Modélisation bayésienne de la lecture

Ali Saghiran

Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **CIA : Ingénierie de la Cognition, de l'interaction, de l'Apprentissage et de la création**

Arrêté ministériel : 25 mai 2016

Présentée par

Ali SAGHIRAN

Thèse dirigée par **Sylviane VALDOIS**, DR CNRS
et codirigée par **Julien DIARD**, CR CNRS

préparée au sein du **Laboratoire de Psychologie et NeuroCognition** et de l'**École doctorale « Ingénierie pour la Santé, la Cognition et l'Environnement »**

Modélisation bayésienne de la lecture
Bayesian modelling of reading

Thèse soutenue publiquement le **10 Juin 2021**
devant le jury composé de :

Monsieur FRANCIS COLAS
CHARGÉ DE RECHERCHE, INRIA NANCY GRAND EST, Rapporteur
Madame CHOTIGA PATTAMADILOK
CHARGÉE DE RECHERCHE, CNRS DÉLÉGATION PROVENCE ET CORSE, Rapporteure
Monsieur STÉPHANE DUFAU
INGÉNIEUR DE RECHERCHE, CNRS DÉLÉGATION PROVENCE ET CORSE, Examinateur
Monsieur LUDOVIC FERRAND
DIRECTEUR DE RECHERCHE, CNRS DÉLÉGATION RHÔNE AUVERGNE, Examinateur
Madame HÉLÈNE LŒVENBRUCK
DIRECTRICE DE RECHERCHE, CNRS DÉLÉGATION ALPES, Examinatrice — Présidente
Madame SYLVIANE VALDOIS
DIRECTRICE DE RECHERCHE, CNRS DÉLÉGATION ALPES, Directrice de thèse
Monsieur JULIEN DIARD
CHARGÉ DE RECHERCHE, CNRS DÉLÉGATION ALPES, Co-Directeur de thèse, Invité

Table des Matières

Liste des Illustrations

Liste des Tableaux

Remerciements

"I want to thank me for believing in me, I want to thank me for doing this hard work, I want to thank me for having no days off, (...)"

— SNOOP DOGG (2018)

Au commencement de cette aventure, un grand doute me hantait quant à ma capacité de mener à bien ce travail. Ma formation antérieure était principalement technique et je n'étais qu'un novice en science cognitive et en psychologie cognitive. Plusieurs rencontres, échanges et épreuves m'ont fait apprendre tant sur moi que sur mon travail, ont changé mes *a priori* et m'ont fait apprécier davantage le domaine des sciences cognitives et la recherche scientifique de manière générale.

Tout d'abord, je tiens à remercier mes deux directeurs de thèse Julien et Sylviane pour l'encadrement, la patience et les discussions théoriques très enrichissantes, ainsi que pour nos réunions qui étaient toujours passionnantes. Vous avez été pour moi un duo complémentaire d'un point de vue scientifique, et avec des qualités humaines remarquables. Je vous remercie pour votre disponibilité permanente, particulièrement dans les moments les plus critiques de ce travail de thèse et ce malgré le contexte pandémique inhabituel et angoissant. Je tiens également à vous remercier de m'avoir proposé ce travail de thèse ainsi que de m'avoir fait confiance tout au long de ce travail. Au fil de ces trois années et demie de thèse, j'ai pu apprendre à vos cotés beaucoup de choses d'un point vue personnel, professionnel et scientifique.

Je tiens à remercier grandement les membres du jury : M. Francis Colas, Mme Chotiga Pattamadilok, M. Stéphane Dufau, M. Ludovic Ferrand et Mme Hélène Lœvenbruck pour leur disponibilité et pour avoir accepté d'évaluer mon travail de thèse. Je les remercie également pour les échanges constructifs que nous avons eu durant la soutenance.

Je tiens à remercier l'ensemble des membres du Laboratoire de Psychologie et de NeuroCognition. Je remercie tout d'abord l'ensemble des membres de l'équipe «BRAID» et «FLUENCE» : Thierry, Svetlana, Emilie, Sonia, Ahmed et Alexandra. Je remercie également toute l'équipe adminitrative : Claire, Guylaine, Sanie et Thierry. Je remercie en particulier tous les doctorants qui sont passés par le bureau D109 : Elise, Célise, Maëlle, Samuel et Olivier. Je remercie également les autres doctorants du LPNC que j'ai pu cotoyés durant ces années de thèse : Elie, Brice, Cynthia, Adeline, Léa, Lucrèce, Laura, Mamady, Merrick, Méline, Martin, Audrey, Rémi, Alexia, Benjamin, Sonja.

Un grand merci à Célise. Merci pour ta disponibilité et ton soutien permanent durant la rédaction de ce manuscrit et certaines périodes difficiles de cette thèse. Merci pour les pauses café. Merci pour ton aide.

Je tiens à remercier l'ensemble de mes amis au Maroc et en France pour leur soutien depuis plusieurs années et bien avant la thèse. Merci à Ahmed, Mohammed Ali, Imad, Mehdi, Boua, Soufian, Anass, Amine, entre autres.

Enfin, un grand merci à mes parents qui m'ont aidé, soutenu et supporté dans tout ce que j'ai entrepris durant tout mon parcours scolaire et universitaire. Un grand merci à mes deux frères Amine et Youssef pour leur soutien et leurs messages d'encouragement incessants.

Merci à toutes les personnes que j'aurais pu oublier de mentionner et qui ont contribué de près ou de loin à ce travail de thèse.

*
* *

Ce travail a été financé dans le cadre du Projet « FLUENCE » : Opération soutenue par l'Etat dans le cadre du volet e-FRAN du Programme d'investissement d'avenir, opéré par la Caisse des Dépôts.

Introduction

Le langage écrit occupe une place importante et indispensable dans la société moderne. Contrairement au langage oral qui s'acquiert naturellement durant les premières années de vie, la manipulation du langage écrit nécessite un apprentissage lent et complexe, mêlant des compétences visuelles pour reconnaître les caractères et les symboles écrits, des compétences phonologiques et linguistiques spécifiques à la langue utilisée, ainsi que des compétences motrices nécessaires pour l'écriture (Dehaene, 2009 ; Seidenberg, 2017). Dès lors que ces compétences sont acquises, le lecteur parvient à traiter le langage écrit de manière efficace et fluente : un lecteur dit « expert » reconnaît un mot en quelques centaines de millisecondes et peut lire jusqu'à 200 mots par minute (Rayner, Pollatsek, Ashby, & Clifton, 2012).

La lecture est un sujet qui a toujours suscité beaucoup d'intérêt en sciences cognitives, que ce soit à travers l'expérimentation ou la modélisation, et en particulier la modélisation computationnelle (Fu, Wang, Guo, Bermúdez-Margaretto, & Martínez, 2020 ; Phénix, Diard, & Valdois, 2016 ; Rayner & Reichle, 2010 ; Yap & Balota, 2015). Modéliser computationnellement la lecture nécessite d'établir un formalisme mathématique et de définir un système qui implémente un certain nombre de postulats théoriques à propos des mécanismes cognitifs en jeu. Ces postulats théoriques spécifient la nature des opérations, les connaissances et les mécanismes mis en jeu dans la lecture et la façon dont ils interagissent. Ils sont ensuite traduits mathématiquement dans des modèles, qui permettent de simuler les comportements de lecture. Ces postulats font l'objet de débats permanents dans la communauté, ce qui explique la multiplicité des modèles proposés et les différentes architectures retenues. La démonstration d'une reproduction fidèle lors des simulations du comportement observé chez l'humain permet d'alimenter les débats, en soutenant les postulats théoriques qui sont à la base du modèle évalué, ainsi que les choix de modélisation associés (Jacobs & Grainger, 1994 ; Norris, 2005). Dans la situation, rare mais idéale, dans laquelle on démontrerait qu'une observation comportementale ne pourrait être soutenue que par un modèle mathématique donné, et donc un jeu de postulats théoriques donné, alors le débat pourrait être « tranché ». Malheureusement, la plupart du temps, les cadres théoriques ne sont pas directement identifiables sur la base d'une simulation discriminante : le débat progresse alors par accumula-

tion d'indices, certains portant sur la parcimonie des modèles, les autres sur leur capacité à rendre compte, aussi précisément que possible, d'un large éventail de données comportementales.

La complexité de la lecture et l'imbrication de ses processus est telle que les recherches menées jusqu'ici n'ont pas eu pour ambition de modéliser la lecture dans son ensemble. Elles se sont plutôt focalisées sur des dimensions spécifiques de la lecture et traitent ces dimensions comme si elles étaient indépendantes les unes des autres (Norris, 2013 ; Rayner & Reichle, 2010 ; voir également Roelofs, 2005). Ceci a abouti au développement de trois champs de recherche voisins qui ont conduit à proposer des modèles computationnels distincts. Le premier champ s'intéresse à l'étude des traitements visuels impliqués dans la reconnaissance de lettres et de mots isolés, et a conduit à développer des modèles computationnels de reconnaissance de mots (Adelman, 2011 ; Davis, 2010 ; Gomez, Ratcliff, & Perea, 2008 ; Grainger & Jacobs, 1996 ; McClelland & Rumelhart, 1981 ; Norris, 2006 ; Whitney, 2001). Le second champ se concentre sur la dénomination orale des mots isolés et le passage du code écrit au code phonologique, conduisant au développement de modèles computationnels dits de lecture à haute voix (Ans et al., 1998 ; Coltheart et al., 2001 ; Perry et al., 2007 ; Perry, Ziegler, & Zorzi, 2010 ; Plaut, McClelland, Seidenberg, & Patterson, 1996 ; Pritchard, Coltheart, Marinus, & Castles, 2018 ; Seidenberg & McClelland, 1989 ; Ziegler, Perry, & Zorzi, 2020). Enfin, le troisième champ de recherche s'est donné comme objectif de prédire les mouvements oculaires lors de la lecture de texte (séquences de mots), on parle alors de modèles du contrôle oculomoteur (Engbert, Nuthmann, Richter, & Kliegl, 2005 ; Reichle, Rayner, & Pollatsek, 2003 ; Snell, van Leipsig, Grainger, & Meeter, 2018 ; Veldre, Yu, Andrews, & Reichle, 2020).

Par ailleurs, il existe des travaux théoriques qui s'intéressent à la lecture et au traitement visuel des mots d'un point de vue neuronal (Dehaene, Cohen, Sigman, & Vinckier, 2005 ; Mechelli et al., 2005 ; Vidyasagar, 2013). Cependant, les modèles qui y sont introduits n'ont pas été implémentés et dépassent le cadre de cette thèse. Il ne seront donc ni présentés ni discutés dans ce qui suit.

Questions de recherche

Les recherches menées dans le cadre de cette thèse se situent dans le champ spécifique des modèles computationnels de la lecture à haute voix. Ainsi, nous définissons trois questions de recherche auxquelles nous nous intéresserons dans ce travail.

Les mécanismes visuels dans les modèles de lecture

La première question met l'accent sur la dispersion des recherches actuelles et le développement des différents champs, de façon indépendante, selon les aspects de la lecture considérés. En effet, les différents champs de la lecture, que nous avons présentés précédemment, se sont développés en relative autonomie les uns par rapport aux autres, c'est-à-dire que les avancées dans l'un des champs ne bénéficient pas nécessairement aux développements ultérieurs dans les autres champs. Ainsi, par exemple, tandis que des modèles de reconnaissance de mots de plus en plus sophistiqués ont été développés au cours des dernières années, les modèles dits de « lecture à haute voix » incluent encore le plus souvent des versions du sous-modèle de reconnaissance de mots

qui seraient jugées obsolètes par les experts de ce champ (par exemple, en reposant sur un codage strict des positions des lettres, ou en faisant l'hypothèse qu'elles sont identifiées immédiatement et sans erreur). De la même façon, alors que les modèles de contrôle oculomoteur incluent tous un composant visuo-attentionnel, la plupart des modèles de reconnaissance de mots ou de lecture à haute voix n'intègrent aucune représentation des mécanismes d'attention visuelle.

La première question concerne donc l'apport de la prise en compte des propriétés des traitements visuels et visuo-attentionnels dans un modèle de lecture. Que peut-on apporter en incluant un sous-modèle sophistiqué de reconnaissance de mots dans un modèle de lecture ? Est-ce que cela n'apporte rien, ce qui légitimerait l'indépendance supposée par les modèles de lecture dominant actuellement dans la littérature, ou bien est-ce que cela permettrait de rendre compte d'effets comportementaux qui leur sont actuellement inaccessibles ?

Lecture de pseudo-mots et architecture à voie unique

La littérature oppose des modèles de lecture possédant une seule voie de traitement (Ans et al., 1998 ; Hendrix et al., 2019 ; Plaut et al., 1996 ; Seidenberg & McClelland, 1989) à des modèles double-voie (Coltheart, 2005 ; Coltheart et al., 2001 ; Perry et al., 2007 ; Perry, Ziegler, & Zorzi, 2010). Les approches de type « voie unique » supposent que la lecture experte met en jeu les mêmes processus et les mêmes connaissances quelles que soient les séquences de lettres à traiter. Ces modèles ne postulent notamment aucun mécanisme de traitement qui serait spécifique à la lecture de pseudo-mots. En revanche, l'approche « double-voie », actuellement dominante, postule que le passage des représentations orthographiques aux représentations phonologiques nécessite deux voies de traitement (une voie sous-lexicale et une voie lexicale) dont chacune fait appel à des processus cognitifs distincts.

La deuxième question est donc de savoir s'il est possible de lire la plupart des pseudo-mots sur la base des seules connaissances lexicales acquises ou s'il est nécessaire, pour rendre compte de la lecture des pseudo-mots, de postuler l'existence de connaissances explicites sur les associations entre unités orthographiques et phonologiques, stockées et mobilisées indépendamment des connaissances lexicales.

Attention visuelle et nature des traitements

Tous les modèles, qu'ils soient à voie unique ou double-voie, postulent que la reconnaissance des mots familiers repose sur un traitement parallèle de la séquence de lettres qui les composent. Ils s'opposent par contre quant à l'existence de traitements sériels pour la lecture des mots nouveaux, l'origine de ces traitements et le niveau auquel ils interviennent. Les modèles de type double-voie postulent que les traitements sériels sont propres à la voie sous-lexicale. Ces traitements interviennent après l'identification en parallèle des lettres du stimulus. Ils permettent le découpage sériel de la séquence de lettres en graphèmes (c'est-à-dire les séquences orthographiques correspondant à un phonème) qui sont successivement associés aux phonèmes correspondants. Du fait de leur architecture, les modèles à voie unique, quant à eux, ne supposent pas que le traitement sériel fasse intervenir des connaissances différentes de celles qui sont impliquées dans le traite-

ment parallèle. Le modèle en Triangle (Seidenberg & McClelland, 1989) n'inclut pas de traitement sériel, sauf dans l'extension proposée par Plaut (1999) dans laquelle ces traitements sont supposés très précoces. Dans le modèle MTM (Ans et al., 1998), la sérialité est liée à la taille de la fenêtre attentionnelle de traitement. Une large fenêtre déployée sur toute la séquence à lire permet de traiter parallèlement toutes les lettres des stimuli familiers. Ce traitement en parallèle échoue pour les mots non familiers, entraînant une réduction de la fenêtre attentionnelle et un traitement séquentiel. Le modèle MTM est le premier modèle de lecture à implémenter une fenêtre attentionnelle ; il postule qu'elle intervient précocement pour l'identification des lettres et qu'elle module le traitement (parallèle ou sériel) du stimulus d'entrée.

La troisième question traitée dans cette thèse est de déterminer dans quelle mesure un modèle de lecture doté d'un composant attentionnel peut rendre compte de certains comportements observés qui sont classiquement attribués aux traitements sériels, tels que les effets de longueur ou la lecture de pseudo-mots.

Objectifs et Contributions

L'objectif de ce travail de thèse est double. Il vise, en premier lieu, à tendre vers un modèle de lecture à haute voix plus intégratif, en se basant sur un sous-module sophistiqué de reconnaissance de mots qui inclut les mécanismes nécessaires à la simulation des fixations et des déplacements visuo-attentionnels pendant la lecture. En effet, le modèle « BRAID » (Phénix, 2018) sur lequel se basera notre travail est un modèle de reconnaissance de mots à la hauteur des derniers développements dans ce champ de recherche. La composante visuo-attentionnelle permet le rapprochement entre, d'une part, la reconnaissance et la lecture de mots et, d'autre part, les comportements visuo-attentionnels d'exploration du stimulus pendant son traitement.

Notre second objectif est de tester la capacité du modèle à rendre compte de la lecture de mots et de pseudo-mots sur la base de connaissances et d'opérations cognitives communes, et donc, avec une architecture en une seule voie. Plus précisément, nous étudierons et évaluerons l'hypothèse selon laquelle un traitement ne mettant en jeu que les connaissances lexicales apprises sur les mots, et donc une architecture à voie unique, est en mesure de rendre compte des relations sous-lexicales entre unités orthographiques et phonologiques, de simuler les effets de longueur dans différents types de tâches et d'opérer un traitement par segments sous-lexicaux pour les mots nouveaux.

Pour mener à bien ce travail de thèse, nous proposons un nouveau modèle computationnel probabiliste de la lecture nommé « BRAID-Phon ». Ce modèle est une extension du modèle computationnel de reconnaissance de mots BRAID (Phénix, 2018) par ajout, principalement, d'un sous-modèle de connaissances phonologiques. Nous utilisons le modèle BRAID-Phon pour étudier la plausibilité d'un système basé sur une architecture à voie unique, c'est-à-dire incluant uniquement des connaissances lexicales orthographiques et phonologiques, pour effectuer une simulation de la lecture. Nous montrerons la capacité de BRAID-Phon à rendre compte des effets de longueur sur les mots dans trois types de tâches (lecture, décision lexicale et *progressive*

demasking) et nous étudierons le rôle des mécanismes implémentés d'attention visuelle sur ces effets. Enfin, nous illustrerons la nécessité d'un processus de traitement par segments, contrôlé par l'attention, pour effectuer la lecture de pseudo-mots et nous évaluerons les conséquences de ce constat en termes de modélisation.

Plan de lecture

Ce document de thèse est composé de cinq chapitres – numérotés de 1 à 5 – dont les Chapitres 2, 3 et 4 présentent les travaux de recherche répondant à notre problématique. Nous donnons ci-dessous une courte description du corps de ce document, chapitre par chapitre.

CHAPITRE 1, Revue de littérature

Dans le premier chapitre, nous présenterons une revue de la littérature consacrée aux modèles double-voie et à voie unique. Nous détaillerons les différences théoriques et les implémentations spécifiques à chacun de ces cadres théoriques. Cela nous conduira à décrire successivement les différents modèles double-voie qui ont été proposés puis les différents types de modèles à voie unique. Nous analyserons ensuite les limites de ces modèles et les questions théoriques qu'ils suscitent.

CHAPITRE 2, Le modèle BRAID-Phon

Dans le second chapitre, nous rappellerons les éléments clés du modèle BRAID (« *Bayesian model of word Recognition with Attention, Interference and Dynamics* »), développé dans la thèse de Thierry Phénix (2018), puis nous présenterons l'extension BRAID-Phon que nous avons développée dans cette thèse. Cette extension présente une mise à jour des traitements sensoriels et un ajout des connaissances et des traitements phonologiques. Enfin, nous définirons mathématiquement la manière de simuler la tâche de lecture dans le modèle BRAID-Phon, et nous l'illustrerons sur quelques exemples.

CHAPITRE 3, Simulation de l'effet de longueur dans des tâches multiples

Dans le troisième chapitre, nous étudierons la capacité du modèle BRAID-Phon à simuler les données comportementales recueillies dans le cadre de la méga-étude Chronolex (Ferrand et al., 2011), sur trois tâches différentes : la lecture (NMG), la décision lexicale (LDT) et le *progressive demasking* (PDM). Nous nous intéresserons plus spécifiquement à la capacité du modèle à rendre compte de l'effet de longueur dans ces différentes tâches. Nous chercherons dans un premier temps à simuler les résultats comportementaux, en utilisant les paramètres par défaut du modèle BRAID-Phon. Puis, dans un second temps, nous étudierons l'impact des paramètres visuo-attentionnels sur les résultats simulés.

CHAPITRE 4, Lecture de pseudo-mots

Dans le quatrième chapitre, nous nous intéressons à la capacité du modèle BRAID-Phon à lire des pseudo-mots sur la seule base de ses connaissances orthographiques sur les mots. Nous verrons que la lecture de pseudo-mots nécessite un traitement par segments. Nous

présenterons les modifications qui doivent être apportées au modèle BRAID-Phon, c'est-à-dire l'ajout d'un mécanisme de traitement des segments, définis par les fixations visuo-attentionnelles. Dans une série de simulations exploratoires, nous illustrerons ensuite le fonctionnement du modèle à travers des exemples de lecture de pseudo-mots ayant des caractéristiques variées.

CHAPITRE 5, Conclusion générale

Dans le dernier chapitre, après un résumé des principaux résultats obtenus au cours de cette thèse, nous discuterons des apports de ce travail, et confronterons notre contribution aux approches existantes. Pour cela, nous traiterons la question de l'architecture des modèles (double-voie *versus* à voie unique) et nous présenterons l'apport du modèle BRAID-Phon dans le cadre de cette comparaison. Nous traiterons ensuite la question de l'attention visuelle et son rôle dans la lecture en lien avec la notion de la taille des unités sous-lexicales. Nous aborderons également le lien entre l'attention visuelle et l'effet de longueur en lecture de mots.

CHAPITRE

1 Revue de littérature

"And so to completely analyze what we do when we read would almost be the acme of a psychologist's achievements, for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilization has learned in all of its history"

— EDMUND B. HUEY (1908)

Les premiers travaux théoriques sur le traitement visuel de mots et la lecture datent des années 1960 (Morton, 1969 ; Morton & Broadbent, 1967). De nos jours encore, ce champ de recherche suscite un grand intérêt étant donné la complexité du système de lecture (Fu et al., 2020). La plupart des théories et des modèles actuels considèrent que lors de la lecture, l'information se propage dans un système qui possède une architecture en deux voies parallèles. D'autres modèles ne recourent pas à cette hypothèse et supposent que l'architecture du système de lecture possède une seule voie de traitement.

Ce chapitre propose une revue de littérature concernant les modèles de lecture à haute voix existants, de façon à clarifier les différences théoriques et d'implémentation entre les modèles double-voie et les modèles à voie unique. Ce chapitre se structure en trois sections. Nous nous focaliserons sur les modèles double-voie dans la première section, puis, dans la deuxième section, nous nous intéresserons aux modèles à voie unique. Enfin, la dernière section nous permettra de mettre en évidence certaines limites et questions théoriques qui restent en suspens et que nous allons aborder dans cette thèse.

1 Modèles Double-Voie (*Dual-Route Models*)

Comme leur nom l'indique, les modèles double-voie supposent l'existence de deux voies de traitement qui font appel à des processus cognitifs distincts. La voie lexicale est spécialisée dans le traitement des mots alors que la voie sous-lexicale permet le traitement des pseudo-mots et des mots nouveaux. L'idée essentielle est que les informations sur les formes orthographique et phonologique des mots sont encodées au sein de la voie lexicale, alors que la voie sous-lexicale représente un système de transcodage direct des unités orthographiques en leurs correspondants phonologiques.

Les premières formulations de la théorie double-voie ont été énoncées majoritairement durant les années 1970 (Forster & Chambers, 1973 ; Marshall & Newcombe, 1973). Cette théorie est devenue rapidement dominante durant les années 1980. Les études de patients dont les lésions cérébrales avaient altéré les compétences en lecture, de différentes manières, ont considérablement motivé cette théorie (Caramazza, 1986 ; Coltheart, 2017).

En lecture, plus spécifiquement, les chercheurs ont identifié différents profils de patients atteints de dyslexie acquise, c'est-à-dire des troubles spécifiques de lecture causés par des lésions cérébrales (Beauvois & Derouesné, 1979 ; Shallice & MacCarthy, 1985). Les patients dits « dyslexiques phonologiques » se caractérisent par un trouble massif en lecture de pseudo-mots[1] alors que la lecture des mots réguliers et irréguliers est largement préservée. Cette dégradation sélective suggère que le mécanisme responsable de la prononciation de nouvelles chaînes de lettres est sélectivement affecté. A l'inverse, les patients dits « dyslexiques de surface » présentent un déficit en lecture de mots irréguliers alors que les mots réguliers et les pseudo-mots peuvent être lus au niveau attendu. Ce trouble serait spécifique à une atteinte des mécanismes qui traitent les stimuli lexicaux uniquement. Cette double dissociation entre lecture de mots irréguliers et de pseudo-mots a été interprétée comme témoignant de l'existence de deux voies de traitement indépendantes et isolables. Il s'agit de l'hypothèse fondamentale des modèles double-voie (Coltheart, Curtis, Atkins, & Haller, 1993 ; Coltheart & Rastle, 1994).

Même si les modèles les plus récents abordent les questions liées à l'apprentissage de la lecture (Perry, Zorzi, & Ziegler, 2019 ; Pritchard et al., 2018), les modèles computationnels issus de cette théorie ont eu pour objectif premier de simuler le plus fidèlement possible les performances du lecteur expert. Ces modèles proposent également des hypothèses explicatives des dyslexies acquises, et permettent de les simuler suite à des « lésions » de certains composants du modèle. Par ailleurs, le cadre théorique des modèles double-voie permet d'expliquer les mécanismes impliqués dans les traitements sémantique et phonologique. En effet, les « partisans » de cette classe de modèles supposent que le traitement via la voie lexicale comporte un accès direct du lexique orthographique vers le lexique phonologique et un accès indirect via la sémantique. En revanche, les implémentations de ces modèles ne considèrent pas les connaissances sémantiques, en général. Nous rappelons que les questions de recherche liées au traitement sémantique des mots

[1] Le terme de pseudo-mot réfère à des séquences orthographiques légales et prononçables. Les séquences « linguistiquement » illégales (par exemple, « RSFAC » ou « NBTSP ») sont des non-mots et ne peuvent donc pas être considérées comme des pseudo-mots.

n'entrent pas dans le cadre de cette revue de littérature ; nous nous intéressons principalement à la reconnaissance visuelle des mots et aux processus qui permettent de passer du code orthographique au code phonologique. Dans cette section, nous décrivons trois modèles computationnels qui relèvent de l'approche double-voie : les modèles DRC (*Dual-Route Cascaded Model* ; Coltheart, 2005 ; Coltheart et al., 2001) et CDP (Connexionniste à Deux Processus ; *Connexionnist Dual Process Model* ; Perry et al., 2007 ; Perry, Ziegler, & Zorzi, 2010, 2014a, 2014b) sont des implémentations directes des modèles conceptuels précurseurs. Le troisième modèle MROM (*Multiple Read-Out Model* ; Grainger & Jacobs, 1996 ; Jacobs et al., 1998) n'adopte pas littéralement cette approche mais conserve le postulat de deux voies de traitement.

1.1 Le modèle DRC

Le premier modèle computationnel ayant implémenté l'approche double-voie est le modèle DRC (Coltheart, 2005 ; Coltheart et al., 2001). Coltheart et al. (2001) ont fourni un ensemble de principes et d'hypothèses sur lesquels repose le développement du modèle DRC. Parmi ces principes, les auteurs reprennent le concept de *nested modelling* (ou en français « modélisation imbriquée ») proposé initialement dans le champ du traitement visuel de mots par Grainger et Jacobs (1996) (voir aussi Grainger & Jacobs, 1998). Selon ce principe, le développement des modèles computationnels doit être cumulatif et chaque nouveau modèle devrait rendre compte de nouveaux phénomènes, en plus de ceux expliqués par les précédents. Le modèle DRC contient donc plusieurs composants inspirés ou repris des modèles antérieurs.

Description générale Le modèle DRC (voir Figure 1.1) comprend des composants qui sont communs aux deux voies de traitement et des composants spécifiques à chacune des deux voies. Les composants communs aux deux voies se situent en entrée et sortie du modèle. L'analyse visuo-orthographique de la séquence de lettres est commune aux deux voies et comprend un mécanisme d'analyse des traits qui composent les lettres ainsi qu'un mécanisme de reconnaissance des lettres. En sortie, les séquences de phonèmes obtenues à partir des traitements lexicaux et/ou sous-lexicaux sont gérées au sein d'un « Système phonémique commun ».

La présentation d'une séquence écrite entraîne des activations dans la première couche de traitement du modèle : le niveau de traitement des traits visuels. Ce niveau est basé sur le système de traits proposé par Rumelhart et Siple (1974). Pour chaque position, les unités de traits visuels sont constituées d'un ensemble de 16 traits (*features*) qui sont activés ou désactivés (1 ou 0) selon qu'ils correspondent, ou non, aux propriétés de la lettre saisie à cette position (voir Figure 1.2). L'information se propage ensuite aux composants des voies lexicale et sous-lexicale. En fin de traitement, le système phonémique est activé par les activations des unités en sortie des deux voies. Le traitement de ces informations phonémiques permet de produire la réponse phonologique correspondant à la séquence écrite présentée à l'entrée du modèle.

Implémentation La voie lexicale (non-sémantique) est implémentée en combinant le modèle *Interactive-Activation* (IA) de McClelland et Rumelhart (1981) et une version simplifiée du modèle de production de la parole de Dell (1986). Le modèle IA a été conçu initialement pour expliquer la

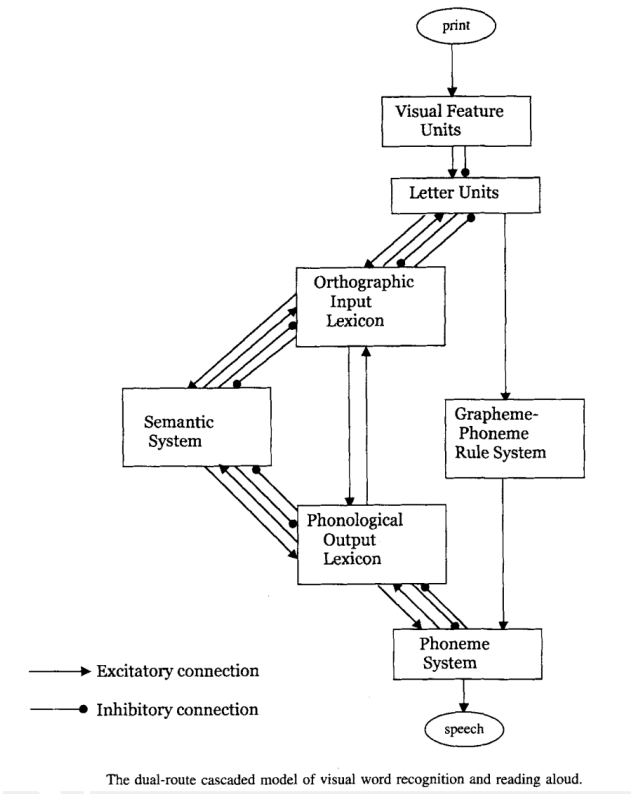


FIGURE 1.1 – **Représentation graphique du modèle DRC (*Dual-Route Cascaded Model*) tirée de l'article de Coltheart et al. (2001).** Dans ce schéma, l'information circule de haut en bas et le modèle possède une architecture en trois voies : une voie sous-lexicale, une voie lexicale et une voie sémantique. La troisième voie, passant par le système sémantique, n'est pas implémentée dans le modèle computationnel.

FIGURE 1.2 – **Le système de traits employé par Rumelhart et Siple (1974) pour la représentation des lettres.** En haut, la police des lettres utilisée dans le programme de simulation et, en bas, l'ensemble des traits visuels utilisés pour les construire.

reconnaissance visuelle de mots. Il se caractérise par un traitement parallèle des lettres et un codage positionnel strict. Les unités « lettres » du mot, activées par le niveau des traits dans chaque position, activent à leur tour les unités « mots » au sein du lexique orthographique. Le modèle IA a été modifié légèrement dans le modèle DRC : (1) Coltheart et al. (2001) utilise un paramétrage différent de celui du modèle IA et (2) les auteurs ont fait usage d'une lettre supplémentaire qui correspond à une « case vide » ou une absence de lettre, étant donné que le modèle IA ne fonctionnait à l'origine que pour des mots de longueur fixe. Le modèle DRC peut donc opérer avec un lexique

de 7 981 mots anglais monosyllabiques de longueur allant de 1 à 8 lettres. L'activation des éléments du lexique orthographique dépend de leur fréquence. Les mots ont une activation initiale proportionnelle au logarithme décimal de leur fréquence. Les unités du lexique orthographique activent les unités du lexique phonologique qui, à leur tour, activent les unités phonémiques correspondantes[2].

La voie sous-lexicale, implémentée symboliquement par des règles de conversion graphème-phonème, permet de générer la forme phonologique du stimulus. Le traitement est dit sériel dans la voie sous-lexicale puisque les règles de conversion s'appliquent séquentiellement de gauche à droite pour l'identification successive des graphèmes et leur recodage phonologique. A cet égard, les auteurs se basent sur les résultats d'un algorithme développé par Coltheart et al. (1993) et qui permet d'apprendre et de définir un ensemble de règles de conversion des graphèmes en phonèmes à partir d'une base de données d'environ 3 000 mots anglais. Le stimulus orthographique est converti à l'aide de cet ensemble de règles : seule la première lettre est convertie en début de traitement sous-lexical, après un nombre fixe de cycles, puis la seconde lettre est présentée et l'ensemble des deux lettres est converti et ainsi de suite jusqu'à la dernière lettre. Le système de conversion privilégie les graphèmes les plus longs aux graphèmes plus courts et les phonèmes activés sont ceux qui sont associés à la séquence de lettres la plus longue. Ce système mime le fonctionnement d'un « *parser* » (système de segmentation) graphémique.

La forme phonologique est générée en plusieurs étapes selon le nombre de lettres du stimulus à lire, et le temps de traitement sous-lexical est, par conséquent, fonction de la longueur de la séquence de lettres. Au contraire, dans la voie lexicale, la forme phonologique est obtenue en une seule étape, par accès à la représentation mémorisée du mot. Dans cette voie, on dit que le traitement est parallèle.

Par ailleurs, le modèle DRC suppose une compétition entre les deux voies de traitement. Ainsi, la sortie phonologique produite pour une entrée donnée est calculée à partir des informations issues soit de la voie sous-lexicale soit de la voie lexicale, selon la vitesse de traitement au sein de chacune des deux voies. Coltheart et al. (2001) postulent *a priori* que la voie lexicale jouit d'un avantage initial, le traitement est le plus souvent plus rapide au sein de cette voie, ce qui est implémenté par un démarrage du traitement légèrement décalé en faveur de la voie lexicale (voir aussi, Coltheart, 2005).

L'asymétrie du temps de traitement en faveur de la voie lexicale favorise notamment la lecture correcte des mots irréguliers tels que « *pint* » , en anglais. Sans cette asymétrie du temps de traitement, le modèle pourrait régulariser la prononciation du mot en appliquant les conversions graphème-phonème (prononcé /pɪnt/ comme le mot anglais « *mint* »). En revanche, on note que l'inhibition initiale, que peut induire la basse fréquence d'un stimulus, désavantage la voie lexicale

[2]Le modèle DRC implémente une transmission en cascade des activations entre les niveaux adjacents. Entre le niveau lettre et celui du lexique orthographique et entre le niveau phonémique et celui du lexique phonologique, les auteurs utilisent des connexions bidirectionnelles, excitatrices et inhibitrices. Lorsqu'une unité est activée par le niveau inférieur via les connexions excitatrices, les connexions inhibitrices, elles, permettent de diminuer l'activité des unités compétitrices.

et augmente la probabilité qu'une prononciation régularisée l'emporte.

Simulations et Résultats La performance du modèle DRC a été évaluée sur les mots de son lexique et une liste de 7 000 pseudo-mots tirés de la base de données de pseudo-mots ARC (Rastle, Harrington, & Coltheart, 2002). Le modèle lit parfaitement presque tous les mots de son lexique (99.98% des mots lus correctement) et obtient un taux d'erreur de l'ordre de 1% pour les pseudo-mots. Ses bonnes performances sont dues, d'une part, à l'encodage localiste de l'orthographe et de la phonologie des mots et, d'autre part, à l'application de règles de conversion graphème-phonème strictes et explicites.

Le modèle DRC a permis de rendre compte d'un très grand nombre d'effets comportementaux, comme les effets les plus classiques, tels que : l'effet de fréquence, c'est-à-dire que les mots fréquents sont lus plus vite que les mots peu fréquents, et l'interaction entre régularité et fréquence, c'est-à-dire que pour les mots peu fréquents, les mots irréguliers sont lus moins vite que les mots réguliers. Dans le modèle, l'interaction est due au fait que des formes phonologiques différentes sont générées par chacune des deux voies pour les mots irréguliers (la forme lexicale canonique et la forme régularisée obtenue par application des règles de conversion graphème-phonème) et que l'activation de la forme canonique au sein du système phonémique est retardée pour les mots de basse fréquence. Ainsi, la probabilité est forte que les formes canonique et régularisée soient en compétition au sein du système phonémique, contrairement aux mots de haute fréquence, dont la forme canonique est produite avant génération de la forme régularisée correspondante.

DRC reproduit également l'effet de la position de l'irrégularité dans le mot. Cet effet consiste en une diminution de l'effet de régularité quand la position de l'irrégularité dans le stimulus est plus éloignée du début du mot. Coltheart et al. (2001) affirment que leur modèle simule cet effet en conséquence directe du caractère sériel de la voie sous-lexicale. Selon leur interprétation, la probabilité de compétition entre les deux voies est plus forte pour les irrégularités situées en début de mot, comme dans « chorale » (le son /ʃ/ est généré très tôt via la voie sous-lexicale) qu'en fin de mot, comme dans « almanach ».

Du fait du caractère sériel de la voie sous-lexicale, le modèle DRC rend compte naturellement de l'effet de longueur en lecture de pseudo-mots. L'effet de longueur, selon lequel les stimuli courts sont lus plus rapidement que les stimuli longs, découle directement du traitement sériel par application des règles de conversion graphème-phonème via la voie sous-lexicale. En revanche, la reproduction de cet effet pour les mots est plus mitigée. Le modèle DRC échoue à reproduire l'effet de longueur lors de la lecture de mots en allemand et en grec (Kapnoula, Protopapas, Saunders, & Coltheart, 2017 ; Perry & Ziegler, 2002). Cet effet est plus marqué, d'après les auteurs, dans des langues transparentes telles que l'allemand et le grec par rapport à des langues opaques telles que l'anglais. Par ailleurs, Perry et Ziegler (2002) montrent qu'une nouvelle version du modèle DRC, forçant un traitement plus rapide via la voie sous-lexicale, permet effectivement de simuler l'effet de longueur observé étant donné la transparence de l'orthographe de la langue allemande. Plus spécifiquement, les auteurs augmentent la valeur d'un paramètre qui contrôle la force d'activation et, donc, la contribution de la voie sous-lexicale. Ceci témoigne à nouveau du fait que les effets de

longueur dans le modèle DRC dépendent de la mise en jeu des traitements de la voie sous-lexicale.

La simulation de la décision lexicale dans DRC se base sur les activations de la voie lexicale. Coltheart et al. (2001) s'inspirent du modèle MROM (voir Section 1.3 de ce chapitre) de Grainger et Jacobs (1996) et réussissent à rendre compte de plusieurs effets dans cette tâche tels que les effets de fréquence du mot, de fréquence du voisinage et de pseudo-homophonie. Cependant, l'implémentation de la décision lexicale hérite aussi de quelques problèmes comme l'incapacité à rendre compte de l'effet de longueur des mots ou de l'effet de voisinage du « corps » (*body neighborhood effect*), le corps étant l'attaque et le noyau de la syllabe dans un mot monosyllabique[3] (par exemple /kla/ dans /klaʃ/) (Grainger & Jacobs, 1996 ; Ziegler & Perry, 1998). On note également que cette implémentation produit des réponses inexactes quand les stimuli sont non-lexicaux. Les auteurs reconnaissent que la procédure par laquelle le modèle DRC prend des décisions lexicales est grossière en ce qui concerne les décisions « non » (Coltheart et al., 2001).

1.2 Le modèle CDP

Le modèle CDP est le modèle le plus récent qui implémente l'approche double-voie. Il s'agit d'une famille de modèles (CDP, CDP+, CDP++), qui a été développée de manière itérative avec pour objectif l'amélioration des performances à chaque nouvelle version. Dans cette section, nous décrivons principalement les versions les plus récentes de cette famille de modèles (CDP+ et CDP++), compte tenu du fait que plusieurs choix d'implémentation, optés dans les premières versions, ont été abandonnés ou redéfinis dans les suivantes.

Description et Architecture Perry et al. (2007) proposent le modèle CDP+ (voir Figure 1.3) qui suit également une architecture en deux voies et qui a pour objectif de fusionner deux modèles antérieurs : DRC et CDP. De manière similaire au modèle DRC, la séquence écrite active premièrement le niveau des traits visuels qui, à son tour, active le niveau des lettres. Les activations à ce niveau représentent l'identité des lettres perçues et alimentent les deux voies de traitement du modèle.

La voie lexicale de CDP+ est identique à celle du modèle DRC et elle repose aussi sur un réseau de neurones localiste de type IA. La nouveauté réside essentiellement dans l'implémentation de la voie sous-lexicale qui est très inspirée de celle du modèle CDP. L'élément central de cette voie est un réseau de neurones associatifs à deux couches, nommé TLA pour « *Two-Layer Assembly network* ». Ce réseau a été proposé par Zorzi, Houghton, et Butterworth (1998a) afin d'extraire les statistiques des correspondances graphème-phonème d'une langue donnée. La couche d'entrée encode l'identité des lettres de la séquence d'entrée et la couche de sortie celle de la séquence

[3] La structure de la syllabe se décompose en deux sous-unités de phonèmes : l'attaque (*onset* en anglais) et la rime. Cette dernière sous-unité se décompose, elle-même, en une sous-unité nommée le noyau (*nucleus* en anglais) et une autre sous-unité nommée la coda. En français, l'attaque et la coda peuvent être absentes dans une syllabe mais le noyau est systématiquement présent. Par exemple, considérons le mot « hiberner ». Celui-ci se compose de trois syllabes : la première syllabe est /i/ et ne contient donc que le noyau; la seconde syllabe est /bɛʁ/ et se compose d'une attaque /b/, d'un noyau /ɛ/ et d'une coda /ʁ/ ; enfin, la troisième syllabe ne contient pas de coda et se compose d'une attaque /n/ et d'un noyau /e/.

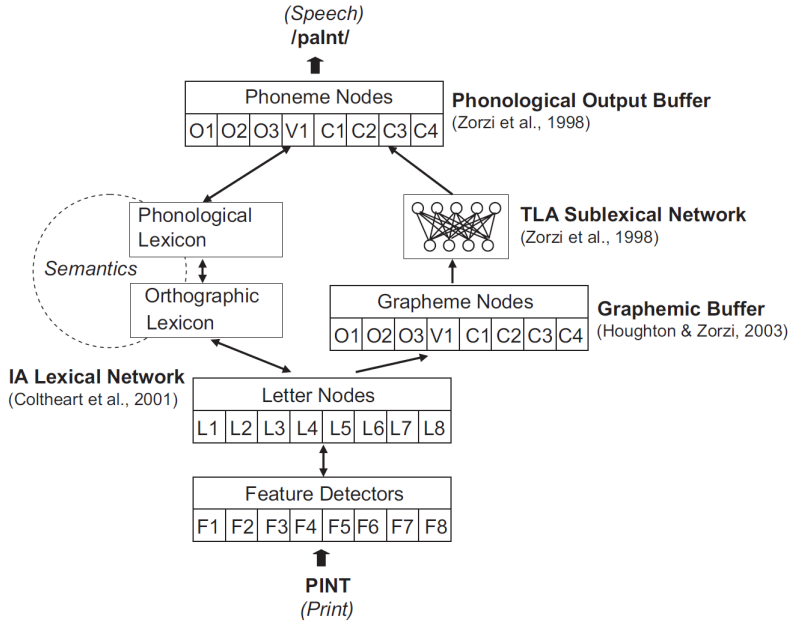


FIGURE 1.3 – **Représentation graphique du modèle CDP+ tirée de l'article de Perry et al. (2007).**

phonologique. Le réseau est entraîné en deux étapes. Dans la première étape, le réseau est exposé à une liste des correspondances graphème-phonème les plus fréquentes puis dans la seconde étape, un corpus de mots est employé pour entraîner le réseau. Les poids des connexions sont ajustés au cours de l'apprentissage à l'aide de la méthode de Widrow-Hoff (1960) qui permet de minimiser l'erreur globale en sortie du réseau.

A l'entrée du réseau TLA, et contrairement au modèle DRC, Perry et al. (2007) font l'usage de deux modules : le *parser* et le *buffer* graphémiques. L'information visuelle provenant du niveau lettre alimente en premier le *parser* graphémique qui permet de découper les mots en chaînes de graphèmes. Le *buffer* graphémique, quant à lui, est un module introduit initialement dans le modèle CDP de Zorzi, Houghton, et Butterworth (1998b) qui code les différents graphèmes en respectant la structure syllabique du mot. Ce module permet de représenter l'identité des lettres en les regroupant en graphèmes et les organisant selon leur position dans la syllabe (attaque, voyelle et coda).

Dans CDP+, le traitement de l'information dans la voie sous-lexicale est de nature sérielle bien que la propagation des activations dans le réseau TLA s'effectue de manière parallèle et simultanée dans tous les nœuds. Ceci est dû à la procédure de segmentation graphémique (*parsing*) qui est supposée impliquer une fenêtre « attentionnelle » qui s'étend sur trois lettres adjacentes et se déplace de gauche à droite sur la chaîne de lettres. Ainsi, les graphèmes du stimulus sont identifiés sériellement au sein du *buffer* graphémique avant d'être convertis en phonèmes.

Conceptuellement, les modèles DRC et CDP semblent reposer sur des principes théoriques identiques bien que certains choix d'implémentation les distinguent. Le fonctionnement « en cas-

cade », fondamental dans DRC, n'est pas fidèlement maintenu dans le modèle CDP. Un seuil de décision est appliqué aux activations au niveau des lettres. L'identité des lettres perçues est déterminée puis introduite dans la voie sous-lexicale. Une autre différence entre les deux modèles concerne l'implémentation du traitement sériel. La voie sous-lexicale dans DRC fonctionne selon un mode lettre-à-lettre tandis qu'elle fonctionne par groupes de trois lettres dans CDP.

A la sortie des deux voies se trouve un *buffer* phonologique, qui emploie également un codage syllabique avec des *slots* pour y représenter les prononciations des séquences orthographiques. Un système de décision permet de « valider » une sortie phonologique (l'identité du phonème) une fois qu'elle a atteint un seuil et s'est stabilisée.

CDP a été par la suite augmenté (CDP++, Perry, Ziegler, et Zorzi 2010) pour simuler la lecture de mots bi-syllabiques. Le modèle CDP++ reste très similaire au modèle CDP+. La voie lexicale peut traiter des stimuli de longueur allant jusqu'à 16 lettres et encode un lexique orthographique et phonologique de 30 000 mots. Quant à la voie sous-lexicale, CDP++ a simplement dupliqué les *buffers* graphémique et phonologique et le réseau TLA. Par la suite, Perry, Ziegler, et Zorzi (2013) se sont penchés sur le mécanisme de segmentation et ont proposé une nouvelle implémentation du *parser* graphémique.

Simulations et Résultats CDP+ démontre de bonnes performances de lecture sur les listes de mots et de pseudo-mots auxquels il est confronté. Sur les 7 383 mots anglais de son lexique, CDP+ en prononce 98,7% correctement et 70% des erreurs sont des régularisations ou des confusions entre homographes hétérophones (c'est-à-dire des mots qui s'écrivent de la même manière mais se prononcent différemment, comme par exemple « fils » en français, prononcé /fil/ si ce sont les fils de couture, et /fis/ pour le fils de sa mère). Sur 592 pseudo-mots anglais, CDP+ fait 37 erreurs. Ce taux d'erreur d'environ 5% est très similaire au taux d'erreur de 7,3% observé chez l'humain (Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994).

Le modèle CDP+ rend compte de l'effet de longueur en lecture de mots et en lecture de pseudo-mots. Cet effet est considéré comme une manifestation de la sérialité du traitement dans la lecture. Perry et al. (2007) notent que la source de la sérialité dans le modèle CDP relève de la nature du traitement dans la voie sous-lexicale alors que lors du traitement des mots, l'influence de la voie lexicale, dont le traitement est parallèle, entraîne une diminution de l'effet de longueur observé pour les mots. Par ailleurs, les auteurs ont comparé le modèle CDP standard à une version « parallélisée » du modèle. La seconde version est incapable de rendre compte de l'effet de longueur en lecture de pseudo-mots et ses prédictions corrèlent moins avec les données comportementales.

CDP+ peut, en principe, simuler la décision lexicale selon le même principe que pour les modèles DRC et MROM. A l'heure actuelle, et à notre connaissance, CDP+ n'a pas été utilisé pour simuler des expériences de décision lexicale.

Par ailleurs, le modèle amélioré CDP++ a été également confronté aux lexiques français et italien (Perry et al., 2014a, 2014b). A l'heure actuelle, le modèle CDP++ est considéré comme le plus puissant en termes de nombre d'effets qu'il est capable de simuler (pour une liste exhaustive des

effets, voir Perry, Ziegler, & Zorzi, 2010, p. 19).

1.3 Modèle MROM

Le modèle MROM (pour *Multiple Read-Out Model*) est le résultat de plusieurs modèles developpés de manière cumulative (Grainger & Jacobs, 1996 ; Jacobs & Grainger, 1992 ; voir aussi Ferrand & Grainger, 1994). Il ne concerne initialement que les traitements orthographiques dans la reconnaissance visuelle de mots et repose sur le modèle IA de McClelland et Rumelhart (1981). Ce modèle a permis de rendre compte de plusieurs effets en reconnaissance de mots et en décision lexicale (par exemple, l'effet de fréquence et l'effet de la densité de voisinage). Sa particularité réside dans sa capacité à rendre compte des distributions des temps de réaction en employant un tirage aléatoire pour définir les seuils d'identification des stimuli (Grainger & Jacobs, 1996). Plus précisément, les auteurs ont échantillonné des valeurs de seuils à partir d'une distribution normale centrée autour d'une valeur préalablement définie. Ce modèle a aussi introduit une méthode pour simuler la décision lexicale à partir de l'activation des représentations lexicales uniquement.

MROM-P (Jacobs et al., 1998) est une extension du modèle MROM permettant de simuler la lecture de mots. Cette extension repose sur le postulat théorique qui considère que l'information phonologique, en plus de l'information orthographique, joue un rôle important dans le traitement et l'identification des mots écrits. Cette théorie occupe une position intermédiaire dans le débat entre les théories de l'accès direct pour l'identification visuelle des mots et celles qui défendent la médiation phonologique (Ferrand & Grainger, 1992). Pour cela, MROM-P inclut des représentations phonologiques (sous-lexicales et lexicales) qui contribuent au traitement. Dans ce contexte, ce modèle postule que l'accumulation d'information phonologique, bien qu'elle soit plus lente par rapport à l'accumulation de l'information orthographique, peut influencer le résultat du traitement. Plus spécifiquement, les informations phonologiques commencent à fournir une entrée excitatrice au lexique et activent donc des mots qui peuvent entrer en conflit avec les mots activés par l'information orthographique. C'est ainsi que ce modèle explique les effets d'amorçage orthographique et phonologique en reconnaissance de mots (Ferrand & Grainger, 1994).

Pour implémenter ce modèle, Jacobs et al. (1998) utilisent deux niveaux de représentation : lexical et sous-lexical (voir Figure 1.4). Dans le niveau lexical, il existe deux types d'unités lexicales interconnectées encodant respectivement les représentations lexicales orthographiques et les représentations lexicales phonologiques. 2 323 unités lexicales phonologiques et 2 494 unités lexicales orthographiques sont utilisées pour représenter un lexique de 2 494 mots monosyllabiques anglais. Le nombre diffère du coté phonologique et orthographique du fait de l'existence d'homophones. Pour chaque unité lexicale, un niveau d'activation de repos est donné à chaque unité pour tenir compte de la fréquence des mots. Le niveau sous-lexical, quant à lui, contient également un ensemble d'unités orthographiques et un ensemble d'unités phonologiques connectées, elles aussi, directement. Le niveau sous-lexical orthographique encode directement les lettres et suit un codage positionnel relatif des lettres, tel que chaque unité encode une lettre dans une position orthographique. Le niveau sous-lexical phonologique suit un codage selon la structure d'une syllabe : 51 unités pour l'attaque, 34 unités pour le noyau et 78 unités pour la coda. Les unités de ce

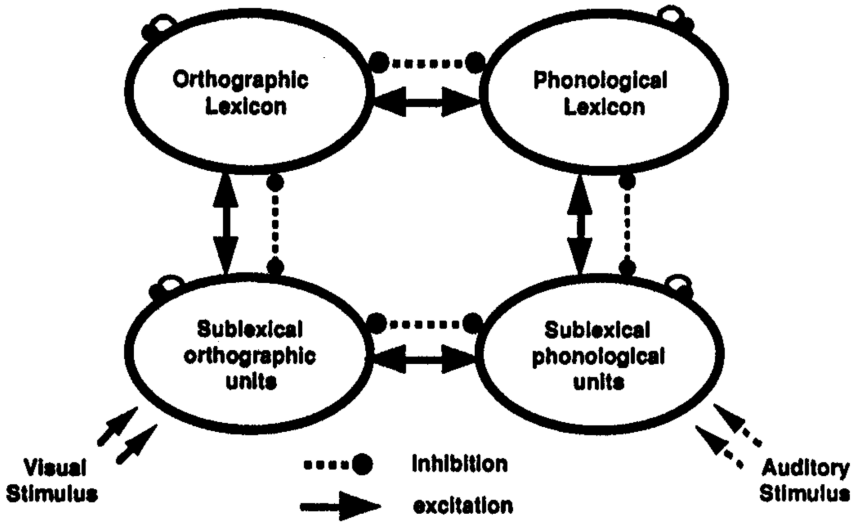


FIGURE 1.4 – **Représentation graphique du modèle MROM-P (Jacobs et al., 1998).**

niveau peuvent contenir un seul phonème ou une séquence de plusieurs phonèmes (voir Jacobs et al., 1998, p. 162, pour plus de détails).

Le modèle MROM-P se caractérise par la distinction entre les lexiques phonologique et orthographique, ainsi que la distinction entre les représentations lexicales et sous-lexicales. Ce choix théorique a permis d'expliquer les effets d'amorçage avec des mots et des pseudo-mots similaires orthographiquement ou phonologiquement dans des tâches de décision lexicale et de lecture de mots. L'approche théorique du modèle MROM-P suppose l'existence de deux voies pour convertir l'orthographe en phonologie : une voie lexicale et une voie sous-lexicale. Cette approche présente donc plusieurs similarités conceptuelles avec les modèles CDP (Grainger & Ziegler, 2011).

2 Modèles à voie unique (*Single-Route Models*)

La seconde classe de modèles que nous passons en revue dans cette section regroupe les modèles de lecture qui font abstraction de l'hypothèse des deux voies de traitement. Il s'agit des modèles à voie unique. La particularité de ces modèles est que le traitement des mots et des pseudo-mots ne fait pas appel à des connaissances et des traitements de nature différentes.

Dans cette section, nous présentons trois modèles se revendiquant de cette approche théorique : le modèle en Triangle (Plaut et al., 1996 ; Seidenberg & McClelland, 1989), le modèle MTM (Ans et al., 1998) et le modèle NDR_A (Hendrix et al., 2019). Ces trois modèles ne suivent pas les mêmes cadres mathématiques de modélisation mais s'accordent sur l'hypothèse d'une seule voie de traitement.

2.1 Modèle en Triangle

Description générale Le modèle en Triangle est un modèle connexionniste distribué de la lecture qui a été introduit pour la première fois par Seidenberg et McClelland (1989). Ce modèle repose sur les principes du traitement distribué parallèle de l'information (*Parallel Distributed Processing* ou PDP). Cette approche suppose que les processus cognitifs peuvent être expliqués par la circulation d'activations entre des unités de traitement (neurones artificiels) entièrement connectées entre-elles, constituant ainsi un réseau de neurones artificiels. Ce dernier s'auto-structure par ajustement graduel des poids de connexions après plusieurs présentations successives d'exemplaires issus de la base de données en phase d'apprentissage. Par conséquent, cette approche se caractérise par le fait que les connaissances mises en jeu pour accomplir un traitement cognitif sont distribuées sur l'ensemble des unités et des poids synaptiques.

Dans ce cadre de modélisation, Seidenberg et McClelland (1989) proposent un modèle théorique de lecture original reposant sur un mécanisme de traitement homogène, unique et complètement parallèle pour lire tous les types de stimuli orthographiques. Les auteurs supposent que la lecture de mots met en jeu principalement trois types de connaissances : orthographiques, phonologiques et sémantiques. Ce modèle est composé de 3 couches d'unités simples qui correspondent respectivement à ces trois types de connaissances, d'où l'appellation « modèle en Triangle » (voir Figure 1.5). Les différentes couches interagissent via des connexions synaptiques et des couches d'unités cachées, encodant collectivement les connaissances du système et les correspondances statistiques entre les différents espaces de représentation.

Comme le postule l'approche PDP, les trois types d'informations sont représentés par des *patterns* d'activité répartis sur plusieurs unités. La lecture d'un stimulus nécessite donc qu'un *pattern* d'activation dans la couche orthographique puisse générer le *pattern* phonologique correspondant.

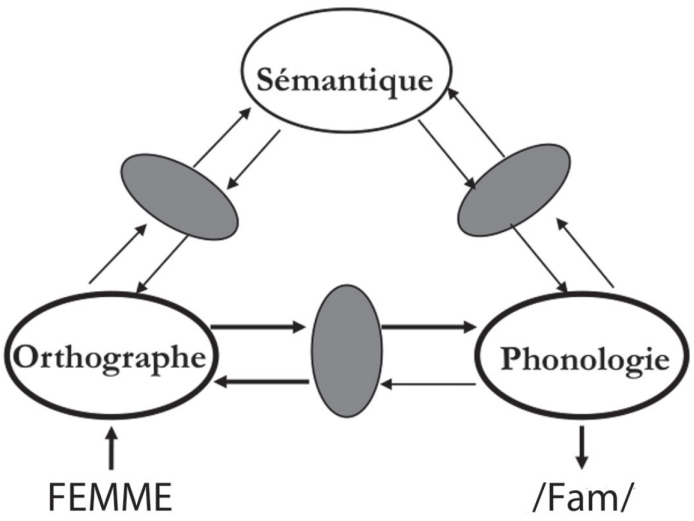


FIGURE 1.5 – **Modèle en Triangle de Seidenberg et McClelland (1989).** Seule la partie en gras a été implémentée. Chaque ellipse désigne une couche cachée de neurones artificiels.

Implémentation Initialement, le modèle en Triangle n'a été implémenté que partiellement par Seidenberg et McClelland (1989) et se limitait au réseau orthographe-phonologie. Ce même modèle a été légèrement modifié dans la version de Plaut et al. (1996) afin d'en améliorer les performances. La composante sémantique n'a été introduite que postérieurement par Harm et Seidenberg (2004). Nous rappelons que nous nous intéressons et décrivons principalement les versions de Seidenberg et McClelland (1989) et de Plaut et al. (1996) étant donné que le traitement sémantique n'est pas traité dans le cadre de cette revue de littérature.

Le modèle de Seidenberg et McClelland (1989) se compose de trois groupes d'unités : 400 unités orthographiques, 200 unités cachées, 460 unités phonologiques. Les interactions sont bidirectionnelles entre la couche orthographique et la couche cachée mais ne sont qu'unidirectionnelles entre la couche cachée et la couche phonologique.

L'entrée orthographique et la sortie phonologique sont implémentées à l'aide d'un codage relatif, dit en triplets de Wickelgren (1969). Par exemple, le stimulus orthographique « CAR » est traité comme 3 triplets de lettres : _CA, CAR et AR_ avec le caractère '_' qui indique le début ou la fin de la séquence. Le niveau phonologique est représenté à l'aide de triplets mais dans un espace phonétique (Rumelhart & McClelland, 1986). Par exemple, le mot anglais « CAR » se compose de trois phonèmes /k/, /ɑ/ et /ɹ/ et peut être vu comme 3 triplets phonémiques : _kɑ, kɑɹ et ɑɹ_. Le codage employé dans le modèle se base sur des triplets de caractéristiques phonétiques tels que chacun contient une caractéristique du premier des trois phonèmes de chaque triplet, une caractéristique du deuxième et une caractéristique du troisième phonème. Ainsi la représentation phonologique du mot « CAR » correspond aux triplets : [_, *stop*, *voyelle*] , [*stop*, *voyelle*, *liquide*] et [*voyelle*, *liquide*, _]. Chaque triplet est encodé par un *pattern* d'activation sur les unités orthographiques ou phonétiques étant donné le caractère distribué des représentations dans le modèle. Ainsi, la prononciation d'un mot est obtenue suite à la propagation des activations du niveau orthographique vers le niveau phonologique via les unités cachées.

On note que le codage des unités phonologiques a été modifié dans la version du modèle proposée par Plaut et al. (1996) afin d'améliorer les performances en lecture de pseudo-mots et d'éviter le problème de dispersion, deux limites relevées dans le modèle de Seidenberg et McClelland (1989). Plaut et al. (1996) montrent qu'en changeant le mode de codage phonologique, le modèle simule correctement les données humaines. La version de Plaut et al. (1996) comporte également un nombre plus petit d'unités de représentations orthographiques et phonologiques, et donc des représentations moins « distribuées » : 166 neurones artificiels au total contre plus de 500 dans la première version du modèle.

Dans la phase d'apprentissage, que ce soit dans la version de Seidenberg et McClelland (1989) ou celle de Plaut et al. (1996), un ensemble de 2 884 mots monosyllabiques anglais, provenant de la liste de Kučera et al. (1967), a été utilisé comme base d'entraînement. Un algorithme de rétropropagation du gradient d'erreur a été utilisé pour l'ajustement des poids synaptiques. Ainsi, le réseau s'auto-structure par renforcement des connexions entre les unités de la couche orthographique et celles de la couche phonologique, de façon à capturer progressivement les régularités

statistiques de la langue. La probabilité qu'un mot soit présenté dans la phase d'apprentissage dépend directement de sa fréquence.

Simulations et Résultats Le modèle rend facilement compte de l'effet de fréquence. Les mots fréquents sont lus plus rapidement étant donné que leurs poids synaptiques sont davantage renforcés que ceux des mots peu fréquents durant la phase d'apprentissage. Le modèle en Triangle fournit également une explication à la différence subtile entre les notions de régularité et de consistance et montre, par le biais de la simulation, comment la consistance permet une meilleure prédiction des temps mesurés que la régularité. La régularité repose sur le respect des règles de conversion et donc applique une dichotomie séparant les mots irréguliers (renfermant des correspondances très rares, comme le « on » dans « monsieur ») et les mots réguliers. La consistance d'un mot est une notion plus nuancée et graduelle qui dépend principalement de la prononciation des mots voisins de celui étudié ; on parle également de consistance du voisinage du corps (*body-neighborhood consistency*) (Glushko, 1979 ; Jared, 2002 ; Peereman & Content, 1997). Cet effet peut être expliqué par la fréquence des mots voisins triés en deux catégories : des mots amis contenant la même correspondance graphème-phonème que celle du mot étudié et des mots ennemis contenant une correspondance graphème-phonème différente et donc concurrente. Ainsi, le temps de réaction pour lire un mot est plus grand quand la somme des fréquences des ennemis est supérieure à la somme des fréquences des amis et *vice versa* (Jared, McRae, & Seidenberg, 1990).

En effet, une séquence de lettres est consistante quand elle se prononce de façon identique dans tous les mots dans lesquels elle apparaît et inconsistante si elle se prononce de différentes façons. Bien que la majorité des mots irréguliers soient inconsistants, la notion de consistance pour les mots se distingue de celle de la régularité. En effet, la consistance d'un mot est évaluée par rapport à son voisinage : un mot est consistant si sa prononciation correspond à celles de ses voisins orthographiques mais inconsistant si sa prononciation ne correspond pas à celles de ses voisins. Par exemple, le mot anglais « *gave* », prononcé /geiv/, est régulier au vu de la correspondance A ↔ /ei/, mais il est inconsistant étant donné qu'il possède un voisin ennemi très fréquent (« *have* » prononcé /hæv/). En revanche, le mot anglais « *hike* » prononcé /haɪk/ est régulier et consistant étant donné que sa prononciation est la même que celle de tous les mots finissant par la rime « IKE » (par exemple « *bike* », « *strike* » et « *like* »).

Au sein du modèle en Triangle, l'effet de consistance découle naturellement du fait que la prononciation de chaque mot est influencée par tous les autres mots appris qui partagent des *patterns* d'activations communs. Les correspondances orthographe-phonologie des mots anglais « *lint* » et « *pink* » (prononcés /lɪŋt/ et /pɪŋk/), acquises par le modèle lors de l'apprentissage, vont participer à traiter un mot proche comme « *link* » (prononcé /lɪŋk/), même s'il n'a jamais été appris. Ces connaissances seront par contre inhibitrices lors du traitement d'un mot proche inconsistant comme « *pint* » (prononcé /paɪnt/).

De plus, le modèle en Triangle réussit à capturer les interactions observées entre la fréquence et les effets de régularité et de consistance : la différence de temps de réponse entre des mots

consistants et inconsistants est réduite quand ces mots sont très fréquents d'une part et, d'autre part, l'effet de fréquence reste assez faible quand on se limite à l'étude des mots réguliers. Ainsi, le modèle en Triangle est celui qui reproduit le mieux les effets de régularité et de consistance observés dans les données expérimentales (Zevin & Seidenberg, 2006).

Cependant, le modèle en Triangle de Seidenberg et McClelland (1989) possède des performances très limitées en lecture de pseudo-mots : en effet, seuls 64% des pseudo-mots sont lus correctement dans les simulations, ce qui est bien inférieur aux performances de sujets normolecteurs. Dans l'implémentation modifiée de Plaut et al. (1996), les performances du modèle pour la lecture des pseudo-mots ont été considérablement améliorées avec 97% pour les pseudo-mots inconsistants issus de l'étude de Glushko (1979) et 85% des pseudo-mots contrôles issus de celle de McCann et Besner (1987).

Afin de mieux rendre compte des effets relevant de la phonologie et des dyslexies liées à un trouble phonologique, Harm et Seidenberg (1999) ont augmenté le modèle en Triangle pour y inclure des aspects articulatoires sous forme d'un réseau d'attraction au niveau des représentations des phonèmes. L'apprentissage des représentations phonologiques a été, par conséquent, facilité et converge plus rapidement. De plus, la dégradation des représentations phonologiques entraîne des performances faibles pour la lecture de pseudo-mots.

En dégradant artificiellement les représentations phonologiques, cette version du modèle réussit à simuler des formes modérées de dyslexie phonologique qui se traduisent comportementalement par une difficulté à lire les pseudo-mots alors que la lecture des mots irréguliers est préservée. On en conclut qu'un déficit phonologique pourrait être responsable de ce type de dyslexie. Néanmoins, le modèle ne réussit pas à simuler les formes sévères de ce trouble en aggravant le déficit phonologique puisque cela affecte aussi bien la lecture des mots irréguliers que celle des pseudo-mots.

Le modèle de Seidenberg et McClelland (1989) ne permet pas de rendre compte de l'effet de longueur observé chez les lecteurs experts. Cet effet a été considéré comme au-delà des centres d'intérêt du modèle initial sous prétexte qu'il serait potentiellement lié à des facteurs visuels ou articulatoires. Pour répondre à cet enjeu, Plaut (1999) propose une implémentation, à l'aide d'un réseau de neurones récurrent doté d'un mécanisme simulant les fixations oculaires, qui permet un traitement séquentiel du stimulus d'entrée. Cette version, reposant sur les mêmes principes fondamentaux du traitement parallèle distribué, a introduit une sérialité partielle du traitement qui n'est pas présente dans les deux implémentations précédentes. Dans ce modèle à voie unique, un seul mécanisme central permet le passage de l'orthographe à la phonologie mais le traitement visuel est différent selon la nature du stimulus. En effet, un mot connu ne nécessite qu'une seule fixation et est donc traité rapidement ; tandis qu'un pseudo-mot ou un mot très peu familier requiert un traitement séquentiel, en plusieurs fixations, et implique une dénomination plus lente. Ainsi, l'effet de longueur pour les mots et les pseudo-mots relève des processus visuels et non des processus centraux impliqués dans la lecture.

En somme, l'originalité du modèle en Triangle est sa capacité à rendre compte d'un ensemble

d'effets, sans avoir recours au postulat fort de la compétition entre deux voies de traitement. Le système de conversion graphème-phonème explicite opérant séquentiellement lors du traitement n'est donc pas indispensable à la lecture des pseudo-mots et les effets de régularité ou de consistance peuvent apparaître au sein d'un système unique de lecture.

Néanmoins, le modèle en Triangle présente plusieurs limites :

1. Seuls des mots monosyllabiques sont utilisés dans les simulations alors que la grande majorité des mots en anglais se compose de plusieurs syllabes.
2. Le modèle simule « avec difficulté » les profils sévères de dyslexies acquises (voir Harm & Seidenberg, 1999).
3. La description se limite aux processus centraux de la lecture et n'intègre que de manière très minimaliste les processus visuels et articulatoires.
4. Les traitements dans le modèle en Triangle sont systématiquement parallèles et ainsi le modèle ne fournit pas d'explications pour la lecture sérielle, excepté dans le travail de Plaut (1999).

2.2 Modèle MTM

Description Générale Le modèle MTM « *Multi-Trace Memory* » est un modèle connexionniste de lecture qui adopte également une structure à voie unique (Ans et al., 1998). Néanmoins, il se distingue des modèles précédents par le fait qu'il est le seul à intégrer un composant d'attention visuelle qui prend la forme d'une fenêtre visuo-attentionnelle. Cette dernière joue un rôle stratégique et permet de moduler la quantité d'information pouvant être traitée lors d'une tâche de lecture. En effet, le déploiement attentionnel module le traitement du stimulus d'entrée qui est soit en une seule capture attentionnelle, soit en plusieurs captures attentionnelles, au sein d'un réseau unique.

Ce modèle se compose d'une couche orthographique d'entrée codant l'identité des lettres et leur position, d'une couche phonologique de sortie, d'une couche « d'écho orthographique » sur laquelle est recréé le pattern d'activation orthographique d'entrée et d'une couche centrale de mémoire épisodique qui encode les connaissances orthographiques. La Figure 1.6 présente un schéma de la structure du réseau.

Le modèle MTM se distingue du modèle en Triangle et des modèles double-voie par le fait qu'il postule l'existence de deux procédures de lecture au sein d'une seule voie de traitement : une procédure globale et une procédure analytique. La procédure globale est la procédure par défaut et suppose une fenêtre attentionnelle qui couvre le stimulus entièrement et permet de générer la forme phonologique correspondante en une seule étape de traitement ; tandis que la procédure analytique implique une fenêtre attentionnelle de taille réduite qui ne couvre qu'une partie du mot à la fois (une syllabe ou quelques lettres) et conduit à générer séquentiellement (en plusieurs étapes de traitement) l'information phonologique correspondante.

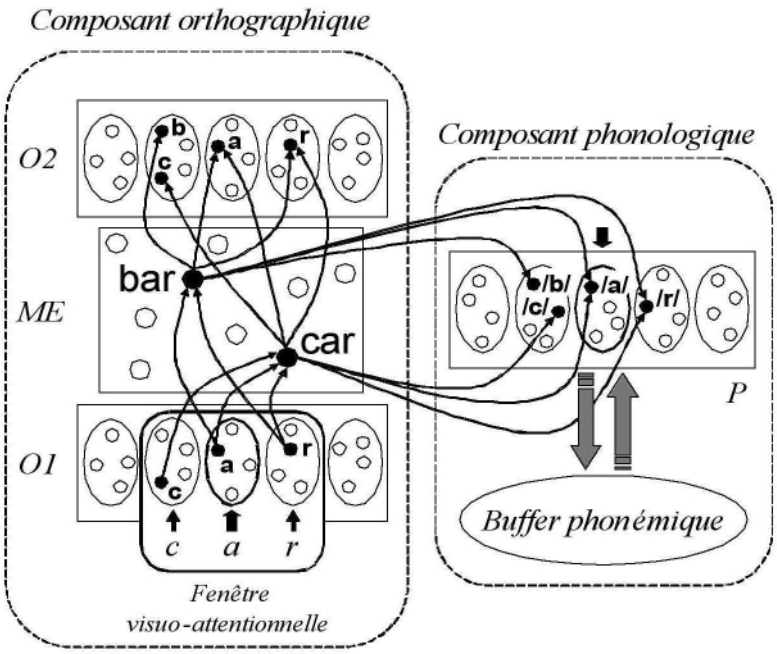


FIGURE 1.6 – **Représentation graphique de l'architecture du modèle MTM tirée de Ans et al. (1998).** O_1 est la couche orthographique d'entrée ; O_2 est une couche orthographique de sortie qui a une structure identique à O_1 ; P est la couche de sortie phonologique ; *ME* est la couche centrale de mémoire épisodique.

Dans le modèle MTM, le traitement est toujours initié en mode global et la procédure analytique se déclenche seulement si la procédure globale échoue à générer une prononciation. Ainsi, la procédure globale (parallèle) permet principalement de lire les mots familiers ou connus, alors que la procédure analytique (séquentielle) est plus souvent mobilisée pour lire les items peu familiers et notamment les pseudo-mots et les mots nouveaux. Toutefois, certains pseudo-mots sont lus correctement via la procédure globale. Le modèle MTM appartient au courant des modèles à voie unique, du fait qu'il met en jeu les mêmes principes computationnels pour lire tous les types de stimuli. Dans les deux procédures (globale et analytique), le traitement se base uniquement sur les relations entre orthographe et phonologie encodées dans les traces de mémoire épisodique (aussi appelées traces mnésiques).

Implémentation Le modèle MTM est composé de quatre couches d'unités neuronales : une couche orthographique d'entrée O_1 codant l'identité des lettres et leur position dans la séquence ; une couche O_2 sur laquelle est recréé le *pattern* d'activation orthographique d'entrée ; une couche phonologique de sortie P et une couche centrale de mémoire épisodique *ME*. Les couches O_1, O_2 et P se composent de plusieurs *clusters* indépendants. Un *cluster* se compose d'un ensemble d'unités élémentaires. Chaque *cluster* correspond à une position orthographique ou phonologique. Chaque unité des *clusters* des couches orthographiques (O_1 et O_2) réfère à une lettre de l'alphabet français et, de manière similaire, chaque unité de la couche phonologique P réfère à un phonème du français[4]. Une fenêtre attentionnelle, centrée autour d'un point de focalisation,

[4] Un caractère supplémentaire est utilisé pour référer au début et à la fin de séquence au niveau phonologique ou orthographique.

est implémentée pour définir le mode du traitement de l'entrée orthographique. Deux modes de traitement sont possibles : dans le mode global, la fenêtre capture l'ensemble des *clusters* qui représentent le stimulus orthographique, tandis que dans le mode analytique, la fenêtre ne capture que les positions orthographiques d'une portion du stimulus.

Dans les couches orthographiques, le codage de la position des lettres est défini relativement à un *cluster* d'origine qui correspond au point de focalisation de la fenêtre attentionnelle. La position de ce point est alignée sur la dernière lettre du premier graphème vocalique du stimulus ou de la portion traitée du stimulus (voir Figure 1.7). La couche phonologique P est divisée en plusieurs ensembles de *clusters* qui encodent successivement la phonologie des syllabes des stimuli traités. Dans un ensemble syllabique, chaque syllabe est représentée par une succession de phonèmes qui sont activés au sein de *clusters*. Chaque *cluster* correspond à une position phonémique déterminée par référence au noyau (phonème vocalique) de la syllabe. Les positions négatives codent les phonèmes appartenant à l'attaque de la syllabe et les positions positives correspondent aux phonèmes de la rime de la syllabe.

L'apprentissage du lexique se base sur un stockage de traces mnésiques au niveau de la couche *ME*. Les traces mnésiques sont créées durant l'apprentissage et peuvent référer soit à un mot entier soit à une portion de mot. En pratique, une entrée orthographique et une sortie phonologique sont

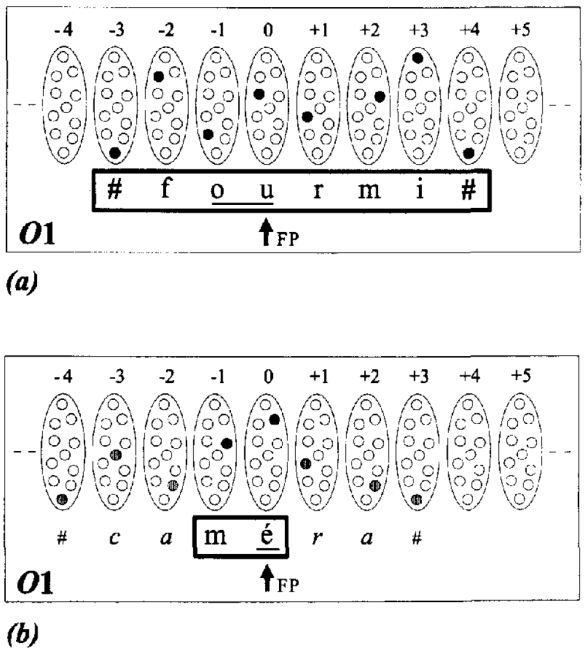


FIGURE 1.7 – **Codage positionnel des représentations orthographiques dans chaque mode de traitement, dans le modèle MTM.** La couche d'entrée orthographique O_1 est composée d'un ensemble de *clusters* de positions constitués d'unités élémentaires codant spécifiquement pour les caractères alphabétiques. La partie encadrée dans la chaîne de lettres d'entrée indique la fenêtre attentionnelle, les autres caractères (en italiques) étant contextuels. Le point de focalisation (FP) est situé sur la lettre la plus à droite du premier graphème vocalique de la séquence (souligné) capturé par la fenêtre attentionnelle. (a) Le mot « fourmi » est traité en mode global et la fenêtre attentionnelle comprend le mot en entier ainsi que les caractères de délimitation (#). (b) Le mot « caméra » est traité en mode analytique dans lequel la fenêtre focale est limitée à une sous-séquence de lettres (Ans et al., 1998).

présentées respectivement aux couches O_1 et P et les unités correspondantes sont activées. Au fur et à mesure de l'apprentissage, une nouvelle unité de la couche *ME* est choisie puis connectée aux unités correspondantes dans les couches O_1 et P (avec des poids de connexions égaux à 1). Un mot est représenté par plusieurs traces mnésiques selon le nombre de contextes différents dans lesquels il est rencontré. Les connexions entre la couche d'écho O_2 et la couche *ME* sont une image des connexions entre cette dernière et la couche d'entrée orthographique O_1.

Pour la phase d'entraînement, Ans et al. (1998) ont utilisé un lexique français de 13 165 mots monosyllabiques et polysyllabiques (de une à cinq syllabes) extraits de la base de données Brulex (Content, Mousty, & Radeau, 1990). Le nombre de présentations d'un mot lors de cette phase est défini en fonction de sa fréquence dans la base de données. Deux modes d'apprentissage sont possibles : global et analytique. Les mots monosyllabiques ont été appris en mode global seulement. Les mots polysyllabiques sont systématiquement appris à la fois globalement et analytiquement, c'est-à-dire en sous-unités codant chacune un segment de mot dans un contexte défini par la lettre qui le précède et celle qui le suit. Pour tous les mots, le nombre de traces mnésiques « globales » est fonction de la fréquence du mot à apprendre. En condition d'apprentissage analytique, les segments sont souvent des syllabes et le nombre de ces traces, donc « syllabiques contextualisées », est fixé *a priori* pour tous les mots et n'est pas directement fonction de la fréquence de chaque syllabe.

Dans la procédure globale, les unités créées encodent les nouvelles traces mnésiques des mots, tandis qu'en procédure analytique elles encodent des segments de mots (syllabes). Dans les deux procédures, seules sont définies les connexions liant les *clusters* positionnels, capturés par la fenêtre attentionnelle, et les *clusters* phonologiques à la trace mnésique correspondante [5]. Ainsi, l'apprentissage permet au modèle de représenter les connaissances lexicales orthographiques de deux façons : soit à l'aide de traces mnésiques globales (mots entiers), soit à l'aide de traces segmentales (une ou plusieurs syllabes contextualisées).

Lors de la lecture d'une séquence, les unités de la couche O_1 activent des traces mnésiques dans la couche *ME* qui, à leur tour, activent des unités de la couche d'écho O_2 et de la couche phonologique P. Le mode de lecture global est le mode par défaut : la fenêtre attentionnelle est large et capture l'ensemble du stimulus. Quand le stimulus présenté correspond à une trace mnésique, l'écho résultant est le plus souvent identique au stimulus. Les activations dans la couche P correspondent à la sortie phonologique et la lecture s'effectue en une étape. Si l'écho résultant n'est pas identique au stimulus, le mode bascule en procédure analytique. Dans cette procédure, la fenêtre attentionnelle se focalise uniquement sur une portion du stimulus à la fois, de manière séquentielle de gauche à droite, et la sortie phonologique est générée et assemblée de manière séquentielle au niveau de la couche P. Plus précisement, Ans et al. (1998) postulent qu'un *buffer* phonologique permet de maintenir temporairement et d'assembler les segments phonologiques successifs mais ce système n'a pas été implémenté. La sortie phonologique est celle qui possède

[5] L'apprentissage dans le modèle MTM est en tout ou rien. Les poids de connexions prennent des valeurs soit nulles, soit égales à 1. Il n'y a pas d'ajustement ou de renforcement de connexions avec l'apprentissage. Pour chaque exemple d'apprentissage, une nouvelle trace mnésique est créée et les poids des connexions correspondants sont fixés à 1.

l'activation maximale et le temps de réponse du modèle est défini en fonction des activations dans la couche de sortie.

Concernant la fenêtre attentionnelle, son fonctionnement est assez grossier. Elle fonctionne comme un filtre attentionnel qui favorise le traitement des lettres qui sont sous le focus de l'attention. Toutes les lettres à l'intérieur de la fenêtre visuo-attentionnelle sont maximalement activées (à 1) ; tandis que toutes les lettres du contexte sont plus faiblement activées (à 0,05).

Principaux résultats Le modèle MTM permet de produire la prononciation de stimuli lexicaux et non-lexicaux en ayant des performances similaires à celles d'un lecteur expert, sans utiliser de système de conversion direct des graphèmes en phonèmes, ni recourir au postulat des deux voies de traitement. MTM suppose les mêmes principes computationnels et les mêmes mécanismes pour traiter tous les stimuli orthographiques. L'attention visuelle, modélisée par la fenêtre attentionnelle, fait partie de ces mécanismes et permet de basculer d'une procédure de traitement à l'autre (globale ou locale). Les pseudo-mots sont traités davantage par la procédure analytique et, au vu du caractère sériel de cette dernière, le temps de réponse estimé par le modèle MTM est plus grand pour les stimuli non lexicaux que pour les stimuli lexicaux.

MTM rend compte aussi des effets de fréquence des mots, des effets de lexicalité, et de la consistance du voisinage. En effet, les mots fréquents activent un plus grand nombre de traces que les mots peu fréquents et la prononciation d'un stimulus est générée à partir de l'activation des traces mnésiques qui lui sont orthographiquement similaires. Par ailleurs, sans recourir à l'hypothèse des deux voies, MTM rend compte des effets de longueur pour les pseudo-mots et de l'effet de la position de l'irrégularité dans les mots. Cependant, il ne prédit pas l'effet de longueur en lecture de mots (Ans et al., 1998).

Si l'on s'intéresse aux pathologies de la lecture, le modèle MTM permet de simuler les dyslexies acquises secondaires à un déficit attentionnel et postule l'indépendance entre traitements phonologique et attentionnel. Il permet de prédire et caractériser les dyslexies causées par un déficit visuo-attentionnel. Le fait de posséder une fenêtre attentionnelle réduite limite la capacité du lecteur à reconnaître le stimulus en une seule étape et force un traitement séquentiel, se traduisant par des effets de longueur et une vitesse de lecture ralentie. Cela mime assez bien les profils de lecture décrits dans certaines pathologies acquises (dyslexies lettre-à-lettre par exemple). Ainsi, un lecteur avec une fenêtre visuo-attentionnelle réduite aura des difficultés pour lire les mots irréguliers qui nécessitent une identification lexicale complète, en une étape, du stimulus. Cela correspond assez bien à certaines formes de dyslexies de surface acquises dont les autres modèles ont du mal à rendre compte. Par ailleurs et bien que les formes développementales de dyslexie n'aient pas été simulées dans l'article original, le modèle permettrait de simuler les formes de dyslexies de l'enfant caractérisées par un déficit de l'empan visuo-attentionnel, soit une fenêtre attentionnelle réduite. D'autres simulations, menées dans le cadre du modèle MTM, ont permis de démontrer la capacité du modèle à reproduire les patterns de performance en lecture et en décision lexicale décrits en contexte développemental suite à un empan réduit (Juphard, Carbonnel, & Valdois, 2004).

2.3 Modèle NDR$_A$

Description générale Un des modèles computationnels récemment proposés, faisant également l'hypothèse d'une voie unique de traitement, est le modèle NDR$_A$ pour *Naive Discriminative Reading Aloud* (Hendrix et al., 2019) (voir Figure 1.8). Ce modèle est une extension du modèle NDR qui est un modèle de reconnaissance de mots (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011). Le modèle NDR$_A$ est présenté comme un modèle connexionniste qui suppose une seule voie de traitement lexicale. La lecture des mots est naturellement effectuée en se basant sur les connaissances lexicales du modèle, mais la lecture de pseudo-mots repose également sur l'activation des connaissances lexicales puisqu'elle est effectuée via l'activation du voisinage orthographique (seules ces représentations lexicales activent les unités phonologiques).

Implémentation Par ailleurs, le modèle NDR$_A$ suppose les niveaux de traitement classiques en lecture et reconnaissance de mots. Le premier niveau de traitement consiste en un mécanisme d'interprétation et de décodage de l'entrée visuelle qui permet d'activer les unités orthographiques du niveau suivant, c'est-à-dire le niveau des représentations orthographiques. Ces unités orthographiques activent les représentations lexicales des mots cibles et des mots orthographiquement similaires. De manière similaire, les unités lexicales activent les unités du niveau phonologique et permettent donc de générer une prononciation du stimulus orthographique.

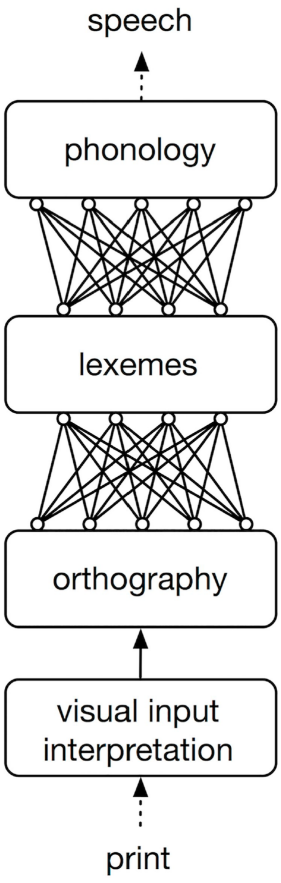


FIGURE 1.8 – **Architecture du modèle NDR$_A$ (*Naive Discriminative Reading Aloud*) (Hendrix et al., 2019).**

Le niveau des traitements visuels n'est pas le champ d'intérêt central du modèle NDR$_A$. Les auteurs supposent donc un système idéalisé et simplifié qui permet le découpage du stimulus visuel en lettres et bigrammes, qui sont une combinaison de deux lettres[6]. Les auteurs introduisent également une mesure de complexité qui quantifie les effets de similarité entre les lettres de l'alphabet.

Le cœur du modèle NDR$_A$ est constitué de deux réseaux de neurones : le premier lie les unités orthographiques aux unités lexicales qui sont à leur tour connectées, via le deuxième réseau, aux unités phonologiques. Le caractère localiste du réseau fait que chaque unité réfère à une représentation précise : une unité orthographique réfère à une lettre de l'alphabet ou à un bigramme, chaque unité lexicale réfère à un mot du lexique et chaque unité phonologique réfère à une demi-syllabe qui se compose soit d'un *onset* et d'une voyelle ou d'une voyelle et d'une coda. Chaque sous-réseau est un réseau de Rescola-Wagner (1972). Au niveau orthographique, l'activation d'une unité lexicale, c'est-à-dire d'un mot du lexique, est égale à la somme des activations des unités orthographiques (lettres et bigrammes) qui la composent. Les poids de connexions phonologiques allant du niveau lexical vers le niveau phonologique sont obtenus suite à un apprentissage supposant des demi-syllabes comme entrée et des mots comme sortie.

L'architecture proposée permet aux unités orthographiques d'activer les représentations lexicales des mots cibles et de diffuser l'activation aux représentations lexicales de mots orthographiquement voisins qui permettent, à leur tour, d'activer les unités phonologiques en sortie. La lecture des pseudo-mots ne fait pas intervenir une représentation lexicale cible mais se fait de manière similaire en se basant uniquement sur les prononciations des mots voisins.

A titre d'exemple, le stimulus anglais « DEAL » implique l'activation des lettres D, E, A et L et des cinq bigrammes '#D', 'DE', EA', 'AL' et 'L#'. Cela permet d'activer l'unité qui réfère au mot *deal* et de manière moins prononcée les unités qui réfèrent à des mots voisins tels que *dead* et *heal*. Au niveau phonologique, les demi-syllabes /di/ et /il/ seront activées par les unités du niveau lexical.

NDR$_A$ est un modèle statique, c'est-à-dire qu'il ne permet pas de simuler l'évolution des activations dans le temps et ne permet pas de simuler des temps de réponse de manière directe. Les temps de réponse prédits par le modèle sont calculés à partir du niveau d'activation des unités de la couche de sortie. Ainsi, une forte activation correspondrait à un temps de réponse court, tandis qu'une faible activation correspondrait à des temps de réponse longs.

Principaux résultats Hendrix et al. (2019) considèrent l'architecture des modèles en une seule voie comme parcimonieuse. Par ailleurs, la performance globale du modèle NDR$_A$ est comparable à celle des modèles double-voie, étant donné qu'il génère des prononciations correctes pour 99,32% des mots et pour 98,49% des pseudo-mots. Néanmoins, notons que toutes les simulations considérées se limitent à des stimuli monosyllabiques.

[6] Pour constituer les bigrammes, toutes les paires de lettres sont considérées. Pour les lettres externes, on utilise un caractère supplémentaire, noté '#'. Les bigrammes permettent d'encoder l'ordre des lettres dans un stimulus à partir de leurs positions relatives.

Les résultats obtenus par le modèle NDR$_A$ montrent que l'architecture à voie unique est capable de capturer un large éventail d'effets documentés dans la littérature de lecture à haute voix, tant pour les mots que pour les pseudo-mots. Ainsi, en raison de la coactivation d'éléments lexicaux ayant des orthographes similaires durant le traitement, le modèle NDR$_A$ explique les effets de régularité et de consistance. En effet, les représentations lexicales coactivées lors du traitement d'un stimulus participent à l'activation ou à l'inhibition de la représentation phonologique du mot cible. La prononciation cible est plus activée si elle est consistante avec la prononciation des mots voisins mais elle sera moins activée si elle est inconsistante.

Le modèle reproduit également un effet de longueur pour les mots et les pseudo-mots mais celui-ci est de plus grande magnitude que dans les données comportementales. Les auteurs attribuent l'effet de longueur principalement à la complexité visuelle du stimulus et considèrent qu'il est au moins partiellement déterminé par des processus extra-linguistiques. Par ailleurs, NDR$_A$ ne réussit pas à simuler l'effet de la position de l'irrégularité ; mais Hendrix et al. (2019) suggèrent que la prise en compte d'un traitement sériel supplémentaire pourrait pallier ce problème.

3 Discussion

La lecture est un champ intensément étudié et plusieurs approches théoriques ont été proposées pour étudier différents aspects de ce domaine. Nous avons décrit les modèles les plus influents en insistant plus particulièrement sur les questions que nous allons traiter dans cette thèse. Dans cette discussion, nous allons comparer ces modèles. Celle-ci se structurera en trois parties. La première partie traitera de l'architecture générale du système de lecture et des processus cognitifs impliqués. Nous nous intéresserons, dans la deuxième partie, au rôle des processus visuels et particulièrement au rôle de l'attention visuelle dans le traitement des mots isolés. Dans la troisième partie, nous discuterons, à la lumière des modèles théoriques, de la nature parallèle ou sérielle des traitements dans la lecture.

3.1 Architecture du système

Niveaux de représentations Dans les modèles computationnels de lecture, les chercheurs définissent plusieurs niveaux de représentations. Chaque niveau regroupe des représentations de même nature. Par exemple, le niveau perceptif réfère aux représentations des lettres, le niveau lexical à celles des mots et le niveau phonologique à celles des phonèmes.

Tous les modèles s'accordent sur l'existence d'un système d'extraction des traits visuels et d'identification des lettres comme première étape du traitement. C'est souvent la seule composante visuelle qui est postulée par la plupart des modèles. Les différentes implémentations des modèles double-voie postulent toutes un système de traits (de type IA) et un codage positionnel strict des lettres. Le modèle en Triangle n'implémente pas de système de traits et utilise un système de codage positionnel relatif en triplets. Le modèle NDR$_A$ détaille relativement plus les traitements entre son premier niveau, celui des traits visuels, et le second niveau orthographique, celui des lettres. Plus spécifiquement, les auteurs, d'une part, incluent la confusion et la similarité

entre les lettres et, d'autre part, ils supposent que l'identité des lettres repose sur un codage non strictement positionnel en lettres et bigrammes. Quant au modèle MTM, il n'inclut pas de niveau de traitement des traits visuels, mais suppose l'existence d'un module d'attention visuelle.

Nous notons qu'il existe un ensemble de phénomènes, bien identifiés dans les études comportementales sur le traitement visuel des séquences orthographiques, qui ne sont pas implémentés dans les modèles de lecture, bien qu'ils soient importants pour rendre compte de certains effets en reconnaissance de lettres et de mots. Ainsi, les modèles de lecture n'ont aucun mécanisme pour rendre compte de l'encombrement visuel ou *crowding* (c'est-à-dire qu'une lettre est plus difficile à identifier lorsqu'elle est entourée de deux autres lettres) et ils ne postulent pas de gradient d'acuité visuelle (Jacobs & Grainger, 1994). Ceci est d'autant plus surprenant que nombre de modèles de reconnaissance de mots (modèles limités aux traitements orthographiques sans prise en compte des traitements phonologiques ; pour une revue voir Phénix et al. 2016) postulent l'existence d'un gradient d'acuité et/ou d'un mécanisme d'interférence latérale, à défaut de proposer un module attentionnel. Malheureusement, comme le souligne Norris (2013), les modèles de lecture à haute voix, même les plus récents, n'intègrent pas les mécanismes postulés par les modèles récents de reconnaissance de mots. Jusqu'ici, aucun des modèle récents de reconnaissance de mots n'a été étendu en modèle de lecture, par ajout d'un composant phonologique. Il est également frappant que la plupart des modèles de lecture de mots isolés ne postulent pas de composants visuo-attentionnels alors que de plus en plus d'études expérimentales suggèrent que l'attention visuelle est indispensable à la reconnaissance des mots écrits et que les modèles de contrôle oculomoteurs postulent que l'attention visuelle sélectionne le mot à traiter dans la phrase, ce qui permet le déplacement de l'œil sur ce mot et initie son traitement (Engbert, Longtin, & Kliegl, 2002 ; Risko, Stolz, & Besner, 2010).

Dans la « chaîne » de traitement permettant la lecture à haute voix, le processus central décrit dans les modèles de lecture est celui qui permet le passage des représentations orthographiques aux représentations phonologiques. C'est le module central sur lequel porte la plupart des débats théoriques et il est, par conséquent, celui où diverses propositions ont été faites. Dans tous les modèles, ce module central inclut un niveau de représentation lexical, qui peut-être implémenté dans un réseau distribué ou de manière localisée. Ces représentations permettent, d'une part, de décider si le stimulus traité est un mot connu ou non (décision lexicale) et, d'autre part, « d'accéder » à la représentation phonologique une fois que le mot orthographique a été reconnu. Les modèles s'accordent également sur le fait que les influences sémantiques ou de contexte devraient également intervenir à ce niveau. Cependant, la contribution de la sémantique, dans ces modèles, n'est simulée que de manière très simplifiée et quasiment uniquement dans le modèle en Triangle. Les modèles computationnels les plus influents (Triangle, DRC et CDP) font également l'hypothèse d'interactions *top-down* entre le niveau lexical orthographique, et le niveau perceptif. Ainsi, les connaissances lexicales peuvent influencer, au cours du traitement, la perception des lettres.

Le niveau phonologique se résume principalement aux représentations phonologiques qui s'activent suite aux calculs effectués dans les niveaux inférieurs. Le modèle DRC suppose également l'existence d'une boucle de rétroaction du niveau phonologique vers le niveau lexical. Cette

hypothèse permet principalement de rendre compte des effets de pseudo-homophonie en lecture : les pseudo-homophones (par exemple, « ROZE » dérivé du mot « rose ») sont plus difficiles à rejeter dans la tâche de décision lexicale que ne le sont des pseudo-mots contrôles (par exemple, « ROVE »).

Ce niveau, dans le cas des modèles double-voie, inclut un *buffer* qui combine les sorties des deux voies de traitement et stocke, à l'aide de *slots*, les phonèmes associés à l'entrée orthographique. Dans les modèles « dynamiques », qui implémentent une évolution temporelle des activations, un système de décision est nécessaire. La façon de l'implémenter diffère selon les différents modèles mais tous les modèles présentés utilisent soit un critère basé sur un seuil d'activation, soit un critère basé sur la stabilité temporelle des résultats du calcul, soit les deux. Les aspects de planification et de réalisation des séquences articulatoires dans la lecture ont un rôle important puisque les mesures comportementales qui évaluent les performances de lecture utilisent classiquement le temps de réaction, c'est-à-dire le temps écoulé entre la présentation du stimulus écrit à l'écran et le début de l'énoncé oral de sa prononciation. Malheureusement, ces aspects ne sont pas encore implémentés dans les modèles computationnels de lecture.

Voies de traitement L'architecture en une ou deux voies s'est posée comme l'élément clé pour distinguer les modèles de lecture. Les approches « double-voie » (ou multiples voies), postulent que le passage des représentations orthographiques aux représentations phonologiques, sur l'ensemble des stimuli possibles, nécessite plusieurs voies dont chacune fait appel à des processus cognitifs distincts. La voie lexicale permet la lecture des mots familiers et des mots irréguliers par une transformation directe et parallèle du stimulus en son, via l'espace lexical, tandis que la voie sous-lexicale nécessaire à la lecture des nouveaux mots et des pseudo-mots, permet une transformation sérielle de sous-séquences du stimulus en sons, sans passer par l'espace lexical.

Les modèles à « voie unique », quant à eux, postulent que la lecture experte peut être décrite par un chemin de traitement unique allant de l'identification des lettres de la séquence orthographique d'entrée à la production de la séquence phonologique correspondante. Cette classe de modèles trouve son origine dans le contexte des théories d'analogie (Glushko, 1979 ; Kay & Marcel, 1981) dans lesquelles la lecture de mots (connus ou non) s'effectue par « analogie » avec des mots du lexique. Ce sont alors les mêmes mécanismes cognitifs qui sont mis en jeu dans tous les types d'items (mots et pseudo-mots) et aucun mécanisme distinct n'est nécessaire à la lecture de pseudo-mots.

On peut ensuite nuancer entre deux sous-positions théoriques dans cette approche. La première position suppose l'existence d'un processus unique de lecture reposant sur la mise en relation des connaissances lexicales orthographiques et phonologiques. Le modèle en Triangle suit cette position puisqu'il s'inscrit dans l'approche de la modélisation PDP et décrit essentiellement l'apprentissage du *mapping* entre l'orthographe et la phonologie. Les connaissances lexicales n'y sont pas implémentées manuellement mais sont acquises au cours de présentations successives de la base de données de façon à ajuster les poids des connexions entre unités orthographiques et phonologiques. Après cette phase d'entraînement, le réseau de neurones encode de manière

distribuée les connaissances lexicales mais peut aussi inclure des correspondances entre unités sous-lexicales[7]. Le modèle NDR_A est également dans ce cas et ne suppose qu'une seule et unique procédure pour traiter les mots et les pseudo-mots.

La seconde position, plus subtile, est celle d'un modèle en une seule voie, et donc ne reposant que sur les connaissances lexicales, mais qui ne traite pas tous les stimuli de la même façon. Le modèle MTM est une implémentation computationnelle de cette position étant donné qu'il suppose une procédure globale pour les mots connus et une procédure analytique pour les pseudo-mots. L'attention visuelle est impliquée dans les deux procédures et c'est la largeur de la fenêtre attentionnelle qui détermine laquelle des deux procédures est choisie. Dans les deux cas, les partisans de la voie unique s'accordent sur le manque de parcimonie de l'hypothèse de deux voies impliquant des principes computationnels distincts.

Les partisans de l'approche double-voie, quant à eux, considèrent que l'hypothèse de l'architecture en deux voies est nécessaire et « indiscutable » pour deux raisons principales (Coltheart, 2005). Premièrement, elle fournit une explication directe de l'existence de profils opposés de dyslexies acquises et développementales, caractérisés soit par une lecture préservée des pseudo-mots et un déficit marqué en lecture de mots irréguliers, soit inversement par une lecture déficitaire des pseudo-mots et des performances normales en lecture de mots irréguliers. Deuxièmement, elle permettrait d'obtenir des modèles computationnels qui simulent un plus large éventail d'effets comportementaux qu'aucun autre modèle de lecture. Coltheart (2005, p. 23) énonce même que les théoriciens de la lecture sont unanimes quant à l'existence de deux procédures distinctes et donc de deux voies de traitement distinctes.

En revanche, l'approche à voie unique présente aussi des contre-arguments forts. Premièrement, Ans et al. (1998) proposent le modèle MTM dans lequel deux procédures (globale et analytique) sont possibles, tout en gardant une seule voie de traitement, le même type de connaissances et les mêmes processus. Deuxièmement, Seidenberg (2012, p. 197) énonce que le succès des modèles double-voie (DRC et CDP), en termes d'ajustement aux données, résulte du fait que plusieurs de leurs composants soient configurés *ad hoc* et en réponse à des données empiriques particulières. Cette approche, selon lui, ne permet pas une généralisation théorique solide (Seidenberg & Plaut, 2006).

Il est clair que l'approche double-voie est dominante dans cette littérature. Cependant, l'approche à voie unique a également montré sa capacité à rendre compte d'une grande partie des effets observés, incluant certains effets (comme la position de l'irrégularité dans le mot) qui étaient annoncés comme impossible à simuler en dehors du postulat double-voie. Il faut aussi noter que les effets de longueur, par exemple, ont été peu étudiés dans le contexte double-voie et sont même souvent omis des effets majeurs dont un modèle devrait rendre compte. Nous constatons que le débat théorique entre les approches double-voie et à voie unique n'est pas encore tranché.

[7] Seidenberg (2012) évoque qu'il est tout à fait possible qu'une spécialisation fonctionnelle se produise pendant l'apprentissage d'un réseau connexionniste et que celui-ci puisse se diviser en deux sous-réseaux : un réseau qui lit les mots réguliers et un autre qui lit les mots d'exception. Cependant, l'analyse des activations des unités pour des mots d'exception et les mots réguliers dans le modèle en Triangle ne confirme pas cette hypothèse (Plaut et al., 1996).

3.2 Attention visuelle et l'effet de longueur en lecture

La grande majorité des modèles de lecture de mots isolés n'inclut pas de composante d'attention visuelle, bien que des modèles de certains champs très connexes soulignent son rôle important, comme les modèles du contrôle des mouvements oculaires en lecture (Reichle et al., 2003 ; Snell et al., 2018) et certains modèles de reconnaissance de mots (LaBerge & Samuels, 1974 ; Mozer & Behrmann, 1990).

Le modèle MTM est le seul modèle de lecture (de mots isolés) qui implémente de manière explicite une fenêtre attentionnelle qui opère sur les unités visuelles orthographiques. Néanmoins la modélisation de l'attention visuelle est relativement grossière. La fenêtre attentionnelle est conçue comme un filtre qui favorise l'identification des lettres sous le focus attentionnel et impacte celle des lettres contextuelles. Chacune des lettres à l'intérieur de la fenêtre attentionnelle reçoit une quantité d'attention identique et uniforme (fixée à 1) alors que les lettres environnantes (hors fenêtre attentionnelle) lors du traitement analytique reçoivent une quantité moindre d'attention (fixée à 0,5). Ainsi, toutes les lettres d'un mot, même très long, sont maximalement activées lors du traitement global, ce qui est peu conforme aux données comportementales. Ce modèle postule que la lecture repose sur deux types de procédures, une procédure globale notamment impliquée dans la lecture des mots familiers et une procédure analytique mobilisée pour les items non-familiers. Ces deux procédures mettent en jeu les mêmes niveaux de traitement mais se distinguent par la taille de la fenêtre attentionnelle qui couvre l'ensemble des lettres de la séquence d'entrée en mode global et se focalise sur des sous-unités de cette séquence en mode analytique. Ainsi, dans le cadre théorique du modèle MTM, les mécanismes attentionnels sont censés jouer un rôle crucial dans l'exécution de la tâche de lecture. Cette hypothèse est appuyée principalement par l'existence de formes de dyslexies développementales qui se caractérisent par un déficit spécifique de l'attention visuelle et démontrent des difficultés lors du traitement des mots irréguliers dont la prononciation est le plus souvent régularisée (Bosse, Tainturier, & Valdois, 2007 ; Zoubrinetzky, Bielle, & Valdois, 2014). L'hypothèse d'une implication attentionnelle en lecture est également soutenue par les données en neuroimagerie démontrant l'implication de régions cérébrales liées à l'attention dans les effets de longueur (Valdois et al., 2006). Le modèle MTM prédit qu'une fenêtre attentionnelle anormalement réduite affectera les performances de lecture en causant des effets de longueur importants, en raison du nombre élevé de déplacements attentionnels et donc de l'augmentation du nombre de captures qui sont nécessaires pour segmenter et traiter le mot.

Au niveau comportemental, chez le lecteur expert, l'effet de longueur est de petite amplitude pour les mots et il est plus prononcé en lecture de pseudo-mots qu'en lecture de mots. Autrement dit, on observe une interaction longueur-lexicalité (Ferrand, 2000 ; Jared & Seidenberg, 1990 ; Weekes, 1997). Le modèle MTM considère cette interaction comme une conséquence de la rapidité du traitement avec la procédure globale utilisée pour les mots connus, contrairement au traitement lent et sériel de la procédure analytique mise en jeu pour lire les pseudo-mots. Si le modèle peut rendre compte de cette interaction, il prédit une absence d'effet de longueur en lecture de mots et en décision lexicale, ce qui ne cadre pas avec les données comportementales. Les

effets de la longueur orthographique sur la latence de dénomination, tant chez les sujets sains que chez les patients cérébrolésés, imposent des contraintes importantes aux théories de lecture des mots, et les modèles existants ne rendent pas suffisamment compte de ces effets.

Les effets de longueur dans le cas des modèles double-voie résultent d'un traitement séquentiel de la voie sous-lexicale. La voie lexicale est, en théorie et par construction, incapable de rendre compte de l'effet de longueur[8]. Dans les modèles CDP et CDP++, le *parser* graphémique, qui segmente le stimulus orthographique en graphèmes et qui est à l'origine du caractère séquentiel de la voie sous-lexicale, est quelquefois considéré comme mettant en jeu des mécanismes d'attention visuelle (Perry et al., 2007 ; Perry, Ziegler, & Zorzi, 2010), sans que ces mécanismes ne soient implémentés (voir cependant Perry et al., 2013). Dans ce modèle, le « mécanisme attentionnel » ne serait dès lors présent que dans la voie sous-lexicale, où il intervient après identification des lettres de la séquence d'entrée. Ce postulat n'est pas en accord avec les données comportementales (Besner et al., 2016 ; Risko et al., 2010) qui montrent une implication attentionnelle en lecture de mots.

Les modèles de lecture en Triangle sont dans l'incapacité de rendre compte des effets de longueur, ce qui a été largement souligné par Seidenberg et Plaut (1998). En revanche, la procédure séquentielle implémentée par Plaut (1999) a permis de résoudre ce problème. Celle-ci correspond à un focus attentionnel qui se déplace de gauche à droite sur la séquence orthographique et modélise le nombre de fixations requises pour analyser l'ensemble du stimulus. Plaut (1999) démontre qu'un réseau doté d'une telle composante visuo-attentionnelle est en mesure de rendre compte des effets de longueur observés lors du traitement des pseudo-mots longs.

3.3 Le traitement parallèle et sériel en lecture

Le traitement parallèle signifie que les lettres d'un stimulus orthographique sont toutes traitées simultanément tout au long du processus de reconnaissance de mots ; tandis que le traitement sériel signifie que le stimulus est découpé en sous-séquences (ensemble arbitraire de lettres, graphèmes ou syllabes) qui sont traitées de manière sérielle, c'est-à-dire l'une après l'autre. La majorité des modèles computationnels qui s'intéressent strictement à la reconnaissance visuelle de lettres et de mots adoptent l'hypothèse d'un traitement parallèle (Adelman, 2011 ; McClelland & Rumelhart, 1981 ; Norris, 2006). En revanche, dans le champ de la modélisation de la lecture, sur lequel nous nous concentrons dans la présente revue, cette question a fait l'objet d'un débat théorique (Coltheart & Rastle, 1994 ; Roberts, Rastle, Coltheart, & Besner, 2003 ; Zorzi, 2000).

Tous les modèles computationnels présentés postulent l'existence d'au moins un niveau de traitement parallèle des lettres et certains d'entre eux supposent l'existence additionnelle d'un mécanisme sériel. Parmi les modèles présentés, deux modèles postulent un traitement uniquement parallèle : le modèle en Triangle et le modèle NDR_A. Dans ces modèles, on postule le traitement parallèle comme principe computationnel et tous les stimuli, lexicaux ou non-lexicaux, sont traités de cette manière.

[8] Adelman et Brown (2008) implémentent un modèle de lecture qui ne contient que la voie lexicale du modèle DRC et redéfinissent toutes les valeurs de ses paramètres en utilisant une procédure d'optimisation. Ce modèle à voie unique, baptisé SRC pour *Single-Route Cascaded*, rend également compte de l'effet de longueur pour les mots.

Les modèles double-voie (DRC et CDP) et le modèle MTM supposent la nécessité des deux types de traitements dans la lecture, mais les hypothèses théoriques sous-jacentes sont différentes. Les modèles double-voie postulent un traitement parallèle dans la voie lexicale et un traitement sériel dans la voie sous-lexicale. Ces deux voies sont en compétition, le traitement démarre simultanément au sein de chaque voie. Le modèle MTM ne suppose qu'une seule voie de traitement mais c'est le contrôle de la fenêtre attentionnelle qui permet soit de traiter le stimulus dans sa globalité en un seul coup soit de le traiter sériellement en plusieurs étapes. Le modèle MTM suppose le recours systématique au traitement global dans un premier temps, et, en fonction de sa réussite ou non, l'emploi ultérieur du traitement sériel.

L'observation de l'effet de longueur et de l'effet de la position de l'irrégularité a suggéré un mécanisme de conversion sériel dans la lecture et a été une des motivations du caractère sériel de la voie sous-lexicale dans les modèles double-voie (Coltheart & Rastle, 1994 ; Kapnoula et al., 2017 ; Weekes, 1997).

Par ailleurs, Plaut et al. (1996) notent que les effets comportementaux sériels peuvent se produire même si le système de conversion de l'orthographe vers la phonologie est purement parallèle. En effet, les mécanismes « en aval », comme la décision phonologique, la planification des séquences phonémiques et leur articulation, pourraient être sériels (Kawamoto, 1993 ; Kawamoto, Kello, Jones, & Bame, 1998). Plaut (1999) implémente un modèle de traitement séquentiel dans le cadre du modèle en Triangle et insiste sur l'importance du traitement sériel pour mieux rendre compte des effets de longueur en lecture de mots. De plus, la présence d'un traitement sériel est suggérée par les résultats d'études sur les mouvements oculaires en lecture de mots isolés, qui montrent que le nombre de fixations oculaires augmente en fonction de la longueur du mot (Rayner, 1998, 2009).

Le traitement sériel en lecture est également motivé par la nature des processus de segmentation qui interviennent en lecture de pseudo-mots. Les modèles parallèles existants se limitent à la lecture des mots monosyllabiques et tous les modèles computationnels de lecture de mots polysyllabiques, à l'exception du modèle MTM, ont eu recours à un système de segmentation.

Dans les modèles double-voie, la segmentation graphémique, réalisée par le *parser* graphémique dans les modèles CDP+ et CDP++, n'intervient que dans la voie sous-lexicale. Les auteurs mentionnent que la segmentation graphémique pourrait se faire sous le contrôle de l'attention visuo-spatiale par déplacement de gauche à droite sur les lettres. Cependant, aucun de ces modèles ne détaille, ni ne modélise ou discute des principes théoriques de ce mécanisme d'attention, pas plus que de la segmentation résultante. En particulier, la segmentation graphémique est supposée capable de s'aligner et de trouver directement, sans erreur, les unités graphémiques. Cependant, celles-ci se définissent par la propriété des phonèmes qui correspondent à ces sous-séquences, puisqu'un graphème est une sous-séquence de lettres qui correspond à un phonème. Idéalement, il faut donc connaître les phonèmes résultants du traitement des sous-séquences pour segmenter correctement ces sous-séquences. Coltheart et al. (1993) décrivent un algorithme d'essai-erreur pour effectuer cette segmentation (augmenter l'ensemble de lettres jusqu'à ne plus

trouver de son unique qui corresponde). Cependant, cette stratégie ne repose sur aucune hypothèse théorique qui soit exposée et discutée. Les conséquences de ce mécanisme de segmentation sur les prédictions du modèle ne sont pas claires, notamment pour les *patterns* d'erreurs ou de difficultés de segmentation qui ne sont, à notre connaissance, jamais discutés.

Le modèle MTM inclut également un système de segmentation basé sur l'attention visuelle. Cette dernière intervient quelle que soit la nature du stimulus et pour les deux types de traitement (global parallèle ou analytique sériel). La segmentation, elle, peut peut être effectué pour tous types de stimuli mais elle est spécifique au traitement sériel. Étant donné que le modèle contient des traces de mots entiers et des traces de segments de mots, la nature du traitement dépend des unités que le modèle est capable de reconnaître dans un stimulus. Ainsi, la fenêtre attentionnelle va se centrer sur le segment reconnu. De plus, le traitement analytique proposé par MTM ne se base pas sur une unité sous-lexicale fondamentale définie *a priori* (graphème ou syllabe). Cependant, il est vrai que, dans l'implémentation du niveau lexical, les traces mnésiques réfèrent soit à des mots entiers, soit à des syllabes. Par conséquent, ce choix d'implémentation favorise arbitrairement un traitement analytique en syllabes.

4 Problématique

Dans cette revue, nous avons comparé les principaux modèles de lecture à haute voix, ce qui a permis de pointer des limites et des questions qui restent en suspend. Nous nous intéresserons particulièrement à trois questions qui sont au cœur des recherches menées dans ce travail.

La première question porte sur l'architecture du système et les connaissances mises en jeu lors de la lecture. Classiquement, on oppose les modèles double-voie aux modèles à voie unique. L'approche double-voie est l'approche dominante dans ce champ de recherche et celle qui a fait l'objet du plus grand nombre d'études. Ainsi, cette approche a permis de rendre compte d'un grand nombre d'effets expérimentaux dans le comportement sain et également dans le comportement pathologique, notamment les dyslexies développementales (voir Perry et al., 2019 ; Ziegler, Perry, & Zorzi, 2014 ; Ziegler et al., 2020). Les modèles à voie unique, quant à eux, ne sont pas aussi « performants » pour reproduire les effets expérimentaux mais ils se présentent comme étant plus parcimonieux, en termes d'hypothèses *ad hoc*, en comparaison aux modèles double-voie. Par ailleurs, la majorité des modèles présentés s'intéressent au problème central du passage de l'orthographe à la phonologie mais ne prennent pas en compte de manière détaillée certains aspects importants des traitements visuels tels que l'attention visuelle, entre autres. Ainsi, la plupart des modèles de lecture de mots isolés incluent un sous-modèle de reconnaissance de mots sous-spécifié et largement obsolète, puisqu'inspiré du modèle IA (McClelland & Rumelhart, 1981), alors même que d'autres modèles de reconnaissance de mots beaucoup plus détaillés et performants ont été proposés depuis les années 1980 (pour des revues, voir Phénix et al., 2016 ; Rayner & Reichle, 2010).

Notre première contribution, en réponse à cette première question, sera la définition d'un modèle de lecture incluant un sous-modèle sophistiqué de reconnaissance de mots. Ce modèle sera basé sur une architecture à voie unique. Nous montrerons que ce modèle est apte à rendre compte

des données expérimentales qui caractérisent à la fois la reconnaissance de mots et la lecture à haute voix, proposant ainsi une vision plus unifiée du vaste champ de recherche qui s'intéresse à l'étude des mécanismes cognitifs en jeu pendant le traitement visuel des mots.

La deuxième question concerne, plus spécifiquement, l'impact de l'attention visuelle en lecture de mots isolés. Parmi les modèles présentés, seul le modèle MTM (Ans et al., 1998) fournit une description formelle de l'attention visuelle et de son rôle en lecture à haute voix. Alors que les modèles double-voie mentionnent, sans l'implémenter, une implication de l'attention limitée à la voie sous-lexicale, le modèle MTM postule que l'attention intervient précocement et qu'elle participe à l'identification des lettres, ce qui est plus en accord avec les données comportementales et les théories attentionnelles actuelles. La plupart des modèles accordent un rôle central à l'attention dans les traitements sériels. Traditionnellement, dans le cadre des modèles double-voie, l'effet de longueur est dû à la lecture sérielle via la voie sous-lexicale. Ces modèles expliquent donc facilement les effets de longueur observés en lecture de pseudo-mots mais peinent à rendre compte des effets de longueur observés sur les mots en lecture à haute voix ou en décision lexicale. Dans le cadre du modèle MTM, qui est un modèle à voie unique, l'effet de longueur est considéré comme un indicateur du nombre de captures de l'attention visuelle lors du traitement de la séquence orthographique d'entrée. Les mots étant le plus souvent traités en mode global, donc en une seule fixation quelle que soit leur longueur, le modèle ne peut rendre compte des effets de longueur qui sont observés sur les mots dans les études expérimentales.

Notre deuxième contribution, en réponse à cette deuxième question, sera de montrer qu'un modèle à voie unique, incluant un mécanisme attentionnel détaillé (notamment, plus sophistiqué que dans le modèle MTM), est capable de rendre compte des effets de longueur pour les mots et ce, dans diverses tâches expérimentales.

La troisième question concerne le rôle de l'attention visuelle lors du traitement sériel de la séquence orthographique, donc essentiellement en contexte de lecture de pseudo-mots. Les modèles double-voie s'accordent sur un traitement sériel (de gauche à droite ou par segments) des pseudo-mots. Ce traitement qui porte sur des unités orthographiques pré-spécifiées (les graphèmes) nécessite une segmentation graphémique pour que chaque graphème puisse ensuite activer le phonème qui lui correspond (mécanisme ou réseau de conversion graphème-phonème spécifique à la voie sous lexicale). Ce mécanisme de segmentation est peu détaillé dans les modèles existants, bien que les chercheurs du domaine s'accordent sur l'implication de l'attention visuelle dans cette segmentation. Une implémentation des mécanismes attentionnels en jeu est proposée dans le modèle MTM. Ils permettent une segmentation essentiellement syllabique même s'ils offrent un peu de flexibilité quant à la taille des sous-unités traitées (graphèmes ou syllabes). En revanche, la distribution homogène et uniforme de l'attention visuelle sur la séquence de lettres, au sein du modèle MTM, ne correspond pas aux conceptions actuelles de l'attention. Celle-ci est plutôt décrite en utilisant la métaphore du « *spotlight* », c'est-à-dire une distribution de l'attention qui s'atténue selon un gradient à partir du point focal (Evans et al., 2011).

Notre troisième contribution, en réponse à cette troisième et dernière question, sera la défini-

tion d'un modèle de lecture à haute-voix des mots et pseudo-mots, reposant sur un mécanisme de traitement de segments basé sur l'attention. Nous montrerons sa capacité à prendre en compte des unités de taille variable selon les caractéristiques du stimulus d'entrée et la nature des connaissances lexicales.

Récapitulatif :

- **Une architecture en une seule voie ou à deux voies ?**
  - Les modèles double-voie supposent des connaissances et des traitements de natures différentes dans chaque voie. La voie lexicale décrit le passage de l'orthographe à la phonologie pour les mots et la voie sous-lexicale décrit le passage des graphèmes aux phonèmes. La première voie est nécessaire pour lire les mots irréguliers tandis que la seconde voie est nécessaire pour lire les pseudo-mots.
  - Les modèles à voie unique supposent une seule voie de traitement. Ainsi, la lecture des mots et des pseudo-mots ne fait pas intervenir des connaissances et des traitements de natures différentes.
  - A l'exception du niveau des lettres, du niveau du mot et du niveau des phonèmes, il n'existe pas de consensus sur les sous-unités orthographiques et phonologiques nécessaires dans la lecture (graphèmes, syllabes ou demi-syllabes).
  - L'attention visuelle intervient dans la lecture et plus spécifiquement dans la segmentation du stimulus. Dans les modèles double-voie, celle-ci est propre à la voie sous-lexicale, tandis que dans certains modèles à voie unique elle intervient dès la perception des lettres quel que soit le stimulus à traiter.
- **Effet de longueur en lecture**
  - L'effet de longueur en lecture est associé au fait que les stimuli longs sont traités plus lentement que les stimuli courts.
  - L'effet de longueur est plus prononcé en lecture de pseudo-mots qu'en lecture de mots.
  - Les modèles double-voie l'interprètent comme une conséquence des traitement sériel spécifique à la lecture de pseudo-mots et donc à la voie sous-lexicale.
  - En dépit du fait qu'ils rendent compte d'un nombre importants d'effets comportementaux, les modèles double-voie ne reproduisent pas l'effet de longueur pour les mots (ni en lecture, ni en décision lexicale).
- **Traitements parallèle et sériel dans la lecture**
  - Les mots sont lus de façon globale et parallèle ; tandis que les pseudo-mots sont lus de façon analytique et sérielle.
  - Dans les modèles double-voie, et par définition, la voie lexicale est parallèle et la voie sous-lexicale est sérielle.
  - Le traitement sériel est communément opéré à travers une fenêtre « attentionnelle ».

Questions de recherche :

→ Quelles sont les limites d'une architecture à voie unique incluant un module de reconnaissance de mots sophistiqué ?

→ Quel est l'impact de l'attention visuelle en lecture de mots isolés ? Et comment permet-elle de rendre compte de l'effet de longueur en lecture de mots ?

→ Quel est le rôle de l'attention visuelle lors du traitement sériel de la séquence orthographique, notamment en contexte de lecture de pseudo-mots ?

CHAPITRE

2 Le modèle BRAID-Phon

"Probability is not really about numbers; it is about the structure of reasoning"

— GLENN SHAFER (CITÉ DANS PEARL, 1988, P. 77)

Dans le Chapitre 1, nous avons fait une revue des modèles de lecture existants. Cela nous a permis de mettre en évidence certaines limites et questions théoriques en suspens que nous allons aborder dans cette thèse. Dans ce chapitre, nous allons présenter le cadre théorique dans lequel nous nous situons, et le modèle computationnel correspondant qui constitue la contribution principale de cette thèse : le modèle BRAID-Phon. BRAID-Phon s'inscrit dans la lignée du modèle BRAID (BRAID signifiant « Bayesian model of word Recognition with Attention, Interference and Dynamics »), développé dans le cadre de la thèse de Thierry Phénix (2018).

Le modèle BRAID original est un modèle de reconnaissance de mots; à ce titre il ne contient pas de composant phonologique. Ce modèle a permis de simuler plusieurs tâches telles que la reconnaissance de lettres, la reconnaissance de mots et la décision lexicale. Il a également permis de reproduire fidèlement une grande variété de données comportementales observées dans ces différentes tâches. BRAID a déjà fait l'objet de plusieurs extensions.

Une première extension du modèle, nommée « BRAID-Learn », a été proposée par Emilie Ginestet (2019) afin de doter le modèle de capacités d'apprentissage orthographique et de simuler l'évolution des mouvements oculaires, en lien avec l'acquisition de nouvelles connaissances orthographiques. « BRAID-Phon » constitue une seconde extension, essentiellement indépendante de l'extension BRAID-Learn, dans laquelle nous allons développer un modèle de lecture à haute voix, en ajoutant un sous-modèle phonologique au modèle BRAID. Plus généralement, le modèle BRAID et ses extensions sont définis dans le cadre de la modélisation bayésienne algorithmique (Diard, 2015).

Dans ce chapitre, nous décrirons le modèle BRAID dans ses grandes lignes. Ensuite, nous dé-

taillerons l'extension BRAID-Phon que nous avons développée. Celle-ci se résume en une mise à jour des traitements sensoriels et un ajout de connaissances et de traitements phonologiques. Enfin, nous illustrerons comment ce modèle « lit » les stimuli qui lui sont présentés.

1 Travaux antérieurs

1.1 Le modèle BRAID

BRAID est un modèle probabiliste et hiérarchique des connaissances et représentations visuelles, attentionnelles et lexicales mises en jeu dans le traitement visuel des mots. La description complète du modèle BRAID est fournie dans les travaux de thèse de Thierry Phénix (Phénix, 2018 ; Phénix, Ginestet, Valdois, & Diard, 2020). Dans les travaux de thèse d'Emilie Ginestet (2019) est fournie la description de l'extension de ce modèle à l'apprentissage orthographique, c'est-à-dire le modèle BRAID-Learn. L'ensemble de ces modèles sont définis dans le cadre de la modélisation bayésienne algorithmique (Diard, 2015). Cette dernière permet de définir des modèles probabilistes structurés et hiérarchisés d'agents cognitifs et elle est héritée directement de la méthodologie de « Programmation Bayésienne » (Bessière, Laugier, & Siegwart, 2008 ; Bessière, Mazer, Ahuactzin, & Mekhnacha, 2013 ; Lebeltel et al., 2004).

Dans cette thèse, nous ne pourrons pas décrire de manière extensive le formalisme et son pouvoir expressif. L'annexe A donnera un descriptif de la méthodologie de la modélisation bayésienne algorithmique.

1.2 Description Conceptuelle

L'objectif de cette sous-partie est de décrire le fonctionnement général du modèle BRAID qui constitue le sous-modèle de reconnaissance de mots de BRAID-Phon.

BRAID est un modèle computationnel, bayésien, dynamique et hiérarchique de la reconnaissance visuelle des mots. En effet, ce modèle est :

- computationnel, dans le sens où il décrit mathématiquement le traitement visuel des mots et permet d'effectuer des simulations informatiques ;
- bayésien, dans le sens « bayésien-subjectiviste », c'est-à-dire qu'on considère que les probabilités représentent des états de connaissance d'un sujet, dans notre cas un sujet qui traite visuellement des mots ;
- dynamique parce qu'il permet de simuler l'évolution temporelle des représentations au fil du traitement ;
- hiérarchique car il suppose plusieurs niveaux de représentations et leurs interactions par échange d'informations, lors du traitement.

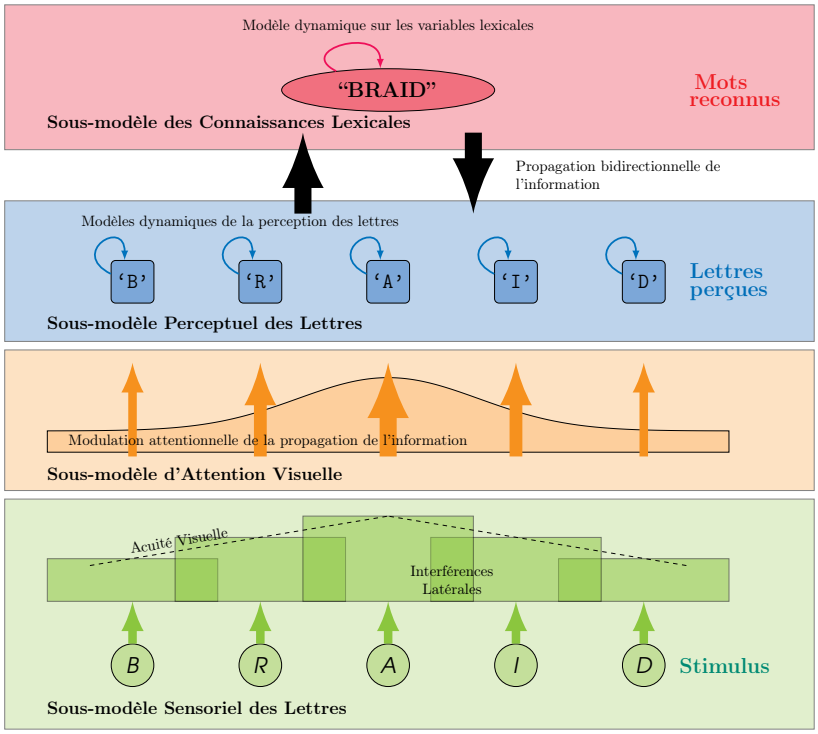


FIGURE 2.1 – **Schéma conceptuel du modèle BRAID.**

BRAID se compose de plusieurs sous-modèles, dont chacun décrit un type de connaissance particulier, impliqués dans la reconnaissance visuelle des mots (voir Figure 2.1). Le premier niveau de traitement, **le niveau sensoriel**, est celui par où l'information sensorielle est introduite, et son rôle est de rendre compte de l'extraction de l'information sensorielle du stimulus orthographique. Cela requiert un ensemble de mécanismes de traitements visuels dit de « bas-niveau » qui sont représentés dans BRAID par une matrice de confusions entre lettres visuellement similaires, un gradient d'acuité et un mécanisme d'interférences latérales entre lettres adjacentes. Ce niveau de traitement modélise également la position de l'œil dans la séquence de lettres du mot.

Le second niveau est **le niveau perceptif des lettres** qui permet l'accumulation et le maintien de l'information sur l'identité des lettres au cours du temps. Pour chaque lettre du stimulus, un processus d'accumulation de l'information sensorielle est réalisé de manière parallèle sur toutes les positions orthographiques. **Le niveau visuo-attentionnel** implémente un mécanisme d'attention qui module la propagation de l'information du niveau sensoriel vers le niveau perceptif des lettres. Ainsi, une région d'intérêt, c'est-à-dire la portion de la scène visuelle sur laquelle l'attention visuelle est focalisée, est traitée plus efficacement au détriment du reste.

Dans le troisième niveau, **le niveau des connaissances lexicales**, sont représentées les connaissances orthographiques des mots connus et donc les lettres qui composent chaque mot. Ce niveau permet également d'accumuler de l'information perceptive au cours du temps sur l'identité du mot correspondant au stimulus visuel d'entrée. L'information circule dans le modèle du niveau sensoriel vers les niveaux hiérarchiquement supérieurs, soit le niveau perceptif des lettres et le niveau lexical. De plus, le flux d'information entre ces deux derniers niveaux est bidirection-

nel. Cela signifie qu'au cours du traitement, les connaissances lexicales orthographiques peuvent aussi influencer le traitement au niveau perceptif des lettres. Enfin, le niveau lexical contient des connaissances et mécanismes pour évaluer l'appartenance lexicale du stimulus, pour évaluer si le stimulus est un mot appartenant au lexique connu, ou non. Cela permet de simuler la tâche de décision lexicale, mais aussi de détecter la nouveauté, dans un contexte d'apprentissage orthographique, comme dans le modèle BRAID-Learn d'Emilie Ginestet (2019).

En somme, BRAID fait l'hypothèse classique de l'existence de trois niveaux de traitement orthographique des mots (sensoriel, perceptif et lexical), qu'on retrouve dans la plupart des modèles de reconnaissance visuelle de mots, en commençant par les travaux de McClelland et Rumelhart (1981) avec le modèle IA. Notons que le traitement des lettres est effectué de manière parallèle sur l'ensemble des positions, et que BRAID adopte un codage positionnel distribué, grâce au mécanisme d'interférences latérales entre les lettres. La principale originalité de BRAID, qui le distingue des modèles précédents, est l'inclusion d'un niveau d'attention visuelle, et une description détaillée des mécanismes visuels de bas niveau (acuité et interférences latérales).

1.3 Description Mathématique

Nous avons exprimé conceptuellement l'architecture et les niveaux de traitement que contient le modèle BRAID. Dans cette partie, nous allons décrire l'implémentation mathématique de ce modèle. La Figure 2.2 représente graphiquement le modèle probabiliste qui traduit le modèle conceptuel de la Figure 2.1 : le graphe illustre la structure de dépendance probabiliste du modèle, avec un nœud du graphe pour chaque variable du modèle, et un arc entre deux nœuds pour relier les variables qui ont une relation de dépendance probabiliste. Chaque niveau est implémenté par un modèle probabiliste qui sera décrit séparément. Nous ne fournissons pas ici une description exhaustive des définitions mathématiques des sous-modèles, et ne donnons que quelques éléments saillants ou utiles pour la compréhension de la suite, et de notre extension BRAID-Phon. La définition mathématique complète du modèle BRAID se trouve dans la thèse de Thierry Phénix (2018).

Le sous-modèle sensoriel des lettres

Ce sous-modèle implémente le niveau sensoriel qui rend compte du processus d'extraction des attributs des lettres du stimulus. Pour cela, ce sous-modèle contient les variables S_n^t qui décrivent le stimulus et les variables I_n^t qui décrivent le résultat du processus complet d'identification d'une lettre. On note que l'indice supérieur t indique les pas de temps du modèle (itérations) et l'indice n la position dans le stimulus. Dans la suite, nous utilisons la notation $X_{1:N}^{1:T}$ pour référer à l'ensemble des variables X_n^t pour les valeurs de temps t comprises entre 1 et T, et les valeurs de position n comprises entre 1 et la position maximale N, qui correspond à la longueur du stimulus orthographique.

Chaque variable de ces deux groupes de variables, $S_{1:N}^{1:T}$ et $I_{1:N}^{1:T}$, est définie dans l'espace discret qui regroupe les 26 lettres de l'alphabet et un caractère supplémentaire de fin de chaîne. Nous notons cet espace $\mathscr{D}_L$. La variable G^t représente la position du regard et les variables $\Delta I_{1:N}^t$ re-

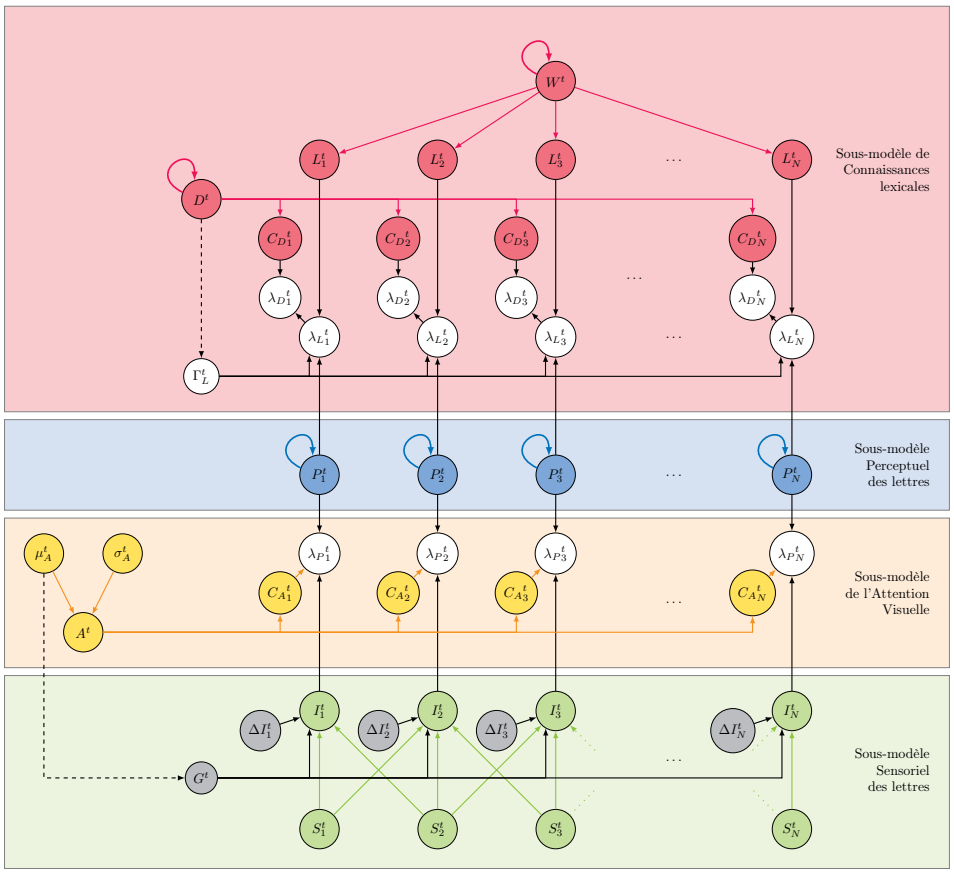


FIGURE 2.2 – **Représentation graphique du modèle BRAID.** Les nœuds représentent les variables du modèle et les flèches qui les relient représentent les dépendances décrites formellement dans la décomposition de la distribution de probabilité conjointe. Les flèches en boucle représentent les dépendances temporelles : une variable à l'instant T , qui est représentée par un nœud avec une flèche en boucle, dépend d'elle même à l'instant $T-1$. La flèche en pointillé représente une affectation de valeur de la variable à l'origine de la flèche et transmise à la variable à la fin de la flèche.

présentent le décalage entre une lettre et ses voisines (variable utile pour le mécanisme d'interférences latérales).

Dans ce sous-modèle, la représentation interne d'une lettre est déterminée à partir de l'identité des lettres du stimulus, de la position du regard et de l'identité des lettres adjacentes dans chaque position. Mathématiquement, la distribution de probabilité conditionnelle $P(I^t_{1:N} \mid S^t_{1:N}\, G^t\, \Delta I^t_{1:N})$ définie pour un pas de temps donné t permet d'exprimer comment les variables $I^t_{1:N}$ dépendent à la fois des variables $S^t_{1:N}$, G^t et $\Delta I^t_{1:N}$.

Maintenant, nous allons décrire en plusieurs étapes ce qu'implémente cette distribution en se focalisant à chaque étape sur l'influence d'un type de variables. Lors de l'identification d'une lettre isolée, la relation entre la variable qui la représente S^t_n et la représentation interne associée I^t_n peut être décrite à l'aide de la distribution de probabilité $P(I^t_n \mid S^t_n)$. On peut représenter cette distribution sous forme d'une matrice de valeurs de probabilité telle que chaque élément de cette matrice $p_{i,s}$ correspond à $P([I^t_n = i] \mid [S^t_n = s])$, c'est-à-dire la probabilité de reconnaître la lettre i si s est présentée dans une position n. La matrice des valeurs $p_{i,s}$ est facilement identifiée à une matrice

de confusion expérimentale, telle qu'on en obtient en observant les performances typiques d'un lecteur expert en reconnaissance de lettres isolées. BRAID utilise une version légèrement adaptée de la matrice de confusion proposée par Townsend (1971).

Le second élément de ce sous-modèle est la position du regard G^t qui permet de modéliser l'effet de l'acuité visuelle. On suppose ici que le traitement d'une lettre du stimulus S_n^t est pénalisé par l'excentricité, c'est-à-dire la distance entre la lettre et la position du regard (mathématiquement $|g^t - n|$). Plus l'excentricité est grande, plus on « dégrade » la matrice de confusion de référence.

Le troisième et dernier facteur qui intervient dans ce sous-modèle lors du traitement d'une lettre cible à l'intérieur d'une séquence est l'identité des lettres adjacentes. La représentation interne des lettres à une position I_n^t ne dépend pas que de la lettre du stimulus à cette position S_n^t mais elle dépend également des lettres voisines S_{n-1}^t et S_{n+1}^t, produisant ainsi un mécanisme d'interférence latérale entre les lettres (permettant de simuler les effets de *crowding*). Ainsi, la matrice de confusion de référence, modifiée par l'acuité, est « mélangée » avec les matrices de confusion des lettres adjacentes : c'est cette forme finale qui définit le terme $P(I_{1:N}^t \mid S_{1:N}^t\ G^t\ \Delta I_{1:N}^t)$.

La structure de dépendance du sous-modèle sensoriel est définie dans le cadre de la modélisation que nous suivons sous forme d'une décomposition de sa distribution conjointe. Mathématiquement, on écrit :

$$P(S_{1:N}^{1:T}\ G^{1:T}\ \Delta I_{1:N}^{1:T}\ I_{1:N}^{1:T}) = \prod_{t=1}^{T}\left[P(G^t)\ \prod_{n=1}^{N}\left[P(S_n^t)\ P(\Delta I_n^t)\ P(I_n^t \mid S_n^t\ G^t\ \Delta I_n^t)\right]\right].$$

Le sous-modèle visuo-attentionnel

Le rôle du sous-modèle visuo-attentionnel est d'agir comme un « filtre » attentionnel qui module spatialement le transfert de l'information sensorielle aux niveaux de traitement supérieurs. Pour implémenter cela, on utilise des variables de « cohérence » $\lambda_{p_{1:N}^{1:T}}$, qui assurent la jonction et le transfert d'information entre le sous-modèle sensoriel et le niveau perceptif des lettres (Gilet, Diard, & Bessière, 2011) ; ces variables de cohérence sont commandées par des variables de « contrôle » $C_{A_{1:N}^{1:T}}$, qui permettent de moduler la quantité d'information transmise du niveau sensoriel vers le niveau perceptif (Phénix, 2018).

Les variables de contrôle dépendent, à leur tour, d'une distribution spatiale de l'attention visuelle représentée par une distribution de probabilité sur la variable A^t. Cette distribution de probabilité est supposée gaussienne, et donc les valeurs de probabilité sur les positions spatiales dépendent d'un paramètre de position, le « foyer attentionnel » (la moyenne μ_A^t de la gaussienne) et d'un paramètre de dispersion (l'écart-type σ_A^t de la gaussienne). La connaissance de l'identité des lettres provenant du sous modèle sensoriel est transmise partiellement au modèle perceptif, et la quantité d'information transmise est proportionnelle aux valeurs de la distribution $P(A^t \mid \mu_A^t\ \sigma_A^t)$. Ainsi le modèle perceptif acquiert de l'information plus rapidement autour du foyer attentionnel μ_A^t, au détriment des autres parties du stimulus.

Mathématiquement, on écrit :

$$P({A^{1:T}}\ {\mu_A}^{1:T}\ {\sigma_A}^{1:T}\ {C_A}_{1:N}^{1:T}) = \prod_{t=1}^{T}\left[P({\mu_A}^t)\ P({\sigma_A}^t)\ P(A^t \mid {\mu_A}^t\ {\sigma_A}^t)\ \prod_{n=1}^{N} P({C_A}_n^t \mid A^t)\right].$$

Le sous-modèle perceptif des lettres

Le sous-modèle perceptif des lettres implémente un processus d'accumulation de l'information sensorielle, indépendamment et en parallèle sur chaque position spatiale. Dans ce sous-modèle, les variables P_n^t réfèrent, à chaque pas de temps t et chaque position n, aux représentations internes des lettres qui résultent de l'accumulation de l'information sensorielle. Ces variables sont définies également sur l'espace des lettres possibles $\mathscr{D}_L$.

Pour chaque position n, la distribution a priori $P(P_n^0)$ définit l'état des connaissances « au repos » sur les lettres, c'est-à-dire avant la stimulation ou en absence prolongée d'information sensorielle. Chacune de ces distributions est mathématiquement définie par une distribution uniforme sur les lettres.

Le caractère dynamique du sous-modèle provient de la dépendance entre les variables $P_{1:N}^{0:T}$ entre deux pas de temps. Ainsi, l'information cumulée à un instant t, formulée mathématiquement par une distribution de probabilité sur la variable $P_{1:N}^t$, est partiellement maintenue du pas de temps t au pas de temps suivant $t+1$, puis elle est modifiée par l'information sensorielle au pas de temps suivant $t+1$. En revanche, en l'absence d'information sensorielle, une perte graduelle de l'information se produit jusqu'à atteindre l'état de repos. Mathématiquement, ces dépendances temporelles des percepts, et par conséquent leur évolution temporelle, sont modélisées à travers la distribution de probabilité conditionnelle suivante :

$$P([P_n^t = p] \mid [P_n^{t-1} = l]) = \begin{cases} \frac{1+Leak_P}{1+|\mathscr{D}_L|\ Leak_P} & \text{si } p^t = p^{t-1} \\ \frac{Leak_P}{1+|\mathscr{D}_L|\ Leak_P} & \text{sinon,} \end{cases}$$

avec $Leak_P$ un paramètre qui détermine la vitesse de déclin.

La distribution conjointe qui définit la structure de dépendance du sous-modèle perceptif des lettres est :

$$P(P_{1:N}^{0:T}) = \prod_{n=1}^{N}\left[P(P_n^0)\ \prod_{t=1}^{T} P(P_n^t \mid P_n^{t-1})\right].$$

Le sous-modèle des connaissances lexicales

L'objectif de ce sous-modèle est de représenter l'orthographe des mots connus, et donc les lettres qui les composent, et d'accumuler l'information sur l'identité du mot contenu dans le stimulus au cours du temps. La composante principale du sous-modèle des connaissances lexicales permet de représenter les mots comme des distributions de probabilité sur des séquences de lettres en utilisant une structure sous forme d'un modèle de fusion naïve de Bayes (pour des propositions similaires, voir McClelland, 2013 ; Norris, 2006).

Pour cela, les variables $W^{0:T}$ sont définies sur un espace de l'ensemble des mots du lexique que nous notons $\mathscr{D}_W$ et les variables $L_{1:N}^{1:T}$ sont définies sur l'espace des lettres $\mathscr{D}_L$. Ainsi, à chaque pas

de temps t, la distribution de probabilité conditionnelle $P(L_n^t \mid [W^t = w])$ est un « quasi-Dirac » qui encode l'identité de la lettre à la position n pour le mot du lexique w : $P(L_n^t \mid [W^t = w])$ est une distribution de probabilité discrète avec une valeur de probabilité élevée pour la lettre à cette position, et des probabilités résiduelles, non-nulles, pour les autres lettres. Mathématiquement, on écrit :

$$P([L_i^t = l] \mid [W^t = w]) = \begin{cases} 1 - (|\mathscr{D}_L| - 1)\,\epsilon & \text{si } l \text{ est la } i\text{ème lettre du mot } w, \\ \epsilon & \text{sinon.} \end{cases}$$

$P(W^0)$ est une distribution a priori qui indique les probabilités des mots à l'état initial. Pour chaque mot du lexique w_i de l'espace $\mathscr{D}_W$, $P([W^0 = w_i])$ (noté par la suite p_{w_i}) correspond à la fréquence de ce mot dans le lexique. Enfin, nous supposons également dans ce sous-modèle, de manière analogue au sous-modèle perceptif des lettres, des dépendances temporelles entre chaque paire de variables W^t et W^{t+1} afin de modéliser l'accumulation d'information sur les mots au cours du temps. Ces dépendances sont conçues de sorte à converger vers $P(W^0)$ en cas d'absence prolongée de stimulus. Mathématiquement, on note :

$$P([W^{t+1} = w^{t+1}] \mid [W^t = w^t]) = \begin{cases} \frac{1+p_{w_i}\ Leak_w}{1+Leak_w} & \text{si } w^t = w^{t-1} \\ \frac{p_{w_i}\ Leak_w}{1+Leak_w} & \text{sinon.} \end{cases}$$

La distribution conjointe de cette première partie du sous-modèle des connaissances lexicales est :

$$P(W^{0:T}\ L_{1:N}^{1:T}) = P(W^0) \prod_{t=1}^{T} \left[P(W^t \mid W^{t-1}) \prod_{n=1}^{N} P(L_n^t \mid W^t) \right].$$

Le sous-modèle des connaissances lexicales contient également des connaissances permettant d'évaluer si une séquence de lettres perçues correspond à un mot connu ou non. Traditionnellement, les modèles de reconnaissance de mots implémentent un mécanisme de ce type, pour simuler la tâche de décision lexicale, et générer une réponse « mot » quand le système de reconnaissance de mots identifie un mot à partir du stimulus et une réponse « non-mot » si aucun mot n'est reconnu dans un délai fixé à l'avance (Coltheart et al., 2001 ; Grainger & Jacobs, 1996).

Dans le modèle BRAID, l'évaluation de l'appartenance lexicale est implémentée de manière plus générale, sans référer spécifiquement à un processus décisionnel spécifique de la tâche de décision lexicale. En effet, le sous-modèle d'appartenance lexicale permet de réaliser une comparaison probabiliste entre l'information du niveau perceptif des lettres (identité des lettres perçues) et celle provenant du niveau lexical (les lettres du mot reconnu). Ainsi, si ces deux informations sont concordantes, une réponse « mot » sera plus probable et, au contraire, si ces informations sont non-concordantes, c'est la réponse « non-mot » qui sera plus probable. Pour définir cette mesure de « concordance », on se base sur l'observation des variables de cohérence entre le sous-modèle perceptif des lettres et le sous-modèle lexical. Ainsi, si un mot du lexique permet de prédire parfaitement les lettres perçues à un instant t, la probabilité que toutes les variables de cohérences soient à 1 est élevée ; en revanche, si le stimulus ne correspond à aucun mot connu, il y a au moins une variable de cohérence dont la probabilité d'être à 1 est faible. Mathématiquement, des patterns de variables de contrôle $C_{D\,1:N}^{1:T}$ représentent ces deux situations, et l'évaluation de la

probabilité de ces patterns informe, à chaque pas de temps, sur la probabilité d'appartenance du stimulus au lexique.

Dans ce sous-modèle, des variables booléennes $D^{0:T}$ représentent l'appartenance lexicale de sorte que la valeur VRAI indique une réponse « mot » et la valeur FAUX une réponse « non-mot ». Ce modèle est également un modèle dynamique possédant, d'une part, un état initial défini par une distribution a priori $P(D^0)$, définie par une distribution de probabilité uniforme et, d'autre part, un terme dynamique défini par la distribution :

$$P([D^t = d^t] \mid [D^{t-1} = d^{t-1}]) = \begin{cases} 1 - Leak_D & \text{si } d^t = d^{t-1} \\ Leak_D & \text{sinon,} \end{cases}$$

avec $Leak_D$ un paramètre qui indique la vitesse de déclin vers l'état initial en cas d'absence du stimulus.

Connexion des sous-modèles

Les sous-modèles de BRAID étant définis, il reste à les connecter pour obtenir le modèle dans son ensemble. Le choix de ces connexions résulte en une architecture globale du modèle probabiliste consistante avec la description conceptuelle du modèle : le sous-modèle sensoriel transmet de l'information au sous-modèle perceptif via le sous-modèle attentionnel, et le sous-modèle perceptif est en contact avec le sous-modèle lexical.

Pour définir ces connexions entre les sous-modèles dans le formalisme probabiliste, nous utilisons des variables probabilistes particulières : des variables de cohérence (Gilet et al., 2011) et des variables de cohérence contrôlées (Phénix, 2018). Elles sont au cœur du sous-modèle attentionnel, et nous les avons déjà mentionnées à ce sujet. Dans ce chapitre, nous n'allons pas décrire leur définition mathématique, et nous soulignons seulement leur usage.

Ainsi, les variables de cohérence peuvent être interprétées comme des « interrupteurs probabilistes », permettant de choisir les portions du modèle dans lesquelles l'information se propage. Avec les variables de cohérence, soit l'information se propage, soit elle ne se propage pas du tout : l'interrupteur permet un choix « tout-ou-rien ». Les variables de cohérence contrôlées, elles, permettent d'ajouter une graduation dans le transfert d'information. Ainsi, dans ces structures, au lieu d'être « ouvert » ou « fermé », l'interrupteur est décrit par une distribution de probabilité entre ces deux états ; ainsi, l'information passe plus ou moins, et le système devient analogue à une « résistance probabiliste variable » (un « potentiomètre probabiliste ») plutôt qu'un interrupteur.

Nous avons vu que le sous-modèle attentionnel repose essentiellement sur cette modulation de la quantité d'information sensorielle transmise au modèle perceptif, qui est différente selon les positions des lettres et la position (et la dispersion) du focus attentionnel. Une couche de variable de cohérence connecte les sous-modèles perceptif et lexical. Cependant, cette couche est double, ce qui permet de rendre asymétrique le transfert remontant (bottom-up) et descendant (top-down) de cet échange bi-directionnel d'information[9]. Ainsi, le transfert d'information re-

[9] Notons que cette asymétrie a été développée dans le modèle BRAID-Learn de la thèse d'Emilie Ginestet (2019), et n'est donc pas présente dans le modèle BRAID dans sa version initiale (Phénix, 2018).

montant est constant, alors que le transfert d'information descendant est modulé par la variable d'appartenance lexicale : lorsque le stimulus est détecté comme un mot connu, les influences lexicales descendantes sont fortes ; en revanche, lorsque le stimulus est détecté comme un mot inconnu, les influences lexicales descendantes sont supprimées.

1.4 Simulation des tâches cognitives

Dans le cadre de la programmation bayésienne, une fois que les connaissances sont entièrement spécifiées dans notre modèle, il est possible de définir des *questions probabilistes* qui permettent de formuler toutes les distributions de probabilité conditionnelles d'intérêt, et de les calculer à l'aide de l'inférence bayésienne. Ainsi, avec le modèle BRAID défini mathématiquement, nous pouvons simuler des tâches cognitives en choisissant des distributions de probabilité particulières à calculer. Cela permet d'obtenir une distribution de probabilité, et son évolution au cours du temps, qui simule les processus dynamiques d'accumulation d'information dans des tâches perceptives. Pour comparer le comportement du modèle, par exemple à des temps de réponse observés, nous ajoutons couramment à nos calculs probabilistes perceptifs un modèle de processus de prise de décision. Dans l'essentiel de nos simulations, ce processus de prise de décision consiste simplement à sélectionner une réponse lorsqu'une valeur de probabilité atteint un seuil.

Par exemple, la première tâche cognitive que nous envisageons est la reconnaissance d'une lettre isolée (dans une position n). Elle est modélisée par la question probabiliste suivante :

$$Q_{P_n}^{T} = P\big(P_n^T \mid s_n^{1:T}\ g^{1:T}\ \mu_A^{1:T}\ {\sigma_A}^{1:T}\ [{\lambda_P}_n^{1:T} = 1]\big),$$

avec $s_n^{1:T}$ qui indique le stimulus orthographique, $\mu_A^{1:T}$ et ${\sigma_A}^{1:T}$ les paramètres de la distribution attentionnelle et $g^{1:T}$ la position du regard. La spécification des variables ${\lambda_P}_{1:N}^{1:T}$, ici en leur donnant la valeur 1 tout au long de la simulation, permet de connecter les sous-modèles impliqués et de propager l'information entre le niveau sensoriel et le niveau perceptif des lettres. Nous ne détaillons pas ici les calculs à mettre en œuvre pour « répondre à la question » de la reconnaissance visuelle d'une lettre ; ils sont décrits par ailleurs (Phénix, 2018). Le calcul de cette distribution de probabilité revient à évaluer, à chaque itération, les probabilités des lettres sachant le stimulus et les paramètres visuo-attentionnels. Pour obtenir l'identité de la lettre perçue et un temps de réaction simulé, nous considérons les distributions de probabilité au cours du temps, et trouvons le premier instant où la probabilité d'une lettre atteint un seuil de décision.

Une variante de cette question permet d'évaluer la perception de lettres en supposant une influence du sous-modèle supérieur (le sous-modèle lexical). En effet, dans le cas d'une lettre isolée, nous supposons que seule l'information des niveaux inférieurs compte et seule l'identité de la lettre à la position d'intérêt est prise en compte dans le calcul. En revanche, dans le cas où la lettre est présentée au sein d'une séquence, nous pouvons considérer les influences lexicales en propageant l'information perceptive vers et depuis le niveau lexical. Mathématiquement, cela repose sur les variables de cohérence ${\lambda_L}_{1:N}^{1:T}$, et la question probabiliste devient :

$${Q'}_{P_n}^{T} = P\big(P_n^T \mid s_{1:N}^{1:T}\ g^{1:T}\ \mu_A^{1:T}\ {\sigma_A}^{1:T}\ [{\lambda_P}_{1:N}^{1:T} = 1]\ [{\lambda_L}_{1:N}^{1:T} = 1]\big)\,.$$

D'une manière similaire, la seconde tâche cognitive que BRAID simule est la reconnaissance de mots. Dans notre cadre probabiliste, cela revient à calculer à chaque itération t la distribution de probabilité sur les mots sachant l'information sensorielle cumulée au cours du traitement. Il s'agit d'une distribution de probabilité conditionnelle, que nous notons :

$$Q_W^T = P\left(W^T \mid s_{1:N}^{1:T}\, g^{1:T}\, \mu_A^{1:T}\, \sigma_A^{1:T}\, [\lambda_P{}_{1:N}^{1:T} = 1]\, [\lambda_L{}_{1:N}^{1:T} = 1]\right).$$

Enfin, la troisième et dernière tâche cognitive simulée est celle de la décision lexicale, dans laquelle un stimulus est présenté et on doit décider s'il correspond à un mot connu ou non. Dans le modèle, on calcule la distribution de probabilité sur la variable d'appartenance lexicale D^t. Mathématiquement, nous notons :

$$Q_{DL}^T = P\left(D^T \mid s_{1:N}^{1:T}\, g^{1:T}\, \mu_A^{1:T}\, \sigma_A^{1:T}\, [\lambda_P{}_{1:N}^{1:T} = 1]\, [\lambda_D{}_{1:N}^{1:T} = 1]\right).$$

2 BRAID-Phon

BRAID-Phon est une extension du modèle BRAID qui inclut des connaissances phonologiques, à la fois dans le sous-modèle lexical, mais aussi avec un sous-modèle phonologique dédié. Dans cette section, nous présentons les ajustements qui permettent d'étendre BRAID vers BRAID-Phon :

- la prise en compte des caractères accentués,
- l'ajout de connaissances phonologiques dans le sous-modèle lexical,
- et l'ajout d'un sous-modèle phonologique perceptif.

2.1 Prise en compte des caractères accentués

Contexte

Dans la version initiale du modèle, seules des lettres non-accentuées étaient représentées. Deux arguments permettaient de justifier ce choix initial. Premièrement, la grande majorité des études sur la confusion ou la similarité des lettres sont effectuées en langue anglaise avec des participants anglais, et donc en se basant strictement sur les 26 lettres de l'alphabet latin (Mueller & Weidemann, 2012). Deuxièmement, les études utilisées pour calibrer et évaluer BRAID n'ont utilisé aucun stimulus avec des caractères accentués : soit il s'agissait d'expériences configurées en anglais, soit en français mais avec des stimuli en majuscule, et donc non-accentués. Il a donc suffi d'utiliser une matrice de confusion des lettres non-accentuées, adaptée de l'expérience de Townsend (1971) (voir Figure 2.3, droite), pour rendre compte des effets comportementaux en reconnaissance de mots en anglais et en français dans ces expériences initiales.

Cependant, ces choix simplificateurs ne sont plus tenables en contexte de lecture à haute voix. En effet, il existe un nombre important de mots qui contiennent des caractères accentués en français et la présence ou non d'accents modifie parfois la prononciation, la fonction grammaticale ou

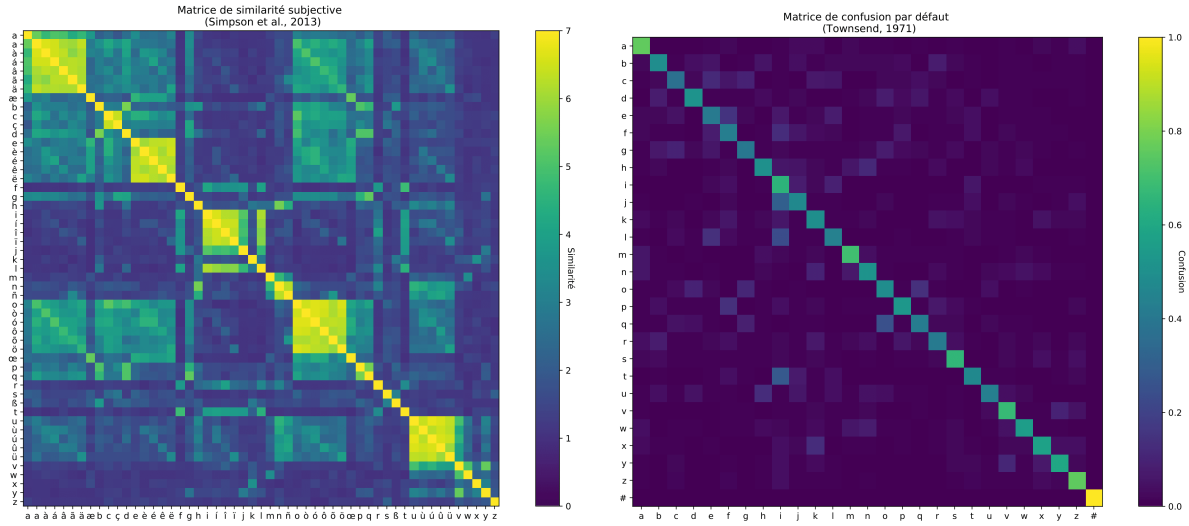


FIGURE 2.3 – **Comparaison entre les matrices de similarité et de confusion.** Matrice de similarité mesurée par Simpson et al. (2012) (à gauche). Matrice de confusion par défaut de BRAID adaptée de celle proposée par Townsend (1971) (à droite).

même le sens d'un mot. On peut citer à titre d'exemple les paires suivantes : « mais » et « maïs », « abîme » et « abimé ». Des exemples similaires existent dans de nombreuses langues, dans lesquelles les signes diacritiques sont utilisés pour faire varier la prononciation des lettres auxquelles ils sont ajoutés. De plus, la plupart des matrices de confusion présentent les lettres en majuscule. Toutefois, la confusion entre deux lettres en majuscule diffère de la confusion entre les mêmes lettres en minuscule (par exemple, « E » et « F » sont visuellement proches mais « e » et « f » ne le sont pas et inversement pour les lettres « B » et « D » *vs.* « b » et « d »). La plupart des études classiques de perception visuelle de mots et de lettres utilisaient des caractères en majuscule pour éviter le recours aux accents. Actuellement, l'utilisation des lettres en minuscule est de plus en plus fréquente étant donné que, dans les situations écologiques, les minuscules sont la règle et les majuscules l'exception.

Simpson et al. (2012) sont les seuls, à notre connaissance, à proposer une matrice de similarité pour l'ensemble des caractères présents dans les langues utilisant l'alphabet latin. Dans cette étude, les auteurs fournissent une mesure « subjective » contrairement à la majorité des études qui fournissent des matrices de confusion « objectives » en se basant sur des résultats comportementaux sur la perceptibilité visuelle des lettres. En effet, la procédure la plus couramment utilisée pour évaluer la confusion entre les lettres d'un alphabet consiste à présenter les caractères et à demander à un observateur de nommer l'identité du caractère présenté. Dans ce paradigme, une matrice de confusion peut être construite en calculant le nombre de réponses données pour chaque paire stimulus-réponse. Le pourcentage de réponses correctes et d'erreurs permet de quantifier objectivement la confusion entre les lettres (Geyer, 1977 ; Townsend, 1971 ; pour une revue voir Mueller et Weidemann 2012). Afin d'éviter que les participants ne fassent aucune erreur et qu'aucune confusion ne puisse être mesurée, les conditions de présentation sont rendues difficiles en fixant par exemple des temps de présentation très courts ou en diminuant le contraste visuel des stimuli.

En revanche, dans l'étude de Simpson et al. (2012), les auteurs demandent aux participants de noter sur une échelle de Likert, allant de 1 à 7, la similarité pour des paires de lettres qui leur sont présentées. Chaque paire est issue d'un tirage aléatoire de l'ensemble des lettres accentuées et non accentuées des langues utilisant l'alphabet latin (53 lettres). Les résultats moyennés sur l'ensemble des paires et sur l'ensemble des participants permettent de produire une matrice de similarité « subjective » et non une matrice de confusion (voir Figure 2.3).

Objectif et Méthode de calibration

Une matrice de confusion « objective » a l'avantage de pouvoir être interprétée comme une table des probabilités de reconnaître une lettre sachant une lettre présentée, ce qui correspond bien à l'information contenue dans la distribution de probabilité $P(I_n^t \mid S_n^t)$ dans le modèle BRAID. En revanche, une matrice « subjective », comme celle proposée par Simpson et al. (2012), présente l'inconvénient de ne pas pouvoir être interprétée directement comme une probabilité de reconnaissance de lettres.

Notre objectif dans cette phase de calibration est donc d'adapter cette matrice de similarité pour exprimer les valeurs de probabilité de la distribution $P(I_n^t \mid S_n^t)$ et de pouvoir rendre compte de la confusion entre les caractères (dont les caractères accentués). La méthode que nous suivons dans cette phase consiste à identifier la transformation mathématique qui lie les paramètres de la matrice de similarité et les paramètres de la matrice par défaut de BRAID. La contrainte que nous fixons pour la nouvelle matrice de confusion est de produire, à travers BRAID-Phon, des dynamiques de reconnaissance de lettres similaires à celles obtenues en utilisant la matrice par défaut. Les valeurs de paramètres de confusion pour les lettres absentes dans la matrice par défaut, seront ensuite estimées en effectuant la même transformation.

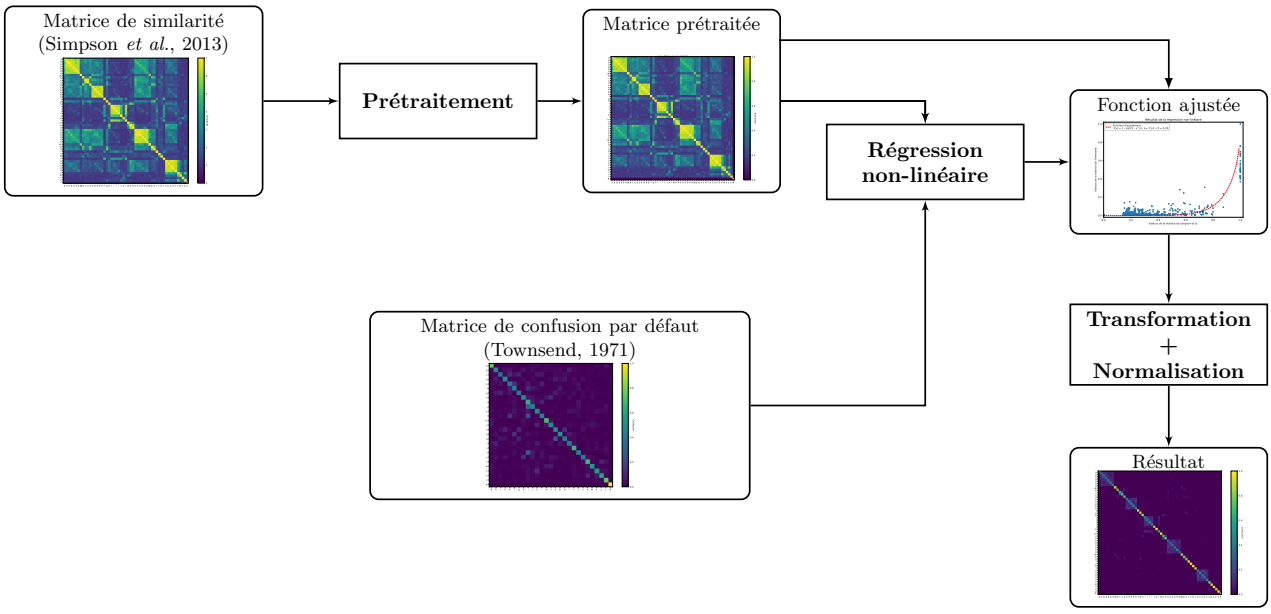


FIGURE 2.4 – **Etapes suivies pour transformer la matrice de similarité de Simpson et al. (2012) en une matrice de probabilité de confusion.**

Nous procédons en quatre étapes (voir Figure 2.4) :

1. **Prétraitement** Dans cette étape, les valeurs de la matrice de Simpson et al. (2012) sont divisées par la valeur de similarité maximale, qui est 7. On obtient donc une matrice dont chaque valeur de similarité est comprise entre 0 et 1.

2. **Régression non-linéaire** A l'aide d'une fonction non-linéaire, nous effectuons une régression afin d'estimer une transformation mathématique entre les valeurs de la matrice par défaut (Townsend, 1971) et la matrice de Simpson et al. (2012) pré-traitée. Dans cette étape, nous nous limitons aux valeurs pour les paires de lettres communes entre les deux matrices.

3. **Transformation de la matrice** Nous ajustons l'ensemble des valeurs de la matrice de Simpson et al. (2012) à l'aide de la fonction obtenue. La matrice résultante est également normalisée de sorte que chaque ligne somme à 1 et qu'elle puisse être interprétée comme une matrice de confusion (chaque valeur peut être interprétée comme la probabilité de reconnaître une lettre sachant la lettre présentée).

4. **Vérification** Nous comparons, à l'aide de l'exécution d'une tâche de reconnaissance de lettres, les courbes d'évolution des probabilités obtenues par la matrice par défaut et la nouvelle matrice de confusion.

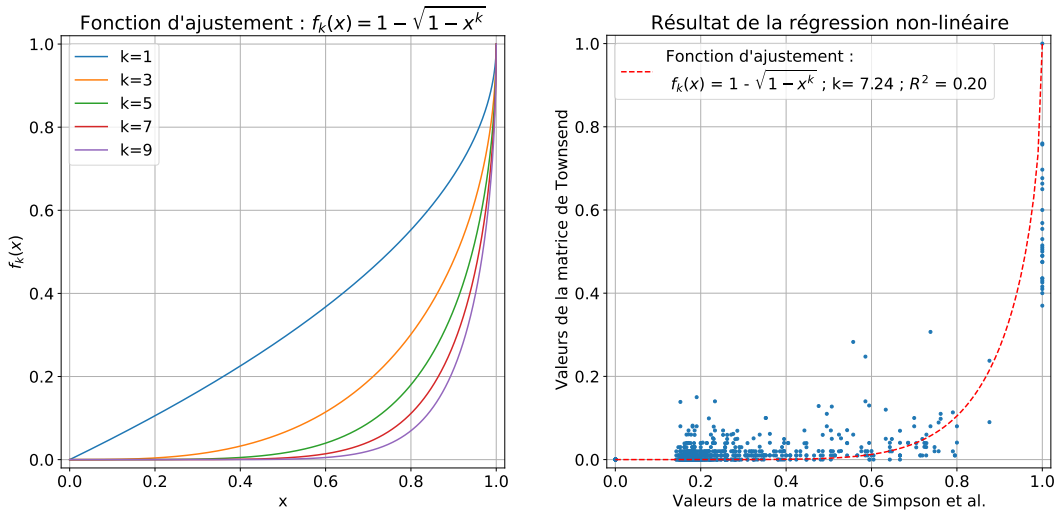


FIGURE 2.5 – **Famille de fonctions utilisée pour ajuster les valeurs de la matrice de similarité de Simpson et al. (2012) à celle de Townsend (1971).** A gauche, nous représentons la famille de fonctions avec différentes valeurs du paramètre k ; toutes les fonctions passent par le point (0,0) et (1,1) mais des valeurs de k de plus en plus élevées impliquent des courbes de plus en plus « affaissées ». A droite, nous visualisons les valeurs à ajuster (en abscisse) et les valeurs cibles (en ordonnées). La fonction d'ajustement, qui résulte de la régression, est représentée par la courbe rouge.

Calibration et Résultats

Nous observons (voir Figure 2.3) que les valeurs dans les matrices de Simpson et al. (2012) et (Townsend, 1971), bien qu'elles soient toutes les deux comprises entre 0 et 1, ne sont pas réparties de la même manière dans cet intervalle. La Figure 2.5 (droite) montre cela par représentation

des valeurs communes entre ces deux matrices, d'un espace relativement à l'autre. Nous observons ainsi que les valeurs de la matrice de Simpson et al. (2012) sont séparées en deux groupes : les premières valeurs sont comprises entre, environ, 0,2 et 0,8, tandis que les valeurs correspondantes dans la matrice de (Townsend, 1971) sont, elles, comprises entre 0 et 0,1. Ces valeurs sont les termes « non diagonaux » des deux matrices. A l'inverse, les valeurs sur la diagonale sont bien réparties entre 0,4 et 0,8 dans la matrice de (Townsend, 1971) alors que les valeurs correspondantes dans Simpson et al. (2012) sont toutes regroupées à 1.

Nous cherchons donc à identifier une transformation pour faire correspondre les valeurs étalées de la matrice de Simpson et al. (2012) à celles de la matrice de (Townsend, 1971) et donc de les « tasser » entre 0 et 0,1, sans affecter les valeurs proches de 1. Après une exploration empirique, nous choisissons une famille de transformations non-linéaires qui répond à ce besoin et qui est définie par :

$$f_k(x) = 1 - \sqrt{1 - x^k},$$

avec un paramètre k qui permet de contrôler l'abaissement de la courbe (voir Figure 2.5). La valeur du paramètre qui permet le mieux d'ajuster les données (obtenue par une méthode standard d'optimisation par moindre carrés non-linéaires) et qui sera utilisée pour la transformation des valeurs de la matrice de similarité est $k = 7{,}24$.

Le décours temporel de la probabilité d'identification des lettres isolées a permis de confirmer la préservation de la dynamique générale des courbes issues des simulations avec les deux matrices (voir Figure 2.6). Nous remarquons que les reconnaissances des lettres non accentuées ont des dynamiques très similaires pour les deux matrices. Cependant, les caractères qui peuvent être accentués (par exemple, le « a », le « e » ou le « i ») ont une dynamique ralentie, à cause de la compétition avec leurs variantes accentuées (par exemple, « â » ou « ä » pour le « a », « é » ou « è » pour le « e », etc.). En moyenne, la probabilité d'identification des lettres atteint 0,95 avant 150 itérations. Ce critère a été adopté lors de la calibration de la matrice d'origine par Phénix (2018) et correspond au taux d'identification des lettres reporté expérimentalement.

Prise en compte des caractères accentués dans BRAID-Phon

Pour utiliser notre nouvelle matrice de confusion étendue aux caractères accentués dans BRAID-Phon, nous adaptons la définition de quelques portions du modèle BRAID. Ainsi, toutes les variables qui avaient pour domaine $\mathscr{D}_L$ (l'ensemble des 26 lettres de l'alphabet anglais plus le caractère inconnu #) voient leur domaine étendu et il devient $\mathscr{D}'_L$ (l'ensemble des 53 lettres latines, accentuées ou non, plus le caractère inconnu #). Cela concerne les variables I_n^t et S_n^t du sous-modèle sensoriel des lettres, les variables P_n^t du sous-modèle perceptif des lettres et les variables L_n^t du sous-modèle lexical. Les domaines des variables I_n^t et S_n^t étant étendus, nous pouvons maintenant re-définir le terme $P(I_{1:N}^t \mid S_{1:N}^t\ G^t\ \Delta I_{1:N}^t)$ du sous-modèle sensoriel des lettres sur la base des paramètres de notre matrice de confusion étendue aux caractères accentués.

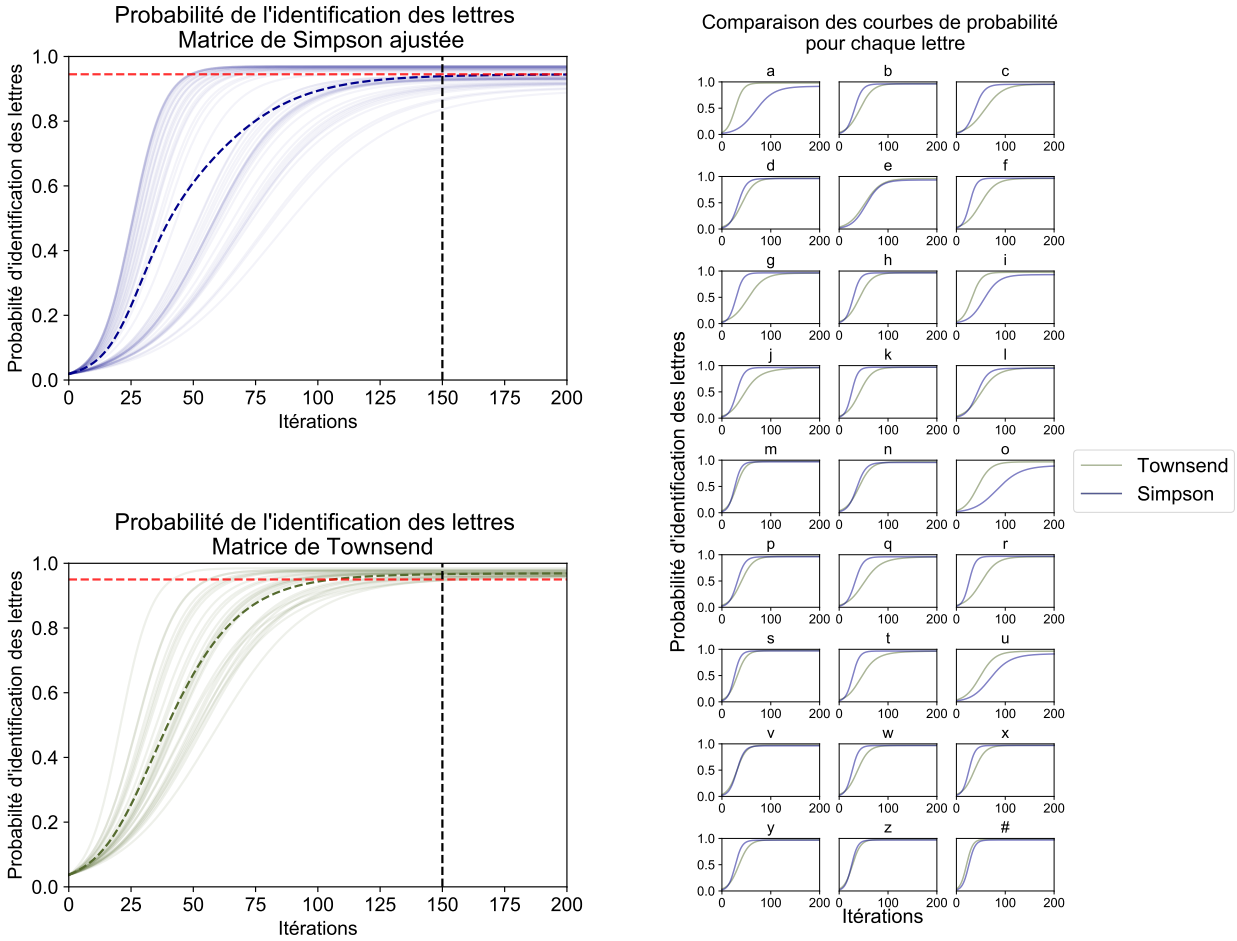


FIGURE 2.6 – **Comparaison des courbes d'évolution de probabilité lors de la reconnaissance des lettres isolées.** Les résultats obtenus en utilisant la matrice de confusion de Simpson et al. (2012) ajustée par notre méthode sont affichés en haut à gauche, et ceux en utilisant la matrice de Townsend (1971) en bas à gauche. Sur tous les graphiques, l'axe des abscisses indique les itérations et l'axe des ordonnées indique la probabilité d'identification de la lettre. A gauche, les courbes en pointillés indiquent la courbe d'évolution moyenne de toutes les lettres. La ligne horizontale rouge indique le seuil de probabilité de 95%. A droite, est affichée la comparaison des courbes de probabilité des lettres, pour chaque lettre non accentuée.

2.2 Connaissances Phonologiques dans le Sous-modèle Lexical

Rappelons que le sous-modèle lexical ne contient, dans le modèle BRAID original, en plus du mécanisme d'évaluation de l'appartenance lexicale, que des connaissances relatives à l'orthographe des mots connus (voir Section 1.3). Il y a deux composantes pour représenter ces connaissances orthographiques lexicales : premièrement, une distribution a priori sur l'espace des mots connus, $P(W^0)$, qui représente la fréquence des mots du lexique, et deuxièmement, une série de distributions de probabilité de la forme $P(L_n^t \mid [W^t = w])$, qui décrivent les connaissances sur la n-ème lettre du mot w (des distributions « quasi-Dirac », donnant une connaissance très certaine sur cette lettre pour ce mot). Le modèle est complété par une chaîne temporelle, reposant sur un terme dynamique $P(W^t \mid W^{t-1})$, pour permettre l'accumulation d'évidence perceptive, à chaque instant, sur l'identité du mot dans le stimulus.

Dans BRAID-Phon, nous développons ce sous-modèle lexical, avec deux objectifs. Notre pre-

mier objectif est d'étendre le sous-modèle lexical pour y inclure, de manière symétrique, des connaissances relatives aux représentations phonologiques des mots. Dans ce travail, nous choisissons, par simplicité, que les représentations phonologiques consistent en des représentations phonémiques : en d'autres termes, les sons pour un mot sont encodés par une séquence de phonèmes (par exemple, nous n'abordons pas ici la question de la nature syllabique ou phonémique, des représentations phonologiques). Notre second objectif est d'implémenter l'hypothèse d'une voie unique entre les représentations orthographique et phonologique du modèle. Par conséquent, nous choisissons la contrainte de ne « lier » les représentations orthographiques et phonologiques qu'à travers l'espace de représentation des mots du lexique.

Formulation mathématique

Dans BRAID-Phon, nous incluons au sous-modèle lexical des variables probabilistes $\Phi_{1:M}^{1:T}$, qui représentent les phonèmes qui composent les mots connus. M est la longueur phonémique la plus grande du lexique. Ces variables sont définies sur l'espace des phonèmes $\mathscr{D}_\Phi$ qui dépend de la langue choisie : nous utilisons 39 phonèmes en français et 57 phonèmes en anglais. De manière analogue aux représentations orthographiques, nous avons ajouté un « phonème » supplémentaire /#/ au domaine des phonèmes, pour représenter la fin de la chaîne phonémique.

Pour répondre à nos deux objectifs initiaux, nous supposons une indépendance entre les représentations phonologiques et orthographiques, conditionnellement à la connaissance du mot considéré. En d'autres termes, il n'y a pas de relation directe entre les variables $L_{1:N}^{1:T}$ et $\Phi_{1:M}^{1:T}$, et toute relation probabiliste entre ces deux représentations sera « médiatisée » par le lexique, c'est-à-dire que les inférences probabilistes mettront en jeu les variables lexicales $W^{1:T}$. La décomposition de la distribution conjointe du sous-modèle lexical étendu aux représentations phonologiques s'écrit alors :

$$P(W^{0:T}\,\Phi_{1:M}^{1:T}\,L_{1:N}^{1:T}) = P(W^0)\prod_{t=1}^{T}\left[P(W^t \mid W^{t-1})\prod_{n=1}^{N}P(L_n^t \mid W^t)\prod_{m=1}^{M}P(\Phi_m^t \mid W^t)\right]. \qquad (2.1)$$

Par rapport au modèle BRAID initial, le sous-modèle lexical de BRAID-Phon contient donc une série de termes $P(\Phi_m^t \mid W^t)$, qui sont les distributions de probabilité qui caractérisent les connaissances sur les phonèmes des mots. Elles sont exprimées de manière symétrique aux distributions $P(L_n^t \mid W^t)$ des lettres, avec des « quasi-Dirac » sur les phonèmes des mots. Nous écrivons :

$$P([\Phi_i^t = \phi] \mid [W^t = w]) = \begin{cases} 1-(|\mathscr{D}_\Phi|-1)\,\epsilon & \text{si } \phi \text{ est le } i\text{ème phonème du mot } w, \\ \epsilon & \text{sinon ,} \end{cases}$$

tel que $\epsilon = 10^{-2}$.

2.3 Sous-modèle Phonologique Perceptif

Nous venons d'étendre le sous-modèle lexical pour y inclure des connaissances phonologiques, et avons procédé par symétrie avec les connaissances orthographiques. En effet, on peut voir le

sous-modèle lexical comme ayant deux espaces de représentations (orthographiques et phonologiques), en « étoile » autour d'un espace pivot, l'espace W des mots du lexique.

Nous poursuivons cette approche « par symétrie » pour inclure un sous-modèle perceptif des phonèmes des mots, en miroir du sous-modèle perceptif des lettres (voir Section 1.3). Le sous-modèle perceptif des lettres de BRAID repose essentiellement sur un ensemble de variables $P_{1:N}^{0:T}$ et des « chaînes de Markov », c'est-à-dire des modèles dynamiques avec des termes $P(P_n^t \mid P_n^{t-1})$ (une telle chaîne pour chaque position orthographique), pour y permettre l'accumulation graduelle d'évidence perceptive.

De manière symétrique, le sous-modèle perceptif des phonèmes repose sur un ensemble de variables probabilistes $\Psi_{1:M}^{0:T}$, définies sur le même espace $\mathscr{D}_\Phi$ que les variables $\Phi_{1:M}^{1:T}$ du sous-modèle lexical étendu. Le sous-modèle phonologique perceptif contient également, des chaînes de Markov (une par position phonologique), avec un même modèle dynamique $P(\Psi_m^t \mid \Psi_m^{t-1})$, et des distributions a priori $P(\Psi_m^0)$ pour les états initiaux de ces chaînes. Ainsi, la décomposition de la distribution conjointe du sous-modèle phonologique perceptif est :

$$P(\Psi_{1:M}^{0:T}) = \prod_{m=1}^{M} \left[P(\Psi_m^0) \prod_{t=1}^{T} P(\Psi_m^t \mid \Psi_m^{t-1}) \right].$$

Les termes $P(\Psi_m^0)$ sont supposés uniformes. Les termes $P(\Psi_m^t \mid \Psi_m^{t-1})$ sont définis de manière similaire aux autres termes dynamiques du modèle :

$$P([\Psi_m^t = \psi_m^t] \mid [\Psi_m^{t-1} = \psi_m^{t-1}]) = \begin{cases} \frac{1+Leak_\Psi}{1+|\mathscr{D}_\Psi| \; Leak_\Psi} & \text{si } \psi_m^t = \psi_m^{t-1} \\ \frac{Leak_\Psi}{1+|\mathscr{D}_\Psi| \; Leak_\Psi} & \text{sinon.} \end{cases}$$

Le sous-modèle phonologique perceptif permet ainsi l'accumulation et le maintien, au cours du temps, des connaissances sur les phonèmes du stimulus, lorsque les autres sous-modèles en fournissent. Inversement, le sous-modèle permet également un déclin graduel de l'information phonologique, au cours du temps, et un retour vers un état initial incertain, en absence d'information phonologique provenant des autres sous-modèles (par exemple, en absence de stimulus visuel). Nous avons paramétré ce sous-modèle empiriquement et $Leak_\Psi = 0{,}02$.

2.4 Distribution Conjointe du modèle BRAID-Phon

Précédemment, nous avons décrit mathématiquement toutes les composantes de BRAID-Phon. Nous présentons maintenant le modèle dans son ensemble. Dans notre formalisme, un modèle bayésien est décrit à l'aide d'une distribution conjointe de toutes les variables qu'il inclut. Ayant défini les distributions conjointes des sous-modèles dans les Sections 1 et 2, la distribution de probabilité conjointe $JD_{BRAID\text{-}Phon}$ du modèle BRAID-Phon est simplement le regroupement de tous les termes décrits précédemment.

Expression de la distribution conjointe

Nous définissons la distribution de probabilité conjointe $JD_{BRAID\text{-}Phon}$ de BRAID-Phon ainsi :

$$JD^{T}_{BRAID\text{-}Phon} = P\left(\begin{array}{c} {\Psi^{0:T}_{1:M}}\ {\lambda_{\Phi}}^{1:T}_{1:M} \\ \\ {\Phi^{1:T}_{1:M}}\ W^{0:T}\ L^{1:T}_{1:N} \\ \\ D^{0:T}\ \Gamma_{L}^{1:T}\ {C_{D}}^{1:T}_{1:N}\ {\lambda_{D}}^{1:T}_{1:N}\ {\lambda_{L}}^{1:T}_{1:N} \\ \\ {P^{0:T}_{1:N}}\ {\lambda_{P}}^{1:T}_{1:N} \\ \\ A^{1:T}\ {\mu_{A}}^{1:T}\ {\sigma_{A}}^{1:T}\ {C_{A}}^{1:T}_{1:N} \\ \\ {S^{1:T}_{1:N}}\ G^{1:T}\ {\Delta I}^{1:T}_{1:N}\ {I}^{1:T}_{1:N} \end{array}\right).$$

La décomposition de cette distribution est le produit des distributions conjointes de chacun des sous-modèles. Nous arrangeons ensuite les termes du produit de façon à exprimer la conjointe de BRAID-Phon $JD_{BRAID\text{-}Phon}$ à l'aide de celle de BRAID, JD_{BRAID}. Nous obtenons :

$$JD^{T}_{BRAID\text{-}Phon} = \left[\begin{array}{c} P(W^0)\ P(D^0)\ \prod_{n=1}^{N} P(P^0_n)\ \prod_{m=1}^{M} P(\Psi^0_m) \\ \\ \prod_{t=1}^{T} \left[\begin{array}{c} \prod_{m=1}^{M}\ P(\Psi^t_m \mid \Psi^{t-1}_m)\ P({\lambda_{\Phi}}^t_m \mid \Psi^t_m\ \Phi^t_m) \\ \\ P(W^t \mid W^{t-1})\ \prod_{m=1}^{M}\ P(\Phi^t_m \mid W^t)\ \prod_{n=1}^{N}\ P(L^t_n \mid W^t) \\ \\ P(\Gamma_L^t \mid D^t)\ P(D^t \mid D^{t-1})\ P({C_D}^t_{1:N} \mid D^t) \prod_{n=1}^{N} P({\lambda_D}^t_n \mid {\lambda_L}^t_n\ {C_D}^t_n) \\ \\ \prod_{n=1}^{N} P(P^t_n \mid P^{t-1}_n)\ P({\lambda_L}^t_n \mid L^t_n\ P^t_n\ \Gamma_L^t) \\ \\ \prod_{n=1}^{N} P({\lambda_P}^t_n \mid P^t_n\ I^t_n\ {C_A}^t_n) \\ \\ P(A^t \mid \mu_A^t\ \sigma_A^t)\ P(\mu_A^t)\ P(\sigma_A^t)\ \prod_{n=1}^{N} P({C_A}^t_n \mid A^t) \\ \\ P(G^t)\ \prod_{n=1}^{N} P(S^t_n)\ P(\Delta I^t_n)\ P(I^t_n \mid S^t_{1:N}\ \Delta I^t_n\ G^t) \end{array}\right] \end{array}\right] \tag{2.2}$$

$$= JD^{T}_{BRAID} \prod_{m=1}^{M} \left[P(\Psi^0_m) \prod_{t=1}^{T} \left[\begin{array}{c} P(\Psi^t_m \mid \Psi^{t-1}_m) \\ \\ P({\lambda_{\Phi}}^t_m \mid \Psi^t_m\ \Phi^t_m) \\ \\ P(\Phi^t_m \mid W^t) \end{array}\right] \right].$$

Etant donné que la structure interne du modèle BRAID est conservée dans l'extension BRAID-Phon, la distribution conjointe du modèle BRAID-Phon est la conjointe de BRAID multipliée par

des termes supplémentaires (qui réfèrent aux sous-modèles de l'extension). Ce modèle possède également une représentation graphique qui décrit visuellement les dépendances entre les variables (voir Figure 2.7).

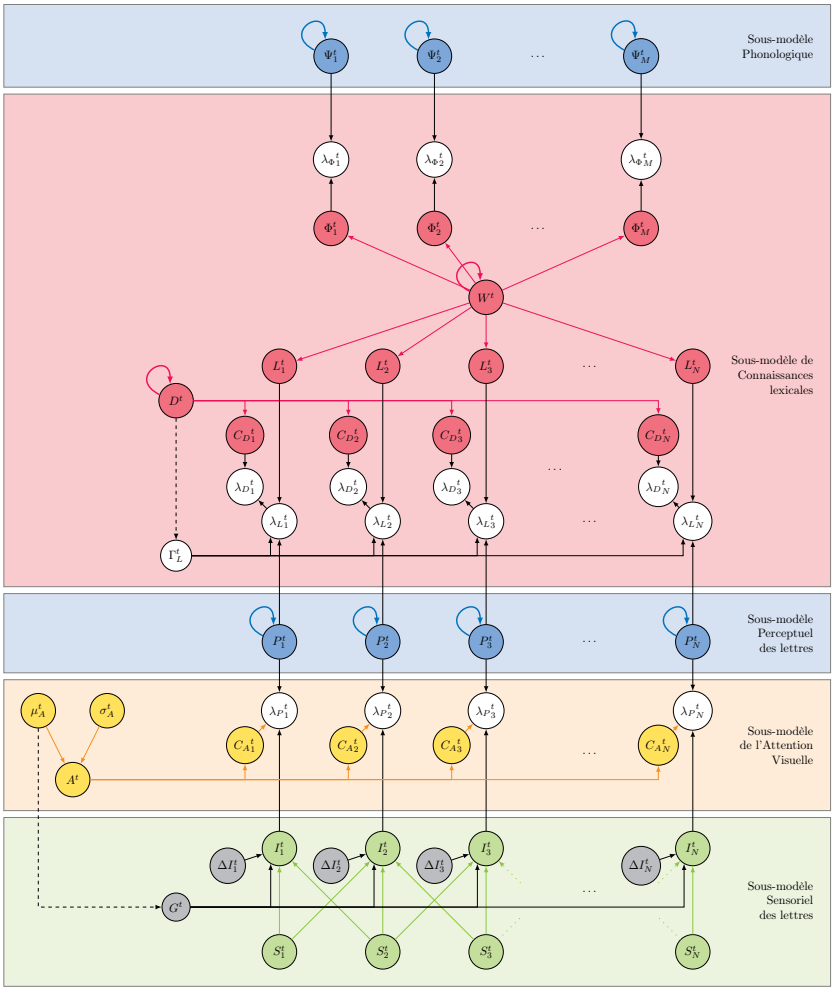


FIGURE 2.7 – **Représentation graphique du modèle BRAID-Phon.** Les nœuds représentent les variables du modèle et les flèches qui les relient représentent les dépendances décrites formellement dans la décomposition de la distribution de probabilité conjointe. Les flèches en boucle représentent les dépendances temporelles : une variable à l'instant T , qui est représentée par un nœud avec une flèche en boucle, dépend d'elle même à l'instant $T-1$. La flèche en pointillé représente une affectation de valeur de la variable à l'origine de la flèche et transmise à la variable à la fin de la flèche.

3 Simulation de la lecture de mot dans BRAID-Phon

La tâche de lecture est modélisée dans BRAID-Phon par une question probabiliste que nous noterons ${Q_\Psi}^T$. Pour calculer la sortie phonologique qui correspond à une séquence de lettres en entrée, nous inférons l'identité des phonèmes à chaque instant T sachant l'identité du stimulus, la position du regard et le profil de l'attention visuelle. Mathématiquement, nous écrivons :

$$ {Q_\Psi}^T = P\left(\Psi_{1:M}^T \mid S_{1:N}^{1:T}\, \mu_A^{1:T}\, \sigma_A^{1:T}\, G^{1:T}\, [\lambda_{\Phi 1:M}^{1:T} = 1]\, [\lambda_{P 1:N}^{1:T} = 1]\right) . \tag{2.3} $$

Dans notre cadre de modélisation, le calcul de la réponse à une question probabiliste s'effectue en appliquant les règles du calcul probabiliste. Notre point de départ est la distribution conjointe $JD^T_{BRAID\text{-}Phon}$ et l'expression de ses termes. Il suffit ensuite de marginaliser sur les variables manquantes (n'apparaissant ni à droite ni à gauche de l'expression de la probabilité conditionnelle). Le résultat obtenu après cette marginalisation est proportionnel à la réponse de la question probabiliste. Dans le cas du calcul de ${Q_\Psi}^T$, nous obtenons :

$$ {Q_\Psi}^T \propto \sum_{\substack{\Psi_{1:M}^{0:T-1} \, \Phi_{1:M}^{1:T} \, W^{0:T} \, L_{1:N}^{1:T} \, \lambda_{1:N}^{1:T} \\ D^{0:T} \, \Gamma_L^{1:T} \, C_{D\,1:N}^{1:T} \, \lambda_{D\,1:N}^{1:T} \\ P_{1:N}^{0:T} \, A^{1:T} \, C_{A\,1:N}^{1:T}}} JD^T_{BRAID\text{-}Phon} \cdot \tag{2.4} $$

Pour rendre cette expression « utilisable », notre objectif est de réorganiser les sommations afin de faire apparaître une formule de récurrence temporelle, c'est-à-dire que nous souhaitons exprimer ${Q_\Psi}^T$ en fonction de la même question au pas de temps précédent, ${Q_\Psi}^{T-1}$. Nous utilisons pour cela la décomposition obtenue dans l'Equation (2.2). Ainsi, nous avons :

$$ {Q_\Psi}^T \propto \sum_{\substack{\Psi_{1:M}^{0:T-1} \, \Phi_{1:M}^{1:T} \\ W^{0:T} \, L_{1:N}^{1:T} \\ D^{0:T} \, \Gamma_L^{1:T} \\ C_{D\,1:N}^{1:T} \, \lambda_{D\,1:N}^{1:T} \\ P_{1:N}^{0:T} \, A^{1:T} \, C_{A\,1:N}^{1:T}}} \left[JD^T_{BRAID} \prod_{m=1}^{M} \left[P(\Psi_m^0) \prod_{t=1}^{T} \begin{bmatrix} P(\Psi_m^t \mid \Psi_m^{t-1}) \\ P([\lambda_{\Phi}{}_m^t = 1] \mid \Psi_m^t \, \Phi_m^t) \\ P(\Phi_m^t \mid W^t) \end{bmatrix} \right] \right] . $$

En isolant les sommes sur les variables spécifiques des connaissances phonologiques, propres au modèle BRAID-Phon, nous obtenons :

$$ {Q_\Psi}^T \propto \sum_{\substack{W^{0:T} \\ \Phi_{1:M}^{1:T}, \Psi_{1:M}^{0:T-1}}} \left[\begin{array}{c} \prod_{m=1}^{M} \left[P(\Psi_m^0) \prod_{t=1}^{T} \begin{bmatrix} P(\Psi_m^t \mid \Psi_m^{t-1}) \\ P([\lambda_{\Phi}{}_m^t = 1] \mid \Psi_m^t \, \Phi_m^t) \\ P(\Phi_m^t \mid W^t) \end{bmatrix} \right] \\ \sum_{\substack{L_{1:N}^{1:T} \, D^{0:T} \, \Gamma_{\Phi}^{1:T} \, \Gamma_L^{1:T} \\ C_{D\,1:N}^{1:T} \, \lambda_{D\,1:N}^{1:T} \, P_{1:N}^{0:T} \, A^{1:T} \, C_{A\,1:N}^{1:T}}} JD^T_{BRAID} \end{array} \right] . \tag{2.5} $$

La sommation interne à cette expression est presque égale au calcul de la reconnaissance de mots dans le modèle BRAID (Phénix, 2018). Cette tâche correspond à la question probabiliste ${Q_W}^T = P\left(W^T \mid K^{1:T}\right)$, avec $K^t = \left\{ S_{1:N}^t \, \mu_A^t \, \sigma_A^t \, G^t \, [\lambda_{P}{}_{1:N}^t = 1] \right\}$: dans BRAID, la tâche de reconnaissance de mot correspond à calculer la distribution de probabilité sur l'espace des mots, sachant un stimulus visuel, la position du regard et de l'attention visuelle, et sachant que l'information se propage dans le modèle entre l'entrée sensorielle et l'espace lexical. Le calcul de la tâche de

reconnaissance de mots admet une solution récursive :

$$Q_{W}{}^{T} = P\left(W^{T} \mid K^{1:T}\right) \tag{2.6}$$

$$\propto \sum_{W^{0:T-1}} \sum_{L_{1:N}^{1:T}} JD_{BRAID}^{T} \tag{2.7}$$

$$\propto \sum_{W^{T-1}} \begin{bmatrix} Q_{W^{T-1}}\, P(W^{T} \mid W^{T-1}) \\ \prod_{n=1}^{N} \left\langle P(L_{n}^{T} \mid W^{T}), P(L_{n}^{T} \mid K^{1:T}) \right\rangle \end{bmatrix}. \tag{2.8}$$

Dans cette expression, le terme $P(L_n^T \mid K^{1:T})$ peut être interprété comme la reconnaissance des lettres du stimulus. Dans certaines de nos équations (Phénix, 2018), il est noté comme la question probabiliste de la reconnaissance des lettres, $Q_P{}^T$.

Nous arrangeons les termes de l'Equation (2.5), pour exprimer $Q_\Psi{}^T$ en fonction de $Q_\Psi{}^{T-1}$ et de la question de reconnaissance de mots $Q_W{}^T$:

$$Q_{\Psi}{}^{T} \propto \sum_{\substack{W^{0:T} \\ \Phi_{1:M}^{1:T}\, \Psi_{1:M}^{0:T-1}}} \left[\prod_{m=1}^{M} \left[P(\Psi_m^0) \prod_{t=1}^{T} \left[\begin{matrix} P(\Psi_m^t \mid \Psi_m^{t-1}) \\ P([\lambda_{\Phi}{}_m^t = 1] \mid \Psi_m^t\, \Phi_m^t) \\ P(\Phi_m^t \mid W^t) \\ Q_W{}^T \end{matrix} \right] \right] \right] \tag{2.9}$$

$$\propto \sum_{W^T\, \Phi_{1:M}^T\, \Psi_{1:M}^{T-1}} \left[\begin{matrix} Q_\Psi{}^{T-1} \\ \prod_{m=1}^{M} \left[\begin{matrix} P(\Psi_m^T \mid \Psi_m^{T-1}) \\ P([\lambda_{\Phi}{}_m^T = 1] \mid \Psi_m^T\, \Phi_m^T) \\ P(\Phi_m^T \mid W^T) \end{matrix} \right] \\ Q_W{}^T \end{matrix} \right]. \tag{2.10}$$

Considérons maintenant une valeur de probabilité particulière de $Q_\Psi{}^T$ pour une séquence de phonèmes donnée $\psi_{1:M}$: $Q_\Psi{}^T(\psi_{1:M}) = P\left([\Psi_{1:M}^T = \psi_{1:M}] \mid S_{1:N}^{1:T}\, \mu_A^{1:T}\, \sigma_A^{1:T}\, G^{1:T} [\lambda_{\Phi}{}_{1:M}^{1:T} = 1]\, [\lambda_P{}_{1:N}^{1:T} = 1]\right)$. Noter cette valeur permet, grâce aux variables de cohérence $\lambda_\Phi{}^{1:T}$ qui sont supposées à 1, de simplifier le terme correspondant, de faire disparaître la sommation sur $\Phi_{1:M}^T$ et de « propager les valeurs $\psi_{1:M}$ aux variables $\Phi_{1:M}^T$ » :

$$Q_{\Psi}{}^{T}(\psi) \propto \sum_{W^T\, \Psi_{1:M}^{T-1}} \left[\begin{matrix} Q_\Psi{}^{T-1} \\ \prod_{m=1}^{M} \left[\begin{matrix} P([\Psi_m^T = \psi_m] \mid \Psi_m^{T-1}) \\ P([\Phi_m^T = \psi_m] \mid W^T) \end{matrix} \right] \\ Q_W{}^T \end{matrix} \right]. \tag{2.11}$$

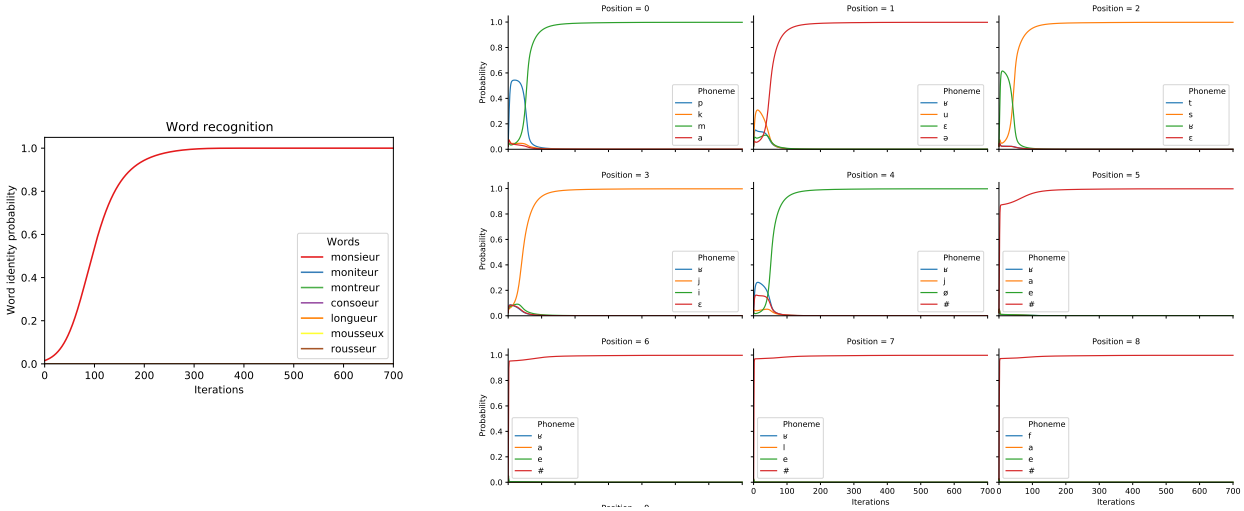
En réorganisant les sommations sur $\Psi_{1:M}^{T-1}$ et sur W^T :

$$Q_{\Psi}{}^{T}(\psi) \propto \sum_{W^T} \begin{bmatrix} \sum_{\Psi_{1:M}^{T-1}} \left[Q_\Psi{}^{T-1} \prod_{m=1}^{M} P([\Psi_m^T = \psi_m] \mid \Psi_m^{T-1}) \right] \\ Q_W{}^T \prod_{m=1}^{M} P([\Phi_m^T = \psi_m] \mid W^T) \end{bmatrix} \tag{2.12}$$

$$\propto \sum_{\Psi_{1:M}^{T-1}} \left[Q_\Psi{}^{T-1} \prod_{m=1}^{M} P([\Psi_m^T = \psi_m] \mid \Psi_m^{T-1}) \right] \cdot \sum_{W^T} \left[Q_W{}^T \prod_{m=1}^{M} P([\Phi_m^T = \psi_m] \mid W^T) \right]. \tag{2.13}$$

Nous pouvons interpréter ce calcul comme une variante de celui de la reconnaissance visuelle de mots. En effet, le calcul de l'identité des phonèmes découle directement de la distribution de probabilité issue de la reconnaissance des mots, ${Q_W}^T$. Mathématiquement, les distributions de probabilité de l'identité des phonèmes proviennent de : 1) la marginalisation du produit $\left[P(\Phi_{1:M}^T \mid W^T)\ {Q_W}^T\right]$ sur W^T, l'espace des mots connus $\mathscr{D}_W$ et 2) la multiplication par le terme dynamique de la chaîne de Markov sur les phonèmes, $\left[{Q_\Psi}^{T-1} \prod_{m=1}^{M} P(\Psi_m^T \mid \Psi_m^{T-1})\right]$. Ainsi, du point de vue des chaînes de Markov sur les variables phonémiques Ψ_m^T, celles-ci reçoivent l'information perceptive sur l'identité des phonèmes dictée par le modèle lexical et sa chaîne de Markov sur l'identité du mot. En d'autres termes, le processus de reconnaissance de mots, et les distributions sur les phonèmes qui en découlent, sont le « modèle capteur » des chaînes de Markov du sous-modèle perceptif des phonèmes.

(a) Lecture du mot « monsieur »



(b) Lecture du mot « mutin »

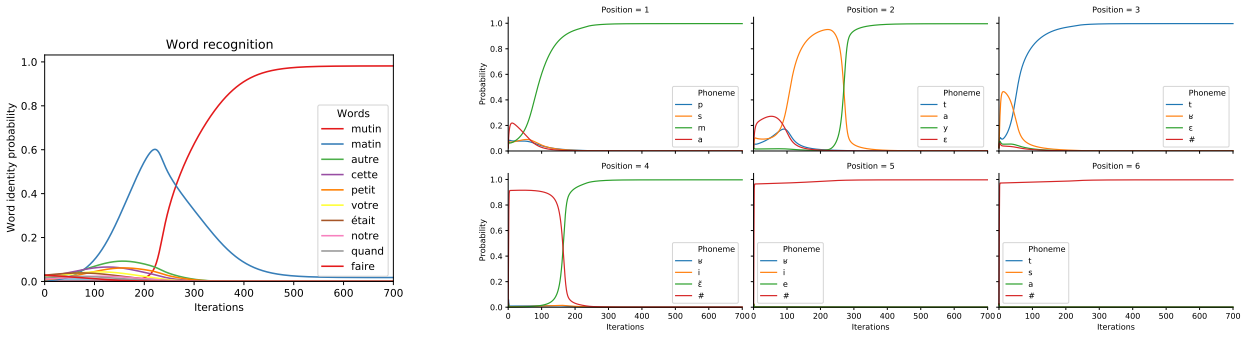


FIGURE 2.8 – **Courbes d'évolution de probabilité lors de la lecture des deux mots : « monsieur » et « mutin ».** Les courbes de probabilité de reconnaissance de mots sont présentées à gauche et celles de production des phonèmes sont présentées à droite. Les courbes relatives aux phonèmes sont visualisées pour plus de positions phonémiques que nécessaire, et le phonème # indique la fin de la suite phonémique produite.

Remarquons que les inférences décrites dans cette section résultent d'une approximation permettant de simplifier les transferts d'informations entre les différentes chaînes de Markov du modèle. En effet, l'inférence bayésienne, dans le cas général des chaînes de Markov couplées, résulte en des boucles de propagation d'information bi-directionnelles. Dans notre cas, l'inférence bayé-

sienne « formelle » indique que nous devrions également considérer les rétro-actions des distributions sur les phonèmes sur le processus de la reconnaissance de mots. Dans le contexte de notre travail, cependant, nous simplifions les calculs pour ne considérer que l'influence de la distribution sur les mots sur la distribution sur les phonèmes. (Plus précisément, c'est le fait d'avoir fait apparaître ${Q_W}^T$ dans notre dérivation qui est « incorrect techniquement », et qui n'est légitime que sous l'hypothèse d'une indépendance conditionnelle simplificatrice, selon laquelle la reconnaissance de mots est considérée indépendante des rétro-actions phonologiques.)

Une conséquence de ce calcul est que la « prononciation » du stimulus dépend de la reconnaissance du mot correspondant à ce stimulus, dans le sens où la distribution de probabilité sur les mots, ${Q_W}^T$, indique directement les distributions de probabilité sur les phonèmes. Par exemple, en tout début de processus, lorsque les lettres sont encore mal identifiées, la distribution sur les mots et donc les distributions de probabilité sur les phonèmes ont une entropie élevée (des distributions « plates »). En revanche, pour un temps de présentation long, si les lettres du stimulus sont parfaitement identifiées et correspondent à un mot connu, alors la distribution sur les mots, et donc sur les phonèmes qui correspondent à ce mot, sont de faible entropie (distributions « piquées »). Enfin, si la distribution de probabilité sur les mots « change d'avis » en cours de route, par exemple lors du traitement d'un stimulus correspondant à un mot de faible fréquence, alors les distributions sur les phonèmes « changeront d'avis » également.

Nous illustrons cette « propriété formelle » sur la Figure 2.8, qui montre des simulations de la tâche de lecture à haute voix pour deux exemples. Dans le premier exemple, le stimulus est le mot « monsieur », et dans le second exemple, le stimulus est le mot « mutin ». Dans le premier exemple, on observe que la reconnaissance de mots, calculée par le terme ${Q_W}^T$, identifie rapidement et sans erreur le mot « monsieur ». En conséquence, les valeurs de probabilité des phonèmes correspondants (/məsjø/) deviennent rapidement élevées elles aussi. L'information sur l'identité du mot est accumulée au cours des itérations, et il en est de même pour l'identité des phonèmes.

Dans le second exemple, en revanche, la reconnaissance du mot a un comportement non-monotone : autour de 250 itérations, c'est le mot « matin » qui est reconnu au lieu de « mutin », car il est plus probable a priori (plus fréquent) lorsque les lettres ne sont pas encore parfaitement identifiées. L'évolution de l'identification des lettres permet de corriger cette erreur transitoire, et la distribution sur les mots converge ensuite vers le mot correct. Nous observons que cette dynamique non-monotone est répliquée à l'identique dans l'espace des phonèmes : le modèle initialement donne des probabilités élevées pour les phonèmes de la séquence /matɛ̃/, puis « change d'avis » et converge vers la prononciation correcte, /mytɛ̃/.

4 Conclusion

Dans ce chapitre, nous proposons une extension d'un modèle de reconnaissance de mots, BRAID (« *Bayesian model of word Recognition with Attention, Inference and Dynamics* »), qui suit une architecture à voie unique. Dans ce travail, nous nous situons dans un cadre bayésien, ce qui permet de formuler chaque tâche cognitive sous forme d'une inférence bayésienne.

L'extension que nous proposons ici, appelée BRAID-Phon, permet de simuler la lecture grâce à une mise à jour du niveau sensoriel des lettres, en prenant en compte les caractères accentués, à l'ajout de connaissances phonologiques dans le niveau lexical et à l'ajout d'un sous-modèle phonologique perceptif. Dans le modèle BRAID-Phon, la sortie phonologique est issue du résultat de la reconnaissance de lettres et de mots et de la fusion avec les connaissances lexicales.

Récapitulatif :

- BRAID est un modèle computationnel de reconnaissance de lettres, de mots et de décision lexicale. Les caractéristiques notables du modèle BRAID :
  - Bayésien : Les connaissances sont décrites par des distributions de probabilité et les tâches cognitives sont simulées par des inférences bayésiennes.
  - Hiérarchique : BRAID suppose les trois niveaux classiques du traitement visuel des mots : le niveau « trait » (notre sous-modèle sensoriel), le niveau « lettres » (notre sous-modèle perceptif) et le niveau « mots » (notre sous-modèle lexical).
  - Attention visuelle : BRAID contient un niveau supplémentaire, représentant l'attention visuelle et permettant de moduler l'accumulation de l'information sensorielle.
- BRAID-Phon : une extension du modèle BRAID, pour la représentation des connaissances phonologiques, obtenue par :
  - la mise à jour du niveau sensoriel pour la prise en compte des lettres accentuées ;
  - la mise à jour du niveau lexical pour y représenter les connaissances phonologiques des mots ;
  - l'ajout d'un sous-modèle phonologique, « miroir » du sous-modèle perceptuel des lettres, pour l'accumulation d'évidences perceptives sur les phonèmes.
- La tâche de lecture de mots est simulée dans le modèle BRAID-Phon.
  - La lecture des mots est définie formellement par une question probabiliste ; répondre à cette question découle de l'application des règles d'inférence bayésienne sur le modèle BRAID-Phon.
  - La lecture des mots produit des phonèmes qui découlent directement de la reconnaissance du mot à lire. En particulier, toute la dynamique de compétition lexicale est reflétée dans la compétition au niveau phonologique.

CHAPITRE

3 Simulation de l'effet de longueur dans des tâches multiples

"Every line is the perfect length if you don't measure it."

—

Dans le Chapitre 2, nous avons défini un nouveau modèle bayésien de lecture à haute voix : le modèle BRAID-Phon. Ce modèle intègre et étend le modèle BRAID, un modèle sophistiqué de reconnaissance de mots. Il est donc conçu pour simuler non seulement la tâche de lecture à haute voix qui implique un transcodage phonologique mais également des tâches plus spécifiquement simulées dans le contexte de modèles de reconnaissance visuelle de mots, comme l'identification de lettres, la reconnaissance de mots et la décision lexicale. Une des originalités du modèle est d'intégrer une composante visuo-attentionnelle, ce qui ouvre, pour la première fois, la possibilité d'évaluer l'impact de l'attention visuelle dans cette grande variété de tâches.

Dans ce chapitre, nous étudions la capacité du modèle BRAID-Phon à simuler les données comportementales recueillies dans le cadre de l'étude Chronolex (Ferrand et al., 2011). Cette méga-étude (en anglais *megastudy*), menée auprès de lecteurs francophones, décrit pour la première fois des données recueillies sur les mêmes items dans trois tâches différentes : une tâche de lecture (NMG), une tâche de décision lexicale (LDT) et une tâche de *progressive demasking* (PDM). Nous nous intéressons plus particulièrement à la capacité du modèle à rendre compte des variations de l'effet de longueur selon la tâche.

Notre premier objectif est d'utiliser les paramètres par défaut du modèle et de déterminer jusqu'à quel point les résultats simulés sont conformes aux données comportementales observées. Le second objectif est d'étudier plus spécifiquement l'impact des paramètres attentionnels sur les résultats simulés. Dans un premier temps, nous tenterons de déterminer si des stratégies visuo-attentionnelles différentes, spécifiques à chaque tâche, permettent de mieux rendre compte des variations observées de l'effet de longueur en termes d'amplitude et de forme. Dans un second temps, nous étudierons de manière exploratoire l'effet de distributions attentionnelles atypiques

sur les différentes tâches.

1 Contexte de l'étude

Les modèles computationnels permettent de simuler les performances observées chez des sujets lors de tâches expérimentales. Dans le domaine du traitement visuel de mots, on mesure habituellement la performance des participants par les taux de bonnes réponses et les temps de réaction observés. Afin de caractériser le système cognitif impliqué dans le traitement visuel des mots, les chercheurs souhaitent « isoler » autant que possible ce système lors des expériences et donc contrôler les autres paramètres expérimentaux qui ont également un effet sur la performance mesurée.

Les caractéristiques des sujets sont un premier élément à contrôler. On appelle habituellement « lecteur expert » un lecteur sain (adulte) ayant de bonnes capacités de lecture, par opposition au lecteur débutant ou au lecteur affecté par une pathologie, comme par exemple un déficit visuel ou une dyslexie.

Le second élément concerne les facteurs psycholinguistiques. Les auteurs établissent des listes de stimuli sélectionnées selon les facteurs d'intérêt (fréquence, longueur orthographique, voisinage, nombre de morphèmes, etc.). Ainsi, ils peuvent estimer la contribution des différents facteurs dans la prédiction des temps de réactions (TR) mesurés et déterminer, parmi ces facteurs, quels sont les meilleurs prédicteurs des performances observées. Durant les dernières décennies, l'approche des méga-études est devenue de plus en plus populaire. Cette approche consiste à collecter des données comportementales sur un grand nombre de sujets et à l'aide d'un très grand nombre de stimuli, allant jusqu'à des dizaines de milliers de mots. Même si des inquiétudes ont été soulevées quant à la stabilité et à la fiabilité relatives des bases de données recueillies (Seidenberg & Plaut, 1998 ; Sibley, Kello, & Seidenberg, 2009), les méga-études ont plusieurs avantages. Elles minimisent le risque de biais liés à l'utilisation de listes restreintes de stimuli et réduisent les incohérences entre les études comportementales.

De plus, certaines méga-études fournissent des TRs mesurés dans plusieurs tâches comportementales. Par exemple, l'ELP, « *English Lexicon Project* » (Balota et al., 2007), met à disposition des mesures recueillies en décision lexicale et en dénomination orale pour un large corpus de mots en anglais. En français, dans « *Chronolex* », Ferrand et al. (2011) fournissent les données pour trois tâches : la décision lexicale (LDT) qui consiste à déterminer si un stimulus est un mot ou un non-mot, la dénomination ou lecture à voix haute (NMG) où l'on demande au participant de prononcer le plus rapidement possible un stimulus et le *progressive demasking* (PDM) qui est une tâche de reconnaissance de mot dans laquelle le stimulus est graduellement affiché, ce qui permet de manipuler le décours temporel de la disponibilité de l'information visuelle.

Ceci nous amène au troisième élément à contrôler : la nature de la tâche expérimentale utilisée. En effet, les tâches mentionnées mobilisent, toutes les trois, les processus cognitifs de la reconnaissance visuelle de mots mais chacune implique des niveaux de traitements distincts. La

dénomination orale mobilise des aspects phonologiques et articulatoires qui sont (relativement) absents dans les deux autres tâches ; la tâche de décision lexicale repose spécifiquement sur les capacités de reconnaissance de mots et la tâche de PDM met l'accent sur les traitements visuels et orthographiques des mots (Carreiras, Perea, & Grainger, 1997 ; Grainger & Jacobs, 1996). Les tâches de LDT et de NMG sont fréquemment utilisées expérimentalement et ont fait l'objet de simulations dans la plupart des modèles de lecture (pour une revue, voir Norris, 2013). Tandis que la troisième, la tâche de PDM, est moins utilisée au niveau expérimental et n'a été exploitée, à notre connaissance, que dans le cadre du modèle MROM (Grainger & Jacobs, 1996).

La comparaison des performances obtenues dans ces trois tâches a permis d'identifier l'influence de facteurs psycholinguistiques communs[10], principalement la fréquence des stimuli dans la langue et leur longueur orthographique (en nombre de lettres). Cependant, si ces facteurs montrent des effets significatifs allant dans la même direction pour ces trois tâches, leur amplitude varie d'une tâche à une autre. Parallèlement, peu d'études ont examiné la capacité des modèles computationnels à rendre compte des effets observés sur des données issues de méga-études (Adelman & Brown, 2008 ; Perry, Ziegler, & Zorzi, 2010) et aucune étude, à notre connaissance, n'a évalué la capacité d'un modèle computationnel à simuler les données d'une méga-étude portant sur plusieurs tâches.

Parmi les facteurs psycholinguistiques mentionnés précédemment, l'effet de fréquence est un effet classique extrêmement stable inter-études et inter-langues (Balota & Chumbley, 1985 ; Brysbaert et al., 2011 ; Murray & Forster, 2004). Plus un mot est fréquent dans la langue, plus son traitement est rapide. Un mot très fréquent est donc lu et reconnu plus rapidement qu'un mot peu fréquent. La fréquence des stimuli est aussi le meilleur prédicteur, en termes de variance expliquée, des TRs moyens lors des tâches de reconnaissance visuelle de mots.

La longueur orthographique a été largement étudiée expérimentalement et a également un effet important sur les TRs moyens. Cet effet a la particularité de soulever des questions fondamentales sur la nature des processus impliqués dans le traitement des mots écrits, et notamment vis-à-vis de leur nature sérielle ou parallèle. Les méga-études se sont emparées de la question, montrant des effets particuliers (New, Ferrand, Pallier, & Brysbaert, 2006). En effet, l'effet de longueur observé en LDT sur les mots n'est pas linéaire en fonction de la longueur orthographique du stimulus. Plus spécifiquement, cet effet est de petite amplitude si l'on se limite aux mots de moins de 5 lettres mais de plus grande amplitude pour les mots plus longs. Par ailleurs, la méga-étude Chronolex (Ferrand et al., 2011) a permis de montrer que cet effet de longueur est présent dans les trois tâches proposées, mais présente des variations d'amplitude et de forme propres à chaque tâche. Ainsi, l'effet de longueur a une plus grande amplitude dans la tâche de PDM que dans la tâche de LDT. Il est de très petite amplitude dans la tâche de NMG et explique une part de variance très limitée dans cette tâche.

[10] Les effets du voisinage orthographique ou phonologique sont fréquemment étudiés expérimentalement. L'intérêt théorique de l'étude de ces effets est de montrer qu'ils sont la conséquence d'une compétition orthographique ou phonologique qui peut être soit facilitatrice soit inhibitrice lors du traitement (Grainger, Muneaux, Farioli, & Ziegler, 2005 ; Ziegler, Muneaux, & Grainger, 2003). Cependant, le voisinage explique très peu de variance en comparaison à la longueur et à la fréquence. C'est pour cette raison que nous ne l'avons pas pris en compte ici.

De plus, l'origine des effets de longueur pour les mots a toujours été objet de débat dans le domaine de la lecture et de la reconnaissance de mots (Coltheart & Rastle, 1994 ; New et al., 2006 ; Weekes, 1997). Cet effet a pu être simulé par certains modèles de lecture à haute voix mais n'a pas suscité beaucoup d'intérêt pour les modèles computationnels qui simulent la LDT, bien qu'il soit considéré comme l'un des prédicteurs majeurs des temps de réaction.

Les modèles double-voie de lecture à haute voix, tels que DRC (Coltheart et al., 2001) et CDP+ (Perry et al., 2007), attribuent les effets de longueur au traitement sériel du stimulus, et donc au processus de conversion graphème-phonème spécifique à la voie sous-lexicale. Par conséquent, dans ce cadre théorique, cette explication ne s'applique que pour le traitement des pseudo-mots ou des mots très peu fréquents.

Les modèles à voie unique ne fournissent pas une explication plus satisfaisante de cet effet. Plus spécifiquement, le modèle en Triangle ne peut pas en rendre compte du fait qu'il ne possède aucun mécanisme de traitement sériel. Le modèle MTM, quant à lui, ne génère un effet de longueur que lorsqu'il recourt à la procédure analytique et au déplacement sériel de la fenêtre attentionnelle, ce qui ne se produit qu'en cas d'échec du mode de traitement global, en l'occurrence lors de la lecture des pseudo-mots et des mots peu fréquents. MTM ne peut donc pas davantage rendre compte de l'effet de longueur en lecture de mots.

Les modèles mentionnés précédemment s'intéressent uniquement à la tâche de NMG. En ce qui concerne la tâche de LDT, aucun des modèles double-voie ou voie unique ne peut simuler l'effet de longueur de mots. Dans le modèle DRC, Coltheart et al. (2001) l'attribuent, comme pour la tâche de NMG, à l'influence de la voie sous-lexicale. Le modèle en Triangle, quant à lui, propose un moyen de simuler la tâche de LDT mais Seidenberg et McClelland (1989) reconnaissent que cette implémentation est très limitée et ne permet pas de rendre compte des performances des lecteurs. Enfin, Juphard et al. (2004) montrent que le modèle MTM ne rend pas compte de l'effet de longueur de mots en tâche de LDT chez les lecteurs experts.

Toutefois, une exception notable pour cette tâche concerne l'étude de Ginestet, Phénix, Diard, et Valdois (2019), dans laquelle l'effet de longueur de mots décrit dans le *French Lexicon Project* (Ferrand et al., 2010) pour des lecteurs experts est simulé dans le cadre du modèle de reconnaissance de mots BRAID. Dans leur étude, les auteurs montrent que les ressources visuo-attentionnelles disponibles sont à l'origine de cet effet. Plus précisément, une distribution attentionnelle (de forme gaussienne) induit un effet de longueur en LDT. Une distribution moins étendue de l'attention visuelle entraîne un effet de longueur plus important même si le traitement du stimulus est bien parallèle (la ressource attentionnelle ne se déplace pas pendant le traitement). Inversement, des ressources attentionnelles illimitées engendrent une disparition, voire une inversion de l'effet de longueur, et donc un traitement plus rapide pour les mots plus longs.

Etant donné le challenge que représente la simulation des effets de longueurs de mots et les questionnements sur l'origine cognitive de ces effets, dont la source est considérée comme phonologique dans certains modèles et visuo-attentionnelle dans d'autres, nous avons choisi d'évaluer la capacité du modèle BRAID-Phon à simuler les données comportementales disponibles pour les

trois tâches étudiées dans le cadre de la méga-étude Chronolex de Ferrand et al. (2011). Le modèle BRAID-Phon, grâce à son module phonologique, qui n'existait pas dans le modèle BRAID, nous permet pour la première fois de simuler l'ensemble des trois tâches (LDT, NMG et PDM). BRAID-Phon inclut les modules et les « briques » nécessaires pour cet objectif. Tout d'abord, le module d'appartenance lexicale permet de simuler la tâche de LDT et le module phonologique permet de simuler la dénomination de mots. Ensuite, grâce à la spécification détaillée des processus visuels, BRAID-Phon permet également la simulation de la tâche de PDM, comme une variante de la tâche de reconnaissance visuelle de mots, dans laquelle on adapte le décours temporel de l'apparition du stimulus visuel.

Dans ce qui suit, nous présenterons les temps de réponses simulés pour les trois tâches par le modèle BRAID-Phon. Nous procéderons en deux étapes. Dans un premier temps, nous évaluerons la capacité du modèle à rendre compte des effets observés dans les trois tâches. Nous imposons la contrainte forte que toutes les simulations utilisent un même jeu de paramètres, correspondant aux valeurs par défaut des paramètres du modèle. Nous évaluerons la capacité du modèle à reproduire fidèlement les effets de longueur et de fréquence des mots. Le fait d'utiliser les paramètres par défaut permet d'étudier le cas dans lequel le modèle n'adopterait aucune stratégie « tâche-spécifique ».

Dans un second temps, nous nous concentrerons sur l'étude de l'effet de longueur. L'objectif est d'étudier l'impact de la distribution visuo-attentionnelle sur cet effet dans les différentes tâches. Nous explorerons donc dans quelle mesure les variations des paramètres attentionnels modulent les effets de longueur selon la tâche. D'une part, nous étudierons l'impact de la position du focus attentionnel (et donc du regard), sans modification de la dispersion attentionnelle. Les déplacements du focus attentionnel peuvent être interprétés comme des stratégies d'exploration visuo-attentionnelle du stimulus, possiblement adaptées aux contraintes de chaque tâche. On évaluera si certaines stratégies visuo-attentionnelles rendent plus fidèlement compte des données observées et, le cas échéant, si une stratégie unique est la meilleure ou si, au contraire, la meilleure stratégie est de s'adapter aux contraintes spécifiques de chaque tâche. Enfin, nous étudierons l'impact de trois positions distinctes du focus attentionnel sur l'effet de longueur dans chacune des trois tâches. Nous vérifierons également les conséquences de deux distributions atypiques de l'attention visuelle sur cet effet.

2 Simulations A : Simulation des effets de longueur et de fréquence

2.1 Objectifs

L'objectif ici est d'évaluer la capacité de BRAID-Phon à rendre compte de l'amplitude des effets de fréquence et de longueur dans les trois tâches d'intérêt. Cette section contient quatre sous-parties, dans lesquelles nous décrirons, successivement et respectivement, les données comportementales, la procédure de simulation, les analyses statistiques employées et les résultats obtenus.

2.2 Données Comportementales

L'étude Chronolex (Ferrand et al., 2011) fournit une base de données de temps de réponse pour 1 482 mots français et pour les trois tâches d'intérêt. Chacun de ces TRs correspond à la moyenne des temps de réponse recueillis auprès de 105 participants. Nous notons que cette base de données ne comporte que des mots monosyllabiques et monomorphémiques de différentes longueurs et que la distribution des mots en fonction de leur longueur n'est pas une distribution uniforme dans la base de données, conformément à la distribution des longueurs dans la langue. En effet, il y a relativement peu de mots monosyllabiques monomorphémiques de longueur 2, 7 et 8 en comparaison aux autres longueurs. Les effectifs pour chaque longueur sont donnés dans le Tableau 3.1.

Tableau 3.1 – **Nombre de mots pour chaque longueur dans la base de données des stimuli de l'expérience Chronolex (Ferrand et al., 2011).**

Longueur	2	3	4	5	6	7	8
Effectif	36	165	405	506	304	61	5

2.3 Procédure

Le formalisme probabiliste dans lequel nous définissons BRAID-Phon permet de simuler les trois tâches en calculant des inférences bayésiennes sur une seule et unique structure de connaissance. Ces calculs fournissent les valeurs numériques des distributions de probabilité pour chaque itération simulée. On représente les résultats de ces calculs sous la forme de courbes d'évolution des distributions de probabilité dans le temps, comme nous l'avons illustré dans le Chapitre 2.

Le modèle BRAID a été préalablement calibré par Phénix (2018) afin qu'une itération simulée corresponde à une milliseconde de temps de traitement humain. Afin d'obtenir des temps de réponses à partir des courbes de probabilités, nous appliquons une procédure de prise de décision simple qui consiste à fixer une valeur de probabilité comme seuil de décision. Le temps de réponse du modèle est le nombre d'itérations nécessaires pour que la courbe de probabilité atteigne ce seuil.

Ginestet et al. (2019) ont défini, dans une étude qui porte sur la simulation des données du *French Lexicon Project* (Ferrand et al., 2010), un seuil de décision pour la tâche de décision lexicale que nous notons τ_{LDT} égal à 0,9. Cette valeur est choisie de manière à minimiser une mesure multicritères, basée sur l'erreur quadratique moyenne entre les temps comportementaux et simulés, sur la différence de variance des temps comportementaux et simulés et, enfin, sur la comparaison des taux de réponse comportementaux et simulés. Par simplicité et parcimonie, nous utilisons la même valeur $\tau_{\text{LDT}} = 0{,}9$ pour la simulation de la décision lexicale dans la présente étude.

Les deux autres tâches (NMG et PDM) n'ayant jamais été simulées précédemment avec le modèle BRAID, nous avons choisi de suivre une méthode analogue à celle de Ginestet et al. (2019) pour définir les seuils de décision en NMG et PDM. La méthode consiste à simuler ces tâches, en utilisant les stimuli de la base de données Chronolex, pour un nombre d'itérations suffisamment long : 1 000 pour la tâche de NMG et 3 000 pour la tâche de PDM. Nous choisissons ensuite des

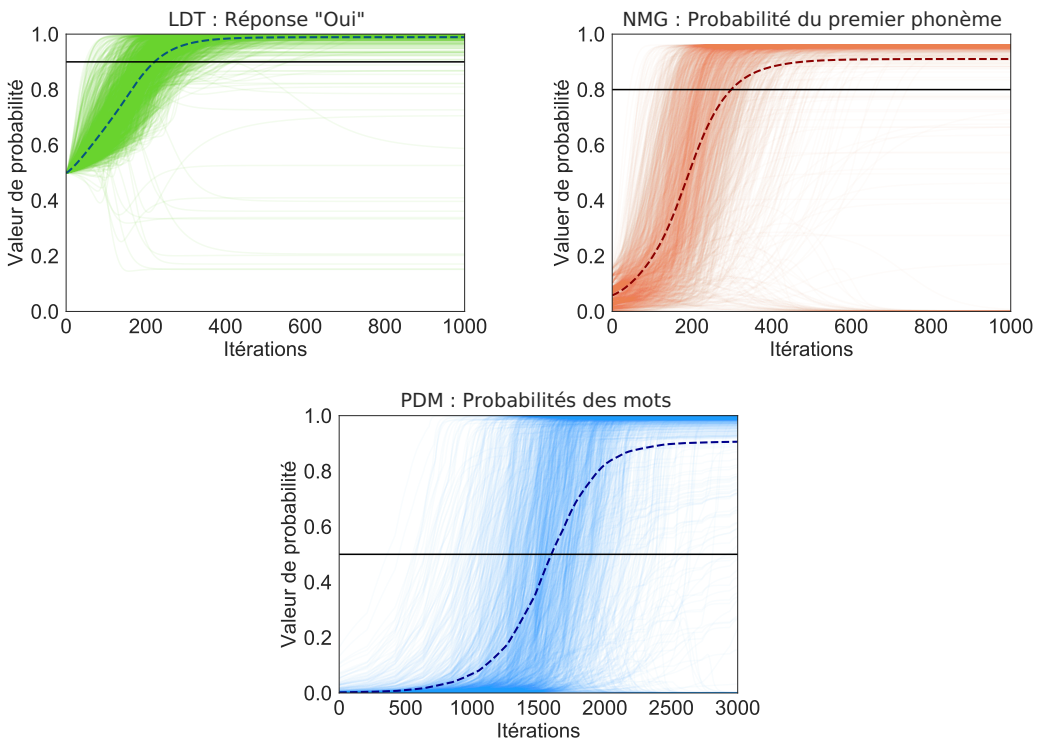


FIGURE 3.1 – **Courbes de probabilités issues de l'ensemble des simulations effectuées.** Les résultats pour la tâche de décision lexicale (LDT) sont présentés en haut à gauche (courbes de la probabilité que le stimulus présenté soit un mot connu), ceux de la dénomination (NMG) en haut à droite (courbes de la probabilité du premier phonème) et ceux de le tâche de *progressive demasking* (PDM) en bas (courbes de la probabilité du mot). Chaque graphique superpose les courbes correspondant à l'évolution de probabilité associée à chaque item du corpus Chronolex. Les courbes noires en pointillés correspondent à la courbe moyenne obtenue par tâche pour l'ensemble des items. La ligne horizontale noire indique la valeur de seuil choisie pour chacune des trois tâches.

valeurs de seuils qui permettent de minimiser l'erreur quadratique moyenne entre les TRs mesurés et simulés et le taux d'erreur du modèle. Cette méthode conduit à choisir comme valeurs $\tau_{\text{NMG}} = 0{,}8$ pour la tâche de NMG et $\tau_{\text{PDM}} = 0{,}5$ pour la tâche de PDM. La Figure 3.1 montre, d'une part, l'ensemble des courbes de probabilité relatives aux simulations effectuées et, d'autre part, les valeurs de seuils choisies pour chaque tâche simulée.

Nous conservons trois valeurs de seuil de décision différentes, pour des raisons d'adéquation empirique, mais aussi pour des raisons mathématiques. En effet, les domaines des variables d'intérêt ont des tailles différentes pour chaque tâche : la variable d'appartenance lexicale D^T est Booléenne et n'a donc que 2 valeurs possibles, alors qu'il y en a 40 pour la variable du premier phonème φ_1^T et 92 117 pour la variable du mot W^T. Ainsi, ces espaces mathématiques sont très différents, et il serait peu probable que les processus de décisions reposent sur une valeur unique, indépendamment de l'espace considéré.

Enfin, pour la calibration du reste des paramètres du modèle, nous utilisons exactement les valeurs par défaut des paramètres du modèle BRAID, et, comme dans la plupart des expériences précédentes avec ce modèle, nous supposons que la position du regard g^T et du focus attentionnel

μ_A^T coïncident, ce qui est par ailleurs justifié par le design expérimental.

Une réponse « correcte » du modèle correspond au fait que la courbe de probabilité de l'élément attendu (la réponse « oui » en LDT, le premier phonème du stimulus en NMG, et le mot cible en PDM) a atteint le seuil de décision. Quand un autre élément atteint ce seuil ou quand aucun élément n'atteint ce seuil, nous considérons la réponse comme « incorrecte ». Dans le cas du NMG, nous vérifions également que les phonèmes cibles sont ceux avec la probabilité la plus élevée dans chaque position phonologique.

2.4 Analyses statistiques

Afin de comparer les TRs simulés aux TRs comportementaux, nous effectuons des régressions linéaires pour chaque tâche, en utilisant deux prédicteurs : la fréquence et la longueur des mots. Pour chaque tâche simulée, seuls les mots dont la réponse du modèle est considérée correcte sont pris en compte dans l'analyse statistique correspondante.

Nous utilisons un modèle linéaire pour analyser les données de chacune des tâches. Celui-ci peut être formulé ainsi :

$$RT \sim intercept + LogFreq + Longueur$$

avec *LogFreq* qui correspond au logarithme décimal de la fréquence des mots. Le nombre d'itérations requis pour atteindre le seuil est considéré comme un indicateur des temps de traitement comparable aux TRs comportementaux.

De plus, il est classique de prendre également en compte la nature du premier phonème comme prédicteur supplémentaire lors de l'analyse des TRs expérimentaux en NMG. En effet, les temps obtenus expérimentalement en NMG sont mesurés à partir de l'enregistrement audio des productions orales des participants. Ils correspondent au temps écoulé entre la présentation du mot à l'écran et le début du signal de parole correspondant à sa prononciation. Il est largement admis que la nature du premier phonème est une source de variabilité importante dans ce type de données, puisque la détection du début de prononciation par l'instrument de mesure (clé vocale ou micro) varie selon que le premier phonème soit une (semi)voyelle ou une consonne et selon la nature du phonème consonantique (e.g., occlusive, fricative, labiale). Il est également probable que les processus moteurs de prononciation, comme la planification du geste articulatoire ou sa réalisation, soient aussi dépendants du premier phonème à prononcer. A titre d'illustration, la prononciation d'un mot commençant par un *cluster* consonantique complexe (par exemple, le mot « strict ») paraît plus difficile que celle d'un mot commençant par une voyelle. Plusieurs études se sont intéressées à l'impact de la nature du premier phonème sur les TRs pour la tâche de NMG et ont proposé d'intégrer des codages particuliers de cette information dans les analyses de régression (Spieler & Balota, 1997 ; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). L'ensemble des études analysant les temps de réactions de longues listes de mots, telles que Chronolex ou encore l'étude de Balota, Cortese, Sergent-Marshall, Spieler, et Yap (2004) sur les mots monosyllabiques en anglais, en ont fait l'usage.

Ce codage consiste à coder, pour chaque mot, la nature du phonème initial de manière dichotomique (0 ou 1, *dummy coding*) selon 13 caractéristiques phonologiques (*phonological features*) : occlusive, fricative, bilabiale, dentale, vélaire, labiodentale, alvéolaire, palatale, nasale, liquide, approximante, uvulaire et voisée. Une valeur 0 indique que le premier phonème du mot en question n'a pas cette caractéristique et une valeur 1 indique qu'il l'a. Par exemple, le mot « verre » commence par le phonème /v/ qui est fricatif, labiodental et voisé. Cet item est codé avec la valeur 1 dans les champs qui correspondent à ces trois caractéristiques et une valeur 0 partout ailleurs. On note que les mots commençant par une voyelle sont codés par la valeur 1 dans le champ « voisé » et la valeur 0 ailleurs. Le modèle linéaire utilisé pour analyser les données comportementales de NMG devient donc :

$$RT_{NMG} \sim intercept + Onset + LogFreq + Longueur$$

avec le symbole *Onset* qui réfère à l'ensemble des 13 variables décrivant le phonème initial du mot (*onset phoneme*).

2.5 Résultats

BRAID-Phon réussit à traiter correctement 98,5% des mots dans la tâche de LDT, 88,5% en NMG et 90,7% en PDM. La Figure 3.2 montre les droites de régression correspondant respectivement à l'effet de fréquence et à l'effet de longueur, et permettent de comparer ces effets entre données comportementales et données simulées selon les tâches. Nous comparerons quantitativement, dans les paragraphes qui suivent, les effets de fréquence et de longueur dans les données de BRAID-Phon et dans celles de Chronolex, pour chacune des trois tâches. Nous baserons nos comparaisons, d'une part, sur l'amplitude des effets qui correspond aux pentes des droites de régression et, d'autre part, sur les parts de variance expliquée. Nous utiliserons le coefficient de détermination des régressions linéaires R^2 pour quantifier la part de variance expliquée par le modèle dans sa globalité. Afin de quantifier la part de variance qu'expliquerait chaque prédicteur, nous calculerons, dans chaque régression linéaire, des coefficients de détermination partiels, avec la formule suivante : $R^2_{partiel} = (SSE_{total} - SSE_{réduit})/(SSE_{réduit})$, avec SSE_{total} la somme des carrés des erreurs du modèle linéaire avec tous les prédicteurs et $SSE_{réduit}$ la somme des carrés des erreurs du modèle avec tous les prédicteurs à l'exception de celui que l'on souhaite isoler et dont on souhaite calculer la contribution (Kutner, 2005).

Décision lexicale Dans la tâche de LDT, les deux facteurs de longueur et de fréquence expliquent 39% de la variance pour les données simulées ($F(2,1458) = 316,4$; $p < 0,001$) ce qui est très proche des 41% de variance expliquée au niveau comportemental ($F(2,1458) = 524,6$; $p < 0,001$). Le modèle produit un effet de fréquence dont l'amplitude, -30 itérations par log(ppm), est proche de celle qui est observée dans les données de Chronolex : -38 ms par log(ppm). Nous observons également le même ordre de grandeur pour l'effet de longueur : 15 itérations par lettre dans la simulation et 8 ms par lettre dans les données Chronolex. Les coefficients de détermination (R^2) partiels correspondants sont également similaires : 0,20 dans la simulation contre 0,34 dans les données comportementales pour l'effet de fréquence et 0,12 contre 0,03 pour l'effet de longueur.

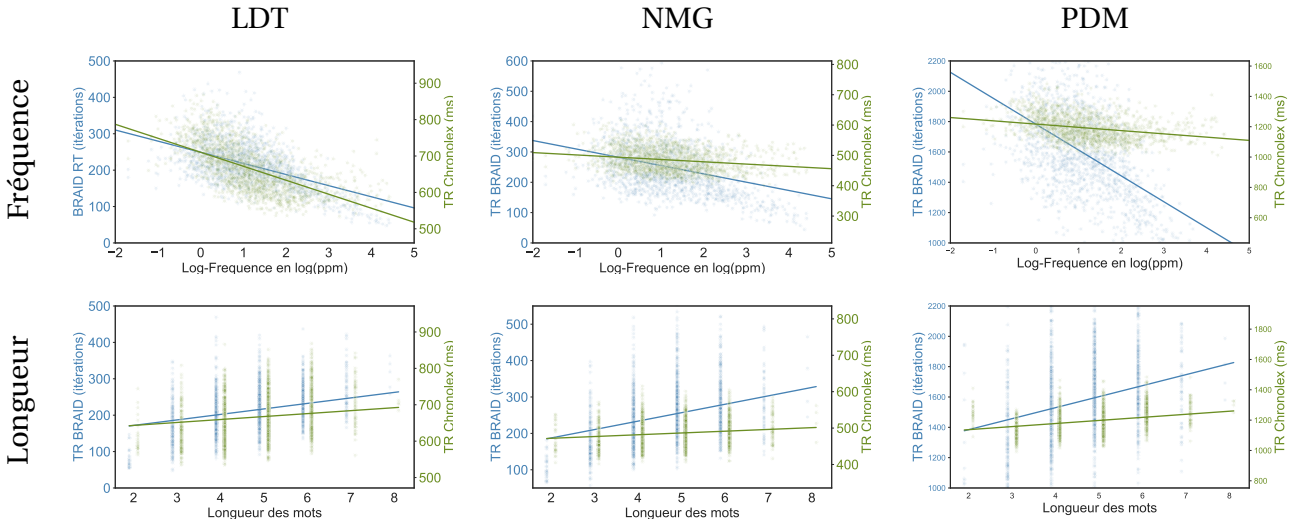


FIGURE 3.2 – **Résultats des régressions linéaires à l'issue des simulations A.** Effet de fréquence (ligne du haut) et de longueur (ligne du bas) sur les TRs simulés par BRAID-Phon (en bleu) et les TRs de Chronolex (en vert), pour les tâches de décision lexicale (LDT, à gauche), dénomination (NMG, centre) et *progressive demasking* (PDM, à droite). Chaque point d'un nuage de points correspond au TR d'un mot de la base de données (résultats présentés item par item, aussi appelé *item-level study* en anglais). Pour faciliter la comparaison des pentes, les axes des ordonnées des TRs simulés sont décalés d'une constante, pour chaque tâche. Cette constante permet de faire coïncider les deux droites de régressions en un point de référence (0 pour la fréquence et 2 pour la longueur).

Dénomination Nous rappelons que nous utilisons la nature du premier phonème du stimulus uniquement lors des analyses des données comportementales en NMG. Ce prédicteur n'est pas pris en compte dans les analyses des données simulées. Pour Chronolex, les deux facteurs et la nature du premier phonème expliquent 53% de la variance ($F(14,1379) = 113$; $p < 0{,}001$) des TRs du NMG. Tandis qu'avec BRAID-Phon la fréquence et la longueur expliquent 24% de la variance des données simulées en NMG ($F(2,1391) = 220{,}4$; $p < 0{,}001$). La pente de l'effet de fréquence est de -27,5 itérations par log(ppm) dans la simulation contre -7,5 ms par log(ppm) dans les données Chronolex. La pente de l'effet de longueur est de 23,1 itérations par lettre contre 4,9 ms par lettre dans les données Chronolex. Pour l'effet de fréquence, les R^2 partiels sont de 0,1 en simulation contre 0,06 dans les données comportementales, et sont respectivement de 0,08 contre 0,03 pour l'effet de longueur. Nous notons que le codage de la nature du premier phonème explique, à lui seul, 49% de la variabilité des TRs en NMG de Chronolex.

Progressive demasking En PDM, les deux prédicteurs que sont la longueur et la fréquence expliquent 38,5% de la variance du nombre d'itérations simulé ($F(2,1354) = 425{,}9$; $p < 0{,}001$) alors qu'ils rendent compte de 19,3% de la variance dans Chronolex ($F(2,1354) = 163{,}2$; $p < 0{,}001$). Les régressions fournissent de grandes valeurs de pente pour les données de simulation : -170,7 itérations par log(ppm) pour l'effet de fréquence et 72,53 itérations par lettre lorsqu'il s'agit de l'effet de longueur (-21,5 et 20,3 respectivement dans les données comportementales). Pour cette tâche, les R^2 partiels des effets de longueur sont égaux (et égaux à 0,07) dans les simulations et dans les données Chronolex. Le R^2 partiel de l'effet de fréquence est de 0,26 en données simulées contre 0,07 dans les données expérimentales.

2.6 Synthèse des résultats des simulations

Ce premier ensemble de simulations avait pour objectif de rendre compte des principaux effets de fréquence et de longueur des mots dans trois tâches différentes de traitement visuel de mots, tout en utilisant un seul modèle computationnel et les mêmes valeurs de paramètres dans les trois tâches. BRAID-Phon parvient à simuler les deux effets étudiés dans la même direction et avec le même ordre de grandeur que ceux observés dans les données comportementales de l'étude Chronolex. Dans la tâche de LDT, les effets de fréquence et de longueur sont reproduits fidèlement par le modèle computationnel. Toutefois, l'amplitude des effets de longueur et de fréquence simulés est plus importante que celle des effets comportementaux en NMG et PDM.

3 Simulations B : Impact des paramètres attentionnels sur les effets de longueur

Dans la Section 2, les simulations ont montré que BRAID-Phon rendait fidèlement compte des effets de longueur et de fréquence pour la tâche de LDT et, au moins qualitativement, pour les tâches de NMG et de PDM. Dans cette section, nous nous focaliserons sur l'effet de longueur afin de mieux comprendre l'impact de la distribution de l'attention visuelle sur cet effet. Pour cela, les simulations seront effectuées en manipulant soit la position du focus attentionnel (paramètre μ_A), soit la dispersion de la distribution attentionnelle (paramètre σ_A). L'objectif est, d'une part, d'évaluer dans quelle mesure la position du focus attentionnel influe sur la forme et l'amplitude de l'effet de longueur (simulations B1), lorsque la valeur par défaut de la dispersion de la distribution attentionnelle reste inchangée. D'autre part, nous étudierons l'impact de deux distributions attentionnelles dispersées de manière atypique; nous considérerons une distribution soit très concentrée (distribution piquée) soit très dispersée (distribution uniforme), et étudierons les effets de longueur qui en résultent pour les trois types de tâches (simulations B2).

3.1 Simulations B1

3.1.1 Objectifs

Ce premier jeu de simulations a pour objectif d'explorer l'effet de la position du focus attentionnel sur les temps de réponse. Pour cela, nous comparons trois conditions qui diffèrent par la position du regard et du focus attentionnel. Dans la première condition (*Centre*), elles sont fixées au centre du stimulus ($\mu_A = g = (N+1)/2$ avec N la longueur du stimulus). Cette condition est identique à celle évaluée précédemment dans les simulations A. Dans la seconde condition, la position du regard et de l'attention est décalée légèrement à gauche du centre (condition *Précentre*, $\mu_A = g = (N-1)/2$). Enfin, dans la troisième condition, la position du regard et de l'attention est fixée entre la deuxième et la troisième lettre du stimulus (condition *2,5*, $\mu_A = g = 2{,}5$). Cette dernière condition correspond au cas dans lequel l'attention porterait approximativement au milieu de la première syllabe. Ainsi, la position du focus attentionnel dans les deux premières conditions dépend de la longueur du stimulus mais elle demeure constante quelque soit la longueur dans la troisième

condition. Pour toutes ces conditions, tous les autres paramètres ont leur valeur par défaut. En particulier, le paramètre de la dispersion attentionnelle a sa valeur habituelle ($\sigma_A = 1{,}75$).

3.1.2 Procédure

La procédure de simulation est identique à celle suivie pour les simulations précédentes. Le seul élément qui change dans ce qui suit est la position du regard et du focus attentionnel (voir Tableau 3.2).

Tableau 3.2 – **Récapitulatif des paramètres de la position du regard et du focus visuo-attentionnel pour les conditions explorées dans les simulations B1 et B2.**

	Simulations B1			Simulations B2	
Conditions	Centre	Précentre	*2,5*	AV. Uniforme	AV. Réduite
$\mu_A = g$	$\frac{N+1}{2}$	$\frac{N-1}{2}$	2,5	$\frac{N+1}{2}$	*lettre à lettre*
σ_A	1,75	1,75	1,75	1000	0,5

3.1.3 Analyses statistiques

Afin de comparer la forme et l'amplitude de l'effet de longueur simulé à celles de l'effet observé expérimentalement dans l'étude Chronolex, nous effectuons une analyse en deux temps, avec tout d'abord une analyse qualitative des effets simulés et observés et ensuite une comparaison quantitative.

Dans la première étape, nous effectuons des régressions *spline* pour chaque tâche et pour chaque condition en utilisant deux prédicteurs : la fréquence et la longueur des mots. Nous utilisons pour cela des modèles additifs généralisés (*Generalized Additive Models*, GAM), introduits par Hastie et Tibshirani (1990) et qui sont une extension des modèles linéaires généralisés. Essentiellement, les modèles GAM correspondent à des modèles linéaires dont les termes sont une somme de fonctions lisses des prédicteurs utilisés. Les modèles additifs généralisés permettent d'ajuster une courbe non linéaire à des données expérimentales et fournissent donc une estimation fine de la forme de l'effet. En outre, ces modèles offrent une grande souplesse et peuvent être utilisés d'une manière similaire à la régression linéaire (Wood, 2006). Tous les modèles sont implémentés en `R` en utilisant le package `mgcv` (Wood, 2019).

Par ailleurs, les modèles GAM ont été utilisés dans plusieurs disciplines qui s'appuient sur des données temporelles, dont la psychologie, lors de l'analyse de données longitudinales ou continues (McKeown & Sneddon, 2014 ; Sullivan, Shadish, & Steiner, 2015). Des analyses similaires ont été employées dans l'étude de Ferrand et al. (2011) pour déterminer la forme des effets étudiés.

Nous nous intéressons dans cette section à l'étude de l'effet de longueur entre les différentes conditions. Nous avons inclus la fréquence des mots dans l'analyse statistique afin de prendre en compte la grande part de variance qu'elle explique ainsi que sa forte corrélation avec la longueur des mots. Par conséquent, les modèles GAM utilisés pour analyser les temps de réponse peuvent

être exprimés de la façon suivante :

$$RT \sim intercept + s(LogFreq) + s(Longueur) ,$$

où *LogFreq* correspond au logarithme décimal de la fréquence et $s(\cdot)$ représente la fonction *spline*. De manière analogue aux analyses effectuées précédemment, l'analyse de régression des données Chronolex de NMG comprend également, comme prédicteur, la nature du premier phonème. Nous écrivons :

$$RT \sim intercept + Onset + s(LogFreq) + s(Longueur) .$$

Dans la seconde étape de l'analyse, nous effectuons une régression linéaire entre les temps de réponses prédits et observés pour chaque longueur. Cette analyse se base sur les courbes ajustées dans la première étape[11] et nous permettra de quantifier la ressemblance entre les effets de longueur prédits par BRAID-Phon dans chaque condition et les effets de longueur observés dans les données Chronolex. Cette comparaison prend en compte les valeurs de trois paramètres pour chaque condition : (1) le coefficient de détermination qui quantifie la qualité de la régression ; (2) la valeur de la pente qui mesure le rapport entre l'effet de longueur observé expérimentalement (en millisecondes) et l'effet simulé (en itérations) ; et (3) la valeur de l'intercept qui indique l'écart à l'origine entre les itérations du modèle et les TRs expérimentaux.

Les deux derniers paramètres nous permettent de mieux caractériser la relation entre les données expérimentales et les données simulées[12] : la valeur de la pente permet de vérifier si l'on peut considérer que l'effet reproduit par le modèle computationnel est calibré temporellement avec les données observées, autrement dit, si une itération de simulation conserve son équivalence avec une milliseconde de TR expérimental. La valeur de l'intercept, quant à elle, fournit une constante qu'il faudrait ajouter pour ajuster l'effet observé et l'effet simulé. Cet constante est un écart à l'origine que nous interprétons comme un coût temporel constant des processus non pris en compte par le modèle computationnel, comme par exemple des temps liés aux processus moteurs et articulatoires pour produire la réponse comportementale.

[11] A première vue, il pourrait sembler inapproprié d'utiliser les valeurs des courbes ajustées pour effectuer une régression linéaire dans le but de comparer les effets prédits et observés. Cependant, les effectifs des différentes catégories de longueur des stimuli ne sont pas équilibrés (voir Tableau 3.1) et, par conséquent, utiliser des TRs moyens « bruts » pourrait biaiser les résultats.

[12] Les valeurs de pente et d'intercept permettent de mieux caractériser la régression puisqu'elles signalent certains cas problématiques dont le coefficient de détermination R^2 ne rend pas compte. Par exemple, la régression linéaire peut produire un R^2 très élevé et en même temps un pente proche de 0, ce qui indique une relation, certes linéaire, mais qui ressemble plus à une droite horizontale. On peut citer comme autre exemple, la comparaison de deux cas dans lesquels on obtiendrait des valeurs de R^2 élevées mais avec, dans un cas, un intercept correspondant à une valeur cohérente positive constante en harmonie avec ce qui est attendu (de l'ordre de 500, si on veut l'interpréter comme un temps de traitement moteur) et, dans l'autre cas, une valeur d'intercept négative ou dans un ordre de grandeur inconsistant, c'est-à-dire très différent de celui attendu.

3.1.4 Résultats

Pourcentages de réponses correctes BRAID-Phon répond correctement pour environ 98% des stimuli en LDT quelle que soit la position du focus attentionnel étudiée, alors que dans les données comportementales de l'étude Chronolex ce taux est de 90,6%. En NMG, le modèle produit correctement la prononciation de 88,5% des mots dans la condition *Centre*, 74,6% dans la condition *Précentre* et 87,6% dans la condition *2,5*. Le pourcentage de prononciations correctes dans l'étude originale est de 92,5%. BRAID-Phon reconnaît correctement 90,6% de l'ensemble des mots en PDM dans la condition *Centre*, 94,3% dans la condition *Précentre* et 87,6% dans la condition *2,5* alors que dans les données comportementales ce taux est de 98%. Les taux de réponses correctes pour les données simulées et comportementales sont montrés dans la Figure 3.3.

En LDT, les erreurs du modèle concernent principalement des mots courts (2 ou 3 lettres) avec des fréquences basses comme, par exemple, les mots « ras », « su » et « nu ». En PDM et en NMG, on observe soit des erreurs similaires à celles observées en LDT soit des erreurs causées par la présence d'un voisin orthographique plus fréquent que la cible. Par ailleurs, nous précisons que le taux de réponses correctes dans chaque tâche des données Chronolex est obtenu en soustrayant le pourcentage de réponses incorrectes (erreurs) et le pourcentage des réponses à valeurs aberrantes (*outliers*).

Comparaison qualitative des effets de longueur Pour la tâche de LDT, les régressions effectuées à l'aide des modèles GAM indiquent que les temps de réponse simulés dans toutes les conditions, ainsi que ceux des données Chronolex, montrent des effets significatifs des termes correspondants à la longueur et à la fréquence (Log-fréquence) des mots. De manière qualitative, la Figure 3.4 montre que la condition *Centre*, dans laquelle le focus attentionnel est au centre du stimulus, est

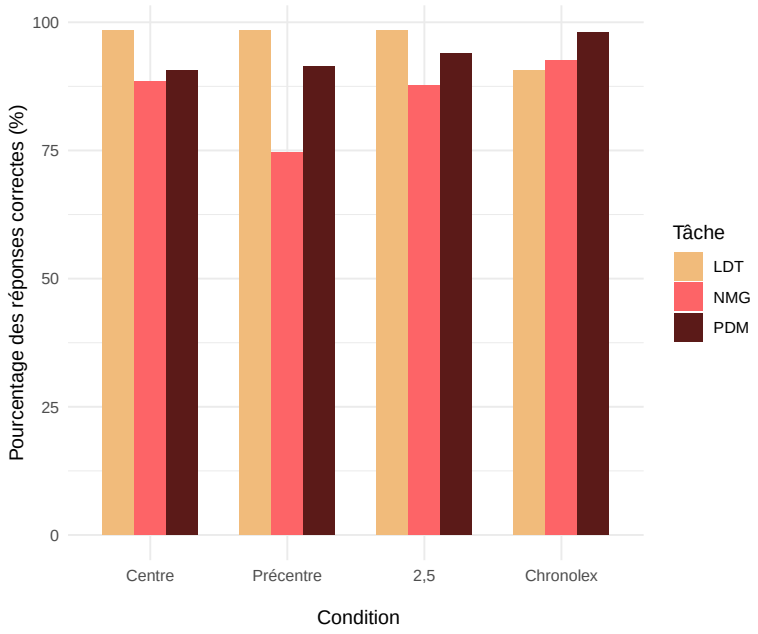


FIGURE 3.3 – **Pourcentages de réponses correctes de BRAID-Phon pour chaque condition des simulations B1 et pour chaque tâche simulée, comparés aux pourcentages de l'étude Chronolex.**

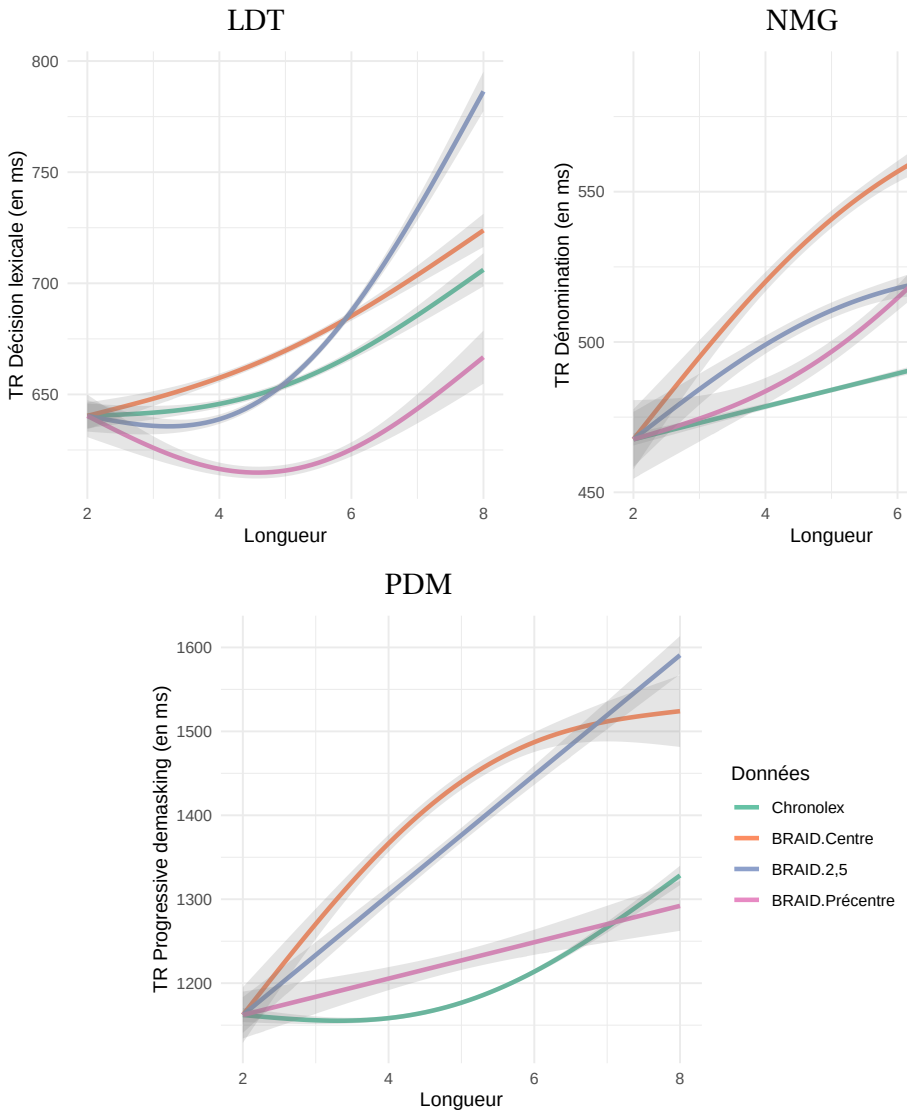


FIGURE 3.4 – **Effet de longueur dans les données comportementales de l'expérience Chronolex et dans les données simulées (simulations B1).** Les courbes des données simulées sont ajustées de sorte que le temps de réponse pour la longueur 2 soit égal à celui des données expérimentales. L'axe des ordonnées représente donc des TRs en millisecondes pour les données Chronolex et des itérations ajustées pour les données simulées. Les « surfaces » d'erreur (bandes grises) correspondent à l'erreur standard des modèles de régression GAM.

celle qui reproduit le plus fidèlement l'effet de longueur observé dans les données Chronolex en tâche de LDT. Dans la tâche de NMG, la condition *2,5* est celle dont la courbe se rapproche le plus de celle de l'effet de longueur comportemental. Néanmoins, l'amplitude de l'effet simulé est légèrement plus grande que celle de l'effet observé. La forme de l'effet simulé pour cette condition atteint un palier pour les mots longs (7 et 8 lettres) tandis que l'effet observé est assez linéaire. Dans la tâche de PDM, les conditions *Centre* et *2,5* montrent des effets simulés ayant des amplitudes plus importantes que celles de l'effet mesuré expérimentalement. Seule la condition *Précentre* produit un effet d'une amplitude similaire. La forme de l'effet observé n'est cependant pas reproduite : l'effet de longueur est assez linéaire pour cette condition de simulation mais il possède une forme d'apparence quadratique dans les données Chronolex.

Tableau 3.3 – **Résultats des régressions linéaires entre les effets de longueur prédits et observés dans chacune des trois tâches et pour chaque conditions de manipulation du focus attentionnel (Simulations B1) ou de la dispersion de l'attention (Simulations B2).**

(a) LDT

	Simulations B1			Simulations B2	
	Centre	Précentre	*2,5*	AV.Uniforme	AV.Réduite
Intercept	477,4	389,4	554,5	483,1	561,2
Pente	0,80	1,04	0,43	0,79	0,36
R^2	0,98	0,58	0,98	0,96	0,99

(b) NMG

	Simulations B1			Simulations B2	
	Centre	Précentre	*2,5*	AV.Uniforme	AV.Réduite
Intercept	408,72	395,36	351,7	402,6	626,32
Pente	0,29	0,33	0,53	0,33	-0,72
R^2	0,96	0,96	0,92	0,96	0,64

(c) PDM

	Simulations B1			Simulations B2	
	Centre	Précentre	*2,5*	AV.Uniforme	AV.Réduite
Intercept	621,9	-940,0	563,2	700,4	718,2
Pente	0,36	1,27	0,39	0,30	0,32
R^2	0,55	0,82	0,82	0,68	0,26

Comparaison quantitative de l'effet de longueur Les paramètres des régressions linéaires entre les données comportementales et les données simulées sont présentés dans le Tableau 3.3a. Pour la tâche de LDT, le coefficient de détermination est très élevé pour la condition *Centre* avec un bon rapport millisecondes/itérations (pente égale à 0,80). Ce rapport est cependant faible dans la condition *2,5* (pente égale à 0,42), bien que l'effet simulé est très corrélé avec l'effet observé ($R^2 = 0{,}98$). Quant à la condition *Précentre*, elle se caractérise par un faible R^2 (égal à 0,58) entre les prédictions et les observations.

En NMG, l'effet simulé présente une forte corrélation avec celui observé dans les données Chronolex ($R^2 = 0{,}96$ dans *Centre* et *Précentre* et $R^2 = 0{,}92$ dans la condition *2,5*). Cependant, cet effet est de plus grande amplitude dans les simulations que dans les données comportementales puisque toutes les pentes sont inférieures à 1. La condition *2,5* est celle qui s'adapte le plus aux données Chronolex avec une pente égale à 0,53 (voir Tableau 3.3b).

Les régressions linéaires pour les simulations de la tâche de PDM (Tableau 3.3c) indiquent des coefficients de détermination élevés dans les conditions *Précentre* et *2,5* ($R^2 = 0{,}82$). Les données simulées et observées sont moins corrélées dans la condition *Centre* ($R^2 = 0{,}55$). La condition *Précentre* est celle qui produit le meilleur rapport millisecondes/itérations (pente égale à 1,27).

En LDT et en NMG, les valeurs d'intercept sont relativement constantes pour les trois conditions : environ 450 millisecondes pour la tâche de LDT et 400 millisecondes pour la tâche de NMG. En tâche de PDM, la valeur de l'intercept est de l'ordre de 600 millisecondes pour les conditions *Centre* et *2,5* mais négative dans la condition *Précentre* (voir Tableau 3.3c).

3.2 Simulations B2

3.2.1 Objectifs

Dans cette section, nous évaluons les conséquences sur les effets de longueur de deux modifications extrêmes de distribution de l'attention (voir Tableau 3.2, Simulations B2). Dans la première condition, nous choisissons une distribution de l'attention dispersée au maximum, ce qui se traduit mathématiquement par une distribution de probabilité uniforme (*AV.Uniforme*), que nous approximons par une distribution gaussienne de très grande variance. Notons que, dans ce cas, la position du regard et du focus visuo-attentionnel ne peut plus être interprétée (quelles que soient ces positions, la distribution attentionnelle est identique), et nous la fixons arbitrairement au centre du stimulus ($\mu_A = g = \frac{N+1}{2}$). Dans la seconde condition, nous simulons une distribution de l'attention extrêmement concentrée, de sorte que la ressource attentionnelle soit presque entièrement allouée à une seule lettre à la fois. Dans cette condition, le focus attentionnel et le regard parcourent le stimulus, en se déplaçant (de manière coïncidente) de lettre en lettre (*AV.Réduite*), toutes les 100 itérations, jusqu'à ce que le seuil de décision de la tâche soit atteint.

3.2.2 Procédure et Analyses statistiques

Nous suivons pour les simulations B2 la même procédure que précédemment et appliquons les mêmes analyses statistiques.

3.2.3 Résultats

Pourcentages de réponses correctes BRAID-Phon reconnaît correctement 98% et 97% de l'ensemble des mots, respectivement, en tâche de LDT et PDM, dans les deux conditions de simulation étudiées. Ces valeurs ne diffèrent pas des valeurs observées dans les conditions étudiées dans les simulations B1. En tâche de NMG, BRAID-Phon génère une grand nombre de prononciations « incorrectes » dans la condition *AV.Réduite*, si bien que 60% seulement des réponses sont considérées correctes. Ce taux est de 89% dans la condition *AV.Uniforme*, ce qui est comparable à la condition *Centre* (voir Figure 3.5).

Comparaison qualitative des effets de longueur Dans les tâche de LDT et de PDM, nous obtenons des effets significatifs[13] des termes correspondants à la longueur et la fréquence dans les deux conditions. Les coefficients de détermination R^2 d'ajustement des modèles GAM en LDT sont

[13]Nous parlons ici, et dans les paragraphes qui suivent, de la significativité approximée des termes des modèles GAM (*Approximate significance of smooth terms*, en anglais) au sens de Wood (2006, chap. 6). Un effet « statistiquement significatif » d'un prédicteur de la régression GAM indique qu'on peut rejeter l'hypothèse nulle selon laquelle le terme de ce prédicteur ($s(\cdot)$) n'influence pas la variable prédite. Nous utilisons ce résultat ici à titre indicatif, d'où le non-report des p-valeurs et des F-valeurs de ces tests.

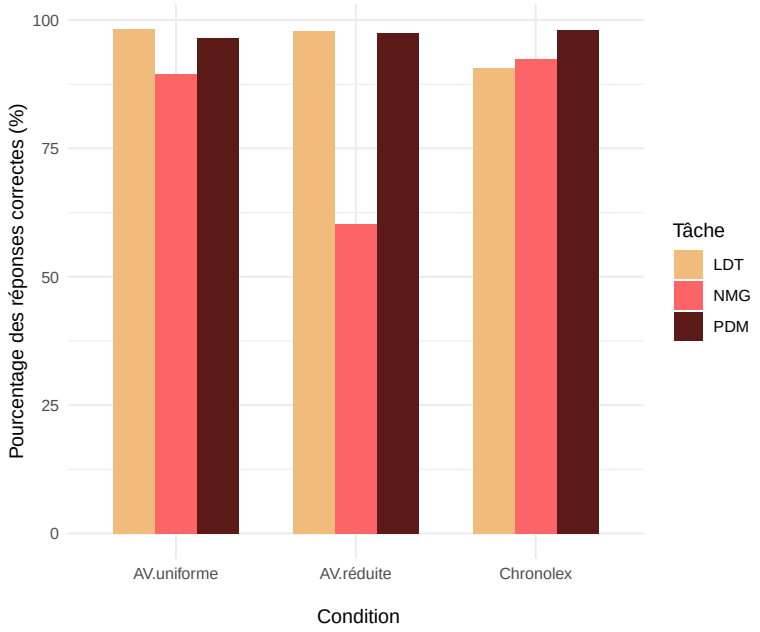


FIGURE 3.5 – **Pourcentages de réponses correctes de BRAID-Phon pour chaque condition des simulations B2 et pour chaque tâche simulée, comparés aux pourcentages de l'étude Chronolex.**

de 0,45 pour la condition *AV.Uniforme* et de 0,49 pour la condition *AV.Réduite* et, respectivement, de 0,46 et de 0,41 en PDM.

Quant à la tâche de NMG, l'analyse des données simulées indique des effets significatifs des termes correspondants à la longueur et la fréquence dans la condition *AV.Uniforme*. La régression correspondante fournit un R^2 égal à 0,31. Dans la condition *AV.Réduite*, seul l'effet du terme de la fréquence est significatif. Le coefficient de détermination du modèle GAM pour cette condition est égal à 0,10.

Sur la Figure 3.6, nous observons que l'amplitude et la forme de l'effet de longueur dans la condition *AV.Uniforme* est assez similaire à celui de la condition *Centre* dans les trois tâches étudiées. La condition *AV.Réduite* génère un effet de grande amplitude en LDT et PDM et un effet inversé en tâche de NMG.

Comparaison quantitative de l'effet de longueur Dans la condition *AV.Uniforme*, nous observons dans les trois tâches que les paramètres de la régression linéaire entre les effets de longueur simulés par BRAID-Phon et les effets observés expérimentalement sont très similaires à ceux de la condition *AV.Centre*. En particulier, les coefficients de détermination sont très élevés pour les tâches de LDT et NMG ($R^2 = 0{,}96$), et moins élevés pour la tâche de PDM ($R^2 = 0{,}68$). En LDT, la condition *AV.Réduite* montre un effet de longueur plus important par rapport aux données comportementales (la pente de la régression linéaire est égale à 0,36). En NMG, la pente de la droite de régression est négative, égale à -0,72, ce qui traduit l'inversion de l'effet de longueur précédemment décrit. De plus, cette régression a un coefficient de détermination moyen ($R^2 = 0{,}64$). En tâche de PDM, nous obtenons une pente et un intercept similaires à ceux de la condition *AV.Uniforme*. En revanche, le coefficient de corrélation est plus petit ($R^2 = 0{,}26$).

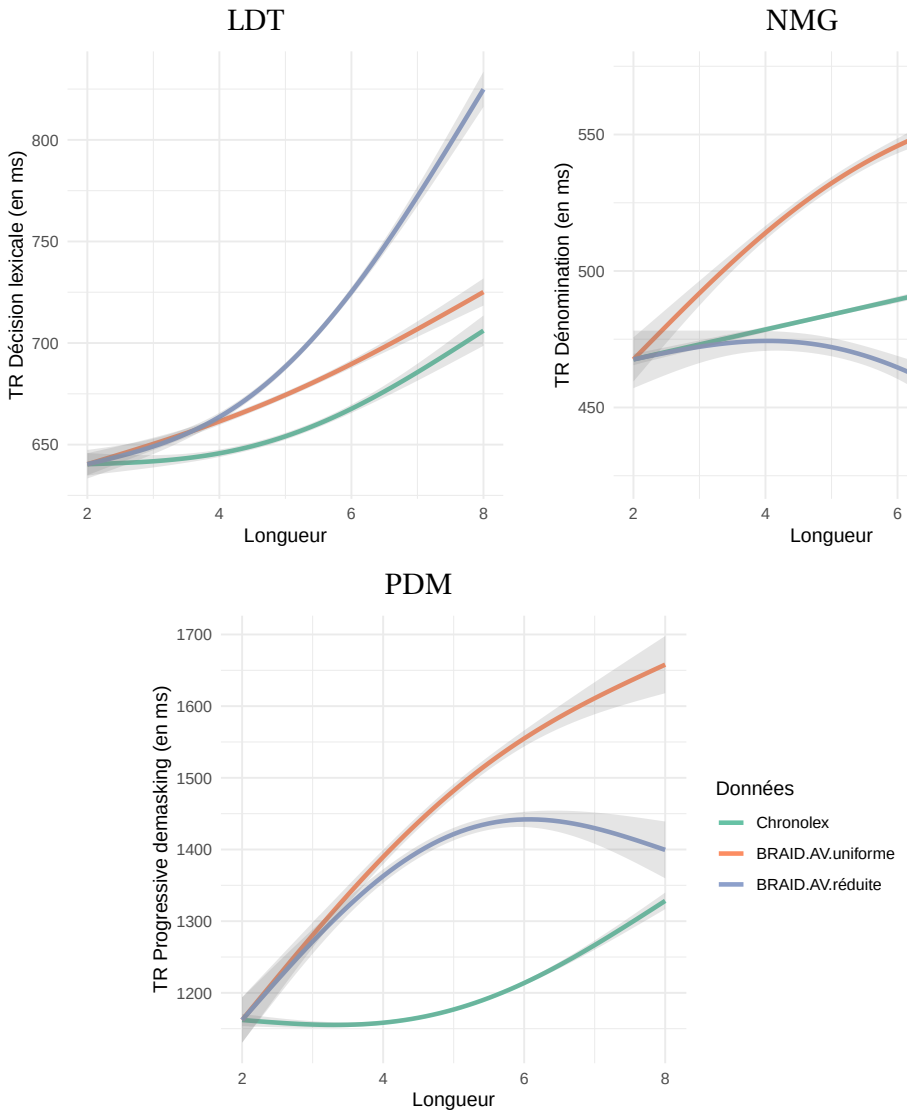


FIGURE 3.6 – **Effet de longueur dans les données comportementales de l'expérience Chronolex et dans les données simulées (simulations B2).** Les courbes des données simulées sont ajustées de sorte que le temps de réponse pour la longueur 2 soit égal à celui des données expérimentales. L'axe des ordonnées représente donc des TRs en millisecondes pour les données Chronolex et des itérations ajustées pour les données simulées. Les « surfaces » d'erreur (bandes grises) correspondent à l'erreur standard des modèles de régression GAM.

3.3 Synthèse des résultats de simulation

Dans cette étude expérimentale, nous avons étudié, à l'aide du modèle BRAID-Phon, les conséquences de manipulations des paramètres attentionnels, sur l'effet de longueur dans les différentes tâches. L'étude des conditions où la position du regard et de l'attention varient (Simulations B1) suggère que l'effet de longueur est mieux reproduit en LDT si la position visuo-attentionnelle est au centre du stimulus, et en NMG quand cette position est fixée entre la deuxième et la troisième lettre du stimulus. Pour la tâche de PDM, c'est la condition *Précentre* qui semble le mieux rendre compte des données comportementales de l'étude Chronolex en termes d'amplitude de l'effet de longueur. Cependant, les TRs du modèle computationnel, dans cette

condition, sont plus grands par rapport aux données expérimentales et aux autres conditions évaluées.

Dans les simulations B2, nous avons étudié l'impact de la dispersion de l'attention sur la prédiction de l'effet de longueur dans les trois tâches. L'utilisation d'une distribution uniforme de l'attention visuelle produit des effets de longueur similaires à ceux de la condition *Centre* : une bonne estimation de cet effet en LDT et une surestimation de son amplitude en NMG et en PDM. Enfin, une distribution très concentrée de l'attention visuelle, conjointe à un parcours sériel du stimulus, produit des résultats différents selon la tâche considérée. En effet, la condition *AV.Réduite* produit un effet de longueur de grande amplitude en LDT et en PDM mais il est non-significatif en NMG. Le taux de bonnes réponses, quant à lui, reste inchangé en LDT et en PDM mais chute considérablement en NMG.

4 Discussion

BRAID-Phon rend compte des principaux effets de la fréquence et de la longueur des mots dans trois tâches différentes de traitement visuel de mots (NMG, LDT et PDM), au moins qualitativement, tout en utilisant un même modèle computationnel et les mêmes valeurs de paramètres (Simulation A). Détaillons ce premier résultat. En effet, il est théoriquement admis que les tâches que nous avons étudiées reposent sur un même système cognitif, c'est-à-dire qu'elles mettent en jeu des connaissances communes qui sont manipulées par des processus communs. Notre approche consiste donc à viser le développement d'un modèle computationnel unique, qui intégrerait le plus de connaissances possibles relatives au traitement visuel de mots, et qui aurait ainsi le potentiel de simuler de nombreuses tâches qui relèvent de ce champ et de rendre compte des effets associés. Cela contraste avec les modèles computationnels existants, qui tendent généralement à se focaliser sur une tâche particulière (Norris, 2013). Bien que plusieurs modèles antérieurs aient tenté de simuler plusieurs tâches (Besner & McCann, 1987 ; Carr & Pollatsek, 1985 ; Jacobs et al., 1998), BRAID-Phon fournit une description plus détaillée des processus et des connaissances mis en jeu dans le traitement visuel des mots que ses prédécesseurs. De plus, les analyses ciblent habituellement la reproduction des effets (significatifs ou non) mais comparent rarement les amplitudes des effets observés et simulés (Adelman & Brown, 2008).

Dans cette étude, nous avons vérifié la capacité du modèle à simuler fidèlement les effets en termes de direction, d'amplitude, de forme et de variance expliquée. Nous avons observé que le modèle reproduit bien, qualitativement les effets attendus : leur direction est toujours correctement prédite, et leur amplitude et leur forme l'est parfois également. En effet, le modèle que nous proposons prouve sa robustesse pour la simulation de la tâche de LDT. En effet, le modèle sur lequel il se base, BRAID, a été précédemment confronté à de nombreuses données expérimentales, et a notamment permis de simuler fidèlement (Ginestet et al., 2019) la méga-étude *French Lexicon Project* (Ferrand et al., 2010), qui repose sur cette tâche. Les résultats que nous avons obtenus sur la tâche de LDT, particulièrement l'amplitude et la forme de l'effet de longueur, sont cohérents avec les résultats de l'étude de Ginestet et al. (2019), bien qu'ils soient issus de deux méga-études

distinctes. En ce qui concerne les tâches de NMG et de PDM, BRAID-Phon réussit à reproduire les effets de longueur et de fréquence attendus mais l'amplitude des effets simulés dans ces deux tâches est plus importante que celle observée dans Chronolex.

Nous avons également montré que les effets de longueur que nous avons obtenus étaient simulés avec des traitements parallèles des lettres, et ne nécessitaient donc pas forcément la mise en œuvre d'un traitement sériel ; ce résultat concorde avec l'étude de Ginestet et al. (2019). En effet, dans le cas particulier de la tâche de NMG, l'effet de longueur est attribué usuellement, et particulièrement dans le contexte des modèles double-voie, au recours à la voie sous-lexicale, et donc aux conversions graphème-phonème. En d'autres termes, le caractère sériel du traitement dans cette voie serait à l'origine de l'effet de longueur (Coltheart et al., 2001 ; Kapnoula et al., 2017 ; Perry & Ziegler, 2002 ; Rastle, Havelka, Wydell, Coltheart, & Besner, 2009). A l'inverse, BRAID-Phon parvient à reproduire l'effet de longueur en NMG (en tout cas aussi bien que les autres modèles, lorsqu'on prend en compte l'effet statistique ajouté par la prononciation du premier phonème), avec une architecture simple-voie et un traitement des lettres des stimuli complètement parallèle.

Nos simulations montrent aussi que, dans le cadre du modèle BRAID-Phon, les paramètres de la distribution visuo-attentionnelle impactent la forme et l'amplitude de l'effet de longueur. Nous avons étudié deux paramètres (la position du focus attentionnel et la dispersion de l'attention visuelle) afin d'évaluer leur effet sur l'ajustement aux données expérimentales.

Premièrement, une attention positionnée au centre du stimulus semble la plus adaptée pour reproduire correctement l'effet de longueur dans les tâches de LDT et de PDM, bien que les résultats soient moins satisfaisants dans la tâche de PDM. A l'inverse, pour la tâche de NMG, une fixation entre la deuxième et la troisième lettre du stimulus semble mieux rendre compte des données expérimentales. Ces différences suggèrent que des « stratégies visuo-attentionnelles » pourraient être déployées par les sujets de manière spécifique à chaque tâche. Dans une tâche de LDT ou de PDM, le sujet doit identifier le stimulus dans son ensemble pour répondre rapidement (identifier si le stimulus est un mot, ou reconnaître le mot). Cependant, en NMG, l'objectif est de prononcer rapidement à haute voix. Ainsi, fixer le début du stimulus afin de produire aussi rapidement que possible le début de la séquence phonémique est une stratégie plausible. Des études comportementales utilisant une mesure des mouvements oculaires fournissent des données consistantes avec cette hypothèse (Brysbaert & Nazir, 2005 ; Kuperman, Drieghe, Keuleers, & Brysbaert, 2013).

Cette stratégie de prononciation « par le début avant d'avoir lu la fin » impliquerait une interaction, dans la suite du processus, entre les processus perceptifs de prise d'information visuelle et les processus moteurs de planification et de production motrice de l'articulation de la parole. Cette interaction pourrait être « en cascade » ou « en pipeline producteur-consommateur », c'est-à-dire qu'on n'attendrait pas la fin des processus perceptifs avant la mise en route des processus moteurs.

Deuxièmement, la manipulation de la dispersion de l'attention visuelle impacte également l'effet de longueur. Dans le modèle BRAID-Phon, le sous-modèle visuo-attentionnel filtre l'information qui se propage du niveau sensoriel aux niveaux de traitement supérieurs. La distribution

d'attention par défaut est gaussienne et accorde plus d'attention, c'est-à-dire une plus grande quantité de ressource attentionnelle, aux lettres situées autour du focus attentionnel. Une distribution uniforme accorde la même quantité d'attention et donc une importance identique aux différentes lettres du stimulus. Dans nos simulations, nous n'observons pas de différences entre une distribution uniforme de l'attention et une attention gaussienne sur la base de données utilisée. Cela est consistant avec le résultat de l'étude de Ginestet et al. (2019), sur la base de donnée FLP, dans laquelle les stimuli avaient des longueurs entre 4 et 11 lettres, alors que notre étude est restreinte à des stimuli de 7 lettres au maximum. Dans leur étude, Ginestet et al. (2019) montrent qu'une répartition uniforme de l'attention génère un effet de longueur en LDT de plus grande amplitude que les données comportementales, mais cela se manifeste surtout pour les stimuli très longs (Ginestet et al., 2019, Figure 6, p 18). Pour les stimuli courts, comme dans notre étude, la différence entre une distribution gaussienne de l'attention dispersée par défaut et une distribution uniforme n'est d'ailleurs, numériquement, pas très grande.

En revanche, une distribution d'attention très piquée génère des simulations bien différentes de la condition *Centre*, et impacte différemment les tâches étudiées. Cette distribution ne permet pas de traiter les lettres du stimulus en parallèle et plusieurs fixations sont nécessaires. Dans les trois tâches étudiées, l'identification correcte du stimulus est nécessaire pour produire des réponses correctes. Dans les tâches de LDT et de PDM, l'attention réduite requiert plus de temps pour identifier les lettres et, par conséquent, augmente l'effet de longueur. En NMG, l'incapacité de traiter l'ensemble du stimulus affecte principalement le taux de bonnes réponses et l'effet de longueur n'est pas simulé dans la bonne direction. La dénomination des mots montre ici une particularité par rapport aux autres tâches : bien que la décision en NMG se base principalement sur l'évolution du premier phonème, l'identification de l'entièreté du stimulus est nécessaire pour le prononcer correctement, en particulier pour les mots qui présentent des irrégularités. Ainsi, la production du premier phonème se produit avant que l'ensemble du stimulus ne soit identifié, ce qui cause le taux d'erreur élevé dans cette condition.

Ces deux simulations exploratoires, sur l'effet des distributions « extrêmes » de l'attention visuelle, illustrent que la modulation de la distribution attentionnelle a un effet massif sur la dynamique d'accumulation d'information perceptive par le modèle, et donc sur la simulation des effets comportementaux. Ainsi, les effets de longueur ne sont pas forcément du ressort de mécanismes sériels du traitement du stimulus. A l'inverse, tous les mécanismes sériels ne génèrent pas forcément un effet de longueur d'amplitude unique, ni même forcément dans la même direction. Ainsi, nos résultats suggèrent, plus généralement, que l'attention visuelle permet de manipuler, de manière très flexible, la prise d'information visuelle sur le stimulus.

Récapitulatif :

- Simulation des données comportementales pour trois tâches (NMG, LDT, PDM) issues d'une méga-étude (Ferrand et al., 2011).
- BRAID-Phon rend compte des principaux effets de fréquence et de longueur des mots dans ces trois tâches (Simulations A).
- Etude des configurations visuo-attentionnelles permettant de rendre compte au mieux de l'effet de longueur (Simulations B1)
  - Reproduction de l'effet de longueur en LDT lorsque la position visuo-attentionnelle est au centre du stimulus.
  - Reproduction de l'effet de longueur en NMG lorsque la position visuo-attentionnelle est située entre la deuxième et la troisième lettre du stimulus.
  - BRAID-Phon rend compte de l'effet de longueur même lorsque le traitement est parllèle et en une fixation pour les trois tâches. Dans ce cadre, l'effet de longueur est une conséquence des ressources attentionnelles limitées.
  - Reproduction de l'effet de longueur, en termes d'amplitude, en PDM dans la condition *Précentre*. Mais les TRs simulés sont plus importants par rapport aux données comportementales.
- Etude des conséquences de la modification de la distribution de l'attention (Simulations B2)
  - Une distribution uniforme de l'attention visuelle permet de produire des effets de longueur similaire à ceux de la condition *Centre*, avec une bonne estimation en LDT et une sur-estimation en NMG et PDM.
  - Une distribution très concentrée de l'attention visuelle (associée à un parcours sériel du stimulus) permet de produire un effet de longueur de grande amplitude en LDT et en PDM dans la condition *AV.Réduite*.
  - Le taux de bonnes réponses est similaire à celui observé dans les données comportementales en LDT et PDM mais chute en NMG.

CHAPITRE

4 Lecture de pseudo-mots

"Some questions will be pondered for all eternity. What is the meaning of life? Where do you go when you die? And even more puzzlingly, what is the right way to pronounce 'GIF'?"

— THE ECONOMIST (2017)

Dans les Chapitres 2 et 3, nous avons proposé et évalué un modèle de lecture de mots, le modèle BRAID-Phon. Il est doté d'un mécanisme d'attention visuelle mais, jusqu'ici, nous n'avons utilisé le modèle que pour traiter des stimuli lexicaux (mots) dans leur globalité, c'est-à-dire avec une unique position pour la distribution visuo-attentionnelle et sans déplacement visuo-attentionnel (à l'exception de la condition *AV.réduite* du Chapitre 3). Nous avons exposé une propriété principale du modèle pendant le traitement du stimulus pour la lecture, qui se résume au fait que les distributions de probabilité sur les phonèmes sont directement influencées par l'évolution de la distribution de probabilité au niveau lexical, qui est elle-même la base de la reconnaissance de mots (voir les exemples présentés dans la Section 3 du Chapitre 2 et la Figure 2.8).

Néanmoins, cette relation très forte entre le processus de reconnaissance de mots et le processus de génération de phonèmes est une limite majeure pour tout modèle ayant l'ambition de rendre compte du système de lecture. En effet, c'est surtout la capacité à rendre compte de la lecture des pseudo-mots qui représente un défi pour les modèles de lecture et donne lieu à débat.

L'approche double-voie a donné lieu au développement de nombreux modèles computationnels qui permettent de simuler tant la lecture des mots connus, réguliers ou irréguliers, que la lecture des pseudo-mots. C'est cette conception double-voie du système de lecture qui prime actuellement dans la littérature. Selon ces modèles, la lecture de pseudo-mots repose sur des connaissances sous-lexicales « stockées indépendamment » des connaissances sur les mots de la langue. Ces connaissances sont le plus souvent représentées par un système de règles ou un réseau de

correspondances graphèmes-phonèmes. Les modèles à voie unique, au contraire, supposent que la prononciation des pseudo-mots ne nécessite pas de connaissances sous-lexicales stockées indépendamment. La lecture des pseudo-mots s'effectuerait par généralisation à partir des connaissances mémorisées sur les mots existants.

Ces deux cadres théoriques postulent l'intervention d'un mécanisme de segmentation permettant d'identifier les unités sous-lexicales pertinentes lors du traitement des pseudo-mots. Pour les modèles double-voie, ce mécanisme permet le passage d'une séquence de lettres à une séquence de graphèmes. Dans les modèles double-voie les plus influents (CDP+, CDP++, et DRC), la segmentation est propre à la voie sous-lexicale. Elle repose généralement sur un *parser*, qui permet d'apparier les lettres ou séquences de lettres du stimulus en des unités graphémiques plus larges, qui sont ensuite converties en phonèmes.

Dans le modèle DRC (Coltheart et al., 2001), le *parser* est en réalité un algorithme basé sur des règles permettant d'identifier les graphèmes puis de les convertir en phonèmes. Dans le modèle CDP+ (Perry et al., 2007), les lettres sont affectées, selon un algorithme de *parsing*, aux *slots* graphémiques à l'entrée du réseau sous-lexical TLA (*Two-Layer assembly network*). Chaque graphème identifié est immédiatement converti en phonème. Dans ces modèles, le *parsing* graphémique est implémenté de manière « algorithmique » et non-plausible, même si l'implication de mécanismes d'attention visuo-spatiale est suggérée. Pour repousser cette limite, Perry et al. (2013) proposent une implémentation du *parser* graphémique par l'intermédiaire d'un réseau récurrent qui traite, à chaque cycle (chaque pas de temps), 5 lettres. Cette séquence fixe de 5 lettres, assimilée à une « fenêtre attentionnelle », parcourt le stimulus en se calant, à chaque cycle, sur la lettre qui suit le graphème reconnu lors du cycle précédent. Les activations qui correspondent aux graphèmes reconnus alimentent, par la suite, le réseau sous-lexical TLA qui, à la suite d'une phase d'apprentissage, permet de convertir chaque unité orthographique (graphémique) en phonème(s).

Dans le cadre des modèles à voie unique, la lecture de pseudo-mots repose sur les mêmes connaissances et mécanismes que la lecture de mots, sans recourir à un quelconque système de conversion graphème-phonème spécifique au traitement sous-lexical. Dans le cadre du modèle MTM (Ans et al., 1998), l'attention visuelle, implémentée sous la forme d'une fenêtre attentionnelle de taille variable, est le mécanisme qui définit la taille des unités à traiter. Celle-ci intervient quelle que soit la nature du stimulus, mot connu ou mot nouveau, et permet un traitement global ou analytique du stimulus (pour plus de détails, voir Section 3 du Chapitre 1). La fenêtre attentionnelle qui est implémentée dans MTM s'adapte initialement à la taille du mot pour en effectuer un traitement global. Ce traitement échoue en général pour les pseudo-mots, conduisant la fenêtre attentionnelle à se focaliser sur la plus grande unité orthographique (graphème ou syllabe) pouvant être reconnue. La fenêtre se déplace ensuite de gauche à droite sur la séquence de lettres pour en effectuer un traitement sériel.

Indépendemment de la structure double-voie ou à voie unique, les modèles cités ci-dessus s'accordent sur l'implication de l'attention visuelle pour identifier les unités sous-lexicales. Néanmoins, on note que le modèle MTM implémente, de manière explicite, ce mécanisme attention-

nel, alors qu'un mécanisme de *parsing* graphémique, attribué à l'attention, est directement implémenté dans les modèles CDP. Par ailleurs, les deux modèles diffèrent fortement quant au rôle de l'attention et au niveau de traitement dans lequel elle intervient. L'attention est impliquée uniquement dans l'identification des unités sous-lexicales dans les modèles CDP alors qu'elle participe au traitement à la fois des mots et des pseudo-mots dans le modèle MTM. Dans les modèles CDP (Perry et al., 2007 ; Perry, Ziegler, & Zorzi, 2010 ; Perry et al., 2013), l'attention interviendrait après que les lettres de la séquence orthographique d'entrée ont été parfaitement identifiées. Dans MTM, l'attention contribue à l'identification des lettres, en favorisant celles qui sont sous le focus attentionnel, au détriment de celles qui ne le sont pas. Cette différenciation est cependant grossière puisque toutes les lettres à l'intérieur de la fenêtre attentionnelle sont maximalement activées alors que toutes celles qui sont à l'extérieur de la fenêtre le sont minimalement.

Dans ce chapitre, nous nous intéressons à la capacité de notre modèle de lecture BRAID-Phon à lire des pseudo-mots. Rappelons que le modèle BRAID-Phon, présenté précédemment, s'inscrit dans une approche à voie unique. Nous lui ajouterons ici un mécanisme de lecture par segments basé sur l'attention visuelle. Notre objectif ne sera pas de confronter le modèle à de larges corpus expérimentaux en tâche de lecture de pseudo-mots ; à la place, nous évaluerons, par le biais de différents exemples, la capacité du modèle à lire des pseudo-mots aux caractéristiques diverses. Par l'étude du traitement de ces exemples par notre modèle, nous aborderons trois questions générales concernant les modèles de lecture : (1) les connaissances lexicales contiennent-elles les connaissances « sous-lexicales » nécessaires à la lecture de pseudo-mots ? (2) Existe-t-il des conditions particulières pour que ces connaissances permettent de générer les bonnes sorties phonologiques ? Enfin, (3) comment l'attention visuelle permet-elle de segmenter les pseudo-mots ?

Pour répondre à ces questions, nous structurerons ce chapitre en trois parties. Dans une première partie, nous illustrerons, à l'aide d'un exemple, le comportement « spontané » du modèle BRAID-Phon afin de mieux spécifier la contrainte posée par l'architecture à voie unique du modèle. Ensuite, nous nous limiterons, dans la deuxième partie, à l'étude du sous-modèle lexical de BRAID-Phon pour montrer comment les relations sous-lexicales y sont représentées et donc répondre aux deux premières questions ci-dessus. Enfin, nous proposerons et évaluerons une extension du modèle BRAID-Phon qui inclut un mécanisme de lecture par segments basé sur l'attention visuelle. Nous montrerons donc le fonctionnement et les propriétés de ce modèle lors de la lecture de pseudo-mots.

1 Lecture d'un pseudo-mot par BRAID-Phon : Comportement spontané

Dans cette section, nous nous intéressons au comportement du modèle BRAID-Phon lors de la lecture de pseudo-mots. Pour cela, nous prenons comme exemple le pseudo-mot « VIRDIN ». Nous explorons d'abord le comportement « spontané » du modèle. Nous entendons par « spontané » le fait que nous utilisons le modèle tel qu'il était décrit dans les Chapitres 2 et 3, les valeurs des différents paramètres étant celles fixées par défaut pour le traitement des mots connus. Cela nous

permettra de montrer les limites d'un traitement « global » (c'est-à-dire en une seule capture attentionnelle) de la forme orthographique d'entrée. Nous explorerons ensuite le comportement du modèle lors de déplacements du focus attentionnel.

1.1 Traitement global : Une fixation centrale

Dans un premier temps, nous utilisons les paramètres par défaut du modèle, c'est-à-dire ceux qui étaient employés pour la lecture des mots. Nous rappelons que dans ce cas, le regard et le focus attentionnel sont positionnés au milieu du stimulus. Puisque le pseudo-mot « VIRDIN » est de longueur 6, nous avons donc $g = \mu_A = 3{,}5$. De plus, la dispersion de la distribution gaussienne de l'attention a sa valeur par défaut, fixée à $\sigma_A = 1{,}75$ (voir Figure 4.1).

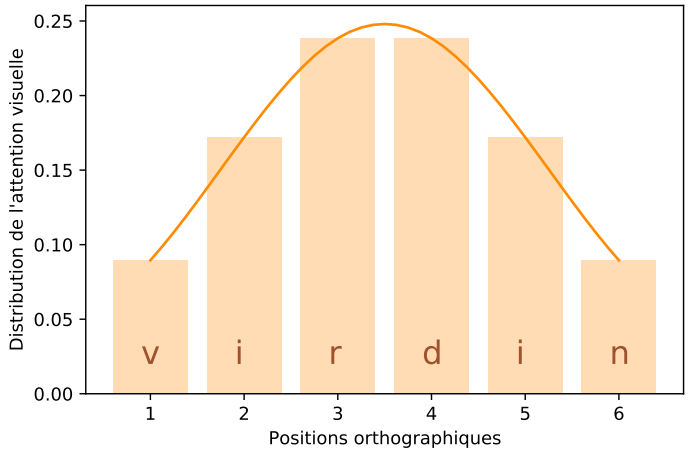


FIGURE 4.1 – **Quantité d'attention visuelle allouée à chaque position lors du traitement « global » du pseudo-mot « VIRDIN ».** La distribution de l'attention visuelle est caractérisée par $P(A \mid [\mu_A = 3{,}5]\ [\sigma_A = 1{,}75])$. Cette distribution modélise une attention visuelle portée au centre du stimulus.

Ce choix est « naïf », puisqu'il suppose un traitement du stimulus dans sa globalité, comme cela est habituellement le cas pour les mots connus courts ou de longueur moyenne. Lors du traitement, l'information perceptive sur l'ensemble des lettres du pseudo-mot s'accumule progressivement au cours du temps, entraînant l'activation des connaissances lexicales mémorisées. Tout un ensemble de mots connus sont simultanément activés, c'est-à-dire que leur probabilité augmente, en fonction de leur proximité orthographique avec le stimulus « VIRDIN ». Le mot « jardin », qui est le mot connu le plus proche orthographiquement du pseudo-mot d'entrée, est le plus activé au niveau lexical, apparaissant ainsi comme le candidat orthographique le plus probable. La distribution de probabilité lexicale entraîne à son tour une accumulation de probabilités, au cours des itérations, quant à la nature des phonèmes correspondants. La suite de phonèmes la plus probable à l'itération finale est la séquence /ʒaʁdɛ̃/ correspondant au mot connu le plus proche du pseudo-mot présenté, le mot « jardin » (voir Figure 4.2).

Nous observons donc une « lexicalisation » par le modèle sur ce stimulus, c'est-à-dire que le modèle le traite comme s'il s'agissait d'un mot connu. Ce résultat pour un pseudo-mot traité en une unique capture visuo-attentionnelle était toutefois prévisible. En effet, hormis les théories

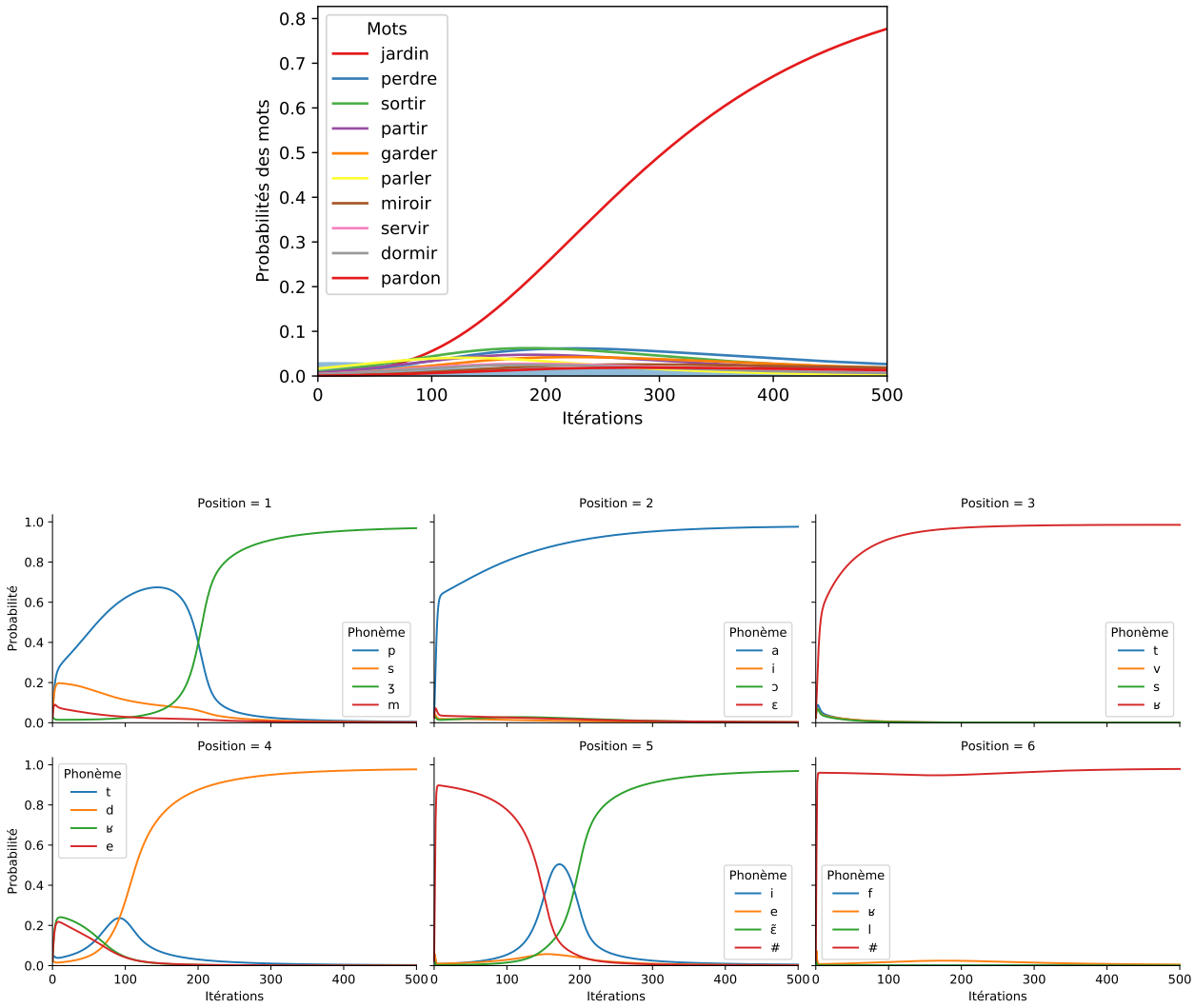


FIGURE 4.2 – **Evolution des distributions de probabilité lexicales et phonémiques pendant le traitement du pseudo-mot « VIRDIN », en une seule fixation attentionnelle.** En haut, courbes d'évolution des probabilités lexicales, c'est-à-dire liées à la reconnaissance de mots ; en bas (6 graphes), courbes d'évolution des probabilités phonémiques. Le mot et les phonèmes les plus probables sont le mot « jardin » et les phonèmes correspondants /ʒaʁdɛ̃/. Les courbes relatives aux phonèmes sont visualisées pour plus de positions phonémiques que nécessaire, et le phonème # indique la fin de la séquence phonémique.

d'analogie (Glushko, 1979 ; Kay & Marcel, 1981) qui se caractérisent par un traitement purement lexical des pseudo-mots et sans aucune sérialité, et dont le modèle Triangle (Plaut et al., 1996 ; Seidenberg & McClelland, 1989) pourrait être considéré comme une implémentation computationnelle, les autres théories de lecture, qu'elles soient à voie unique ou double-voie, reconnaissent que la lecture de pseudo-mots repose essentiellement sur une procédure analytique séquentielle. Cette procédure est censée reposer sur l'application de relations graphème-phonème stockées indépendamment des connaissances lexicales dans les modèles double-voie. Dans les modèle à voie unique (voir Ans et al., 1998 ; Hendrix et al., 2019), le traitement du pseudo-mot active des mots connus, qui contribuent à la lecture du pseudo-mot sans toutefois entraîner de lexicalisation. Ces modèles contiennent un mécanisme de « composition » de portions des activations phonémiques

relatives à chaque mot connu activé. Ainsi, qu'on procède par analogie ou par une procédure analytique, il apparaît qu'un traitement séquentiel de la séquence d'entrée, et/ou de la séquence de sortie, est nécessaire au traitement des pseudo-mots pour éviter l'accumulation d'erreurs de lexicalisation.

1.2 Traitement séquentiel : Deux fixations successives

Etudions maintenant le comportement du modèle pendant le traitement séquentiel d'un pseudo-mot. Pour cela, nous choisissons arbitrairement de traiter le stimulus « VIRDIN » en suivant une procédure séquentielle en deux captures attentionnelles (voir Figure 4.3). Plus spécifiquement, l'attention est dirigée pendant les 250 premières itérations sur la première portion du stimulus ($g = \mu_A = 2$ et $\sigma_A = 1{,}75$), puis sur la seconde portion ($g = \mu_A = 5$ et $\sigma_A = 1{,}75$) pour les 250 itérations suivantes.

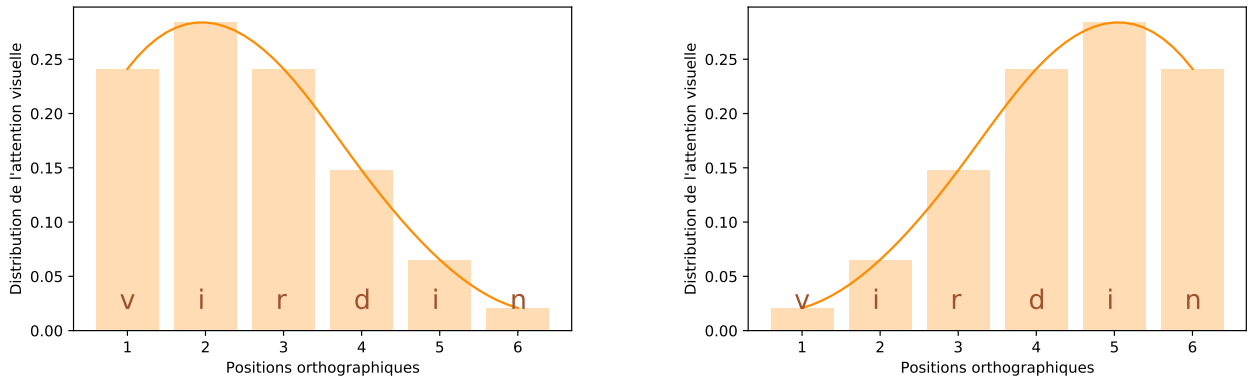


FIGURE 4.3 – **Quantité d'attention visuelle allouée à chaque position lors du traitement du pseudo-mot « VIRDIN », en deux fixations attentionnelles.** Le premier graphique (à gauche) correspond à la distribution visuo-attentionnelle $P(A \mid [\mu_A = 2]\ [\sigma_A = 1{,}75])$ et le second (à droite) correspond à la distribution visuo-attentionnelle $P(A \mid [\mu_A = 5]\ [\sigma_A = 1{,}75])$. Ces distributions modélisent une attention visuelle portée, respectivement, sur la première et la seconde portion du stimulus.

Lorsque l'attention se porte sur le début du pseudo-mot « VIRDIN », le mot « virage » et les phonèmes correspondants (/viʁaʒ/) sont les plus probables à l'issue des 250 premières itérations (voir Figure 4.4). Pendant cette période, le modèle n'accumule clairement de l'information perceptive que sur les premières lettres de la séquence orthographique d'entrée et les mots, et donc les phonèmes, sont activés sur la seule base de cette information sensorielle. En d'autres termes, le modèle « perçoit » un mot de 6 lettres qui commence, en gros, par « VIR », et correspond le plus probablement au mot « virage ». Le même mécanisme se reproduit lorsque l'attention et la position du regard sont décalées vers la droite : le modèle accumule alors de l'information perceptive sur « DIN », les dernières lettres du stimulus; le mot « jardin » gagne la compétition au niveau lexical et les phonèmes correspondants /ʒaʁdɛ̃/ sont activés. Nous observons que, bien que les phonèmes les plus activés aient bien été ceux qui correspondent à « VIR » durant la première fixation et à « DIN » durant la seconde fixation, l'état des distributions de probabilité à la

fin du traitement correspond au traitement de la dernière fixation attentionnelle, conduisant à lire /ʒaʁdɛ̃/ pour « VIRDIN ».

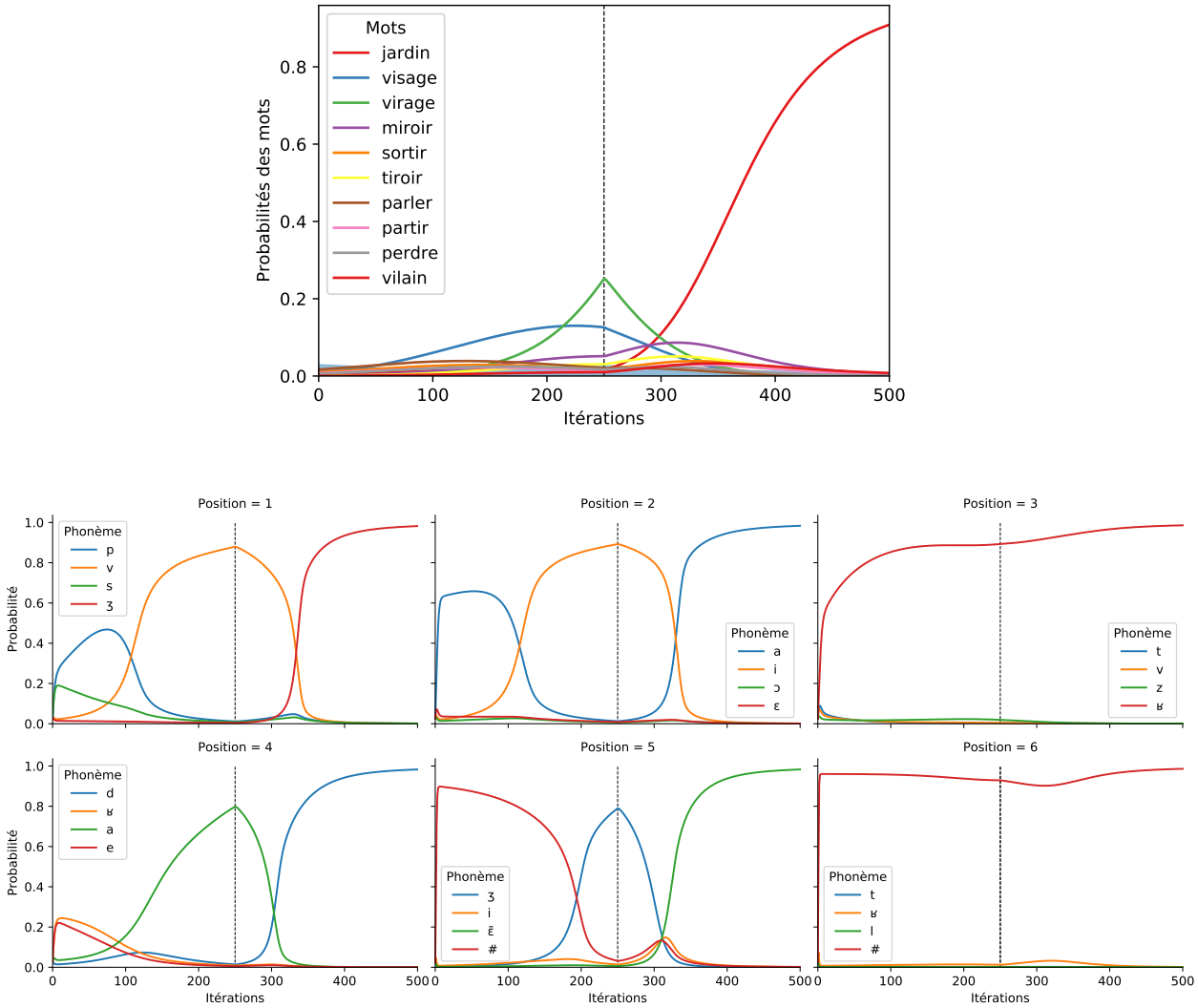


FIGURE 4.4 – **Évolution des distributions de probabilité lexicales et phonémiques pendant le traitement du pseudo-mot « VIRDIN », en deux fixations attentionnelles.** En haut, courbes d'évolution des probabilités lexicales, c'est-à-dire liées à la reconnaissance de mots ; en bas (6 graphes), courbes d'évolution des probabilités phonémiques. Le première fixation attentionnelle (sur la gauche du stimulus) dure pendant les 250 premières itérations, et la seconde (sur la droite du stimulus) dure pendant les 250 itérations suivantes (le pointillés verticaux délimitent ces deux fixations). Les courbes relatives aux phonèmes sont visualisées pour plus de positions phonémiques que nécessaire, et le phonème # indique la fin de la suite phonémique produite.

1.3 Analyse du comportement « spontané » du modèle

Cet exemple montre que, lorsqu'il est confronté à un pseudo-mot traité de façon globale (distribution attentionnelle par défaut), BRAID-Phon tend à produire un mot orthographiquement proche à la place du pseudo-mot présenté : il lexicalise. Lorsque le pseudo-mot est traité séquentiellement en deux captures attentionnelles, la séquence phonologique à l'itération finale dépend des

traitements effectués lors de la dernière fixation, ce qui conduit également à générer les phonèmes correspondants à un mot existant. Le mot produit est alors celui qui partage le plus de similarités avec les informations orthographiques accumulées au cours de cette seconde fixation. Dans le cas de l'exemple « VIRDIN », les sorties phonologiques générées après une fixation unique ou après deux fixations sont identiques et correspondent au mot « jardin ». Ceci n'est pas une règle générale, avec d'autres exemples de pseudo-mots, deux mots différents sont parfois produits lors des traitements global et séquentiel.

Dans notre exemple de traitement séquentiel, on remarque que l'information perceptive accumulée à chaque fixation permet de générer des séquences phonologiques partiellement compatibles avec la sortie phonologique attendue pour le pseudo-mot « VIRDIN ». En revanche, à la fin du traitement après la seconde fixation, le modèle ne dispose plus que de l'information phonologique correspondant au dernier mot le plus probable. Les portions phonologiques « pertinentes » obtenues à la fin de la première fixation sont perdues. En effet, dans le modèle BRAID-Phon, l'aspect dynamique dans les différents niveaux de représentations est modélisé par des chaînes de Markov. Chacune de ces chaînes permet d'accumuler de l'évidence perceptive sur « UNE séquence de lettres », ou sur « UN mot », ou sur « UNE séquence de phonèmes ». Au niveau lexical en particulier, la seconde fixation active « jardin », et écrase ainsi le mot « virage » préalablement activé. Il y a un seul espace lexical de probabilité, si bien qu'augmenter la probabilité du mot « jardin » diminue mécaniquement celle du mot « virage », d'autant que celui-ci n'est plus soutenu, puisqu'il n'est plus consistant avec les lettres perçues lors de la seconde capture attentionnelle.

Notons que cette propriété, d'avoir un unique espace lexical alimenté par une unique représentation perceptive des lettres, est critique pour que le modèle puisse rendre compte des observations comportementales issues d'expériences avec amorçage (Ginestet, 2016 ; Phénix et al., 2020). En effet, dans ces expériences, le stimulus change après quelques itérations, et l'espace lexical est « laissé dans un certain état » par cette amorce, ce qui accélère la reconnaissance de la cible, lorsqu'elle est congruente avec l'amorce, et la retarde dans le cas contraire.

A partir de cet exemple, nous constatons que BRAID-Phon ne parvient pas à produire la phonologie des stimuli non lexicaux avec son comportement « spontané ». En effet, le modèle tend à lexicaliser bien que, lors du traitement séquentiel, les sorties phonologiques soient pertinentes, portion par portion, c'est-à-dire vis-à-vis de la portion du stimulus sur laquelle porte l'attention visuelle. Ceci suggère qu'une information sous-lexicale est extraite des connaissances lexicales du modèle, et qu'un traitement par portions pourrait permettre de lire correctement les pseudo-mots. Par la suite, nous utilisons le terme « segment » pour référer à une portion de mot correspondant à une lettre ou une séquence de plusieurs lettres d'un stimulus donné. Ainsi, nous conservons la généralité de notre analyse et de notre modèle, sans présupposer que les segments sont syllabiques, graphémiques ou correspondent à une quelconque unité sous-lexicale définie *a priori*.

Cette analyse nous permet de spécifier les questions auxquelles nous essayerons de répondre dans la suite du chapitre :

1. Est-ce que le sous-modèle lexical de BRAID-Phon permet de formuler mathématiquement la lecture d'un segment de stimulus et, donc, de convertir une séquence de lettres en une séquence de phonèmes ?
2. Sous quelles conditions ces traitements sous-lexicaux sont-ils pertinents ?
3. Comment implémenter dans BRAID-Phon un traitement par segment qui réussisse à générer des sorties phonologiques correctes et qui ne lexicalise pas (systématiquement) lors de la lecture de pseudo-mots ?

Pour répondre aux deux premières questions, nous allons étudier le comportement du modèle en nous restreignant, temporairement, au sous-modèle lexical de BRAID-Phon. Cela permettra notamment d'alléger les notations mathématiques pour étudier la capacité du modèle à extraire les relations entre l'orthographe et la phonologie, lorsque des segments sont traités. Pour répondre à la dernière question, nous proposerons, dans le cadre du modèle BRAID-Phon, un mécanisme de lecture par segments. Nous illustrerons son rôle et son impact dans le passage de l'orthographe à la phonologie des pseudo-mots.

2 Correspondances orthographe-phonologie dans le sous-modèle lexical

Dans la Section 3 du Chapitre 2, nous avons défini formellement comment simuler la tâche de lecture à haute voix d'un mot lors d'un traitement global portant sur l'ensemble de la séquence de lettres avec le modèle BRAID-Phon. Nous avons supposé que toutes les lettres du stimulus étaient traitées simultanément et, donc, que l'information sur l'ensemble du stimulus était générée en fin de traitement. Dans cette section, nous nous intéresserons à ce que le modèle produirait lorsque seulement un segment du stimulus est présenté. L'objectif est d'étudier les correspondances générées sur la base des connaissances lexicales entre unités orthographiques et unités phonologiques. Une question corollaire concernera la taille des unités sous-lexicales considérées pour ces correspondances.

La relation entre unités orthographiques et phonologiques joue un rôle central en lecture, particulièrement dans la lecture de nouveaux mots et de pseudo-mots. En effet, la lecture de pseudo-mots nécessite de convertir des séquences de lettres en unités phonologiques, alors qu'il n'existe pas de « point intermédiaire unique dans l'espace lexical », par définition même du fait d'être un pseudo-mot. Ce passage des lettres aux phonèmes n'est pas trivial dans de nombreuses langues, dans lesquelles la prononciation d'une lettre ou d'un groupe de lettres n'est pas unique, et dépend souvent du contexte. Par exemple, si l'on considère les mots « plage », « plain » et « plaine », prononcer correctement la voyelle représentée par la lettre « A » requiert l'accès à l'identité des lettres voisines. De plus, le nombre de lettres représentant un phonème n'est pas constant. En effet, il n'existe pas (en français, en anglais et dans un ensemble de langues écrites dites « opaques ») de correspondances un-à-un entre les lettres et les phonèmes. Le recours à des unités orthographiques intermédiaires, telles que le graphème (suite de lettres correspondant à un seul phonème)

ou la syllabe permet de pallier cette difficulté. Cependant, ces unités sous-lexicales intermédiaires n'ont pas de longueur (nombre de lettres) prédéfinie, et leur identification par le système de lecture ne parait pas évidente. Ainsi, nous ne présupposerons pas ici la disponibilité de telles unités, et étudierons, à la place, le traitement à l'aide de segments de différentes tailles. Nous montrerons que cette flexibilité en termes du choix de la taille des segments (de l'unité sous-lexicale intermédiaire) est utile pour tenir compte du contexte et dériver la sortie phonologique pertinente.

Dans cette section, nous nous limiterons à étudier les propriétés du sous-modèle lexical, pour alléger la notation mathématique tout en permettant de se focaliser sur le lien établi entre les représentations orthographiques et phonologiques dans le modèle BRAID-Phon. Nous nous intéresserons donc à la sensibilité de la conversion de segments orthographiques en phonèmes à l'aide de ce sous-modèle et donc à la « potentielle » flexibilité de traitement qu'il prodigue. Nous illustrerons ceci à l'aide d'exemples.

2.1 Modèle

Nous considérons donc ici un modèle qui est une version « épurée » du sous-modèle lexical de BRAID-Phon, défini dans le Chapitre 2 (Figure 2.7), que nous appellerons par la suite le « modèle lexical statique ». Pour cela, nous dissocions la composante dynamique du modèle (la chaine de Markov qui permet la mise à jour, au fur et à mesure des itérations, de la distribution de probabilité sur l'espace lexical) de sa composante statique (le modèle décrivant les lettres et les phonèmes qui correspondent aux mots connus). Nous ne conservons que la composante statique, ce qui permet d'ôter les indices temporels des variables. Ainsi, le modèle lexical statique se limite aux variables $\Phi_{1:M}$, $L_{1:N}$ et W, et décrit uniquement les connaissances orthographiques et phonologiques des mots connus (voir Figure 4.5).

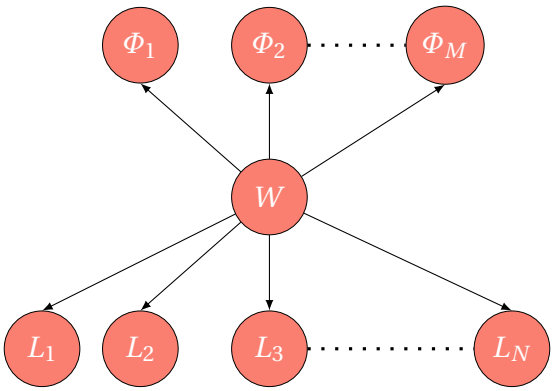


FIGURE 4.5 – **Représentation graphique des relations de dépendance dans le sous-modèle lexical statique.** Les dépendances inférieures, entre les variables W et $L_{1:M}$, représentent les connaissances orthographiques et les dépendances supérieures, entre W et $\Phi_{1:M}$, représentent les connaissances phonologiques.

La distribution conjointe du modèle lexical statique est donc définie par :

$$\begin{aligned} P(W\ L_{1:N}\ \Phi_{1:M}) =& P(W)\ P(L_{1:N} \mid W)\ P(\Phi_{1:M} \mid W) \\ =& P(W) \prod_{n=1}^{N} P(L_n \mid W) \prod_{m=1}^{M} P(\Phi_m \mid W)\ . \end{aligned} \tag{4.1}$$

Nous gardons les définitions de ces distributions telles qu'elles ont été introduites précédemment. Pour rappel, $P(W)$ est un prior sur l'espace lexical $\mathscr{D}_W$, qui représente la fréquence des mots du lexique, et $P(L_n \mid W)$ et $P(\Phi_m \mid W)$ sont des distributions quasi-Dirac qui indiquent les lettres et les phonèmes de chaque mot du lexique (c'est-à-dire, la bonne lettre ou le bon phonème a une probabilité très élevée, et il reste une petite probabilité résiduelle pour les alternatives).

En pratique, nous utilisons un lexique français de 142 694 mots (issu de Lexique 3.8 ; New, 2006) ou un lexique anglais de 38 839 mots (issu de l'*English Lexicon Project* ; Balota et al., 2007). Dans les exemples de cette section, nous ne nous restreignons pas à une longueur spécifique de mots. Toutes les longueurs sont considérées. Dans ce cadre, N correspond à la longueur orthographique maximale dans le lexique (c'est-à-dire la longueur du mot le plus long orthographiquement) et M à la longueur phonologique maximale (c'est-à-dire la longueur du mot le plus long phonologiquement). Afin de gérer la différence de longueur entre les mots, nous utilisons un symbole spécial (noté #) pour représenter l'absence d'une lettre ou d'un phonème dans une séquence ou pour représenter la marque de fin de séquence (*padding*). Dans le cas du lexique français, nous avons $M = 21$ et $N = 25$. Par exemple, le mot « pain » est représenté orthographiquement par la séquence de 4 lettres « PAIN » suivi de 21 « # » et phonologiquement par la séquence de 2 phonèmes /pɛ̃/ suivi de 19 « # ».

2.2 Calculs d'inférence

2.2.1 Cas général

Après avoir défini tous les termes de la distribution conjointe et, avec eux, la structure et les connaissances du modèle, l'inférence bayésienne permet de calculer les distributions de probabilité conditionnelles que l'on souhaite. Considérons, comme premier cas, la conversion d'une séquence complète de lettres en une séquence complète de phonèmes. Mathématiquement, on calcule la distribution de probabilité sur les phonèmes $\Phi_{1:M}$ sachant les lettres en entrée $L_{1:N}$:

$$\begin{aligned} P(\Phi_{1:M} \mid L_{1:N}) \propto& \sum_{W} P(W\ L_{1:N}\ \Phi_{1:M}) \\ \propto& \sum_{W} \left(P(W) \prod_{m=1}^{M} P(\Phi_m \mid W) \prod_{n=1}^{N} P(L_n \mid W) \right) . \end{aligned} \tag{4.2}$$

Dans ce calcul, nous observons qu'aucun terme ne relie directement les lettres $L_{1:N}$ aux phonèmes $\Phi_{1:M}$. A la place, des termes $P(L_n \mid W)$ relient les lettres à l'espace lexical, et des termes $P(\Phi_m \mid W)$ relient l'espace lexical aux phonèmes. Ceci découle de l'hypothèse structurale centrale du sous-modèle lexical, selon laquelle le lien entre connaissances orthographiques et phonologiques s'effectue uniquement par l'intermédiaire de l'espace lexical. C'est l'hypothèse de l'architecture à

voie unique. Mathématiquement, cette hypothèse se traduit par une hypothèse d'indépendance conditionnelle : on suppose les lettres $L_{1:N}$ indépendantes des phonèmes $\Phi_{1:M}$, conditionnellement à la connaissance du mot W.

2.2.2 Prononcer une suite de lettres correspondant à un segment de stimulus

Jusqu'ici, nous avons considéré le calcul de l'identité des phonèmes sachant la séquence complète des lettres du stimulus. Nous souhaitons, maintenant, calculer la distribution de probabilité sur l'identité des phonèmes, mais en ne disposant, en entrée, que d'une sous-séquence des lettres (un segment du stimulus).

Pour cela, nous devons tout d'abord définir le segment d'entrée (la sous-séquence de lettres) et le segment en sortie (la sous-séquence de phonèmes) que nous considérons. Considérons, pour simplifier, un segment d'entrée de taille quelconque et un segment de sortie restreint à un seul phonème. Notons donc m la position du phonème dont on souhaite calculer la probabilité et $\{i, i+1, \cdots, i+k\}$ les positions des lettres que l'on considère connues, avec i la position de la première lettre du segment (ce segment est donc de longueur $k+1$). En utilisant le modèle que nous avons décrit précédemment et en appliquant les règles d'inférence probabiliste, nous calculons la probabilité du phonème en position m sachant l'identité d'une suite de lettres :

$$P(\Phi_m \mid [L_{i:i+k} = l_{i:i+k}]) = \sum_W \left(P(W)\ P(\Phi_m \mid W) \prod_{j=i}^{i+k} P([L_j = l_j] \mid W) \right). \tag{4.3}$$

Interprétons cette expression. Tout d'abord, rappelons que les termes associant les lettres aux mots, et les phonèmes aux mots, sont des distributions discrètes quasi-Dirac. Ainsi, leur valeur est très proche de 1 « lorsque c'est la bonne lettre ou le bon phonème pour cette position et pour ce mot », et très proche de 0 sinon. Ainsi, dans la sommation sur l'espace lexical ($\sum_W \ldots$), le produit des probabilités des lettres est élevé pour les mots qui ont cette sous-séquence de lettres, et proche de 0 pour les autres. On peut donc interpréter cette partie du calcul comme une « sélection » des mots dans W qui ont les lettres $l_{i:i+k}$ entre les positions i et $i+k$.

Par exemple, s'il n'y a qu'un seul mot possédant cette sous-séquence de lettres, alors il donne directement une très haute probabilité au phonème qu'il a en position m, et la distribution de probabilité finale sur les phonèmes est dictée par les connaissances phonologique de cet unique mot. En revanche, si plusieurs mots possèdent cette sous-séquence de lettres, alors la distribution finale est une somme (pondérée par la fréquence des mots) des phonèmes associés à ces mots. Considérons tout d'abord le cas où tous les mots concernés associent le même phonème en position m : alors, là aussi, la distribution finale donne une très haute probabilité pour ce phonème. Considérons maintenant le cas où plusieurs phonèmes différents sont des associations possibles pour cette séquence de lettres (dans le cas d'une langue non-transparente) : alors le résultat de l'inférence est une distribution reflétant toutes les correspondances phonémiques possibles pour cette séquence de lettres, et la valeur de probabilité de chaque phonème résulte de la « fréquence relative » des correspondances possibles entre ces lettres et les phonèmes dans la langue.

Ainsi, on comprend que l'Equation (4.3) calcule la correspondance entre une sous-séquence de lettres et le phonème en position m, en fonction des statistiques de ces correspondances dans la langue. Dans le cas spécifique où les sous-séquences de lettres ne correspondraient qu'à des graphèmes, alors on pourrait interpréter l'Equation (4.3) comme « calculant à la volée » les conversions graphèmes-phonèmes. Ainsi, nous voyons que le modèle BRAID-Phon contient bien les connaissances nécessaires pour le calcul des conversions des unités orthographiques en unités phonologiques, et plus spécifiquement des conversions de segments de lettres et segments de phonèmes, mais sans un « stockage explicite » de ces conversions en mémoire.

2.2.3 Sensibilité positionnelle de la conversion lettres-phonèmes

Etudions maintenant certaines limites du calcul des conversions lettres-phonèmes par segment que nous avons proposé. Nous les exprimons mathématiquement sous forme d'inégalités, afin de cerner ce que permet et ne permet pas la structure de connaissances lexicales que nous supposons dans ce chapitre.

Premièrement, dans le modèle lexical statique, le codage des positions des phonèmes et des lettres est strict. Nous observons, à partir de l'Equation (4.3), une forte dépendance à la position considérée. En effet, le calcul de l'identité du phonème dépend toujours de la position dans laquelle elle est calculée. Nous pouvons affirmer à titre d'exemple que :

$$P(\Phi_m \mid [L_{i:i+k} = l_{i:i+k}]) \neq P(\Phi_{m+1} \mid [L_{i+1:i+1+k} = l_{i:i+k}]) \,.$$

En d'autres termes, la distribution de probabilité des phonèmes dans une position sachant un segment dépend des positions orthographiques et phonologiques sur lesquelles on effectue le calcul. Ainsi, si l'on souhaitait calculer, grâce au modèle, des conversions graphème-phonème indépendantes de la position, il faudrait les calculer en toute position, puis les agréger (par exemple, en moyennant sur les positions). Nous n'aurons pas spécialement besoin de ce mécanisme dans la suite de notre raisonnement.

Deuxièmement, le calcul d'inférence dépend de l'identité des lettres du segment pris conjointement. Lorsqu'on effectue une conversion des lettres vers les phonèmes, il faut noter que le résultat obtenu au niveau phonologique (identité des phonèmes) dépend de l'identité de toutes lettres et ne peut être décomposé. Mathématiquement, nous avons :

$$P(\Phi_{1:M} \mid L_{1:N}) \neq \prod_{j=1}^{M} \prod_{i=1}^{N} P(\Phi_j \mid L_i) \,.$$

Par conséquent, nous ne pouvons pas envisager le calcul d'une inférence, sachant l'ensemble des lettres d'un stimulus, comme équivalente à une multiplication des résultats des inférences pour deux segments orthographiques qui le composeraient. Ainsi, mathématiquement :

$$P(\Phi_{1:M} \mid L_{1:N}) \neq \prod_{j=1}^{M} \left[P(\Phi_j \mid L_{1:k}) \, P(\Phi_j \mid L_{k+1:N}) \right] \,,$$

tel que k, compris entre 1 et N, indique la longueur du premier segment et $N-k$ indique la longueur du deuxième segment. Calculer les sorties phonologiques dépend alors de la segmentation choisie pour le stimulus et par conséquent de la longueur de ces segments.

Notons que chacune de ces inégalités peut évidemment être contredite par un cas particulier, c'est-à-dire avec un lexique particulier et un choix de segment d'entrée ou de sortie particulier. Par exemple, ces inégalités ne seraient plus vraies pour une langue complètement transparente, dans les deux sens (un seul son par lettre, une seul lettre par son), et dans laquelle les conversions seraient, de plus, totalement indépendantes de la position. Nous considérons, dans la suite, le cas général dans lequel ces inégalités sont vraies.

2.3 Illustrations de la lecture de segments de pseudo-mots

Résumons les observations mathématiques de la section précédente. Nous avons montré que, dans le sous-modèle lexical statique, la structure de dépendance probabiliste entre les espaces lexicaux, orthographiques et phonologiques implique que :

1. les relations orthographiques-phonologiques peuvent être calculées à partir des connaissances lexicales ;
2. le résultat du processus de conversion entre sous-séquences de lettres et sons dépend de la manière de segmenter le stimulus, et, en particulier, ce résultat ne peut pas être exprimé en considérant la prononciation de chaque lettre isolée.

Dans cette section, nous présenterons plusieurs exemples nous permettant d'illustrer ces deux propriétés. Premièrement, nous présenterons l'exemple de pseudo-mots commençant par un segment ambigu (tel que « CH » en français et « TH » en anglais) afin de montrer comment les relations orthographe-phonologie ambiguës sont représentées dans le modèle lexical statique. Puis, nous présenterons un exemple à l'aide duquel nous comparerons le résultat quand un segment donné, en l'occurrence « AI », est en début ou au milieu du mot. Enfin, nous présenterons quelques cas d'irrégularité pour comparer le résultat de la conversion orthographe-phonologie obtenue par segments avec celui obtenu par traitement global.

2.3.1 Segments ambigus : « CH », « TH » et « ON »

Afin d'illustrer ces traitements sous-lexicaux, nous examinons tout d'abord le cas de segments orthographiques qui présentent une ambiguïté. Pour cela, nous considérons trois exemples de segments : « CH », « TH » et « ON ».

Commençons par le premier exemple, le cas de « CH ». En effet, selon son voisinage orthographique, la lettre « C » peut être associée à plusieurs phonèmes en français : /s/, /k/ ou /ʃ/ (quand elle est suivie d'un « H »). Elle peut également être muette, comme dans « tabac ». Nous utilisons ici notre version statique du sous-modèle lexical pour calculer la distribution de probabilité sur les phonèmes pour cette lettre.

Lorsque l'entrée est uniquement la lettre « C », et que nous considérons qu'elle est en première position, alors la distribution de probabilité $P(\Phi_1 \mid [L_1 = \text{'C'}])$ calculée dans le modèle donne aux trois phonèmes /s/, /k/ et /ʃ/ des probabilités élevées, et dont les valeurs reflètent essentiellement leur fréquence d'occurrence dans le lexique (voir Figure 4.6).

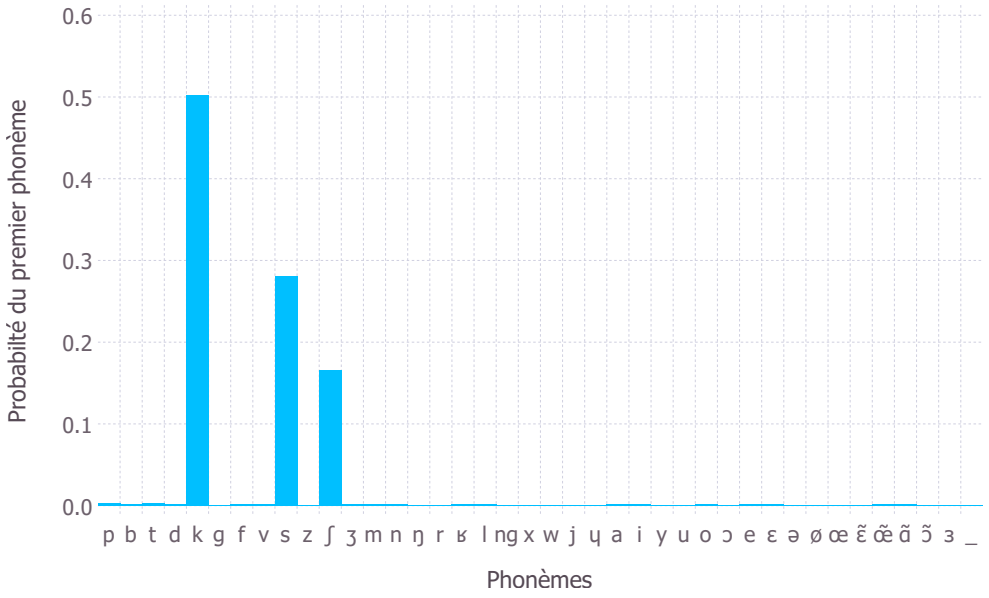


FIGURE 4.6 – **Distribution de probabilité des phonèmes sachant quel la première lettre est un « C ».** En abscisse, les phonèmes possibles ; en ordonnée, la valeur de probabilité de la distribution $P(\Phi_1 \mid [L_1 = \text{'C'}])$ pour chaque phonème.

Si on considère maintenant des éléments contextuels en rajoutant, au fur et à mesure, des informations sur les lettres qui suivent, alors la distribution de probabilité sur les phonèmes calculée va changer. Nous illustrons cela sur la séquence de lettres « CHORAT », et nous commençons par un segment d'une seule lettre puis nous augmentons la taille du segment jusqu'à six lettres. Ainsi, nous calculons $P(\Phi_1 \mid [L_1 = \text{'C'}])$, $P(\Phi_1 \mid [L_{1:2} = \text{'CH'}])$, etc, jusqu'à $P(\Phi_1 \mid [L_{1:6} = \text{'CHORAT'}])$ (voir Figure 4.7a). Les probabilités des phonèmes avec une seule lettre sont « incertaines » (et identiques à la distribution montrée dans la Figure 4.6), puis les segments « CH » et « CHO » donnent un avantage au phonème /ʃ/ dans la première position (comme dans « chocolat »). L'identité du premier phonème redevient incertaine (partagée entre /ʃ/ et /k/) lorsque l'on présente au modèle le segment « CHOR ». Enfin, lorsque l'on présente au modèle « CHORA » ou « CHORAT », la certitude du modèle augmente et la probabilité que le premier phonème soit un /k/, comme dans le mot « chorale », augmente ; tandis que les prononciations alternatives (/s/ ou /ʃ/) sont très peu probables.

Dans un second exemple (le cas de « TH »), avec un lexique anglais, nous obtenons des résultats similaires (voir Figure 4.7b). La lettre « T » isolée, en entrée et en première position, induit une forte probabilité pour le phonème /t/ dans la première position phonologique, comme dans le mot « *table* ». Le segment « TH » induit, lui, une forte probabilité de retrouver le phonème /ð/, comme dans le mot « *there* ». En se rapprochant du mot anglais « *thief* », le segment « THIE » rend le phonème /θ/ est le plus probable. En revanche, un segment très grand, comme « THIELD »

(a) *Lexique en français*

(b) *Lexique en anglais*

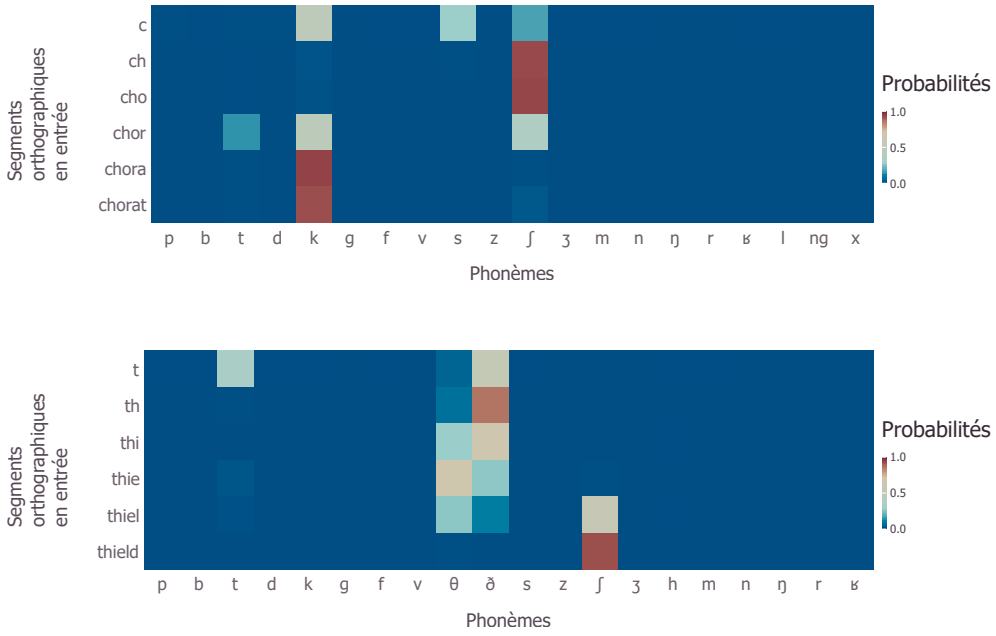


FIGURE 4.7 – **Distributions de probabilité du premier phonème pour des entrées orthographiques étant des séquences de lettres de taille variable.** (a) Exemple en français, où l'entrée va d'une unique lettre « C » à une séquence de six lettres « CHORAT ». (b) Exemple en anglais où l'entrée va d'une unique lettre « T » à une séquence de six lettres « THIELD ». La couleur de chaque case représente la probabilité de ce phonème. Seuls les phonèmes consonantiques, et les probabilités des consonnes les plus élevées sont représentées, pour des raisons de lisibilité.

provoque une lexicalisation. En effet, le phonème /ʃ/ devient le plus probable à cause de la similarité avec le mot « *shield* ». En effet, il n'existe aucun mot, dans le lexique anglais, possédant cette sous-séquence de lettres, et la « médiation forcée par l'espace lexical » induit la lexicalisation.

Enfin comme précédemment, dans le cas de la séquence « ON », la prononciation générée dépend fortement du contexte (voir Figure 4.8). Si l'on compare les distribution de l'identité des phonèmes pour les segments « BO », « BON », « BONN », et « BONT », nous constatons que l'identité des phonèmes générés en sortie dépend là-encore du nombre de lettres contextuelles prises en compte. La prononciation du segment « BON » reste relativement ambiguë lorsque le contexte qui suit immédiatement n'est pas connu. La prononciation /õ/ est légèrement dominante mais elle est fortement concurrencée par la prononciation /ɔ/. Par contre, la prononciation est clairement désambiguïsée dès que le contexte immédiat est précisé, « BONN » ou « BONT », donnant alors clairement la prononciation /ɔ/ dans le premier cas et /õ/ dans le second. On voit par ailleurs que le seul segment « BO » présenté en entrée sans aucune information contextuelle, conduit à une forte incertitude quant à sa prononciation. Il génère alors comme prononciation possible /w/ (comme dans « bois »), /u/ comme dans « bout », /ɔ/ comme dans « botte » ou /õ/ comme dans « bonté » et très peu probablement /o/.

En résumé, l'information contextuelle d'un segment orthographique est indispensable pour désambiguïser la sortie phonologique correspondante. De plus, le choix de la longueur du seg-

FIGURE 4.8 – **Distributions des probabilités du deuxième phonème pour chaque segment orthographique en entrée, notée $P(\Phi_2 \mid L_{1:k})$.** Nous comparons ici les distributions en sortie pour différents segments (« BO », « BON », « BONN » et « BONT »).

ment est crucial. En effet, deux lettres vont le plus souvent suffire pour désambiguïser la prononciation d'un segment « simple » (par exemple, un « C » sera prononcé /k/ s'il est suivi de « R, L, A, O, ou U » /s/ s'il est suivi de « I, E ou Y », et /ʃ/ s'il est suivi de « H ». Cependant, un segment plus large (et donc plus d'information contextuelle) conduit à une sortie phonologique plus proche de celle des mots du lexique contenant ce segment. Cela peut générer une prononciation « par analogie avec un mot irrégulier », comme c'est le cas pour « CHORA » où « CH » est prononcé /k/, ou entraîner une lexicalisation, comme c'est le cas pour « THIELD ». Etant donné la nature opaque de l'orthographe anglaise, l'ambiguïté du premier phonème reste élevée même lorsque l'information contextuelle est importante mais la présence d'un voisin orthographique (*shield*) conduit finalement à une lexicalisation et attribuant « TH » le phonème /ʃ/.

2.3.2 Position du segment : Le cas de « AI »

Dans cet exemple, nous évaluons l'influence de la position orthographique lors de la production d'un segment donné. Prenons ici l'exemple de « AI ». Dans un premier temps, nous nous limitons à la prononciation de la lettre « A » selon la position dans laquelle elle se trouve (voir Figure 4.9). Lorsque « A » est présenté en troisième position (comme dans le cas de « PLAINI ») et qu'on calcule la distribution de probabilité des phonèmes dans cette position, la distribution est « incertaine » et la probabilité se « répartit » entre les phonèmes /ɛ/, /ɛ̃/ ou /ɑ̃/ (voir Figure 4.9a). Cependant, lorsqu'on calcule la distribution de probabilité des phonèmes pour cette même lettre mais présentée en première position, on obtient alors une probabilité élevée pour le phonème /a/ (voir Figure 4.9b). Ainsi, le choix entre les différentes alternatives va être déterminé par l'augmentation de la taille du segment, ce qui revient à donner plus d'information sur le contexte orthographique de la lettre.

Dans un second temps, considérons la prise en compte d'une lettre contextuelle, ce qui correspond au segment « AI », en position initiale ou au milieu du mot. Ce segment est susceptible d'être associé aux phonèmes /e/ (comme dans « aimant »), /ɛ/ (comme dans « aide » ou « plaine »), /a/ (comme dans « ail » ou « braille ») ou encore /ɛ̃/ (lorsqu'il est suivi de « n », comme dans « ainsi » ou « plainte »). Dans le cas de « PLAINI » (voir Figure 4.9a), le segment « PLA » induit une probabilité élevée de retrouver les phonèmes /a/ ou /ɛ/ en troisième position. L'ajout d'une lettre supplémentaire « I », fait basculer la probabilité vers les phonèmes /e/, /ɛ/ ou /ɛ̃/. En se rapprochant donc des mots comme « plaine » , « plaisir », « plaie », les probabilités que le phonème en troisième position

FIGURE 4.9 – **Distributions des probabilités des phonèmes pour chaque segment orthographique en entrée.** Nous affichons la distribution à la troisième position phonologique $P(\Phi_3 \mid L_{1:k})$ dans le premier cas (a) ; tandis que nous affichons celle à la première position $P(\Phi_1 \mid L_{1:k})$ dans le second cas (b). k est la longueur du segment considéré.

soit /e/ ou /ɛ/ sont plus fortes. Le segment « PLAIN » correspond à un mot existant dans le lexique (mot peu fréquent, signifiant plat et sans relief), d'où la forte probabilité que le troisième phonème soit un /ɛ̃/. Enfin, la présentation de « PLAINI » induit une augmentation de la probabilité que le phonème soit /e/.

Dans le cas de « AI » placé en début de mot, la probabilité de retrouver /ɛ/ est forte (voir Figure 4.9b). L'ajout d'informations, à la suite de ces deux premières lettres, modifie la distribution de probabilités sur les phonèmes. En effet, selon la présentation qui est faite en entrée, le choix du phonème en première position varie (et donc la probabilité de retrouver un phonème plutôt qu'un autre) : la présentation de « AIN » donne le phonème /ɛ̃/ ; la présentation de « AIL » donne /a/ ; et la présentation de « AILE » donne /ɛ/. Ainsi, l'ajout d'informations contextuelles induit des changements dans les distributions de probabilité phonémiques calculées.

Ces exemples nous montrent que le segment choisi en entrée fait varier la distribution de probabilité phonémique calculée, non seulement en fonction de sa longueur, mais également en fonction de sa position. En d'autres termes, dans notre modèle statique, la prononciation d'un pseudo-mot dépend de la taille du segment considéré en entrée (et donc du contexte orthographique) et de la position de ce segment dans le pseudo-mot.

2.3.3 Lecture par segments de mots irréguliers

Nous considérons ici des exemples de stimuli irréguliers, c'est-à-dire des mots dont la prononciation ne découle pas de l'application des règles habituelles de conversion entre graphèmes et phonèmes, comme les mots « huit » et « monsieur » en français. Nous évaluons l'effet de la segmentation dans ces cas puis nous comparons la sortie phonologique des mots irréguliers à celle produite pour des mots réguliers.

Le cas de « UIT » : entre nuit et huit Ici, nous nous intéressons au cas de la sous-séquence de lettres « UIT ». Ce segment est irrégulier dans le cas de « huit » (prononcé /ɥit/, le « T » n'est pas muet) mais régulier dans le cas de « nuit », « cuit », « suit » ou « fruit » (dans lesquels le segment « UIT » se prononce /ɥi/, le « T » étant muet).

Nous présentons trois exemples. Dans le premier exemple, seul « UIT » est fourni au modèle (voir Figure 4.10a), en indiquant que ce segment est entre les positions 2 et 4, c'est-à-dire qu'il y a une lettre qui précède, en position 1, mais dont on ne donne pas l'identité au modèle. Les distributions de probabilité calculées pour le premier et le dernier phonème (position 1 ou 4) sont alors assez incertaines. En effet, dans la première position, la probabilité des phonèmes se partage entre /n/, /k/ et /s/ (vraisemblablement à cause des mots « nuit », « cuit » et « suit »). En quatrième position, la probabilité se partage entre /e/, /t/ ou /#/ (le phonème supplémentaire indiquant la fin de la séquence). En revanche, les distributions de probabilité sur les phonèmes pour la portion « centrale » (positions 2 et 3) sont plus concentrées, avec une valeur élevée de probabilité pour /ɥi/, qu'on peut considérer comme la production correcte.

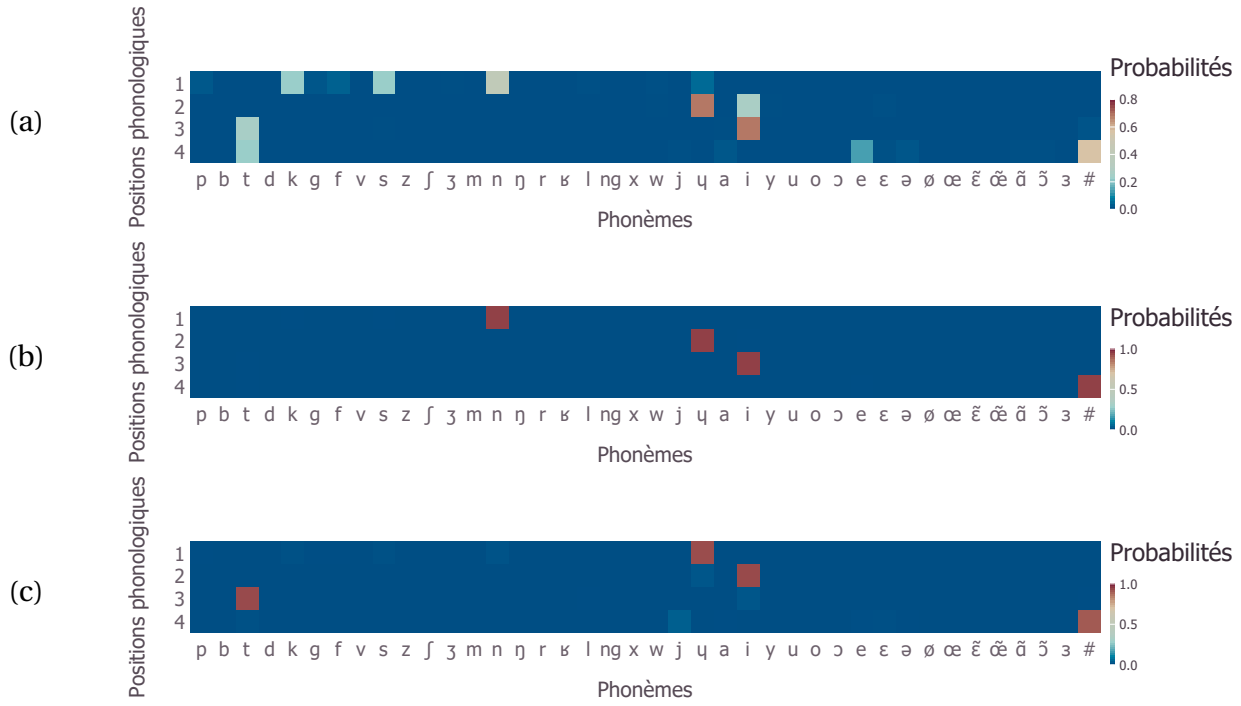


FIGURE 4.10 – **Distributions des probabilités des phonèmes pour différents segments.** Seules les probabilités des cinq premières positions phonologiques ($\Phi_{1:5}$) pour les segments « -UIT » (a, $L_{2:4}$ = *'UIT'*) ; « NUIT » (b, $L_{1:4}$ = *'NUIT'*) ; « HUIT » (c, $L_{1:4}$ = *'HUIT'*) sont affichées ici.

Dans les deux autres exemples, le modèle reçoit soit « NUIT », soit « HUIT » comme séquence de lettres en entrée (voir Figure 4.10b, c). Nous constatons que spécifier l'identité de la première lettre du stimulus permet de diminuer l'incertitude dans les distributions de probabilité calculées, en particulier pour la prononciation de la fin du mot. En effet, dans cet exemple, la lettre « T » en quatrième position peut être muette ou non. Ainsi, le fait de donner tout le stimulus en entrée permet de désambiguïser et donc d'augmenter la probabilité d'observer, en quatrième position, le phonème /t/ pour « HUIT » et /#/ pour « NUIT » (le « T » est muet).

Le cas de forte irrégularité : « monsieur » et « enivrer » Nous proposons ici une autre illustration du phénomène de prononciation correcte, ou non, d'une séquence de lettres par analogie avec un mot irrégulier de la langue. Dans cet exemple, nous considérons le mot irrégulier « monsieur ». En effet, en ne présentant au modèle que le segment « MON », il est lu /mɔ̃/ par analogie aux mots réguliers du lexique français (par exemple, « mon », « mont » ou « monde ») ; tandis que lorsqu'on donne « MONSIEUR » en entrée, les phonèmes correspondants aux trois premières positions phonologiques sont lus /məs/ (voir Figure 4.11).

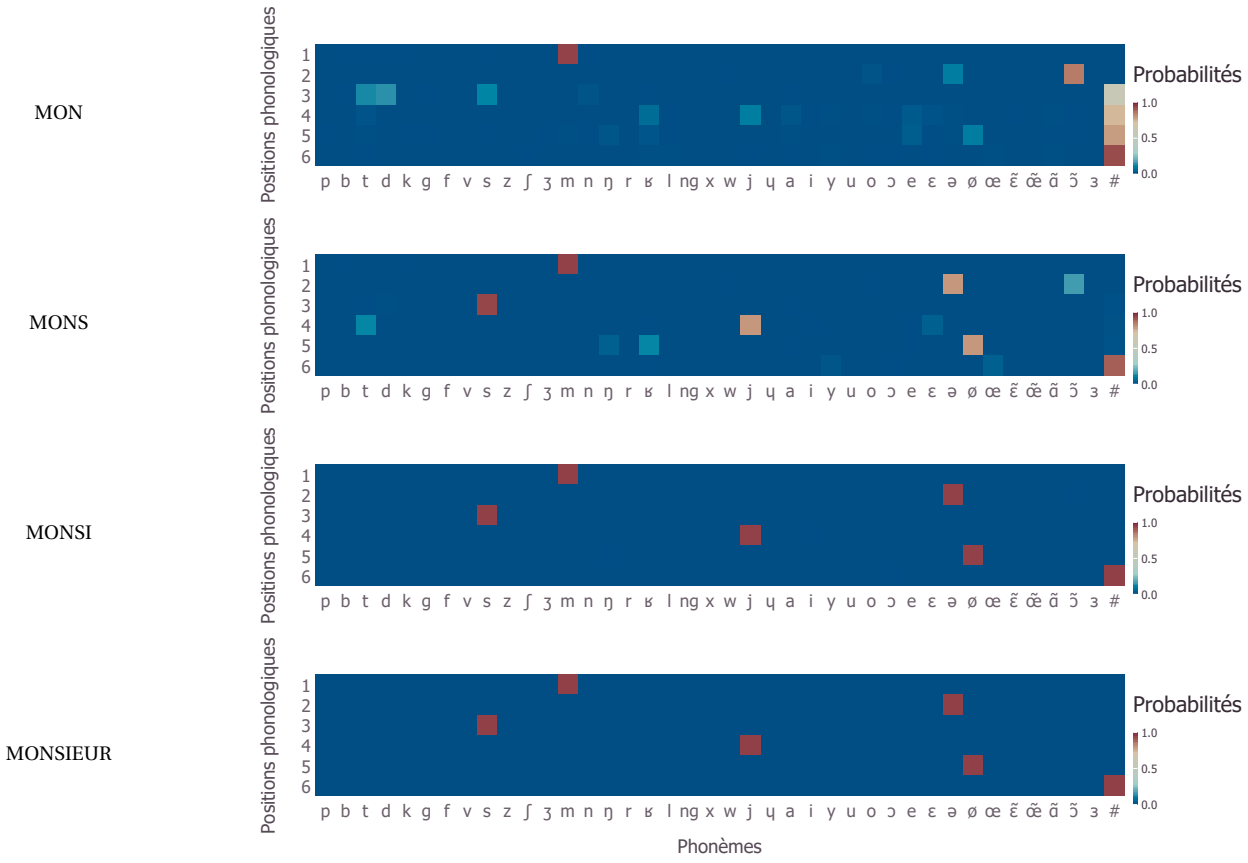


FIGURE 4.11 – **Distributions des probabilités des phonèmes pour les six premières positions phonologiques ($\Phi_{1:6}$).** En entrée, nous utilisons les segments « MON », « MONS » et « MONSI » puis le stimulus « MONSIEUR » en entier

Nous remarquons, de plus, qu'il n'est pas nécessaire d'avoir la séquence de lettres « MONSIEUR » complète en entrée pour produire la sortie phonologique /məs/. En effet, avec « MONSI » en entrée, le modèle est déjà capable d'identifier avec une très haute probabilité que l'entrée correspond au mot « monsieur », et donc d'inférer la prononciation correcte correspondante. Remarquons, enfin, que cela permet non seulement de « bien prononcer le ON irrégulier » du début de la séquence, mais également de « prédire la fin du mot », et donc à la fois la fin de la séquence phonémique, mais aussi, la fin de la séquence orthographique (même si nous ne le montrons pas ici, ni mathématiquement, ni dans la Figure 4.11). Présenter au modèle l'entrée « MONS » produit un résultat intermédiaire, avec une probabilité qui a déjà son maximum pour la « prononciation irrégulière » /məs/, même si la prononciation régulière /mɔ̃/ est encore en compétition.

Le nombre de lettres qu'il faut présenter en entrée pour obtenir la prononciation irrégulière

correcte dépend évidemment du mot considéré. Nous montrons Figure 4.12 l'exemple du mot « enivrer » dans lequel très peu de lettres suffisent. En effet, si beaucoup de mots commencent par « EN », et si le phonème correspondant pour ces lettres est habituellement /ɑ̃/, dès que le modèle a en entrée les lettres « ENI », il devient très probable que la séquence débute par les phonèmes /ɑ̃n/, c'est-à-dire que le « N » se prononce également (alors qu'il est nasalisé dans la prononciation régulière). L'observation des distributions de probabilité calculées sur les phonèmes suggère que le modèle, sur la base de l'entrée « ENI », donne déjà une probabilité élevée au mot « enivrer », et donc à la prononciation irrégulière correspondante.

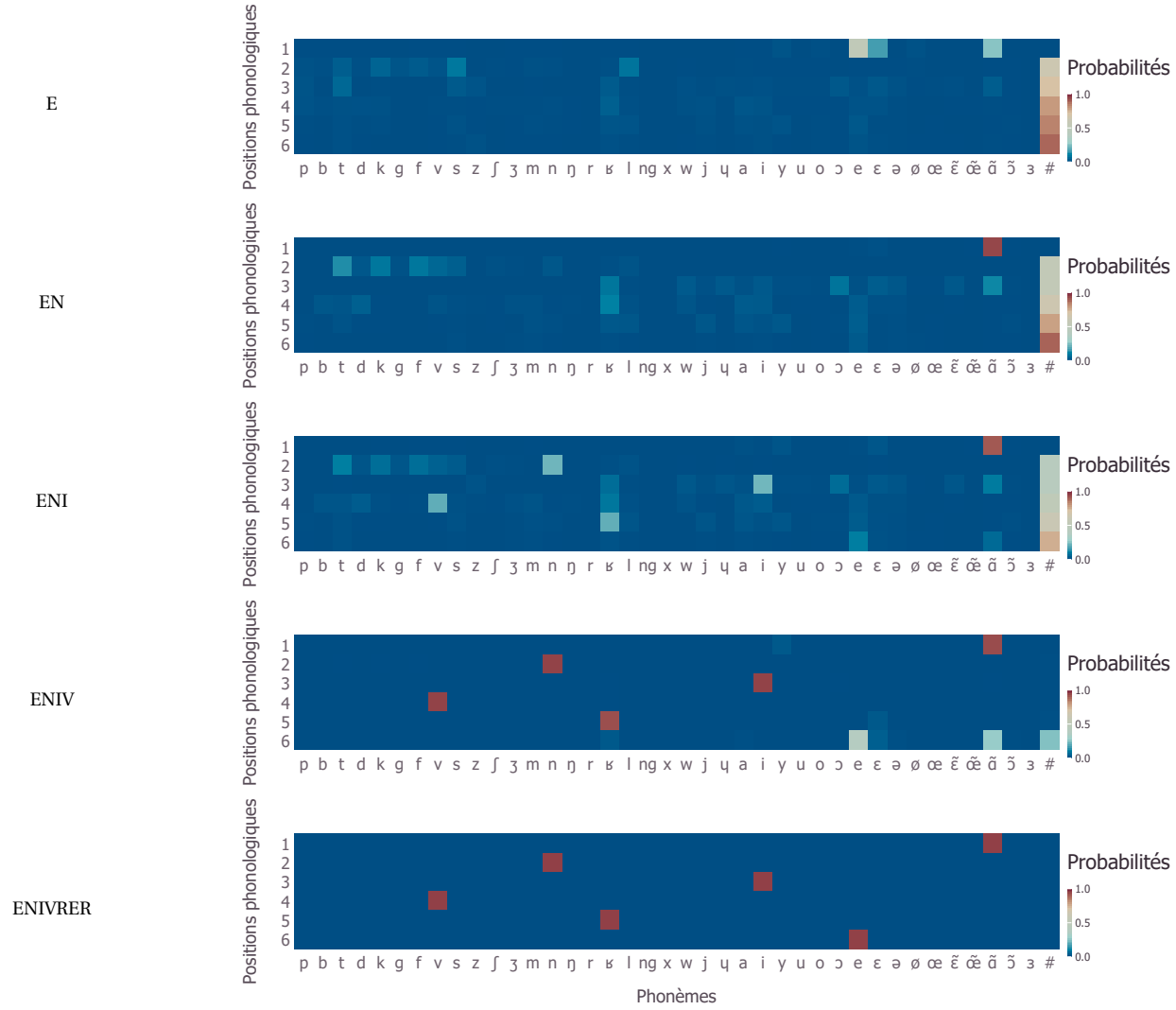


FIGURE 4.12 – **Distributions des probabilités des phonèmes pour les six premières positions phonologiques ($\Phi_{1:6}$).** En entrée, nous utilisons les segments « E », « EN », « ENI » et « ENIV » puis le stimulus « ENIVRER » en entier

Ainsi, les deux exemples présentés ici (« huit »/« nuit » et « monsieur », avec le support de l'exemple « enivrer ») nous montrent que pour lire correctement un mot irrégulier, il n'est pas nécessaire d'avoir la séquence complète des lettres en entrée. En effet, il suffit d'avoir suffisamment de lettres pour « désambiguïser suffisamment » au niveau lexical : « MONSI » suffit pour identifier le mot « monsieur » et la prononciation résultante. Il n'est donc pas nécessaire de considérer

qu'un traitement portant sur l'ensemble des lettres du mot est indispensable à la lecture des mots irréguliers. Une lecture par segment peut fonctionner, tant que le segment est suffisamment long pour inclure le point d'unicité distinguant ce mot de tous les autres mots du lexique.

A l'inverse, considérer des segments trop courts induiraient des erreurs de régularisation. En effet, si on fournit au modèle trois segments à la suite (MON-----, ---SI--- puis -----EUR), alors on obtient la prononciation « régularisée » et donc incorrecte /mɔ̃siœʁ/. Cela illustre, une fois de plus, la relation forte entre la segmentation et la prononciation qui en résulte.

2.4 Synthèse intermédiaire

Dans la Section 1 de ce chapitre, nous avons étudié le comportement « spontané » de BRAID-Phon lors de la lecture de pseudo-mots. Celui-ci se caractérisait par une lexicalisation systématique des stimuli. La lexicalisation se produit non seulement lorsque le pseudo-mot est traité en une seule capture attentionnelle mais également lorsqu'il est traité sériellement en plusieurs captures attentionnelles.

L'analyse du calcul opéré par le modèle a montré que cette lexicalisation systématique était due, d'une part, au passage « obligatoire » par l'espace lexical, entre les lettres et les phonèmes et, d'autre part, à l'opération du calcul au niveau de l'unité mot. Cela questionne la capacité de cette structure de connaissance « en une seule voie » à effectuer le traitement sur des unités plus petites que l'unité mot afin de générer une séquence phonologique plausible, lorsqu'il n'y a pas de point de l'espace lexical permettant cette conversion.

Dans la Section 2, nous avons donc étudié la capacité du modèle à produire les correspondances entre lettres et phonèmes lorsqu'il est alimenté par des sous-séquences de lettres en entrée. Nous avons montré que cela permettait le calcul de correspondances sous-lexicales entre segments orthographiques et phonèmes. Nous avons observé qu'en limitant le traitement à des segments, les connaissances lexicales conduisent à générer les phonèmes attendus. Ainsi, la lecture par segments permet la prononciation des pseudo-mots et des « nouveaux stimuli », comme les mots à apprendre.

Cependant, notre raisonnement mathématique et les simulations qui les illustrent montrent que le mécanisme de traitement sous-lexical est sensible au choix des segments. Plus spécifiquement, la génération phonologique appropriée implique la prise en compte de segments de taille variable selon les stimulus. Ainsi, si le segment de lettres est très grand, le modèle risque de lexicaliser ; au contraire, si le segment de lettres est très petit et ne fournit pas suffisamment d'information sur le contexte, le modèle risque de faire des erreurs (par exemple, en se limitant au « C » dans le traitement d'un stimuli contenant « CH ») ou de régulariser si le mot à lire est irrégulier (par exemple, en prononçant /ʃɔʁal/ le mot « chorale »). De plus, la position du segment dans le stimulus est un autre facteur qui affecte la prononciation. En effet, les phonèmes générés quand un segment est au début de stimulus diffèrent de ceux générés quand le même segment est au milieu du stimulus.

3 Lecture par segments avec BRAID-Phon

Nous avons explicité précédemment les problèmes relatifs à la lecture de pseudo-mots par le biais d'un modèle lexical statique limité à la structure des connaissances lexicales du modèle BRAID-Phon. Dans cette section, nous considérons le modèle initial complet, en « reconnectant le reste du modèle », et décrivons comment effectuer la lecture de pseudo-mots dans le modèle BRAID-Phon.

Pour cela, nous présentons, dans la suite de ce chapitre, un mécanisme de lecture basé sur le traitement séquentiel de plusieurs segments pour le modèle BRAID-Phon. Celui-ci repose sur la manipulation des paramètres de la distribution attentionnelle pour « ne traiter qu'un segment du stimulus ». La définition des segments reposant sur la distribution attentionnelle, les segments seront donc définis de manière flexible et adaptable au stimulus et aux conditions de son traitement. Pour réaliser ce mécanisme, nous décrivons tout d'abord comment le traitement d'un segment du stimulus s'opère dans le modèle complet, puis nous introduisons au modèle un sous-modèle supplémentaire (le sous-modèle d'attention phonologique), pour « coordonner la sortie phonologique » avec la distribution visuo-attentionnelle, et donc s'assurer que le segment phonologique en sortie correspond au segment en cours de traitement visuel. Enfin, nous montrons, à travers quelques exemples de lecture de pseudo-mots, certaines propriétés et limites de ce mécanisme.

3.1 Modèle

Le mécanisme que nous allons introduire ici et qui permet la lecture par segments se compose de deux éléments : le premier élément est un mécanisme de traitement de segments au niveau lexical, et le second élément est un nouveau module attentionnel au niveau phonologique (défini de manière symétrique au sous-modèle de l'attention visuelle). De plus, ces deux éléments seront pilotés par le module de l'attention visuelle. Ainsi, nous pourrons simuler, dans le modèle, le calcul séquentiel de segments phonologiques, sur la base de segments orthographiques.

3.1.1 Segments dans le sous-modèle lexical

Comme nous l'avons montré précédemment, la lecture de pseudo-mots nécessite d'effectuer un traitement par segments du stimulus. Il est aisé de représenter, dans le modèle, un segment orthographique ou phonologique : il suffit d'imaginer que toutes les lettres, ou tous les sons, du pseudo-mot présenté en entrée ne sont pas donnés, ou ne sont pas perçus. Mathématiquement, on ne spécifie alors les variables P ou Ψ que pour quelques positions, ou, de manière équivalente, on donne des distributions uniformes pour les autres positions. En revanche, nous avons vu que le sous-modèle lexical lexicalise facilement, c'est-à-dire qu'il infère efficacement sur l'espace des mots possibles, ce qui l'amène à obtenir des distributions de probabilité précises sur les phonèmes en toutes positions, même sur la base d'un seul segment orthographique. Notre objectif est à présent de formuler mathématiquement les mécanismes permettant de limiter, dans le modèle, la capacité à inférer « le reste du mot » lorsqu'on ne fournit qu'un segment orthographique.

Pour cela, nous choisissons de définir les segments à considérer par le sous-modèle lexical grâce aux variables de cohérence $\lambda_{L1:N}$. Ces variables permettent de contrôler le flux d'information

entre le niveau lexical et le niveau perceptif des lettres et fonctionnent comme des interrupteurs ON/OFF permettant à l'information de circuler ou non (Gilet et al., 2011) entre ces deux niveaux[14]. Dans les calculs probabilistes que nous considérons, ces variables sont supposées avoir une valeur de 1, impliquant une circulation complète de l'information du sous-modèle perceptif des lettres vers le sous-modèle lexical, quelle que soit la position considérée.

Pour restreindre le transfert d'information aux positions correspondant à un segment donné, il faut ne transférer l'information orthographique au modèle lexical que pour ces positions là. Ainsi, nous fixons à 1 la valeur des variables λ_L dans le segment, et à 0 celle des variables λ_L en dehors du segment[15].

Afin d'illustrer comment nous contrôlons ces variables de cohérence λ_L, reprenons notre exemple initial du stimulus « VIRDIN ». Pour restreindre le traitement lexical uniquement aux trois premières lettres, nous fixons les variables $\lambda_{Li} = 1$ pour i égal à 1, 2 ou 3. Ainsi, la totalité de l'information accumulée au niveau de ces trois positions orthographiques transite au niveau lexical. En revanche, la valeur 0 est attribuée aux variables λ_{L4}, λ_{L5} et λ_{L6}, empêchant ainsi la circulation de l'information sur les positions orthographiques respectives. A l'inverse, pour traiter le second segment sur les lettres en positions 4, 5 et 6, alors on supposera $\lambda_{L4:6} = 1$ et $\lambda_{L1:3} = 0$.

3.1.2 Sous-modèle de l'attention phonologique

Au cours d'un traitement par segment, le modèle doit générer uniquement les phonèmes dans les positions phonologiques qui correspondent au segment orthographique traité. Par exemple, si le modèle traite le segment « VIR » du stimulus « VIRDIN », on vise à calculer l'identité des premiers phonèmes du stimulus, les trois premiers dans l'idéal, et non pas les derniers phonèmes.

L'information sur l'identité des phonèmes est accumulée au niveau du sous-modèle phonologique et portée par les variables $\Psi^t_{1:M}$ à chaque itération t. BRAID-Phon, tel que décrit dans le Chapitre 2, suppose un transfert complet de l'information entre le sous-modèle lexical et le sous-modèle phonologique. L'introduction d'un sous-modèle d'attention phonologique a pour objet de permettre un transfert au niveau phonologique de la seule information pertinente pour le traitement en cours, celle qui correspond au segment orthographique traité. Cela implique que le sous-modèle d'attention phonologique soit couplé au sous-modèle d'attention visuelle.

D'un point de vue computationnel, ce sous-modèle est analogue au sous-modèle d'attention visuelle et il décrit de la même manière une distribution d'attention qui contrôle le transfert d'in-

[14] Pour être précis, il y a deux couches de variables de cohérence entre le niveau perceptif des lettres et le niveau lexical, pour contrôler indépendamment les flux montant et descendant d'information. Le flux descendant représente les influences *Top-Down* du niveau lexical vers le niveau perceptif des lettres et il est contrôlé par le résultat de la décision lexicale via la variable Γ_{TD} (Ginestet, 2019). Ici, nous ne considérons que le flux montant et donc l'information transitant de l'orthographe vers le lexique.

[15] Une valeur de λ_{Li} égale à 0, pour une position i, signifie que l'on suppose que, lors du calcul de l'inférence, les identités des lettres portées au niveau de la variables P_i et L_i sont différentes. En d'autres termes, cela fait l'hypothèse que les lettres dans les positions hors-segment sont différentes de celles potentiellement perçues. Une alternative consisterait à ne pas propager l'information du tout entre P_i et L_i, en ne supposant pas de valeur connue pour la variable λ_i.

formation. A chaque itération t, cette distribution est portée sur la variable ${A_\Phi}^t$ et elle est supposée gaussienne telle que ${\mu_\Phi}^t$ est sa moyenne et ${\sigma_\Phi}^t$ son écart-type (sa dispersion). Les variables de contrôle $C_{\Phi 1:M}$ permettent de filtrer, en accord avec la distribution de l'attention, la quantité d'information phonologique transmise du sous-modèle lexical vers le sous-modèle phonologique. Ceci est implémenté en agissant sur des variables de cohérence $\lambda_{\Phi 1:M}$, de manière similaire à l'implémentation dans le sous-modèle d'attention visuelle (voir Section 1 du Chapitre 2, page 48).

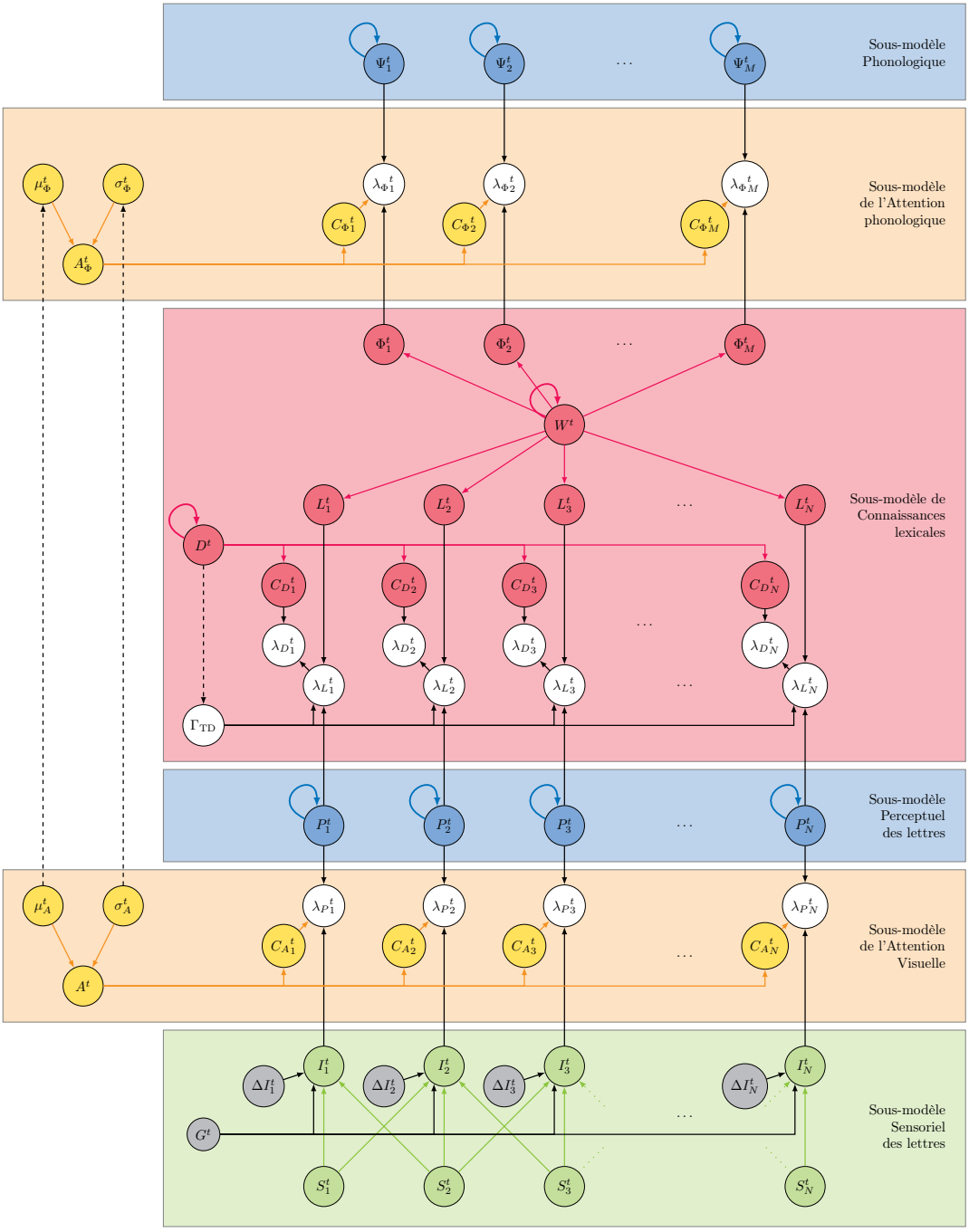


FIGURE 4.13 – **Représentation graphique du modèle BRAID-Phon contenant le module d'attention phonologique.** Les nœuds représentent les variables du modèle et les flèches en trait plein représentent les dépendances probabilistes entre les variables. Les flèches en pointillés indiquent une affectation de valeur de la variable du nœud d'arrivée se basant sur l'état de la variable dans le nœud du départ.

3.1.3 Modèle entier

Le lien entre l'attention phonologique et l'attention visuelle du modèle est établi en fixant les valeurs des variables μ_Φ et σ_Φ à partir de μ_A et σ_A (voir le schéma du modèle global Figure 4.13). Cette relation entre les paramètres de la distribution phonologique et de la distribution visuo-attentionnelle est formulée à l'aide d'une fonction linéaire que nous notons f_A (voir Figure 4.14). Celle-ci représente le lien entre la longueur phonologique et la longueur orthographique. Elle est obtenue après la réalisation d'une régression linéaire entre les longueurs orthographiques et phonologiques de l'ensemble des mots du lexique français Lexique 3 (New, 2006). La fonction $f_A(n)$ obtenue est $f_A(n) = 0{,}12 + 0{,}73 * n$: la longueur phonologique est, en moyenne, 73 % de la longueur orthographique. Ainsi, nous pourrons admettre, à partir du résultat de cette régression, qu'un mot qui possède n lettres (un mot de longueur orthographique n) possède une longueur phonologique moyenne de $f_A(n)$ phonèmes.

Rappelons que les variables μ_A^t et σ_A^t sont définies sur un espace continu. Il est donc possible de définir la position du focus visuo-attentionnel sur une position orthographique donnée (par exemple, pour la 3ème lettre d'un stimulus $\mu_A^t = 3$) ou entre deux positions orthographiques (par exemple, pour avoir une attention visuelle fixée entre les 2ème et 3ème positions orthographiques, nous définissons μ_A^t à la valeur 2,5). De la même manière, nous considérons des positions non-entières également au niveau phonologique. Dans ce cas, une attention visuelle centrée sur un μ_A^t correspond à centrer l'attention phonologique sur $\mu_\Phi^t = f_A(\mu_A^t)$ et une dispersion σ_A^t de l'attention visuelle correspond à une dispersion σ_Φ^t de l'attention phonologique égale à $f_A(\sigma_A^t)$ phonèmes du coté phonologique (voir Figure 4.15).

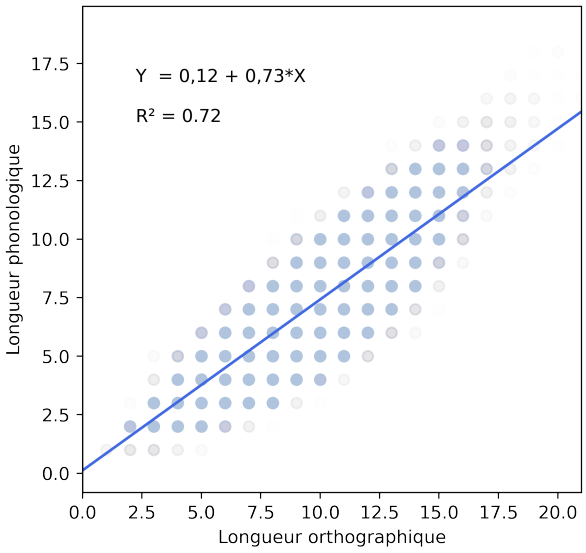


FIGURE 4.14 – **La fonction f_A issue de la régression linéaire entre longueur phonologique et longueur orthographique du lexique français.** La fonction correspond à la droite bleue et le nuage de points représente tous les mots du lexique français.

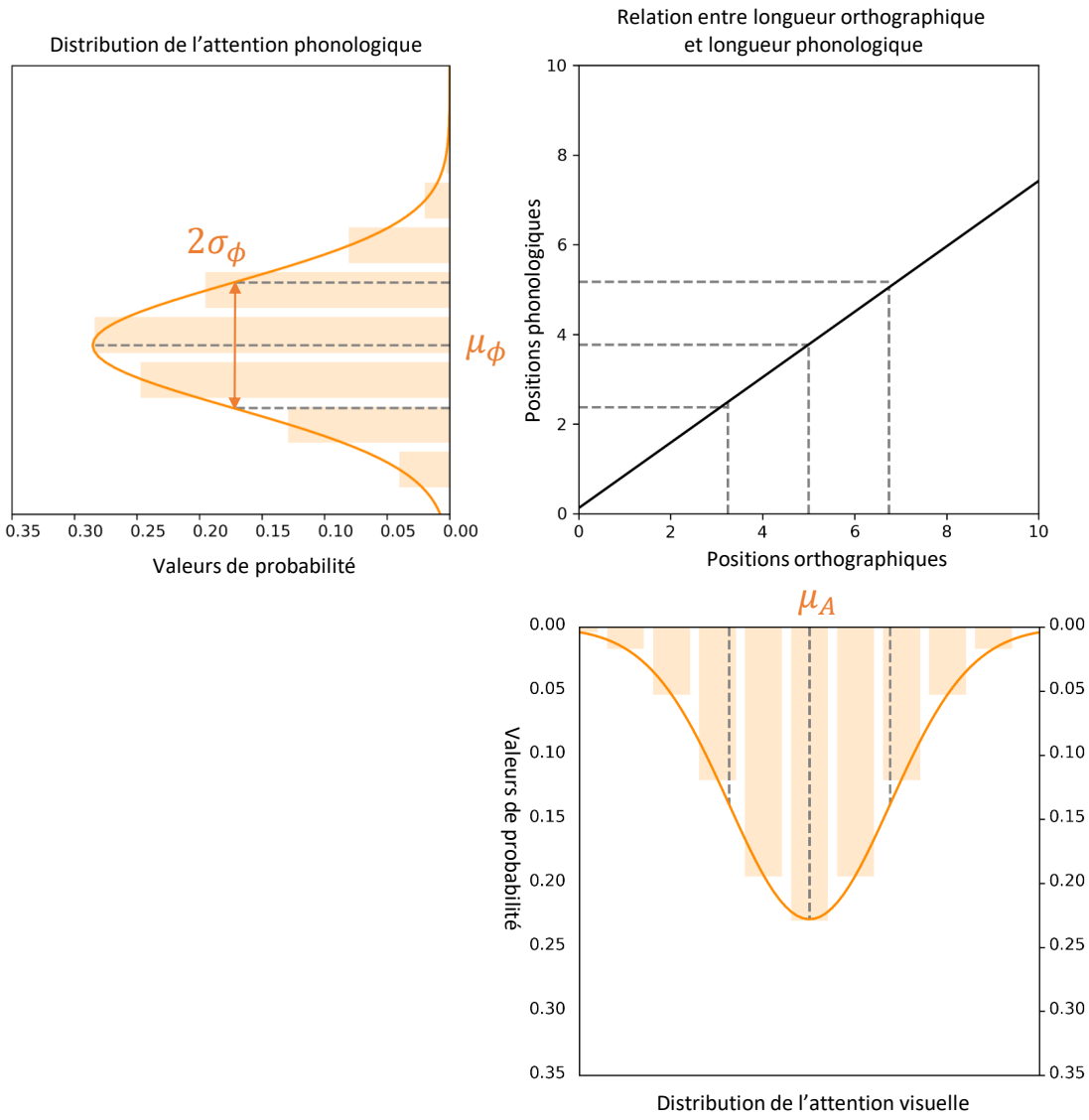


FIGURE 4.15 – **Lien entre les paramètres de l'attention visuelle et de l'attention phonologique.** La fonction de lien f_A est tracée en haut à droite. Nous représentons, en bas, la distribution de l'attention visuelle et, en haut à gauche, la distribution de l'attention phonologique correspondante. Les courbes en cloche sont les tracés de gaussiennes de paramètres μ et σ tandis que les bâtons représentent la « quantité » d'attention accordée à chaque position, phonologique ou orthographique, étant donné le caractère discret de ces grandeurs.

3.2 Inférence

3.2.1 Paramètres pour représenter une séquence de segments

Pour traiter une séquence de segments, il nous faut la définir formellement. Ainsi, nous devons définir leur nombre, et leurs étendues, en termes de positions et de durées. Pour définir la position d'un segment, nous partons des paramètres des captures attentionnelles, c'est-à-dire la position μ_A^t du focus visuo-attentionnel et la dispersion visuo-attentionnelle σ_A^t. Les paramètres de l'attention phonologique en découlent directement (voir Section 3.1.3). Nous définissons les valeurs des variables de cohérence λ_L pour limiter l'accès sous-lexical sur les positions correspondant au focus visuo-attentionnel (dans les exemples illustratifs qui suivent, cela est fait au cas par cas, pour

ajuster les λ_L mis à 1 en fonction de la distribution visuo-attentionnelle). Enfin, nous choisirons « à la main » le nombre de segments, pour chaque exemple. Pour simplifier, nous supposerons que chaque segment est traité pour une durée fixe, que nous noterons T_f.

3.2.2 Question probabiliste

Afin de lire un stimulus, nous calculons la distribution de probabilité des phonèmes à chaque itération. Il s'agit, comme dans le cas « spontané » (décrit dans les Chapitre 2 et 3 ainsi que dans la Section 1 de ce chapitre), d'une distribution de probabilité conditionnelle pour laquelle on définit à chaque itération t, le stimulus $S^t_{1_N}$ et les paramètres visuo-attentionnels g^t, μ^t_A et σ^t_A. Ici, nous précisons, de plus, les valeurs attribuées à $\lambda_{L^t_{1:N}}$, pour définir de manière opérationnelle les segments en cours de traitement (cela contraste avec le cas par défaut ou elles sont toutes fixées à 1 et dans lequel le stimulus est toujours considéré dans son entièreté).

Pour illustrer ceci, reprenons le cas du stimulus « VIRDIN ». Si l'on souhaite lire ce pseudo-mot et si l'on considère deux segments (« VIR » puis « DIN »), la question probabiliste permettant le calcul, à chaque pas de temps T, des distributions de l'identité des phonèmes devient :

$$Q_\Psi^{1:2T_f} = P\left(\Psi_{1:M}^{1:2T_f} \;\middle|\; \begin{array}{c} S_{1:N}^{1:2T_f}\; G^{1:T_f}\; {\sigma_A}^{1:T_f}\; {\mu_A}^{1:T_f}\; [\lambda_{L_{1:3}}^{1:T_f} = 1]\; [\lambda_{L_{4:6}}^{1:T_f} = 0] \\ G^{T_f+1:2T_f}\; {\sigma_A}^{T_f+1:2T_f}\; {\mu_A}^{T_f+1:2T_f}\; [\lambda_{L_{1:3}}^{T_f:2:T_f} = 0]\; [\lambda_{L_{4:6}}^{T_f+1:2T_f} = 1] \end{array}\right). \quad (4.4)$$

3.2.3 Décision phonologique

Le résultat du traitement est obtenu en inférant les variables $\Psi^t_{1:M}$ pour chaque itération t (voir Equation (4.4)). La prononciation finale est décidée sur la base des courbes d'évolution des probabilités. Nous définissons un moyen pour identifier les phonèmes les plus probables tout au long du traitement et non seulement à la dernière itération. En effet dans un traitement séquentiel de plusieurs segments orthographiques, les phonèmes « gagnants » à la fin du traitement (ceux avec la plus grande probabilité) ne reflètent que la prononciation du dernier segment. Ainsi, afin de pallier ce problème, nous calculons une moyenne temporelle des probabilités des phonèmes dans chaque position. Mathématiquement,

$$\psi^T_{gagnant} = \arg\max_{\Psi_m}\left(\frac{1}{T}\sum_{t=0}^{T} Q^t_{\Psi_m}\right),$$

avec $Q^t_{\Psi_m}$ la question probabiliste de l'Equation (4.4) qui correspond à l'inférence de l'identité des phonèmes dans la position m et pour l'itération T.

3.3 Propriétés de la lecture par segments des pseudo-mots avec BRAID-Phon

Nous illustrons maintenant le fonctionnement du modèle BRAID-Phon pour la lecture de pseudo-mots par segments. Le premier exemple reprend l'exemple du pseudo-mots « VIRDIN », et montre que le mécanisme proposé corrige le comportement « spontané » du modèle (voir Section 1). Les trois exemples suivants explorent l'effet du choix des segments sur la prononciation produite.

3.3.1 Exemple 1 : Illustration du traitement par segments

Reconsidérons maintenant l'exemple « VIRDIN ». Nous devons tout d'abord déterminer le nombre de fixations, leurs paramètres visuo-attentionnels, les durées des fixations et les segments qui y sont traités. Nous choisissons d'effectuer deux fixations qui durent, chacune, 250 itérations ($N_f = 2$ et $T_f = 250$). La première fixation porte sur la première moitié du stimulus $\mu_A^{1:T_f} = 2$ et la seconde sur la seconde moitié $\mu_A^{T_f+1:2T_f} = 5$. De façon analogue, pour les segments d'accès lexical, nous choisissons de traiter le segment « VIR » ($\lambda_{L_{1:3}}^{1:T_f} = 1$ et $\lambda_{L_{4:6}}^{1:T_f} = 0$) durant la première fixation et « DIN » lors de la deuxième ($\lambda_{L_{1:3}}^{T_f+1:2T_f} = 0$ et $\lambda_{L_{4:6}}^{T_f+1:2T_f} = 1$). Les distributions de l'attention visuelle, de l'attention phonologique et d'accès lexical pour cet exemple sont présentées sur la Figure 4.16.

Le résultat de la simulation de lecture par segments du mot « VIRDIN » est présenté sur la

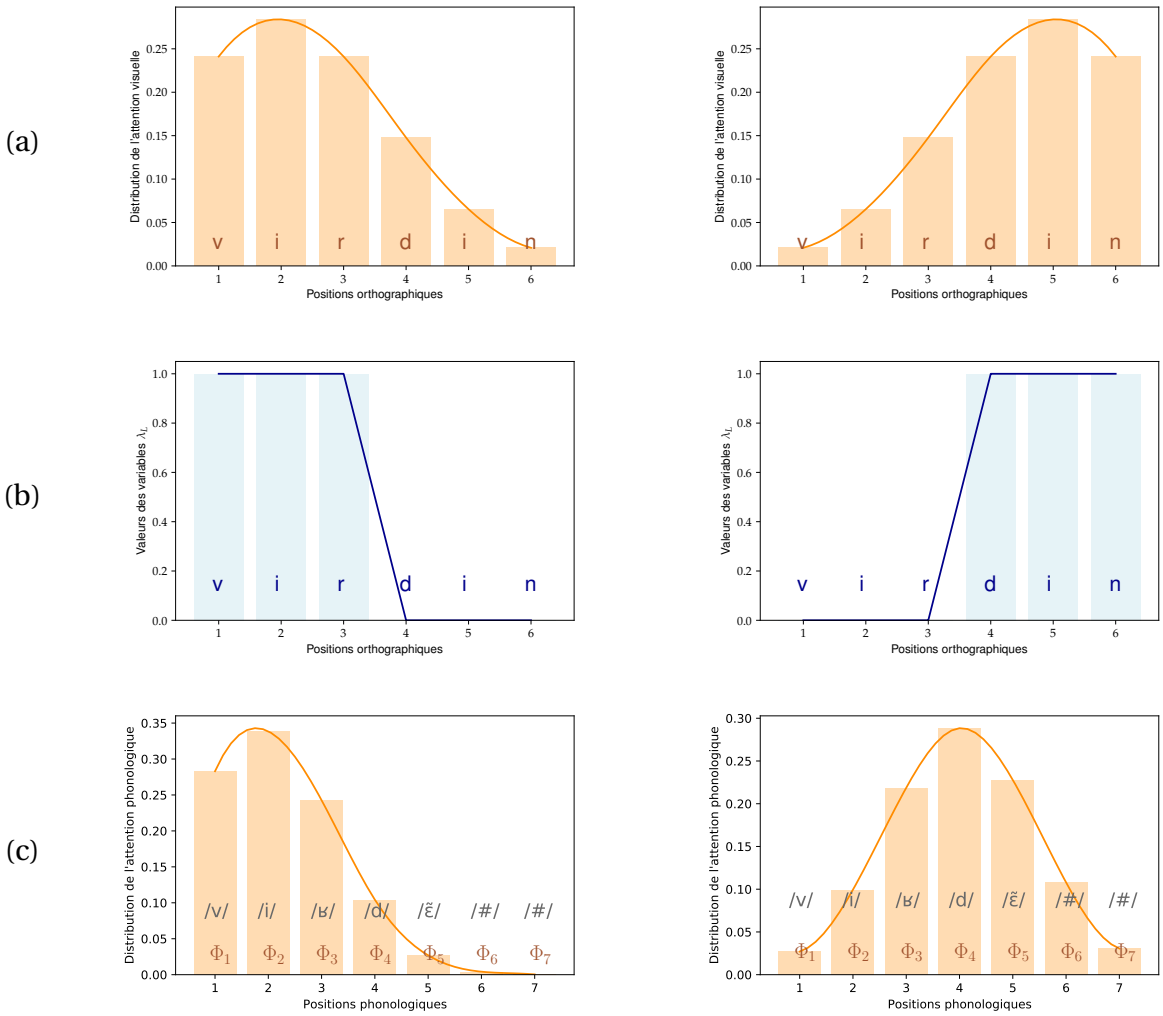


FIGURE 4.16 – **Les distribution de l'attention visuelle et phonologique et les segments considérés dans le cas de « VIRDIN ».** A gauche, sont représentés : la distribution de l'attention visuelle (a), le segment considéré (b) et la distribution de l'attention phonologique (c), supposés pour la première fixation. A droite, sont représentés leurs équivalents pour la deuxième fixation.

Figure 4.17. La prononciation produite par le modèle est /viʁdɛ̃/, ce qui correspond à la prononciation attendue, contrairement à ce que nous avons observé lors de la lecture de « VIRDIN » par le modèle BRAID-Phon sans mécanisme de lecture par segments.

Dans le comportement « spontané » du modèle, étant donné que ce sont des mots différents du lexique qui contribuent au stimulus reconnu pendant la première et la seconde fixation attentionnelle, l'information que ceux-ci fournissent est conflictuelle. Par exemple, le /ʒaʁ/ de « JARDIN » peut « écraser » le /viʁ/ reconnu lors du traitement du premier segment du stimulus « VIRDIN » (voir Figure 4.4). En revanche, pendant la lecture par segments, le recours à la segmentation visuo-attentionnelle, couplée à l'attention phonologique, permet de limiter l'« écrasement » de l'information construite lors des premières itérations et le résultat final ne reflète pas uniquement les phonèmes issus de la lecture du dernier segment. Dans l'exemple de « VIRDIN », nous observons que le phonème /a/ ressort en deuxième position mais avec une probabilité plus faible que lorsque l'on évalue le comportement « spontané » du modèle BRAID-Phon. Nous notons, par ailleurs, que le phonème /ʒ/ ressort en première position phonologique mais garde une probabilité très faible puisque la quantité d'attention portée sur cette position phonologique durant le traitement du second segment est très faible si nous la comparons avec celle portée sur la seconde position phonologique (voir Figure 4.16c).

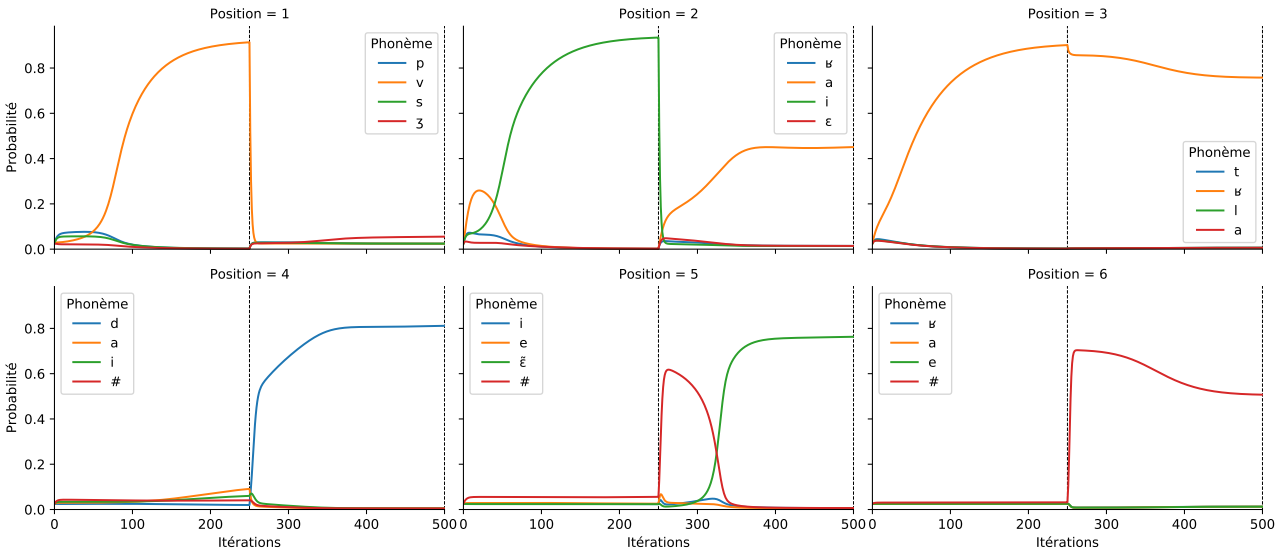
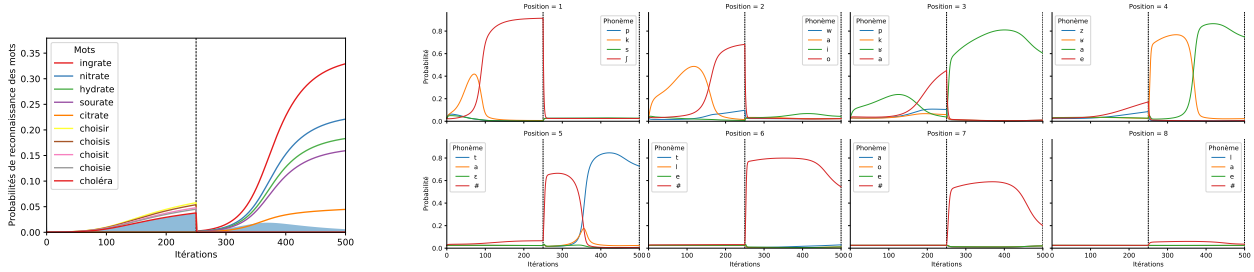


FIGURE 4.17 – **Distributions de probabilité phonologiques au cours du temps pour la lecture par segment du pseudo-mot « VIRDIN » par le modèle BRAID-Phon.** Les courbes d'évolution des phonèmes sont affichées pour chaque position phonologique. Les lignes verticales en pointillées noires indiquent la fin de chaque fixation.

3.3.2 Exemple 2 : Choix des segments

Dans ce deuxième exemple, nous simulons la lecture par segments du pseudo-mot « CHORATE » avec BRAID-Phon. Ce pseudo-mot présente une ambiguïté sur les trois premières lettres « CHO », que l'on peut prononcer /ʃo/ par analogie avec des mots comme « chômage » ou « choquer » ou /ko/ comme dans « choléra » ou « chorale ». La suite du mot ne pose pas problème puisque « RATE » se prononce toujours /ʁat/.

(a) Stimulus segmenté en « CHO » puis « RATE »



(b) Stimulus segmenté en « CHORA » puis « TE »

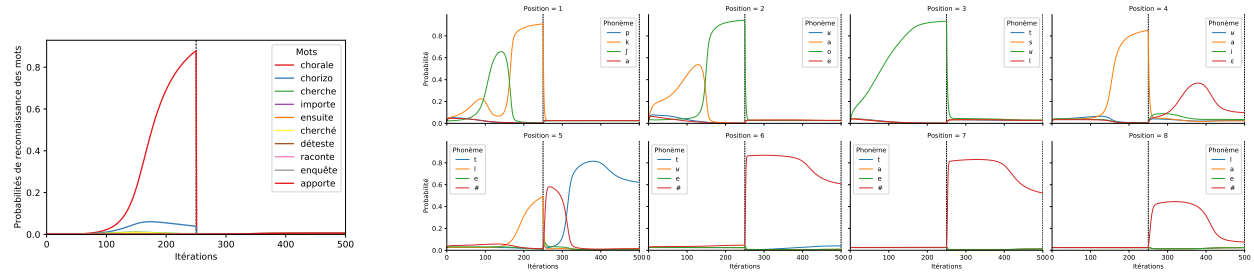


FIGURE 4.18 – **Lecture du stimulus « CHORATE » selon deux segmentation du stimulus.** le traitement est soit effectué en segmentant en « CHO » puis « RATE » (a), soit en « CHORA » puis « TE » (b). Les courbes d'évolution des probabilités de la reconnaissance des mots sont représentées à gauche, tandis qu'à droite sont représentées les courbes d'évolution des phonèmes pour chaque position phonologique. Les lignes verticales en pointillées noires indiquent la fin de chaque fixation.

Nous comparons deux manières de segmenter le stimulus. Dans les deux cas, nous choisissons un traitement en deux segments. Dans le premier cas, le premier segment est « CHO » et le second est « RATE ». Tandis que dans le second cas, le premier segment est « CHORA » et le second segment est « TE ». Ce choix de segmentation est implémenté manuellement à travers les variables λ_L d'une part et les paramètres de focus de l'attention visuelle μ_A d'autre part (voir Tableau 4.1).

Tableau 4.1 – **Variables λ_L et paramètres de focus de l'attention visuelle μ_A considérés pour la lecture du pseudo-mot « CHORATE » en deux fixations.** Deux segmentations sont présentées : « CHO » puis « RATE » dans le premier cas et « CHORA » puis « TE » dans le second cas.

	1er cas	2ème cas
1ère fixation	$\lambda_{L_{1:3}}^{1:T_f} = 1$ et $\lambda_{L_{4:7}}^{1:T_f} = 0$ $\mu_A^{1:T_f} = 2$	$\lambda_{L_{1:5}}^{1:T_f} = 1$ et $\lambda_{L_{6:7}}^{1:T_f} = 0$ $\mu_A^{1:T_f} = 2{,}5$
2ème fixation	$\lambda_{L_{1:3}}^{T_f+1:2T_f} = 0$ et $\lambda_{L_{4:7}}^{T_f+1:2T_f} = 1$ $\mu_A^{T_f+1:2T_f} = 5{,}5$	$\lambda_{L_{1:5}}^{T_f+1:2T_f} = 0$ et $\lambda_{L_{6:7}}^{T_f+1:2T_f} = 1$ $\mu_A^{T_f+1:2T_f} = 6{,}5$

Les résultats montrent que la prononciation privilégiée pour la première syllabe est /ʃo/ dans le premier cas tandis que c'est la prononciation /ko/ qui l'est dans le deuxième cas (voir Figure 4.18). Nous pouvons remarquer, également, que les mots qui contribuent à la génération des phonèmes dans le premier cas de figure sont différents de ceux qui contribuent dans le deuxième cas. Ainsi, cet exemple montre que, selon la segmentation choisie, la prononciation ne sera pas la même.

3.3.3 Exemple 3 : Segments avec ou sans chevauchement

Dans ce troisième exemple, nous simulons la lecture d'un pseudo-mot par le modèle BRAID-Phon en considérant deux autres manières de segmenter le stimulus, selon que les segments choisis se chevauchent ou non. Pour cela, nous choisissons le stimulus « VERDULIN ». Nous traitons le pseudo-mot dans les deux cas en 3 fixations, donc en 3 segments. Dans le premier cas (segments qui ne se chevauchent pas), chaque segment correspond à une syllabe (VER----- , ---DU--- puis -----LIN). Dans le second cas (segments qui se chevauchent), nous choisissons une segmentation qui ne respecte pas les frontières syllabiques, avec des segments qui recouvrent un peu plus d'une syllabe. Plus précisément, nous choisissons d'inclure dans chaque segment une syllabe et la lettre qui la précède et celle qui la suit, lorsque c'est possible. Le stimulus « VERDULIN » sera donc segmenté ainsi : VERD---- , --RDUL-- puis ----ULIN. Ces deux types de segmentation sont implémentés à travers les variables λ_L et le paramètre μ_A (voir Tableau 4.2).

Dans le cas des segments sans chevauchement (voir Figure 4.19a), parmi les mots qui contribuent à la prononciation du stimulus, on retrouve les mots « vertueux » et « vertige » lors du traitement du premier segment « VER » ; « conduire » lors du traitement du second segment « DU » ; et « orphelin » et « tremplin » lors du traitement du dernier segment « LIN ». Cependant, dans ce cas, la ressemblance phonologique est limitée et la production finale faite par le modèle est /vɛʁdl/. Dans le cas des segments qui se chevauchent (voir Figure 4.19b), on obtient la prononciation attendue (/vɛʁdylɛ̃/) et parmi les mots y contribuant on retrouve les mots « verdures », « verdâtre » et « verdicts » lors du traitement du premier segment « VERD » ; « bordures » lors du traitement du deuxième segment « RDUL » ; et « masculin » lors du traitement du dernier segment « ULIN ».

Nous observons donc que des segments plus grands qu'une syllabe permettent de reconnaître des mots plus similaires (qui partagent plus de lettres dans les mêmes positions orthographiques) et donc plus pertinents par rapport au stimulus. Les segments courts, limités aux syllabes, permettent la contribution d'un nombre important de mots dans la production des phonèmes mais la sortie peut être incorrecte. Sur cet exemple, l'utilisation de segments qui se chevauchent permet

Tableau 4.2 – **Variables λ_L et paramètres de focus de l'attention visuelle μ_A considérés pour la lecture du pseudo-mot « VERDULIN » en trois fixations.** Deux segmentations sont présentées : « VER », « DU » puis « LIN » dans le premier cas et « VERD » , « RDUL » puis « ULIN » dans le second cas.

	1er cas	2ème cas
1ère fixation	$\lambda_{L_{1:3}}^{1:T_f} = 1$ et $\lambda_{L_{4:8}}^{1:T_f} = 0$ $\mu_A^{1:T_f} = 2$	$\lambda_{L_{1:4}}^{1:T_f} = 1$ et $\lambda_{L_{5:8}}^{1:T_f} = 0$ $\mu_A^{1:T_f} = 2{,}5$
2ème fixation	$\lambda_{L_{1:3/6:8}}^{T_f+1:2T_f} = 0$ et $\lambda_{L_{4:5}}^{T_f+1:2T_f} = 1$ $\mu_A^{T_f+1:2T_f} = 4{,}5$	$\lambda_{L_{1:2/7:8}}^{T_f+1:2T_f} = 0$ et $\lambda_{L_{3:6}}^{T_f+1:2T_f} = 1$ $\mu_A^{T_f+1:2T_f} = 4{,}5$
3ème fixation	$\lambda_{L_{1:5}}^{2T_f+1:3T_f} = 0$ et $\lambda_{L_{6:8}}^{2T_f+1:3T_f} = 1$ $\mu_A^{2T_f+1:3T_f} = 7$	$\lambda_{L_{1:4}}^{2T_f+1:3T_f} = 0$ et $\lambda_{L_{5:8}}^{2T_f+1:3T_f} = 1$ $\mu_A^{2T_f+1:3T_f} = 6{,}5$

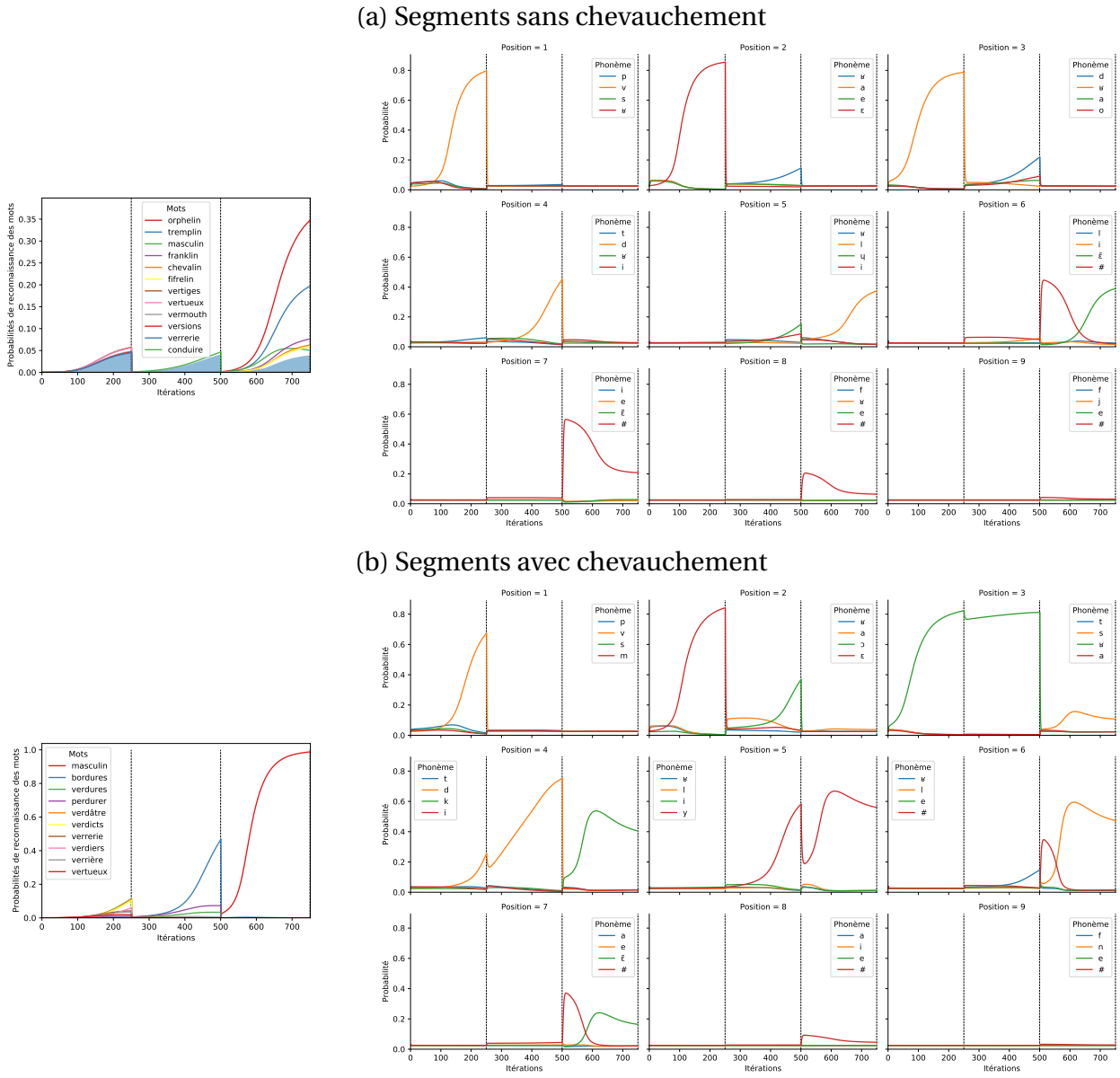


FIGURE 4.19 – **Lecture du stimulus « VERDULIN » en trois fixations.** Le traitement est soit effectué à l'aide de trois segments qui ne se chevauchent pas (a), soit à l'aide de trois segments qui se chevauchent (b). A gauche, les courbes d'évolution des probabilités de la reconnaissance des mots sont représentées et la surface bleue désigne la probabilité maximale des mots non affichés. A droite, sont représentées les courbes d'évolution des phonèmes pour chaque position phonologique. Les lignes verticales en pointillées noires indiquent la fin de chaque fixation.

donc une meilleure lecture du pseudo-mot.

3.3.4 Exemple 4 : Dépendance à la position

Dans ce dernier exemple, nous illustrons une importante limite dans la structure de connaissance du modèle qui est la dépendance à la position. Pour cela, nous considérons les pseudo-mots « MISTOUDIN » et « XAINE ». Dans le cas de « MISTOUDIN », nous traitons le stimulus en trois segments : « MIS », « STOU », « UDIN ». Alors que dans le cas de « XAINE », nous traitons le stimulus

en deux segments : « XAI », « INE ». Ces segments sont implémentés à travers les variables λ_L et le paramètre μ_A (voir Tableau 4.3).

Tableau 4.3 – **Variables λ_L et des paramètres de focus de l'attention visuelle μ_A considérés pour la lecture des pseudo-mots « MISTOUDIN » et « XAINE » en trois et deux segments.**

	« MISTOUDIN »	« XAINE »
1ère fixation	$\lambda_{L_{1:3}}^{1:T_f} = 1$ et $\lambda_{L_{4:9}}^{1:T_f} = 0$ $\mu_A^{1:T_f} = 2$	$\lambda_{L_{1:2}}^{1:T_f} = 1$ et $\lambda_{L_{3:5}}^{1:T_f} = 0$ $\mu_A^{1:T_f} = 2$
2ème fixation	$\lambda_{L_{1:2/7:8}}^{T_f+1:2T_f} = 0$ et $\lambda_{L_{3:6}}^{T_f+1:2T_f} = 1$ $\mu_A^{T_f+1:2T_f} = 4{,}5$	$\lambda_{L_{1:2}}^{T_f+1:2T_f} = 0$ et $\lambda_{L_{3:5}}^{T_f+1:2T_f} = 1$ $\mu_A^{T_f+1:2T_f} = 4$
3ème fixation	$\lambda_{L_{1:5}}^{2T_f+1:3T_f} = 0$ et $\lambda_{L_{6:9}}^{2T_f+1:3T_f} = 1$ $\mu_A^{2T_f+1:3T_f} = 7{,}5$	— —

Dans le cas de « MISTOUDIN » (voir Figure 4.20), les deux premiers segments (« MIS » et « STOU », prononcés /mistu/) sont correctement lus, avec la contribution de « misérable » et « ristourne », respectivement. Le dernier segment, quant à lui, n'est pas lu correctement et les mots qui contribuent sont « applaudit » et « applaudir ». Notons que, dans le calcul que nous effectuons ici, le lexique du modèle BRAID-Phon ne contient que les mots de la même longueur orthographique que le stimulus. Ainsi, puisqu'aucun mot de la même longueur que « MISTOUDIN », dans le lexique, ne présente « UDIN » dans les mêmes positions orthographiques, ce segment ne peut pas être lu correctement.

De la même façon, pour « XAINE » (voir Figure 4.21), le premier segment « XAI » ne peut pas être lu correctement et c'est le mot « faire », entre-autres, qui contribue. En revanche, le second segment « INE » est lu correctement et c'est le mot « peine » qui contribue. Dans ce cas, de nouveau, le premier segment ne possède aucun voisin lexical parmi les mots de la même longueur dans le lexique, ce qui rend incorrect la lecture de ce pseudo-mot. En résumé, le codage positionnel, orthographique et phonologique, utilisé dans BRAID-Phon pour l'appariement lexical est empêche de généraliser les correspondances de façon indépendante de la position. Autrement dit, S'il n'y a pas de voisins lexicaux qui partagent des segments communs dans les mêmes positions orthographiques que le pseudo-mot à lire, alors la production des phonèmes correspondants ne peut pas être correcte.

4 Conclusion

Dans ce chapitre, nous avons proposé d'évaluer la capacité du modèle BRAID-Phon, doté d'un mécanisme de lecture par segments, à lire des pseudo-mots. Rappelons que nous souhaitions répondre à trois questions :

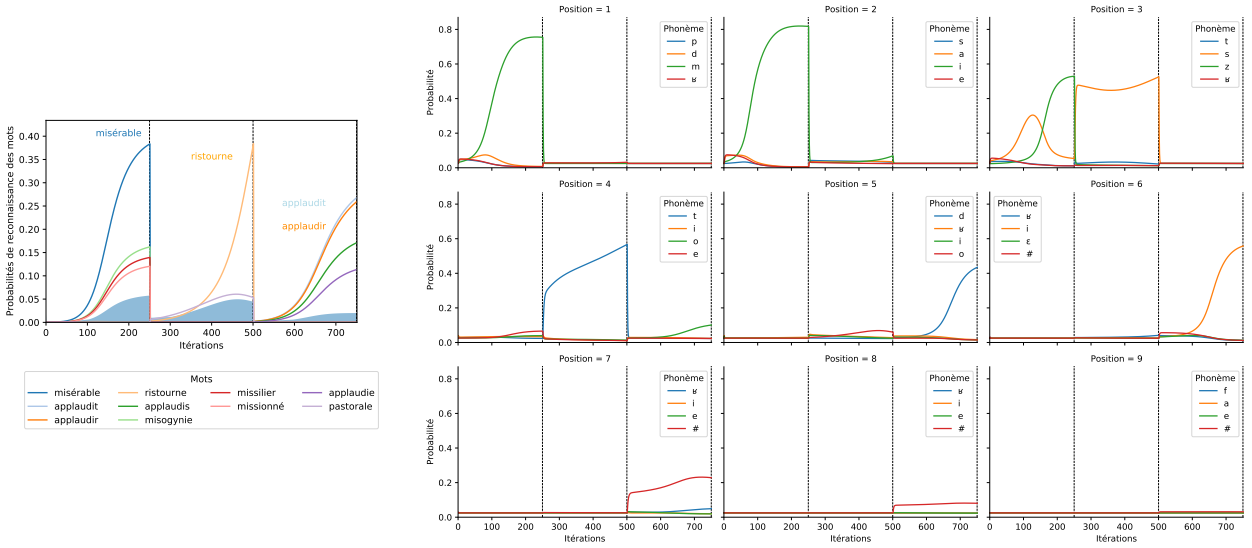


FIGURE 4.20 – **Lecture du stimulus « MISTOUDIN » en trois fixations.** A gauche, sont représentées les courbes d'évolution des probabilités de la reconnaissance des mots avec la surface bleue qui désigne la probabilité maximale des mots non affichés. A droite, sont représentées les courbes d'évolution des phonèmes pour chaque position phonologique. Les lignes verticales en pointillées noires indiquent la fin de chaque fixation.

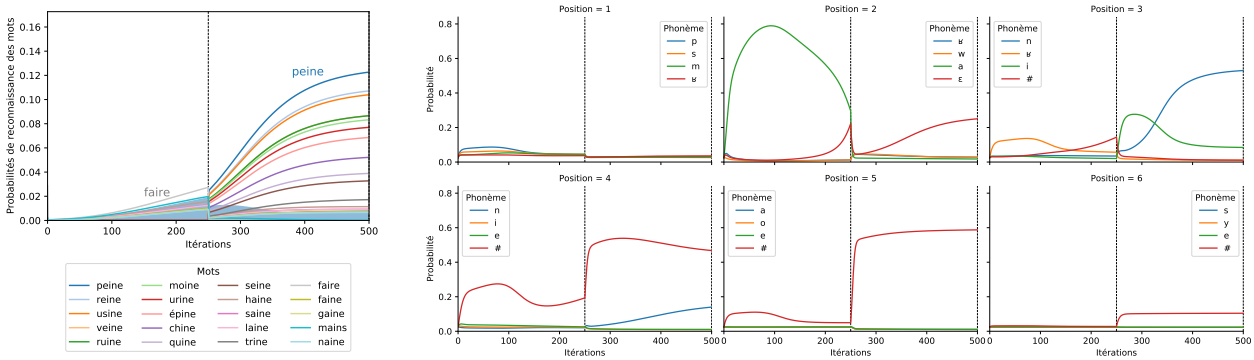


FIGURE 4.21 – **Lecture du stimulus « XAINE » en deux fixations.** A gauche, sont représentées les courbes d'évolution des probabilités de la reconnaissance des mots avec la surface bleue qui désigne la probabilité maximale des mots non affichés. A droite, sont représentées les courbes d'évolution des phonèmes pour chaque position phonologique. Les lignes verticales en pointillées noires indiquent la fin de chaque fixation.

1. Est-ce que le sous-modèle lexical de BRAID-Phon permet de formuler mathématiquement la lecture d'un segment de stimulus et, donc, de convertir une séquence de lettres en une séquence de phonèmes ?

2. Sous quelles conditions ces traitements sous-lexicaux sont-ils pertinents ?

3. Comment implémenter dans BRAID-Phon un traitement par segment qui réussisse à générer des sorties phonologiques correctes et qui ne lexicalise pas (systématiquement) lors de la lecture de pseudo-mots ?

Nous avons montré que la capacité à effectuer un traitement par segments est une condition

nécessaire à la lecture de pseudo-mots. En effet, la lecture de pseudo-mots par le modèle BRAID-Phon, sans mécanisme de lecture par segments, entraîne une lexicalisation systématique des stimuli. Lorsque le stimulus est considéré dans son entièreté, le modèle reconnaît le mot du lexique le plus proche comme correspondant au stimulus d'entrée et produit donc la prononciation de ce mot et non celle du stimulus. Traiter le stimulus en plusieurs fixations ne règle pas le problème ; cela permet d'observer l'apparition des phonèmes corrects, mais ces phonèmes ne sont plus activés lors des fixations ultérieures. En d'autres termes, le comportement « spontané » du modèle peut, au mieux, produire de manière transitoire des portions de prononciation correcte.

En revanche, si nous présentons au modèle des segments du stimulus (séquences de lettres faisant partie du stimulus), celui-ci génère, à partir des connaissances lexicales, la correspondance sous-lexicale recherchée pour le segment orthographique considéré, ce qui permet de générer les phonèmes attendus. C'est ce que nous avons pu montrer avec le modèle BRAID-Phon doté d'un mécanisme de lecture par segments, c'est-à-dire un mécanisme basé sur la définition de segments, l'ajout d'un module d'attention phonologique et sa liaison au module d'attention visuelle.

Nous avons cependant noté deux types de propriétés qui sont à prendre en compte lors « du choix des segments », et sont relatives à la taille du segment considéré et la dépendance à la position des correspondances orthographe-phonologie. Premièrement, la taille du segment affecte la sortie phonologique. En effet, un segment limité à la taille du graphème (le « O » de « BONNE » par exemple) ne donne, le plus souvent, pas suffisamment d'information sur le contexte orthographique pour permettre une sortie phonologique correcte. Mais à l'inverse, considérer un segment très long (comme « THIELD ») peut augmenter la probabilité de lexicalisation. En fait, la prise en compte d'un segment plus ou moins long, au sein d'un même pseudo-mot, modifie sa prononciation. Cela peut conduire, dans certains cas, à des prononciations dites "irrégulières", comme dans le cas de « CHORATE » où la prise en compte du segment « CHO » conduit à privilégier la prononciation /ʃɔʁat/, alors que le traitement du segment « CHORA » conduira à privilégier /kɔʁat/. Il apparaît donc que la prononciation produite en lecture est sensible au choix des segments.

Les quelques exemples que nous avons illustrés suggèrent très fortement que la taille des segments appropriée lors du traitement est variable d'un item à l'autre et qu'elle dépasse en général la taille du graphème. La prise en compte de tailles de segments variables pour un même item pourraient rendre compte de la variabilité observée chez les normo-lecteurs en lecture de pseudo-mots. Elle pourrait également expliquer les inconsistances de traitement pour le même item lors de rencontres différentes. Ainsi par exemple, la probabilité de prononciation irrégulière (comme /kɔʁat/) augmente pour les pseudo-mots voisins de mots irréguliers mais les prononciations "régulières" (comme /ʃɔʁat/) sont également observées sur ces mêmes items chez des individus différents, voire chez le même individu lors de lectures différentes. Il semble aussi vraisemblable que la rencontre avec un pseudo-mot, par exemple s'il est inattendu dans une phrase, débute par une stratégie de lecture globale, qui serait le mode par défaut pour les mots connus. Cette lecture globale serait en difficulté pour un pseudo-mot, et le système basculerait ensuite sur une lecture par segments. Le choix des segments permettant une prononciation correcte pourrait être limité par le niveau de lecture (peut-être lettre à lettre pour un lecteur débutant, par exemple), voire, il pour-

rait être obtenu par une stratégie exploratoire (le système tenterait des segments de plus en plus grands jusqu'à trouver une prononciation probable). Ainsi, dans notre modèle, comme chez les lecteurs, le choix des segments affecte la prononciation finale. Dans le modèle, nous proposons que les paramètres visuo-attentionnels soient le « mécanisme flexible » qui permette de définir des segments de taille différente selon le stimulus, et selon les propriétés de la langue.

Deuxièmement, la prononciation dépend des positions orthographiques et phonologiques où se produisent les correspondances recherchées dans le lexique. Plus spécifiquement, le modèle lit les stimuli par analogie aux mots dont il « connaît » l'orthographe et la phonologie. De plus, les positions orthographiques et phonologiques des lettres et des phonèmes des mots du lexique sont représentées de manière stricte, et la recherche de correspondance opère aussi de manière stricte, d'une part en ne considérant que les mêmes positions, et pour les mots de même longueur uniquement. Par conséquent, il est incapable de lire correctement un segment lorsque celui-ci nécessite une correspondance qu'on ne peut retrouver dans le lexique à la même position (exemple de « XAINE » et « MISTOUDIN ») dans un mot de même longueur. Nous considérons que cette propriété résulte de notre choix d'éviter un coût calculatoire élevé pour la réalisation informatique de nos simulations, et qu'elle ne résulte pas d'une hypothèse théorique. Par ailleurs, il faut noter que la « conversion » d'un segment orthographique en phonèmes peut dépendre de la position du segment dans le stimulus. Le traitement du système de lecture d'un sujet expert ne serait pas complètement indépendant de la position des segments de lettres qu'il traite mais il serait assez flexible pour prendre en compte quand il est possible, l'information du contexte et de la position.

L'inclusion du mécanisme de lecture par segments a nécessité une flexibilisation de BRAID-Phon pour ne traiter que des segments de stimulus. Ceci a été obtenu par le biais du contrôle des variables de cohérence au niveau lexical et l'introduction d'un sous-modèle attentionnel au niveau phonologique. En revanche, comme nous l'avons illustré, le choix des segments a toujours été supposé *a priori* dans chaque simulation et configuré manuellement. En effet, nous n'avons dans ce travail pas proposé de mécanisme de segmentation proprement dit, c'est-à-dire de mécanisme qui sélectionne les segments à traiter. Un tel mécanisme pourrait par exemple sélectionner une lecture lettre-à-lettre, par graphème, par syllabe, voire avec des segments syllabiques augmentés de leur contexte, voire encore en choisissant un segment unique, c'est-à-dire une lecture globale. Nous avons proposé que l'attention visuelle soit le mécanisme pour implémenter les segments ; nous n'avons pas proposé de mécanisme de segmentation qui piloterait l'attention visuelle.

Récapitulatif :

- La lecture de pseudo-mots par le modèle BRAID-Phon sans mécanisme de lecture par segments entraîne une lexicalisation systématique des stimuli, en raison du passage par l'espace lexical et la non considération d'unités sous-lexicales.
- Après ajout, au modèle BRAID-Phon, d'un mécanisme de lecture par segments et l'ajout d'un module d'attention phonologique et de sa liaison au module d'attention visuelle :
  - Les connaissances lexicales permettent de générer les correspondances entre orthographe et phonologie pour des segments de stimulus.
  - La taille des segments et leurs positions ont un impact sur la prononciation générée par le modèle.
  - Le modèle BRAID-Phon génère une lecture correcte de pseudo-mots sur plusieurs exemples.
- Le choix des segments a été fixé manuellement dans BRAID-Phon pour chaque simulation. Le développement d'un mécanisme de segmentation qui piloterait l'attention visuelle permettrait une lecture « automatique » de stimuli mot et non-mot, par l'adaptation des paramètres visuo-attentionnels en fonction des conditions de lecture.

CHAPITRE

5 Conclusion générale

"You cannot answer a question that you cannot ask, and you cannot ask a question that you have no words for."

— JUDEA PEARL (2018)

1 Principales contributions

Dans ce travail de thèse, notre objectif a été de développer un modèle de lecture à haute voix, nommé BRAID-Phon, qui se base sur un modèle sophistiqué de reconnaissance visuelle de mots, nommé BRAID (Phénix, 2018). Au travers de ce modèle BRAID-Phon, nous avons eu pour but d'évaluer l'hypothèse selon laquelle un traitement, ne mettant en jeu que les connaissances lexicales apprises sur les mots, serait en mesure de rendre compte des relations sous-lexicales entre les unités orthographiques et les unités phonologiques.

Une revue critique de la littérature dans le domaine de la lecture, effectuée dans le Chapitre 1, nous a permis de pointer les désaccords théoriques et les limites des différents modèles qui ont été proposés afin de rendre compte de la lecture à haute voix. Ceci nous a permis de formuler et de spécifier les questions de recherche auxquelles nous nous sommes intéressés dans le cadre de cette thèse. Nous avons, dans un premier temps, souligné le fait que trois grands champs théoriques se sont développés en relative indépendance pour étudier soit la reconnaissance de mots isolés, avec un intérêt particulier pour les processus visuo-orthographiques ; soit la lecture à haute voix de mots isolés, avec un intérêt particulier pour le passage de l'orthographe à la phonologie ; soit la lecture de phrases, avec un intérêt particulier pour le contrôle oculomoteur (pour des revues, voir Jacobs & Grainger, 1994 ; Norris, 2013 ; Rayner & Reichle, 2010). Dans le cadre des modèles de lecture à haute voix, nous avons notamment présenté et discuté l'opposition entre les modèles de type double-voie (Coltheart et al., 2001 ; Jacobs et al., 1998 ; Perry et al., 2007 ; Perry, Ziegler, & Zorzi, 2010) et les modèles dits à voie unique (Ans et al., 1998 ; Hendrix et al., 2019 ; Plaut

et al., 1996 ; Seidenberg & McClelland, 1989). Cette revue de question nous a amené à effectuer plusieurs constats.

Premièrement, les modèles de lecture à haute voix qui sont dominants, de nos jours, dans la littérature sont des modèles de type double-voie. Ces modèles intègrent tous un sous-modèle de reconnaissance de mots. Or, même dans les modèles de lecture les plus récents, les sous-modèles de reconnaissance de mots retenus sont fortement inspirés du modèle IA de McClelland et Rumelhart (1981). Ceux-ci ne tiennent donc pas compte des développements récents dans le champ de la reconnaissance visuelle de mots (Adelman, 2011 ; Gomez et al., 2008 ; Mozer & Behrmann, 1990 ; Norris, 2005 ; Whitney, 2001). Ainsi, les sous-modèles orthographiques des modèles double-voie obéissent à des principes qui ont été remis en question, améliorés et mieux spécifiés par les chercheurs du domaine du traitement visuel de lettres et de mots (pour une revue, voir Grainger, 2018).

Deuxièmement, alors que de très nombreuses recherches comportementales insistent sur le rôle de l'attention visuelle en lecture et que l'attention visuelle est considérée comme un composant indispensable pour expliquer le contrôle oculomoteur en lecture de phrases, les modèles de reconnaissance de mots ou de lecture à haute voix qui incluent un composant visuo-attentionnel restent l'exception (Ans et al., 1998 ; Mozer & Behrmann, 1990 ; voir aussi LaBerge & Samuels, 1974). Ces modèles défendent une architecture à voix unique, en postulant que l'attention visuelle intervient à un niveau précoce de traitement. Ainsi, la prise en compte de l'attention visuelle permet de modéliser soit un traitement global, lorsque l'attention porte sur la séquence entière du mot à lire, soit un traitement analytique, lorsqu'elle effectue une analyse sérielle des sous-unités orthographiques (Ans et al., 1998).

Troisièmement, les modèles double-voie postulent que seule la voie sous-lexicale incorporant des traitements spécifiques, de nature sérielle, et impliquant un système de conversion graphème-phonème, peut rendre compte des traitements sériels en lecture (Coltheart, 2005 ; Coltheart et al., 2001 ; Perry et al., 2007). Etant donné le caractère parallèle du traitement dans la voie lexicale, l'effet de longueur en lecture de mots n'est pas reproduit par les modèles double-voie (Kapnoula et al., 2017), sauf lorsque l'on augmente le poids de la voie sous-lexicale (Perry & Ziegler, 2002). Ainsi, selon les partisans des modèles double-voie, rendre compte d'un comportement sériel en lecture, tel que l'effet de longueur de mots, représenterait un défi pour les modèles à voie unique qui traitent les lettres de manière parallèle (Coltheart & Rastle, 1994).

Enfin, quatrièmement, les modèles de lecture double-voie font l'hypothèse que le système de conversion graphème-phonème (soit basé sur des règles, soit sur les statistiques de la langue) est indispensable à la lecture de pseudo-mots (Coltheart et al., 1993 ; Perry, Ziegler, Braun, & Zorzi, 2010). Ils font donc l'hypothèse forte que le graphème est l'unité privilégiée lors des traitements sériels sous-lexicaux. Par ailleurs, même les versions computationnelles les plus récentes de ces modèles n'implémentent pas les mécanismes qui permettent l'identification et le traitement des séquences de graphèmes au sein de la séquence de lettres du pseudo-mot à lire. Seul le modèle CDP++ l'implémente formellement et explicitement mais il se réduit à une description simpliste

de la fenêtre attentionnelle supposée impliquée dans la segmentation graphémique (voir Perry et al., 2013).

Ces différents points nous ont amené à formuler trois questions de recherche :

1. Quelles sont les limites d'une architecture à voie unique incluant un module de reconnaissance de mots sophistiqué ?
2. Quel est l'impact de l'attention visuelle en lecture de mots isolés ? Et comment permet-elle de rendre compte de l'effet de longueur en lecture de mots ?
3. Quel est le rôle de l'attention visuelle lors du traitement sériel de la séquence orthographique, notamment en contexte de lecture de pseudo-mots ?

1.1 Le modèle BRAID-Phon

Comme première contribution, nous avons décrit, dans le Chapitre 2, un modèle de lecture à haute voix qui est une extension d'un modèle récent de reconnaissance de mots, le modèle BRAID, et suit une architecture à voie unique. Nous nous situons, pour cela, dans un cadre bayésien, ce qui permet de décrire les connaissances par des distributions de probabilité et de simuler les tâches cognitives par des inférences bayésiennes.

Le modèle BRAID (Phénix, 2018) est un modèle de reconnaissance de mots qui comporte trois niveaux classiques de traitement visuel : le niveau « trait » (sous-modèle sensoriel), le niveau « lettres » (sous-modèle perceptif) et le niveau « mots » (sous-modèle lexical). Son originalité réside dans l'implémentation d'un sous-modèle d'attention visuelle qui intervient précocement entre le niveau sensoriel et le niveau perceptif et permet de moduler l'accumulation d'information sensorielle sur l'identité des lettres au cours du traitement.

Le modèle BRAID a été légèrement modifié pour les besoins de cette thèse et il a également été adapté au français. En effet, dans la version initiale du modèle, seules des lettres non-accentuées étaient représentées puisque la grande majorité des études sur la confusion ou la similarité des lettres étaient effectuées à partir de corpus de langue anglaise. Dans la thèse, une nouvelle matrice de confusion visuelle des lettres a été implémentée et calibrée, dans le sous-modèle sensoriel. Contrairement à la précédente, celle-ci prend en compte les lettres accentuées propres au français.

Le développement du modèle de lecture BRAID-Phon a impliqué deux ajouts par rapport au modèle de reconnaissance des mots BRAID. D'une part, nous avons inclus des connaissances sur la représentation phonologique des mots au sein du sous-modèle lexical ; d'autre part, nous avons implémenté un sous-modèle phonologique permettant l'accumulation d'information sur l'identité des phonèmes au cours du traitement.

La lecture des mots dans le modèle BRAID-Phon, comme les autres tâches, précédemment, dans le modèle BRAID, est définie formellement par une question probabiliste et répondre à cette question découle de l'application de règles d'inférence bayésienne. Nous avons effectué ces calculs et avons ainsi obtenu une définition mathématique, et donc une simulation informatique,

de la tâche de lecture. Nous avons observé que notre simulation de la lecture de mots génère une sortie phonologique en se basant sur le résultat de la reconnaissance de mots. En particulier, toute la dynamique de la compétition lexicale est reflétée dans la compétition au niveau phonologique. En conséquence, si le stimulus correspond à un mot connu et s'il est reconnu, alors il est prononcé correctement par le modèle.

1.2 Effet de longueur en traitement visuel des mots

Comme deuxième contribution, nous avons montré, dans le Chapitre 3, qu'un modèle de lecture à voie unique tel que BRAID-Phon permettait de simuler les effets de longueur décrits dans la littérature. Cette seconde contribution présente deux originalités : le fait d'avoir confronté le modèle à plusieurs tâches, tout en maintenant fixes les valeurs de paramètres utilisées, et le fait d'avoir simulé un large corpus de données comportementales. En effet, nous avons étudié la capacité de BRAID-Phon à simuler les effets de longueur dans trois tâches : la dénomination orale (NMG), la décision lexicale (LDT) et le *progressive demasking* (PDM). Les données comportementales dont nous disposons, pour ces trois tâches, sont issues de la méga-étude Chronolex (Ferrand et al., 2011).

Nous avons montré que BRAID-Phon permet de rendre compte des principaux effets de fréquence et de longueur de mots dans ces trois tâches (Simulations A). Dans la tâche de LDT, les effets de fréquence et de longueur sont reproduits fidèlement par le modèle ; dans les tâches de NMG et de PDM, les effets de longueur et de fréquence simulés sont de plus grande amplitude que les effets observés au niveau comportemental. Les tailles des effets (mesurés par les coefficients R^2) reproduits par le modèle BRAID-Phon sont néanmoins comparables à ceux simulés dans le cadre d'autres modèles (Adelman & Brown, 2008 ; Adelman, Marquis, Sabatos-DeVito, & Estes, 2013), pour les tâches que ceux-ci ont simulées.

Dans la mesure où l'attention visuelle est particulièrement flexible et en mesure de s'adapter aux contraintes des différentes tâches, nous avons évalué l'hypothèse selon laquelle certaines distributions visuo-attentionnelles pourraient permettre de mieux rendre compte des effets de longueur, selon la tâche considérée (Simulations B1). Les résultats des simulations montrent que la position du focus attentionnel module les performances du modèle et l'amplitude des effets de longueur. Nous avons montré que l'effet de longueur observé en décision lexicale (LDT) était le plus fidèlement reproduit lorsque le focus de l'attention visuelle est situé au centre du stimulus. Les résultats pour la dénomination orale (NMG) pourraient suggérer un avantage lorsque l'attention est placée au début du stimulus pour les items les plus longs (condition *2,5*). En *progressive demasking* (PDM), la condition avec le focus de l'attention visuelle au centre du stimulus semble rendre compte de l'effet de longueur, en termes d'amplitude, mais les temps de réaction simulés sont plus grands dans cette condition que dans les données observées. Ces résultats suggèrent que des stratégies visuo-attentionnelles différentes pourraient être adoptées selon la tâche. Néanmoins, une manipulation systématique de la position du focus attentionnel serait nécessaire pour véritablement déterminer si certaines stratégies attentionnelles sont mieux adaptées à chaque tâche.

Nous nous sommes également intéressés à l'impact des variations de la dispersion de l'attention visuelle sur les effets de longueur selon la tâche (Simulations B2). Nous avons choisi deux conditions contrastées, caractérisées soit par une distribution uniforme de l'attention sur toutes les positions soit par une distribution très piquée correspondant à une dispersion attentionnelle réduite qui pourrait correspondre à certains déficits de lecture. Nous avons montré que les effets de longueur simulés en condition uniforme sont très proches de ceux précédemment décrits avec une distribution gaussienne de l'attention (et un focus attentionnel au centre du stimulus). Par contre, une distribution piquée (associée à un parcours sériel du stimulus) modifie fortement les performances. On obtient alors un effet de longueur de plus grande amplitude en LDT et en PDM et un taux d'erreurs élevé en NMG seulement.

1.3 Lecture de pseudo-mots

Un des défis à relever pour tout modèle de lecture à voie unique est de montrer sa capacité à lire des pseudo-mots. Il s'agit de la troisième et dernière contribution dans cette thèse. Il est communément admis que la lecture de pseudo-mots repose sur un traitement sériel par sous-unités. Dans la thèse, nous avons utilisé le terme de « segment » pour parler de ces sous-unités, afin de ne pas référer à une unité spécifique, ni même à une séquence de lettres correspondant à une des unités orthographiques classiquement identifiées.

Dans le Chapitre 4, nous avons donc procédé par étapes afin d'étudier la capacité du modèle à lire des pseudo-mots. Nous avons décrit, dans un premier temps, les limites du modèle BRAID-Phon, initialement dépourvu de capacités de lecture par segments. Ce modèle (tel qu'il est décrit dans le Chapitre 2) entraîne une lexicalisation systématique des pseudo-mots présentés. Lorsque le traitement porte sur l'ensemble des lettres du pseudo-mot (traitement en une seule fixation), le modèle reconnaît le mot du lexique le plus proche orthographiquement et c'est la phonologie correspondant à ce mot qui est produite à la place du pseudo-mot. Ce problème persiste lorsque le pseudo-mot est traité séquentiellement (en plusieurs fixations). On observe alors que la prononciation correspondant à chacune des fixations est correcte, mais la séquence phonologique à l'itération finale dépend uniquement des traitements visuo-orthographiques effectués lors de la dernière fixation. C'est alors la phonologie correspondant au mot qui partage le plus de similarités avec les informations orthographiques accumulées au cours de la dernière fixation sur le pseudo-mot qui est produite. Ceci témoigne du fait que, sans mécanisme de lecture par segments, le modèle BRAID-Phon est capable, au mieux, de produire des portions de prononciations correctes mais seulement de manière transitoire au cours du traitement.

Dans un second temps, nous avons doté le modèle BRAID-Phon d'un mécanisme de lecture par segments. Ceci a conduit à implémenter un module d'attention phonologique qui est couplé au module d'attention visuelle. Les simulations portant sur la lecture de pseudo-mots par cette version finale du modèle BRAID-Phon montrent que lorsque des segments de stimulus sont présentés en entrée, le modèle génère la correspondance sous-lexicale recherchée pour chacun des segments orthographiques considérés, conduisant à produire une séquence phonologique plausible lors de la lecture des pseudo-mots. Plus spécifiquement, l'utilisation du mécanisme de sé-

lection des segments et du sous-modèle attentionnel phonologique permet d'accumuler les bons phonèmes dans les bonnes positions tout en empêchant que ceux-ci soient « écrasés » par d'autres phonèmes qui ne sont pas pertinents. Ceci permet d'éviter ce que nous observions préalablement avec le modèle dépourvu de mécanisme de lecture par segments : les trois premiers phonèmes /viʁ/ de « VIRDIN » étaient « écrasés » à la fin de la deuxième fixation par /ʒaʁ/ de « jardin » (voir Figure 4.4 et Figure 4.17 du Chapitre 4). Cependant, nous avons noté certaines propriétés de la lecture par segments qui sont particulièrement intéressantes.

Premièrement, la taille des segments influence la sortie phonologique obtenue. En effet, un segment trop court ne donne, en général, pas suffisamment d'informations orthographiques contextuelles et un segment trop long augmente la probabilité de lexicalisation, ce qui induit une prononciation incorrecte du stimulus dans les deux cas. Les simulations effectuées suggèrent que la taille des segments permettant de générer une sortie phonologique plausible varie selon les pseudo-mots présentés et que la prise en compte de segments de tailles différentes peut conduire à des prononciations différentes et plausibles pour un même pseudo-mot. Par ailleurs, la sortie phonologique produite est également influencée par la position des correspondances orthographe-phonologie dans la séquence du pseudo-mot. La prononciation des segments au sein d'un pseudo-mot dépend de la prononciation de ces mêmes segments dans les mots du lexique, à position identique. Les simulations montrent qu'un même segment peut être plus souvent associé à une prononciation en début de mot et à une autre en milieu ou fin de mot. Ceci a pour conséquence de modifier la prononciation proposée par le modèle selon la position du segment dans le pseudo-mot. Une conséquence délétère de ce même principe est que le modèle est incapable de générer la prononciation correcte de segments qu'on ne peut trouver dans la même position dans les mots de même longueur dans le lexique.

2 Discussion générale

Dans cette section, nous discutons de notre contribution, des nouveaux questionnements et des réponses qu'elle fournit aux questions théoriques que nous avons posées suite à notre revue de littérature. Pour rappel, la première question traite de l'architecture du système de lecture, la deuxième question traite du rôle de l'attention visuelle dans la lecture de mots, et notamment de son rôle vis-à-vis des effets de longueur, et la troisième question traite du rôle de l'attention visuelle dans le traitement des pseudo-mots.

Dans un premier temps, nous aborderons la question de l'architecture du système de lecture et nous situerons BRAID-Phon dans ce cadre, ainsi que son apport dans la comparaison entre ces deux types d'architecture. Dans un second temps, nous aborderons la question de l'attention visuelle en détaillant son rôle dans la lecture, son implication dans la lecture par segments et son lien avec l'effet de longueur pour les mots. Nous nous intéresserons également au module d'attention phonologique, introduit pour la lecture par segments, et à son lien supposé avec l'attention visuelle. Enfin, nous discuterons de ce que nous apporte la considération de segments de taille variable.

2.1 Architecture en une seule voie plutôt qu'en deux voies

Un modèle plus intégratif

Les modèles double-voie sont les plus dominants dans la littérature, de nos jours. Ceci est dû, d'une part, au fait que ces modèles se sont développés les uns après les autres et de manière « cumulative » (c'est-à-dire, l'un étant une extension du précédent). D'autre part, ils permettent de rendre compte d'un grand nombre d'effets comportementaux en lecture. Il faut également noter que l'ancienneté de ces modèles est également un élément déterminant. En effet, du fait de leur ancienneté ils ont pu être confrontés à un grand nombre d'aspects de la lecture (experte, développementale et pathologique), ce qui est intéressant pour la communauté des chercheurs travaillant, de près ou de loin, sur ce sujet. Cependant, les partisans de l'approche à voie unique considèrent que plusieurs critiques peuvent être faites concernant les modèles double-voie.

Tout d'abord, les modèles double-voie, en dépit de leur caractère « cumulatif » et de leur volonté de rendre compte du plus grand nombre d'effets possibles, se focalisent sur les traitements phonologiques et adoptent des niveaux de modélisation simplistes des traitements visuo-orthographiques. Ainsi, ils ne détaillent pas suffisamment les processus visuels et restent sur des niveaux de modélisation simplistes de ces mécanismes (Norris, 2013). Plus spécifiquement, les modèles computationnels de lecture ayant une architecture double-voie ne rendent pas compte de certains effets observés en reconnaissance de mots. A titre d'exemple, l'effet de longueur en reconnaissance de mots n'a pas été traité dans le cadre du modèle IA (McClelland & Rumelhart, 1981), sur lequel reposent les modèles double-voie. Cependant cet effet a été étudié dans le cadre de modèles plus récents de reconnaissance de mots comme le modèle SERIOL (Whitney, 2001). Nous pouvons également mentionner d'autres effets « visuels » tels que les effets observés lors d'expériences comportementales se servant du paradigme d'amorçage (ou *priming*) (Ginestet, 2016), les effets dus au codage positionnel incertain des lettres (Gomez et al., 2008), l'effet de la position optimale du regard (*Optimal viewing Position*, OVP) (Phénix, 2018 ; Phénix et al., 2020) et les effets de la fréquence du voisinage (Phénix, Valdois, & Diard, 2018). Une autre critique que nous pouvons formuler concerne les mécanismes attentionnels. Ceux-ci sont assez peu décrits et rarement inclus de façon formelle dans la plupart des modèles de lecture, bien que certains auteurs ne rejettent pas l'implication de l'attention à certains niveaux de traitement (Perry, Ziegler, & Zorzi, 2010 ; Perry et al., 2013).

Dans cette thèse, nous avons essayé de montrer comment un modèle à voie unique, BRAID-Phon, permettait de proposer un début de réponse à certaines des limitations présentes dans les modèles double-voie. Grâce à l'inclusion d'un modèle détaillé de reconnaissance visuelle des lettres et des mots et d'un module attentionnel, BRAID-Phon présente un caractère intégratif auquel ses prédécesseurs ne prétendent pas. Le modèle BRAID-Phon intègre un modèle de reconnaissance de mots digne des développements les plus récents dans ce domaine. Ceci lui permet de rendre compte d'effets comportementaux en décision lexicale (Ginestet et al., 2019 ; Phénix, 2018). De plus, parce qu'il inclut un composant d'attention visuelle, BRAID-Phon permet de briser la dichotomie entre les modèles de reconnaissance de mots et de lecture d'une part, et les modèles de contrôle oculomoteur d'autre part (Engbert et al., 2005 ; Reichle et al., 2003 ; Snell et al., 2018 ;

voir aussi Veldre et al., 2020). En effet, dans des travaux menés en parallèle de ceux présentés ici, Ginestet (2019) a développé le BRAID-learn (une autre extension de BRAID), concernant l'étude de l'apprentissage orthographique, et a proposé un système de contrôle des paramètres visuo-attentionnels basé sur l'optimisation de l'accumulation de l'information orthographique perçue. Plus spécifiquement, un critère basé sur le gain d'entropie des distributions de probabilité sur les lettres, au cours du traitement, permet de « choisir » les paramètres visuo-attentionnels permettant d'accumuler rapidement de l'information sur l'identité des lettres, et ainsi, de construire efficacement une nouvelle trace orthographique pendant l'apprentissage. En couplant ce mécanisme et l'adaptant au modèle BRAID-Phon, ce dernier devrait pouvoir opérer une ou plusieurs fixations pendant la lecture des mots ou des pseudo-mots et donc rendre compte du comportement oculomoteur en lecture de mots ou de pseudo-mots.

Un modèle plus parcimonieux

L'architecture à voie unique du modèle BRAID-Phon, à notre avis, offre un cadre plus parcimonieux pour décrire le système cogntifs de lecture et les connaissances qu'il contient. Premièrement, d'un point de vue purement computationnel, les deux voies de traitement que les modèles double-voie supposent semblent contenir des informations « redondantes ». Autrement dit, la voie lexicale encode les connaissances des représentations orthographiques et phonologiques des mots du lexique ; tandis que la voie sous-lexicale encode les connaissances des correspondances sous-lexicales (graphèmes-phonèmes), qui sont techniquement apprises ou extraites des mots du lexique. Ces connaissances sous-lexicales encodent donc les correspondances entre unités sous-lexicales en référence aux propriétés du lexique considéré. Ainsi, si la branche lexicale du modèle contient déjà l'information des correspondances sous-lexicales, il semble, à première vue, non nécessaire de les « stocker ailleurs », en l'occurrence dans une autre branche du modèle qui fonctionne selon des principes computationnels différents. Toutefois, il faut noter que les chercheurs, y compris les partisans des modèles à voie unique, ne contestent pas l'existence des redondances dans le système cognitif. La redondance de l'information peut même être un mécanisme important dans les traitements cognitifs ; les chercheurs parlent parfois d'encodage de l'information dans de multiples niveaux de représentations (Conrad, Carreiras, Tamm, & Jacobs, 2009 ; Kim, Taft, & Davis, 2004 ; Taft, 1991). Au niveau algorithmique de Marr (1982) où se situe notre analyse, l'architecture double-voie ne parait donc pas nécessaire. En revanche, l'architecture double-voie n'est pas incompatible avec une implémentation physique par deux voies totalement ou partiellement redondantes. Ainsi, des modèles du niveau implémentation de Marr (1982) pourraient avoir une architecture double-voie, légitimée par une correspondance des voies avec des corrélats neurophysiologiques. A notre connaissance, malheureusement, le débat entre les architectures ne repose pas sur des arguments de cette nature, et se contente essentiellement à favoriser la classe de modèles la mieux établie, et qui rend compte du plus grand nombre d'effets.

Deuxièmement cette redondance peut paraître excessive lorsque l'on place l'approche double-voie dans un cadre théorique plus large qui est celui des théories de représentations duales (*Dual Theory of representation*). Pour rendre compte à la fois de la lecture et de la production écrite de mots (orthographe lexicale ou *spelling* en anglais), ce cadre théorique postule une distinction

entre les connaissances lexicales mises en jeu en lecture des connaissances lexicales mises en jeu en dictée (Baxter & Warrington, 1985 ; Patterson, 1986 ; Patterson & Shewell, 1987 ; voir aussi Figure 5 de l'article de Coltheart et al., 2001). Cette approche est critiquable sous l'angle d'un autre cadre théorique (théories du lexique commun ou des représentations partagées) et d'autres observations qui suggèrent qu'un seul lexique et une seule base de connaissances lexicales serait plus parcimonieux et plus en accord avec les données (Allport & Funnell, 1981 ; Angelelli, Marinelli, & Zoccolotti, 2010 ; Behrmann & Bub, 1992 ; Houghton, 2018 ; Rapp & Lipka, 2011 ; voir aussi Coltheart & Funnell, 1987 ; Hillis & Rapp, 2004, pour des comparaisons entre les deux approches théoriques).

Nous avons montré avec le modèle BRAID-Phon et à l'aide du formalisme bayésien, qu'avec une structure de connaissance purement lexicale, il est possible d'inférer « à la volée » les correspondances sous-lexicales, donc sans supposer qu'elles sont « stockées indépendamment » et, surtout, sans faire l'hypothèse qu'elles sont manipulées avec d'autres mécanismes computationnels. Dans le cadre de BRAID-Phon, des simulations préliminaires effectuées en marge de la thèse permettent de penser que BRAID-Phon pourrait simuler à la fois la lecture, où l'on convertit l'orthographe en phonologie, et l'opération inverse (l'écriture ou la dictée) qui consiste à convertir la phonologie en orthographe (Saghiran, Valdois, & Diard, 2019). Dans ce travail de thèse sur le modèle BRAID-Phon, nous avons étudié le mécanisme de lecture à haute voix, en orientant donc nos inférences pour aller du stimulus visuel vers les représentations phonologiques. Les propriétés du calcul bayésien permettent d'envisager la transformation inverse de manière symétrique. En effet, on peut voir le théorème de Bayes comme une opération permettant de calculer en « inversant les connaissances », puisqu'on calcule $P(A \mid B)$ en fonction de $P(B \mid A)$. Ainsi, le modèle BRAID-Phon peut, tout aussi bien, être utilisé pour calculer les correspondances des phonèmes vers les graphèmes, c'est-à-dire pour inférer les lettres à partir des phonèmes. Si l'on applique BRAID-Phon au domaine de l'orthographe, cette propriété le mettrait dans le « camp » des théories qui postulent un lexique commun pour la lecture et la dictée.

2.2 Une attention visuelle flexible plutôt qu'une deuxième voie sérielle

Dans la littérature, la plupart des modèles théoriques de lecture ou de reconnaissance de mots n'incluent pas de composants attentionnels. Ce désintérêt des modèles pour l'attention visuelle contraste avec les résultats des études comportementales ou des études de neuroimagerie qui suggèrent que l'attention visuelle est impliquée tant dans la reconnaissance des mots connus, même familiers (Besner et al., 2016 ; Lachter, Forster, & Ruthruff, 2004 ; Risko et al., 2010 ; Waechter, Besner, & Stolz, 2011) que dans la lecture des pseudo-mots (Facoetti et al., 2006 ; Franceschini et al., 2020 ; Franceschini, Gori, Ruffino, Pedrolli, & Facoetti, 2012 ; Valdois et al., 2006 ; Valdois, Roulin, & Bosse, 2019). Les modèles double-voie considèrent, au mieux, que l'attention visuelle pourrait être un mécanisme spécifique à la voie sous-lexicale permettant de segmenter les unités graphémiques une fois la séquence de lettres du mot clairement identifiée (Perry, Ziegler, & Zorzi, 2010 ; Perry et al., 2013). Selon ces modèles, l'intervention de l'attention est donc propre à la voie sous-lexicale et relativement tardive.

Le rôle de l'attention visuelle en lecture

Contrairement à ces modèles, BRAID-Phon suppose que l'attention visuelle intervient très précocement dans le traitement de l'entrée orthographique, puisque le module attentionnel se situe entre le niveau sensoriel et le niveau perceptif et affecte la façon dont les lettres sont identifiées. Ceci est conforme à la façon dont l'attention visuelle a été implémentée dans le modèle MORSEL de reconnaissance de mots (Mozer & Behrmann, 1990) et dans le modèle MTM de lecture (Ans et al., 1998). Une intervention précoce de l'attention pour l'identification des lettres du mot est également postulée par les modèles de contrôle oculomoteur (voir Leinenger & Rayner, 2015, pour une revue). Cependant, dans la plupart de ces modèles, l'attention visuelle est uniformément répartie sur toutes les lettres qui sont sous le focus attentionnel. BRAID-Phon fait un choix différent puisque la distribution de l'attention est de forme gaussienne. Ainsi, il se rapproche de OB1-Reader, un modèle récent de contrôle oculomoteur en lecture de phrases (Meeter, Marzouki, Avramiea, Snell, & Grainger, 2020 ; Snell et al., 2018). Nous notons, par ailleurs, certaines différences quant à la conception et la formulation mathématique de l'attention visuelle entre BRAID-Phon et OB1-Reader. En effet, le modèle OB1-Reader suppose une attention largement distribuée (qui capture plusieurs mots dans une séquence de mots), et dont la dispersion n'est pas flexible. Ceci contraste avec la distribution d'attention de BRAID-Phon qui, elle, est sur l'échelle d'un seul mot et dont la dispersion peut varier (*via* la manipulation du paramètre σ_A).

Une autre originalité du modèle BRAID-Phon est de postuler que l'attention visuelle est impliquée, non seulement dans les traitements sériels en lecture, et donc dans la lecture de pseudo-mots, mais également pour expliquer l'effet de longueur observé en lecture de mots. En effet, la distribution visuo-attentionnelle permet de déterminer la nature du traitement. Le traitement est parallèle lorsque l'attention est répartie sur toutes les lettres du stimulus ; tandis que le traitement est sériel ou séquentiel[16] lorsque l'attention permet de capturer, de manière séquentielle, des portions du stimulus. Cette propriété de l'attention visuelle en lecture a également été utilisée par d'autres modèles incluant un composant attentionnel (Ans et al., 1998 ; LaBerge & Samuels, 1974 ; Mozer & Behrmann, 1990). Ceci suggère que, dès lors que l'on introduit un composant attentionnel dans les modèles, on peut expliquer que le traitement soit parallèle lors de la lecture de mots connus et sériel lors de la lecture de pseudo-mots ou de mots nouveaux, sans recourir à l'hypothèse de deux voies de traitement. Ainsi, les cadres théoriques et les modèles cités précédemment, grâce à leur représentation de l'attention visuelle et sa manipulation pour adapter le traitement aux caractéristiques du stimulus, remettent en question l'hypothèse centrale des modèles double voie.

L'attention visuelle et l'effet de longueur

L'effet de longueur a été habituellement interprété comme une manifestation des traitements sériels en lecture et a été observé dans plusieurs langues (italien, français, grec et allemand) lors de la lecture de mots (Juphard et al., 2004 ; Kapnoula et al., 2017 ; Perry & Ziegler, 2002 ; Spinelli et

[16] Nous notons que nous préférons utiliser ici le terme « séquentiel » plutôt que le terme « sériel » pour mettre l'accent sur le fait qu'il s'agisse d'un traitement par séquence et pas forcément d'un traitement lettre à lettre. La séquence en question peut être une séquence de segments orthographiques.

al., 2005). Il a également été observé dans la tâche de décision lexicale dans différentes langues (anglais, français et néerlandais) (Balota et al., 2007 ; Brysbaert, Stevens, Mandera, & Keuleers, 2016 ; Ferrand et al., 2010 ; New et al., 2006) et dans la tâche de *progressive demasking* (Ferrand et al., 2011). Il faut noter que dans le cadre des modèles double-voie, l'effet de longueur n'est induit que par le traitement sériel de la voie sous-lexicale. Une des limitations du cadre théorique des modèles double-voie est donc le fait qu'elle permet d'expliquer l'interaction lexicalité-longueur dans la lecture et non de l'effet de longueur dans les mots. Cela pose un problème étant donné le caractère parallèle du traitement des mots et notamment pour les mots courts.

Dans le cadre de notre modèle, cet effet ne relève pas d'un traitement sériel puisqu'il intervient, dans BRAID et BRAID-Phon, même lors d'un traitement parallèle. En effet, les ressources attentionnelles sont limitées et sont distribuées (distribution gaussienne) sur les lettres quelle que soit la longueur du stimulus. Ainsi les mots courts sont plus rapidement traités que les stimuli longs, étant donné que plus d'attention est attribuée sur chaque lettre et que l'ensemble des lettres sont capturées quand le stimulus est court. Néanmoins, nous montrons également qu'une dispersion de l'attention réduite, avec un traitement sériel, augmente l'effet de longueur en décision lexicale. De plus, lorsque l'attention visuelle est considérée comme étant impliquée dans la lecture, l'interprétation de l'effet de longueur change. Dans le cadre de simulations avec le modèle BRAID, Ginestet et al. (2019) ont montré qu'un traitement séquentiel (en deux captures) permettait de diminuer l'effet de longueur observé dans les temps de réponse simulées. Ainsi, de manière contre-intuitive, la séquentialité du traitement ne ralentit pas forcément le traitement ; dans certains cas, elle peut même l'accélérer, par exemple lorsque l'attention visuelle se déplace de façon à aller chercher efficacement l'information sur les portions de stimulus encore peu traitées.

L'effet de longueur nous semble donc être un point important permettant l'évaluation de l'implication de l'attention visuelle dans le traitement des mots. Whitney (2018) mentionne même un effet de longueur facilitateur en décision lexicale, lorsqu'elle simule une configuration spécifique du traitement sériel de stimulus (voir aussi Whitney, 2011). Ainsi, ces résultats montrent que la présence ou l'absence de l'effet de longueur n'est pas une « preuve indubitable » du caractère sériel ou parallèle du traitement.

En lecture de pseudo-mots, nous postulons un traitement séquentiel par segments qui est piloté par l'attention. En effet, les mécanismes attentionnels de BRAID-Phon permettraient de sélectionner, de manière séquentielle, des segments orthographiques du stimulus qui pourraient être « convertis », par analogie aux mots du lexique, en segments phonologiques. Par ailleurs, la position que nous prenons ici avec le modèle BRAID se distingue des affirmations d'autres cadres théoriques. En effet, il existe deux cas de figure. Soit les auteurs n'envisagent pas de traitement sériel (Seidenberg & McClelland, 1989) et considèrent qu'il s'agit d'une limite majeure, quitte à proposer, par la suite, une variante qui effectue un traitement sériel en recourant à une fenêtre attentionnelle (Plaut, 1999). Soit les auteurs revendiquent que les traitements sériels sont spécifiques à la voie sous-lexicale, qui est la seule voie à faire intervenir l'attention visuelle (Perry et al., 2013). Dans les deux cas, l'attention visuelle est modélisée de façon minimaliste et se réduit à une fenêtre de traitement (tout-ou-rien).

L'attention visuelle et l'attention phonologique

Nous avons introduit, dans le modèle BRAID-Phon, un sous-modèle d'attention phonologique impliqué dans la lecture par segments. Ce module d'attention phonologique est couplé au module d'attention visuelle de façon à obtenir, pour chaque segment orthographique, une sortie phonologique plausible. Plus spécifiquement, nous avons fait l'hypothèse que l'attention visuelle avait un rôle prépondérant étant donné que les paramètres de la distribution attentionnelle phonologique sont asservis à ceux de la distribution de l'attention visuelle.

L'ajout d'un sous-modèle d'attention phonologique représente une forte contribution dans cette thèse. En effet ce sous-modèle affecte largement le fonctionnement du modèle puisque, de manière analogue à l'attention visuelle qui permet de sélectionner des segments orthographiques, l'attention phonologique permet de sélectionner des segments d'unités phonologiques.

De la même façon que l'attention visuelle n'est, en général, pas présente dans les modèles de lecture ou de reconnaissance de mots, aucune représentation de l'attention phonologique n'a été proposée, jusqu'ici, dans les modèles de lecture. Ainsi, nous nous interrogeons sur la légitimité des choix que nous avons faits et deux questions se posent. (1) Est-il pertinent de proposer un mécanisme d'attention phonologique ? (2) Est-il raisonnable de supposer un couplage entre attention visuelle et attention phonologique ?

Concernant le mécanisme d'attention phonologique, celui-ci permet d'identifier des segments phonologiques au sein de la représentation phonologique du mot. Ce mécanisme pourrait donc intervenir dans les tâches métaphonologiques qui demandent d'identifier et de manipuler des unités phonologiques, comme la syllabe, la rime ou le phonème, au sein des mots produits à haute voix. En accord avec cette hypothèse, plusieurs études suggèrent que l'attention sélective est impliquée dans les épreuves de conscience phonologique (Pattamadilok, Perre, & Ziegler, 2011 ; Yoncheva, Maurer, Zevin, & McCandliss, 2013 ; Yoncheva, Zevin, Maurer, & McCandliss, 2009). L'implication de l'attention et son lien avec les traitements phonologiques est également compatible avec les théories de la mémoire où on envisage la mémoire verbale à court terme comme la partie de la mémoire à long-terme contrôlée par l'attention (Engle, Kane, & Tuholski, 1999). Dans le modèle de mémoire à court terme de Baddeley (Baddeley, 1996, 2003), l'attention représentée par le *central executive*, appelé également système de contrôle attentionnel (ou *attentional control system*), joue un rôle crucial et agit sur le système de stockage et de manipulation de l'information phonologique (boucle phonologique ou *phonological loop*).

Par ailleurs, le système de contrôle attentionnel dans les modèles de mémoire intervient non seulement dans le traitement des informations phonologiques mais également dans celui de l'information visuo-spatiale (calepin visuospatial ou *visuospatial sketchpad*). Notre postulat d'un couplage entre attention visuelle et attention phonologique est également compatible avec les données qui témoignent d'un impact de l'orthographe sur la segmentation phonologique (Pattamadilok et al., 2011). Nous avons fait ici l'hypothèse que l'attention visuelle piloterait l'attention phonologique, suggérant que l'identification des segments au niveau orthographique modulerait la façon dont le système identifie les segments au niveau phonologique. Ceci rejoint une

hypothèse formulée par Vidyasagar et Pammer (2010) selon laquelle des problèmes de segmentation graphémique pourraient entraîner des problèmes de segmentation phonémique chez les personnes dyslexiques. Notre postulat selon lequel l'attention visuelle jouerait un rôle à la fois dans la segmentation orthographique et dans la segmentation phonologique est également intéressant par rapport aux théories de l'apprentissage de la lecture. Il est, en effet, largement admis qu'il existe un lien entre conscience phonémique et apprentissage de la lecture (pour une revue, voir Melby-Lervåg, Lyster, & Hulme, 2012) mais ce lien n'est pas considéré comme causal (pour une revue, voir Castles & Coltheart, 2004). Au contraire, c'est la confrontation à l'écrit et l'apprentissage explicite de la lecture qui permettraient le développement de la conscience des phonèmes. Ceci suggère que l'identification des unités orthographiques pourrait bien être le déclencheur de l'identification des unités correspondantes au niveau phonologique. Les données des études de rééducation vont dans le même sens. Il a été largement démontré que l'entraînement à la conscience phonémique n'est pas efficace lorsqu'il repose uniquement sur la forme phonologique des mots. Par contre, cet entraînement est efficace lorsqu'il propose sur la segmentation simultanée de la séquence écrite et de la séquence orale des items entraînés (pour une revue, voir Ehri et al., 2001).

Ainsi, BRAID-Phon est le premier modèle de lecture qui implémente des mécanismes directement impliqués dans les épreuves métaphonologiques. Le sous-modèle d'attention phonologique n'a pas été introduit pour rendre compte des performances métaphonologiques et son lien avec le sous-modèle d'attention visuelle se justifie indépendamment pour répondre à la question de la segmentation en unités sous-lexicales. Ce faisant, le modèle BRAID-Phon offre un cadre théorique nouveau qui explicite de façon claire la nature du lien possible entre lecture et conscience phonologique. En implémentant des mécanismes attentionnels dynamiques aux niveaux phonologique et visuel, le modèle propose également un module de mémoire parfaitement intégré aux processus de lecture, sans avoir à postuler de composants spécifiques tel que les *buffer* graphémique ou phonologique des modèles double-voie.

2.3 Des segments orthographiques plutôt que des graphèmes

Pour réaliser la lecture de pseudo-mots, nous avons été confrontés à deux problèmes dans le cadre du modèle BRAID-Phon : celui de la taille des segments et celui de la sensibilité de la conversion du segment à sa position dans le stimulus. Ces problèmes ne sont pas spécifiques à nos postulats théoriques mais se posent également dans le cadre des modèles double-voie.

Dans la littérature, et notamment au sein des modèles double-voie, il existe un accord quant à la pertinence du graphème, étant donné que cette unité est nécessaire pour procéder à des conversions graphème-phonème. Pour autant, cette unité sous-lexicale ne semble pas fournir une réponse satisfaisante aux deux problèmes précédents. En effet, les règles graphèmes-phonèmes qui ne dépendent pas du contexte sont considérées non suffisantes pour la lecture de mots chez des lecteurs débutants (Schmalz, Robidoux, Castles, & Marinus, 2020 ; Steacy et al., 2018). En effet, la nécessité de considérer des segments de petite et de grande taille, notamment dans des langues opaques telles que l'anglais, est probablement une conséquence inévitable des correspondances entre l'orthographe et la phonologie. Les correspondances entre petits segments sont faciles à en-

seigner, car les lettres individuelles sont explicitement représentées dans l'entrée écrite. En revanche, se focaliser sur de petits segments ne permet pas de dériver de manière non-ambiguë leur prononciation. Ils peuvent être prononcés de plusieurs façons, et leur prononciation peut être épelée de plusieurs façons. Par ailleurs, même une lettre simple *a priori*, la lettre « O » en français, nécessite l'information sur les lettres adjacentes, et donc l'information contextuelle, pour savoir s'il s'agit d'un graphème, et donc pour pouvoir la prononcer correctement." Comme nous l'avons illustré dans le Chapitre 4, la prononciation de celui-ci dépend considérablement de l'identité des lettres qui l'avoisine. Nous considérons donc que le recours à des segments plus grand que le graphème est quasi-indispensable pour désambiguïser la sortie phonologique. La taille des unités des segments peut dépendre de la tâche expérimentale, de l'âge ou de la langue (Brown & Deavers, 1999 ; Goswami, Ziegler, Dalton, & Schneider, 2003 ; Ziegler & Goswami, 2005).

En revanche, la flexibilité de la taille des unités à considérer est très peu considérée dans les modèles récents de lecture à haute voix, et ils reposent, au contraire, sur un type d'unité unique et pré-supposé. La possibilité de supposer des segments de taille flexible ou variable contraste fortement avec l'hypothèse de segmentation en graphèmes prônée par les modèles double-voie (Coltheart et al., 2001 ; Perry et al., 2013). Selon ces modèles, le graphème est l'unité à privilégier. Or cette segmentation en graphème n'est pas sans poser de problème, même dans ces modèles. Plus spécifiquement, ces modèles nécessitent d'accompagner les graphèmes par de l'information contextuelle pour prononcer correctement les pseudo-mots [17]. Plus précisément, le modèle DRC fait l'usage de règles dites « contextuelles » qui peuvent changer la correspondance en fonction des lettres qui sont voisines au graphème à convertir. Le modèle CDP++, (Perry, Ziegler, & Zorzi, 2010), lui, recourt à un codage syllabique des graphèmes, permettant ainsi au réseau de conversions graphème-phonème TLA de prendre en compte l'information contextuelle du graphème (au niveau de la syllabe ou il apparaît). Ainsi, l'information contextuelle permet de désambiguïser la sortie phonologique lorsque les correspondances graphème-phonème sont appliquées.

Le mécanisme d'attention, dans le modèle BRAID-Phon, possède la propriété d'être flexible par nature. Ainsi, l'attention pourrait adapter la taille des segments et des unités qu'elle capture en fonction du contexte et des caractéristiques du traitement, pendant celui-ci. Par exemple, on considère, dans de nombreuses situations, écologiques ou expérimentales, qu'il est vraisemblable que le traitement initial d'un stimulus est « global », ce qui fonctionne en général pour des mots, lorsqu'ils sont fréquents par exemple, et bien prédits par le contexte environnant. Ensuite, si ce traitement global « semble ne pas fonctionner », alors on bascule sur un traitement séquentiel. L'attention visuelle, typiquement, permet de décrire ces modes de traitement par l'adaptation des paramètres de la distribution attentionnelle.

De plus, les unités sous-lexicales considérées dans notre approche ne sont pas figées *a priori* et pourraient être décrites par la taille des segments définis par l'attention visuelle. Il est facile

[17] Le terme « correctement » signifie ici que le modèle produit des prononciations de pseudo-mots qui sont similaires à celles observés chez des sujets. Il faut noter également que, notamment dans une langue opaque comme l'anglais, plusieurs prononciations d'un seul pseudo-mot peuvent être considérés comme « correctes », contrairement à des langues transparentes comme l'espagnol ou l'italien (Coltheart & Ulicheva, 2018 ; Pritchard, Coltheart, Palethorpe, & Castles, 2012).

d'imaginer que, pour un lecteur débutant, les segments soient plus petits que pour un lecteur expert. Il est également facile d'imaginer que la taille du segment choisi soit également fonction de « ce que l'on sait après la capture globale » : si le stimulus est inconnu mais semble « facile » (par exemple, un pseudo-mot légal comme « VIRDIN »), alors on peut tenter des segments avec quelques lettres ; en revanche, si le stimulus semble « difficile » (par exemple, une séquence de consonnes ou un pseudo-mot illégal), alors on se rabat sur des segments petits et un décodage sériel lettre-à-lettre. Pour mesurer la « difficulté » du stimulus, le système pourrait utiliser l'état des distributions de probabilité sur les lettres, ou sur l'espace des mots ; on peut également imaginer que le choix des segments pourrait se baser sur un critère lié aux distributions de probabilités sur la sortie phonologique. Ainsi, l'unité sous-lexicale orthographique pertinente serait, par exemple, celle qui produirait la sortie phonologique la moins ambiguë ou la plus certaine.

Un tel mécanisme de sélection automatique des unités à traiter constituerait un « mécanisme de segmentation » du stimulus. Nous suggérons donc ici que ce mécanisme soit basé sur l'état des connaissances au cours du traitement, et affecte les paramètres de la distribution attentionnelle, pour opérationnaliser la sélection des segments. Dans ce travail, nous n'avons pas développé ce mécanisme de segmentation, et suggérons que cela constitue une perspective importante de développement du modèle BRAID-Phon. En effet, un tel mécanisme serait en accord avec les travaux effectués dans le cadre de la thèse d'Emilie Ginestet (2019), qui se sont intéressés à la modélisation de l'apprentissage orthographique, et des trajectoires d'exploration visuelle pendant cette apprentissage, et qui ont décrit un mécanisme d'adaptation des paramètres de la distribution visuo-attentionnelle pour optimiser le gain entropique dans le sous-modèle perceptif des lettres.

Plus en accord avec la théorie de la *grain size* (Ziegler & Goswami, 2005), le modèle BRAID-Phon, avec sa capacité à simuler plusieurs langues différentes (en alphabet latin), permettrait d'étudier les tailles de segments pertinentes pour la lecture de pseudo-mots et l'apprentissage de mots nouveaux à travers les langues. Autrement dit, nous pourrions vérifier l'hypothèse selon laquelle des langues transparentes mettraient en jeu des segments de plus petite taille que des langues opaques.

3 Perspectives

Dans cette section nous allons aborder quelques perspectives et axes d'améliorations et généralisation de ce travail de thèse. Dans la Section 2 de ce chapitre nous avons formulé certaines limites qui sont propres au modèle BRAID-Phon. Nous avons également identifié d'autres limites qui ne sont pas spécifiques à nos postulats théoriques. Dans cette section nous présenterons, dans un premier temps, l'intérêt de fusionner un modèle de parole et un modèle de lecture exhaustifs et détaillés. Dans un second temps nous traiterons la prise en compte et la modélisation de la variabilité dans le cadre de la modélisation bayésienne. Enfin, dans un troisième temps, nous présenterons comment BRAID-Phon pourrait être étendu afin de modéliser et étudier l'apprentissage de la lecture.

3.1 Modèle de lecture incluant un modèle de parole

Comme nous l'avons mentionné dans la revue de littérature (Chapitre 1), les modèles de lecture se focalisent sur les processus centraux permettant la conversion du code orthographique en code phonologique et n'incluent pas de sous-modèle de production de parole prenant en compte les aspects de programmation motrice et l'articulation (Kawamoto & Kello, 1999). En revanche, les données comportementales évaluant la lecture à haute voix reposent en grande partie sur des mesures de temps de réaction pour lire un stimulus. Ce type de données, bien qu'il ait permis d'identifier plusieurs effets dû à des caractéristiques psycholiguistiques pertinentes dans la lecture, dépend fortement des caractéristiques articulatoires lors de la production de parole. Dans un angle plus proche des travaux effectués en cours de cette thèse, nos simulations des données Chronolex (Ferrand et al., 2011) indiquent l'importance de la dimension articulatoire dans la lecture, étant donné que la plus grande part de variance dans les données comportementales est expliquée par la nature articulatoire du premier phonème. Ce caractère est spécifique au temps de réponses de la tâche de lecture (ou dénomination, NMG) qui sont mesurés à partir du signal de parole mesuré. De façon plus générale, on sait qu'environ 50% de la variabilité des temps de réponse en lecture à haute voix est prédite par la nature du premier phonème (Balota et al., 2004 ; Cortese & Khanna, 2007 ; Rastle & Davis, 2002 ; Spieler & Balota, 1997).

Actuellement, BRAID-Phon n'inclut pas de sous-modèle de production de parole. En se basant sur d'autres travaux de modélisation algorithmique bayésienne (Diard, 2015), effectués dans les domaines de la perception, l'apprentissage et la production de parole (Barnaud, 2018 ; Laurent, 2014 ; Laurent, Barnaud, Schwartz, Bessière, & Diard, 2017 ; Patri, 2018 ; Saghiran, 2017), de futurs travaux de recherche et de modélisation pourraient assurer la jonction entre ces différents champs. Un modèle de ce type, plus intégratif, permettrait de mieux appréhender la relation entre la représentation phonologique et articulatoire dans la lecture et de fournir de meilleures explications et prédictions des données comportementales dans les tâches de lecture. Plus spécifiquement, l'inclusion d'un modèle de perception de la parole dans le modèle permettrait de formuler, dans un cadre unique, la propagation de l'entrée acoustique vers le niveau orthographique. Cela pourrait permettre de rendre compte des influences et des interactions croisées entre les traitements phonologiques et les traitements orthographiques observés expérimentalement (Frost & Ziegler, 2007).

De plus, la jonction entre l'apprentissage de la parole et de la lecture permettrait de poser la question des unités pertinentes intervenant lors de la lecture, dans un cadre plus large qui aurait comme objectif d'étudier l'apprentissage et le développement des unités sous-lexicales orthographiques et phonologiques. Plus spécifiquement, un modèle de parole privilégierait vraisemblablement des unités syllabiques dans les représentations phonologiques avant l'apprentissage de la lecture. La lecture et le fait de traiter initialement des segments orthographiques courts pourrait induire des traitements sur des unités phonologiques plus petites telles que les phonèmes (par exemple, Dehaene, Cohen, Morais, & Kolinsky, 2015). Inversement, si les syllabes sont des unités existantes au niveau phonologique, elles pourraient également être à l'origine de la prise en compte d'unités syllabiques au niveau orthographique.

Les unités sous-lexicales orthographiques, telles que les graphèmes ou les syllabes orthographiques, sont des sous-unités apprises et identifiées par le système de lecture. La littérature actuelle ne se questionne pas quant à l'origine de ces unités mais les inclut *a priori*. Ainsi, les modèles actuels de lecture à haute-voix pré-supposent les unités sous-lexicales orthographiques (graphèmes ou syllabes orthographiques ou les deux), et donc peuvent difficilement aborder la question de leur développement.

3.2 Modéliser la variabilité

Le cadre bayésien de modélisation permet de décrire et de formuler mathématiquement l'incertitude, non seulement au niveau des connaissances mais également au niveau des observations (mesures effectuées sur le modèle). Actuellement, nous exploitons uniquement la formulation de l'incertitude au niveau des connaissances représentées et manipulées dans le modèle, mais nous ne l'avons pas exploitée pour modéliser la variabilité des sorties phonologiques ou des temps de réponse. Cela nous parait être une limite importante, puisque l'évaluation quantitative des modèles peut se baser sur la comparaison des distributions de temps de réaction (simulés et observés), et pourrait se baser sur la comparaison des distributions des sorties phonologiques (simulés et observés) (Jacobs & Grainger, 1994)).

En effet, ni la forme, ni l'entropie de ces distributions n'ont été mentionnées dans les travaux présentés dans cette thèse. L'entropie, par exemple, permet de quantifier le degré d'incertitude des distributions de probabilité du modèle en sortie. La forme des distributions de probabilité des phonèmes, elle, peut nous fournir une estimation de la distribution des prononciations observées chez les lecteurs. En effet, il est connu que les prononciations produites pour des pseudo-mots sont, notamment dans les langues opaques, multiples (Coltheart & Ulicheva, 2018 ; Pritchard et al., 2012).

Par ailleurs, nous avons établi des temps de réaction du modèle en définissant un algorithme de décision qui se base sur un seuil de probabilité fixe. Il est tout à fait possible d'enrichir ce mécanisme de calcul des temps de réaction par un processus aléatoire, ce qui permettrait de simuler une première dimension de la variabilité des temps de réponse. Ainsi, le modèle BRAID-Phon produirait une distribution de temps de réponses au lieu d'une valeur unique. Cela rapprocherait les modèles BRAID et BRAID-Phon des modèles qui étudient les distributions des temps de réponse, par exemple dans le contexte de la décision lexicale (Dufau, Grainger, & Ziegler, 2012 ; Ratcliff, Gomez, & McKoon, 2004 ; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008).

En outre, introduire de l'incertitude au niveau des paramètres du modèle, ce qui est possible théoriquement dans le cadre des modèles bayesiens, pourrait nous permettre de modéliser la variabilité observée, par exemple entre les participants. Nous savons également qu'une grande variabilité peut exister au niveau de la mesure des temps de réponse entre les individus (par exemple, voir Adelman et al., 2013), voire, pour un même individu, entre plusieurs essais. De plus, les études comportementales, y compris les *megastudy*, montrent une corrélation toute relative entre elles (Sibley et al., 2009), suggérant une variabilité entre les études, qui pourraient correspondre à des

variations de paramètres, soit liées aux participants de ces études, soit induites par les conditions expérimentales (protocole expérimental, répartition statistique et temporelle des stimuli, formulation de la consigne, etc.).

3.3 L'apprentissage de la lecture

Le modèle BRAID-Phon, tel que nous l'avons développé, est un modèle permettant de simuler la lecture chez un lecteur expert, c'est-à-dire un lecteur adulte sain ayant de bonnes capacités de lecture et donc possédant un lexique orthographique et phonologique riche. Notons que la dimension de l'apprentissage n'est pas prise en compte dans le travail présenté dans cette thèse. Cependant, comme énoncé précédemment, Emilie Ginestet (2019), dans ses travaux de thèse, s'est intéressée à la modélisation de l'apprentissage orthographique et a décrit un mécanisme d'apprentissage de nouvelles traces orthographiques, et un modèle d'exploration visuo-attentionnelle pendant cet apprentissage, permettant donc de simuler dans le modèle l'acquisition des représentations orthographiques de nouveaux mots.

Le modèle BRAID-Phon nous a permis d'expliciter les mécanismes phonologiques en lien avec la lecture de mots et également de développer une piste vers la lecture de pseudo-mots. Ainsi, le modèle « prépare le terrain » pour s'intéresser à l'implémentation, dans le cadre bayésien, de l'hypothèse de l'auto-apprentissage de Share (1995). Notons que le travail de thèse d'Alexandra Steinhilber, actuellement en cours, s'oriente dans cette direction et a pour objectif de développer une extension du modèle BRAID-Phon, baptisée BRAID-Acq.

Cette hypothèse d'auto-apprentissage peut être formulée de la façon suivante. A partir d'une nouvelle entrée orthographique, les lettres sont reconnues et décodées phonologiquement. Ainsi, des traces orthographiques et phonologiques sont construites. Dans l'hypothèse d'auto-apprentissage, un décodage phonologique réussi est une condition nécessaire pour réaliser l'apprentissage orthographique. Cette hypothèse n'a été implémentée que dans le cadre des modèles de lecture double-voie (Pritchard et al., 2018 ; Ziegler et al., 2014, 2020). BRAID-Acq permettra de simuler et d'étudier cette situation d'apprentissage dans le cadre d'un modèle de lecture à voie unique.

Par ailleurs, en incluant des mécanismes attentionnels à plusieurs niveaux, le modèle BRAID-Acq pourrait également évaluer une alternative à l'apprentissage de Share (1995). Celle-ci provient des travaux de LaBerge et Samuels (1974) qui proposent que le contrôle attentionnel serait impliqué sur différents niveaux de traitement lors de l'apprentissage (orthographe, phonologie, sémantique). Les ressources attentionnelles sont dirigées de manière sélective, sur un niveau à la fois, en fonction du degré d'automatisation des processus de lecture (reconnaissance, décodage et compréhension). Ainsi, une formulation mathématique des différents composants de l'attention dans le système cognitif permettrait de simuler cette hypothèse d'apprentissage et de la comparer avec l'hypothèse de Share (1995).

Bibliographie

Adelman, J. S. (2011). Letters in time and retinotopic space. *Psychological Review*, *118*(4), 570–582. doi: 10.1037/a0024811

3 citations, pages 4, 36 & 132

Adelman, J. S., & Brown, G. D. (2008). Methods of testing and diagnosing model error : Dual and single route cascaded models of reading aloud. *Journal of Memory and Language*, *59*(4), 524–544. doi: 10.1016/j.jml.2007.11.008

4 citations, pages 36, 71, 88 & 134

Adelman, J. S., Marquis, S. J., Sabatos-DeVito, M. G., & Estes, Z. (2013). The unexplained nature of reading. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *39*(4), 1037–1053. doi: 10.1037/a0031829

2 citations, pages 134, 147

Allport, D. A., & Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society of London*, *295*(1077, Series B), 397–410. doi: 10.1098/rstb.1981.0148

cité page 139

Angelelli, P., Marinelli, C. V., & Zoccolotti, P. (2010). Single or dual orthographic representations for reading and spelling? A study of Italian dyslexic–dysgraphic and normal children. *Cognitive Neuropsychology*, *27*(4), 305–333. doi: 10.1080/02643294.2010.543539

cité page 139

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*(4), 678–723. doi: 10.1037/0033-295x.105.4.678-723

17 citations, pages v, 4, 5, 6, 19, 24, 25, 26, 27, 28, 34, 39, 94, 97, 131, 132 & 140

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438–481. doi: 10.1037/a0023851

cité page 29

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences*, *93*(24), 13468–13472. doi: 10.1073/pnas.93.24.13468

cité page 142

Baddeley, A. (2003). Working memory and language : An overview. *Journal of Communication Disorders*, *36*(3), 189–208. doi: 10.1016/S0021-9924(03)00019-4

cité page 142

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task : Lexical access and/or production? *Journal of Memory and Language*, *24*(1), 89–106. doi: 10.1016/0749-596X(85)90017-8

cité page 71

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology : General*, *133*(2), 283–316. doi: 10.1037/0096-3445.133.2.283

2 citations, pages 76, 146

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, *39*(3), 445–459. doi: 10.3758/bf03193014

3 citations, pages 70, 103 & 141

Barnaud, M.-L. (2018). *Modèlisation bayésienne du développement conjoint de la perception, l'action et la phonologie* (Thèse de doctorat). Université Grenoble-Alpes, Grenoble, France.

cité page 146

Baxter, D. M., & Warrington, E. K. (1985). Category specific phonological dysgraphia. *Neuropsychologia*, *23*(5), 653–666. doi: 10.1016/0028-3932(85)90066-1

cité page 139

Beauvois, M.-F., & Derouesné, J. (1979). Phonological alexia : Three dissociations. *Journal of Neurology, Neurosurgery & Psychiatry*, *42*(12), 1115–1124. doi: 10.1136/jnnp.42.12.1115

cité page 10

Behrmann, M., & Bub, D. (1992). Surface dyslexia and dysgraphia : dual routes, single lexicon. *Cognitive Neuropsychology*, *9*(3), 209–251. doi: 10.1080/02643299208252059

cité page 139

Besner, D., & McCann, R. S. (1987). Word frequency and pattern distortion in visual word identification and production : An examination of four classes of models. In M. Coltheart (Ed.), *Attention and performance XII : The psychology of reading* (p. 201-219). Hillsdale, NJ : Lawrence Erlbaum Associates, Inc.

cité page 88

Besner, D., Risko, E. F., Stolz, J. A., White, D., Reynolds, M., O'Malley, S., & Robidoux, S. (2016). Varieties of attention. *Current Directions in Psychological Science*, *25*(3), 162–168. doi: 10.1177/0963721416639351

2 citations, pages 36, 139

Bessière, P., Laugier, C., & Siegwart, R. (Eds.). (2008). *Probabilistic reasoning and decision making in sensory-motor systems* (Vol. 46). Berlin : Springer.

2 citations, pages 44, 169

Bessière, P., Mazer, E., Ahuactzin, J. M., & Mekhnacha, K. (2013). *Bayesian programming*. Boca Raton, FL : CRC Press.

2 citations, pages 44, 170

Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia : The visual attention span deficit hypothesis. *Cognition*, *104*(2), 198–230. doi: 10.1016/j.cognition.2006.05.009

cité page 35

Brown, G. D., & Deavers, R. P. (1999). Units of analysis in nonword reading : Evidence from children and adults. *Journal of Experimental Child Psychology*, *73*(3), 208–242. doi: 10.1006/jecp.1999.2502

cité page 144

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*, *58*(5), 412–424. doi: 10.1027/1618-3169/a000123

cité page 71

Brysbaert, M., & Nazir, T. (2005). Visual constraints in written word recognition : Evidence from the optimal viewing-position effect. *Journal of Research in Reading*, *28*(3), 216–228. doi: 10.1111/j.1467-9817.2005.00266.x

cité page 89

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times : Evidence from the dutch lexicon project 2. *Journal of Experimental Psychology : Human Perception and Performance*, *42*(3), 441–458. doi: 10.1037/xhp0000159

cité page 141

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance : The case for single-patient studies. *Brain and cognition*, *5*(1), 41–66. doi: 10.1016/0278-2626(86)90061-8

cité page 10

Carr, T., & Pollatsek, A. (1985). Recognizing printed words : A look at current models. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research : Advances in theory and practices 5* (p. 1-82). San Diego, CA : Academic Press.

cité page 88

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition : Cross-task comparisons. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *23*(4), 857–871. doi: 10.1037/0278-7393.23.4.857

cité page 71

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. doi: 10.1016/S0010-0277(03)00164-1

cité page 143

Colas, F., Diard, J., & Bessiere, P. (2010). Common bayesian models for common cognitive issues. *Acta Biotheoretica*, *58*(2-3), 191–216. doi: 10.1007/s10441-010-9101-1

cité page 169

Coltheart, M. (2005). Modeling reading : The dual-route approach. In *The science of reading : A handbook* (p. 6–23). Blackwell Publishing Ltd. doi: 10.1002/9780470757642.ch1

5 citations, pages 5, 11, 13, 34 & 132

Coltheart, M. (2017). The assumptions of cognitive neuropsychology : Reflections on Caramazza (1984, 1986). *Cognitive Neuropsychology*, *34*(7-8), 397–402. doi:

10.1080/02643294.2017.1324950

cité page 10

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud : Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*(4), 589–608. doi: 10.1037/0033-295X.100.4.589

4 citations, pages 10, 13, 37 & 132

Coltheart, M., & Funnell, E. (1987). Reading and writing : One lexicon or two? In A. Allport, D. G. MacKay, & W. Prinz (Eds.), *Language perception and production : Relationships between listening, speaking, reading and writing* (pp. 313–339). San Diego, CA : Academic Press.

cité page 139

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud : Evidence for dual-route models of reading. *Journal of Experimental Psychology : Human Perception and Performance*, *20*(6), 1197–1211. doi: 10.1037/0096-1523.20.6.1197

5 citations, pages 10, 36, 37, 72 & 132

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC : A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. doi: 10.1037/0033-295X.108.1.204

16 citations, pages v, 4, 5, 11, 12, 13, 14, 15, 50, 72, 89, 94, 131, 132, 139 & 144

Coltheart, M., & Ulicheva, A. (2018). Why is nonword reading so variable in adult skilled readers? *PeerJ*, *6*, e4879. doi: 10.7717/peerj.4879

2 citations, pages 144, 147

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and bigrams : Orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *Journal of Experimental Psychology : Human Perception and Performance*, *35*(2), 461–479. doi: 10.1037/a0013480

cité page 138

Content, A., Mousty, P., & Radeau, M. (1990). Brulex. une base de données lexicales informatisée pour le français écrit et parlé. *L'année psychologique*, *90*(4), 551–566. doi: 10.3406/psy.1990.29428

cité page 27

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables : An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, *60*(8), 1072–1082. doi: 10.1080/17470210701315467

cité page 146

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*(3), 713–758. doi: 10.1037/a0019738

cité page 4

Dehaene, S. (2009). *Reading in the brain : The science and evolution of a human invention.* New York, NY : Viking.

cité page 3

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate : Behavioural and cere-

bral changes induced by reading acquisition. *Nature Reviews Neuroscience*, *16*(4), 234–244. doi: 10.1038/nrn3924

cité page 146

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words : a proposal. *Trends in cognitive sciences*, *9*(7), 335–341. doi: 10.1016/j.tics.2005.05.004

cité page 4

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. doi: 10.1037/0033-295X.93.3.283

cité page 11

Diard, J. (2015). *Bayesian algorithmic modeling in cognitive science* (Habilitation à Diriger des Recherches (HDR)). Université Grenoble Alpes.

4 citations, pages 43, 44, 146 & 169

Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say "no" to a nonword : A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *38*(4), 1117–1128. doi: 10.1037/a0026948

cité page 147

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read : Evidence from the national reading panel's meta-analysis. *Reading Research Quarterly*, *36*(3), 250–287. doi: 10.1598/rrq.36.3.2

cité page 143

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621–636. doi: 10.1016/S0042-6989(01)00301-7

cité page 32

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT : A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. doi: 10.1037/0033-295x.112.4.777

2 citations, pages 4, 137

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In *Models of working memory* (pp. 102–134). Cambridge University Press. doi: 10.1017/CBO9781139174909.007

cité page 142

Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., … Wolfe, J. M. (2011). Visual attention. *Wiley Interdisciplinary Reviews : Cognitive Science*, *2*(5), 503–514. doi: 10.1002/wcs.127

cité page 39

Facoetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., … Mascetti, G. G. (2006). The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive Neuropsychology*, *23*(6), 841–855. doi: 10.1080/02643290500483090

cité page 139

Ferrand, L. (2000). Reading aloud polysyllabic words and nonwords : The syllabic length effect reexamined. *Psychonomic Bulletin & Review*, *7*(1), 142–148. doi: 10.3758/bf03210733

cité page 35

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., ... Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking : Evidence from chronolex. *Frontiers in Psychology*, *2*, 306, 1–10. doi: 10.3389/fpsyg.2011.00306

12 citations, pages viii, 7, 69, 70, 71, 73, 74, 80, 91, 134, 141 & 146

Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition : Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology Section A*, *45*(3), 353–372. doi: 10.1080/02724989208250619

cité page 18

Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology Section A*, *47*(2), 365–382. doi: 10.1080/14640749408401116

cité page 18

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., ... Pallier, C. (2010). The French Lexicon Project : Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. doi: 10.3758/brm.42.2.488

4 citations, pages 72, 74, 88 & 141

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. doi: 10.1016/S0022-5371(73)80042-8

cité page 10

Franceschini, S., Bertoni, S., Puccio, G., Mancarella, M., Gori, S., & Facoetti, A. (2020). Local perception impairs the lexical reading route. *Psychological Research*. doi: 10.1007/s00426-020-01326-z

cité page 139

Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012, may). A causal link between visual spatial attention and reading acquisition. *Current Biology*, *22*(9), 814–819. doi: 10.1016/j.cub.2012.03.013

cité page 139

Frost, R., & Ziegler, J. C. (2007). Speech and spelling interaction : The interdependence of visual and auditory word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. Oxford, UK : Oxford University Press. doi: 10.1093/oxfordhb/9780198568971.013.0007

cité page 146

Fu, Y., Wang, H., Guo, H., Bermúdez-Margaretto, B., & Martínez, A. D. (2020). What, where, when and how of visual word recognition : A bibliometrics review. *Language and Speech*, 002383092097471. doi: 10.1177/0023830920974710

2 citations, pages 3, 9

Geyer, L. H. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics*, *22*(5), 487–490. doi: 10.3758/BF03199515

cité page 54

Gilet, E., Diard, J., & Bessière, P. (2011). Bayesian action-perception computational model : Interaction of production and recognition of cursive letters. *PLoS ONE*, *6*(6), e20387. doi: 10.1371/journal.pone.0020387

3 citations, pages 48, 51 & 116

Ginestet, E. (2016). *Modélisation probabiliste de reconnaissance visuelle des mots : simulation d'effets d'amorçage et de compétition lexicale* (Mémoire de Master non publié). Univ. Grenoble Alpes.

2 citations, pages 100, 137

Ginestet, E. (2019). *Modélisation bayésienne et étude expériementale du rôle de l'attention visuelle dans l'acquisition des connaissances lexicales orthographiques* (Thèse de doctorat non publiée). Université Grenoble-Alpes.

8 citations, pages 43, 44, 46, 51, 116, 138, 145 & 148

Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical decision : The role of visual attention. *Vision Research*, *159*, 10–20. doi: 10.1016/j.visres.2019.03.003

7 citations, pages 72, 74, 88, 89, 90, 137 & 141

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology : Human Perception and Performance*, *5*(4), 674–691. doi: 10.1037/0096-1523.5.4.674

4 citations, pages 22, 23, 33 & 97

Gomez, P., Ratcliff, R., & Perea, M. (2008). The Overlap model : A model of letter position coding. *Psychological Review*, *115*(3), 577–600. doi: 10.1037/a0012667

3 citations, pages 4, 132 & 137

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies : How flexible is the choice of reading units? *Applied Psycholinguistics*, *24*(2), 235–247. doi: 10.1017/S0142716403000134

cité page 144

Grainger, J. (2018). Orthographic processing : A 'mid-level' vision of reading : The 44th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, *71*(2), 335–359. doi: 10.1080/17470218.2017.1314515

cité page 132

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition : A multiple read-out model. *Psychological Review*, *103*(3), 518–565. doi: 10.1037/0033-295x.103.3.518

6 citations, pages 4, 11, 15, 18, 50 & 71

Grainger, J., & Jacobs, A. M. (1998). On localist connectionism and psychological science. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 1–38). L. Erlbaum Associates.

cité page 11

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 981–998. doi: 10.1080/02724980443000386

cité page 71

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 54, 1–13. doi: 10.3389/fpsyg.2011.00054

cité page 19

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia : Insights from connectionist models. *Psychological Review*, *106*(3), 491–528. doi: 10.1037/0033-295x.106.3.491

2 citations, pages 23, 24

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading : Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720. doi: 10.1037/0033-295X.111.3.662

cité page 21

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall/CRC.

cité page 80

Hendrix, P., Ramscar, M., & Baayen, H. (2019). NDR_A : A single route model of response times in the reading aloud task based on discriminative learning. *PLOS ONE*, *14*(7), e0218802. doi: 10.1371/journal.pone.0218802

8 citations, pages v, 5, 19, 29, 30, 31, 97 & 131

Hillis, A. E., & Rapp, B. C. (2004). Cognitive and neural substrates of written language comprehension and production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences iii*. Cambridge, MA : MIT Press.

cité page 139

Houghton, G. (2018). Action and perception in literacy : A common-code for spelling and reading. *Psychological Review*, *125*(1), 83–116. doi: 10.1037/rev0000084

cité page 139

Huey, E. B. (1908). *The Psychology and Pedagogy of Reading*. Oxford, UK : Macmillan.

cité page 9

Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology : Human Perception and Performance*, *18*(4), 1174–1188. doi: 10.1037/0096-1523.18.4.1174

cité page 18

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition : Sampling the state of the art. *Journal of Experimental Psychology : Human Perception and Performance*, *20*(6), 1311–1334. doi: 10.1037/0096-1523.20.6.1311

4 citations, pages 3, 32, 131 & 147

Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-p : An interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (p. 147–188). Lawrence Erlbaum Associates Publishers.

6 citations, pages v, 11, 18, 19, 88 & 131

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*(4), 723–750. doi: 10.1006/jmla.2001.2827

cité page 22

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*(6), 687–715. doi: 10.1016/0749-596x(90)90044-z

cité page 22

Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology : Human Perception and Performance*, *16*(1), 92–105. doi: 10.1037/0096-1523.16.1.92

cité page 35

Jaynes, E. T. (2003). *Probability theory : The logic of science*. Cambridge, UK : Cambridge University Press.

cité page 170

Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision : Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition*, *55*(2), 332–340. doi: 10.1016/j.bandc.2004.02.035

3 citations, pages 28, 72 & 140

Kapnoula, E. C., Protopapas, A., Saunders, S. J., & Coltheart, M. (2017). Lexical and sublexical effects on visual word recognition in Greek : Comparing human behavior to the Dual Route Cascaded model. *Language, Cognition and Neuroscience*, *32*(10), 1290–1304. doi: 10.1080/23273798.2017.1355059

5 citations, pages 14, 37, 89, 132 & 140

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity : A parallel distributed processing account. *Journal of Memory and Language*, *32*(4), 474–516. doi: 10.1006/jmla.1993.1026

cité page 37

Kawamoto, A. H., & Kello, C. T. (1999). Effect of onset cluster complexity in speeded naming : A test of rule-based approaches. *Journal of Experimental Psychology : Human Perception and Performance*, *25*(2), 361–375. doi: 10.1037/0096-1523.25.2.361

cité page 146

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation : Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *24*(4), 862–885. doi: 10.1037/0278-7393.24.4.862

cité page 37

Kay, J., & Marcel, A. (1981). One process, not two, in reading aloud : Lexical analogies do the work of non-lexical rules. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 397–413. doi: 10.1080/14640748108400800

2 citations, pages 33, 97

Kim, J., Taft, M., & Davis, C. (2004). Orthographic–phonological links in the lexicon : When lexical and sublexical information conflict. *Reading and Writing*, *17*(1/2), 187–218. doi: 10.1023/B:READ.0000013805.77364.cd

cité page 138

Kučera, H., Francis, W., Bell, L., Twaddell, W., Carroll, J., & Marckworth, M. (1967). *Computational analysis of present-day american english*. Providence, RI : Brown University Press.

cité page 21

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, *66*(3), 563–580. doi: 10.1080/17470218.2012.658820

cité page 89

Kutner, M. (2005). *Applied linear statistical models.* Boston, MA : McGraw-Hill Irwin.

cité page 77

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. doi: 10.1016/0010-0285(74)90015-2

4 citations, pages 35, 132, 140 & 148

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958) : Still no identification without attention. *Psychological Review*, *111*(4), 880–913. doi: 10.1037/0033-295X.111.4.880

cité page 139

Laurent, R. (2014). *COSMO : un modèle bayésien des interactions sensori-motrices dans la perception de la parole* (Thèse de doctorat non publiée). Université de Grenoble, Grenoble, France.

cité page 146

Laurent, R., Barnaud, M.-L., Schwartz, J.-L., Bessière, P., & Diard, J. (2017). The complementary roles of auditory and motor information evaluated in a bayesian perceptuo-motor model of speech perception. *Psychological Review*, *124*(5), 572–602. doi: 10.1037/rev0000069

cité page 146

Lebeltel, O., Bessière, P., Diard, J., & Mazer, E. (2004). Bayesian robot programming. *Autonomous Robots*, *16*(1), 49–79. doi: 10.1023/B:AURO.0000008671.38949.43

5 citations, pages vii, 44, 169, 170 & 171

Leinenger, M., & Rayner, K. (2015). Eye movements and visual attention during reading. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (Vol. 281, pp. 282–300). MIT Press.

cité page 140

Marr, D. (1982). *Vision. a computational investigation into the human representation and processing of visual information.* New York, NY : W.H. Freeman and Company.

2 citations, pages 138, 169

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia : A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*(3), 175–199. doi: 10.1007/BF01067101

cité page 10

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones : Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology : Human Perception and Performance*, *13*(1), 14–24. doi: 10.1037/0096-1523.13.1.14

cité page 23

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks : A historical and tutorial review. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00503

cité page 49

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception : I. an account of basic findings. *Psychological Review*, *88*(5), 375–407. doi: 10.1037/0033-295X.88.5.375

8 citations, pages 4, 11, 18, 36, 38, 46, 132 & 137

McKeown, G. J., & Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological Methods*, *19*(1), 155–174. doi: 10.1037/a0034282

cité page 80

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Lambon Ralph, M. A., Patterson, K., … Price, C. J. (2005). Dissociating reading processes on the basis of neuronal interactions. *Journal of Cognitive Neuroscience*, *17*(11), 1753–1765. doi: 10.1162/089892905774589190

cité page 4

Meeter, M., Marzouki, Y., Avramiea, A. E., Snell, J., & Grainger, J. (2020). The role of attention in word recognition : Results from OB1-reader. *Cognitive Science*, *44*(7), e12846. doi: 10.1111/cogs.12846

cité page 140

Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read : A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. doi: 10.1037/a0026744

cité page 143

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178. doi: 10.1037/h0027366

cité page 9

Morton, J., & Broadbent, D. E. (1967). Passive versus active recognition models, or is your homunculus really necessary. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 103–110). Cambridge, MA : The MIT Press.

cité page 9

Mozer, M. C., & Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge : A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, *2*(2), 96–123. doi: 10.1162/jocn.1990.2.2.96

3 citations, pages 35, 132 & 140

Mueller, S. T., & Weidemann, C. T. (2012). Alphabetic letter identification : Effects of perceivability, similarity, and bias. *Acta Psychologica*, *139*(1), 19–37. doi: 10.1016/j.actpsy.2011.09.014

2 citations, pages 53, 54

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access : The rank hypothesis. *Psychological Review*, *111*(3), 721–756. doi: 10.1037/0033-295X.111.3.721

cité page 71

New, B. (2006). Lexique 3 : Une nouvelle base de données lexicales. In *Actes de la conférence traitement automatique des langues naturelles (taln 2006)*.

2 citations, pages 103, 118

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual

word recognition : New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45–52. doi: 10.3758/BF03193811

3 citations, pages 71, 72 & 141

Norris, D. (2005). How do computational models help us develop better theories? In A. Cutler (Ed.), (pp. 331–346). Mahwah, NJ : Lawrence Erlbaum.

2 citations, pages 3, 132

Norris, D. (2006). The bayesian reader : Explaining word recognition as an optimal bayesian decision process. *Psychological Review*, *113*(2), 327–357. doi: 10.1037/0033-295x.113.2.327

3 citations, pages 4, 36 & 49

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, *17*(10), 517–524. doi: 10.1016/j.tics.2013.08.003

6 citations, pages 4, 32, 71, 88, 131 & 137

Patri, J.-F. (2018). *Bayesian modeling of speech motor planning : variability, multisensory goals and perceptuo-motor interactions* (Thèse de doctorat non publiée). Université Grenoble Alpes, Grenoble, France.

cité page 146

Pattamadilok, C., Perre, L., & Ziegler, J. C. (2011). Beyond rhyme or reason : ERPs reveal task-specific activation of orthography on spoken language. *Brain and Language*, *116*(3), 116–124. doi: 10.1016/j.bandl.2010.12.002

cité page 142

Patterson, K. (1986). Lexical but nonsemantic spelling? *Cognitive Neuropsychology*, *3*(3), 341–367. doi: 10.1080/02643298608253363

cité page 139

Patterson, K., & Shewell, C. (1987). Speak and spell : Dissociations and word-class effects. In *The cognitive neuropsychology of language* (pp. 273–294). Hillsdale, NJ : Lawrence Erlbaum Associates, Inc.

cité page 139

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems : Networks of plausible inference*. San Mateo, CA : Morgan Kaufmann Publishers.

cité page 43

Pearl, J. (2018). *The book of why : The new science of cause and effect*. New York, NY : Basic Books.

cité page 131

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming : Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language*, *37*(3), 382–410. doi: 10.1006/jmla.1997.2516

cité page 22

Perry, C., & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology : Human Perception and Performance*, *28*(4), 990–1001. doi: 10.1037/0096-1523.28.4.990

4 citations, pages 14, 89, 132 & 140

Perry, C., Ziegler, J. C., Braun, M., & Zorzi, M. (2010). Rules versus statistics in reading aloud : New evidence on an old debate. *European Journal of Cognitive Psychology*, *22*(5), 798–812. doi:

10.1080/09541440902978365

cité page 132

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories : The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. doi: 10.1037/0033-295x.114.2.273

13 citations, pages v, 4, 5, 11, 15, 16, 17, 36, 72, 94, 95, 131 & 132

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables : Large-scale modeling of reading aloud with the connectionist dual process (CDP++) model. *Cognitive Psychology*, *61*(2), 106–151. doi: 10.1016/j.cogpsych.2010.04.001

12 citations, pages 4, 5, 11, 17, 18, 36, 71, 95, 131, 137, 139 & 144

Perry, C., Ziegler, J. C., & Zorzi, M. (2013). A computational and empirical investigation of graphemes in reading. *Cognitive Science*, *37*(5), 800–828. doi: 10.1111/cogs.12030

9 citations, pages 17, 36, 94, 95, 133, 137, 139, 141 & 144

Perry, C., Ziegler, J. C., & Zorzi, M. (2014a). CDP++.italian : Modelling sublexical and supralexical inconsistency in a shallow orthography. *PLoS ONE*, *9*(4), e94291. doi: 10.1371/journal.pone.0094291

2 citations, pages 11, 17

Perry, C., Ziegler, J. C., & Zorzi, M. (2014b). When silent letters say more than a thousand words : An implementation and evaluation of CDP++ in french. *Journal of Memory and Language*, *72*, 98–115. doi: 10.1016/j.jml.2014.01.003

2 citations, pages 11, 17

Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale computational models. *Psychological science*, *30*(3), 386–395. doi: 10.1177/0956797618823540

2 citations, pages 10, 38

Phénix, T. (2018). *Modélisation bayésienne algorithmique de la reconnaissance visuelle de mots et de l'attention visuelle* (Thèse de doctorat non publiée). Université Grenoble-Alpes.

16 citations, pages 6, 7, 43, 44, 46, 48, 51, 52, 57, 63, 64, 74, 131, 133, 137 & 169

Phénix, T., Diard, J., & Valdois, S. (2016). Les modèles computationnels de lecture. In *Traité de neurolinguistique* (p. 167–182). De Boeck Supérieur.

3 citations, pages 3, 32 & 38

Phénix, T., Valdois, S., & Diard, J. (2018). Reconciling opposite neighborhood frequency effects in lexical decision : Evidence from a novel probabilistic model of visual word recognition. In *40th Annual Conference of the Cognitive Science Society (CogSci 2018)* (p. 2238-2243). Madison, WI.

cité page 137

Phénix, T., Ginestet, E., Valdois, S., & Diard, J. (2020). *Visual attention matters during word recognition : A Bayesian modeling approach.* (Article soumis pour publication)

3 citations, pages 44, 100 & 137

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia : Extension to sequential processing. *Cognitive Science*, *23*(4), 543–568. doi: 10.1207/s15516709cog2304_7

6 citations, pages 6, 23, 24, 36, 37 & 141

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading : Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. doi: 10.1037/0033-295X.103.1.56

9 citations, pages 4, 5, 19, 21, 23, 34, 37, 97 & 132

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), 1–49. doi: 10.1111/cogs.12571

3 citations, pages 4, 10 & 148

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading : Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology : Human Perception and Performance*, *38*(5), 1268–1288. doi: 10.1037/a0026703

2 citations, pages 144, 147

Rapp, B., & Lipka, K. (2011). The literate brain : The relationship between spelling and reading. *Journal of Cognitive Neuroscience*, *23*(5), 1180–1197. doi: 10.1162/jocn.2010.21507

cité page 139

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology : Human Perception and Performance*, *28*(2), 307–314. doi: 10.1037/0096-1523.28.2.307

cité page 146

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords : The ARC nonword database. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1339–1362. doi: 10.1080/02724980244000099

cité page 14

Rastle, K., Havelka, J., Wydell, T. N., Coltheart, M., & Besner, D. (2009). The cross-script length effect : Further evidence challenging PDP models of reading aloud. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *35*(1), 238–246. doi: 10.1037/a0014361

cité page 89

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182. doi: 10.1037/0033-295x.111.1.159

cité page 147

Rayner, K. (1998). Eye movements in reading and information processing : 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi: 10.1037/0033-2909.124.3.372

cité page 37

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture : Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi: 10.1080/17470210902816461

cité page 37

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. J. (2012). *The psychology of reading*. New York, NY : Psychology Press.

cité page 3

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews :*

Cognitive Science, *1*(6), 787–799. doi: 10.1002/wcs.68

4 citations, pages 3, 4, 38 & 131

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading : Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–476. doi: 10.1017/s0140525x03000104

3 citations, pages 4, 35 & 137

Rescorla, R., & Wagner, A. (1972). A theory of pavlovian conditioning : Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning ii : Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

cité page 30

Risko, E. F., Stolz, J. A., & Besner, D. (2010). Spatial attention modulates feature crosstalk in visual word processing. *Attention, Perception, & Psychophysics*, *72*(4), 989–998. doi: 10.3758/APP.72.4.989

3 citations, pages 32, 36 & 139

Roberts, M. A., Rastle, K., Coltheart, M., & Besner, D. (2003). When parallel processing in visual word recognition is not enough : New evidence from naming. *Psychonomic Bulletin & Review*, *10*(2), 405–414. doi: 10.3758/BF03196499

cité page 36

Roelofs, A. (2005). From Popper to Lakatos : A case for cumulative Computational Modeling. In A. Cutler (Ed.), (pp. 313–330). Mahwah, NJ : Lawrence Erlbaum.

cité page 4

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of english verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing : Explorations in the microstructure of cognition : Psychological and biological models* (Vol. 2). MIT Press.

cité page 21

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review*, *81*(2), 99–118. doi: 10.1037/h0036117

3 citations, pages v, 11 & 12

Saghiran, A. (2017). *COSMO WordPhon, une modélisation de l'influence de l'information lexicale dans l'apprentissage des catégories phonémiques* (Mémoire de Master non publié). Grenoble-INP, Phelma.

cité page 146

Saghiran, A., Valdois, S., & Diard, J. (2019). Bayesian modeling of word and pseudo-word reading in a single-route architecture [poster presentation]. In *The International Convention for Psychological Science (ICPS 2019)*. doi: 10.13140/RG.2.2.28090.95689

cité page 139

Schmalz, X., Robidoux, S., Castles, A., & Marinus, E. (2020). Variations in the use of simple and context-sensitive grapheme-phoneme correspondences in english and german developing readers. *Annals of Dyslexia*, *70*(2), 180–199. doi: 10.1007/s11881-019-00189-3

cité page 143

Seidenberg, M. S. (2012). Computational models of reading : Connectionist and dual-route approaches. In *The cambridge handbook of psycholinguistics* (pp. 186–203). Cambridge Uni-

versity Press. doi: 10.1017/CBO9781139029377.010

cité page 34

Seidenberg, M. S. (2017). *Language at the speed of sight.* New York, NY : Basic Books.

cité page 3

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. doi: 10.1037/0033-295x.96.4.523

12 citations, pages v, 4, 5, 6, 19, 20, 21, 23, 72, 97, 132 & 141

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level : Matching the grain of theory and data. *Psychological Science*, *9*(3), 234–237. doi: 10.1111/1467-9280.00046

2 citations, pages 36, 70

Seidenberg, M. S., & Plaut, D. C. (2006). Progress in understanding word reading : Data fitting versus theory building. In S. Andrews (Ed.), *From inkmarks to ideas : Current issues in lexical processing* (pp. 25–49). New York, NY : Psychology Press.

cité page 34

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology : Human Perception and Performance*, *20*(6), 1177–1196. doi: 10.1037/0096-1523.20.6.1177

cité page 17

Shallice, T., & MacCarthy, R. (1985). Surface dyslexia. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia : Neuropsychological and cognitive studies of phonological reading* (p. 361–397). doi: 10.4324/9781315108346

cité page 10

Share, D. L. (1995). Phonological recoding and self-teaching : sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218. doi: 10.1016/0010-0277(94)00645-2

cité page 148

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (2009). Error, error everywhere : A look at megastudies of word reading. In *21th Annual Conference of the Cognitive Science Society (CogSci 2009)*.

2 citations, pages 70, 147

Simpson, I. C., Mousikou, P., Montoya, J. M., & Defior, S. (2012). A letter visual-similarity matrix for latin-based alphabets. *Behavior Research Methods*, *45*(2), 431–439. doi: 10.3758/s13428-012-0271-4

6 citations, pages v, 54, 55, 56, 57 & 58

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-reader : A model of word recognition and eye movements in text reading. *Psychological Review*, *125*(6), 969–984. doi: 10.1037/rev0000119

4 citations, pages 4, 35, 137 & 140

Snoop Dogg. (2018). Snoop Dogg thanks himself as he gets star on Hollywood Walk Of Fame. ([Online; posted 20-November-2018] https://news.sky.com/story/snoop-dogg-thanks-himself-as-he-gets-star-on-hollywood-walk-of-fame-11558636)

cité page 1

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the

item level. *Psychological Science*, *8*(6), 411–416. doi: 10.1111/j.1467-9280.1997.tb00453.x

2 citations, pages 76, 146

Spinelli, D., Luca, M. D., Filippo, G. D., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading : Role of reading experience and reading deficit in italian readers. *Developmental Neuropsychology*, *27*(2), 217–235. doi: 10.1207/s15326942dn2702_2

cité page 141

Steacy, L. M., Compton, D. L., Petscher, Y., Elliott, J. D., Smith, K., Rueckl, J. G., ... Pugh, K. R. (2018). Development and prediction of context-dependent vowel pronunciation in elementary readers. *Scientific Studies of Reading*, *23*(1), 49–63. doi: 10.1080/10888438.2018.1466303

cité page 143

Sullivan, K. J., Shadish, W. R., & Steiner, P. M. (2015). An introduction to modeling longitudinal data with generalized additive models : Applications to single-case designs. *Psychological Methods*, *20*(1), 26–42. doi: 10.1037/met0000020

cité page 80

Taft, M. (1991). *Reading and the mental lexicon.* Hillsdale, N.J : Lawrence Erlbaum Associates.

cité page 138

The Economist. (2017). How do you pronounce "GIF"? ([Online; posted 29-June-2017] https://www.economist.com/graphic-detail/2017/06/29/how-do-you-pronounce-gif?)

cité page 93

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, *9*(1), 40–50. doi: 10.3758/BF03213026

7 citations, pages v, 48, 53, 54, 56, 57 & 58

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of english orthography. *Journal of Experimental Psychology : General*, *124*(2), 107–136. doi: 10.1037/0096-3445.124.2.107

cité page 76

Valdois, S., Carbonnel, S., Juphard, A., Baciu, M., Ans, B., Peyrin, C., & Segebarth, C. (2006). Polysyllabic pseudo-word processing in reading and lexical decision : Converging evidence from behavioral data, connectionist simulations and functional MRI. *Brain Research*, *1085*(1), 149–162. doi: 10.1016/j.brainres.2006.02.049

2 citations, pages 35, 139

Valdois, S., Roulin, J.-L., & Bosse, M. L. (2019). Visual attention modulates reading acquisition. *Vision Research*, *165*, 152–161. doi: 10.1016/j.visres.2019.10.011

cité page 139

Veldre, A., Yu, L., Andrews, S., & Reichle, E. D. (2020). Towards a complete model of reading : Simulating lexical decision, word naming, and sentence reading with über-reader. In *42nd Annual Conference of the Cognitive Science Society (CogSci 2020)* (p. 151-157). Toronto, ON.

2 citations, pages 4, 138

Vidyasagar, T. R. (2013). Reading into neuronal oscillations in the visual system : implications for developmental dyslexia. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00811

cité page 4

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia : A deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, *14*(2), 57–63. doi: 10.1016/j.tics.2009.12.003

cité page 143

Waechter, S., Besner, D., & Stolz, J. A. (2011). Basic processes in reading : Spatial attention as a necessary preliminary to orthographic and semantic processing. *Visual Cognition*, *19*(2), 171–202. doi: 10.1080/13506285.2010.517228

cité page 139

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*(1), 140–159. doi: 10.1016/j.jml.2007.04.006

cité page 147

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology Section A*, *50*(2), 439–456. doi: 10.1080/713755710

3 citations, pages 35, 37 & 72

Whitney, C. (2001). How the brain encodes the order of letters in a printed word : The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, *8*(2), 221–243. doi: 10.3758/bf03196158

3 citations, pages 4, 132 & 137

Whitney, C. (2011). Location, location, location : How it affects the neighborhood (effect). *Brain and Language*, *118*(3), 90–104. doi: 10.1016/j.bandl.2011.03.001

cité page 141

Whitney, C. (2018). When serial letter processing implies a facilitative length effect. *Language, Cognition and Neuroscience*, *33*(5), 659–664. doi: 10.1080/23273798.2017.1404115

cité page 141

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, *76*(1), 1–15. doi: 10.1037/h0026823

cité page 21

Widrow, B., & Hoff, M. E. (1960). Adaptive Switching Circuits. In *Institute of Radio Engineers, Western Electronic Show and Convention record, Part 4* (pp. 96–104). New York, NY : Institute of Radio Engineers.

cité page 16

Wood, S. (2006). *Generalized Additive Models : An introduction with R*. Boca Raton, FL : Chapman & Hall/CRC Press.

2 citations, pages 80, 85

Wood, S. (2019). `mgcv` : Mixed GAM Computation Vehicle with Automatic Smoothness Estimation [Manuel de logiciel]. Consulté sur `https://cran.r-project.org/web/packages/mgcv/index.html` (version 1.8-31)

cité page 80

Yap, M. J., & Balota, D. A. (2015). Visual word recognition. In *The oxford handbook of reading* (p. 26-43). New York, NY : Oxford University Press. doi: 10.1093/oxfordhb/9780199324576.013.4

cité page 3

Yoncheva, Y. N., Maurer, U., Zevin, J. D., & McCandliss, B. D. (2013). Effects of rhyme and spelling patterns on auditory word ERPs depend on selective attention to phonology. *Brain and Language*, *124*(3), 238–243. doi: 10.1016/j.bandl.2012.12.013

cité page 142

Yoncheva, Y. N., Zevin, J. D., Maurer, U., & McCandliss, B. D. (2009). Auditory selective attention to speech modulates activity in the visual word form area. *Cerebral Cortex*, *20*(3), 622–632. doi: 10.1093/cercor/bhp129

cité page 142

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming : A comparison of current models. *Journal of Memory and Language*, *54*(2), 145–160. doi: 10.1016/j.jml.2005.08.002

cité page 23

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages : A psycholinguistic Grain Size Theory. *Psychological Bulletin*, *131*(1), 3–29. doi: 10.1037/0033-2909.131.1.3

2 citations, pages 144, 145

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition : Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*(4), 779–793. doi: 10.1016/S0749-596X(03)00006-8

cité page 71

Ziegler, J. C., & Perry, C. (1998). No more problems in coltheart's neighborhood : resolving neighborhood conflicts in the lexical decision task. *Cognition*, *68*(2), B53–B62. doi: 10.1016/s0010-0277(98)00047-x

cité page 15

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. doi: 10.1098/rstb.2012.0397

2 citations, pages 38, 148

Ziegler, J. C., Perry, C., & Zorzi, M. (2020). Learning to Read and Dyslexia : From Theory to Intervention Through Personalized Computational Models. *Current Directions in Psychological Science*, *29*(3), 293–300. doi: 10.1177/0963721420915873

3 citations, pages 4, 38 & 148

Zorzi, M. (2000). Serial processing in reading aloud : No challenge for a parallel model. *Journal of Experimental Psychology : Human Perception and Performance*, *26*(2), 847–856. doi: 10.1037/0096-1523.26.2.847

cité page 36

Zorzi, M., Houghton, G., & Butterworth, B. (1998a). The development of spelling-sound relationships in a model of phonological reading. *Language and Cognitive Processes*, *13*(2-3), 337–371. doi: 10.1080/016909698386555

cité page 15

Zorzi, M., Houghton, G., & Butterworth, B. (1998b). Two routes or one in reading aloud? a connectionist dual-process model. *Journal of Experimental Psychology : Human Perception and*

Performance, *24*(4), 1131–1161. doi: 10.1037/0096-1523.24.4.1131

cité page 16

Zoubrinetzky, R., Bielle, F., & Valdois, S. (2014). New insights on developmental dyslexia subtypes : Heterogeneity of mixed reading profiles. *PLoS ONE*, *9*(6), e99337. doi: 10.1371/journal.pone.0099337

cité page 35

ANNEXE

A Méthodologie de la Programmation Bayésienne

La modélisation Bayésienne algorithmique permet de définir des modèles probabilistes structurés et hiérarchisés d'agents cognitifs. Cette méthodologie, héritée directement de la programmation bayésienne (Bessière et al., 2008 ; Lebeltel et al., 2004), permet de définir des programmes robotiques avec des modèles probabilistes. Ce cadre offre un langage mathématique pour exprimer et manipuler des connaissances, dans des modèles arbitrairement complexes. Cette méthodologie a été appliquée pour la programmation d'agents cognitifs artificiels (Colas, Diard, & Bessiere, 2010). On peut voir la modélisation bayésienne algorithmique comme une description, au niveau algorithmique de Marr (1982), d'agents cognitifs naturels (Diard, 2015).

Dans cette méthodologie, les processus de traitement de l'information d'un sujet cognitif sont modélisés par l'intermédiaire de programmes probabilistes qui permettent de : 1) décrire, de manière probabiliste, l'ensemble des connaissances que le système étudié possède ; et (2) de pouvoir décrire l'utilisation de ces connaissances pour le raisonnement et pour générer un comportement à l'aide de l'application du calcul probabiliste.

Cette annexe donne un bref descriptif de la méthodologie de la modélisation Bayésienne algorithmique et une petite application dans le champ de la reconnaissance de mots à partir de lettres. L'objectif est donc d'illustrer sur un exemple simple comment est formulé et appliqué ce formalisme de modélisation.

NOTE :

Le contenu de cette annexe est un extrait du Chapitre 3 de la thèse de Thierry Phénix (2018).

1 Méthodologie de la Programmation Bayésienne

Le cadre de modélisation bayésienne que nous suivons repose particulièrement sur les principes de la programmation bayésienne des robots (Bessière et al., 2013 ; Lebeltel et al., 2004). Cette programmation est elle-même basée sur une interprétation subjectiviste des probabilités (Jaynes, 2003).

1.1 Concepts de base

Variable probabiliste Une variable probabiliste est l'association entre un symbole (le nom de la variable) et un ensemble de valeurs possibles pour cette variable (un domaine de définition). Par exemple, nous définissons la variable X sur l'espace discret $\mathscr{D}_X$. Ainsi, chaque expression $[X = x]$ peut être considérée comme une proposition logique avec x qui est une valeur du domaine $\mathscr{D}_X$. Le cadre bayésien subjectiviste permet d'attribuer une valeur de probabilité à chaque proposition afin d'exprimer le degré de certitude à l'égard de sa vérité. On note :

$$P([X = x]) = p_x, \text{ tel que } p_x \in [0,1] .$$

Règles de Calcul Les distributions de probabilité suivent deux règles principales : la règle de la somme et la règle du produit. La règle de la somme, aussi appelée la règle de normalisation, indique que la somme des probabilités des cas possibles est 1. On note :

$$\sum_{x \in \mathscr{D}_X} P([X = x]) = 1 .$$

La règle du produit permet de décomposer une distribution conjointe en un produit de termes :

$$P(X\ Y) = P(X \mid Y)\ P(Y) = P(Y \mid X)\ P(X) .$$

La règle de Bayes n'est qu'un corollaire de ce qui précède :

$$P(X \mid Y) = \frac{P(Y \mid X)\ P(X)}{P(Y)} .$$

1.2 Programme Bayésien

La programmation bayésienne est une méthodologie pour définir de manière formelle les programmes bayésiens. Un programme est construit à partir d'une description et d'une question (voir Figure A.1). Dans la description, on définit les connaissances préalables dans le modèle. Mathématiquement, cela se traduit par la déclaration des variables probabilistes et de leurs espaces de définition. Dans la phase de décomposition, nous introduisons les *hypothèses d'indépendance conditionnelle* pour simplifier l'expression de la conjointe en un produit de distributions plus simples. Ces hypothèses permettent également de spécifier la structure de dépendance entre les variables probabilistes.

Cette première phase de description se termine par la définition des formes paramétriques des distributions qui sont exprimées explicitement ou calculées à partir d'une autre programme

bayésien. Il est tout-à-fait possible d'imaginer un programme bayésien dont la définition requiert l'appel à des sous-programmes bayésiens.

Une fois le modèle spécifié, il est possible de « faire apprendre » au modèle des connaissances supplémentaires, en spécifiant les valeurs de paramètres libres restants, à partir de données observées. L'objectif dans cette première partie déclarative du modèle est que tous les termes de la décomposition soient entièrement définis et, par conséquent, que la distribution de probabilité conjointe soit également pleinement définie.

La question probabiliste correspond à la distribution de probabilité qui nous intéresse dans le problème. Mathématiquement, cela correspond à calculer, à l'aide d'une inférence bayésienne, une distribution de probabilité conditionnelle d'un ensemble de variables d'intérêt ou recherchées (*Cherchées*) sachant un ensemble de variables connues (*Connues*) et un ensemble de variables libres (*Libres*).

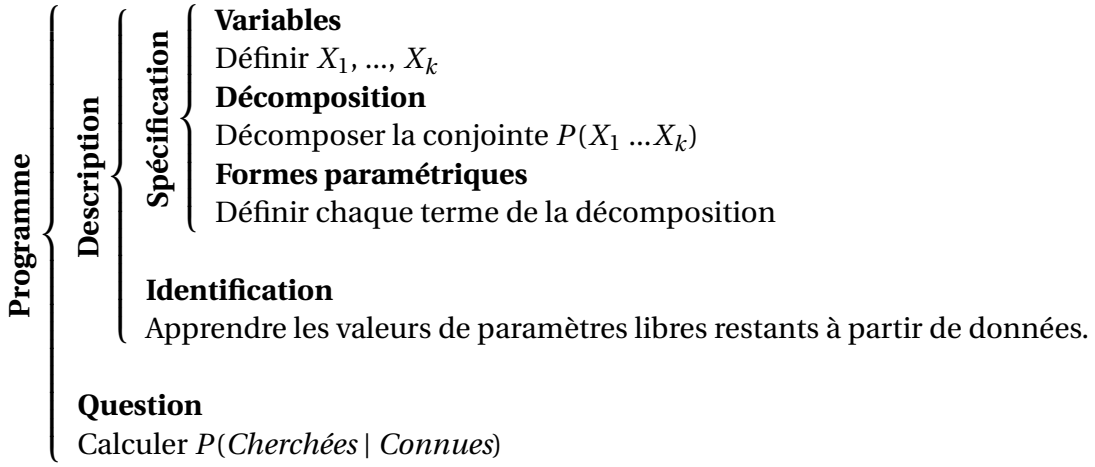


FIGURE A.1 – **Structure générale d'un programme bayésien adapté de Lebeltel et al. (2004).**

2 Exemple simple de reconnaissance de mots

Dans cette section, nous utiliserons un exemple simple de reconnaissance de mots de façon à illustrer et montrer dans un cadre moins abstrait le formalisme de modélisation adopté. L'exemple sera présenté en étapes, telles qu'elles sont introduites dans la programmation bayesienne.

2.1 Variables du modèle

Dans cet exemple, les domaines des variables probabilistes que nous manipulons sont discrets et finis (par exemple, l'ensemble des lettres de l'alphabet). Cette notion s'oppose à la notion de variable continue pour lesquelles le domaine de définition est représenté par un ensemble infini, ou un intervalle réel, donc non dénombrable (par exemple n'importe quelle valeur réelle entre 0,25 et 0,75). Dans notre exemple, les mots sont composés de trois lettres. Nous avons ainsi quatre variables probabilistes.

- La variable W représente l'identité du mot. Son domaine est l'ensemble des mots possibles : dans notre exemple, considérons qu'il n'y a que trois mots possibles, et $W = \{w_1, w_2, w_3\}$.
- Trois variables L_1, L_2 et L_3 qui chacune représente uen des lettres du mot. chacune de ces variables a un domaine qui décrit les lettres possibles de l'alphabet. Par la suite, nous utiliserons un domaine, $\mathscr{D}_L = \{a, b, c, \cdots, z, \#\}$, avec une valeur possible pour chaque lettre de 'a' à 'z' et un symbole particulier, '#', pour représenter une lettre illisible ou manquante.

Remarquons la notation utilisée ici et par la suite. Lorsqu'elles sont indicées, les variables sont « localisées spatialement » : ainsi, dans notre exemple, L_1 est la première lettre du mot, L_2 est la deuxième lettre du mot, et L_3 est la dernière lettre du mot. Plus tard, nous décrirons l'évolution temporelle des variables, et les annoterons par des exposants : ainsi, L_2^t représentera la lettre en seconde position au temps t. Introduisons ici quelques raccourcis de notation :

- $X^{1:T}$ est la conjonction des variables X^t pour les valeurs de temps t comprises entre 1 et T, c'est-à-dire $X^{1:T} = (X^1, X^2, \cdots, X^T)$;
- ce même raccourci de notation s'applique, pour un pas de temps t donné, aux positions spatiales. Ainsi $X^t_{1:N}$ est la conjonction des variables X^t_n, pour les positions n allant de 1 à N, c'est-à-dire $X^t_{1:N} = (X^t_1, X^t_2, \cdots, X^t_N)$;
- ce raccourci se combine pour lister simultanément des variables sur plusieurs instants de temps et plusieurs positions, par exemple :

$$X^{1:T}_{1:N} = (X^1_1, X^2_1, \cdots, X^T_1, X^1_2, X^2_2, \cdots, X^T_2, \cdots, X^1_N, X^2_N, \cdots, X^T_N) ;$$

- nous notons $[X^t_n = x^t_n]$ l'assignation de la valeur x^t_n à la variable probabiliste X^t_n , et, lorsque celle-ci est implicite, nous l'abrégeons en x^t_n.

2.2 Distribution conjointe et décomposition

La seconde étape de la modélisation bayésienne algorithmique consiste à définir la distribution de probabilité conjointe du modèle, c'est-à-dire la distribution de probabilité sur l'espace conjoint formé par toutes les variables listées précédemment. Ainsi, dans notre exemple, c'est la distribution de probabilité conjointe $P(W\ L_1\ L_2\ L_3)$ que nous devons définir. Les distributions de probabilité conjointe portent souvent sur des espaces trop compliqués pour être définies directement. Ainsi, en appliquant la règle du produit, nous décomposons la distribution de probabilité conjointe en un produit de termes. Dans notre exemple, nous choisissons de définir :

$$P(W\ L_1\ L_2\ L_3) = P(W)\ P(L_1 \mid W)\ P(L_2 \mid L1\ W)\ P(L_3 \mid L2\ L1\ W)\ .$$

À cette étape, nous introduisons des *hypothèses d'indépendance conditionnelle* pour simplifier certains des termes de ce produit. On dit que deux variables A et B sont indépendantes, conditionnellement à une variable C, lorsque la probabilité de A connaissant la valeur de la variable C ne

dépend pas des valeurs de la variable B, ce qui s'écrit : $P(A \mid B\ C) = P(A \mid C)$. Dans notre exemple, nous faisons l'hypothèse que la principale source d'information sur les lettres du mot, est l'identité w du mot. En d'autres termes, si on suppose le mot w connu, on peut prédire chacune de ses lettres, indépendamment de la connaissance des autres lettres. Ainsi, le terme $P(L_2 \mid L_1\ W)$ se simplifie en $P(L_2 \mid W)$: on fait l'hypothèse que L_2 est indépendant de L_1, conditionnellement à la connaissance de W. Une hypothèse similaire simplifie $P(L_3 \mid L_2\ L_1\ W)$ en $P(L_3 \mid W)$. Nous obtenons une définition « sous hypothèse » de la distribution de probabilité conjointe :

$$P(W\ L_1\ L_2\ L_3) = P(W)\ P(L_1 \mid W)\ P(L_2 \mid W)\ P(L_3 \mid W)\ . \tag{A.1}$$

On vérifie aisément que les hypothèses d'indépendance conditionnelle permettent de réduire la complexité du modèle : le terme $P(L_3 \mid L2\ L_1\ W)$ a été approximé par $P(L_3 \mid W)$, qui est une forme mathématique plus simple. En effet, dans $P(L_3 \mid L_2\ L_1\ W)$, il y a autant de distributions de probabilités sur L_3 qu'il y a de combinaisons possibles de valeurs de L_2, L_1 et W (dans notre exemple, 27*27*3), alors que dans $P(L_3 \mid W)$, il y a seulement trois distributions de probabilités sur L_3.

2.2.1 Formes paramétriques

Pour rendre opérationnelle la définition du modèle, il faut donner un moyen de calculer chacun des termes qui apparaissent dans la décomposition de la distribution conjointe. Nous associons à chacun de ces termes une forme paramétrique, c'est-à-dire une famille paramétrée de distributions de probabilités. Par exemple, nous choisissons que tel terme est une distribution gaussienne de probabilités (de paramètres libres μ, σ), que tel terme est une table de probabilités conditionnelles (avec autant de paramètres que de cases dans les tables), que tel terme est une distribution uniforme de probabilité (sans paramètre libre), etc. Dans notre exemple, nous devons choisir quatre formes paramétriques, une pour chacun des termes de l'Équation (A.1). Nous choisissons par exemple de définir le terme $P(W)$ par une distribution uniforme : ainsi,

$$P([W = w_1]) = P([W = w_2]) = P([W = w_3]) = \frac{1}{3}\ .$$

Considérons ensuite le terme $P(L_1 \mid W)$: pour un mot w, il n'y a qu'une lettre possible en première position, et cette lettre est connue sans aucune incertitude. Ainsi, $P(L_1 \mid [W = w])$ est une distribution de probabilité Dirac, c'est-à-dire que la probabilité est 1 sur la bonne lettre, et 0 partout ailleurs.

Dans notre exemple, imaginons que le mot w_1 est le mot « bal », alors :

$$P([L1 = \text{`}b\text{'}] \mid [W = w]) = 1\ ,$$

et

$$\forall l \neq \text{`}b\text{'}, P([L_1 = l] \mid [W = w]) = 0\ .$$

Définissons que w_2 et w_3 sont les mots « bol » et « lac » : tous les termes $P(L_1 \mid W)$, $P(L_2 \mid W)$ et $P(L_3 \mid W)$ sont facilement définis de manière similaire.

Avec ces choix de formes paramétriques pour notre exemple, il ne reste aucun paramètre libre dans la définition de la distribution de probabilité conjointe. Dans le cas général, cependant, il peut rester à cette étape des paramètres libres, et on complète la définition du modèle en indiquant la manière de les calculer à partir, le plus souvent, d'une base de données expérimentale. C'est une étape d'apprentissage de paramètres ou de calibrage, dont nous nous passons dans notre exemple.

Notre modèle introductif est donc entièrement spécifié, et sa distribution de probabilité conjointe est calculable en tout point. La phase déclarative de la méthodologie, c'est-à-dire la phase dans laquelle on définit sous forme probabiliste l'ensemble des connaissances du modèle, est alors terminée.

2.2.2 Inférence bayésienne

Le modèle étant entièrement spécifié, nous entrons maintenant dans une phase procédurale, c'est-à-dire que nous allons maintenant mettre en oeuvre le modèle, et les connaissances qu'il contient, pour simuler des tâches cognitives. Pour simuler une tâche, nous définissons une question probabiliste, c'est-à-dire la distribution de probabilité d'intérêt qui modélise cette tâche. Dans notre exemple, nous considérons les tâches de reconnaissance du mot, au vu d'une ou des lettres qui le composent. Par exemple, si la tâche est de reconnaître le mot alors qu'on sait toutes les lettres, la question probabiliste est $P(W \mid L_1\ L_2\ L_3)$. Si on ne connaît que la première lettre, la question est $P(W \mid L_1)$; si on connaît la première et la dernière lettre, mais pas la seconde, la question est $P(W \mid L_1\ L_3)$, etc.

Remarquons que ces questions sont des distributions de probabilités qui ne sont pas contenues dans la définition de la distribution conjointe du modèle (voir Équation (A.1)). Il faut donc les calculer, par inférence bayésienne, c'est-à-dire appliquer les règles du calcul probabiliste pour, à partir de la distribution conjointe, obtenir une expression qui réponde à la question probabiliste (un théorème montre que, la conjointe étant définie, on peut ainsi répondre à n'importe quelle question probabiliste qui porte sur les variables de la conjointe). Cette étape d'inférence ne requiert plus l'intervention du modélisateur, c'est-à-dire que, la question et la conjointe étant posées, la réponse est unique et mathématiquement entièrement déterminée (et trouvable automatiquement).

Considérons par exemple la question $P(W \mid L_1\ L_2\ L_3)$. L'inférence bayésienne donne :

$$\begin{aligned} P(W \mid L_1\ L_2\ L_3) &= \frac{P(W\ L_1\ L_2\ L_3)}{P(L_1\ L_2\ L_3)} \\ &= \frac{1}{Z}\ P(W\ L_1\ L_2\ L_3)\ . \end{aligned}$$

En effet, le terme au dénominateur, $P(L_1\ L_2\ L_3)$ est une constante, les valeurs de L_1, L_2 et L_3 étant connues. On peut ainsi noter $Z = P(L_1\ L_2\ L_3)$. Il est souvent plus aisé de ne pas calculer cette constante Z explicitement, et de l'obtenir en fin de calcul par renormalisation de la distribution. On peut donc procéder qu'en calculer « à un facteur près », c'est-à-dire en ne s'intéressant qu'au numérateur : en notant $\propto$ la relation « est proportionnel à », on réécrit :

$$P(W \mid L_1\ L_2\ L_3) \propto P(W\ L_1\ L_2\ L_3)\ .$$

Ensuite, on remarque que le calcul n'implique que la distribution conjointe, et on la remplace par sa définition :

$$P(W \mid L_1\ L_2\ L_3) \propto P(W)\ P(L_1 \mid W) P(L_2 \mid W) P(L_3 \mid W)\ .$$

Ici, tous les termes de la distribution conjointe interviennent, mais pour certaines questions, il est possible de simplifier. Par exemple, considérons la question de la reconnaissance du mot ne connaissant que sa première lettre :

$$\begin{aligned}
P(W \mid L_1) &= \frac{\sum_{L_2,L_3} P(W\ L_1\ L_2\ L_3)}{P(L_1)} \\
&\propto \sum_{L_2,L_3} P(W\ L_1\ L_2\ L_3) \\
&\propto \sum_{L_2,L_3} P(W)\ P(L_1 \mid W)\ P(L_2 \mid W)\ P(L_3 \mid W) \\
&\propto P(W)\ P(L_1 \mid W) \sum_{L_2} P(L_2 \mid W) \sum_{L_3} P(L_2 \mid W) \\
&\propto P(W)\ P(L_1 \mid W)\ .
\end{aligned}$$

Remarquons que l'on peut interpréter ces calculs comme des processus d'inversion probabiliste de connaissance : le modèle contient des termes qui « prédisent » les lettres pour chaque mot, et, dans ces inférences, la règle de Bayes permet de calculer la relation inverse, dans laquelle, une ou plusieurs lettres étant connues, on « reconnaît » le mot. Pour conclure notre exemple, imaginons que la première lettre est un '*b*', et calculons $P(W \mid [L_1 = \text{'}b\text{'}])$. Nous obtenons une distribution dans laquelle les mots w_1 (« bal ») et w_2 (« bol ») sont équiprobables, de probabilité 0,5, et le mot w_3 (« lac ») de probabilité de 0.

Title — Bayesian modeling of reading

Abstract — Most computational models of reading adopt a two-route architecture that assumes that the knowledge involved in serial processing differs fundamentally from the knowledge involved when processing is parallel. Thus, the reading of new words or pseudowords involves a system of grapheme-phoneme correspondences via the sublexical channel that operates in a serial way, whereas the reading of known words relies on the activation of lexical knowledge of the lexical route which is parallel. These models also suppose that the grapheme-phoneme conversion system is explicit, independently stored and necessary for the reading of pseudowords. They assume, moreover, that this conversion system is preceded by a system of segmentation of the word into subunits to be converted independently.

However, other reading models assume that parallel and serial processing involve the same type of knowledge. These two classes of models agree on the explanation of parallel processing in reading but do not agree in the description of serial processing. They therefore differ on their explanation of length effects, *i.e.*, the observation of longer treatment durations for long stimuli. Indeed, dual-route models can only interpret these length effects through serial decoding of the sublexical channel. Moreover, even if all models agree on the role of visual attention in serial processing, the corresponding mechanisms are not well described, especially mathematically, in the literature.

The aim of this thesis is to evaluate the hypothesis that a treatment involving only lexical knowledge about words is able to account for sublexical relationships between orthographic and phonological units, to simulate length effects in different types of tasks and to perform sublexical segmentation of new words.

For this purpose, we propose a new probabilistic computational model of reading called "BRAID-Phon". This model is an extension of the computational word recognition model of the "BRAID" model by adding a phonological knowledge sub-model. We use the BRAID-Phon model to study the plausibility of a system based on a "single-route" architecture, *i.e.*, including only orthographic and phonological lexical knowledge to perform a reading simulation. We show the ability of BRAID-Phon to account for length effects on words in three types of tasks (reading, lexical decision and progressive demasking) and study the role of implemented visual attention mechanisms on these effects. Finally, we illustrate the need for an attention-controlled segment-based processing to perform pseudoword reading.

Keywords — Computational modeling, Bayesian modeling, expert reading-aloud, visual attention, lexical decision.

Titre — Modélisation bayésienne de la lecture

Résumé — Les modèles computationnels de la lecture adoptent pour la plupart une architecture double-voie qui suppose que les connaissances impliquées dans les traitements sériels diffèrent fondamentalement des connaissances mises en jeu lorsque le traitement est parallèle. Ainsi, la lecture des mots nouveaux ou pseudo-mots implique un système de correspondances graphème-phonème via la voie sous-lexicale qui opère de façon sérielle alors que la lecture des mots connus repose sur l'activation de connaissances lexicales via la voie lexicale qui est parallèle. Ces modèles postulent également que le système de conversion graphème-phonème est explicite, stocké indépendamment et indispensable à la lecture des pseudo-mots. Ils supposent, de plus, que ce système de conversion est précédé d'un système de segmentation du mot en sous-unités à convertir indépendamment.

Cependant, d'autres modèles de lecture postulent que les traitements parallèles et sériels font intervenir le même type de connaissances. Ces deux classes de modèles s'accordent pour expliquer les traitements parallèles en lecture mais s'opposent quant à la description des traitements sériels. Ils diffèrent donc sur leur explication des effets de longueur, c'est-à-dire l'observation de durées de traitement plus longs pour les stimuli longs. En effet, les modèles double-voie ne peuvent interpréter ces effets de longueur que par le biais du décodage sériel de la voie sous-lexicale. De plus, même si tous les modèles s'accordent à propos du rôle de l'attention visuelle dans le traitement sériel, les mécanismes correspondants sont assez peu décrits, notamment mathématiquement, dans la littérature.

Cette thèse a pour but d'évaluer l'hypothèse selon laquelle un traitement ne mettant en jeu que les connaissances lexicales apprises sur les mots est en mesure de rendre compte des relations sous-lexicales entre unités orthographiques et unités phonologiques, de simuler les effets de longueur dans différents types de tâches et d'opérer une segmentation sous-lexicale des mots nouveaux.

Pour cela, nous proposons un nouveau modèle computationnel probabiliste de la lecture nommé « BRAID-Phon ». Ce modèle est une extension du modèle computationnel de reconnaissance de mots du modèle « BRAID » par ajout d'un sous-modèle de connaissances phonologiques. Nous utilisons le modèle BRAID-Phon pour étudier la plausibilité d'un système basé sur une architecture « simple-voie », c'est-à-dire incluant uniquement des connaissances lexicales orthographiques et phonologiques pour effectuer une simulation de la lecture. Nous montrons la capacité de BRAID-Phon à rendre compte des effets de longueur sur les mots dans trois types de tâches (lecture, décision lexicale et *démasquage progressif*) et étudions le rôle des mécanismes implémentés d'attention visuelle sur ces effets. Enfin, nous illustrons la nécessité d'un processus de lecture par segments, contrôlé par l'attention, pour effectuer la lecture de pseudo-mots.

Mots clés — Modèle computationnel, modélisation bayésienne, lecture experte, attention visuelle, décision lexicale.