



QUAND L'IA NE NOUS COMPREND PAS : INTERFERENCES MULTILINGUES ET LIMITES DE LA RECONNAISSANCE VOCALE DU FRANÇAIS AU NIGERIA

OLUWATOYIN DEBORAH LAWRENCE

debbielaw4real@gmail.com

Résumé

L'essor de l'intelligence artificielle (IA) dans le domaine de l'enseignement des langues étrangères a profondément modifié les méthodes d'évaluation de la compétence orale, notamment grâce aux systèmes de reconnaissance automatique de la parole (Automatic Speech Recognition – ASR). Ces technologies offrent aux apprenants des retours immédiats concernant leur prononciation, leur fluidité et certains aspects de leur production linguistique. Cependant, leur efficacité dépend largement des données utilisées pour entraîner les modèles algorithmiques, lesquelles demeurent souvent centrées sur des variétés standardisées de la langue cible et représentent insuffisamment la diversité linguistique mondiale. Cette étude examine les limites des systèmes ASR dans la reconnaissance du français langue étrangère (FLE) produit par des apprenants nigériens vivant dans un environnement fortement plurilingue. Elle s'intéresse particulièrement aux interférences linguistiques résultant du contact entre l'anglais, les langues nigérianes et le français, ainsi qu'à leurs conséquences sur la fiabilité des évaluations phonétiques automatisées. À partir d'une méthodologie mixte combinant l'analyse quantitative des erreurs de transcription et une approche qualitative fondée sur l'expérience des apprenants, les résultats montrent une diminution significative de la performance des systèmes ASR face aux variations phonétiques, prosodiques et accentuelles propres aux locuteurs nigériens du français. Ces limites ne relèvent pas uniquement de contraintes techniques ; elles traduisent également l'existence de biais liés à la composition des corpus d'apprentissage et aux normes linguistiques privilégiées par les modèles d'intelligence artificielle. Cette recherche met ainsi en évidence la nécessité de concevoir des technologies de reconnaissance vocale plus inclusives, capables d'intégrer la diversité des pratiques langagières. Elle invite également à repenser l'évaluation de la compétence orale en FLE en accordant davantage d'importance à l'intelligibilité et à l'efficacité de la communication plutôt qu'à une conformité stricte au modèle du locuteur natif.

Mots-clés : intelligence artificielle, reconnaissance automatique de la parole, français langue étrangère, plurilinguisme, interférences linguistiques, Nigeria.

ABSTRACT

The increasing deployment of artificial intelligence (AI) in language education has transformed the assessment of second language oral production, particularly through Automatic Speech Recognition (ASR) systems that provide immediate feedback on pronunciation, fluency, and accuracy. Despite their pedagogical value, the performance of these systems remains highly dependent on the linguistic composition of their training datasets, which are often dominated by native or standardized speech forms. This raises concerns regarding their reliability in multilingual and non-standard speech environments. This study examines the performance of ASR systems in processing French as a foreign language (FLE) produced by Nigerian learners in a highly multilingual context shaped by English and indigenous



languages such as Yoruba, Igbo, and Hausa. It specifically investigates how multilingual interference and accent variation affect transcription accuracy and automated pronunciation evaluation. A mixed-methods approach was employed, combining quantitative error analysis of ASR outputs with qualitative insights drawn from learner experiences. Three ASR systems were tested using three oral production tasks: word reading, sentence reading, and spontaneous speech elicitation based on visual prompts. The results show a systematic decline in recognition accuracy across systems as task complexity increases, with word-level recognition performing highest (75–82%), sentence-level performance decreasing moderately (65–73%), and spontaneous speech yielding the lowest accuracy (49–58%). Error analysis reveals that phonetic substitutions (38%), lexical mismatches (29%), omissions (18%), and prosodic errors (15%) are the most frequent categories. These errors are strongly associated with cross-linguistic influence from the learners' multilingual repertoires, particularly in phonological and prosodic transfer patterns. The findings suggest that ASR limitations are not solely technical but also reflect structural biases in model training data, which underrepresent multilingual and accented speech varieties. The study concludes that improving ASR fairness and pedagogical effectiveness requires diversification of training corpora and a shift in pronunciation assessment toward intelligibility-oriented frameworks rather than native-speaker conformity.

KEYWORDS

Artificial Intelligence; Automatic Speech Recognition (ASR); French as a Foreign Language (FLE); Multilingualism; Nigerian Learners; Language Interference; Pronunciation Assessment; Speech Recognition Accuracy; Algorithmic Bias.

1. Introduction

L'intelligence artificielle occupe désormais une place essentielle dans les environnements éducatifs contemporains, notamment dans le domaine de l'apprentissage et de l'enseignement des langues étrangères. Le développement des systèmes de reconnaissance automatique de la parole (ASR) a permis l'apparition de nouveaux outils pédagogiques capables d'analyser les productions orales des apprenants et de fournir un retour instantané sur la prononciation, la fluidité et la précision linguistique (Warschauer et Healey).

Ces innovations constituent une avancée importante, particulièrement dans les contextes où l'accès à des enseignants spécialisés ou à des interactions régulières avec des locuteurs compétents demeure limité. Les applications basées sur l'intelligence artificielle permettent aux apprenants de pratiquer leur expression orale de manière autonome et de recevoir des évaluations personnalisées. Toutefois, la fiabilité de ces technologies dépend principalement de la qualité et de la diversité des données utilisées pour entraîner les modèles de reconnaissance vocale.

Dans de nombreux cas, les corpus vocaux employés dans le développement des systèmes ASR sont majoritairement constitués d'enregistrements provenant de locuteurs natifs ou de variétés



considérées comme standards de la langue étudiée. Cette situation crée un déséquilibre susceptible d'affecter la reconnaissance des productions issues de contextes plurilingues ou caractérisées par des variations phonologiques particulières (Blodgett et al.).

Le Nigeria constitue un terrain d'étude particulièrement pertinent pour analyser cette problématique. Pays marqué par une grande diversité linguistique, il compte plusieurs centaines de langues en coexistence avec l'anglais, qui joue le rôle de langue officielle et de scolarisation. Les apprenants nigériens du français développent donc leurs compétences orales dans un environnement où plusieurs systèmes phonologiques interagissent simultanément. Les influences du yoruba, de l'igbo, du hausa et de l'anglais peuvent produire des caractéristiques phonétiques spécifiques qui ne correspondent pas toujours aux normes prises en compte par les systèmes ASR.

Lorsque ces systèmes interprètent incorrectement les productions des apprenants, la question ne relève pas uniquement d'un problème technologique. Elle soulève également des enjeux liés à la justice linguistique, à la représentation de la diversité des locuteurs et aux pratiques pédagogiques en français langue étrangère. Une évaluation automatique incapable de reconnaître certaines variations linguistiques peut conduire à une perception erronée des compétences réelles des apprenants.

Ainsi, cette recherche vise à analyser les limites des systèmes de reconnaissance vocale dans le traitement du français parlé par des apprenants nigériens et à examiner les implications pédagogiques et sociolinguistiques de ces limites. L'objectif est également de proposer une réflexion sur le développement d'outils d'intelligence artificielle plus inclusifs et mieux adaptés à la pluralité des pratiques linguistiques contemporaines.

2. Cadre théorique

2.1 Intelligence artificielle et apprentissage des langues

L'intégration de l'intelligence artificielle dans l'enseignement des langues étrangères a profondément modifié les pratiques pédagogiques contemporaines. Grâce aux progrès réalisés dans les domaines de l'apprentissage automatique, du traitement automatique du langage naturel et de la reconnaissance automatique de la parole (Automatic Speech Recognition – ASR), les



apprenants disposent aujourd'hui d'outils capables d'offrir un entraînement personnalisé, une correction immédiate et une évaluation automatique de leur production orale.

Les systèmes ASR occupent une place croissante dans l'apprentissage assisté par ordinateur des langues étrangères. Ils permettent notamment d'évaluer la prononciation, la fluidité, l'intonation et certains aspects de l'intelligibilité de la parole. Cependant, malgré les progrès considérables réalisés grâce aux réseaux neuronaux profonds et aux modèles d'apprentissage à grande échelle, la fiabilité de ces systèmes demeure variable selon les profils linguistiques des utilisateurs.

Les travaux récents de Farrús (2023) montrent que les technologies ASR représentent un potentiel important pour l'apprentissage des langues secondes, mais que leur efficacité dépend fortement de leur capacité à prendre en compte la diversité des productions orales des apprenants. De même, les recherches de Sun (2023) démontrent que l'utilisation des systèmes ASR peut améliorer la prononciation et les compétences orales des apprenants, tout en soulignant que le feedback automatisé doit être interprété avec prudence et accompagné d'une approche pédagogique adaptée.

Par ailleurs, les recherches actuelles sur l'évaluation automatique de la prononciation indiquent que les systèmes fondés sur l'intelligence artificielle ont connu des progrès significatifs dans l'analyse des caractéristiques segmentales et suprasegmentales de la parole. Néanmoins, plusieurs défis persistent, notamment l'évaluation équitable des locuteurs non natifs, la prise en compte des différentes variétés linguistiques et la distinction entre une variation acceptable et une véritable erreur de prononciation.

2.2 Plurilinguisme, diversité accentuelle et limites des systèmes ASR

Le plurilinguisme représente un défi majeur pour les technologies de reconnaissance vocale. Dans les sociétés multilingues comme le Nigeria, les apprenants de français développent leur compétence orale sous l'influence simultanée de plusieurs systèmes linguistiques, notamment l'anglais et les langues locales telles que le yoruba, l'igbo et le hausa. Ces interactions peuvent se manifester par des transferts phonologiques, des variations prosodiques et des particularités accentuelles qui constituent des manifestations naturelles du contact des langues.



Les recherches récentes sur les performances des systèmes ASR montrent que ces technologies ne reconnaissent pas tous les groupes de locuteurs avec le même niveau de précision. Feng et al. (2024) démontrent que les biais de reconnaissance vocale peuvent être liés à l'accent, à l'âge, au genre et aux caractéristiques linguistiques des utilisateurs. Ces travaux montrent également que les différences de prononciation n'expliquent qu'une partie des erreurs observées, ce qui indique l'existence de limites liées aux architectures des modèles et aux données utilisées pour leur entraînement.

De même, les études portant sur la reconnaissance de la parole de locuteurs non natifs soulignent que les performances des systèmes ASR diminuent lorsque les productions s'éloignent des variétés linguistiques majoritairement représentées dans les corpus de référence. Cette situation est particulièrement problématique dans les contextes éducatifs, où une mauvaise interprétation de la parole peut conduire à une évaluation inexacte des compétences réelles des apprenants.

2.3 Normes linguistiques, intelligibilité et justice algorithmique

Les approches traditionnelles de l'enseignement de la prononciation ont longtemps valorisé une conformité à un modèle de locuteur natif. Toutefois, les perspectives contemporaines en didactique des langues privilégient davantage l'intelligibilité, c'est-à-dire la capacité d'un apprenant à communiquer de manière claire et efficace, indépendamment de la présence d'un accent lié à son parcours linguistique.

Dans le contexte des systèmes d'intelligence artificielle, cette réflexion rejoint la question de la justice algorithmique. Les modèles ASR sont construits à partir de bases de données vocales qui ne représentent pas toujours l'ensemble des communautés linguistiques mondiales. Ainsi, les systèmes peuvent reproduire ou renforcer des inégalités existantes en favorisant les variétés linguistiques les plus présentes dans leurs corpus d'apprentissage.

Les recherches sur les biais algorithmiques dans les technologies vocales ont mis en évidence la nécessité de développer des systèmes plus inclusifs, capables de reconnaître une plus grande diversité d'accents, de langues et de pratiques discursives. Une telle évolution est essentielle pour garantir une utilisation équitable de l'intelligence artificielle dans l'enseignement du français langue étrangère et dans les environnements plurilingues.



3. Problématique et objectifs de la recherche

Malgré les progrès significatifs réalisés dans le domaine de la reconnaissance automatique de la parole, plusieurs études démontrent que les performances des systèmes ASR restent variables selon les caractéristiques linguistiques des utilisateurs. Les locuteurs provenant de contextes plurilingues ou utilisant des variétés non standardisées de la langue peuvent rencontrer davantage d'erreurs de transcription et recevoir des évaluations moins fiables.

Dans le contexte nigérian, cette problématique est particulièrement pertinente en raison de la coexistence de plusieurs langues qui influencent la prononciation du français langue étrangère. Les outils d'intelligence artificielle utilisés pour l'apprentissage du FLE risquent ainsi de confondre des particularités liées au plurilinguisme avec des erreurs de compétence linguistique.

Cette étude cherche donc à répondre aux questions suivantes :

1. Quels types d'erreurs les systèmes de reconnaissance automatique de la parole produisent-ils face au français parlé par des apprenants nigériens ?
2. Dans quelle mesure ces erreurs sont-elles associées aux interférences résultant du contact entre l'anglais, les langues nigérianes et le français ?
3. Quels effets pédagogiques ces limites technologiques peuvent-elles avoir sur l'apprentissage et la perception des compétences orales des apprenants ?
4. Comment développer des systèmes de reconnaissance vocale plus inclusifs et adaptés à la diversité linguistique mondiale ?

Les objectifs principaux de cette recherche sont d'identifier les limites actuelles des systèmes ASR dans un contexte plurilingue, d'analyser l'origine des erreurs observées et de proposer des pistes d'amélioration pour une utilisation plus équitable de l'intelligence artificielle dans l'enseignement du FLE.



4. Méthodologie

4.1 Approche de recherche

Cette étude adopte une méthodologie mixte combinant une approche quantitative et une approche qualitative. L'approche quantitative permet d'évaluer le niveau de performance des systèmes de reconnaissance vocale à travers l'analyse des erreurs de transcription, tandis que l'approche qualitative vise à comprendre les expériences, les perceptions et les difficultés rencontrées par les apprenants lors de l'utilisation de ces technologies.

4.2 Participants

L'échantillon de l'étude est constitué de trente étudiants universitaires nigériens apprenant le français comme langue étrangère. Les participants sont âgés de 18 à 26 ans et possèdent un niveau intermédiaire en français.

Leur profil linguistique reflète la diversité caractéristique du contexte nigérien :

- langues premières : yoruba, igbo et hausa ;
- langue de scolarisation et langue seconde : anglais ;
- langue étrangère étudiée : français.

Cette diversité permet d'observer l'influence des différents parcours linguistiques sur les interactions entre les apprenants et les systèmes de reconnaissance automatique de la parole.

4.3 Collecte des données

La collecte des données repose sur trois types d'activités orales destinées à solliciter différents niveaux de production linguistique :

- la lecture de phrases en français afin d'évaluer la reconnaissance d'une production guidée ;
- la répétition de mots isolés pour examiner la précision phonétique au niveau lexical ;



- la production orale spontanée à partir d'images afin d'analyser les performances des systèmes ASR dans une situation de communication plus naturelle.

Les enregistrements vocaux obtenus ont ensuite été analysés par trois systèmes de reconnaissance automatique de la parole, désignés anonymement comme systèmes A, B et C afin de préserver la neutralité de l'étude.

4.4 Analyse des données

L'analyse quantitative s'est concentrée sur plusieurs indicateurs de performance, notamment le taux global d'erreurs de transcription, les substitutions lexicales, les omissions et les erreurs liées aux caractéristiques prosodiques.

Parallèlement, une analyse qualitative fondée sur des entretiens semi-directifs avec les participants a permis d'examiner leurs expériences face aux évaluations fournies par les systèmes ASR, leur perception de la fiabilité des corrections automatiques et l'impact de ces outils sur leur confiance dans leur capacité à communiquer en français.

5. Résultats

5.1 Performance générale des systèmes de reconnaissance automatique de la parole

L'analyse des données recueillies met en évidence une diminution progressive des performances des systèmes ASR en fonction du degré de complexité des tâches orales proposées aux participants. Les systèmes obtiennent les meilleurs résultats lors de la reconnaissance de mots isolés, tandis que leur précision diminue considérablement lorsque les apprenants produisent un discours spontané.

Tableau 1 : Taux de précision des systèmes ASR selon le type de tâche orale

Type de tâche	Système A	Système B	Système C
Mots isolés	82 %	78 %	75 %
Lecture de phrases	73 %	69 %	65 %

**Type de tâche****Système A Système B Système C**

Production orale spontanée 58 % 53 % 49 %

Ces résultats indiquent que les systèmes de reconnaissance vocale rencontrent davantage de difficultés lorsqu'ils doivent traiter une parole naturelle comportant des variations de rythme, d'intonation et de prononciation. La baisse de performance observée dans les productions spontanées suggère que les modèles utilisés sont moins adaptés aux caractéristiques linguistiques propres aux locuteurs plurilingues.

5.2 Typologie des erreurs observées

L'étude des transcriptions générées par les systèmes ASR révèle plusieurs catégories d'erreurs récurrentes. Les substitutions phonétiques représentent la catégorie la plus fréquente, suivies des erreurs lexicales, des omissions et des difficultés liées aux caractéristiques prosodiques.

Tableau 2 : Répartition des principaux types d'erreurs

Type d'erreur	Pourcentage
Substitutions phonétiques	38 %
Erreurs lexicales	29 %
Omissions	18 %
Erreurs prosodiques	15 %

Les substitutions phonétiques sont souvent associées aux différences entre les systèmes phonologiques du français et ceux des langues déjà maîtrisées par les apprenants. Certaines particularités de prononciation influencées par le yoruba, l'igbo, le hausa ou l'anglais peuvent être interprétées incorrectement par les systèmes ASR, entraînant ainsi des transcriptions inexactes.

5.3 Influence des profils linguistiques des apprenants

Les résultats montrent également des variations modérées dans les performances des systèmes ASR en fonction des profils linguistiques des participants. Bien que ces différences ne



permettent pas d'établir une hiérarchie entre les langues premières, elles démontrent que les caractéristiques phonologiques propres à chaque parcours linguistique influencent la manière dont les systèmes d'intelligence artificielle interprètent les productions orales.

Ces observations confirment que les technologies de reconnaissance vocale ne sont pas totalement neutres et que leur efficacité dépend largement de la diversité et de la représentativité des données utilisées lors de l'entraînement des modèles.

6. Discussion

Les résultats obtenus confirment que les systèmes de reconnaissance automatique de la parole présentent des limites importantes lorsqu'ils sont appliqués à des contextes plurilingues tels que celui du Nigeria. La diminution des performances observée dans les tâches de production spontanée démontre que les modèles actuels éprouvent des difficultés à gérer la variabilité naturelle de la parole humaine.

Ces difficultés ne doivent pas être interprétées uniquement comme des défaillances techniques. Elles reflètent également des biais présents dans les bases de données utilisées pour développer les modèles d'intelligence artificielle. Lorsque les corpus d'entraînement sont principalement constitués de productions correspondant aux normes linguistiques dominantes, les systèmes risquent de moins bien reconnaître les accents et les variétés provenant de communautés linguistiques moins représentées (Blodgett et al.; Koenecke et al.).

Sur le plan pédagogique, cette situation soulève des enjeux importants pour l'enseignement du français langue étrangère. Un retour automatique incorrect peut amener les apprenants à considérer certaines caractéristiques de leur parole comme des erreurs, alors qu'elles peuvent simplement relever de leur identité linguistique et de leur parcours plurilingue. Il devient donc nécessaire d'adopter une approche de l'évaluation de la prononciation qui valorise davantage l'intelligibilité et l'efficacité de la communication plutôt qu'une imitation absolue du modèle du locuteur natif (Levis).

De plus, le développement futur des technologies éducatives fondées sur l'intelligence artificielle doit intégrer une plus grande diversité de locuteurs, d'accents et de situations communicatives.



Une conception plus inclusive des systèmes ASR contribuerait à réduire les biais linguistiques et à offrir une expérience d'apprentissage plus juste et adaptée aux réalités sociolinguistiques mondiales.

7. Conclusion

Cette recherche a analysé les limites des systèmes de reconnaissance automatique de la parole dans l'évaluation du français langue étrangère produit par des apprenants nigériens issus d'un environnement plurilingue. Les résultats montrent que les performances des systèmes ASR diminuent considérablement lorsque les productions orales s'éloignent des modèles linguistiques standardisés ayant principalement servi à leur entraînement.

Les erreurs observées, notamment les substitutions phonétiques, les erreurs lexicales et les omissions, sont en partie liées aux interactions entre le français, l'anglais et les langues premières des apprenants. Ces difficultés révèlent que les limites des technologies ASR sont autant d'ordre technique que sociolinguistique, puisqu'elles mettent en évidence une représentation insuffisante de la diversité linguistique au sein des modèles d'intelligence artificielle.

Cette étude invite ainsi les chercheurs, les concepteurs de technologies éducatives et les spécialistes du français langue étrangère à repenser les critères d'évaluation de la compétence orale. L'objectif ne doit pas uniquement être la conformité à une norme phonétique unique, mais la capacité des apprenants à communiquer de manière claire et efficace.

Enfin, l'avenir de l'intelligence artificielle appliquée à l'enseignement des langues dépendra de sa capacité à reconnaître et à intégrer la pluralité des voix, des accents et des expériences linguistiques qui caractérisent les sociétés contemporaines.

WORKS CITED

Baker-Bell, April. *Linguistic Justice: Black Language, Literacy, Identity, and Pedagogy*. Routledge, 2020.

Blodgett, Su Lin, Solon Barocas, Hal Daumé III, and Hanna Wallach. "Language (Technology) Is Power: A Critical Survey of 'Bias' in NLP." *Proceedings of the 58th Annual Meeting of the*



Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 5454–5476.

Bourdieu, Pierre. *Ce que parler veut dire : L'économie des échanges linguistiques*. Fayard, 1982.

Bourdieu, Pierre. *Language and Symbolic Power*. Harvard University Press, 1991.

Calvet, Louis-Jean. *La sociolinguistique*. Presses Universitaires de France, 1993.

Derwing, Tracey M., and Murray J. Munro. “Second Language Accent and Pronunciation Teaching: A Research-Based Approach.” *TESOL Quarterly*, vol. 39, no. 3, 2005, pp. 379–397.

Derwing, Tracey M., and Murray J. Munro. *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. John Benjamins Publishing Company, 2015.

El Kheir, Yassine, Ahmed Ali, and Shammur Absar Chowdhury. “Automatic Pronunciation Assessment: A Review.” *Findings of the Association for Computational Linguistics: EMNLP 2023*, Association for Computational Linguistics, 2023.

Farrús, Mireia. “Automatic Speech Recognition in L2 Learning: A Review Based on PRISMA Methodology.” *Languages*, vol. 8, no. 4, 2023, article 242.

Flege, James E. “Second Language Speech Learning: Theory, Findings, and Problems.” *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange, York Press, 1995, pp. 233–277.

Feng, Siyuan, et al. “Towards Inclusive Automatic Speech Recognition.” *Computer Speech & Language*, vol. 84, 2024, article 101567.

Koenecke, Allison, et al. “Racial Disparities in Automated Speech Recognition.” *Proceedings of the National Academy of Sciences*, vol. 117, no. 14, 2020, pp. 7684–7689.

Levis, John M. “Changing Contexts and Shifting Paradigms in Pronunciation Teaching.” *TESOL Quarterly*, vol. 39, no. 3, 2005, pp. 369–377.



Levis, John M. *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge University Press, 2018.

Munro, Murray J., and Tracey M. Derwing. "Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners." *Language Learning*, vol. 45, no. 1, 1995, pp. 73–97.

Sun, Weina. "The Impact of Automatic Speech Recognition Technology on Second Language Pronunciation and Speaking Skills of EFL Learners: A Mixed Methods Investigation." *Frontiers in Psychology*, vol. 14, 2023.

UNESCO. *AI and Education: Guidance for Policy-Makers*. UNESCO Publishing, 2021.

Warschauer, Mark. "Computer-Assisted Language Learning: An Introduction." *Multimedia Language Teaching*. Edited by Sandra Fotos, Logos International, 1996, pp. 3–20.

Warschauer, Mark, and Deborah Healey. "Computers and Language Learning: An Overview." *Language Teaching*, vol. 31, no. 2, 1998, pp. 57–71.