

## Présentation

**Raquel AMARO**

<https://orcid.org/0000-0002-4923-7186> ; [raquelamaro@fcsh.unl.pt](mailto:raquelamaro@fcsh.unl.pt)  
Université Nova de Lisboa (PORTUGAL)

**Jorge BAPTISTA**

<https://orcid.org/0000-0003-4603-4364> ; [jbaptis@ualg.pt](mailto:jbaptis@ualg.pt)  
Université de l'Algarve (PORTUGAL)

**Thomas FRANÇOIS**

<https://orcid.org/0000-0001-9415-7639> ; [thomas.francois@uclouvain.be](mailto:thomas.francois@uclouvain.be)  
Université Catholique de Louvain (BELGIQUE)

**Marcos GARCÍA**

<https://orcid.org/0000-0002-6557-0210> ; [marcos.garcia.gonzalez@usc.gal](mailto:marcos.garcia.gonzalez@usc.gal)  
Université de Santiago de Compostela (ESPAGNE)

**Rodrigo WILKENS**

<https://orcid.org/0000-0003-4366-1215> ; [r.wilkens@exeter.ac.uk](mailto:r.wilkens@exeter.ac.uk)  
Université d'Exeter (ROYAUME-UNI)



© des auteurs

Citation suggérée : AMARO, R., BAPTISTA, J., FRANÇOIS, T., GARCÍA, M., WILKENS, R. (2025), Présentation, *Langue(s) & Parole*, **10**, 7-10, <https://doi.org/10.5565/rev/languesparole.156>.

Cette section thématique de ce numéro de *Langue(s) & Parole* réunit trois articles développés dans les thématiques structurantes du projet européen *iRead4Skills (Intelligent Reading Improvement System for Fundamental and Transversal Skills Development)*, un projet de recherche et d'innovation financé par la Commission européenne dans le cadre du programme *Horizon 2020*. Coordonné par l'Université Nova de Lisboa (Portugal), ce projet associe plusieurs institutions partenaires<sup>1</sup>. Son objectif est de promouvoir le développement des compétences en lecture et en littératie fonctionnelle chez les adultes, à travers la mise au point d'un système intelligent capable d'évaluer la complexité textuelle et de

---

<sup>1</sup> En Belgique, l'Université Catholique de Louvain ; en Espagne, l'Universidade de Santiago de Compostela et l'Universitat Autònoma de Barcelona ; au Luxembourg, LISER - Luxembourg Institute of Socio-Economic Research ; et au Portugal, l'INESC-ID Lisboa - Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa, Mindshaker Serviços Informáticos, Lda, et le Ministério da Educação e Ciência.

recommander des lectures adaptées aux niveaux de compétence des utilisateurs. Au-delà de sa dimension technologique, *iRead4Skills* s'inscrit dans une perspective humaniste et inclusive, visant à favoriser la participation sociale et professionnelle des adultes peu alphabétisés grâce à l'accès à des ressources linguistiques et éducatives adaptées.

Les trois contributions présentées dans cette section, placée sous la coordination scientifique de Raquel Amaro, Jorge Baptista, Thomas François, Marcos Garcia et Rodrigo Wilkens, témoignent de la complémentarité entre recherche linguistique, ingénierie du langage et innovation pédagogique. Chacune aborde, selon un angle spécifique, la question de la complexité textuelle et de l'apprentissage linguistique dans un contexte multilingue européen.

Le premier article, « *L'effet facilitateur de la transparence interlinguistique sur l'acquisition lexicale du français et de l'espagnol chez les apprenants sinophones* », de Keran Mu, étudie les effets croisés de la transparence lexicale entre langues romanes et de la connaissance de l'anglais dans l'apprentissage du français et de l'espagnol par des étudiants sinophones. En s'appuyant sur la comparaison de plusieurs lexiques – élaborés à partir de manuels destinés aux étudiants chinois et d'un corpus authentique issu du projet *iRead4Skills* – cette recherche met en évidence le rôle des correspondances morphologiques et orthographiques interlinguistiques dans le transfert lexical et l'acquisition du vocabulaire.

Le deuxième article, « *A Multi-Stage Heuristic Filtering Pipeline for Refining a Spanish Legal Corpus for Natural Language Processing* », par Nikolai Tiurin et Xavier Blanco, propose une méthodologie computationnelle innovante pour le traitement automatique de grands corpus juridiques. Les auteurs conçoivent une chaîne heuristique de filtrage en plusieurs étapes destinée à nettoyer et normaliser les textes du *Boletín Oficial del Estado* (BOE), afin d'éliminer les erreurs issues de la numérisation et les segments non linguistiques. Ce travail, qui introduit la notion de *Combined Borderline Score* (CBS), offre une contribution méthodologique précieuse pour la constitution de corpus de qualité en espagnol et pour le développement d'applications de simplification automatique dans le domaine juridique.

Le troisième article, « *Framework of Textual Complexity for Low-Literacy Adults: Levels and Descriptors within the iRead4Skills Project* », de Raquel Amaro

*et al.*, présente le cadre linguistique et conceptuel de la complexité textuelle élaboré dans le projet *iRead4Skills*. Ce cadre propose trois niveaux de complexité – *très facile*, *facile* et *clair* – correspondant approximativement aux niveaux A1 à B1 du Cadre Européen Commun de Référence pour les Langues (CECRL), et spécifiquement adaptés aux besoins des adultes à faible niveau de littératie. Fondé sur l'intégration de mesures lexicales, syntaxiques et discursives, et validé par des experts de la formation des adultes, ce travail contribue à rapprocher la recherche théorique en linguistique des pratiques pédagogiques et des outils numériques d'évaluation et d'adaptation textuelle.

Ces trois études, complémentaires dans leurs approches et leurs objets, illustrent la richesse des questionnements portés par *iRead4Skills* autour des relations entre complexité linguistique, apprentissage et accessibilité textuelle. Ensemble, elles montrent que la réflexion sur la complexité textuelle dépasse le seul cadre de la linguistique descriptive : elle engage une approche interdisciplinaire qui associe sciences du langage, sciences cognitives, ingénierie du traitement automatique du langage et sciences de l'éducation. Dans ce sens, les travaux présentés dans cette section contribuent à l'élaboration d'une approche scientifique de la complexité textuelle à la fois empirique, explicative et socialement engagée.

**Raquel AMARO** est Professeure Auxiliaire à l'Université NOVA de Lisbonne (NOVA FCSH) et chercheuse au Centre de Linguistique de l'Université NOVA de Lisbonne (CLUNL) depuis 2015. Elle a fait partie du Conseil d'Administration du CLUNL de 2017 à 2025, période durant laquelle l'unité de recherche a atteint la classification « Excellent », et elle est devenue, en 2025, la coordinatrice du groupe Lexicologie, Lexicographie et Terminologie du CLUNL. Elle est actuellement coordinatrice du projet Horizon *iRead4Skills* (2023-2026) et chercheuse au projet Horizon MultiPod (2024-2027). Elle a participé à plus de trente projets dans les domaines de la linguistique de corpus, de la sémantique lexicale computationnelle, de la lexicographie et des technologies de la langue.

**Jorge BAPTISTA** est Professeur Associado avec *Agregação* (Habilitation) à l'Université de l'Algarve (FCHS) et chercheur intégré à l'INESC-ID Lisboa, dans l'équipe du Human Language Technology Lab (HLT). Il coordonne la participation portugaise à plusieurs projets nationaux et européens, dont *iRead4Skills*, et il est le responsable linguistique du système STRING pour le Traitement Automatique du Portugais. Il dirige aussi l'initiative *Lexicalized Meaning Representation* (LMR). Ses travaux portent sur le lexique, la syntaxe et la sémantique du portugais, les expressions polylexicales, les lexiques computationnels et le traitement automatique du langage naturel.

**Thomas FRANÇOIS** est Professeur Associé en linguistique appliquée et en traitement automatique des langues à l'Université Catholique de Louvain (UCLouvain), en Belgique. Ses travaux portent sur les approches computationnelles de l'évaluation automatique de textes, de la mesure de la lisibilité, de la simplification textuelle et de l'apprentissage des langues assisté par ordinateur. Il est l'auteur de plus d'une centaine de publications scientifiques. Il est également fondateur du projet CEFRLex, il a récemment dirigé l'équipe scientifique qui a développé FIDELIA en collaboration avec France Éducation International.

**Marcos GARCÍA** est chercheur associé en linguistique computationnelle au Centre de Recherche en Technologies Intelligentes (CITIUS) et professeur de linguistique à l'Université de Saint-Jacques-de-Compostelle. Il coordonne le groupe de recherche en linguistique computationnelle et est actuellement le chercheur principal de plusieurs projets portant sur la lisibilité, la modélisation du langage et la traduction automatique. Il a publié plus de cent articles de recherche évalués par les pairs dans des revues et des conférences internationales, et il a également organisé plusieurs conférences et ateliers internationaux dans ce domaine.

**Rodrigo WILKENS** est Lecturer (Maître de Conférences) en informatique à l'Université d'Exeter. Ses recherches portent sur le traitement automatique du langage naturel (TAL), avec un intérêt particulier pour la lisibilité, la simplification textuelle et l'intelligence artificielle explicable. Il travaille à l'intersection de la linguistique et de l'intelligence artificielle, en explorant les similitudes entre les décisions prises par les modèles computationnels et celles prises par les humains. Il possède une vaste expérience en recherche multilingue portant sur le portugais, le français et l'anglais, et collabore avec des équipes internationales sur les questions de lisibilité et d'accessibilité de textes en TAL.