {
```

1. cette requête et les résultats peuvent également être visualisés ici : <https://naija.grew.fr/?custom=664205b96cfe>

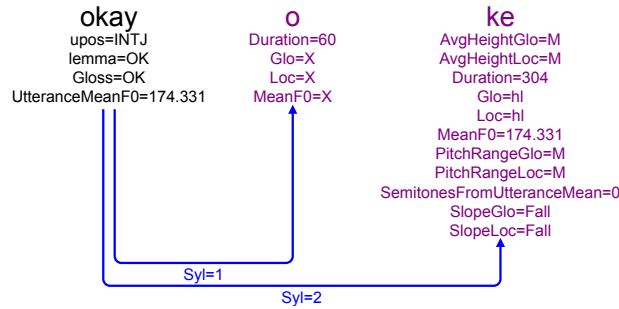


FIGURE 1 – Un token bisyllabique du corpus NaijaSynCor-Prosody

```

G01 -[comp:aux]-> G02;
G01 [form="go"]; G02 [form="go"];
G01 -[Syl=1]-> S1; G02 -[Syl=1]-> S2 }

```

Les deux premières lignes de cette requête localisent deux tokens (G01 et G02) reliés par une relation `comp:aux` (complément auxiliaire), qui portent tous deux la forme GO. Le reste de la requête garantit que les deux tokens sont associés à des de syllabes distinctes (S1 et S2), ce qui exclut les cas où les deux mots sont prononcés comme une seule syllabe fusionnée.

Les constructions résultantes peuvent ensuite être regroupées selon qu'elles remplissent ou non la condition $S2.MeanF0 > S1.MeanF0$, ce qui signifie une F0 plus élevée sur la syllabe associée au second GO. Étant donné que chaque paire représente deux mots prononcés par le même locuteur dans le même énoncé, nous pouvons comparer directement leurs valeurs moyennes de F0 en Hz. Ce test vise à identifier d'éventuelles différences de hauteur entre les deux parties du discours dans ce contexte particulier. Si, comme le décrit la littérature, le marqueur de futur doit avoir une hauteur inférieure au verbe de mouvement, on s'attendrait à ce que la forme auxiliaire soit plus basse que la forme verbale dans les contextes où les deux formes coïncident. Sur les 89 occurrences d'un GO verbal lié à un GO auxiliaire, 80 instances (89.9%) présentent une F0 plus élevée sur le verbe. Les résultats corroborent donc partiellement la littérature décrivant une différence de hauteur entre les deux utilisations.

Il est toutefois intéressant de noter qu'en modifiant cette requête pour comparer la durée relative de chaque instance de GO, on constate qu'en plus d'avoir une F0 plus basse, les auxiliaires sont majoritairement plus courts. Dans 73 cas sur 89 (82%), l'auxiliaire est plus court que le verbe. À notre connaissance, cette différence de durée n'a jamais été documentée dans les études sur cette langue. Une interprétation possible est que les syllabes qui portent un ton bas sont phonétiquement plus courts que celles qui portent un ton haut, mais que cette différence joue un rôle minimal dans la perception. Une autre interprétation est que la paire minimale examinée dans cet article n'est pas distinguée par le ton, mais par une combinaison d'une hauteur et d'une durée élevées. Dans une langue à accent lexical, comme l'anglais, les auxiliaires et d'autres grammèmes monosyllabiques ne seraient pas accentués, alors que les verbes et d'autres mots lexicaux portent toujours un accent. Si le pidgin nigérian est aussi une langue à accent lexical, les différences observées entre le GO grammatical et le verbe de mouvement seraient équivalentes à celles qui existent entre *butt* et le grammème non-accentué *but* en anglais.

Le simple test présenté dans cet article est loin d’être suffisant pour fournir des preuves solides pour ou contre cette analyse. Cependant, cela ouvre la voie à de nouvelles questions sur le rôle de la hauteur et de la durée dans la désambiguïsation de certains mots et constructions. Au fur et à mesure que nous explorons ce corpus, de telles hypothèses peuvent être validées par l’utilisation de l’outil présenté dans la section suivante, un système de synthèse vocale conçu pour des expériences perceptives.

3 NaijaTTS : vers un système de synthèse pour valider les hypothèses

Dans la section précédente, nous avons vu comment l’exploration de cette ressource peut mener à de nouvelles hypothèses sur le rôle de la durée et de la F0 en pidgin nigérian, et de la typologie prosodique de cette langue. Si d’autres tests révèlent une tendance généralisée et statistiquement significative dans l’ensemble du corpus, elle mérite d’être confirmée dans un cadre expérimental contrôlé. Cette section présente NaijaTTS, un système de synthèse (TTS) conçu pour tester les hypothèses faites pendant les explorations du treebank NaijaSynCor-Prosody. Cette section adapte [Strickland et al. \(2023a\)](#) et explique sa relation avec le corpus décrit dans la section précédente.

NaijaTTS a été entraîné sur les fichiers *gold* du corpus NaijaSynCor, correspondant à environ 7,5 heures de parole et 80 locuteurs. Pour garantir la fiabilité de nos données d’entraînement, nous avons limité ce projet à ces fichiers, car leur transcription et alignement syllabique ont été vérifiés par des annotateurs. Les données d’entraînement sont donc basées sur les mêmes fichiers sons et alignements utilisés pour produire le corpus NaijaSynCor-Prosody.

Nous avons basé notre système de synthèse vocale sur la plateforme FastSpeech 2 (FS2) ([Ren et al., 2022](#)), un système de TTS neuronal de bout en bout non autorégressif entraîné directement sur des entrées `.wav`, dont il extrait des descripteurs de durée, de hauteur et d’énergie pour les utiliser lors de l’entraînement. L’architecture de FS2 inclut un encodeur qui convertit la séquence d’encastrement des phonèmes (*phoneme embedding sequence*) en une séquence cachée des phonèmes. Ensuite, un *variance adaptor* ajoute des informations relatives à la durée, la hauteur, et l’intensité prédits pour chaque phonème. Ensuite, cette séquence cachée est convertie en mel-spectrogrammes par un décodeur. FS2 a été choisi principalement pour le *variance adapter*. Nous intervenons dans cette partie du pipeline pour remplacer les valeurs prédites par nos propres vecteurs contenant une valeur de hauteur, de durée, et d’intensité pour chaque phonème. Cela permet un contrôle direct et précis de la prosodie au niveau phonémique.

Des efforts ont été déployés pour s’assurer que NaijaTTS était capable de produire un discours naturel et compréhensible dans un cadre expérimental. Cela s’est avéré être l’un des plus grands défis dans le développement de NaijaTTS, car les fichiers audio n’ont pas été enregistrés dans l’optique de la création d’un système de TTS. Les fichiers ont souvent été enregistrés dans de mauvaises conditions acoustiques, contiennent des bruits de fond importants, et les énoncés présentent de nombreuses dysfluences. En raison de ces facteurs, les premières itérations ont produit des sorties de mauvaise qualité, difficiles à comprendre et contenant de nombreux artefacts. De gains significatifs ont été faits grâce à un processus de filtrage de données visant à entraîner le modèle sur les fichiers enregistrés dans les meilleures conditions. Pour ce faire, nous avons brièvement écouté chaque session d’enregistrement et évalué leur qualité sonore sur une base informelle est purement perceptuelle. Nous avons ensuite ré-entraîné le modèle avec les enregistrements que nous avons jugés les meilleures.

En ce qui concerne les dysfluences, nous avons exploité les informations annotées dans le treebank pour exclure tous les énoncés contenant des pauses remplies, des réparations, des mots abandonnés ou des segments jugés incompréhensibles par les transcrip-teurs.

Ces contraintes ont réduit la taille de notre jeu de données à 3,7 heures de parole et 52 locuteurs. Cependant, malgré la taille réduite du corpus d'apprentissage, nous avons constaté une nette amélioration de la qualité de la parole générée. L'exclusion des dysfluences et des fichiers enregistrés dans de mauvaises conditions semblent avoir contribué de manière comparable à l'amélioration de la qualité de la parole synthétisée. Des gains significatifs en termes de qualité de la parole ont également été obtenus en affinant le vocodeur HiFi-GAN (Kong *et al.*, 2020) à l'aide des fichiers `.wav` utilisés pour l'entraînement du modèle.

Nous avons également veillé à ce que la parole de sortie ne corresponde pas aux voix des locuteurs figurant dans le corpus NaijaSynCor. Nous avons utilisé deux approches distinctes pour anonymiser notre modèle. La première a consisté à séparer les locuteurs par genre, afin de produire 2 modèles voix selon leur genre. Cette approche a permis de produire de voix masculines et féminines de haute qualité, bien que nous ayons remarqué que les caractéristiques vocales perçues changeaient parfois en fonction de l'énoncé généré. Nous soupçonnons que certaines séquences rares de phonèmes ont amené le modèle à adopter les caractéristiques des fichiers sonores dans lesquels ces séquences apparaissent de manière disproportionnée. Une approche prometteuse, inspirée par les travaux de Meyer *et al.* (2023) consistait à enregistrer brièvement un membre du projet et à extraire un embedding anonymisé en utilisant une méthode d'extraction de xvector fournie par Speechbrain (Ravanelli *et al.*, 2021), puis à le fournir directement au modèle en tant qu'entrée. Cela a produit une parole de haute qualité avec des caractéristiques vocales stables.

Dans les sections précédentes, nous avons vu que la F0 et la durée semblaient être des indicateurs fiables pour déterminer si une occurrence donnée de GO fonctionnait comme un marqueur de futur ou comme un verbe de mouvement. Cependant, il n'est pas clair si c'est la F0 seule ou une combinaison de F0 et de durée qui permet aux locuteurs du pidgin nigérian de distinguer ces deux utilisations. C'est un domaine dans lequel NaijaTTS peut être utilisé pour compléter les techniques exploratoires présentées dans la section précédente. Dans un contexte expérimental contrôlé, NaijaTTS nous permettrait de générer une série d'énoncés différant seulement par la hauteur et la durée assignées à GO. Nous pourrions donc voir si la hauteur est bien le seul paramètre qui permet aux locuteurs de distinguer les deux usages, ce qui suggérerait l'existence d'un système de ton lexical, ou si la durée est aussi perceptivement saillante, ce qui suggérerait plutôt un système d'accent lexical.

Plusieurs variantes de cette expérience peuvent être réalisées. L'une d'entre elles consisterait à utiliser des constructions syntaxiquement non ambiguës telles que *my father go go house* « mon père ira à la maison », qui sont clairement composées d'un verbe de mouvement précédé d'un marqueur de futur. Dans de tels cas, nous testerons si la hauteur et la durée contribuent de manière similaire à la perception du caractère naturel de l'énoncé. Si une hauteur plus basse mais une durée plus élevée sur le GO grammatical sont perçus comme naturelles, mais pas un hauteur plus élevée et une durée plus courte, cela indiquerait que la durée n'est pas aussi perceptivement saillante et que le pidgin nigérian pourrait être une langue à tons comme l'ont supposé des analyses antérieures. Cependant, si les deux contribuent de manière significative au caractère naturel perçu de l'énoncé, il se peut que la langue n'ait pas de ton lexical mais plutôt un système d'accent lexical. D'autres tests pourraient impliquer l'interprétation sémantique de constructions syntaxiquement ambiguës comme *my father go eat* qui peut signifier « mon père mangera » ou « mon père est parti manger » en fonction de la forme de GO utilisée. Lors de ces tests, les participants seraient interrogés sur leur interprétation de la phrase plutôt

que sur leur opinion quant à son caractère naturel. Dans ces cas, nous verrons si c'est la durée, la hauteur ou une combinaison des deux qui contribuent à l'interprétation du sens. Comme l'autre test, celui-ci permettrait de mieux comprendre si le pidgin nigérian utilise un système d'accent lexical, ce que les données du corpus suggèrent mais ne prouvent pas de manière définitive.

4 Conclusion

Dans cet article, nous avons présenté deux outils produits à partir des mêmes données, conçus pour remplir deux rôles dans une méthodologie scientifique partagée. NaijaSynCor-Prosody est un corpus novateur qui permet de faire des hypothèses sur l'intonosyntaxe en pidgin nigérian. NaijaTTS représente un système de synthèse conçu pour tester ces hypothèses dans un environnement expérimental contrôlé. Même si ces deux outils ont été spécifiquement développés pour le pidgin nigérian, cette méthodologie peut également être appliquée à d'autres langues si une quantité de données adéquate est disponible.

Références

- BIGI B., ABIOLA O. S. & CARON B. (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*, p. 164–173 : Springer.
- COURTIN M., CARON B., GERDES K. & KAHANE S. (2018). Establishing a language by annotating a corpus. In *annDH 2018 Annotation in Digital Humanities*, volume 2155, p. 7–11 : CEUR.
- ELUGBE B. O. & OMAMOR A. P. (1991). *Nigerian Pidgin : Background and Prospects*.
- FARACLAS N. (1996). *Nigerian Pidgin*. Routledge.
- GERDES K., GUILLAUME B., KAHANE S. & PERRIER G. (2018). SUD or Surface-Syntactic Universal Dependencies : An annotation scheme near-isomorphic to UD. In *Universal dependencies workshop 2018*.
- GUILLAUME B. (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*, p. 168–175.
- GUILLER K. (2020). Analyse syntaxique automatique du pidgin-créole du nigeria à l'aide d'un-transformer (bert) : Méthodes et résultats.
- KAHANE S., CARON B., STRICKLAND E. & GERDES K. (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)*, p. pp–35 : Association for Computational Linguistics.
- KONG J., KIM J. & BAE J. (2020). Hifi-gan : Generative adversarial networks for efficient and high fidelity speech synthesis. *Advances in Neural Information Processing Systems*, **33**, 17022–17033.
- LACHERET A. & VICTORRI B. (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.

- LACHERET-DUJOUR A., KAHANE S. & PIETRANDREA P. (2019). *Rhapsodie : A prosodic and syntactic treebank for spoken French*, volume 89. John Benjamins Publishing Company.
- LIU L., LACHERET-DUJOUR A. & OBIN N. (2019). Automatic Modelling and Labelling of Speech Prosody : What's New with SLAM+ ? In S. CALHOUN, P. ESCUDERO, M. TABAIN & P. WARREN, Édts., *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. HAL : [hal-02119926](https://hal.archives-ouvertes.fr/hal-02119926).
- MAFENI B. (1971). Nigerian pidgin. *The English Language in West Africa*, p. 95–112.
- MANFREDI S., CARON B., GERDES K. & COURTIN M. (2021). Naijasyncor : a syntactic treebank, a parser and a wictionary for naija. In *Summer Conference of the Society of Pidgin and Creole Linguistics*.
- MEYER S., TILLI P., DENISOV P., LUX F., KOCH J. & VU N. T. (2023). Anonymizing speech with generative adversarial networks to preserve speaker privacy. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, p. 912–919. DOI : [10.1109/SLT54892.2023.10022601](https://doi.org/10.1109/SLT54892.2023.10022601).
- OBIN N., BELIAO J., VEAUX C. & LACHERET A. (2014). SLAM : Automatic stylization and labelling of speech melody. In *Speech prosody*, p. 246.
- PIETRANDREA P., KAHANE S., LACHERET-DUJOUR A. & SABIO F. (2014). The notion of sentence and other discourse units in corpus annotation. *Spoken corpora and linguistic studies*, p. 331–364.
- RAVANELLI M., PARCOLLET T., PLANTINGA P., ROUHE A., CORNELL S., LUGOSCH L., SUBAKAN C., DAWALATABAD N., HEBA A., ZHONG J., CHOU J.-C., YEH S.-L., FU S.-W., LIAO C.-F., RASTORGUEVA E., GRONDIN F., ARIS W., NA H., GAO Y., MORI R. D. & BENGIO Y. (2021). SpeechBrain : A general-purpose speech toolkit. arXiv :2106.04624.
- REN Y., HU C., TAN X., QIN T., ZHAO S., ZHAO Z. & LIU T.-Y. (2022). FastSpeech 2 : Fast and high-quality end-to-end text to speech.
- STRICKLAND E., AUBAKIROVA D., DONCENCO D., TORRES D. & EVRARD M. (2023a). NaijaTTS : A pitch-controllable TTS model for Nigerian Pidgin. In *ISCA Speech Synthesis Workshop*, Grenoble, France. HAL : [hal-04183972](https://hal.archives-ouvertes.fr/hal-04183972).
- STRICKLAND E., EVRARD M. & LACHERET-DUJOUR A. (2023b). SLAM 3 : An updated stylization model for speech melody. In *International Congress of Phonetic Sciences*, Prague, Czech Republic. HAL : [hal-04171671](https://hal.archives-ouvertes.fr/hal-04171671).
- STRICKLAND E., LACHERET-DUJOUR A., EVRARD M., KAHANE S., QUENNHEN P., EGBO-KHARE F. & GUILLAUME B. (2024). New Methods for Exploring Intonosyntax : Introducing an Intonosyntactic Treebank for Nigerian Pidgin. In *LREC-Coling*, Turin, Italy.
- WELLS J. C. (1995). Computer-coding the ipa : a proposed extension of sampa. *Revised draft*, 4(28), 1995.