

# Modélisation bayésienne des interactions entre lecture et écriture : simulations avec un même modèle computationnel de tâches de lecture de mots et de production orthographique

Sujet de stage de M2 en Sciences Cognitives

Année universitaire 2023-2024

## 1 Cadre théorique

L'apprentissage de la lecture et de l'écriture occupe une place centrale dans le développement de l'enfant, marquant notamment son entrée dans le milieu scolaire. Les supports écrits, qu'ils soient manuscrits ou imprimés, deviennent ensuite indispensables à sa scolarité, de la lecture des manuels pédagogiques à la rédaction de textes argumentés. Bien que la maîtrise de la littérature constitue la base de la plupart des apprentissages ultérieurs et soit essentielle à la vie scolaire et professionnelle de chaque individu, son acquisition demeure une étape complexe, générant de nombreuses inégalités tout au long de la scolarité. Les performances moyennes des enfants en lecture et en compréhension de texte en Europe déclinent depuis plusieurs années (DEPP, 2021), témoignant ainsi de la difficulté inhérente à l'acquisition de cette compétence fondamentale.

Dans ce contexte, la question des méthodes d'enseignement de la lecture et de l'écriture revêt une importance capitale dans la recherche contemporaine. Bien que certaines méthodes, telles que les approches phoniques, semblent faire consensus (Castles et al., 2018; Ramus, 2018), de nombreuses questions connexes restent en suspens. Par exemple, il convient d'explorer les liens entre la lecture et l'écriture, et de se demander comment enseigner au mieux aux enfants le principe alphabétique et les relations qui existent entre les mots à l'oral et à l'écrit.

La modélisation computationnelle permet d'étudier cette question sous un angle nouveau. Cette méthodologie permet de dépasser les limites des modèles théoriques traditionnels (les modèles « boîtes-flèches »), par la spécification en détail des mécanismes et processus théorisés. De plus, ils permettent de confronter différentes hypothèses dans un milieu contrôlé, afin d'évaluer les conséquences de chacune à la lumière de données comportementales.

Néanmoins, les modèles computationnels produits sont restreints à une ou quelques-unes des tâches comportementales permettant l'étude de la lecture (Pritchard et al., 2018; Seidenberg & McClelland, 1989; Ziegler et al., 2014) ou de la production orthographique (Bullinaria, 1997; Houghton & Zorzi, 2003; Olson & Caramazza, 1994), et ne permettent donc pas d'avoir un modèle théorique unifié, permettant de rendre compte d'effets dans ces tâches variées. Le développement du modèle BRAID-Spell répond à cette problématique, par l'élaboration d'un modèle probabiliste simple-voie capable d'effectuer un panel de tâches à la fois en lecture et en production orthographique (c'est-à-dire en dictée de mots). À l'aide du modèle nouvellement implémenté, un certain nombre de simulations devraient permettre de mettre en évidence la capacité du modèle à reproduire à la fois des effets liés à l'activité de lecture et à celle de la production orthographique.

Ce modèle ouvre donc la porte à l'étude conjointe de ces deux activités, notamment dans le cadre de l'apprentissage typique et pathologique de la littératie.

## 2 Objectifs

Les objectifs de ce stage de M2 sont expérimentaux en simulation. Il s'agit d'étudier par des simulations la capacité d'un modèle computationnel unique à rendre compte des activités de lecture et de production orthographique, notamment dans le cadre de l'apprentissage orthographique.

## 3 Travail proposé

La prise en main du modèle BRAID-Spell permettra une première approche de la méthodologie computationnelle en sciences cognitives, afin de savoir comment sont réalisées les simulations au cours du stage. Dans le même temps, un travail bibliographique des données comportementales existantes permettra d'identifier les effets d'intérêt à simuler. Par exemple, nous nous intéressons à la simulation de l'effet de consistance orthographique à la fois dans le sens de la lecture et de la production orthographique, étant donné que la langue française a la particularité d'être davantage consistante dans le sens de la lecture que dans celui de la production orthographique. Enfin, il faudra implémenter et réaliser les simulations et l'analyse des résultats produits dans le contexte de la littérature, notamment par la proposition si possible de prédictions expérimentales testables ultérieurement.

## 4 Contexte pratique

Le travail sera réalisé au sein d'une équipe pluridisciplinaire alliant modélisateurs et spécialistes de l'apprentissage de la lecture, au sein du LPNC (Laboratoire de Psychologie et NeuroCognition) à Grenoble. Le stage sera encadré conjointement par Julien Diard (modélisation bayésienne) et Camille Charrier (doctorante).

## 5 Compétences demandées

Le candidat devra idéalement avoir une formation en programmation (langages Python et R) et un intérêt pour la modélisation mathématique en Sciences Cognitives. Néanmoins, le sujet pourra être adapté à d'autres profils (n'hésitez pas à nous contacter). Des connaissances préalables en probabilités sont un plus, mais ne sont pas indispensables.

## 6 Contact

- Julien DIARD, Chargé de recherche, LPNC, CNRS : [julien.diard@univ-grenoble-alpes.fr](mailto:julien.diard@univ-grenoble-alpes.fr)
- Camille CHARRIER, doctorante EDISCE, LPNC, UGA : [camille.charrier@univ-grenoble-alpes.fr](mailto:camille.charrier@univ-grenoble-alpes.fr)
- Marie-Line BOSSE, Professeure des universités, LPNC, UGA : [marie-line.bosse@univ-grenoble-alpes.fr](mailto:marie-line.bosse@univ-grenoble-alpes.fr)

## 7 Financement

Les indemnités de stage seront assurées par le projet TRANS3.

## References

- Bullinaria, J. A. (1997). Modeling reading, spelling, and past tense learning with artificial neural networks. *Brain and Language, 59*(2), 236–266.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest, 19*(1), 5–51.
- DEPP. (2021). PIRLS 2021, Évaluation internationale des élèves de CM1 en compréhension de l'écrit. *Note d'Information, 23*(21).
- Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive neuropsychology, 20*(2), 115–162.
- Olson, A., & Caramazza, A. (1994). Representation and connectionist models: The NETspell experience. *Handbook of spelling: Theory, process and intervention, 337–363*.
- Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A Computational Model of the Self-Teaching Hypothesis Based on the Dual-Route Cascaded Model of Reading. *Cognitive Science, 42*(3), 722–770. <https://doi.org/10.1111/cogs.12571>
- Ramus, F. (2018). *Lire au CP* [Séminaire des IEN 1er degré, Apprentissages fondamentaux, ESENER].
- Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*, 523–568. <https://doi.org/10.1037/0033-295X.96.4.523>
- Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling Reading Development through Phonological Decoding and Self-Teaching: Implications for Dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120397. <https://doi.org/10.1098/rstb.2012.0397>