Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

SCIENCES HUMAINES & SOCIALES
Perspectives Européennes

ÉCOLE DOCTORALE 519

EA 4440 LPC

THÈSE présentée par :

Marie-Ange LECERF

soutenue le : **11 Juillet 2023**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Psychologie

An investigation of three sublexical variables and their influence on L2 visual word recognition.

THÈSE dirigée par :

Pr DEMONT Elisabeth	Supervisor, université de Strasbourg
Dr COMMISSAIRE Eva	Co-Supervisor, université de Strasbourg
Pr CASALIS Séverine	Co-supervisor, université de Lille

RAPPORTEURS :

Dr VAN HEUVEN Walter	External Examiner, University of Nottingham
Dr APARICIO Xavier	External Examiner, EPHE, Université Paris-Est Créteil

AUTRES MEMBRES DU JURY :

Pr LETE Bernard	External Reviewer, Université Lumière Lyon 2
Pr PONCELET Martine	External Reviewer, Université de Liège

“It’s still magic even if you know how it’s done.”

Terry Pratchett, A Hat Full of sky

Acknowledgements

The story behind the document you are holding begins 5 years ago when I received a call from Séverine proposing me a fully funded PhD position with the LPC laboratory from Strasbourg, “on a topic relating to L2-acquisition and bilingualism”. That was, to me, a dream proposition… Except that the application for the grant had to be completed by the end of the day … and that I was currently moving out from my apartment. That’s when I met Eva, Anne-Sophie and Elisabeth via e-mail. They took over most of the work that day, and managed to apply on time. A week or so later, I met them in person when they prepared me (which required a lot of patience considering I didn’t know anything about the project previously to that) for the oral presentation following which we obtained the grant! That was the start of a long adventure in which they all continued to play an important role. So, as I am writing this PhD thesis, I am constantly reminded that without you I wouldn’t have had this incredible opportunity and I want to thank you very much, for the opportunity but also for your guidance, your advice and your support.

Talking of support, I want to thank my family and my friends. I want to thank them for supporting me even though they did not understand why any sane human being would inflict a PhD upon themselves. First, I want to thank my mom, who patiently listened to most of my evening rants (even though most of them were about statistics). I want to thank my sister that always took my side no matter the battle I was imagining myself to be leading. I want to thank Perrine who encouraged me to take time away from this thesis and kept me (mostly) sane. I also want to thank Gary(-sha) that single-handedly got me through the first lockdown. And finally I want to dedicate a special mention to Yohann who took it upon himself to ask THE sensitive “So, how is your PhD going?“ at EVERY family gathering. There had to be one person, and I’m glad it was you. Merci à vous tous!

Next, I wish to thank all the young researchers with whom I worked along the years, be it from Lille or Strasbourg, from upstairs or downstairs, you all played a role in shaping the researcher I hope to become. Thanks to Claire for our daily venting sessions and for being the best front-desk-neighbour I could hope to have. Thanks to Jad, Marine, Thibaut, Cécile, Séléna and Emery for the many late-Friday conversations. Thanks to Marc, Florian, Camille and Fanny for their long-distance support. And thanks Gary (the human one) for your guidance.

I also want to extend my thanks to the entirety of the LPC lab for their warm welcome (with a special mention to Florence who took part to my PhD committee).

Thanks to every middle-school teacher (and particularly to Madame Vogel) that agreed to work with me even though the data collected could not be included in this document.

Finally, I want to thank the members of the jury for accepting to review this work. I look forward to meeting you all!

So that’s it, but before we get down to it, I want to thank every single one of you one last time. As you can see, I may have written those (so many) pages on my own but you all made your contributions, and for that I am so very grateful.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

RÉSUMÉ EN FRANÇAIS

Le contexte de mondialisation dans lequel nous baignons a rendu nécessaire l'apprentissage d'une deuxième langue. En effet, nous sommes confrontés quotidiennement aux langues étrangères, que ce soit dans des contextes formels comme un entretien d'embauche, ou dans des contextes moins formels comme sur les panneaux de publicité. La maîtrise d'une langue seconde est alors devenue une obligation pour une bonne qualité d'intégration sociale.

Toutefois, l'apprentissage d'une langue seconde représente un défi aux multiples facettes pour chaque individu qui l'entreprend. En effet celui-ci doit acquérir du nouveau vocabulaire mais aussi se familiariser avec de nouvelles règles de syntaxe et de grammaire qui doivent alors co-exister avec leurs équivalents en langue maternelle. Pour une majorité de la population cet apprentissage est d'autant plus ardu qu'il est entrepris tardivement, aux alentours du début de l'adolescence, ce qui implique une asynchronie entre l'intégration d'un concept et l'apprentissage du mot L2 qui lui est associé. Cet apprentissage prend également place dans des conditions peu écologiques, i.e., une salle de classe. Dans cette salle de classe, les enseignants sont alors confrontés à un groupe d'individus auxquels ils devront inculquer les rudiments d'une langue qui leur est étrangère en mettant à profit les quelques heures qui leur sont accordées au cours d'une semaine. Pour relever ce défi, ces professeurs mettent donc à contribution les forces de leurs élèves et s'appuient sur leur capacité de lecture en leur présentant une grande quantité de matériel visuel (i.e., manuels, listes de vocabulaire, sous-titres) dans lesquels les mots de la langue seconde sont présentés sous leurs formes écrites, ce qui engendre une asymétrie entre l'exposition aux formes écrites et l'exposition aux formes orales des mots de langue seconde.

Il est donc loin d'être surprenant qu'une littérature riche et ample se soit développée autour de la reconnaissance visuelle de mots en L2, donnant naissance à plusieurs modèles rendant compte des mécanismes impliqués dans un tel processus. Les modèles les plus cités sur la question, le modèle BIA (van Heuven et al., 1998) et le modèle BIA+ (Dijkstra & van Heuven, 2002; rejoint récemment par le modèle MULTILINK; Dijkstra et al., 2019), font l'hypothèse d'un lexique bilingue unique, dans lequel les représentations lexicales de la langue maternelle et de la langue seconde sont intégrées (i.e., le lexique bilingue intégré), et d'un accès non-spécifique à ce lexique (i.e., l'accès lexical non-sélectif), qui impliquerait qu'un individu bilingue ne serait pas capable de sélectionner la langue dans laquelle il souhaiterait interroger son lexique.

Ces hypothèses de lexique intégré et d'accès non-sélectif sont soutenues par les quelques études ayant testé la reconnaissance de mot en fonction de la taille de leur voisinage orthographique interlangue (Dirix et al., 2017; Grossi et al., 2012; Midgley et al., 2008; van Heuven et al., 1998). Ces études ont mis en évidence que les mots possédant de nombreux voisins orthographiques dans l'autre langue connue par un bilingue (PASTE ; 10 voisins orthographiques en français) étaient reconnus moins rapidement que les mots possédant peu de voisins orthographiques dans l'autre langue (FINGER ; 1 seul voisin orthographique en français). Ce qui suggère que les mots ayant de nombreux voisins orthographiques interlangues rencontrent plus de compétition lexicale que les mots ayant peu de voisins orthographiques interlangues (Voir aussi pour des résultats divergents Commissaire, Audusseau, et al., 2019; Mulder et al., 2018).

De même, l'étude des réponses associées à des mots cognates, qui partagent leur forme orthographique ET leur signification à travers les deux langues d'un bilingue (e.g., TRAIN), et celles associées à des homographes interlangues, qui partagent leur forme orthographique MAIS PAS leur signification à travers les langues (e.g., FOUR), a mis en évidence des résultats en accord avec l'existence d'un lexique intégré et d'un accès lexical non-sélectif. En effet, les mots cognates suscitent des temps de réaction plus courts que des mots contrôles (e.g., HOUSE), tandis que les homographes interlangues provoquent des temps de réaction plus longs que les mots contrôles (Dijkstra et al., 1999; Poort & Rodd, 2017; Vanlangendonck et al., 2019). Ces résultats semblent indiquer qu'en présence d'un mot cible écrit, les individus bilingues activent toutes les représentations qui lui sont associées, qu'elles soient L1 ou L2, ce qui peut faciliter le traitement du mot cible dans les rares occasions où ces représentations se recoupent totalement comme dans le cas des cognates, mais peut également ralentir le traitement d'un mot si ce recoupement est associé à un conflit sur l'une des représentations mobilisées comme dans le cas des homographes interlangues.

Ainsi il semble que la reconnaissance visuelle d'un mot, qu'il provienne de la L1 ou de la L2, dépende de l'activation automatique des représentations appartenant à la même langue, i.e., des activations intra-langues, mais aussi de représentations provenant de l'autre langue, i.e., des activations inter-langues. Cette activation automatique se fait sur la base de similarités formelles (e.g., sur la forme orthographique) des différentes représentations avec l'item cible.

Cependant, les modèles BIA et BIA+ (ainsi que le modèle MULTILINK) ont en grande partie ignoré les aspects sublexicaux qui font référence aux unités telles que les graphèmes, les phonèmes ou les bigrammes qui construisent un mot. Cependant, l'exposition aux formes écrites d'une langue seconde implique non seulement l'acquisition de nouvelles représentations lexicales mais également l'apprentissage d'un nouveau système de correspondances graphèmes-phonèmes permettant de traduire un mot écrit en sa forme orale.

Dans le cas des langues européennes qui partagent un même alphabet (ou tout du moins une grande partie de cet alphabet) mais qui diffèrent dans leur utilisation de celui-ci, cet apprentissage sublexical amène l'individu à acquérir de nouveaux patterns orthographiques dans lesquels des lettres avec lesquelles i.e.l est déjà familier vont être réarrangées pour créer des bigrammes (ou trigrammes etc…) illégaux ou très rares dans leur langue maternelle mais qui sont assez courants dans leur langue seconde (e.g., le KN pour un apprenant de l'anglais ayant le français en tant que langue maternelle ou le EU pour un apprenant du français ayant l'anglais en tant que langue maternelle). Ces patterns orthographiques (e.g. le KN pour un lecteur français-anglais ou le EU pour un lecteur anglais-français), ainsi que les rares lettres spécifiques à un langue (e.g., le ß allemand), vont alors constituer ce que l'on appelle des marqueurs orthographiques d'une langue (i.e., orthographic markers) et vont avoir un impact sur le traitement d'un mot par un bilingue, notamment en ce qui concerne l'identification de sa langue d'appartenance (Borragan et al., 2020; Jared et al., 2013; Oganian et al., 2016). De plus, les apprenants sont aussi conduits à mettre à jour leur répertoire de Correspondances Graphèmes-Phonèmes (CGP) pour y inclure les CGP propres à la langue seconde (e.g., le graphème EA réfère au phonème \i\ en anglais) bien que certaines d'entre-elles entrent directement en contradiction avec les CGP acquises lors de l'apprentissage de la lecture dans sa langue maternelle (e.g., le graphème I peut-être associé au phonème \aɪ\ en anglais, mais il est associé au phonème \i\ en français). Bien que peu nombreuses, les quelques études (Brysbaert, 2001; Commissaire, Duncan, et al., 2019; Hevia-Tuero et al., 2021) sur le sujet suggèrent que les lecteurs activent toutes les CGP dont i.e.ls sont familiers, peu importe la langue d'appartenance du mot cible ou de la correspondance. Ainsi, lorsqu'on lui présente un mot tel que HAIR, un lecteur familier du français et de l'anglais activera non-seulement les CGP anglaises \eə\ et \eɪ\ mais aussi la CGP française associée \ɛ\.

Le premier objectif de cette thèse consiste donc à explorer les conséquences de l'apprentissage sublexical décrit ci-dessus. Nous proposons dans un premier temps d'examiner et d'isoler l'impact propre de trois variables sublexicales sur le traitement et la reconnaissance de mots de langue seconde présentés visuellement. Pour cela, notre première étude décrite dans le chapitre III, se compose de trois paires d'expériences (constituées de décisions lexicales et de tâches de lecture à voix haute), qui ont chacune été construites autour d'une variable sublexicale d'intérêt. La première variable, le marquage orthographique fait référence à la présence ou à l'absence de marqueurs orthographiques (i.e., patterns orthographiques spécifiques à une langue) dans les items que nous avons sélectionné. La deuxième variable, la congruence interlangue des CGP, renvoie à la variance ou à la non-variance des CGP entre les deux langues d'un individu bilingue, tandis que la troisième variable, la consistance L2 des CGP, renvoie à la variance ou à la non-variance des CGP dans la langue seconde du même individu.

La première paire de tâches (expériences 1A et 2A), qui se constitue d'une décision lexicale et d'une tâche de lecture à voix haute, a été conçue pour explorer l'impact du marquage orthographique sur le traitement des mots L2 présentés visuellement. Pour cela, les réponses à 80 items (40 mots et 40 pseudomots) marqués, contenant un bigramme spécifique de la L2 (e.g., le OW de TOWN) ont été comparées aux réponses à 80 items non-marqués, ne contenant que des bigrammes partagés entre les deux langues (e.g., HOUSE). Il est important de noter que tous les mots utilisés dans cette tâche sont consistants, c'est-à-dire qu'ils ne contiennent que des rimes qui ne varient pas en anglais[1], le -OWN de TOWN rime avec DOWN et le -OUSE de HOUSE rime avec MOUSE. De la même façon un contrôle de la congruence a été implémenté sur les mots non-marqués qui sont tous incongruents, c'est-à-dire qu'ils ne contiennent que des CGP qui différent de celles fréquemment utilisées en langue maternelle. Ainsi, le OU de HOUSE ne se lit pas \u\ comme en français mais \aʊ\. Ce contrôle avait pour but de s'assurer que tout effet de marquage que l'on pourrait mettre en évidence ne puisse pas être attribué à l'ajout d'une CGP spécifique à la L2 que requièrent les marqueurs orthographiques.

Les résultats récoltés par la décision lexicale n'ont pas permis de mettre en évidence un effet du marquage orthographique. Ceci n'est pas en accord avec l'étude de Commissaire et al., (2019) qui a révélé un effet facilitateur du marquage orthographique dans une décision lexicale similaire alors même que le profil des participants et la fréquence lexicale des mots dans cette étude antérieure sont proches des paramètres de l'expérience décrite ici. Il semble

[1] ou tout du moins contiennent les CGP les plus fréquentes associées à un graphème FIND vs. WIND.

donc que l'absence d'un effet du marquage orthographique dans notre étude s'explique par le contrôle des CGP que nous avons implémenté.

De la même façon, les résultats de la tâche de lecture à voix haute n'ont pas révélé d'effet de marquage orthographique, ce qui contraste avec l'effet facilitateur mis en évidence par l'autre étude sur le sujet ayant utilisé une tâche similaire (Oganian, 2016). Cependant, cette divergence peut être attribuée à une différence méthodologique. En effet, tandis qu'Oganian et al. se sont intéressés à la prononciation de pseudomots inconnus par le participant, nous nous sommes attachés à étudier la prononciation de mots déjà connus par le participant.

Ensemble, ces résultats semblent indiquer que l'utilisation de mots connus par les participants, en association avec un contrôle strict des CGP, et notamment par la comparaison entre mots marqués et mots non-marqués mais incongruents, pourrait avoir masqué l'effet du marquage orthographique. Il semble donc que l'impact du marquage orthographique sur le traitement visuel des mots de L2 serait au moins en partie lié aux aspects phonologiques via l'association des marqueurs avec de nouvelles CGP.

Ensuite, la deuxième paire de tâche (expérience 1B et 2B) cherchait à explorer les conséquences de la congruence interlangue des CGP pour le traitement visuel des mots de langue seconde. Pour cela, les réponses à 80 items congruents (i.e., dont les CGP sont proches en Français et en Anglais ; e.g., FISH) ont été comparées aux réponses à 80 items incongruents (i.e., dont l'une des CGP diverge entre le français et l'anglais ; e.g., le I de TIME). Pour s'assurer du caractère congruent ou incongruent de nos items, nous avons demandé à une trentaine de participant natif du français d'évaluer la proximité des CGP anglaise avec les CGP française pour chacun des mots que nous avons sélectionnés (e.g., on pouvait par exemple leur demander si lire le mot TIME « à la française » donnerait le même résultat que si on le lisait en s'appliquant à utiliser les règles de lecture de l'anglais). Les résultats à ce pré-test coïncidaient avec notre catégorisation initiale. Encore une fois, un contrôle strict des variables sublexicales d'intérêt a été appliqué, puisque dans cet échantillon, aucun item ne contenait de marqueurs orthographiques et tous les mots étaient constitués d'une rime consistante.

Les résultats de la décision lexicale indiquent que bien que les participants y soient sensibles (cf. résultats du pré-test), la congruence interlangue des CGP n'impacte pas la reconnaissance visuelle des mots de langue seconde. En revanche, la tâche de lecture à voix-haute révèle un effet significatif de cette variable sur la prononciation des mots de L2. De façon surprenante, cet effet est inhibiteur, c'est-à-dire que les mots congruents sont traités avec moins d'aisance que les mots incongruents. Ce pattern pourrait s'expliquer par une mauvaise perception des formes phonologiques associées aux mots congruents. Ainsi, les participants que nous avons enrôlés pour cette étude pourraient avoir perçu les phonèmes inclus dans les mots congruents comme étant exactement similaires aux phonèmes français alors qu'il existe une différence objective entre la prononciation française et anglaise d'un même graphème, ce qui les inciterait à l'erreur.

Il est intéressant de mettre en relation nos résultats avec ceux obtenus par Jared & Kroll, (2001) dans une autre tâche de lecture à voix haute dans laquelle les chercheuses s'étaient attachées à comparer les réponses à des mots sans aucun « ennemis » du point de vue des CGP à travers les langues (POKE) aux réponses à des mots dont certains voisins orthographiques interlangues faisaient appel à des CGP différentes de celles utilisées en langue maternelle (BAIT et LAIT). Ces derniers pourraient alors être rapprochés de nos mots incongruents bien qu'ils engendrent de moindres performances lorsqu'ils sont comparés aux mots contrôles. Cependant, les mots contrôles de l'étude de Jared & Kroll., ne peuvent pas être considérés comme congruents à proprement parler, ce qui pourrait expliquer la divergence de résultats.

D'autre part, deux études plus récentes sur l'impact de la congruence des CGP ayant utilisé des tâches de détection de lettre ont mis en évidence un effet facilitateur de la congruence (Commissaire et al., 2014; Hevia-Tuero et al., 2021). A la lumière des résultats présentés ci-dessus, il semble donc que les apprenants d'une langue seconde activent automatiquement l'ensemble des CGP avec lesquelles i.e.ls sont familier.e.s peu importe la langue d'appartenance du mot qui leur est présenté. Ensuite il semblerait que la congruence des CGP n'impacte pas les processus lexicaux impliqués dans la reconnaissance visuelle de mots L2 *per se* (voir Experiment 1b) mais pourrait impacter les traitements sublexicaux impliqués dans la lecture à voix haute (voir Experiment 1b et Jared & Kroll.) ainsi que ceux impliqués dans la détection de lettre (Commissaire et al., 2014; Hevia-Tuero et al., 2021) bien que la nature de cet impact, qu'il soit positif ou négatif, dépende des demandes de la tâche ainsi que des propriétés de la catégorie d'item à laquelle sont comparés les mots incongruents.

Enfin, la dernière paire d'expérience (expérience 1C et 2C) avait pour but d'étudier l'impact de la consistance des CGP de la langue seconde sur le traitement visuel des mots de langue seconde. Pour cela, les réponses à 40 items (20 mots et 20 pseudomots) contenant des rimes consistantes en anglais (e.g., le -EEP de SHEEP et KEEP etc.) ont été comparées aux réponses à 40 items contenant des rimes inconsistantes en anglais (e.g., le -EAR de DEAR et BEAR etc.). De plus, la moitié de ces items contiennent un marqueur orthographique tandis que l’autre moitié n’en contient aucun, ce qui nous a permis d’explorer l’effet du marquage orthographique en fonction de la consistance de nos items.

Les résultats ont mis en évidence un effet facilitateur de la consistance des CGPs dans la tâche de lecture à voix haute, ce qui est en accord avec les résultats d’une étude similaire conduite par Jared & Kroll, (2001) dans laquelle des mots L2 sans « ennemis» en L2 (e.g., BUMP- JUMP) proches de nos mots consistants, engendraient de meilleures performances de lecture à voix haute que des mots L2 avec « ennemis » en L2 (e.g., BEAD – DEAD) proches de nos mots inconsistants. De plus, comme notre étude de la variable congruence des CGP, notre exploration de la variable consistance des CGP suggère que la tâche de lecture à voix haute est plus à même de révéler un effet de variable liées au CGP que la tâche de décision lexicale, puisque l’effet de la consistance, significatif dans la tâche de lecture à voix haute, n’est que tendanciel dans la tâche de décision lexicale. Néanmoins, ces résultats indiquent que les apprenants d’une langue seconde sont sensibles à la consistance des CGPs dans leur langue seconde et confirment qu’i.e.ls activent toutes les CGPs avec lesquelles i.e.ls sont familiers, en accord avec les conclusions de l’étude sur la variable congruence interlangue.

Ces deux dernières expériences (1C et 2C) ont également mis en évidence un effet inhibiteur du marquage orthographique sur les résultats à la décision lexicale et à la tâche de lecture à voix haute. Ces résultats, bien qu’en contradiction avec les études antérieures citées plus haut, ont mis en évidence des implications importantes pour la suite de l’exploration du marquage orthographique. En effet, il semble que l’effet des marqueurs orthographiques dépend de la consistance des CGP du mot qui les contient, ce qui conforte notre hypothèse selon laquelle l’impact du marquage orthographique serait étroitement lié avec les aspects CGP de la reconnaissance visuelle de mots L2.

Plus précisément, dans la tâche de décision lexicale, l'effet du marquage orthographique n'est significatif que lorsqu'on considère uniquement les mots inconsistants (vs. consistants). Ainsi, ces résultats sont facilement réconciliables avec ceux obtenus lors de la première décision lexicale (expérience 1A) puisque celle-ci ne contenait que des mots consistants qui semblent peu adaptés à l'apparition d'un effet du marquage orthographique. Ensemble, ces deux décisions lexicales (expérience 1A et 1C) semblent indiquer que les processus impliqués dans la reconnaissance visuelle de mots L2 doivent être ralentis, par l'introduction de CGP inconsistantes, pour pouvoir observer un impact du marquage orthographique.

En revanche, on retrouve le pattern inverse dans les résultats à la tâche de lecture à voix haute, l'effet du marquage orthographique n'est significatif que si l'on considère uniquement les mots consistants (vs. inconsistants). A première vue, ces résultats contredisent ceux obtenus dans la première tâche de lecture à voix haute (Expérience 2A) qui ne contenait que des mots consistants. Cependant, il est important de préciser que les items de l'expérience 2C étaient exclusivement monosyllabiques tandis que ceux de l'expérience 2A étaient pour la moitié d'entre eux bisyllabiques. Ainsi, la mesure de la consistance que nous avons utilisée (qui prend en compte la rime du mot) pourrait-être moins adaptée aux mots de l'expérience 2A qu'à ceux de l'expérience 2C. La sélection des mots consistants dans la première tâche de lecture à voix haute (Expérience 2A) pourrait avoir été moins précise que celle appliquée pour la dernière tâche de lecture à voix haute (Expérience 2C), expliquant que la première n'ait pas révélé d'effet de marquage orthographique. Néanmoins, les résultats de ces deux expériences (Expérience 2A et 2C) suggèrent que le traitement sublexical impliqué dans la lecture à voix haute doit être encouragé par l'absence (complète) de CGP inconsistantes, pour qu'un effet du marquage orthographique soit observable.

En résumé, les expériences du chapitre III nous ont permis d'observer l'impact distinct du marquage orthographique, de la congruence interlangue des CGP et de la consistance intralangue des CGP de la L2, sur le traitement et la reconnaissance de langue seconde présentés visuellement. Cette première étude met en évidence l'importance d'étudier chaque variable indépendamment des autres pour démêler le nœud de facteurs impliqués dans la reconnaissance visuelle de mots L2, mais aussi pour mieux comprendre la dynamique qui existe entre ces différentes variables. Ainsi cette étude a permis de montrer que des apprenants tardifs d'une L2 sont néanmoins sensibles aux variations des CGP qu'elles soient interlangues ou propres à la L2. Cette sensibilité est par ailleurs accompagnée d'un impact négatif de la congruence et d'un impact positif de la consistance sur les processus de lecture à voix haute. D'autre part, bien que l'effet du marquage orthographique n'ait pas été clairement retrouvé dans cette étude, les résultats obtenus laissent penser que l'impact du marquage orthographique sur les processus de traitement et de reconnaissance visuelle de mots L2 serait très étroitement lié aux propriétés des CGP auxquelles le participant doit faire appel.

De plus, cette première étude souligne l'intérêt pour le recours à une combinaison de tâches complémentaires dans leur examen d'un même phénomène. Elle met tout particulièrement en avant l'intérêt pour l'association d'une tâche de décision lexicale et d'une tâche de lecture à voix haute pour défaire le nœud de variables sublexicales et leur impact sur l'accès lexical bilingue. En effet, les tâches de lecture à voix haute semblent être plus à même de mettre en évidence un effet des variations des CGP que les tâches de décision lexicale, ce qui suggère que ces variables n'ont pas d'impact particulier sur la reconnaissance visuelle de mots de langue seconde mais impactent leur prononciation. D'autre part, aucune des deux tâches n'a permis d'observer un effet du marquage orthographique. Cependant, leur combinaison, ainsi que leur association avec un contrôle strict des variables CGP, a permis de révéler les conditions requises pour mettre en évidence un tel effet sur chaque processus étudié ici.

Pour notre deuxième étude, décrite dans le chapitre IV, nous avons choisi d'explorer plus précisément l'impact du marquage orthographique sur la reconnaissance visuelle de mots L2, en l'abordant sous un autre angle. En effet, le marquage orthographique est considéré comme un indice fiable sur la langue d'appartenance d'un item (Borragan et al., 2020; Casaponsa et al., 2014; Jared et al., 2013; Oganian et al., 2016; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012). Ainsi, un pseudomot contenant un marqueur orthographique (e.g., KN) est plus souvent catégorisé comme appartenant à la langue à laquelle appartient le marqueur (i.e., Anglais) qu'à l'autre langue maitrisée par l'individu bilingue (i.e., Français). De même, la catégorisation selon la langue d'appartenance est plus rapide pour les mots contenant un marqueur orthographique (e.g., KNIFE) que pour les mots ne contenant pas un tel indice (e.g., HOUSE). En outre, les résultats obtenus par Casaponsa et al. en 2014, 2016 et 2019, témoignant d'une facilitation de la reconnaissance visuelle de mots marqués comparativement aux mots non-marqués, ont conduit ces auteurs à faire l'hypothèse d'un rôle de guide à l'accès lexical pour les marqueurs orthographiques (et à la proposition d'une extension au modèle BIA+ ; Casaponsa et al., 2019). Toutefois, cette position n'est pas unanime, et certaines données empiriques, notamment celles récoltées par Van Kesteren et al., (2012) sont plutôt en défaveur de cette hypothèse, au bénéfice d'un impact stratégique des marqueurs orthographiques en lien avec les contraintes de la tâche à réaliser (ici aussi une extension du modèle BIA+ a été proposé).

Il est intéressant de noter qu'un autre indice renseignant sur la langue d'appartenance d'un item a également suscité des résultats hésitants lorsqu'il est question de son impact sur l'accès lexical. En effet, le modèle BIA+ fait l'hypothèse que le contexte linguistique, tel que celui apporté par une phrase, pourrait guider l'accès lexical et le rendre moins non-sélectif. Cependant, les études empiriques ayant testé cette hypothèse n'ont pas permis de proposer une conclusion claire à cette question (voir Lauro & Schwartz, 2017 pour une méta-analyse).. En effet, les études sur la question ont en majorité permis de répliquer les effets d'interférences interlangues, bien que l'item d'intérêt (soit un cognate, soit un homographe interlangue) était inclus dans une phrase (Schwartz & Kroll, 2006; Van Assche et al., 2009). Ces résultats tendent alors à indiquer que l'accès lexical est profondément non-sélectif. C'est-à-dire que les représentations lexicales des deux langues sont automatiquement activées et ce même si le contexte linguistique est spécifique à une langue.

Toutefois, une étude récente, conduite par Hoversten & Traxler, (2020) ayant exploité le language switch (i.e., le coût associé au passage d'une langue à l'autre) a permis de mettre en évidence des résultats en faveur d'un accès lexical partiellement sélectif. Selon cette nouvelle hypothèse, les représentations lexicales des 2 langues sont en effet accessibles à un individu bilingue. Cependant, le degré d'activation des représentations des différentes représentations L1 et L2 dépend du contexte linguistique dans lequel se trouve l'individu (ce qui se rapproche du bilingual mode hypothesis de Grosjean, 2001). Ainsi, dans un contexte majoritairement L1, le niveau d'activation au repos des représentations lexicales L2 est très bas, il faudra alors plus de temps pour atteindre le seuil d'identification pour les mots L2 que pour les mots L1.

Dans cette deuxième étude nous avons émis l'hypothèse que l'association des marqueurs orthographiques au contexte linguistique potentialiserait l'observation de l'impact de ce dernier sur l'accès lexical bilingue. Plus précisément, la présence d'un marqueur orthographique dans un contexte linguistique tel qu'une phrase (e.g., These fl**ow**ers gr**ow** in the jungle), faciliterait la détection de la langue du contexte comparativement à un contexte linguistique non-marqué (e.g., These spiders live in the jungle). Les marqueurs orthographiques permettraient alors aux contextes linguistiques qui les contiennent de restreindre l'accès lexical aux représentations de la langue cible de manière plus efficace que les contextes ne contenant pas de tels indices.

Dans cette deuxième étude, nous nous sommes proposé de tester cette hypothèse. Cependant, la manipulation de phrases complètes nous a semblé trop ambitieuse pour une première étude sur la question. Nous avons donc décidé d'utiliser un paradigme de flanker innovant tel que proposé par Declerck et al., dans leur article de 2017. Dans cette tâche, les participants doivent réaliser une décision de langue ou une décision lexicale sur un item cible (HOUSE) présenté au centre de deux items flankers (KNIFE-HOUSE-KNIFE). Ces items flankers, représentant le contexte linguistique, peuvent provenir de la même langue que l'item cible, on parlera alors d'essais congruents, ou de l'autre langue connue par le participant (JUPE-HOUSE-JUPE), on parlera alors d'essais incongruents. Le marquage orthographique quant à lui est manipulé au niveau des cibles qui peuvent être marquées pour l'anglais (KNIFE), pour le français (FLEUR) ou non-marquées (HOUSE ou JUPE). Le marquage est également manipulé au niveau des flankers, qui peuvent être marquées pour l'une des deux langues ou non-marquées. De plus, deux durées de présentation ont été appliquées, une condition courte (170ms), et une condition longue (3000ms) de manière à obtenir une image dynamique des effets que nous explorons.

La première expérience a donc consisté en une tâche de décision de langue, associée au paradigme de flanker, et réalisée selon 2 conditions de présentation (courte et longue). L'objectif était à la fois d'explorer l'impact du marquage et du contexte linguistique sur l'identification de la langue d'appartenance d'un mot cible, mais aussi de s'assurer qu'il est possible de mettre en évidence un effet du marquage orthographique en ayant recours à ce paradigme, avant de l'associer à une décision lexicale. Ainsi, dans cette tâche, il était demandé aux 83 participants d'indiquer la langue d'appartenance d'une série de mots cibles (80 mots français et 80 mots anglais) marqués ou non-marqués (40 mots marqués dans chaque langue) tout en ignorant les mots flankers qui les entouraient (40 mots français et 40 mots anglais) qui pouvaient eux aussi être marqués ou non. Il est important de noter que dans la moitié des essais, un conflit était induit quant à la langue d'appartenance des items à l'écran (e.g. cible L1 vs. flankers L2) qui représente ici une information critique pour la prise de décision.

Tout d'abord, les résultats de cette expérience ont permis de mettre en évidence un effet facilitateur du marquage orthographique de la cible, ce qui est en accord avec la littérature sur la question. En effet on a retrouvé de meilleures performances pour les décisions de langue concernant des cibles marquées que pour des cibles non-marquées. Cet effet principal était par ailleurs modulé par la durée de présentation des items, puisque dans la condition longue il concernait non seulement les cibles L2 mais aussi les cibles L1, tandis que dans la condition courte il n'apparaissait significatif que pour les cibles L2. Ce pattern semble indiquer que les participants de cette étude, de manière consistante avec la littérature

(Casaponsa et al., 2014; Duñabeitia et al., 2020; Vaid & Frenck-Mestre, 2002), seraient plus sensibles à la présence de marqueurs orthographiques provenant de la L2 qu'à la présence de marqueurs orthographiques provenant de la L1. Toutefois, la condition longue révèle que lorsqu'il est accordé suffisamment de temps aux participants, i.e.ls seraient capables de détecter des marqueurs orthographiques de la L1.

Ensuite, les résultats de cette première expérience ont également révélé de meilleures performances aux essais congruents qu'aux essais incongruents, confirmant les résultats de Declerck et al., (2018). Cet effet facilitateur de la congruence de l'essai était également modulé par la durée de présentation des items. En effet, dans la condition longue, le bénéfice de la congruence de l'essai se retrouvait aussi bien sur les cibles L1 que sur les cibles L2. Dans la condition courte en revanche, seul les mots cibles L2 étaient impactés par la congruence de l'essai. Ce pattern de résultats peut être interprété comme indiquant que dans la condition courte, les informations de langue contenues dans les mots de la L2, probablement extraites à partir des unités sublexicales, sont trop faibles pour résister à la compétition imposée par les mots flankers (quand le mot L2 est une cible) ou pour provoquer de la compétition (quand le mot L2 est un flanker). Dans la condition longue, la compétition se ferait entre les représentations lexicales associées aux items cibles et aux items flankers, ce qui expliquerait que les mots cibles soient similairement impactés par la présence de flankers peu importe leur langue d'appartenance.

De façon intéressante, dans la condition courte, l'effet de la congruence de l'essai était modulé par le marquage orthographique des items flankers. Effectivement, cet effet n'était significatif que lorsque les items flankers contenaient un marqueur orthographique de la L2 (KNIFE-HOUSE-KNIFE vs. KNIFE-JUPE-KNIFE). Cette modulation suggère que la présence de marqueurs orthographiques L2 (dans les items flankers) rend la langue du contexte linguistique saillante, ce qui amplifierait la compétition entre informations de langue provenant de la cible et informations de langue provenant du contexte. D'autre part, dans la condition longue, l'effet de congruence de l'essai dépendrait principalement de la présence de flankers L2 (vs. L1) peu importe si ces flankers contiennent un marqueur orthographique ou pas, ce qui est en accord avec notre hypothèse d'une compétition entre des informations d'ordre lexical pour la condition longue.

En s'appuyant sur les résultats de la première expérience qui ont pu mettre en évidence un effet du marquage orthographique des mots cibles dans une décision de langue, nous avons alors entrepris d'utiliser le paradigme de flanker en association avec une décision lexicale généralisée. Dans cette deuxième expérience, les 97 participants ont été amenés à indiquer si les items cibles formaient un mot de la L1 (80 mots français ; réponse OUI), de la L2 (80 mots anglais ; réponse OUI) ou n'existaient dans aucune des deux langues (160 pseudomots; réponse NON). Ces items pouvaient contenir un marqueur orthographique de la L1 (40 mots marqués pour la L1, 40 pseudomots marqués pour la L1), un marqueur orthographique de la L2 (40 mots et 40 pseudomots) ou ne contenir aucun marqueur orthographique (40 mots L1 non-marqués, 40 mots L2 non-marqués et 80 pseudomots non-marqués). De même les items flankers, qui devaient être ignorés par les participants, pouvaient être marqués pour la L1 (20 mots et 20 pseudomots) pour la L2 (20 mots et 20 pseudomots) ou non-marqués (40 mots et 40 pseudomots). Comme précédemment, dans la moitié des essais, un conflit était induit quant à la langue d'appartenance des items à l'écran. Cependant, il est important de noter que les mots cibles étaient toujours associés à des mots flankers et jamais à des pseudomots flankers (l'inverse est vrai pour les pseudomots cibles).

De ce fait, le conflit provoqué par les essais incongruents ne portait que sur les informations de langue d'appartenance et pas sur la lexicalité des items qui représentait ici l'information critique sur laquelle la prise de décision devait se focaliser.

Tout d'abord, les résultats à cette deuxième expérience ne sont pas parvenus à mettre en évidence un effet du marquage orthographique des cibles aussi évident que celui dégagé par la première expérience. Si l'on se limite aux réponses aux mots cibles, l'effet du marquage était tout simplement absent, peu importe la condition de présentation des items (courte ou longue). Cette absence d'effet du marquage sur les réponses aux mots cibles est en désaccord avec les résultats obtenus par Thomas & Allport, (2000) dans une décision lexicale généralisée classique, ce qui suggère que la présence d'item flanker dans notre tâche pourrait avoir effacé l'effet du marquage des cibles. Cependant, les réponses aux pseudomots cibles ont révélé un effet inhibiteur de la présence de marqueurs orthographiques L2 qui se trouve être en accord avec la littérature (Commissaire et al., 2018 ; Lemhöfer & Radach, 2008) et qui confirme l'aspect asymétrique de la sensibilité aux marqueurs orthographique. De plus, cet effet négatif des marqueurs orthographiques L2 n'était significatif que dans la condition de présentation courte, ce qui suggère que la présence de tels marqueurs n'a d'impact sur une décision lexicale seulement si un traitement plus en profondeur est empêché.

L'effet de congruence de l'essai quant à lui est tout à fait absent des résultats de cette deuxième expérience. A première vue, ces résultats semblent indiquer que les participants à cette tâche n'étaient pas en mesure d'extraire les informations de langue provenant des items flankers. Cependant, cette explication nous semble peu probable lorsque l'on considère les résultats à la décision de langue présentés ci-dessus. Il nous semble plus probable que cette information sur la langue d'appartenance, une fois extraite des items flankers, n'exerce aucune influence sur le processus impliqué dans la décision sur le statut lexical de l'item cible. Ces résultats sont en contradiction avec l'effet facilitateur de la congruence de l'essai mis en évidence par Declerck et al. (2018). Toutefois, cette différence dans les résultats peut être attribuée à une différence méthodologique. En effet, tandis que Declerck et al. ont eu recours à une décision lexicale par bloc de langue, nous avons utilisé une décision de langue généralisée dans laquelle les mots cibles L1 pouvaient succéder à des mots cibles L2 dans un même bloc. Cette différence de méthodologie a pu engendrer des divergences dans les attentes des participants quant à la langue des items cibles qu'i.e.ls étaient susceptibles de rencontrer (voir bilingual mode hypothesis Grosjean, 2001). Ainsi dans l'étude de Declerck et al., dans laquelle les participants ne s'attendaient qu'à des mots provenant d'une langue spécifique par bloc, l'incongruence de l'essai aurait eu plus d'impact que dans l'étude présente dans laquelle les participants s'attendaient à trouver des mots provenant des deux langues.

Afin de mieux comprendre la dynamique qui s'installe entre les informations de langue provenant des items cibles et des items flankers dans notre deuxième expérience, nous nous sommes à nouveau penché sur l'existence d'un effet de congruence de l'essai en fonction du marquage des items flankers. Cette analyse a révélé un effet facilitateur de la présence de flankers L1 sur la décision lexicale. La présence de flankers L2 en revanche, a un effet inhibiteur. Ce pattern suggère que les représentations lexicales associées aux mots L1 sont assez « fortes » pour apporter de l'aide aux participants lors de la décision lexicale tandis que les représentations lexicales associées aux mots L2 sont trop « faibles », ce qui ralentit le traitement et implique un ralentissement de la décision lexicale. De plus, la temporalité de l'effet inhibiteur des flankers L2 dépend du marquage de celles-ci. En effet, les flankers L2 non-marquées exercent leur impact sur la décision lexicale dans la condition courte tandis que les flankers L2 marquées n'ont un impact significatif que dans la condition longue. Pour rendre compte de ce pattern de résultats nous proposons que les représentations lexicales des mots L2 non-marquées sont accessibles rapidement, ce qui leur permet d'avoir un impact précoce sur la décision lexicale, tandis que les représentations lexicales associées avec les mots L2 marquées ne sont pas accessibles rapidement et ont un impact tardif sur la décision lexicale (voir Bartolotti & Marian, 2017).

Pour résumer, les expériences du chapitre IV avaient pour objectif d'explorer plus en profondeur l'impact du marquage sur la reconnaissance visuelle de mots de langue seconde en l'étudiant sous l'angle de son interaction avec une autre source potentielle de modulation de l'accès bilingue que représente le contexte linguistique. Cette étude a permis de mettre en évidence un effet de marquage de la cible lorsque celui-ci était associé à un avantage pour la prise de décision (i.e., dans la tâche de décision de langue mais pas dans la tâche de décision lexicale). Ce pattern supporte l'hypothèse de van Kesteren et al., (2012) selon laquelle le marquage orthographique serait utilisé de manière stratégique par les participants. De même, l'effet de la congruence de l'essai est présent dans la tâche de décision de langue mais pas dans la tâche de décision lexicale. Les participants à ces deux tâches semblent capables d'extraire les informations de langue mais pas de les utiliser pour faciliter une décision sur le statut lexical. Enfin, il semble que la présence de marqueurs L2 dans le contexte ait pu moduler l'effet de congruence de l'essai, notamment dans la condition courte de la décision de langue. Ces résultats semblent indiquer que les informations sur la langue d'appartenance d'un mot de langue seconde sont détectées plus tôt en présence d'un marqueur orthographique qu'en son absence, tandis que les représentations lexicales associées aux mots L2 marqués sont disponibles seulement tardivement. Ainsi, sur le plan méthodologique, cette étude a permis de révéler l'intérêt de l'utilisation du paradigme flanker pour explorer les questions de marquage orthographique. Elle a également permis de confirmer l'importance d'associer différentes tâches et différentes durées de présentation pour obtenir une image complète et dynamique du phénomène étudié.

En conclusion, dans le cadre de ce travail de thèse, nous avions pour objectif d'examiner les conséquences de l'apprentissage sublexical associé à l'exposition aux formes écrites de mots L1 et L2 sur la reconnaissance visuelle de mots de langue seconde. Pour remplir cet objectif, nous avons conduit deux études. La première étude, rapportée dans le chapitre III, nous a permis d'étudier l'impact respectif de 3 variables sublexicales que sont le marquage orthographique, la congruence interlangues des CGP et la consistance intralangue (en L2) des CGP. Les résultats de cette étude ont révélé que les participants, ici tous apprenants tardifs de leur L2, étaient sensibles aux variations inter- et intra-langue des CGPs. Ces résultats ont également révélé que l'effet du marquage orthographique était en partie lié aux propriétés des CGP auxquelles les marqueurs étaient liés (consistance) et/ou comparés (congruence). La deuxième étude, rapportée dans le chapitre IV, nous a permis d'explorer le marquage orthographique sous l'angle de son interaction avec un autre indice sur la langue d'appartenance d'un item (i.e., le contexte linguistique). Les résultats de cette deuxième étude indiquent que la présence de marqueurs orthographiques, et particulièrement la présence de marqueurs orthographiques L2, pouvaient avoir un impact sur la détection de la langue d'appartenance du mot qui le contient, mais n'avait pas d'influence sur l'identification de son statut lexical. Ces résultats suggèrent que l'information de langue associée à un mot de langue seconde contenant un marqueur orthographique était accessible plus rapidement que pour les mots de langue seconde ne contenant pas de tels indices. Cependant, la représentation lexicale associée à un mot de langue seconde contenant un marqueur orthographique ne semble être accessible que tardivement.

Bien que l'objectif principal du travail de thèse présenté ici ne consiste pas à répondre au débat sur la nature de l'effet du marquage orthographique, les résultats aux deux études rapportées ici semblent soutenir la position « stratégique » de Van Kesteren et al. au détriment de la position « automatique » de Casaponsa et al. L'effet du marquage orthographique apparaît clairement seulement lorsque l'information sur la langue d'appartenance de l'item est au centre de la tâche à accomplir (i.e., dans une tâche de décision de langue ; Chapitre IV, expérience 1). Lorsque cette information sur la langue n'est pas nécessaire pour prendre une décision (i.e., dans une tâche de décision lexicale ; Chapitre III, expérience 1A, 1C et Chapitre IV, expérience 2 ou dans une tâche de lecture à voix haute ; Chapitre III, expérience 2A et 2C), l'effet du marquage orthographique n'apparaît pas ou selon des conditions particulières (Chapitre III, expérience 1C et 2C). De plus, à l'inverse des hypothèses (et des résultats) de Casaponsa et al., l'effet du marquage sur l'accès lexical mis en évidence dans nos deux études semble être inhibiteur.

Enfin, sur le plan méthodologique, les résultats présentés dans notre chapitre III suggèrent que les variables sublexicales étudiées forment un nœud de variables difficile à démêler. Pour étudier le marquage orthographique, il sera par exemple nécessaire de tenir compte de la congruence (ou non) des CGP contenues dans les mots non-marqués auxquels seront comparés les mots marqués. Il sera également nécessaire de prendre en compte la consistance des CGP L2 de ces 2 types de mots (marqués et non-marqués) en accord avec la question de recherche formulée et avec les contraintes de la tâche sélectionnée. Ces mêmes résultats encouragent d'ailleurs l'utilisation d'une combinaison de tâches, et particulièrement l'association d'une décision lexicale avec une tâche de lecture à voix haute, pour faire sens de ce nœud de variable. Cet intérêt pour l'association de plusieurs tâches dans le but de répondre à une unique question de recherche est par ailleurs appuyé par les résultats présentés dans le chapitre IV, qui associe la décision lexicale avec une tâche de décision de langue. Le chapitre IV souligne également l'avantage de la multiplication des durées de présentation pour mieux comprendre la dynamique des effets étudiés.

Ainsi, les résultats mis en évidence dans ce document soulignent l'intérêt que représente l'exploration des variables sublexicales pour la compréhension des mécanismes impliqués dans le bilinguisme. Ces résultats encouragent également les futures études sur le sujet à considérer avec soin chacune des variables sublexicales mentionnées ci-dessus ainsi que la manière dont elles interagissent entre-elles. De même ces futures études sont incitées à multiplier les tâches et les paramètres qui y sont associés pour obtenir une image la plus complète et dynamique possible du problème qu'elles se sont fixées d'étudier.

GENERAL INTRODUCTION

In recent years, second language acquisition has become somewhat of a requirement to cope with the ever-increasing exposition to foreign languages in our daily life. Yet, the road toward L2-knowledge is paved by many challenges going from vocabulary acquisition to grammar rules appropriation. This second language is often acquired through the school system (at least for the European population), bringing its own challenges to the table. In fact, when faced with the arduous task of teaching a foreign language to a group of eleven year-old (or even older), teachers tend to play on their pupil's strength and rely on their reading ability by presenting them with a wide range of visual medias in the target language. Consequently, those individuals are rapidly and heavily exposed to their second language orthography, which will require them to overcome additional difficulties arising from the co-existence of the two written languages in their mind, such as integrating new orthographic patterns to their repertory, e.g., the KN bigram when learning English for a French native speaker, but also updating their set of grapheme-to-phoneme correspondences (GPC) to include those that are specific to the new language, e.g., the OU grapheme can be pronounced \aʊ\ in English but not in French, even though those GPCs may conflict with already established correspondences in the native language, e.g., the OU grapheme is pronounced \u\ in French.

The aim of this PhD thesis is to explore the implications of this "sublexical" learning tied to L2 acquisition for later L2-word processing. In fact, as we are about to see, even though the literature on bilingual visual word recognition is abundant, leading to the proposition of several models to describe the mechanisms involved in this process, the sublexical aspect of L2-visual word processing has been rather neglected.

To achieve that we will start by reviewing the literature available on visual word recognition both on the monolingual and the bilingual end of the spectrum and including considerations about the lexical and the sublexical aspect of this process. Thus, we will discuss the particularity of the bilingual lexical access and the cross-language interactions that derive from it, which will lead us to interrogate whether and how bilinguals manage this additional source of competition. Throughout this theoretical overview we will introduce several sublexical variables that emerge from the co-existence of two written languages in the bilingual mind and might either be a source of cross-language interactions or a source of control for the bilingual lexical access.

In the first experimental chapter, three of those sublexical variables, namely orthographic markedness, GPC cross-language congruency and L2 within-language GPC consistency, and their impact on bilingual lexical access will be investigated using three lexical decision tasks paired with three naming tasks, allowing us to have access to the recognition and to the pronunciation aspect of L2 word processing. In the second experimental chapter, a focus will be placed on orthographic markedness, which will be further investigated through its interaction with another source of language membership information, that is, the linguistic context. To do that we will introduce an innovative method consisting of pairing a flanker paradigm with either a language or a lexical decision task, allowing us to explore target word recognition while providing a minimal linguistic context.

THEORETICAL OVERVIEW

CHAPTER I / From modeling monolingual visual word recognition to modeling the bilingual lexical access

Nowadays, learning a second language has become necessary to reach a good quality of integration in an increasingly globalized world where an individual is likely to be exposed to foreign languages daily, in a variety of situations ranging from job interviews to the packaging of everyday items. To meet this need, at least for the European population, a particular emphasis has been put on second language (L2) acquisition by the education system. As a result, for those populations, the second language is often acquired through a school setting in which its written form plays an important role which presents some unique challenges. Consequently, the number of L2 learners is ever increasing and with it the need to explore the ramification of acquiring a second language through its visual forms for bilingual lexical processing and particularly for bilingual visual word recognition.

In this chapter our aim is to paint a picture of the bilingual processing of words presented visually based on the literature available on the topic. However, since the exploration of this question takes root in the literature about visual word recognition in a monolingual setting, e.g., models of L2 visual word recognition originates from models of visual word recognition in a monolingual situation, it seems necessary to present this literature. Similarly, the literature about expert reading introduces some important concepts for understanding the literature about visual word recognition, and therefore needs to be developed here. Hence, in this chapter, we will first present a brief overview of the literature about expert reading and L1 visual word recognition, in which we will develop the necessary concepts required to fully grasp the ins and outs of the following description of theoretical positions and experimental data available about L2 visual word recognition including considerations about lexical and sublexical units.

I. Monolingual visual word recognition: sublexical and lexical influences

1. Reading across languages

Every writing system aims to translate the spoken language into a visual representation through a universal principle, namely the universal phonological principle introduced by Perfetti and Dunlap, (2008) which states that across languages the act of reading consists into mapping these visual representations with the sounds of the spoken language and not directly at least with meaning. However, this same principle claims that every writing system will map a visual unit with the smallest phonological unit allowed by the language it represents. This will lead to differences across writing systems as to their particular mapping principle, and especially on the size/type of units they encode. Some systems encode whole morphemes like the Chinese logographs (猫= Māo meaning CAT), some systems encode syllables (the Japanese hiraganas ね=ne and こ=ko; neko meaning CAT) and others encode phonemes, i.e., the smallest phonological unit, this is the case of alphabetical systems like French or English (C=\k\, A=\æ\ and T=\t\; CAT). Reading is therefore ruled by general principles but also guided by language specificities, which we will endeavour to relate in the following section.

A. Modelling expert reading: general principles

Mostly based on European alphabetical systems, the Dual-Route Cascaded model of reading (Coltheart et al., 2001; see figure 1) describes two main pathways a reader may use to translate a written word into its phonological counterpart. The sublexical (or indirect) and the lexical (or direct) pathways. Both pathways are assumed to be activated parallelly during reading, but each pathway is at its most efficient for different type of words.

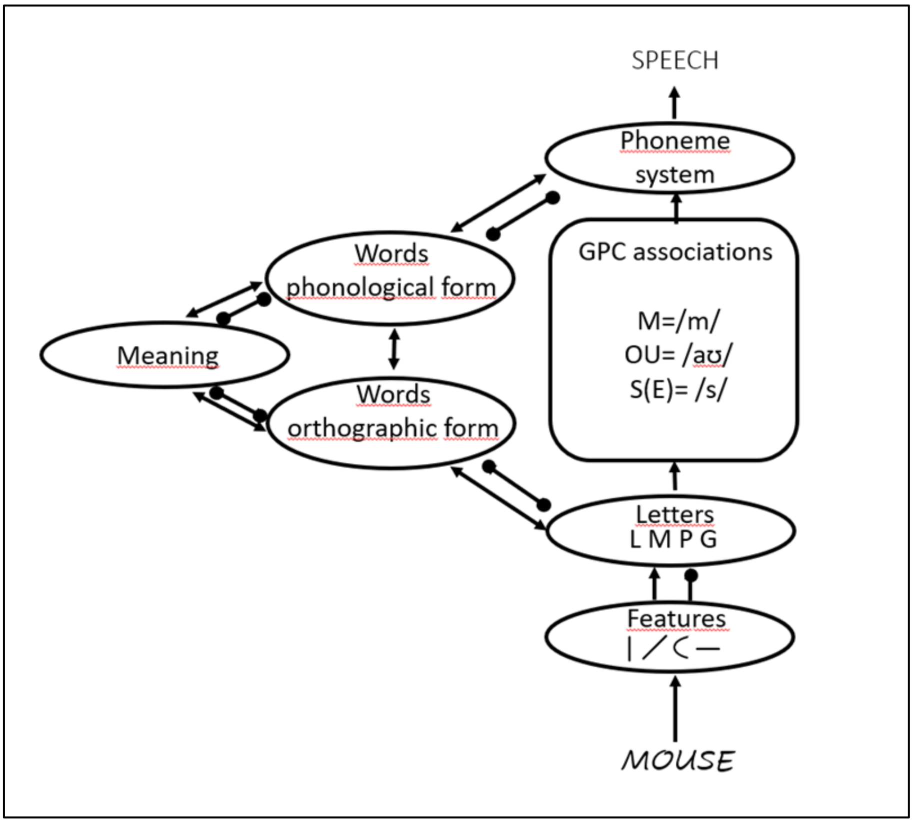


Figure 1: Dual Route Cascaded Model (Coltheart, 2001)

On the one hand, following the lexical pathway, the phonological output is retrieved from memory and is activated directly from whole-word representations, therefore this pathway is often the fastest route available to an expert reader. However, since this pathway relies on lexical knowledge, it is not always available for rare words, and is not useful for new words. For example, following this pathway, to read the word CAT, the reader needs to have already encountered it and to have stored both orthographic and phonological representations in their lexicon. On the other hand, the sublexical pathway implies the use of a set of rules that connects written units with their phonological counterpart, namely the Grapheme-to-Phoneme Correspondences (GPCs). Following this pathway, the reader will break down the word into graphemes, i.e., orthographic units that may refer to a phoneme, and associate each grapheme with the correct phoneme, which is a slower process than the one involved in the lexical pathway but has the advantage to be applicable to new words. Indeed, to read the word CAT following this pathway, the reader does not need to know the word CAT, but only the set of GPCs involved, e.g., C = \k\.

Thus both pathways are activated upon the presentation of a word. However, since the lexical pathway is more direct than the sublexical pathway, at least whenever the word is already known by the reader, it often ends up providing the response. When the word is not known on the other hand, the lexical pathway cannot provide a response, therefore the sublexical pathway takes over and the reader will have to break down the word into graphemes in order to retrieve the phonological representation of the item. This model of reading is assumed to be applicable to every European language, but as we are about to see, even though European languages may share this same basis, the actual way it will be used will depend on language orthographic specificities.

B. Modelling expert reading: language specificities

In fact, even though European alphabetical systems share more or less the same mapping principle, i.e., a grapheme is associated with a phoneme, they differ from one another in terms of orthographic depth, which refers to both (1) the complexity (size) and (2) the consistency (reliability) of these associations (Schmalz et al., 2015). Some languages such as English and French use complex graphemes, often pairing a couple of letters (e.g., OU) with a phoneme (\u\ in French or \u:\ in English), whereas some other languages such as Spanish or Italian will mainly pair a single letter with a phoneme (e.g., U = \u\ in both Spanish and Italian, i.e., simple graphemes). Two languages can also differ on the consistency of such GPCs[2]. For example, Spanish, which has a transparent orthography will map one grapheme with only one phoneme (U= \u\ in every occurrence like UNO, ULTIMA, USTED). On the other hand, an opaque orthography like English will map one grapheme to several phonemes with little consistency (OU= \u\ in YOU or OU= \aʊ\ in HOUSE etc…). The particular orthographic depth of a language has been hypothesized to impact both the reading strategy (orthographic depth hypothesis, Katz & Frost, 1992) and the size of the unit (grain-size hypothesis, Ziegler & Goswami, 2005) used by expert readers.

In fact, because GPCs are not completely reliable in opaque orthographies, their readers should be discouraged to use the sublexical pathway. That is not the case of transparent orthography readers for whom GPCs are highly consistent, and the sublexical pathway is quite reliable. Therefore, the orthographic depth hypothesis (ODH; Katz & Frost, 1992) assumes that readers of a shallow orthography will rely more heavily on sublexical strategies than readers of a deep orthography, and inversely, readers of a deep orthography will rely more heavily on lexical strategies than readers of a shallow orthography.

---

[2] *The inconsistency can come from the Grapheme-to-Phoneme association like described here, in which case it will be referred to as feedforward consistency, or it can come from the Phoneme-to-Grapheme association, in which case it will be referred to as backward consistency. For example, the phoneme \ɛ\ can be transcribed by several graphemes in French, e.g. AI in MAISON, È in ÉLÈVE, ER in TERMINER etc…. In this document we will focus on feedforward consistency and mostly ignore backward consistency.*

Evidence for this hypothesis can be found in various forms. A first example is that the lexical status, i.e., the comparison between words and nonwords and between high and low frequency words, and the semantic priming (facilitative, e.g., cat-DOG), exhibit a stronger effect in opaque orthographies than in transparent ones (Frost, 1994; Frost et al., 1987). But more importantly, the comparison of the reaction times in a lexical decision task and a naming task, in two versions of the same orthographic system (pointed and unpointed Hebrew), highlighted some interesting results in favour of the ODH. In fact, while in the opaque version of the Hebrew orthographic system (unpointed Hebrew), the reaction times to the lexical decisions were similar to those obtained in the naming task, hinting at the adoption of a lexical strategy in both tasks for the opaque orthography, the difference between lexical decision and naming task was significant in the transparent version of this same orthographic system (pointed Hebrew; Naming < LDT), suggesting the adoption of a sublexical strategy in the naming task for the transparent orthography. These results have been gathered through within- (by using pointed and unpointed Hebrew orthographies; Frost, 1994) and cross-language (by using unpointed Hebrew, English and Serbo-Croatian; Frost et al., 1987) studies focusing on the consistency aspect of the orthographic depth factor.

More recently, a study from Schmalz et al., (2016) has considered both the consistency and the complexity aspect of orthographic depth and their influence over reading times by using predictable and unpredictable English words and complex or simple French words that they presented to native readers to read aloud. The results showed that English readers took more time to read aloud unpredictable words, for which at least two pronunciations were possible, i.e., the G in GHOST could be hard \g\ or soft\ʤ\, than to read aloud predictable words for which only one pronunciation was possible, i.e., the G in FORGE can only be pronounced \ʤ\, indicating that unpredictable words were likely processed based on lexical knowledge more than on GPC parsing. French readers on the other hand, did not differ in their response time for complex, i.e., containing a complex grapheme like GATEAU, and simple, i.e., containing only simple graphemes like GARNIR, words. However, there was an interaction between complexity and frequency, the frequency effect being stronger for complex words than it is for simple words, indicating that the lexical pathway was more involved for complex words than it was for simple words. Results to the two experiments led to the conclusion that both complexity and (in)consistency increased the reliance on the lexical route by impairing the sublexical pathway.

Regarding the sublexical reading pathway, the Grain Size Hypothesis (GSH; Ziegler & Goswami, 2005) goes further in assuming that language specificities may lead to slight accommodations of this process. As we have seen, in an opaque orthography, using the basic grapheme-to phoneme correspondences can lead to a lot of mistakes, the EA grapheme for example can be pronounced \i\ or \eɪ\ in English. To reduce this inconsistency, opaque orthography readers can take larger units into account. For example, when the EA grapheme is embedded in a rhyme, it gains some consistency, the rhyme -EAM for instance is consistent in English, it is always pronounced \im\ as in SCREAM-DREAM or CREAM. Thus, since it is beneficial for their accuracy, opaque orthography readers are assumed to apply larger grain mapping than transparent orthography readers do.

Evidence for this hypothesis has first been gathered with children in cross-language studies where they were asked to read aloud some pseudowords. Goswami et al., 1998 for example, showed that English readers (opaque orthography) benefited more from the presence of a familiar rhyme (e.g., DAKE derived from LAKE vs. DAIK which contains an unfamiliar rhyme), than French readers (less opaque orthography) did, indicating that opaque orthography readers relied more heavily on larger units, here the rhyme unit, compared to less opaque orthography readers. Also, using similar stimuli, they found that while English children performed better when large units (DAKE) and small units (DAIK) pseudowords were blocked by size than when they were mixed together in the same list, German children (transparent orthography) performed similarly whether both items were mixed or blocked. This indicated that while children familiar with an opaque orthography used both small and large sublexical units to decode those pseudowords, entailing some processing cost in the mixed condition, children familiar with a transparent orthography relied more heavily on smaller sublexical units.

Still, the most striking evidence for the Grain Size Hypothesis, comes from Ziegler et al., 2001, who asked expert readers of an opaque language, i.e., English, and expert readers of a transparent language, i.e., German, to read identical words, e.g., SAND, of varying length (4 to 7 letters) and word-body neighbourhood size (a few vs. a lot of rime related words, SAND: BAND, LAND, HAND etc.). They found that the strategy readers used to read the same words depended on the orthographic depth of their language. On the one hand, readers of German exhibited a stronger length effect, suggesting a greater reliance on smaller sublexical units, than English readers. On the other hand, readers of English exhibited a stronger word-body neighbourhood size effect, suggesting a greater reliance on larger sublexical units than German readers.

To sum up, every writing system share the same goal, that is to visually represent the sound of the spoken language, this general principle allowed researchers to propose a model of reading that could apply to every writing system even though it was based on the investigation of European language reading. Yet, the particular constraints imposed by each language shape their writing system and more particularly the associations they make between phonological and orthographic units. Even languages that share a mapping principle, i.e., European languages that use the roman alphabet, may present some discrepancies as to the size and consistency, i.e., the orthographic depth, of those associations. In turn, these orthographic specificities are thought to encourage readers to adapt their reading strategies, resulting in cross-language differences in terms of reading strategies and particularly in terms of the size of the units used by readers[3].

2. Identifying whole words

Yet, once a word is acquired, the lexical pathway should provide the fastest response no matter the orthographic depth an expert reader is presented with. Even in transparent orthographies a reader should benefit from a direct access to the whole-word orthographic and phonological representations compared to the slow process of parsing and activating every single GPC contained in a word. Therefore, exploring whole-word visual recognition is paramount to the understanding of expert reading. Consequently, numerous propositions have been made to model this process, i.e., logogen model (Morton, 1969), Activation-Verification model (Paap et al., 1982) or the MROM model (Grainger & Jacobs, 1996) , and one of those model, the IA model (McClelland & Rumelhart, 1981) in particular was later used in an attempt at accounting for the various orthographic and phonological effects that have been shown to impact visual word recognition.

[3] Orthographic depth also impacts reading acquisition (Seymour et al., 2003; Ziegler & Goswami, 2005)

A. *Modelling visual word recognition*

So, the most relevant model for our purpose, but also the most popular model of word recognition in expert readers is the Interactive Activation model which was first described by McClelland and Rumelhart in 1981, and is still the basis for many recent assumptions about visual word recognition (see figure 2). Structurally, the IA model is constituted of three levels, each encoding a different type of unit. The first level encodes letter features, which are distinctive visual traits found in letters, e.g., |; \; ⌡; Ɔ; —; etc., the second level encodes abstract letter identities, e.g., L, G, P etc., and acts as a mental alphabet, the third level encodes whole word units, e.g., POMME.

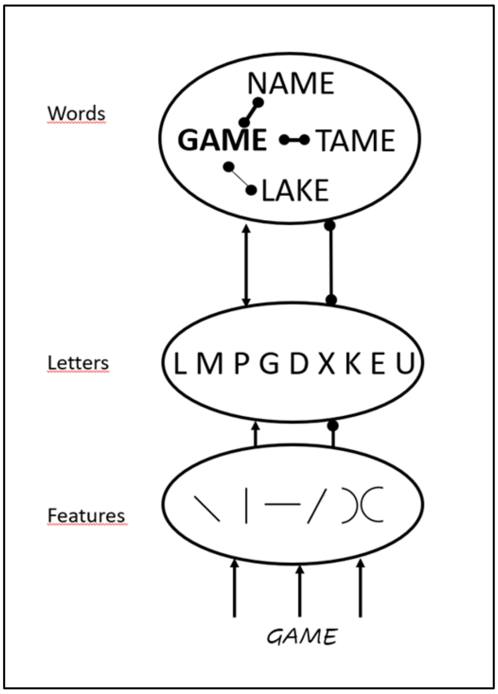


Figure 2: Interactive Activation model (McClelland and Rumelhart, 1981)

Those three levels communicate with each other through bottom-up connections. Therefore, a visual trait like a vertical line, i.e., I, could send activation toward letters containing such a trait, e.g., L, K or D (among others), and at the same time it could also send inhibition toward letters like S or X which do not contain this distinctive trait. In turn, the letters that were activated can send forward the activation toward words that contain those letters, the letter L would activate the word LAKE, LOOK and LUCK (among others) and inhibit words that do not contain it e.g., MAKE, RICE, TOOK etc.

Importantly, as it was first designed to account for the word superiority effect[4] (WSE) exhibited for the first time by Reicher (1969) and Wheeler (1970), and indicating that lexical knowledge impacts letter perception, the IA model also contains top-down connections between the word and the letter level. Finally, at the word level, several units will be activated, those units will then need to compete for the word recognition to be complete, as the first word that reaches a threshold of activation will be selected for recognition. This is represented within the IA model by a complex network of resting and activation levels (depending on frequency of occurrence etc.) and inhibitory connections between word units.

B. Lexical competition and orthographic overlap

Interestingly, the IA model has been successful at accounting for the orthographic neighborhood effect, which originates from partial orthographic overlap between words, and refers to the finding that words with a lot of neighbors like MAKE (17 neighbors) lead to different identification rate than words with only a few neighbors like HOUR (6 neighbors). In fact, since in the IA model the strength of the connections between representations is dependent on the similarity between them, orthographic neighbors[5] (LAKE and MAKE) will

---

[4] *The word superiority effect refers to the fact that a letter, e.g. D, is better perceived when it is included in a word, e.g. WORD, compared to when it is included in a pseudoword, e.g. OWRD, or even more surprisingly, when it is presented in isolation.*

[5] *As a result of the strict letter coding adopted by the IA model to explain how words like ART and RAT are distinguished quite efficiently even if they share the same letters, the IA model describes an orthographic*

inhibit each other more importantly than words that do not overlap orthographically (LOOK and HEAD). Consequently, still following the IA model, the amount of lexical competition generated by the recognition of a target word should depend on the number of orthographic neighbors it possesses, i.e., on orthographic neighborhood density: a word coming from a large neighborhood (e.g., MAKE; 17 neighbors) should receive more inhibition than a word coming from a small neighborhood (e.g., HOUR; 6 neighbors).

To shed some light on this question a lot of studies have used the priming paradigm (Davis & Lupker, 2006; De Moor & Brysbaert, 2000; De Moor & Verguts, 2006; Grainger & Ferrand, 1994; Segui & Grainger, 1990). In this paradigm, the presentation of a target word (MAIL) follows the brief presentation of a prime word that can be related orthographically (FAIL), phonologically (MALE) or even semantically (POST) with the target word, or not at all (i.e., control; e.g., JUMP). Participants have to take a decision on the target word only, and their responses to the related condition are compared to their responses to the unrelated condition. Critically for us, following the IA's hypothesis the presentation of an orthographic neighbour as a prime word (i.e., fail-MAIL) is expected to slow down the recognition of the subsequent target compared to an unrelated prime-target pair (i.e., jump-MAIL). This inhibitory orthographic priming effect was found no matter the Orthographic Neighborhood size (De Moor & Brysbaert, 2000) but did depend on the relative frequency of the prime/target pair (rather than their absolute frequency, Davis & Lupker, 2006; Grainger & Ferrand, 1994; Segui & Grainger, 1990). This Orthographic Neighborhood frequency effect, in line with the IA

---

neighbors as a word that share every letter, in the same slot/position, with the target word except one. For example, the word LATE is an orthographic neighbor of the word LAKE, but the word TALE isn't.

model's assumptions, show that partial orthographic overlap may impact visual word recognition in expert readers.

C. Phonological overlap and sublexical influences

When considering visual word recognition, another important question pertains to the role phonological information plays in the process. Several positions have been held on this issue, phonological information has been considered in turn to be completely superfluous (Morton, 1969; Paap et al., 1982) or to be absolutely necessary (Rubenstein et al., 1971) to the visual word recognition process. Critically, the DRC model presented above (Coltheart et al., 2001), adopted an intermediate position, in which phonological information was not always required to complete visual word recognition (direct access) but could guide the word recognition process under some circumstances where the orthographic information is not sufficient (indirect access).

To explore the role of the phonological information in visual word recognition, researchers have used homophones, which are words that sound the same as another word but are spelled differently and refer to another meaning, e.g., MAIL and MALE. They also used pseudohomophones, which are pseudowords that are constructed to sound like an existing word based on GPCs, e.g., CHARE and CHAIR. The rationale behind the use of those stimuli is that if phonological information is indeed automatically activated upon the presentation of an orthographic form, then responses to those items should differ from responses to control words and pseudowords that present no such phonological overlap with another word. Overall, homophones and pseudohomophones are recognized slower than control items (MALE vs. TAKE; Ferrand & Grainger, 1992, 2003; Rubenstein et al., 1971; Ziegler, Jacobs, et al., 2001) but may facilitate target word recognition when used as primes (MALE-mail vs.

TAKE-mail; Brysbaert, 2001; Ferrand & Grainger, 1992, 1993; Grainger & Ferrand, 1994, 1996; Perfetti & Bell, 1991; Rastle & Brysbaert, 2006); either way these results support the hypothesis of an automatic activation of the phonological code during visual word processing.

Even more importantly, since the pseudohomophone's phonological overlap with a preexisting word derives from the Grapheme-to-Phoneme Correspondences it is constituted of, the pseudohomophone effect indicates that expert readers not only retrieve lexical phonological information but are also able to generate the phonological code based on orthographic information through a sublexical parsing of the string of letters. This is in line with the DRC model presented above, the homophone effect illustrating the lexical/direct access to the phonological representation, and the pseudohomophone effect illustrating the sublexical/indirect access to the phonological representations through GPCs conversion. However, those effects also highlight the need for the IA model to better integrate the phonological information, both lexical and sublexical, to account for visual word recognition. This was the aim behind the later BIAM model proposed by Grainger and Ferrand in 1994 and 1996, which can be considered as an extension to the IA model that account for orthographic and phonological, lexical and sublexical contributions to visual word recognition.

3. Intermediate summary:

In every language, the act of reading consists in mapping visual representations to their phonological counterparts. This universal principle (Perfetti & Dunlap, 2008) allowed researchers to propose a model of reading that may apply to every situation: i.e., the DRC model (Coltheart et al., 2001). Following the DRC's assumptions, readers may retrieve the phonological output either (1) based on lexical knowledge, through the direct/lexical pathway or (2) by associating each grapheme to its phonological counterpart (phoneme), through the indirect/sublexical pathway. Still, every writing system must deal with the particular constraints set by the language they represent, encouraging their readers to adapt their reading strategies, in particular for the sublexical processing (Katz & Frost, 1992). The lexical processing on the other hand should give fast and accurate responses (upon known word presentation) for every reader no matter the orthographic depth of their language. Thus, models of whole word visual word recognition should also apply to every language.

The IA model (McClelland & Rumelhart, 1981) proposes a computational account of the visual word recognition process by introducing three representation levels, each encoding for a different unit going from visual features to whole-words via letter units, and describing the link connecting every level to the next. Interestingly, the IA model has been able to account for the orthographic neighborhood effect which refers to the fact that a word with a lot of neighbors (MAKE) generate more lexical competition than a word with only a few neighbors (HOUR) and suggests that the recognition of a target word is impacted by its orthographic proximity with other words. The homophone and particularly the pseudohomophone effect on the other hand underscore the automatic activation of phonological information involved in visual word recognition and led to the proposition of an extension to the IA model, i.e., the BIAM model (Grainger & Ferrand, 1994, 1996). To sum up, the data presented in this section seem to indicate that visual word recognition is influenced by orthographic and by phonological information coming from the lexical and from the sublexical level.

II. Bilingual visual word recognition: lexical and sublexical influences

When learning a second language (L2) an individual is confronted to a new system of communication, including new vocabulary. Learning this L2 vocabulary poses distinct challenges compared to learning new first language (L1) vocabulary.

First, it is accompanied by an asynchrony between the acquisition of the concept and the lexical form. Most of the time, when learning a new L1 word (e.g., FERN), the new lexical form refers to a new concept (e.g., a flowerless plant) for the learner. Hence, both representations will be bound together and added to the lexicon at the same time. Learning a new L2 word (e.g., APPLE) on the other hand, implies associating a new lexical representation to an already well established network of representations, containing both the concept (e.g., a round fruit) and the L1 lexical representation (e.g., POMME). This asynchrony between the acquisition of the concept and the acquisition of the lexical form in L2-learners elicited the need to investigate L2 word recognition *per se* instead of merely generalizing monolingual findings to bilingual processing.

Furthermore, since L2 vocabulary acquisition is often imparted through the school system, which heavily implies the use of visual supports (i.e., use of workbooks and subtitles), most L2 learners encounter new L2 words through their written from before they hear their phonological form for the first time. This is not the case for native language vocabulary learning, where new vocabulary is often encountered orally before the introduction of its printed form. The asymmetry between orthographic and phonological form of L2 words leads to an emphasis on bilingual visual word recognition over spoken word recognition (Bassetti, 2008; Cornut, 2021).

In the first section of this chapter we have seen that visual word recognition in one's native language is a complex process that requires the activation of a lot of representations (orthographic but also semantic and phonological) on a lexical but also a sublexical level. Considering the asynchrony between lexical and semantic integration but also the asymmetry between orthographic and phonological information, recognising a word in a second language should therefore be at least as complex as L1 word recognition, if not more so. In the next section we will endeavour to describe the diverse consequences of L2-acquisition and subsequent bilingualism on visual word recognition both on lexical and sublexical processing.

1. Two languages, twice the competition? Lexical aspect of bilingual visual word recognition

Models of L2 vocabulary acquisition, such as the BIA-d model and the RHM model (Grainger et al., 2010; Kroll & Stewart, 1994 respectively), state that at first, the direct connection between the L2 lexical representation and the semantic representation is weak (see figure 3). Thus, access to the meaning of the L2 word will need to call on its excitatory connection with the L1 lexical representation of its translation equivalent (i.e., indirect access). With increasing L2 exposure, the direct link with the semantic representation will become stronger and the connections between L1 and L2 lexical representations will become inhibitory, reflecting their integration into a single bilingual lexicon. This implies the existence of two lexical representations for one concept with comparably strong connections in highly-proficient bilingual individuals.

Therefore, when considering bilingual processing, a crucial issue pertains to the mechanisms involved in order to access the correct representation for the task at hand. Since bilinguals hardly ever make mistakes, we could intuitively assume that bilinguals select the language in which they will address their lexicon. That is the assumption of the selective access hypothesis (Macnamara & Kushnir, 1971). However, given the cross-language interaction effects widely found in visual word processing (see Dijkstra, 2007), the most widely accepted position about the bilingual lexical access is the non-selective access hypothesis, which assumes that bilinguals activate lexical representations from both languages non-preferentially. It is the position assumed by the BIA model (van Heuven et al., 1998) and its extensions BIA+ (Dijkstra & van Heuven, 2002) and BIA-d (Grainger et al., 2010) which are the most influential models of bilingual visual word recognition.

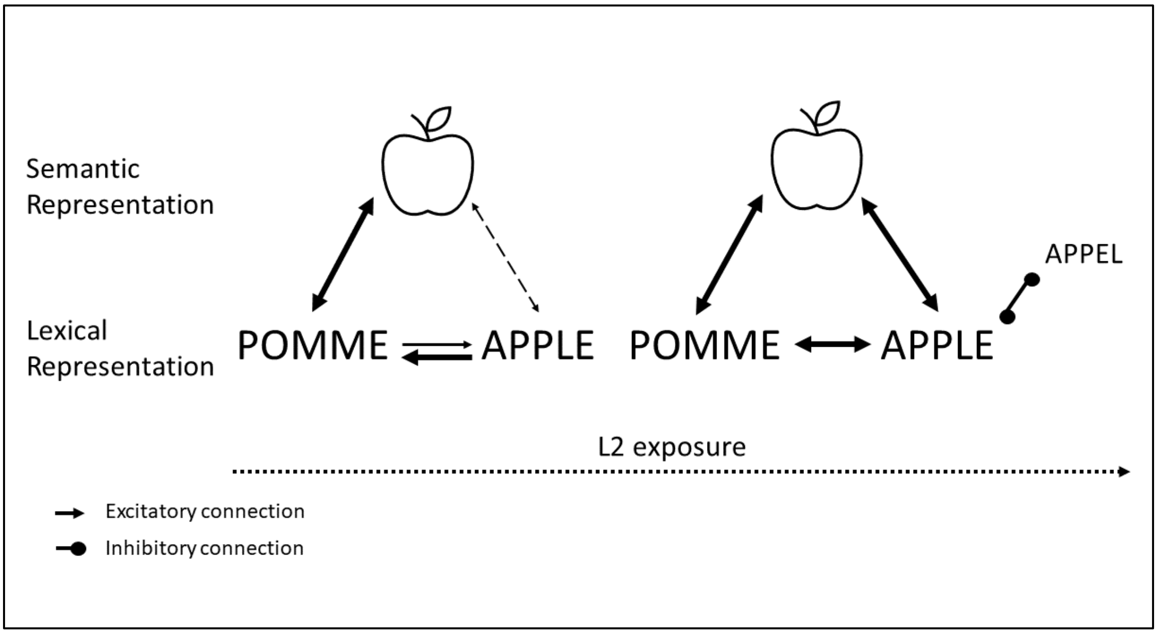


Figure 3: L2-vocabulary acquisition according to the BIA-d model (Grainger et al.,2010) and the RHM model (Kroll & Stewart, 1994)

It is important to mention that a recent proposition has been made to account for L2 visual word recognition in the form of the Multilink model (Dijkstra et al., 2019). However, since we aimed at exploring the sublexical aspect of L2-visual word recognition and since we wanted to look at the interaction between word units, and considering that both elements were absent from the Multilink model, we will not develop this model further in this thesis.

A. Common ground in modeling the bilingual lexical access

Building on the IA model of monolingual lexical access, the BIA models (represented in figure 4) assume that the bilingual lexicon is integrated and that the bilingual lexical access is language non-selective. Meaning that both L1 and L2 lexical representations will be stored in a unique lexicon and associated with their corresponding language node. While the two models differ on the particular function of such nodes (see chapter II), both models agree to define them as language membership representations that accumulate activations from their associated lexical representations, allowing bilinguals to retrieve the language membership information associated with a particular word, although only after complete word identification (i.e., post-lexically). Therefore, according to those models, bilinguals are able to determine to which language a particular word belongs to, but this information will not be used to access lexical representations preferentially, leading to predictions of cross-language interactions in bilingual visual word recognition. When presented with a written word (e.g., POOL), a bilingual will not only activate within-language lexical representations (e.g., TOOL) but also cross-language lexical representations based on the form overlap (either orthographic e.g., POIL(hair), phonological e.g., POULE(hen) or semantic e.g., PISCINE(pool)) with the printed word, and this will influence its processing.

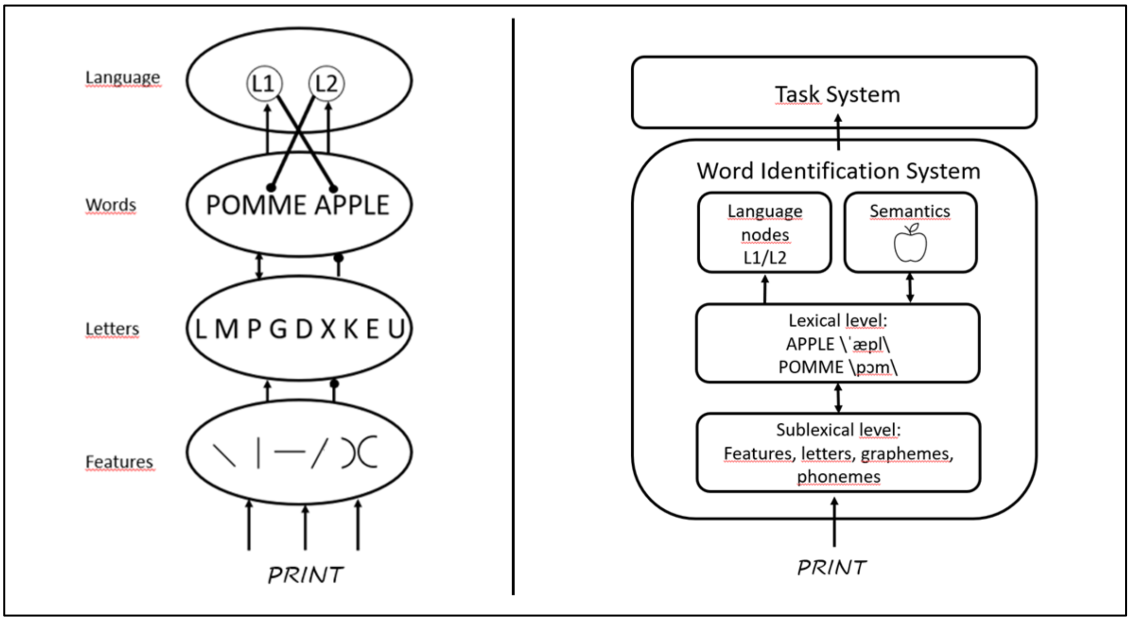


Figure 4: On the left the BIA model (Van Heuven et al., 2008), on the right the BIA+ model (Dijkstra et al., 2002)

B. Evidence for cross-language interactions coming from isolated word studies

Following those assumptions, the BIA models predict that bilingual word recognition will be influenced by lexical competition, not only within language but also across languages. For example, the visual presentation of a word like FIRE will lead bilinguals to activate not only the target lexical representation but also those orthographic neighbors in the same language (i.e., within-language neighbor; FIVE) and in the other language known by the participants (i.e., cross-language neighbor; LIRE). This is supported by the finding that the recognition of a target word can be delayed by the presentation of a cross-language orthographic neighbor as a prime (LIRE-fire vs. LUNE-fire; Bijeljac-babic et al., 1997; Commissaire, 2021; Dijkstra et al., 2010). Furthermore, still following the BIA assumptions, the size of the cross-language neighborhood should impact target word recognition. A target word with a few cross-language neighbors (e.g., FINGER has only 1 French neighbor) should experience less lexical competition than a target word with many cross-language neighbors (e.g., PASTE has 10 French neighbors).

Although few in numbers and not consistent in their results, studies on Cross-Language Neighborhood Density (Commissaire, Audusseau, et al., 2019; Dirix et al., 2017; Grossi et al., 2012; Midgley et al., 2008; Mulder et al., 2018; van Heuven et al., 1998) provide some interesting data for understanding the architecture and dynamics of the bilingual lexicon.

The earliest study investigating the cross-language orthographic neighborhood density effect (van Heuven et al., 1998) used words coming from 1) small within- and small cross-language neighborhood (SISTER; 3 within and 2 cross-language neighbors) 2) small within- but large cross- language neighborhood (GARDEN; 3 within and 6 cross-language neighbors) 3) large within- but small cross- language neighborhood (FINGER; 8 within and 1 cross-language neighbors) and from 4) large within- and large cross- language neighborhood (TRADE; 5 within and 7 cross-language neighbors) in several word identification tasks. Large cross-language neighborhood (TRADE and GARDEN) led to slower visual word recognition than small cross-language neighborhood (SISTER and FINGER), when participants had to identify an L2 word that was gradually revealed to them in a progressive demasking task (PMD). It was also true when they had to indicate whether an item, either a word or a pseudoword existed in an L2 LDT and a generalized LDT (GLDT; in which both L1 and L2 words should be associated with a YES-response).

There has been some attempts at replicating the cross-language orthographic neighborhood density effect (Commissaire et al., 2019; Dirix et al., 2017, experiment 1; Grossi et al., 2012; Midgley et al., 2008; Mulder et al., 2018, experiment 1), however their results were mixed. While several later studies successfully found an effect of cross-language orthographic neighborhood density either by using ERP recordings (Grossi et al., 2012; Midgley et al., 2008) and GLDTs (Dirix et al., 2017), Mulder et al., (2018) and Commissaire et al., (2019), were not able to replicate the cross-language orthographic neighborhood density effect in an L2 lexical decision task. The results from Mulder et al., are even more surprising considering that they used the exact same material as the one used in van Heuven et al., (1998). However, they explained those discrepancies by an increase in L2 proficiency between the two cohorts of participants involved in the two studies, without rejecting the non-selective access hypothesis.

Thus, even though the cross-language orthographic neighborhood density effect is not consistently replicated, every study on the topic argues for a non-selective bilingual lexical access. Cross-language orthographic neighbors (PRISE) are assumed to be activated upon the presentation of a target word (PRICE). They will then compete, alongside within-language neighbors (PRIDE) for selection and this will slow-down recognition of the target word depending on the task requirements and the L2-proficiency of the individuals (see figure 5).

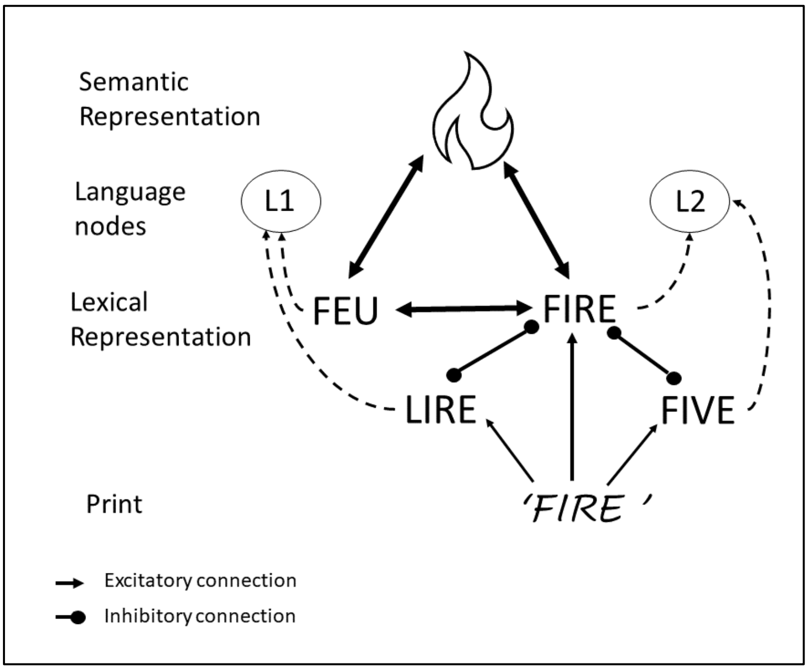


Figure 5: Connections between cross-language neighbors

Another way to test the non-selective lexical access, which has been more largely adopted than the investigation of the cross-language orthographic neighborhood density effect, comes from the use of language-ambiguous items. In fact, the two languages known by a bilingual rarely differ on each and every word, some cross-language overlap is to be expected, words may share some of their features across the two languages. For example, some words will overlap both on orthographic form and on meaning across languages, those words are called cognates. Two types of cognates are described in the literature; identical cognates, which completely overlap across languages (e.g., TRAIN in English and French) and near cognates, which partially overlap on orthographic forms across languages (e.g., the pair BANANE-BANANA also in French and English). Sometimes this cross-language orthographic overlap can be misleading, some words will overlap on orthographic form but will refer to two distinct meanings depending on the language, those words are referred to as interlingual homographs. For example, the word FOUR exist in both French and English, but it refers to the number 4 in English while it means OVEN in French. Also some words will not overlap on their orthographic form, but will overlap on their phonological form across languages, the English word POOL for example is pronounced like the French word POULE (hen). These various types of cross-language overlap may have an impact on bilingual visual word recognition. In fact, if the bilingual lexical access is truly non-selective, those ambiguous words should be processed differently, or at least at a different pace, than control words (which do not overlap across languages, e.g., APPLE) are.

For instance, the BIA models assume that an identical cognate such as TRAIN, would share a unique lexical representation across languages and unlike control words e.g., APPLE, would benefit from an already strong association with the semantic representation (see figure

6). This is supported by the facilitative cognate effect (faster RT for cognates than for controls), which has been widely replicated across a large panel of tasks, such as a L1 LDT (van Hell & Dijkstra, 2002)[6], some L2 LDTs (Dijkstra et al., 1999; Dijkstra, Miwa, et al., 2010; Lemhöfer et al., 2004; Poort et al., 2016; Poort & Rodd, 2017; Vanlangendonck et al., 2019), a GLDT (Lemhöfer et al., 2004) and some PMD task (Dijkstra et al., 1999; Dijkstra, Miwa, et al., 2010), and is already present with children at early L2-learning stages (Brenders et al., 2011). Furthermore, the cognate effect has not only been found through behavioural measures but also on ERP measures (the N400 is less negative for cognates than controls, Peeters et al., 2013)[7]. Yet, it is important to distinguish between identical and near cognates (Arana et al., 2022; Dijkstra, Miwa, et al., 2010; Vanlangendonck et al., 2019), since near cognates like BANANA-BANANE, are less likely to share a unique lexical orthographic representation. This is supported by the finding that, while both cognate types will be recognised faster than control words, the difference will be more important for identical cognates than for near-cognates (Dijkstra, Miwa, et al., 2010; Vanlangendonck et al., 2019).

[6] *The cognate effect depends on non-target language proficiency: found for L1-L2 cognates, and for L1-L3 cognates but only if participants are HP in L3*
[7] *where it is further modulated by the frequency of the cognate in both the target (L2) and the non-target (L1) language*

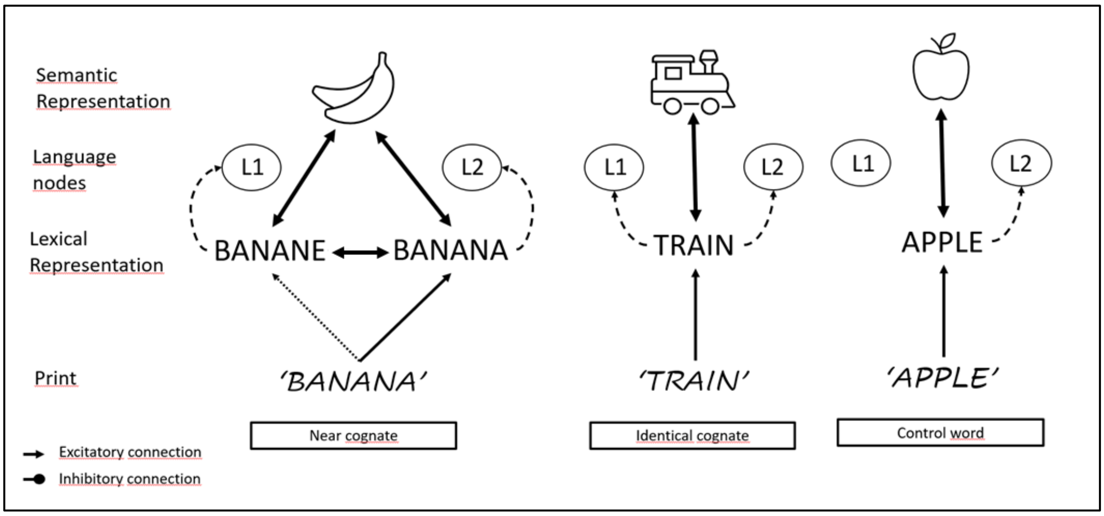


Figure 6: Identical and Near cognate representations compared to a control word

Interestingly, cognates (e.g., TRAIN or BANANA/BANANE) have been found to be inhibitory in a language decision task (Dijkstra, Miwa, et al., 2010), confirming that those words are ambiguous when it comes to their language membership identity. To account for this finding, the BIA models state that cognate words are associated with both L1 and L2 language nodes (vs. control words that are associated to only one language node) leading to conflict when detecting the language membership of these items. Similarly, Schwartz et al., 2007, in a study using a naming task found that identical cognates with different pronunciations across languages (i.e., SO cognates, e.g., TRAIN) led to slower naming responses than cognates with closer phonological forms across languages (i.e., SOP cognates, e.g., PIANO), a pattern that suggests that cognates may be associated with two different lexical phonological representations[8].

[8] *Still, both SOP and SO cognates are still responded to faster than controls in L2 LDTs (Dijkstra, 1999; Lemhofer, 2004), in a L2 PMD (Dijkstra, 1999) and in a generalised LDT (Lemhofer, 2004), indicating that the existence of two conflicting phonological representation did not hinder cognate processing when they were not needed to complete the task.*

Research on interlingual homographs (e.g., FOUR), contrary to that of cognate items, have led to inconsistent results. While most of the studies using L2 LDTs have found an inhibitory effect of interlingual homographs over control words (Brenders et al., 2011; Dijkstra, de Bruijn, et al., 2000; Poort & Rodd, 2017; van Heuven et al., 2008; Vanlangendonck et al., 2019) some have failed to find such an effect, (Dijkstra, de Bruijn, et al., 2000; Vanlangendonck et al., 2019, see van Heuven et al., 2008 for a similar result using a generalized LDT), and some have even found a facilitative effect (Dijkstra et al., 1999; Lemhöfer et al., 2004). Those discrepancies, while they still mostly corroborate the cross-language interaction account, poses difficulties when it comes to drawing a clear conclusion as to how interlingual homographs are represented in the bilingual lexicon. Interlingual homographs may share a unique orthographic representation just like cognates do, or they might be represented by two language-specific lexical representation (Dijkstra et al., 1999; Lemhöfer et al., 2004). However, no matter the actual lexical representation configuration (single vs. language-specific representation) interlingual homographs would be associated with two competing semantic representations (see figure 7).

Unsurprisingly, interlingual homographs have been found to be inhibitory in several tasks investigating language membership detection, such as a language decision task and two go/no-go task (Dijkstra, Timmermans, et al., 2000), indicating that interlingual homographs, just like cognates, are associated with two conflicting language nodes. It is also likely that interlingual homographs are associated with two phonological representations, as they are usually pronounced differently depending on the language at hand, e.g., the word FOUR is pronounced \fɔː\ in English and \fuʁ\ in French. This is supported by the inhibitory effect found for interlingual homographs over control words in a naming task (Jared & Szucs, 2002)[9].

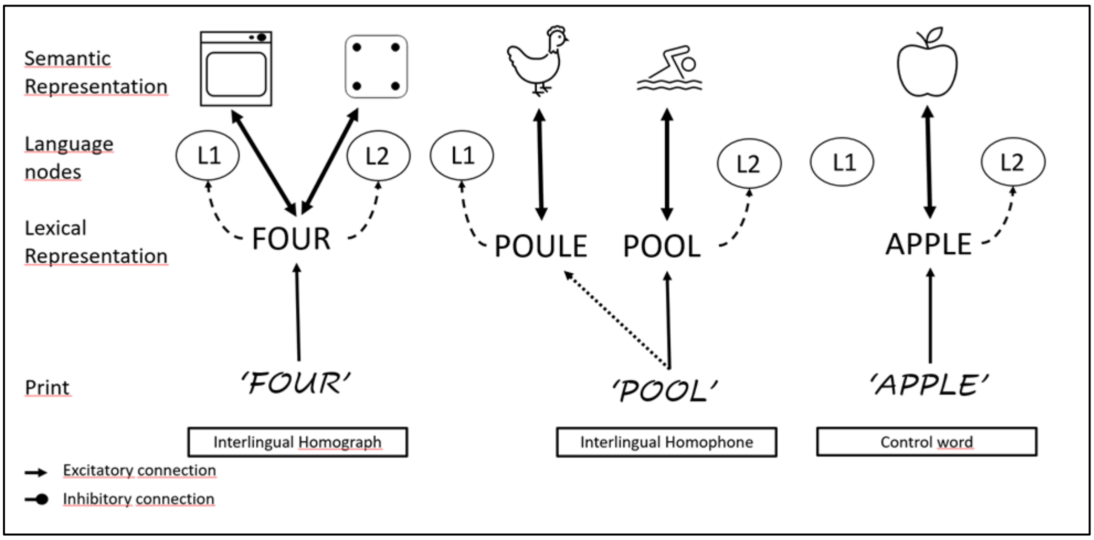


Figure 7: Interlingual homographs and Interlingual homophone representations compared to a control word

Furthermore, as we have seen so far with studies using cognate and interlingual homographs, cross-linguistic phonological overlap has an impact on bilingual visual word processing (Dijkstra et al., 1999; Jared & Szucs, 2002; Lemhöfer et al., 2004) suggesting that

---

[9] *Even though it depends on the status of the language at hand (dominant or not) and on recent experience with the other language.*

bilinguals activate both L1 and L2 phonological representations in a language specific situation. This is confirmed by studies using interlingual homophones, such as POOL (homophone of POULE) for French-English bilinguals. In fact, interlingual homophones have been found to be named faster than control words by young bilingual readers (Jared et al., 2012), to be identified faster than L2 control words by adult bilinguals[10] (Haigh & Jared, 2007; experiment 1), and to elicit reduced N400 compared to control words (Carrasco-Ortiz et al., 2012). When used as primes, interlingual homophones have been found to facilitate the identification of an L2 target word, both when the prime was a direct homophone of the target (e.g., poule-POULE; Brysbaert et al., 1999) or in a more roundabout way, when it was an L2 homophone of the L1 translation of the L2-target[11], (e.g., pool(poule)-HEN; Duyck, 2005; experiment 5) suggesting that bilinguals activate both meanings associated to an interlingual homophone automatically upon presentation.

2. From orthography to phonology: Sublexical aspects of bilingual visual word recognition

To sum up, both complete and partial cross-language form overlap and their impact on visual word recognition in bilinguals massively support the non-selective access hypothesis which is assumed by the most influential models of bilingual visual word recognition (Dijkstra & van Heuven, 2002; van Heuven et al., 1998). When presented with a printed word (e.g FIRE) a bilingual will activate both within-language (e.g English) and cross-language (e.g French) lexical representations, based on form similarity (i.e., orthographic, phonological, semantic) with the target word, those lexical representations will then compete for selection. However,

[10] *At least when the phonology-based strategy was not discouraged.*
[11] *L1 target on the other hand did not benefit from the presence of an L1 homophone of its L2 translation as a prime (e.g. raide(red)-ROUGE, Duyck, 2005; experiment 6).*

cross-language lexical competition is not the only challenge a L2 reader is faced with when acquiring their L2. In fact, the exposition to a second written language will require the reader to cope with cross-language overlap, but also language specificities at the sublexical level. Yet, the sublexical units have been mostly overlooked when considering bilingual visual word recognition. For instance, the sublexical level has not been included in the initial BIA model (except for a letter level), and even when it was included in the BIA+ model, its role in the overall word recognition process has not been much described.

For example, the GPCs used in the 2nd language might overlap with those acquired for L1 processing, e.g., OU might refer to the phoneme \u\ in both French and English, or they might generate some cross-language discrepancies, e.g., the same grapheme OU might also refer to the phoneme \aʊ\ in English, whereas this association is not possible in French. Therefore, the acquisition of an L2 via its orthographic form, not only requires participants to learn new vocabulary but also requires them to update the set of GPCs they had previously acquired for L1 processing.

A. Complexity, consistency and cross-language reading strategies

As we have seen in the first section of this chapter, most European alphabetical systems share a similar mapping principle, but will likely differ in terms of orthographic depth, which refers to both (1) the complexity (or size) of the orthographic units referring to a phoneme and (2) the reliability with which a grapheme will be associated to a phoneme and vice-versa, i.e., on their GPC consistency (Schmalz et al., 2015), resulting in a tuning of reading processes to the particular depth of the language's orthography (Orthographic Depth Hypothesis; Katz & Frost, 1992, see above). Therefore, familiarity with two such orthographic systems that differ on their orthographic depth could raise some unique challenges as to the particular strategies used to read words in both the 1st and/or the 2nd language.

Studies exploring reading in bilinguals with two languages that differ on orthographic depth (i.e., mixed depth bilinguals) have found that the consistency of the 1st language affects later L2 processing (Paulesu et al., 2021) and also that bilinguals of a deep-shallow pair would use different reading strategies depending on the orthographic depth of the language at hand. In fact, mixed-depth bilinguals (e.g., French-German bilinguals) have been shown to engage different neural networks when reading words from the deep language, (or even pseudowords in the deep language context) compared to when they are reading words from the shallow language (or even pseudowords in the shallow language context), suggesting that bilinguals engage different neural networks depending on the orthographic depth of the language at hand (Buetler et al., 2014, 2015). This is further supported by eye-tracking measures that found that a shallow linguistic context induces a more local reading strategy whereas a deep linguistic context induces a more global reading strategy (de León Rodríguez et al., 2016).

Furthermore, mixed-depth bilinguals have been shown to distribute their visual attention more evenly across the letter-string than monolinguals of shallow orthography (Lallier et al., 2013) and readers of two shallow orthographies (Antzaka et al., 2018), even though those differences may depend on the stimulus properties, namely the presence of a complex grapheme (Lallier et al., 2021). These results led to the grain size accommodation hypothesis (Lallier & Carreiras, 2018) which assumes that some cross-language transfer would take place in mixed-depth bilinguals who would therefore use a hybrid orthographic grain size for sublexical processing. This hypothesis leads to two main predictions, (a) mixed-depth bilinguals should rely on smaller grains/units while reading in their deep orthography than monolinguals of this same deep orthography would, (b) those same bilinguals should rely on larger grains/units while reading in their shallow orthography than monolinguals of this same shallow orthography would.

B. Pseudohomophones, GPC Congruency and automatic cross-language activation

Another source of differences across languages comes from the GPC identity. In fact, while some GPCs are shared across languages, e.g., the pronunciation of the letter I found in the English word FISH (mostly) overlaps with its French pronunciation in BISE, others will not overlap, e.g., the same I in the English word TIME. A first line of enquiry into those cross-language GPC incongruencies consisted in testing whether a bilingual used one or both set of GPCs in a language-specific context. One of the first study providing some information on the subject, even though it wasn't its main objective (Nas, 1983), found that L1 words e.g., BEELD (Dutch word for PICTURE) homophone to an L2 word e.g., BUILD, if L2 (English) GPCs were used, were more difficult to reject than control pseudowords when performing an L2 LDT (experiment 1). They also found that pseudohomophones, which are pseudoword that were constructed to sound like an L1 word when using L2 GPCs e.g., DROAL (homophone to the French word DRÔLE meaning funny) were slower to reject than classical pseudowords when performing an L2 LDT (experiment 2).

Subsequent studies built on this by using pseudohomophones as primes or as target items in several LDTs. When used as primes both L2-pseudohomophone of an L1 target (e.g., VARE-vert; Van Wijnendaele & Brysbaert, 2002), and L1-pseudohomophone of an L2 target facilitated word recognition (e.g., GRINE-green; Brysbaert, 2001). Recognition of a target word could also be facilitated by the presentation of a pseudohomophone of its translation equivalent as a prime (e.g., GREAN-vert or VAIR-green) or even by a pseudohomophone of a related word in the other language (e.g., BLOU-green), although this latter effect was significant only with L2 targets (Duyck, 2005). When used as target items on the other hand, pseudohomophones seem to be rather inhibitory. In fact in a series of L2 LDTs, Commissaire, Duncan, et al., 2019, found that L1-pseudohomophones of an L2 word (e.g., GRINE for the word GREEN) but also L2-pseudohomophones of an L2 word (e.g., DREEM for the word DREAM) were rejected slower than classical pseudowords, albeit this was significant only with an older group of L2-learners. L2-pseudohomophone of an L1 word (e.g., DROAL for the word DRÔLE meaning funny) were inhibitory compared to pseudowords, this time for all participant groups. Taken together those results indicate that despite being in a language specific context, both within and cross-language GPCs are activated by bilinguals and young L2-learners.

Consequently, when investigating the GPC set of an individual familiar with two languages it is important to take into account a cross language variation of GPC consistency variable which we will refer to as GPC-Congruency from now on. GPCs that are shared across languages, (e.g., OU=\u\ for French-English speakers) will be referred to as congruent GPCs, while those that differ across languages, (e.g., OU= \aʊ\) will be referred to as incongruent GPCs.

Only two studies have explored the direct impact of GPC congruency (Commissaire et al., 2014; Hevia-Tuero et al., 2021). They have both done so by using a letter detection task (in which a participant had to indicate whether a particular letter was included in a target word), with words containing a congruent GPC (e.g., the “a” in PARK), or containing an incongruent GPC (e.g., “a” in GAME), yet they found a critically different result pattern. While Commissaire et al., working with young and older French learners of English (middle-school and university students), found that the facilitative effect of congruency over reaction times was present only for adult learners of English, Hevia-Tuero et al., found that the facilitative effect of congruency was found on mouse trajectories but only with younger L2 learners (7 y.o).

The fact that an effect of congruency was found, even though not following the same pattern in each experiment still suggests that L2-learners activate both L1 and L2 phonological codes while identifying a word in their non-dominant language. However, while results from Commissaire et al., suggest that L2 GPCs become strong enough to compete with L1 GPCs only after a certain amount of L2 exposure, results from Hevia-Tuero et al., suggests that with longer exposition to L2, bilinguals were more efficient at inhibiting the cross-lingual interferences between the two sets of GPCs. However, several differences that could explain the discrepancy have been highlighted by the second study; first, participants from Hevia-Tuero et al., were exposed to L2 earlier than those from Commissaire et al., and while Commissaire et al., participants knew two deep languages, Hevia-Tuero et al., participants knew a shallow orthography and a deep orthography, and this may have modulated their sensitivity to GPC congruency. Nevertheless, both investigation into GPC congruency confirmed the earlier finding that individuals familiar with two languages will activate every GPCs they know even in a language-specific context, the particular developmental dynamic around this effect on the other hand, still need to be uncovered.

C. Intermediate summary

The familiarity with two written languages (in the case of two European languages using an alphabet derived from the roman alphabet) implies the knowledge of two sets of GPCs which might differ on several aspects, such as within language depth (either on complexity or on consistency), and most likely cross-language (in)consistency, i.e., GPC (in)congruency. Those discrepancies may impact bilingual visual word processing in various ways. First, the difference in orthographic depth across the two languages known to a bilingual implies that they will need to modulate the reading strategy they will adopt in each language (Buetler et al., 2014, 2015; de León Rodríguez et al., 2016) but also results in the use of an hybrid grain-size by bilingual readers in both their languages (i.e., grain-size accommodation hypothesis, Lallier & Carreiras, 2018). Second, and most importantly, studies investigating the discrepancies as to the particular Grapheme-to-Phoneme associations have showed that bilinguals will activate both target and non-target GPCs even in a language specific context (Commissaire et al., 2014; Hevia-Tuero et al., 2021; Sauval et al., 2017). This is evidence that cross-language interactions also happen at the sublexical level and it shows that the phonological side of the sublexical level needs to be taken into account when investigating L2-visual word recognition.

III. Conclusion

In this present chapter, we have gathered some evidence suggesting that the acquisition of a second language, particularly when imparted through the school setting and thus relying heavily on prints, implies that learners make some adjustments to the reading and visual word recognition processes they had previously developed for native-language reading/processing.

In fact, since monolingual reading is tuned to the orthographic depth of the reader's language (Katz & Frost, 1992), bilingual readers are encouraged to adapt their reading strategies to the language being presented to them (Buetler et al., 2014, 2015; de León

Rodríguez et al., 2016), but also to adapt the size of the units they will consider for sublexical processing (Lallier & Carreiras, 2018). Therefore, mixed-depth bilinguals (e.g., English-Italian) will favor local reading strategies when presented with their shallow orthography (Italian) and global strategies when presented with their deep orthography (English). Besides, the unit adopted by mixed-depth bilinguals for sublexical processing in both their languages is supposed to be an hybrid grain size, not quite as small as the grain size used in the shallow orthography (Italian) but also not as large as the one used in the deep orthography (English).

Considering visual word recognition *per se*, both monolingual and bilingual processes imply the automatic activation of lexical and sublexical representations based on form similarity with the target item. Yet, it appears that bilinguals are unable to select the language in which they will call on their lexicon, resulting in some cross-language interaction effects such as cognate facilitative effect (Dijkstra et al., 1999; Dijkstra, Miwa, et al., 2010; Lemhöfer et al., 2004; Poort et al., 2016; Poort & Rodd, 2017; Vanlangendonck et al., 2019) and interlingual homograph inhibitory effect (Brenders et al., 2011; Dijkstra, de Bruijn, et al., 2000; Poort & Rodd, 2017; van Heuven et al., 2008; Vanlangendonck et al., 2019). Thus it seems that upon presentation of a printed word, e.g., BISE (French word for kisses), a bilingual will not only activate within language GPCs, e.g., I = \i\, and lexical representations, e.g., MISE, but also cross-language GPCs, e.g., I= \ai\, and lexical representations, e.g., BIKE, and those representations will influence target word recognition.

CHAPTER II / How do bilinguals manage cross-language competition

In the previous chapter we have seen that experimental evidence largely supported the non-selective access hypothesis which is assumed by the most influential models of bilingual visual word recognition (Dijkstra & van Heuven, 2002; van Heuven et al., 1998). When presented with a printed word (e.g., FIRE) a bilingual will activate both within-language (e.g., English) and cross-language (e.g., French) lexical, and sublexical, representations, based on form similarity (i.e., orthographic, phonological, semantic) with the target word, those representations will then compete for selection. In some rare cases these cross-lingual interactions may help processing (cognate effect; Dijkstra et al., 1999), but most of the time it hinders visual word recognition (Cross-language Neighborhood Density effect; van Heuven et al., 1998; interlingual homograph effect; Dijkstra, de Bruijn, et al., 2000). Consequently, the non-selective access hypothesis implies the existence of a language control mechanism that would allow bilinguals to sort through the lexical activations and select the correct representation according to the linguistic context at hand.

In this chapter we aim to investigate how bilinguals manage this increased competition, by first considering the experimental evidence of bilingual language control and how they were used by the BIA and BIA+ to each build up their own account of the mechanism. Besides, since the linguistic context is assumed to be a source of modulation of the bilingual lexical access by the BIA+ model, we will explore evidence of its impact on target word recognition in a second section. Then in a third section we will review the literature surrounding a sublexical variable which results from the familiarity with two written languages, namely the orthographic markedness. This variable and particularly its potential impact over the bilingual lexical access has indeed recently sparked some interest.

I. Is there such a thing as a language control mechanism?

Having thoroughly investigated the potential control bilinguals may exert over their lexical access, the speech processing literature has provided the first description of the particular mechanism involved in such a control: the inhibitory control model (IC; Green, 1997)[12]. In this model, the language control comes from the interaction of task schemas, that set the strategy used to perform a particular task, and language nodes, which collect language membership information from lexical representations. Those language nodes then impact lexical activation through top-down activations.

It may be hypothesized that the same system is used for visual word recognition (Peeters et al., 2019). However, since visual word recognition is about perception rather than production, a strong language control such as described in the IC model, might not be applicable. In fact, as described in the first part of this chapter, bilinguals are not able to inhibit lexical representations based on language membership from the very beginning. Thus, the language control locus and particular mechanism involved for visual word recognition needs

[12] *More recently, a neural network of this language control mechanism was even proposed (Abutalebi & Green, 2016).*

to be specified. But first we need to find evidence that some language control does indeed happen on visual word recognition.

1. Evidence for language control in visual word recognition- list composition effect

A first evidence for language control over visual word recognition comes from Grainger and Beauvillain, (1987) who found that word recognition was slower when the items were presented in language-mixed lists, containing words from both languages, than in a language-specific list, containing words from only one language. This suggests that a constant shift in focus from one language to the other is onerous for bilinguals and that stimulus list composition may impact bilingual visual word recognition.

Interestingly, the effects described earlier as evidence for a non-selective lexical access are also modulated by stimulus list composition[13]. For instance, the facilitative interlingual homophone effect can be reversed when some pseudohomophone are introduced in the list, discouraging a phonology based strategy (Haigh & Jared, 2007; Mcquade, 1981). Besides, the inhibitory effect of interlingual homograph just like the cognate facilitation effect are sensitive to the presence of L1 control words that need to be rejected (Dijkstra, Bruijn, et al., 2000; van Heuven et al., 2008; Vanlangendonck et al., 2019).

This is illustrated quite well in the study from Dijkstra, Bruijn, et al., (2000), in which participants were presented with a list of L2 words containing interlingual homographs, cognates and other L2 words (during the whole duration of the experiment) but also some L1 words (second part of the experiment only) in a L2 LDT. The inhibitory effect of interlingual

[13] *Stimulus list composition, and particularly the presence of hermit words (i.e. words without any neighbors in either languages; CHALK) also impacted the Cross-language neighborhood density effect (Mulder et al., 2018)*

homographs emerged only in the second part of the experiment, which is to say as soon as some L1 words that needed to be rejected by the participants were introduced. Similarly, Vanlangendonck et al., (2019), using two separate lists, only one of them containing L1 words, obtained similar results to those of Dijkstra, Bruijn, et al., (2000). More interestingly they found that the cognate effect was facilitatory in the L2-only list but became inhibitory in the mixed language list. Thus, it appears that by increasing the response competition, the authors were able to exhibit the interlingual homograph effect which is otherwise not very consistent, and to suppress (Dijkstra, Bruijn, et al., 2000) or completely reverse the cognate effect which is generally quite consistent (Vanlangendonck et al., 2019, see table 1 for clarifications)[14]. This is in line with results from van Heuven et al., (2008), who found an inhibitory effect of interlingual homograph when performing an L2 LDT inducing response conflict, but not in a Generalised LDT in which there was no response conflict.

[14] *Interestingly, Peeters et al., (2019) by contrasting Interlingual Homographs to control words and by contrasting a pure list (list 1 from Vanlangendonck et al.,) with a mixed list (list 2), uncovered the network involved in the language control for a visual word recognition task, which is close to the one uncovered for language control in a production task (Abutalebi & Green, 2016).*

Table 1: Stimulus list composition and the cross-language interaction effects associated in Dijkstra, Bruijn et al., 2000 and in Vanlangendonck et al., 2019's studies including several Lexical Decision Tasks

<table>
<tr><th>L2 LDT from :</th><th>List</th><th>Items</th><th>Response associated</th><th>Effect (vs. L2 control)</th></tr>
<tr><td rowspan="4">Dijsktra et al., 2000a</td><td rowspan="2">1st part</td><td>Interlingual Homographs
Cognates
L2 controls</td><td>YES</td><td rowspan="2">No effect of interlingual homographs</td></tr>
<tr><td>Pseudowords</td><td>NO</td></tr>
<tr><td rowspan="2">2nd part</td><td>Interlingual Homographs
Cognates
L2 controls</td><td>YES</td><td rowspan="2">Inhibitory effect of interlingual homographs</td></tr>
<tr><td>L1 controls
Pseudowords</td><td>NO</td></tr>
<tr><td rowspan="4">Vanlangendonck et al., 2019</td><td rowspan="2">1st list</td><td>Interlingual Homographs
Cognates
L2 controls</td><td>YES</td><td rowspan="2">No effect of interlingual homographs
&
Facilitative effect of cognates</td></tr>
<tr><td>Pseudowords</td><td>NO</td></tr>
<tr><td rowspan="2">2nd list</td><td>Interlingual Homographs
Cognates
L2 controls</td><td>YES</td><td rowspan="2">Inhibitory effect of interlingual homographs
&
Inhibitory effect of cognates</td></tr>
<tr><td>L1 controls
Pseudowords</td><td>NO</td></tr>
</table>

2. Evidence for language control in visual word recognition- language switch cost

Thus, introducing words from the non-target language to a list of words has an impact on bilingual word processing, particularly when those "intrusions" lead to response conflicts (Dijkstra, Bruijn, et al., 2000; van Heuven et al., 2008; Vanlangendonck et al., 2019). Still, mixing L1 and L2 words in a single list also ask the question of the particular trial-to-trial processing cost of switching from one language to the next. This is referred to as the language switching cost and results from the comparison of language-repeated trials (trial *n* and *n+1* involve the same language) and language-switched trials (trial *n+1* involves another language compared to trial *n,* see figure 8).

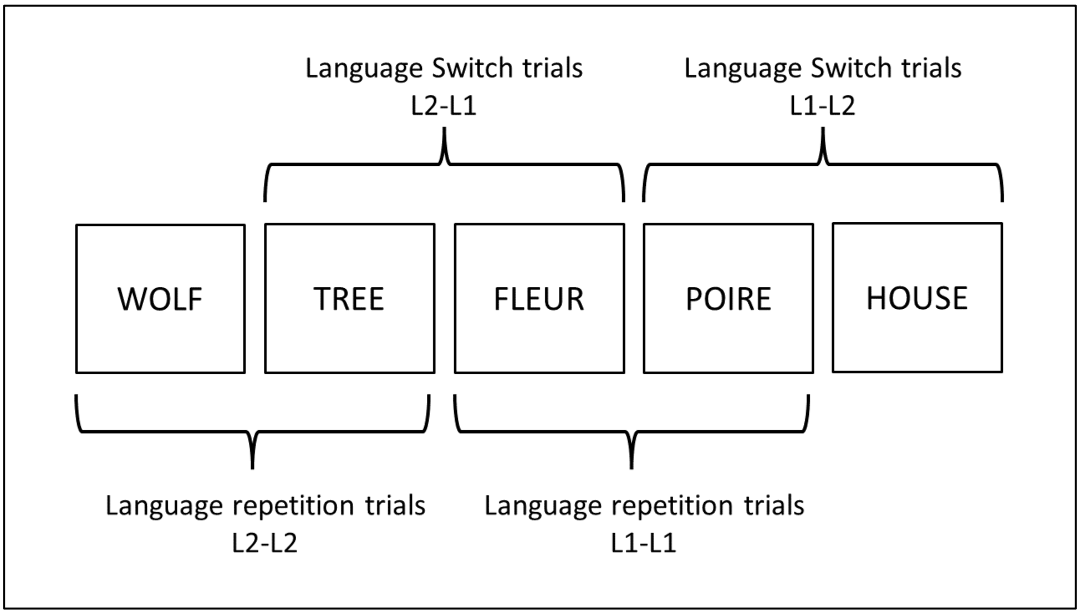


Figure 8: Illustration of a language switching paradigm for a French-English bilingual

The switch cost, i.e., longer reaction times for language-switched trials than for language-repeated trials, has been consistently found on language production studies such as naming studies (Meuter & Allport, 1999; Olson, 2016; Timmer et al., 2019), and also on language comprehension studies, through lexical decision tasks (Grainger & Beauvillain, 1987;

Orfanidou & Sumner, 2005; Thomas & Allport, 2000; von Studnitz & Green, 1997), and categorization tasks (Declerck & Grainger, 2017; Macizo et al., 2012; von Studnitz & Green, 2002) where it has been interpreted as reflecting the inhibition a bilingual exerts upon non-target language representations when processing words in a target language. Interestingly, several of those studies have exhibited an asymmetrical switch cost (Meuter & Allport, 1999; Timmer et al., 2019 but see Gade et al., 2021), the direction of which seems to be dictated by the linguistic context (Timmer et al., 2019) indicating that the amount of activation of a particular language depends on contextual information such as the extent to which this language is used in the task at hand[15].

Furthermore, studies exploring the switch cost on visual word recognition (Orfanidou & Sumner, 2005; Thomas & Allport, 2000; von Studnitz & Green, 1997, 2002), and particularly the study from von Studnitz & Green, (1997), argue for a source of language control outside the bilingual lexicon, at the response level. In fact, von Studnitz & Green, found a more important switch cost in a specific lexical decision task (participants had to indicate whether the item belonged to the language of the trial at hand) than in a general decision task (participants had to indicate if the item was a word in either known language), indicating that the language switch cost should therefore be associated to a task system that would specify strategies and set decision thresholds depending on the task requirements. Following this assumption, in a general decision task, both first and second language could be considered as target languages by the task system, and would be attributed the same threshold, whereas in an L2-specific LDT, the first language might be considered a non-target language, and

[15] *This is in line with the adaptive control hypothesis of Green & Abutalebi, (2013), coming from speech processing literature.*

therefore would be attributed a higher activation threshold than the second language (target language), resulting in a more important switching cost in the L2-specific LDT.

3. Accounting for language control in the BIA and BIA+ models

The stimulus list composition effects and the language-switch cost described above constitute evidence for the existence of a language control mechanism that should be included in models of bilingual visual word recognition. Both the BIA and BIA+ present some attempt at implementing a language control mechanism. Yet each model does so quite differently from the other. In the first BIA model, the language control is attributed to the language nodes, that are able to inhibit lexical representations according to the language membership information they collect (e.g., when activated, the L1 language node can inhibit L2 lexical representations). However, although the later BIA+ model kept the language nodes, they lost their direct top-down connections with the lexical representations. This time the language control was attributed to the task system (absent from the BIA model), which, like in the IC model, was used to implement strategies and set some task schemas according to the task demands. Thus, following the BIA+ model, bilinguals can only control their decision (e.g., by setting different decision threshold for L1 and L2 depending on task requirements), but not their lexical access.

Interestingly, the distinction between the task decision system and the word identification system introduced in the BIA+ model is accompanied by a distinction between linguistic context (e.g., sentence context) and non-linguistic context (e.g., task requirements). Moreover, the BIA+ model makes some assumptions as to the interaction between those two sources of contextual influences and the two systems it described. According to this model, while the task decision system could be influenced by the non-linguistic context, the word identification system and therefore the lexical access, could only be impacted by the linguistic context.

More precisely the BIA+ model assumes that the word recognition system is sensitive to syntactic and semantic context information: “such linguistic context (sentence) information may exert serious constraints on the degree of language selective access that may be observed”(Dijkstra & van Heuven, 2002, p.187). Hence, following the BIA+ model, even though bilinguals are not able to actively control their lexical access through top-down connections between the task system and the word identification system, this (bilingual) lexical access might be modulated by linguistic context information directly in the word identification system.

4. intermediate summary

The shift in focus from one language to another has an impact on bilingual visual word recognition. For instance, the introduction of words from the non-target language to the list of items a bilingual has to process modulates the cross-language interaction effects usually found with this population, particularly when the response competition is increased by this manipulation (Dijkstra, Bruijn, et al., 2000). Taken together with the trial-to-trial processing cost of switching from one language to the next in language mixed list (i.e., language switch cost, von Studnitz & Green, 1997), those results hint at the existence of a post-lexical mechanism that allows bilinguals to sort through their lexical activations.

To account for this apparent language control mechanism, the two BIA models made different assumptions. While the BIA model conferred their language membership nodes direct inhibitory connections to lexical representations, the later BIA+ model removed the direct link between language membership nodes and lexical representations but implemented a task system that could exert some control over the decisions made by bilinguals while keeping the lexical access completely non-selective. Thus, it appears that even though bilinguals do not select the language in which they will access their lexicon, they are able to sort through their lexical activations to select the correct representation according to the task at hand. Interestingly, even though the BIA+ model adopt a strong non-selective access position where bilinguals can control their decisions but not their lexical access, the same model also assumes that the linguistic context provided by a sentence, might modulate the degree of non-selectivity adopted by a bilingual when consulting their lexicon.

II. Sources of lexical access modulation: sentence context

Since the BIA+ model assumes that the linguistic context may modulate the bilingual lexical access, we might expect cross-language interferences to be reduced when a linguistic context, such as a sentence, is provided. Therefore, as we are going to see, a large part of the studies exploring this assumption have done so by trying to replicate, or rather to not replicate, the cognate and the interlingual homograph effect while including those words in a sentence.

The cognate effect was replicated in a variety of tasks including an L2 lexical decision task on the last word of a sentence (e.g., They were drawing a LION/FROG; Duyck et al., 2007) and a naming task with L1 and L2 sentences (Gullifer et al., 2013). It was also found on eye-tracking measures with L1 sentences (Van Assche et al., 2009) and with L2 sentences (Bultena et al., 2014; Duyck et al., 2007). Moreover, even though significantly less studies have investigated the interlingual homograph effect, it was replicated when performing an L2 LDT on the last word of a sentence (e.g., They picked up the dirty COIN/SOCK), at least when there was an overlap on word class across languages (Baten et al., 2011). Therefore, at first glance, the linguistic context provided by a sentence appears insufficient to narrow down the lexical access to the target language. Yet, following studies on the topic brought an interesting variable to the table, namely the semantic context implicated in sentence processing.

1. Investigating the impact of the linguistic context on bilingual lexical access through ambiguous word recognition

To do that, some studies have manipulated the semantic constraints implied by the sentences they used. The semantic constraint refers to the relative ease with which readers will be able to predict a target word based on the semantic information included in a sentence. In a low semantic constraint sentence, readers will not be able to easily/accurately predict the target word based on the semantic cues: "They watched an animated movie about a LION". In a high semantic constraint sentence however, predictions on the target item could more easily/accurately be made: "They watched a wildlife documentary about a LION". The monolingual literature has already shown that a target word is more easily processed in a semantically congruent context than in a neutral context (Simpson & Krueger, 1991), meaning that a more constraining context restrained the lexical search. Bilinguals are also able to make some predictions on the upcoming words based on semantic context (Foucart et al., 2014; Zirnstein et al., 2018 but see Martin et al., 2013). Therefore, bilinguals may restrain lexical search to representations that fit the semantic constraints, and this may successfully reduce the impact of cross-language interaction effects.

Not surprisingly, the cognate effect was massively reproduced in low semantic constraint sentences across a large array of tasks, since it was found in an L2 naming task (Gullifer et al., 2013; Schwartz & Kroll, 2006), in L1 and L2 picture naming tasks (Starreveld et al., 2014), in L2 LDTs (Dijkstra et al., 2015; Van Assche et al., 2009), in an L1 LDT (Dijkstra et al., 2015) and in backward and forward translation tasks (van Hell & de Groot, 2008). It was also found on eye-tracking measures (Libben & Titone, 2009; Pivneva et al., 2014; Titone et al., 2011; Van Assche et al., 2009) where cognate words had a facilitative impact on both early and late eye-tracking measures in an L2 reading task (Libben & Titone, 2009; Pivneva et al., 2014; Van Assche et al., 2009), which was reduced to early eye-tracking measures when reading in L1 (Titone et al., 2011)[16]. These results confirm that linguistic cues given by a low semantic constraint sentence was not enough to restrain the bilingual lexical access.

Meanwhile, in high semantic constraint sentences, the cognate effect tended to disappear in an L2 naming task (Schwartz & Kroll, 2006), and in an L2 LDT (van Hell & de Groot, 2008)[17]. It was also reduced in an L1 LDT (Dijkstra et al., 2015) and in translation tasks (van Hell & de Groot, 2008) and it was even restricted to L2 naming in a L1 and L2 picture naming task (Schwartz & Kroll, 2006). Besides, the cognate effect was also significantly reduced in an L1 sentence reading task (Titone et al., 2011). These results are in line with the hypothesis we introduced above, stating that a more constraining semantic context could restrain the lexical search. However, cognates still impacted early and late eye-tracking measures in L2 sentence reading (Pivneva et al., 2014; Van Assche et al., 2009). The only study that found a modulation of the cognate effect on eye-tracking measures by the high semantic constraints comes from Libben and Titone, (2009) who found that the cognate status impacted only early eye-tracking

[16] *Although the cognate impact on eye-tracking measures was modulated by the presence or absence of L2 filler sentences.*
[17] *However, the cognate effect could be reproduced in a language mixed context (Dijkstra et al., 2015)*

measures[18]. As concluded by a recent meta-analysis (Lauro & Schwartz, 2017), the results described above suggest that even a highly constraining semantic context do not turn the bilingual lexical access into a selective one, but it can reduce its non-selectivity depending on the language at hand (i.e., L2 vs. L1) and the task requirements (i.e., overt response vs. eye-tracking measures).

Still, cognates are a very particular type of cross-language homographs that overlap not only on their orthographic form but also on their meaning across languages. Hence a bilingual will not benefit from narrowing its lexical access to the target language in order to process a cognate. On the other hand, since they do not share their meaning across languages, Interlingual Homographs are more likely to benefit from a linguistic context. For instance, knowing that the upcoming word will refer to a sort of currency may help bilinguals to access the appropriate L2 meaning of the word COIN (meaning corner in French), compared to a neutral context where both L1 and L2 meanings are most likely to be activated at the same time.

A few studies have explored the impact of semantic constraints on the interlingual homograph effect. They reported that the manipulation of the semantic constraint did not impact the interlingual homograph effect on overt responses, since its absence in an L2 naming task (Schwartz & Kroll, 2006), and its presence in an L1 LDT (Lagrou et al., 2015), did not depend on the strength of the semantic constraints.

Interestingly, interlingual homographs impacted only late eye-tracking measures in both low and high semantic constraint contexts, in an L1 reading task (Titone et al., 2011),

---

[18] *Although, this result must be taken with caution, as Pivneva et al., (2014) using the same material, did found an impact of cognate status on both measures.*

suggesting that the L2 meaning of an interlingual homograph is activated quite late during L1 processing, no matter the semantic constraints. Even more interestingly, in an L2 reading task, interlingual homographs impacted both early and late measures in low semantic constraints but could be restricted to early measures in high semantic constraints (Libben & Titone, 2009)[19]. This pattern led the authors to interpret their results as indicating that the non-target L1 meaning is activated at early stages, when reading in L2, but can be inhibited by a highly constraining context so as not to impact late reading stages. Thus, semantic constraints seem to only have a limited impact over cross-language interactions, even when considering interlingual homographs which are more likely to benefit from semantic cues provided by a sentence context.

2. Investigating the impact of the linguistic context on bilingual lexical access through non-ambiguous word recognition

As we have seen so far, the literature on the impact of a linguistic context on the bilingual lexical access is in line with a strong non-selective lexical access. Yet, the use of language-ambiguous words (cognates and interlingual homographs, whose form is shared across languages) in every study reviewed until now may have encouraged the persistence of a language non-selective process, even in a linguistic context such as embedding these words in a sentence context. Therefore, the use of *language non-ambiguous words* (in a sentence) may be more suited to induce a language selective access guided by the linguistic context.

Following this reasoning, some studies have compared the processing of a monolingual sentence “Peter ate an apple” to the processing of a sentence containing a code-switch: “Peter

---

[19] *Although this may interact with the participant’s executive control abilities Pivneva et al., 2014.*

ate une pomme". The results showed that the language switch still exhibited a cost on ERP recordings (Moreno et al., 2002, experiment 2), self-paced reading (Bultena et al., 2015a; Litcofsky & Van Hell, 2017) and with a shadowing task (Bultena et al., 2015b), when reading L1 sentences (Moreno et al., 2002) or L2 sentences (Van Der Meij et al., 2011). However, the switch cost was larger from L1 to L2 (Bultena et al., 2015b, 2015a; Hoversten & Traxler, 2020; Litcofsky & Van Hell, 2017) than it was from L2 to L1[20]. This asymmetric switching cost may reflect the greater difficulty experienced by bilinguals when suppressing L1 activations in an L2 context (vs. when suppressing L2 activations in an L1 context). Still, the existence of a language switch cost in those sentences implies that the non-target language was less activated than the target language was. Thus, a certain amount of control might be exerted by bilinguals on their lexical access, depending on the linguistic context they find themselves in.

Furthermore, in an interesting variation of this paradigm, Hoversten and Traxler, (2020), compared the processing of a sentence "we saw that his ___ had a horrible scar" containing (1) a non-switched word "hand", (2) a code-switched word "mano" (meaning hand in Spanish) or (3) a pseudoword "erva" (see table 2 for a summary). When switching from L1 to L2, the code-switch word was always processed better than the pseudoword and worse than the non-switch word on all eye-tracking measures. Meaning that it was less easily accessed than the non-switch word but still accessible compared to the pseudoword. Moreover, in a second experiment, the authors introduced the code-switch only as a preview condition, so that participants were not aware of any introduction from the non-target

---

[20] *which is the reverse pattern of what was found in the isolated word study by Meuter and Allport, (1999).*

language. Hence, the preview words, either non-switched (valid: hand, invalid: boss, pseudoword: shup), or code-switched (invalid: jefe (meaning boss), pseudoword: erva) were replaced by the target word "hand" before the participants' gaze landed on it. This time, code-switched words did not differ from pseudowords, indicating that in a fully language-specific context, the non-target language was as inaccessible to bilinguals as pseudowords were. These results, presented in table 2 below, were interpreted by the authors, as reflecting a partially selective lexical access during reading. Following this hypothesis, both target and non-target language lexical representations are activated during target language processing, accounting for non-target language influences, albeit at different degrees depending on the linguistic context. As a consequence, in a fully monolingual context, the non-target language lexical representations were not accessible to the bilingual[21].

Table 2: Examples of sentences used in Hoversten & Traxler, (2020) and the results obtained on skip rate in a sentence reading task (NS= Non-switched; CS=Code-Switched; PSW=Pseudoword)..

Experiment	**Word type**	**Example**	**Results**
Experiment 1	Non-Switched	We saw that his HAND had a horrible scar	NS>**CS>PSW** on skip rate
	Code Switched	We saw that his MANO had a horrible scar	
	Pseudoword	We saw that his ERVA had a horrible scar	
Experiment 2	Identical	We saw that his HAND/hand had a horrible scar	NS>**CS=PSW** on skip rate
	Non-Switched	We saw that his BOSS/hand had a horrible scar	
	Code Switched	We saw that his MANO/hand had a horrible scar	
	Pseudoword	We saw that his ERVA/hand had a horrible scar	

[21] *This is supported by findings from Elston-Güttler et al., (2005), where the two meanings associated with an interlingual homograph could be activated only if the testing condition was bilingual enough and for a short time.*

3. Intermediate summary

Studies inserting language ambiguous words in sentences have not been successful at suppressing cross-language interactions. Cognate and Interlingual homograph effects can only ever be reduced or delayed by the semantic manipulations introduced in the sentence but they still have an impact on visual word recognition when included in sentences with high-semantic context (Lauro & Schwartz, 2017; Schwartz & Kroll, 2006; Titone et al., 2011; van Hell & de Groot, 2008). This pattern of result favours a weak version of the BIA+ assumption that the linguistic context might modulate the non-selectivity of the bilingual lexical access, since the sentence context only has a limited impact over the apparition of cross-language interactions.

Yet recent studies have been able to replicate the language switch cost in sentence contexts indicating that depending on the linguistic context, non-target language representations are made less accessible than target language representations. Even more importantly, when the language switch was not perceptible by participants, non-target language representations behaved just like pseudowords did in the same situation, suggesting that they were not accessible at all (Hoversten & Traxler, 2020). This has been interpreted by the authors as evidence for a partially selective bilingual lexical access where the languages known by a bilingual will be activated more or less according to the linguistic context at hand, and where the non-target language could only be shut off in a fully monolingual context containing only language non-ambiguous words (in line with the language mode hypothesis; Grosjean, (2001).

III. Sources of lexical access modulation: orthographic markedness

Even if bilinguals do not access their lexicon based on language membership information, it is critical for them to be able to retrieve this information, seeing as this information could help them resolve ambiguities such as that introduced by interlingual homographs. Despite their differences, both BIA models assume that language membership information becomes available only post-lexically, after complete identification of the word, preventing it from shutting off completely the non-target language representations. Yet, recent studies have shown that languague membership information may be retrieved pre-lexically (Casaponsa et al., 2019; Hoversten et al., 2015, 2017). Those studies claim that the sublexical level, the level below the word unit level, containing smaller units such as letters, bi- and trigrams and graphemes, could therefore be an interesting source of language membership information, mostly through the use of orthographic markers (see below). More importantly, this language membership cue might be retrieved early enough to guide the bilingual lexical access.

1. The orthographic markedness, an interesting cue for language detection

When acquiring their 2nd written language, some bilinguals are required to learn a new script (e.g., English-Japanese) or a new alphabet (e.g., English-Greek), providing them with a low-level visual information that they can use to determine to which language they are being exposed. On the contrary, other bilinguals are introduced to a second language that shares most or even all of its alphabet letters with their native language (e.g., English-French). Yet, the coexistence of the two written languages will most likely result in the emergence of language-specific orthographic regularities, revered to as orthographic markers. Interestingly, those language-specific orthographic regularities are integrated after only short exposure to a new language (Chetail, 2017; Lemhofer et Radach, 2009), and may provide bilinguals with some sublexical language membership cues.

For instance, the most obvious type of orthographic markers are language-specific letters, the ß being found in German does not exist in French nor in English, and therefore may "mark" the word WEIß (white) as belonging to German. Other markers may come from letter combinations, a bi- or trigram (e.g., KN) might be legal in one language, (e.g., English) but not in the other language (e.g., French), known by a bilingual. Besides, orthographic markers can also be conceptualized using a more continuous approach, according to which a letter combination that is much more frequent in one language compared to the other language known by a bilingual (e.g., the EU bigram is more frequent in French than it is in English), can also constitute an orthographic marker[22] (see Table 3 for a summary).

Table 3: Summary of orthographic markers' type

Orthographic markers	**Definition**	**Example**
Language specific letters	A letter that can be found in only one of the languages known by an individual	The letter ß for a French-German bilingual
Language specific bigrams	A bigram (group of two letters) that is legal in only one of the languages known by an individual	The bigram KN for a French-English bilingual
Frequency-based markers	A bigram that is much more frequent in one of the languages (compared to the other languages) known by an individual	The bigram EU for an English-French bilingual

[22] *Orthographic markers are the product of the comparison of the two languages known by a bilingual, the same bigram will not constitute a marker for every language pairs, the KN bigram for example, will not constitute a marker for German-English bilinguals because it occurs in both languages at the same extent (KNIFE and KNOPF).*

A. Switch cost

Historically, the first studies using orthographic markedness as a variable were studies investigating the language mixing effect and the language switching cost (Grainger & Beauvillain, 1987; Orfanidou & Sumner, 2005; Thomas & Allport, 2000; von Studnitz & Green, 1997). Orthographic markedness was then used as a way to uncover the locus of language control, since it was first thought that an impact of orthographic markedness upon switch cost would constitute unequivocal evidence in favor of the word recognition system as a source of language control (vs. the task system). But as we are about to see both the results and the conclusions as to the nature of the language control derived from those studies are unclear.

While these studies (using both generalised or language-specific LDTs), reliably found that Marked words such as KNIFE were recognised faster than Non-Marked words such as CARE (for a French-English speaker; Casaponsa et al., 2019; Grainger & Beauvillain, 1987; Orfanidou & Sumner, 2005; Thomas & Allport, 2000), and that Marked pseudowords (e.g., KNIDE) were rejected slower than Non-Marked pseudowords (e.g., VARE; von Studnitz & Green, 1997), it remains unclear whether the presence of orthographic markers influenced the switching cost. Although some studies found larger (or exclusive) switching cost for Non-Marked items than for Marked items (Grainger & Beauvillain, 1987; Orfanidou & Sumner, 2005), other studies have found that orthographic markedness did not interact with switching cost (Orfanidou & Sumner, 2005; Thomas & Allport, 2000; von Studnitz & Green, 1997) and others have even found larger switch-cost with Marked items than with Non-Marked items (Casaponsa et al., 2019)[23].

[23] *Although studies that found this larger switch-cost for marked items compared to non-marked items used a masked priming paradigm and therefore the language switch was not conscious.*

The 1987's study by Grainger and Beauvillain can be considered to be the very first one to investigate the question. Using a generalized lexical decision task, they found a significant mixing cost (i.e., longer reaction times for words in a language mixed list than in a language-specific list) that was significant only for the Non-Marked word condition (i.e., when the list contained only Non-Marked items), which they interpreted as reflecting the fact that the word recognition system was the source of language control. Following their interpretation, in Non-Marked list, language membership information was too scarce to inhibit non-target language representations resulting in a language switch cost, cost that did not emerge when language markers were present.

However, a subsequent study by Thomas and Allport (2000), using a similar task, did not replicate this interaction between markedness and language mixing effects. This discrepancy was attributed to the introduction of a decision advantage bias for Marked words in Grainger & Beauvillain's experiment, which used only Non-Marked pseudowords. Consequently, when they were available, i.e., in the Marked lists, orthographic markers were always associated with a YES-response. When they fixed that bias, by adding Marked pseudowords to their list, Thomas and Allport did not find any difference between the switch cost in the Marked list and the Non-Marked list in both a generalized and in a language-specific lexical decision task. This suggests that orthographic markers could dampen switch-cost only if they correctly predicted the response, indicating that the language control would come from the task system.

Furthermore, Orfanidou & Sumner, (2005) managed to replicate the results from both Grainger & Beauvillain, (1987), and from Thomas & Allport, (2000). In their first experiment, they used completely Marked (containing both Marked words and Marked pseudowords) or completely Non-Marked lists (containing both Non-Marked words and Non-Marked pseudowords), they found that the switch cost was less important for Marked items than for Non-Marked items. In their second experiment, they used a mixed list, containing Marked and Non-Marked items, and again orthographic markedness did not interact with the switch-cost. Hence, taken together those three studies seem to indicate that orthographic markedness does not modulate the switch cost effect when markers are mixed with Non-Marked items (Orfanidou & Sumner, experiment 2; Thomas & Allport,) but can reduce this cost when every item in a list is Marked (Orfanidou & Sumner, experiment 1, Grainger & Beauvillain,). Consequently, even if their differences prevent them from drawing a clear conclusion about the locus of language control, studies investigating the impact of orthographic markedness on the language switch cost suggest, although in a roundabout way, that orthographic markers could constitute an interesting cue when a bilingual needs to retrieve the language membership of an item.

B. L1 tuning

The role such orthographic markers play when attributing/detecting the language an item belongs to has been investigated more straightforwardly through language decision tasks, in which bilinguals are openly asked to retrieve the language membership of an item.

First, these studies have shown that, when available, pseudowords were attributed a language according to their markers (Borragan et al., 2020; Jared et al., 2013; Oganian et al., 2016). Meaning that a French-English bilingual will attribute a pseudoword like KNIDE (KN being an English marker) as belonging to English more often than a pseudoword such as LUVRE (VR being a French marker), or a pseudoword such as POUSE (Non-Marked item, every bigram is shared across French and English). Besides, the study from Oganian et al., 2016, showed that for pseudowords containing only legal bigrams in both languages, the response rate tended to be modulated by the relative bigram frequency in both languages, meaning that a pseudoword containing a bigram that was more frequent in L2 than L1 (EA more frequent in English: 2395 per million than French: 438 per million according to the CELEX database) was sorted as belonging to L2 faster than a pseudoword containing a bigram that was equally frequent in both languages (OU in English: 3127 per million; in French: 3813 per million)[24]. Therefore, orthographic markers, such as language specific letters (e.g., ß), language specific letter combinations (e.g., KN) and frequency-based markers (e.g., EU) are used to attribute a language to a novel word[25].

Moreover, orthographic markedness has also been found to impact language detection using word items, since a Marked word like KNIFE is responded to faster than a Non-Marked word like HOUSE (Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012) in a language decision task, indicating that participants use orthographic markers when determining the language membership of an item, even when lexico-semantic information is available. This effect has been found with proficient bilinguals (Casaponsa et al., 2014; Oganian et al., 2016; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012), but also with monolinguals (Borragan et al., 2020; Casaponsa et al., 2014), suggesting that

---

[24] *The reverse was true for pseudoword containing a bigram more frequent in L1 than L2 (EU is legal in both French and English but more frequent in French: 3153 per million than in English: 189 per million).*

[25] *Interestingly, in a second experiment using a naming task as an implicit language decision task, marked words containing an illegal letter combinations were pronounced according to the language they were marked for (KNIDE read according to English GPCs) more often than neutral (TOUSE) or marked for the other language pseudowords (LUVRE), and frequency based markers (EU or EA) speeded responses.*

participants did not need to be highly proficient to extract and use orthographic markers to determine the language of an item.

Interestingly, bilinguals seem to be more sensitive to L2 orthographic markers than they are to L1 orthographic markers. French-English bilingual will benefit more from the presence of the KN- bigram when classifying a word like KNIFE as belonging to English, than they will benefit from the presence of the VR bigram when classifying LIVRE as belonging to French (Casaponsa et al., 2014; Duñabeitia et al., 2020; Vaid & Frenck-Mestre, 2002). More precisely, it seems that bilinguals are sensitive to a wider range of orthographic markers in L2 than in L1 (van Kesteren et al., 2012). In fact, both L1 and L2 words containing language-specific letters are responded to faster than Non-Marked words (containing only legal letter combinations). However, while L2 words containing legal letter combinations that are more frequent in L2 than in L1 (e.g., EA) are responded to faster than L2 words containing legal letter combinations that are as frequent in L1 and in L2 (e.g., OU), this difference was not found for L1 words (i.e., EU vs. OU)[26].

So far, the results coming from language decision tasks, i.e., the markedness benefit for bilinguals AND monolinguals, for pseudowords AND words, and the asymmetrical sensitivity to L1 and L2 markers, could easily be explained if we suggest that in order to perform a language decision/attribution task, an individual will compare every item to their L1 orthotactic distribution. L2-markers, either L2-specific bigrams or L2-specific letters will constitute violations to the L1 orthotactic rules, while bigrams that are more frequent in L2

---

[26] *This enhanced sensitivity to L2 markers has also been exhibited by studies using pseudowords, L2 marked pseudowords are responded to more quickly than L1 marked pseudowords (Borragan et al., 2020; Duñabeitia et al., 2020; Oganian et al., 2016), even though this difference tends to be reduced with growing L2 exposure (Borragan et al., 2020)*

than in L1 would constitute deviations from the L1 orthotactic patterns, making them very much salient no matter whether a lexical form is available or not. L1-markers on the other hand would not stand out, and only the most salient, such as L1-specific letters would impact responses.

Furthermore, this L1-tuning is comforted by the fMRI data collected by Oganian et al., (2015), but seems to be modulated by L2 exposure, since Borragan et al., (2020), found that participants became more sensitive to L1 markers with L2 exposure. Indeed, before completing an L2 learning program, participants (older monolinguals) were shown to be more sensitive to L2 markers than to L1 markers, however after completion of this program, the difference between L1 and L2 markers was reduced.

To sum up, the studies presented above show that orthographic markedness constitutes a sublexical cue that bilinguals are able to use to attribute language membership information to a novel word (Borragan et al., 2020; Jared et al., 2013; Oganian et al., 2016), and to ease language membership retrieval from an established word (Casaponsa et al., 2014; Duñabeitia et al., 2020; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012). This is further supported by the finding that under specific circumstances, orthographic markedness also modulates well-researched effects that are derived from language membership detection such as language mixing effect and language switch cost (Grainger & Beauvillain, 1987; Orfanidou & Sumner, 2005). Moreover, since orthographic markedness effect has been found for both bilinguals and monolinguals (Borragan et al., 2020; Casaponsa et al., 2014), and given that there is an asymmetry between L1 and L2 markers (Casaponsa et al., 2014; Duñabeitia et al., 2020; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012), this language detection advantage found for orthographic marker could reflect the detection of L1 violations.

2. … that could guide lexical access

This finding of a language detection advantage for orthographic markers constitutes the first step toward understanding the influence of orthographic markedness on bilingual visual word processing. The next important question pertains to the particular locus of orthographic markedness effect in the bilingual visual word recognition process. In fact, following the BIA+ model, the finding that orthographic markers help language detection could easily be explained by a decision advantage arising in the task system. If orthographic markedness impacts lexical identification on the other hand, then this effect must be accounted for in the word identification system, since the BIA+ model assumes that the task system is not able to restrain lexical access. Thus, investigating the impact orthographic markedness may have on word recognition/identification *per se* is paramount in locating the orthographic markedness effect but also in testing the assumptions of the BIA+ model.

A. Depending on the task requirements

To explore this issue of whether the orthographic markedness can be used not only to detect the language membership of an item, but also to help word recognition, van Kesteren et al., (2012), used two lexical decisions, one in L1, the other in L2 (see table 4 for a summary). Although orthographic markedness impacted responses in both L1 and L2 LDTs, a closer look at the data suggests that it was used strategically by participants. In fact, van Kesteren et al., manipulated the predictive value of L1 and L2 markers in both tasks, and this manipulation led to different response patterns in each task. In the L1 LDT, a set of L1 words, Marked or Non-Marked, was associated with YES-responses. A set of L1 pseudowords and a set of L2 words, also Marked or Non-Marked, were associated with NO-responses. Meaning that in the

L1 task, the L2 markers were always associated with a NO-response, while L1 markers could be associated with either YES or NO-responses, therefore L2 markers were predictive of the response while L1 markers were not. The reverse was true for the L2 task, (L1 markers were predictive of the response, while L2 markers were not). The L1 LDT exhibited a facilitative effect of orthographic markedness over the rejection of L2 words, but not for the recognition of L1 words nor for the rejection of L1 pseudowords. The results from the L2 LDT closely mirrored those of the L1 LDT, with a facilitative effect over the rejection of L1 words, but not for the recognition of L2 words nor for the rejection of L2 pseudowords. These results suggest a strategic use of orthographic markedness, since markers were used only when they held a predictive value[27].

Following their results, van Kesteren et al., proposed an extension to the BIA+ model that accommodates the data reviewed this far (see figure 9). To account for the fact that bilinguals use both lexical and sublexical information to detect the language membership of an item, the BIA-extended model contains two sources of language membership information, represented by one pair of language nodes at the lexical level, and one pair of language nodes at the sublexical level, the latter being an addition to the original BIA+ model. The two sets of nodes, both lexical and sublexical, collect language membership information that they send forward to the task/decision system, which will in turn select the response and/or the correct strategy based (partly) on this information. Importantly, the sublexical language nodes do not

---

[27] *This is in line with another study from Lemhöfer and Radach (2009), in which they explored the impact of language-likeness of pseudowords depending on the language of the task at hand. They found that L1-like pseudowords were more difficult to reject in an L1 LDT, in which they were not predictive of the response than in an L2 LDT, and that L2-like pseudowords were more difficult to reject in an L2 LDT, in which they were not predictive of the response than in an L1 LDT. However, it is difficult to disentangle Orthographic Markedness from Orthographic Neighborhood in this study.*

have a direct association with the lexicon, consequently it cannot impact lexical activations. Thus, in this extended model, the orthographic markedness effect takes place in the task decision system. In other words, in the BIA-extended model, orthographic markedness cannot directly impact lexical access.

Table 4: Stimulus list composition for the lexical decision tasks in van Kesteren et al., 2002's study

L2 LDT:			
Item type	Example	Response required	markedness Effect
L2 words (Marked or Non-Marked)	Wise, Brush vs. Frame	YES	No effect
L2 pseudowords (Marked or Non-Marked)	Wist, Pish vs. Flime	NO	No effect
L1 words (Marked or Non-Marked)	Slør, Sjal vs. Brud	NO	Facilitation
L1 LDT:			
Item type	Example	Response required	markedness Effect
L1 words (Marked or Non-Marked)	Slør, Sjal vs. Brud	YES	No effect
L1 pseudowords (Marked or Non-Marked)	Føst, Nesj vs. Slemp	NO	No effect
L2 words (Marked or Non-Marked)	Wise, Brush vs. Frame	NO	Facilitation

Yet, we could argue that the lexical decision tasks used by van Kesteren et al., were just another way to investigate the decision advantage provided by orthographic markers. In fact, even though the decision had to be made on lexical information, participants in these two tasks were still required to retrieve language membership information to perform the task at hand, seeing as there were non-target language words to reject. Consequently, just like in a language decision task, markers were giving crucial information to the task system in order to make accurate decisions. Therefore, these tasks do not allow to explore the orthographic markedness effect independently from its decision advantage component.

Hence, the question about whether orthographic markedness can impact word recognition even when language membership is not useful for the task at hand still stands, and with it the question about the locus of this effect in the bilingual word recognition system is left open.

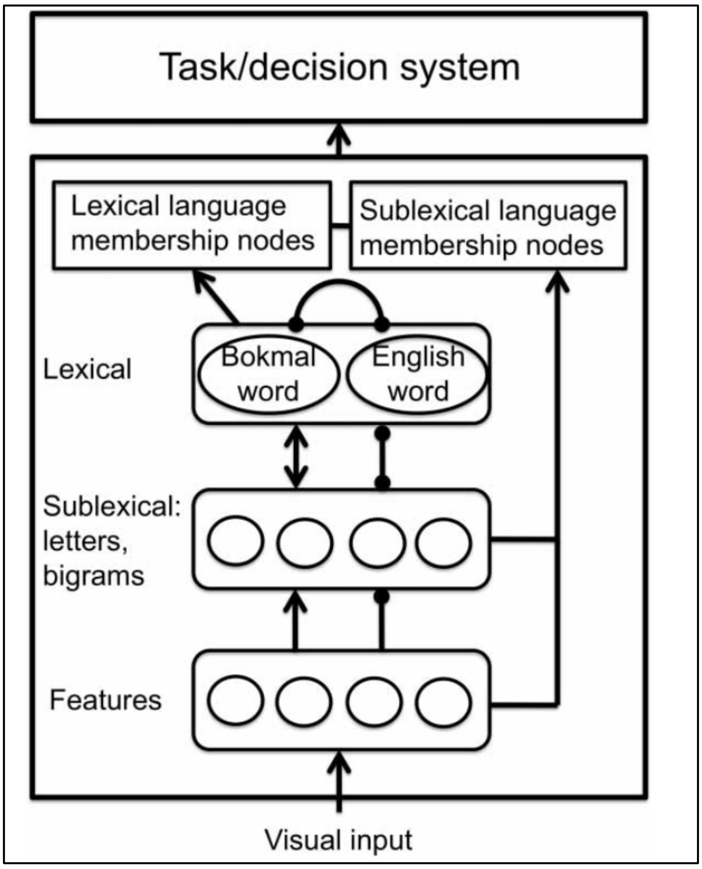


Figure 9: the BIA extended model from van Kesteren et al., 2002

B. When no strategic advantage can be gained…

In two recent studies, Casaponsa et al., in 2015 and later in 2019, instructed their participants to indicate when they detected that a target word, always presented in L1, referred to an animal. Those target words could either be preceded by masked-L1-primes (non-switched pairs) or by masked-L2-primes (language-switched pairs). Critically, L2-primes could either be Marked (e.g., containing an illegal bigram for L1, TX) or Non-Marked (e.g., containing only legal bigram for L1). Consequently, by using a semantic categorization task paired with a masked priming paradigm and by taking ERP measures, they provided their participants with what was essentially a monolingual context, in which there was no benefit in using orthographic markers to give accurate responses, thus successfully disentangling the potential orthographic markedness effect from its decision advantage component.

The results of those studies showed that, when considering both a group of bilinguals and a group of monolinguals, L2 Marked primes elicited a language switch cost (i.e., more negativity for switched pairs vs. non-switched pairs) for both monolingual and bilingual participants, whereas L2 Non-Marked primes elicited a language switch cost for bilingual participants only (Casaponsa et al., 2015). Furthermore, when focusing on a bilingual group, only language-switched pairs containing L2 Marked primes elicited more negative ERPs than non-switched pairs (Casaponsa et al., 2019).

The impact orthographic markedness exerts on such language switch cost is in line with results obtained with other paradigms providing the participants with a monolingual context, such as a letter identification task (inhibitory markedness effect; Casaponsa & Duñabeitia, 2016), a lexical decision task (facilitative effect for word items, inhibitory for pseudoword items, Commissaire, Audusseau, et al., 2019) or a progressive demasking task (facilitative markedness effect; Casaponsa et al., 2014) and suggests that orthographic markers are used automatically by bilingual participants (and by monolingual participants) even when language membership information was not strategically useful for the task at hand.

Besides, the markedness effect found by Casaponsa et al., (2019) was exhibited as early as the N250 component which is considered to represent an index of sublexical to lexical mapping (Grainger & Holcomb, 2009). This early orthographic markedness effect is further supported by another recent study from Hoversten et al., (2017), where orthographic markers were used to investigate the time course of language detection.

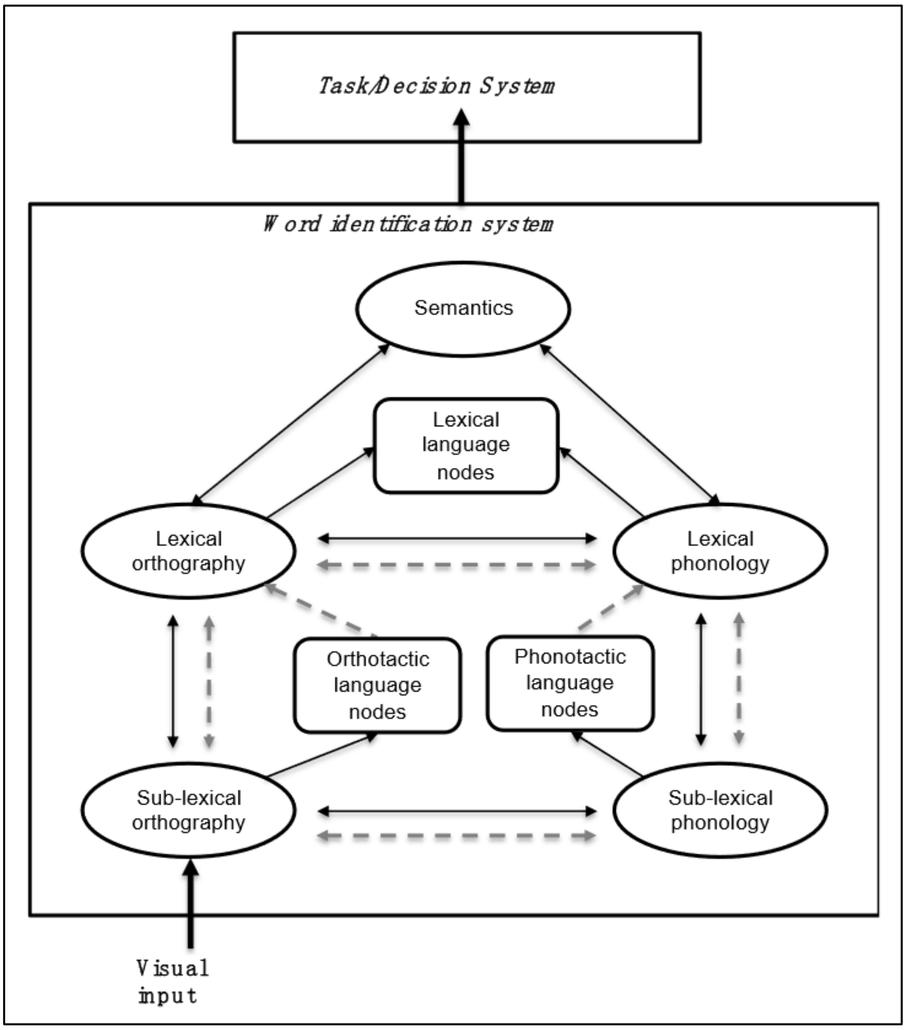


Figure 10: the BIA-s model from Casaponsa et al., (2019)

In this experiment, the authors used an oddball paradigm, in which participants were asked to detect words coming from a target language in a "base language context". Meaning that participants could either be presented with an L1 (i.e., base language) dominant condition in which they were asked to detect L2 words (i.e., target language) or they could be presented with an L2 dominant condition in which they were asked to detect L1 words. In each condition, 70% of the trials were constituted of words coming from the ‘base’ language, 10% were constituted of L1-like pseudowords, 10% of L2-like pseudowords and finally 10% of the trials were constituted of target-language words, which required a response. Critically, words and pseudowords could either be Marked or not according to their language membership. Therefore, the word to be detected could either be Marked (e.g., KNIFE) or Non-Marked (e.g., HOUSE) for the target language. Results showed that target words and target-like pseudowords had an effect on the N200 and on the P300, which suggests that the presence of an orthographic marker could help bilinguals to retrieve language membership information even before full lexical identification. The pre-lexical nature of the orthographic markedness effect comforts the hypothesis of an automatic use of orthographic markers, it also allows to further hypothesize that the orthographic markers may impact lexical access, since information on language membership coming from those markers would be available early enough to impact lexical activations.

These findings of an early and automatic orthographic markedness effect, led Casaponsa et al., (2019) to propose their own extension of the BIA model, the BIA-s model (see figure 10). Just like the BIA extended model from van Kesteren et al., (2012), the BIA-s contains two sources of language information, the lexical AND sublexical language nodes. But unlike the previous proposition, the sublexical language nodes are able to influence lexical activation via bottom-up connections that are assumed to be inhibitory, allowing for some language-selectivity.

3. Orthographic Neighborhood effect and the particular mechanisms by which orthographic markedness may impact bilingual lexical access

Before we dive in the particular mechanisms described to account for a potential impact of orthographic markedness on lexical access, which are heavily based on the interpretation of the interaction between the orthographic markedness effect and the orthographic neighborhood effect, it is crucial to ensure that both effects are not confounded. In fact, Marked words tend to have less cross-language neighbors than Non-Marked words do, casting doubt on subsequent modulation of the orthographic neighborhood effect by the markedness status of a word. This clarification was the aim of the study from Commissaire, Audusseau, et al., (2019).

In their experiment, participants performed a lexical decision task on three types of items, (1) Marked and Small cross-language Orthographic Neighborhood items (WRECK; 0 French Neighbor[28]); (2) Non-Marked and Small cross-language Orthographic Neighborhood items (SPARE; 2 French Neighbors); and (3) Non-Marked and Large cross-language Orthographic Neighborhood items (PRIDE; 8 French Neighbors). A significant Orthographic markedness effect was found, with Marked words being responded to faster than Non-Marked words (WRECK<SPARE), and with Marked pseudowords being rejected slower than Non-Marked pseudowords (WRICK>SPAFE). However, the cross-language orthographic neighborhood density effect was not impactful (SPARE=PRIDE). Thus, the orthographic markedness effect was successfully disentangled from the orthographic neighborhood effect. Consequently, any modulation of cross-language neighborhood effect by the orthographic

---

[28] *Data coming from Clearpond (Marian et al., 2012)*

markedness factor cannot be explained solely by the fact that Non-Marked words have generally more cross-language neighbors than Marked words do.

More to the point, using two paradigms, a letter detection task and a masked translation priming paradigm paired with an L2 lexical decision task, Casaponsa & Duñabeitia, (2016) found a pattern of results suggesting a stronger cross-language activation for Non-Marked items compared to Marked items. In fact, letters were more easily identified in Non-Marked items than in Marked items (i.e., containing an illegal bigram), while the masked translation priming effect (inhibitory) was found to be significant only for Non-Marked primes (e.g., ilargi-LUNA meaning MOON vs. txapel-BOINA meaning BERET). This led them to propose that orthographic markers would modulate bilingual lexical access by (1) promoting the selection of target language lexical representations and/or (2) by inhibiting non-target language lexical representations.

Following the first assumption, marked bigrams, would send activation only toward language-congruent/within language lexical representations (i.e., KN toward KNOW, KNIFE, KNIGHT, KNOT etc.)[29], while a Non-Marked bigram would send activation to both L1 and L2 lexical representations (i.e., LA toward LATER, LAKE, LAVE, LARME) resulting in a bias toward within-language lexical representations for Marked items. Following the second assumption, the marker would send activation specifically toward one language node (i.e., KN toward the English sublexical language node), whereas Non-Marked bigrams would send activation to both language nodes (i.e., LA toward both French and English sublexical language nodes), the

[29] *Similarly a frequency based marker such as the EU bigram would send activation toward more French lexical representations (MEUBLE, JEU, PEUR, HEURE, NEUTRE…) than toward English lexical representations (NEUTRAL…)*

sublexical language node that receives information from the marker would then receive more information than the non-target node and would therefore be able to inhibit cross-language lexical representations[30].

Surprisingly, in a recent study by Commissaire, (2021) using another priming paradigm with French-English speaker, a stronger inhibitory (orthographic) priming effect was found in the Marked prime condition and not in the Non-Marked prime condition. At surface level, this is not consistent with the results from Casaponsa & Duñabeitia, (2016) described above, since it would suggest more cross-language activation in the case of Marked words than Non-Marked words. However, while Casaponsa & Duñabeitia, (2016), used primes that had a semantic overlap (e.g., txapel-BOINA vs. ilargi-LUNA), Commissaire, used primes with an orthographic overlap (e.g., wrap-DRAP vs. trap-DRAP). Therefore, in the first study the competition is indirect whereas in the second study the competition is direct, maybe WRAP activates less cross-language neighbors than TRAP does, but in both cases the cross-language target DRAP is among those neighbors.

The particular mechanism by which orthographic markedness influences lexical access may therefore depend on the task requirements, when the lexical competition between the prime and the target is indirect, markers may activate more within language neighbors than Non-Marked items do AND inhibit cross-language lexical representation, when the lexical competition between the prime and the target is direct, then maybe the 1st mechanism is disabled, and the 2nd mechanism comes fully into play. In fact, both results are still in line

[30] *Interestingly, the first proposition is in line with the proposition coming from the monolingual literature, that a bigram of low frequency will activate only a few candidates at the lexical level which will result in a weak lexical competition (Chetail, 2015) while the second proposition mirrors the role of language nodes in the original BIA model.*

with the hypothesis of an earlier inhibition of non-target language representations when markers are presented.

4. A word on L2-proficiency and orthographic markedness

While sensitivity to orthographic markers does not require any experience with a second language, it has been shown to be modulated by L2 experience (Borragan et al., 2020; Duñabeitia et al., 2020). Yet the language proficiency factor has been mostly overlooked when investigating the impact of orthographic markedness on bilingual lexical access. In fact, most of the studies discussed above were based on data coming from highly-proficient bilinguals. To our knowledge, only two studies took L2 proficiency as a factor when investigating the effect of orthographic markedness on lexical access.

While Commissaire, Audusseau, et al., (2019) did not find an interaction between L2 proficiency and orthographic markedness, Casaponsa et al., (2014) on the other hand, showed that even though monolinguals and bilinguals benefited from the presence of an orthographic marker in a language decision task, they behaved differently when they had to perform a progressive demasking task[31], in which the language membership of the item did not matter so much as its lexical identity. The results from the unbalanced bilingual are really interesting here, since they behave like monolinguals when it comes to the comparison between L2 and L1 words but behave like balanced bilinguals when it comes to L2 Marked and Non-Marked word comparison.

---

[31] *L2 marked words (KNIFE) were identified faster than L2 non-marked words (HOUSE) for both bilingual groups (balanced and unbalanced), but the reverse was found for monolinguals (HOUSE<KNIFE). Also marked L2 words (KNIFE) were recognized slower than control L1 words (JUPE meaning skirt in French) for unbalanced bilinguals AND monolinguals but no difference was found for balanced bilinguals.*

This pattern of results suggests that at first the presence of L2 markers is inhibitory for word recognition but acquiring some knowledge of an L2 allows participants to use the orthographic markers to speed up their responses. However, bilinguals need to be perfectly balanced in order to match L2-Marked words with their L1 counterparts. Thus, although evidence is scarce, the orthographic markedness on lexical access may be modulated by L2 proficiency.

5. Intermediate summary:

To sum up, the investigation of orthographic markedness has brought up some questions about the current models of bilingual lexical access, leading to some updates. The first one concerns the language membership information. While in the BIA models this information is retrieved post-lexically, studies on orthographic markedness have shown that language membership information could be extracted from sublexical cues, before complete word identification, which led to the addition of sublexical language nodes to the extension of the BIA+ model (BIA extended; van Kesteren et al., 2012, BIA-s; Casaponsa et al., 2019) .

Importantly, if markers facilitate word identification only when associated with a decision advantage, this would support the language control mechanism described in the BIA+ model. On the other hand, if markers influence word identification automatically, some further updates are required for the BIA+ model to account for this finding. The particular locus of the orthographic markedness effect is a matter of much debate. Still, some evidence points at an impact of orthographic markers on bilingual lexical access, and some attempts have been made to describe the particular mechanisms by which they may do so (Casaponsa & Duñabeitia, 2016; Commissaire, 2021).

IV. Conclusion

Current models of bilingual visual word recognition assume a “deeply” language non-selective access to an integrated lexicon. To manage the increased competition arising under those assumptions, these models each implemented a language control mechanism that differed on its particular process, even though it was considered to take place post-lexically in both models. Most tellingly, the BIA+ model introduced the language control mechanism in the task system, suggesting that bilinguals will control their decisions but not their lexical access *per se*. Nevertheless, in this chapter we discussed two variables that might modulate the bilingual lexical access, namely the linguistic context and the orthographic markedness.

Interestingly, the linguistic context was already considered by the BIA+ model to be a source of modulation of the bilingual lexical access, since it was assumed to impact directly the word recognition system. Yet, actual evidence supporting this claim is scarce. Cross-language interactions are still observed when ambiguous words are integrated to a sentence context, no matter the strength of the semantic information it may introduce (Lauro & Schwartz, 2017; Schwartz & Kroll, 2006; Titone et al., 2011; van Hell & de Groot, 2008). On the other hand, studies investigating the language switch-cost appear to be more successful at unearthing any modulation of the bilingual lexical access by the linguistic context (Bultena et al., 2015b, 2015a; Hoversten & Traxler, 2020; Litcofsky & Van Hell, 2017). Particularly, Hoversten & Traxler, (2020), have found evidence for a partially selective access hypothesis, where the non-target language was not accessible to the bilingual when the linguistic context was essentially fully monolingual.

Unlike the linguistic context, orthographic markedness was overlooked by the BIA+ model, consequently no particular assumptions as to its impact on the bilingual visual word recognition were described in this model. Yet, this sublexical variable has since gathered interest and has even led to the proposition of some updates to the BIA+ model. Orthographic markedness has been unanimously found to be a reliable language membership cue (Casaponsa et al., 2014; Oganian et al., 2016; van Kesteren et al., 2012). Even though its particular locus is still a matter of debate, orthographic markedness seems to impact bilingual visual word recognition either strategically (van Kesteren et al., 2012) or automatically (Casaponsa et al., 2019), suggesting that the presence of orthographic markers may guide bilingual lexical access by either favouring the activation of target language representations, or by inhibiting non-target language lexical representations (Casaponsa & Duñabeitia, 2016).

To conclude, even though models of bilingual visual word recognition propose a strong language non-selective lexical access hypothesis, accompanied by a post-lexical language control mechanism, (tenuous) evidence for an early modulation of the bilingual lexical access emerges from the literature investigating the impact of the linguistic context and from the literature exploring the impact of orthographic markedness on bilingual visual word recognition.

OBJECTIVES

As we have seen thus far, the globalized world that surrounds us, which encourages the close cohabitation of several languages in our communities has sparked scientific interest in the mechanisms underlying bilingualism. Moreover, the important role played by the written language in second-language acquisition, at least for school-taught bilinguals, has steered the attention of psycholinguists toward L2 visual word recognition.

As a result of the numerous studies on the subject, several models (BIA, BIA+, RHM, Multilink etc.) have been proposed to account for visual word recognition in bilinguals. The prevailing propositions, i.e., the BIA and BIA+ models (Dijkstra & van Heuven, 2002; van Heuven et al., 1998) make two important assumptions about the bilingual lexicon: those of (1) an integrated lexicon and of (2) a language non-selective access. Besides, even though they differ on the particular mechanism involved in language control, both BIA and BIA+ models argue for its post-lexical nature.

On the other hand, neither model provides a detailed description of the sublexical level nor any predictions about the impact sublexical units may exert on L2 visual word recognition. This oversight reflects the lack of investigation into this particular issue in the literature. Still, the few studies that did focus on sublexical units and their variations across languages have revealed some interesting results that could further our understanding of bilingual visual word recognition.

For instance, the co-existence of two set of GPCs (e.g., I refers to \i\ in French but can refer to \aɪ\ in English) in the bilingual brain leads to some cross-language interactions at the sublexical level (Brysbaert, 2001; Commissaire, Duncan, et al., 2019; Commissaire et al., 2014; Hevia-Tuero et al., 2021; Nas, 1983; Van Wijnendaele & Brysbaert, 2002) indicating that bilinguals activate not only target but also non-target GPCs even in a language-specific context. This finding comforts a deeply language non-selective access to the bilingual lexicon. Besides, the variation in orthotactic distribution and the orthographic markers (e.g., KN or OW are English markers for a French-English reader) that derive from it appear to facilitate language membership detection (Borragan et al., 2020; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012). Those markers are even conceived by some authors as a source of modulation in bilingual lexical access (Casaponsa et al., 2019).

In this thesis our aim was twofold.

First, we wanted to explore three sublexical variables arising from the co-existence of two written languages in the bilingual mind, i.e., Orthographic markedness, Cross-language GPC congruency and L2 within-language GPC consistency, and to identify their separate impact on L2 visual word processing. This was achieved through 6 experiments, including 3 lexical decision tasks that were paired with 3 naming tasks. This multi-task approach allowed us to gain some insight on two complementary aspects of L2 visual-word processing, that are visual word recognition and pronunciation. This study and the results it ensued are described in Chapter III.

Our second objective was to further investigate the impact of orthographic markedness through its interaction with the linguistic context, that also constitutes a potential source of modulation of the bilingual lexical access (Dijkstra & van Heuven, 2002). To do that we paired, first a language decision task, then a generalized lexical decision task, with an innovative flanker paradigm (Declerck et al., 2018), that allowed us to tightly control for the markedness of the target and of the linguistic context. This new methodology, along with the results it revealed are described in Chapter IV.

EMPIRICAL CONTRIBUTION

CHAPTER III / An exploration of three sublexical variables and their separate impact on L2 visual word processing

Exposition to a second language through visual media requires individuals to update their set of sublexical units to accommodate their L2's orthographic system and its specificities. For instance, L2-learners will need to include new L2-specific orthographic units, e.g., KN bigram for a French native reader being exposed to English, which are referred to as orthographic markers. They will also need to update their GPC repertory to include L2 GPC rules, which might overlap with L1 in some cases, e.g., OU can be pronounced \u\ in both French and English, but might also differ between L1 and L2, e.g., OU can be pronounced \aʊ\ in English but not in French, giving rise to cross-language incongruencies. Moreover, as we can see with the OU bigram, a consistent GPC in L1 may need to be updated to include 2 different GPCs in L2 and thus become inconsistent when considering L2 within-language consistency. Those three sources of sublexical variations brought up by L2 acquisition may well impact L2 visual word processing. Still, except for the orthographic markedness effect that has recently gained some popularity, the impact of these three sublexical variables on L2 visual word processing has scarcely been explored.

Moreover, those three sublexical variables are tightly tied together. For instance, GPC congruency variations are inherently exclusive to shared bigram (vs. markers) and might constitute a confounding variable for orthographic markedness. Similarly, the two sources of GPC inconsistency (GPC congruency and consistency) can co-occur, which might lead them to interact with each other. Hence, in this chapter we will endeavor to explore the separate impact of each of these sublexical variables on L2 visual word recognition while carefully controlling for the other two sources of variations. First, we propose to conduct three lexical decision tasks that will allow us to gain a direct insight on the lexical access process while providing us with a baseline of what to expect for later explorations. We will then conduct three naming tasks which we believe would encourage sublexical processing in reading, and by doing so give us a closer look at those sublexical variables.

Experiment 1: Investigating sublexical variables through lexical decision tasks

As stated above, in this first experiment our aim was to investigate the impact of these three sublexical variables in three separate lexical decision tasks in order to disentangle them from each other. For instance, the first lexical decision task was designed to explore orthographic markedness while controlling for GPC components, by selecting only incongruent and consistent words. Thus, if the orthographic markedness effect is solely due to relative bigram frequency differences from language to language, we should be able to replicate the facilitative orthographic markedness effect found by Commissaire et al., (2019) with similar items and participants. Similarly, the second lexical decision task aimed at investigating the impact of GPC congruency on actual L2 visual word recognition (vs. letter detection, as previously investigated by Commissaire, Duncan, et al., 2019 and Hevia-Tuero et al., 2021) while using only Non-Marked and consistent words, allowing us to further disentangle both cross-language variables (GPC congruency and orthographic markedness) from each other. Finally, the third lexical decision task was designed to explore the impact of L2 within-language GPC consistency and its interaction with orthographic markedness.

Moreover, since the reliance on those sublexical variables might depend on language proficiency (Casaponsa et al., 2014; Commissaire et al., 2014; Jared & Kroll, 2001), and considering the particular profile of the L2-learner population we were interested in, namely French late-learners of English, we recruited participants that learned English through a school setting, and we measured each participants English proficiency using the LEXTALE test (Lemhöfer & Broersma, 2012) in order to explore the interaction between L2 proficiency and each of our selected variables. Interestingly in this study, we used a new database; i.e., the APPREL2 database; that better reflected French English-learners' actual experience with English words compared with classical databases.

We expected the first lexical decision task to exhibit a facilitative orthographic markedness effect based on Commissaire et al., (2019) results and on Casaponsa et al., (2019) assumptions, with better performances for Marked words than for Non-Marked words. Similarly, based on studies from Commissaire et al., (2014) and from Hevia-Tuero et al., (2021), we expected the second lexical decision task to show a facilitative GPC congruency effect. That is, better performances for words containing only congruent GPCs than for words containing incongruent GPCs. And finally based on monolingual studies (Jared et al., 1990; 1997) we expected the third lexical decision task to reveal a facilitative GPC consistency effect. Moreover, since higher-proficiency participants should have a more thorough knowledge of orthographic distributions and of Grapheme-to-Phoneme correspondences, those main effects should all be modulated by L2 proficiency, leading in our opinion to larger differences for higher-proficiency participants than for lower-proficiency participants. Finally, in the third lexical decision task we looked for a modulation of the orthographic markedness effect depending on L2 within-language GPC consistency, but with no particular assumptions about the dynamic we might uncover.

Experiment 1A: Investigating orthographic markedness

Method

Participants

A total of 62 university students aged between 18 and 37 (mean age=22; sd=3.82) participated in this study. They all reported that their native and dominant language was French and they all learned English through a school setting. Their L2 proficiency was measured using the LEXTALE score (Lemhöfer & Broersma, 2012; mean score=73.72; sd=9.02; min score= 52.5; max score= 95). This sample contained 16 male and 45 female participants, 6 of them were left-handed.

Material

Words were selected from the APPREL2 database for this task. The APPREL2 database was designed to provide a source of material adapted to young L2-learners of English. This database contains 19671 English words coming from 53 textbooks frequently used for English courses in French middle schools. Hence, the data given by this database reflects more accurately the experience of young French learners of English in terms of exposition to the orthographic forms of English words.

Words. We selected 80 English words (mean zipf=5.19; sd=0.41) from the APPREL2 database. The word length varied from 3 to 8 letters. 40 were monosyllabic words and 40 were dissyllabic words. Half of these words contained a frequency-based orthographic marker for the English language, i.e., the 'ea', 'oo', 'ee' , 'oa', 'ow' or 'aw' bigram that was much more frequent in English than it was in French (bigram frequency between 10 and 438 per million in French and between 388 and 2395 per million in English). Those items are referred to as Marked words. An example of such a word is BOAT. The other half contained only shared bigrams between English and French, i.e., the ‘ou’, ‘oi’, ‘an’, ‘in’ or ‘un’ bigrams that were as frequent in English as they were in French (bigram frequency between 456 and 8270 per million in French and between 468 and 7156 per million in English). Those items are referred to as Non-Marked words. An example of such a word is HOUSE (the material and their pairing parameters can be found in Appendix 1A and 1B).

First, to calculate orthographic markedness we used the minimum bigram frequency measure from the CELEX database (Baayen, R H. et al., 1995) which is a good indicator of orthographic markedness. We found that while Non-Marked words were matched on minimum bigram frequency across languages (mean min bigram frequency in English= 860, SD=571; mean min bigram frequency in French=843, SD=805; p=.911), Marked words had a significantly higher minimum bigram frequency in English than in French (mean min bigram frequency in English=682, SD=497; mean min bigram frequency in French=91, SD=144; p<.001).

Moreover, minimum bigram frequency as calculated by the APPREL2 database, did not differ between Marked and Non-Marked words (min bf for Marked words=3582, SD=2757; min bf for Non-Marked words=3134, SD=2715; p=.233), meaning that if there is an orthographic markedness effect it does not come from a limited exposition to markers in English courses, French learners of English were exposed to English markers at the same extent as they were to other bigrams. The difference should come from the fact that English markers are indeed more present in English than they are in French.

Marked and Non-Marked words were matched on lexical frequency, (mean zipf for Marked words=5, SD=0.42; mean zipf for Non-Marked words=5, SD=0.39; *p*=.372), on distance of Levenshtein with French words (mean for Marked words=3, SD=0.50; mean for Non-Marked words=3, SD=0.41; *p*=.057) as indicated by the APPREL2 database. We also ensured that they were matched on mean orthographic neighborhood frequency in both English (mean for Marked words=66, SD=102; mean for Non-Marked words=86, SD=221; *p*=.605) and French (mean for Marked words=64, SD=288; mean for Non-Marked words=74, SD=410; *p*=.899) as given by the Clearpond database (Marian et al., 2012). This was done to avoid neighbourhood effects and to further ensure that Marked and Non-Marked words were similar on the lexical level. We also ensured that both Marked and Non-Marked words were matched in terms of mean reaction times for English-native speakers using the British Lexicon Project database (Keuleers et al., 2012; mean for Marked words=528.59, SD=31.99; mean for Non-Marked words=533.93, SD=24.51; *p*=.405), ensuring us that any difference between Marked and Non-Marked words should come from the fact that English was not the first language of our participants.

Every word selected for this experiment contained a consistent rhyme (within-language wise), or at least referred to the most frequent pronunciation of their rhyme. A word like "Sleep" is consistent because there is only one way to read the 'eep'. A word like "Door" (\dɔː\) is considered consistent here despite the existence of "Poor" (\pʊə\), because the \ɔː\ pronunciation is considerably more frequent than the other pronunciations. Also, every Non-Marked word was considered incongruent when considering its GPCs, meaning that (at least) one of their bigram was associated with a different pronunciation in L2 compared to L1. For instance, every 'ou' occurrences refers to a pronunciation which is different from the French pronunciation, namely \aʊ\ in HOUSE (vs. \u\ in French). This was done so that Marked and Non-Marked words would be comparable in that both conditions were associated with an L2-specific GPC.

Pseudowords. We created 80 pseudowords by changing one letter from the selected words while taking care to keep the bigram of interest (both Marked and Non-Marked bigrams) intact. Consequently, half of those pseudowords were Marked pseudowords, since they contained a frequency-based marker. The other half were Non-Marked pseudowords that contained only bigrams which exist in both English and French. Marked pseudowords had a significantly higher minimum bigram frequency in English (mean=675; sd=539) than in French (mean=83; sd=139; p<.001), while the Non-Marked set did not significantly differ as to their minimum bigram frequency in both languages (minimum bigram frequency in English= 828; sd=690, minimum bigram frequency in French= 748, sd=814; p=.639).

On the other hand, Marked and Non-Marked pseudowords were matched on orthographic neighborhood frequency in both English (mean Orthographic neighborhood frequency in English for the Marked pseudowords=86; sd=78; mean Orthographic neighborhood frequency in English for the Non-Marked pseudowords=158; sd=253; p=.088) and French (mean Orthographic neighborhood frequency in French for the Marked pseudowords=41; sd=194; mean Orthographic neighborhood frequency in French for the Non-Marked pseudowords=161; sd=515; p=.172).

Procedure

Participants were tested in an isolated room. First, they were asked to perform two positioning test in English (DIALANG, University of Lancaster, and LEXTALE, Lemhöfer & Broersma, 2012) and to read a text in French ("le pollueur", Gola-Asmussen, et al., 2011). Then they were asked to perform three lexical decision tasks followed by three naming tasks, the order of which was counterbalanced across participants.

Every one of these lexical decisions and naming tasks was created using DMDX (Forster & Forster, 2003). Items appeared in white on a black background with a font size of 30. First, a fixation cross was displayed for 500ms. Then the item was presented until a response was given or for a maximum of 3000ms. Participants were asked to indicate if a given item was a real English word by clicking on a response key with their preferred hand. If the item was not a real English word, they had to click on another response key with their non-preferred hand. No feedback was given as to the accuracy of the response.

Overall, the completion of these tasks took a little under 2 hours with around 30 minutes that were dedicated to the completion of the lexical decision tasks.

Results

Table 5: Mean Lexical Decision Latencies (in ms) and Accuracy (in %) to Marked and Non-Marked items presented in Experiment 1A

	Experiment 1A					
	Word Data			**Pseudoword Data**		
	Marked	Non-Marked	Total	Marked	Non-Marked	Total
Reaction times (in ms)	665(177)	649(150)	657(164)	853(278)	886(273)	870(276)
Accuracy (in %)	97(0.16)	98(0.12)	98 (0.14)	95(0.22)	96(0.19)	96(0.20)

Data trimming and analysis

Erroneous responses, timed out responses as well as responses above or below three times the standard deviation both by participants and by items in each condition (Marked vs. Non-Marked) were removed from the reaction time analyses. Responses under 300ms and timed-out responses were removed from the accuracy analyses. Finally, data coming from two participants (which had a low accuracy score on pseudoword rejection= 66% accuracy or were extremely slow= 1400ms for words and 2000ms for pseudowords), as well as the items AWFUL, LAND, TROW, SLEND and NOICE (which were responded between 47% and 89% accuracy with an accuracy mean at 95%) had to be removed, resulting in the exclusion of 3.5% of the word data, and 1.9% of the pseudoword data.

Reaction times (inverse transformed) and accuracy data on both words and pseudowords were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). In the reaction time analysis, we were able to include random effects that take into account the deviation from the intercept and from the slope for our participants and from the intercept only for our items (i.e., model=Y~X (1+X|Participant)+(1|Item)). In the accuracy analysis, the random effects only include by-participant and by-item random intercepts (i.e., model=Y~X(1|Participant)+(1|Item)) to ensure model convergence.

We used a stepwise model comparison approach in which we compared 4 models. Model 1 contained the random effects. Model 2 added the effect of L2 proficiency (as a continuous variable based on LEXTALE scores). Model 3 added the effect of orthographic markedness (Marked or Non-Marked). Model 4 added the interaction between L2 proficiency and orthographic markedness (details about model comparisons can be found in Appendix 2A and 2B).

Words- Reaction Times

Model 2 containing the effect of L2-proficiency was the best fit (AIC=-59900; χ2(1,6)=7.92; *p*=.004). A closer look at model 2 confirmed this effect of L2 proficiency (b2=-8.4e-06; SE= 2.7e-06; t=3.14), the increase in L2-proficiency is paired with faster reaction times (see table 6). No model containing the orthographic markedness effect improved fit.

Table 6:Final model for the words' latency data in Experiment 1A (1~ L2proficiency).

Fixed Effects	Estimate	Standard Error	t
Intercept	9.6^{e-04}	1.9^{e-04}	4.84
L2 Proficiency	8.4^{e-06}	2.7^{e-06}	3.14
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	1.16^{e08}	1.1^{e-04}	
Intercept Participants	3.34^{e-08}	1.8^{e-04}	
+markedness	1.27^{e-09}	3.6^{e-05}	

Words- Accuracy

Model 3 was the best fit for accuracy data (AIC=931; $\chi^2(1,5)=5.16$; *p*=.023). Looking closer (see table 7), there was a main effect of L2-proficiency (b_2=0.03; SE= 0.02; z=1.92), with higher-proficiency participants being more accurate than lower-proficiency participants, and a main effect of markedness (b_3=0.61; SE=0.27; z=2.29), with Marked words being responded to less accurately than Non-Marked words (Marked= 97% vs. Non-Marked words= 98%).

Table 7: Final model for the words' accuracy data in Experiment 1A (1~ L2proficiency+ Orthographicmarkedness)

Fixed Effects	Estimate	Standard Error	z
Intercept	1.70	1.21	1.40
L2 Proficiency	0.03	0.02	1.92
Orthographic markedness	0.61	0.27	2.29
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	0.44	0.66	
Intercept Participants	0.50	0.71	

Pseudowords- Reaction Times

The model comparison revealed that the second model containing L2 proficiency significantly improved fit, (AIC=-58771; $\chi^2(1,7)=12.12$; *p*<.001). No model containing orthographic markedness improved fit (*p*>.10). Looking closer at model 2, the main effect of L2 proficiency was significant ($b_2=8.9^{e-06}$; $SE=2.5^{e-06}$; t=3.61), with higher-proficiency participants rejecting pseudowords faster than lower-proficiency participants (HP=794ms vs. LP=897ms).

Pseudowords- Accuracy

Model 3, containing both the effect of L2 proficiency and the effect of orthographic markedness best fitted the accuracy data (AIC=1619; $\chi^2(1,5)=2.9$; *p*=.089). Looking at this model (see table 8) we found a significant effect of L2 proficiency ($b^2=0.05$; SE=0.015, z=3.58) and a trend toward an effect of orthographic markedness ($b_3=0.38$; SE=0.22; z=1.70). Higher proficiency participants rejected pseudowords more accurately than lower-proficiency participants (HP=97% vs. LP=94%), and Marked pseudowords tended to be rejected less accurately than Non-Marked pseudowords (M=95% vs. NM=96%).

Table 8: Final model for the pseudowords' accuracy data in Experiment 1A (1~ L2proficiency+ Orthographicmarkedness)

Fixed Effects	Estimate	Standard Error	z
Intercept	-0.59	1.10	-0.54
L2 Proficiency	0.05	0.02	3.58
Orthographic markedness	0.38	0.22	1.70
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	0.47	0.69	
Intercept Participants	0.58	0.76	

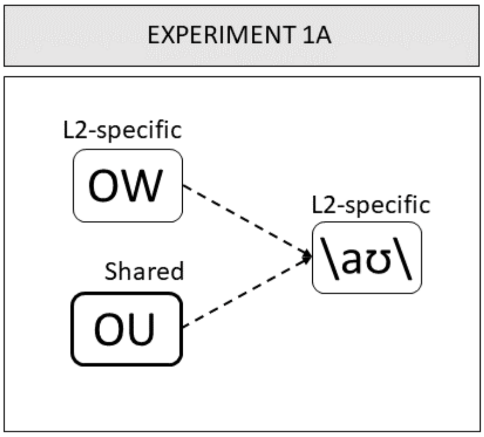
Discussion

To sum up, the only effect of orthographic markedness found in this experiment was an inhibitory effect on accuracy for both word and pseudowords with Marked items leading to lower accuracy compared to Non-Marked items. These results are not in line with the literature on the subject, since even the study closest to our experiment (Commissaire et al., 2019) in terms of material, procedure and participants found a facilitative effect of orthographic markedness on reaction times (see figure 11).

Several explanations may be proposed to account for our lack of orthographic markedness on reaction times. First, the strategic orthographic markedness point of view states that orthographic markers are used only if they are tied to a decision advantage. Thus, our particular design, i.e., a language-specific lexical decision task, in which the orthographic markers did not provide any decision advantage, might be the reason why we did not find any effect of orthographic markers. Second, our participants were all late-learners of English who learned their L2 in a school setting and they may not have been able to benefit from orthographic markers to perform lexical decisions, just as the study from Casaponsa et al., (2014) suggests. However, even though we cannot completely cast those explanations out, they are not likely to be the whole reason for the lack of orthographic markedness effect in our experiment. In fact, the study from Commissaire et al, that exhibited a facilitative effect of orthographic markedness, also used a language-specific lexical decision task with French-English speakers drawn from a similar population, i.e., French Psychology students, that we can safely assume to be L2-late learners of English for the most part and to have acquired this second language through a school setting.

We could also explain our lack of markedness effect by the fact that the database we used, that was originally designed to test L2 visual word processing in young L2-learners, may have driven us to select words that were too frequent for any orthographic markedness effect to appear in our adult participants. In fact, when we look at the lexical frequency in a published database, the word we used in this experiment were highly frequent (mean lexical frequency based on the BLP database= 4.9; sd=0.54). Hence, our participants may have been too familiar with the material we used, discouraging a more thorough sublexical processing and preventing us from observing any effect of orthographic markedness. Yet, our set of material was comparable with the one from Commissaire et al., in terms of lexical frequency as measured in a classical database (mean lexical frequency based on the BLP database= 4.6; sd=0.81). However, the previous study from Commissaire et al., exhibited a facilitative effect of orthographic markedness on accuracy, while we on the other hand found a negative impact of orthographic markedness on accuracy, suggesting that the words used in each experiment differed in terms of subjective frequency for the participants involved.

Still, the most likely explanation of our pattern, in our opinion, must come from the control of GPC consistency and congruency we implemented in our study. For example, we only included consistent words in our material list. This may have altered the effect of orthographic markedness. In fact, the lack of GPC inconsistency combined with the fact that our words were pretty familiar to our participants, may have discouraged a thorough sublexical processing and therefore erased the orthographic markedness effect. Moreover, since Marked words always referred to L2 specific GPCs (at least on the orthographic component), we wanted to select Non-Marked words that also referred to L2 specific GPCs, hence every Non-Marked word used here were incongruent. Thus, maybe when Marked words are compared to incongruent Non-Marked words, no benefit might be derived from orthographic markedness because both types of words refer to an L2-specific GPC either on the orthographic and/or on the phonological component. This interpretation implies that the orthographic markedness effect could be explained in part by the association between orthographic markers and language specific GPCs.



	WORDS	PSEUDOWORDS
REACTION TIMES	L2 proficiency (t=3.14) No Orthographic markedness effect	L2 proficiency (z=3.61) No Orthographic markedness effect
ACCURACY	L2 proficiency (z=1.92) Orthographic markedness effect (z=2.29)	L2 proficiency (z=3.58) Trend for orthographic markedness (z=1.70)

Figure 11: Summary of the results from Experiment 1a

Experiment 1B: Investigating cross-language GPC congruency

Method

Participants

Same as experiment 1A

Material

Words. We selected 80 English words (mean zipf=5.14; sd=0.35) from the APPREL2 database. The word length varied from 3 to 9 letters, half of them were monosyllabic words and the other half were dissyllabic words. We ensured that the items used in this task were overall not Marked for English using the minimum bigram frequencies from the CELEX database (Min Bigram Frequency in English= 641, sd= 503; Min Bigram Frequency in French=

562, sd= 711; *p*=.419). Critically, half of these words were congruent with the French GPC conversion rules, meaning that if the French GPCs were applied to them, the results would be close to the English original pronunciation. For example, the word FLAG would be read \flag\ instead of \flæg\. The other half of the words were incongruent, meaning that if the French GPCs were applied to them, the results would be quite different from the original English pronunciation (see figure 12). The word LATE would be read \lat\ and not \leɪt\ (the material and its pairing parameters can be found in Appendix 3A and 3B).

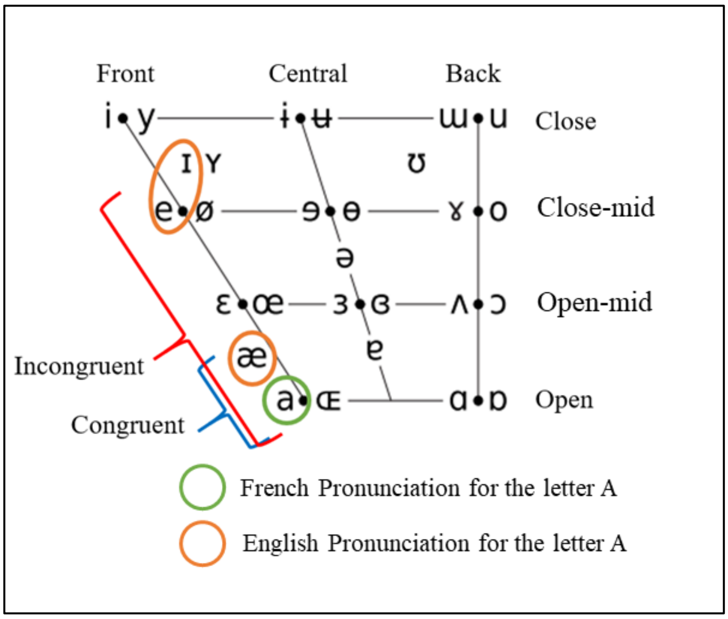


Figure 12: Distance between the English and the French pronunciation for the grapheme 'A' (congruent for FLAG, and incongruent for LATE) adapted from the International Phonetic alphabet charts (2015).

We used a survey to check whether our participants did perceive the two pronunciations (the French and the English one) associated with a word to be truly different for the incongruent words and rather similar for congruent words. In this survey French native late-learners of English were asked to rate the degree of similarity in the pronunciation of each words when comparing French and English GPC rules. They had to rate those items on a Lickert scale going from 1 (really different to French) to 5 (really similar to French). Data coming from this survey showed that congruent words were rated as significantly more similar to the French pronunciation than incongruent words (mean for congruent words= 4.16, sd=1.07; mean for incongruent words= 1.99, sd=0.97; $p<.001$).

Congruent and Incongruent words were matched on lexical frequency in the APPREL2 database (mean zipf for congruent words= 5, sd=0.37; mean zipf for incongruent words= 5, sd=0.32; $p=.926$), on distance of levenshtein (mean for congruent word=3, sd= 0.43; mean for incongruent words= 3, sd= 0.43; $p=.589$) and on orthographic neighborhood frequency in both English (mean for congruent words=72,69, sd=133,62; mean for incongruent words= 89,85, sd=127,51; $p=.559$) and French (mean for congruent words= 327.39, sd=1481.53; mean for incongruent words= 65.42, sd=327.71; $p=.281$) as given by the Clearpond database (Marian et al., 2012). This was done to avoid neighbourhood effects and to further ensure that congruent and incongruent words were similar on the lexical level. We also ensured that both congruent and incongruent words were matched on mean English bigram frequency as indicated by the APPREL2 database (mean bigram frequency for congruent words= 7633, sd= 3014; mean bigram frequency for incongruent words= 6801, sd= 3430; $p=.236$). A later check in a database published with native speakers confirmed that match (CELEX; mean bigram frequency for congruent words= 6772, sd= 3767; mean bigram frequency for incongruent words= 6468, sd= 4167; $p=.733$). Finally, we ensured that the two categories were matched in terms of mean reaction times for native speakers using the British Lexicon Project database (Keuleers et al., 2012; 549.06, sd=34.51; mean for incongruent words= 542.64, sd=30.78; $p=.383$), ensuring us that any difference between Marked and Non-Marked words should come from the fact that English was not the first language of our participants.

Pseudowords. 80 pseudowords were created by changing one letter from the selected words. Consequently, half of those pseudowords were considered to be congruent pseudowords, since reading those items based on their lexical neighbours would lead the grapheme of interest to be associated with the congruent GPC; i.e., the I in SISH should be assimilated to the one in FISH and therefore should sound \ɪ\ like it does in French. The other half of the pseudowords were considered to be incongruent pseudowords since reading those items based on their lexical neighbours would lead the grapheme of interest to be associated with the incongruent (language-specific) GPC; i.e., the I in DIRL should be assimilated to the one in GIRL and therefore should sound \ɜ\ which is impossible in French.

Congruent and Incongruent pseudowords were matched for orthographic neighborhood frequency in both English (p=.735) and French (p=.665). Plus, the bigrams included in the pseudowords for this experiment were equally frequent in English and French (p=.178).

Procedure

Same as the first lexical decision task.

Results

Table 9: Mean Lexical Decision Latencies (in ms) and Accuracy (in %) to Congruent and Incongruent items presented in Experiment 1B

Experiment 1B						
	Word data			**Pseudoword data**		
	Congruent	Incongruent	Total	Congruent	Incongruent	Total
Reaction times (in ms)	678(186)	661(168)	670(177)	858(266)	864(266)	861(266)
Accuracy (in %)	97(0.16)	98(0.14)	97(0.15)	95(0.22)	94(0.23)	95(0.22)

Data trimming and analysis

Erroneous responses, timed out responses as well as responses above or below three times the standard deviation both by participants and by items in each condition (Congruent, Incongruent) were removed from the reaction time analyses. Responses under 300ms and timed-out responses were removed from the accuracy analyses. Finally, data coming from two participants (which had a low accuracy score on word responses= 79% accuracy or on pseudoword rejections= 72%) the items GRID, TIDY and CRIP (which were responded to with low accuracy, between 42% and 68% with a mean accuracy at 96% for words and 94% for pseudowords) had to be removed, resulting in the exclusion of 2.1% of the word data, and 3.4% of the pseudoword data.

Reaction times (inverse transformed) and accuracy data on both words and pseudowords were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). In the reaction time to word items analysis, we were able to include random effects that take into account the deviation from the intercept and from the slope for our participants and from the intercept only for our items (i.e., model=Y~X (1+X|Participant)+(1|Item)). In the reaction time to pseudoword analysis and both accuracy analyses, the random effects only includes by-participant and by-item random intercepts (i.e., model=Y~X(1|Participant)+(1|Item)) to ensure model convergence.

We used a stepwise model comparison approach in which we compared 4 models. Model 1 contained the random effects. Model 2 added the effect of L2 proficiency (based on LEXTALE scores). Model 3 added the effect of GPC congruency (Congruent or Incongruent). Finally, Model 4 added the interaction between L2 proficiency and GPC Congruency (more detail about the model comparisons can be found in Appendix 4A and 4B).

Words- Reaction times

Model 2 containing the effect of L2-proficiency was the best fit for the reaction times (AIC=-59715; $\chi^2(1,7)=7.26$; $p=.007$). A closer look confirmed this effect of L2-proficiency ($b_2=7.8^{e-06}$; $SE=2.9^{e-06}$; $t=2.74$), with higher-proficiency participants responding faster than lower-proficiency participants.

Words- Accuracy

Again the second model was the best fit for the accuracy data (AIC=991; $\chi^2(1,4)=13.37$; $p<.001$), a closer look revealed that the L2-proficiency effect was significant ($b_2=0.06$; $SE=0.02$; $z=3.64$) with better performances for higher-proficiency participants than for lower-proficiency participants.

Pseudowords- Reaction times

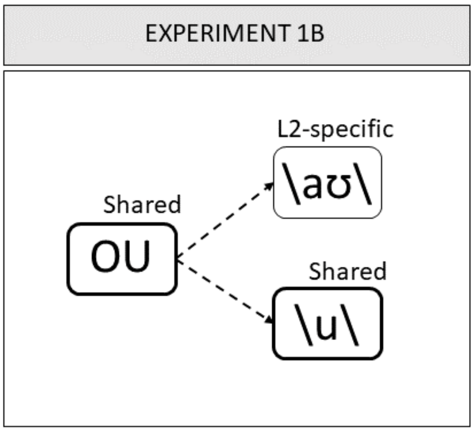
No models improved fit (all *ps*>.10).

Pseudowords- Accuracy

The second model tended to improve fit compared to the first model (AIC=1793; $\chi^2(1,4)=3.63$; $p=.057$), a closer look revealed that the L2-proficiency effect was trending ($b_2=0.41$; $SE=0.26$; $z=-1.59$) with better performances for higher-proficiency participants than for lower-proficiency participants.

Discussion

The results from this second experiment indicate that cross-language GPC congruency did not impact lexical decisions for our group of participants, no matter their L2-proficiency. At first glance this lack of a congruency effect seems to indicate that our participants did not perceive the GPC incongruencies. But since the congruency or incongruency of the actual words used in this experiment was verified through a GPC similarity rating questionnaire performed by a sample of participants coming from a similar population, we can reject this interpretation quite confidently. It is rather more likely that this pattern indicates that the addition of an L2 specific GPC (at least on the phonological component) does not hinder L2-word recognition for adult participants that acquired their L2 through a school setting (see figure 13).



	WORDS	PSEUDOWORDS
REACTION TIMES	L2 proficiency (t=2.74) No GPC congruency effect	No L2 proficiency No GPC congruency effect
ACCURACY	L2 proficiency (z=3.64) No GPC congruency effect	Trend for L2 proficiency (z=1.59) No GPC congruency effect

Figure 13: Summary of the results from Experiment 1b

Hence, it is difficult to reconcile these results with the literature on the impact of GPC congruency on L2 visual word processing. In fact, considering our population and our measures, i.e., reaction time and accuracy on a key-press response, we expected to replicate the facilitative effect of GPC congruency found by Commissaire et al., (2014). However, our lack of a GPC congruency effect is closer to the results of Hevia-Tuero et al., (2022) also on adult participants but from another linguistic population (i.e., Spanish-English readers with more experience with their L2) and using a different measure (i.e., mouse trajectories). Still, the discrepancy between the results from Commissaire et al., and our own experiment could originate in the task requirements. In fact, previous studies on GPC congruency used a letter-detection task, which unlike the LDT used in the present experiment, encourages a sublexical processing of their items. Thus, taken together these results might indicate that GPC congruency impacts sublexical processes such as letter-identification but would no longer exerts an influence on word recognition once the lexical process is engaged.

Experiment 1C: Investigation of L2 grapheme-to-phoneme consistency

Method

Participants

Same as experiment 1A and 1B.

Material

Words. We selected 40 English monosyllabic words (mean zipf=5.25, sd=0.46) from the APPREL2 database. The word length varied from 3 to 6 letters. Critically, half of these words were consistent on their pronunciation with every other iteration of its rhyme in the database[32]. The word "Sheep" for example is consistent because every other word ending in 'eep' in the dataset, Sleep, Deep, or Keep refers to the same pronunciation of 'eep' = \ip\. The other half were inconsistent on their pronunciation with some iterations of its rhyme within the database. The word "Break" where 'eak' is read \eɪk\ is inconsistent because other words such as "Speak" refers to another pronunciation of 'eak' namely \ik\. Importantly, to make the inconsistencies more salient to our participants we decided to select a pair of items for each rhyme, thus the -EEP rhyme will be presented in SHEEP but also in KEEP whereas the rhyme

---

[32] *It would be more accurate to talk about print-to-sound mapping, since we focus on rhyme consistency. However, to ensure the flow of information, we will keep referring to this variable as GPC consistency for the rest of the document.*

EAK will be presented in BREAK and SPEAK in this experiment (the material and its pairing parameter can be found in Appendix 5A and 5B).

Importantly, once data were gathered we decided to check our stimuli for GPC consistency by comparing our sample to the list of rhymes included in Ziegler's study of feedforward consistency (Ziegler et al., 1997). This comparison revealed that all our inconsistent rhymes could indeed be considered to be inconsistent according to Ziegler's definition and computations. All consistent rhymes could also be considered to be consistent according to Ziegler's list, except for the -ARE rhyme. In fact, this rhyme which was considered to be inconsistent since it could be read \ɑ\ when presented in the word ARE and \eə\ when presented in words such as CARE, DARE etc. This is a singular case in which only one frequent word (at least when considering what words an English L2 learner is likely to know) is responsible for the GPC inconsistency. Thus, since we recruited only late-learners of English and words in which the -ARE rhyme constituted only part of the word and not the whole-word, we considered it to be consistent.

Consistent and inconsistent words were matched on lexical frequency, (mean zipf for consistent words=5, sd=0.40; mean zipf for inconsistent words=5, sd=0.69; p=.752), and on distance of Levenshtein with French words (mean for consistent words=3, sd=0.13; mean for inconsistent words=3, sd=0.38; p=.198) as indicated in the APPREL2 database. They were also matched on orthographic neighbourhood in both English (mean for consistent words=91, sd=108; mean for inconsistent words=178, sd=255; p=.168) and French (mean for consistent words=473, sd=1045; mean for inconsistent words=458, sd=1915; p=.976) based on Clearpond data (Marian et al., 2012). They were also matched for bigram frequency, in terms of mean English bigram frequency (mean bigram frequency for consistent words=9731, sd=5343; mean bigram frequency for inconsistent words=13586, sd=7414; *p*=.068) and minimum English bigram frequency (min bigram frequency for consistent words=4605, sd=4271; min bigram frequency for inconsistent words=4632, sd=2547; *p*=.490) as indicated by the APPREL 2 database, but also according to the CELEX database (mean bigram frequency for consistent words=6305, sd=3282; mean bigram frequency for inconsistent words=7191, sd=2748; *p*=.180). This time as we were exploring an intralingual variable we expected native speakers to exhibit shorter reaction times for consistent words compared to inconsistent words, and that is what the data from the British lexicon project database (Keuleers et al., 2012) showed (mean for consistent words=524, sd=20; mean for inconsistent words= 540, sd=26; *p*=.032).

Furthermore, half of these words were Marked words (i.e., they contained a bigram that was frequent in English but not frequent in French; e.g., BREAK), the other half were non-Marked words (i.e., they contained bigrams that were quite frequent both in English and French; e.g., MOVE). Therefore, we had Marked-Consistent words (e.g., SHEEP), Marked-Inconsistent words (e.g., BREAK), Non-Marked-Consistent words (e.g., SAME) and Non-Marked-Inconsistent words (e.g., MOVE).

Marked and Non-Marked words were matched on lexical frequency in the APPREL2 database, (mean zipf for Marked words= 5, sd=0.47, mean zipf for Non-Marked words= 5, sd=0.65; *p*=.499) and on orthographic neighborhood frequency in both English (mean for Marked words= 135.53, sd= 92.60; mean for Non-Marked words= 133.40, sd= 267.94; *p*=.973) and French (mean for Marked words= 244.08, sd= 664.41; mean for Non-Marked words= 686.76, sd= 2052.67; *p*=.368) as given by the Clearpond database (Marian et al., 2012). This was done to avoid neighbourhood effects and to further ensure that Marked and Non-Marked words were similar on the lexical level. We also ensured that the two categories were matched in terms of mean reaction times for native speakers using the British Lexicon Project database (Keuleers et al., 2012; mean for Marked words=535.12, sd=19.63; mean for Non-Marked words= 529.54, sd=28.32; *p*=.474), ensuring us that any difference between Marked and Non-Marked words should come from the fact that English was not the first language of our participants.

Moreover, both Marked and Non-Marked words were matched on mean English bigram frequency as indicated by the APPREL2 database (mean bigram frequency for Marked words=10673, sd=3716; mean bigram frequency for Non-Marked words=12643, sd=8687; *p*=.179). A later check using the CELEX database confirmed this match (mean bigram frequency for Marked words=6347, sd=3082; mean bigram frequency for Non-Marked words=7149, sd=2984; *p*=.204). Finally, we calculated markedness using the minimum bigram frequency measure from the CELEX database (Baayen, R H. et al., 1995) which is a good indicator of orthographic markedness. We found that while Non-Marked words were matched on minimum bigram frequency across languages (mean min bigram frequency in English=959, sd=567; mean min bigram frequency in French=759, sd=705; *p*=.329), Marked words had a significantly higher minimum bigram frequency in English than in French (mean min bigram frequency in English=913, sd=604; mean min bigram frequency in French=208, sd=195; *p*<.001).

Pseudowords. We created 40 pseudowords by changing one letter from the selected words. Half of those pseudowords were considered to be consistent pseudowords, since they contained a consistent rhyme (EEP in SNEEP), and the other half was considered to be inconsistent pseudowords since they contained an inconsistent rhyme (EAR in MEAR), even though we couldn't be sure that participants would consider them inconsistent. Moreover, half of these items contained a marked bigram (EE or OW) while the other half did not contain such markers (i.e., Non-Marked pseudowords; TOVE).

Consistent and Inconsistent pseudowords were matched on orthographic neighborhood frequency in both English (p=.433) and French (p=.204). Moreover, while the English bigram frequency differed from the French bigram frequency in the Marked plot (p<.001) it did not in the Non-Marked plot (p=.630).

Procedure

Same as the first lexical decisions.

Results

Table 10: Mean Lexical Decision Latencies (in ms) and Accuracy (in %) to Consistent and Inconsistent items presented in Experiment 1C

			Word data			
	Reaction Times (in ms)			Accuracy (in %)		
	Consistent	Inconsistent	Total	Marked	Non-Marked	Total
Marked	659(165)	640(161)	650(163)	97(0.16)	98(0.13)	98(0.15)
Non-Marked	693(180)	664(187)	679(184)	97(0.16)	98(0.16)	98(0.15)
Total	676(173)	652(174)	664(174)	97(0.16)	98(0.14)	98(0.15)
			Pseudoword data			
	Reaction Times (in ms)			Accuracy (in %)		
	Consistent	Inconsistent	Total	Consistent	Inconsistent	Total
Marked	909(276)	936(329)	923(303)	89(0.32)	87(0.34)	88(0.33)
Non-Marked	874(259)	916(333)	895(296)	95(0.23)	87(0.33)	91(0.28)
Total	892(268)	926(331)	909(300)	92(0.55)	87(0.34)	90(0.45)

Data trimming and analysis

Erroneous responses, timed out responses as well as responses above or below three times the standard deviation both by participants and by items in each condition (Marked-Consistent, Marked-Inconsistent, Non-Marked-Consistent and Non-Marked-Inconsistent) were removed from the reaction time analyses. Responses under 300ms and timed-out responses were removed from the accuracy analyses. Finally, data coming from the items COAT and NOSE (which led to low accuracy, between 85 and 87% accuracy for a mean accuracy at 97% accuracy) had to be removed, resulting in the exclusion of 3,2% of the word data, and 2,2% of the pseudoword data.

Reaction times (inverse transformed) and accuracy data on both words and pseudowords were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). In both the reaction time and the accuracy analysis, the random effects only includes by-participant and by-item random intercepts (i.e., model= Y~X(1|Participant)+(1|Item)) to ensure model convergence.

First, we used a stepwise model comparison approach in which we compared 8 models. Model 1 contained the random effects. Model 2 added the effect of L2 proficiency (based on LEXTALE scores). Model 3 added the effect of orthographic markedness (Marked or Non-Marked). Model 4 added the effect of GPC Consistency (Consistent vs. Inconsistent). Model 5 added the interaction effect between L2 proficiency and orthographic markedness, Model 6, the interaction between L2 proficiency and GPC consistency and Model 7 the interaction between orthographic markedness and GPC consistency. Finally, Model 8 added the full interaction (L2 proficiency, Orthographic markedness, GPC consistency). Since models containing the interaction between L2 proficiency and the other variables did not converge in the accuracy data, they were dropped from the comparison (details about model comparisons can be found in Appendix 6A, 6B and 6C).

Words – Reaction Times

Model 4, containing the main effects of L2-proficiency, orthographic markedness and of GPC consistency was the best fit for the reaction times data, (model 4: AIC=-29955; $\chi^2(1,7)=3.42$; $p=.064$). A closer look at the model (see table 11) revealed a significant effect of L2 proficiency ($b_2=7.8^{e-06}$; $SE=2.7^{e-06}$; $t=2.91$), a significant effect of orthographic markedness ($b_3=-6.5^{e-05}$; $SE=2.9^{e-05}$; $t=2.28$), and a trend toward an effect of GPC consistency ($b_4=-5.3^{e-05}$; $SE=2.8^{e-05}$; $t=-1.83$). Increase in L2 proficiency led to shorter reaction times (Lower Proficiency participants= 684ms vs. Higher-Proficiency participants=644ms), Marked words were responded to slower than Non-Marked words (Marked words= 676ms vs. Non-Marked words= 652ms), and inconsistent words also tended to lead to slower reaction times than consistent words (Consistent words=650ms vs. Inconsistent=679ms).

Table 11: Final model for the words' latency data in Experiment 1C (1~ L2proficiency+Orthographicmarkedness+GPCconsistency)

Fixed Effects	Estimate	Standard Error	t
Intercept	1.0^{e-03}	2.0^{e-04}	4.99
L2 Proficiency	7.8^{e-06}	2.7^{e-06}	2.91
Orthographic markedness	6.5^{e-05}	2.9^{e-05}	2.28
GPC Consistency	-5.3^{e-05}	2.8^{e-05}	-1.83
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	3.4^{e-08}	1.8^{e-04}	
Intercept Participants	6.5^{e-09}	8.0^{e-05}	
Residuals	7.8^{e-08}	2.8^{e-04}	

Going further: analysis by subplot.

To better understand the potential dynamic between orthographic markedness and GPC consistency, we investigated these data further by analyzing subplots separated on the GPC consistency variable (only consistent vs. only inconsistent words). For both subplots, we compared four models, including (1) random effects, (2) adding the L2 proficiency effect, (3) adding the orthographic markedness effect and (4) adding the interaction between L2 proficiency and orthographic markedness.

For both subplots the model containing L2 proficiency improved fit (see table 12 and 13; model 2 consistent: AIC=-15027; $\chi^2(1,5)=7.60$; $p=.006$; model 2 inconsistent: AIC=-14823; $\chi^2(1,5)=7.98$; $p=.004$). However, while the model containing the orthographic markedness effect improved the fit further in the inconsistent subplot (model 3 inconsistent: AIC=-14825; $\chi^2(1,6)=3.97$; $p=.046$) it did not do so in the consistent subplot (model 3 consistent: AIC=-15026; $\chi^2(1,6)=1.45$; $p=.228$). This is confirmed when we look closer at both models. There was an orthographic markedness in model 3 on inconsistent words ($b_{3a}=8^{e-05}$; $SE=3.9^{e-05}$; t=-2.02) but not in model 3b on consistent words ($b_{3b}=4.9^{e-05}$; $SE=4.2^{e-05}$; t=1.18). Inconsistent Marked words tended to be recognized slower than Inconsistent Non-Marked words (Inconsistent Marked=693ms vs. Inconsistent Non-Marked=665ms), but this difference was not found for consistent words (Consistent Marked=660ms vs. Consistent Non-Marked=640ms).

Table 12: Final model for the latency data in the consistent word subplot from Experiment 1C (1~ L2proficiency+Orthographicmarkedness)

Fixed Effects	Estimate	Standard Error	t
Intercept	1.0^{e-03}	1.9^{e-04}	5.26
L2 Proficiency	7.4^{e-06}	2.6^{e-06}	2.81
Orthographic markedness	4.9^{e-05}	4.2^{e-05}	1.18
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	3.1^{e-08}	1.8^{e-04}	
Intercept Participants	7.3^{e-09}	8.5^{e-05}	
Residuals	7.5^{e-08}	2.7^{e-04}	

Table 13 : Final model for the latency data in the inconsistent word subplot from Experiment 1C (1~ L2proficiency+Orthographicmarkedness)

Fixed Effects	Estimate	Standard Error	t
Intercept	9.1^{e-04}	2.2^{e-04}	4.23
L2 Proficiency	8.3^{e-06}	2.9^{e-06}	2.88
Orthographic markedness	8.0^{e-05}	3.9^{e-05}	2.02
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	3.7^{e-08}	1.9^{e-04}	
Intercept Participants	6.1^{e-09}	7.8^{e-05}	
Residuals	8.1^{e-08}	2.8^{e-04}	

Words – Accuracy

No model improved fit (all *ps*>.10).

Pseudowords – Reaction Times

Model 3 was the best fit for the reaction times data (model 3: AIC=-29976; $\chi^2(1,6)=2.25$; *p*=.134). Looking closer at the model, we found a trend toward an L2-proficiency effect ($b_2=4.4^{e-06}$; $SE=2.5^{e-06}$; t=1.73), with higher-proficiency participants rejecting pseudowords faster than lower-proficiency participants. The effect of orthographic markedness on the other hand was not significant ($b_3=3,8^{e-05}$; $SE=2,5^{e-05}$; t=1.49).

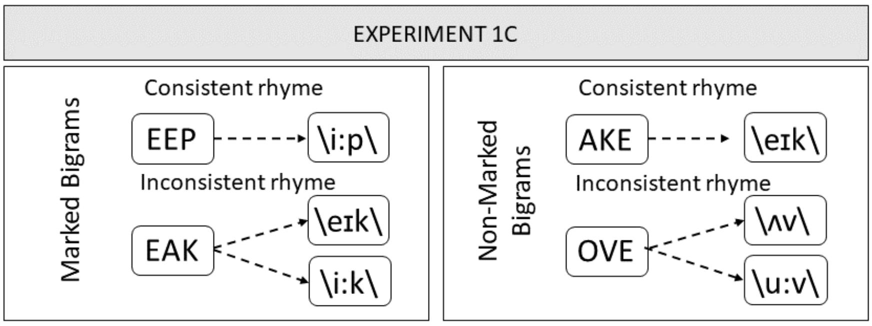
Pseudowords – Accuracy

Model 2 containing the effect of L2-proficiency improved fit (model 2: AIC=1480; $\chi^2(1,6)=7.42$; $p=.006$). A closer look at this model confirmed the effect of L2-proficiency ($b_2=0.04$; SE=0.01; z=2.85), increase in L2-proficiency led to higher accuracy (Higher-proficiency=98% vs. Lower-proficiency=97%).

Discussion

The results presented here show that the GPC consistency effect was only trending and was restricted to the latency measures on lexical decisions to words (see figure 14). As no other studies have explored the L2 readers' sensitivity to L2 GPC inconsistencies and its impact on L2 visual word recognition we cannot compare our results to the literature. Yet this trend constitutes a first indication that L2 readers are indeed slowed down by L2 GPC inconsistencies when processing L2 words. Moreover, this trend might transform into a fully-fledged significant effect with a better item selection. Indeed, here in order for the inconsistency to appear more salient to our participants, we chose to use both the dominant GPC (EAR= \ir\) and the subordinate GPC (EAR=\er\) associated with an inconsistent GPC. However, words containing an inconsistent but dominant GPC (DEAR) may not be slowed down by the activation of its competitor GPC as much as an item containing an inconsistent and subordinate GPC (BEAR), which is why our effect may only be trending and not significant.

On another note, in this last experiment, we replicated the orthographic markedness effect, even though it was restricted to the latency measures word responses. At first look, this seems to be in line with the literature on the orthographic markedness, except Commissaire, 2019 found a facilitative effect of orthographic markedness, while we found an inhibitory effect of orthographic markedness. This also differs from the results of our first lexical decision task, despite the fact that the paradigm and the participants were exactly the same. As stated earlier (see Experiment 1A - Discussion), the discrepancy between our results and those of Commissaire et al, 2019 are difficult to account for by invoking demographic and/or lexical variations between the two studies. It is all the more unlikely that this inconsistency between the results in our first and our third lexical decision task might be explained by those factors.



	WORDS	PSEUDOWORDS
REACTION TIMES	L2 proficiency (t=2.91) Orthographic markedness effect (only in the Inconsistent word subplot: t=2.02 vs. Consistent subplot : t= 1.18) Trend for a GPC consistency effect (t=1.83)	Trend toward L2 proficiency (t=1.73) No Orthographic markedness effect No GPC consistency effect
ACCURACY	No L2 proficiency No Orthographic markedness effect No GPC consistency effect	L2 proficiency (t=2.85) No Orthographic markedness effect No GPC consistency effect

Figure 14: Summary of the results from Experiment 1c

Interestingly, a closer look at the orthographic markedness effect revealed that it was significant only in the inconsistent word subplot, suggesting that the orthographic markedness effect did in fact depend on the (un)predictability of the GPC. While this result is rather difficult to interpret, it appears that the use of inconsistent words slowed down word recognition enough for an orthographic markedness effect to appear in this experiment. Conversely, the presence of an L2-specific orthographic pattern consistently associated with a unique phonological representation does not impact L2 visual word recognition. This reconciles our third lexical decision task with our first one. In fact, the lack of any orthographic markedness effect on reaction times in the first lexical decision task might be explained by our exclusion of inconsistent words from the stimuli list.

General Discussion for Experiment 1

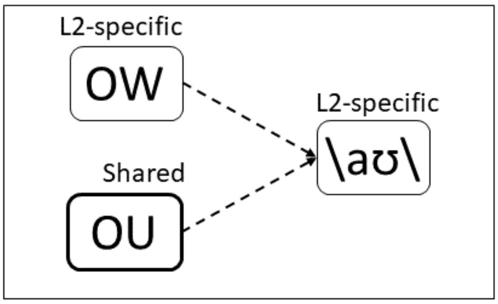
Overall, the three lexical decisions presented here aimed to explore several sublexical variables such as GPC consistency and congruency but also orthographic markedness and their separate impact over L2 visual word recognition by late learners of an L2 (see figure 15).

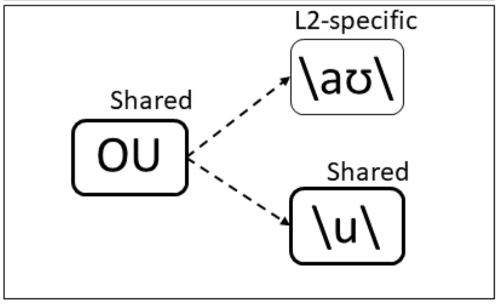
The first experiment, in which we manipulated orthographic markedness while controlling for GPC consistency and for GPC congruency failed to replicate the orthographic markedness effect. As discussed before, this pattern diverges from the results presented in a study using similar items in a similar task with similar participants (Commissaire et al., 2019). Therefore, it is unlikely that the absence of orthographic markedness effect could be accounted for by the particular linguistic experience of our participants, by the characteristics of our items and/or by the dissociation between orthographic markedness and any decision advantage we implemented in our experiment. Rather, it is more likely the control of GPC variables that prevented orthographic markedness from impacting word recognition in this experiment.

As for the GPC aspect of L2 visual word processing, the data we gathered in the second and third lexical decision tasks do not reveal an overwhelming effect of either GPC congruency or L2 GPC consistency. In fact, results to the second lexical decision task indicated that while L2-learners were able to perceive GPC incongruencies, i.e., they were able to categorize words as being congruent or incongruent in an additional survey, their word recognition processes were not impacted by this variable. Furthermore, results to the third lexical decision task suggested that L2-learners were not only sensitive to L2 GPC consistency and that their word recognition processes tended to be slowed down by the presence of an L2 within-language GPC inconsistency.

Moreover, while the third lexical decision task shed some light on the interaction between orthographic markedness and L2 within-language GPC consistency, the second lexical decision task did not explore the dynamic existing between GPC congruency and either orthographic markedness or GPC consistency. Future investigations of sublexical variables in L2 word recognition should look into those interactions. Most importantly, they should compare responses to Marked words to the responses to a set of Non-Marked and incongruent words but also to a set of Non-Marked and congruent words (see Commissaire & Demont, 2022).

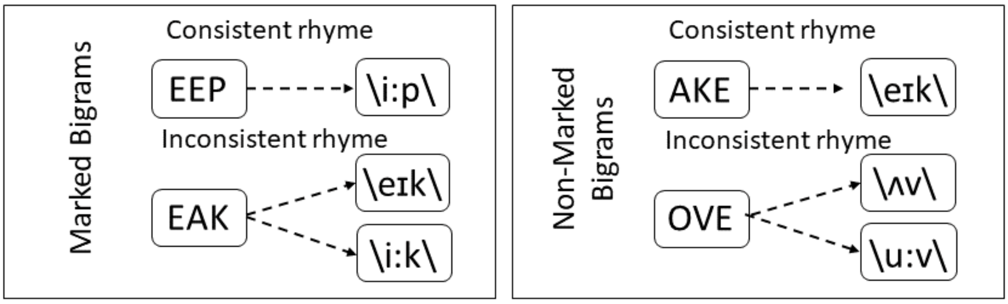
EXPERIMENT 1A			EXPERIMENT 1B		





	EXPERIMENT 1A: WORDS	EXPERIMENT 1A: PSEUDOWORDS	EXPERIMENT 1B: WORDS	EXPERIMENT 1B: PSEUDOWORDS
REACTION TIMES	L2 proficiency (t=3.14) No Orthographic markedness effect	L2 proficiency (z=3.61) No Orthographic markedness effect	L2 proficiency (t=2.74) No GPC congruency effect	No L2 proficiency No GPC congruency effect
ACCURACY	L2 proficiency (z=1.92) Orthographic markedness effect (z=2.29)	L2 proficiency (z=3.58) Trend for orthographic markedness (z=1.70)	L2 proficiency (z=3.64) No GPC congruency effect	Trend for L2 proficiency (z=1.59) No GPC congruency effect

EXPERIMENT 1C



	WORDS	**PSEUDOWORDS**
REACTION TIMES	L2 proficiency (t=2.91) Orthographic markedness effect (only in the Inconsistent word subplot: t=2.02 vs. Consistent subplot : t= 1.18) Trend for a GPC consistency effect (t=1.83)	Trend toward L2 proficiency (t=1.73) No Orthographic markedness effect No GPC consistency effect
ACCURACY	No L2 proficiency No Orthographic markedness effect No GPC consistency effect	L2 proficiency (t=2.85) No Orthographic markedness effect No GPC consistency effect

Figure 15: Summary of the results for experiments 1A to 1C.

Experiment 2: Investigating sublexical variables through three naming tasks

To complete our exploration of sublexical information on L2 visual word processing, we decided to conduct three naming tasks. In fact, just like lexical decision tasks, naming tasks are able to probe into the lexical access process and as such are able to exhibit similar lexical effects of word frequency (Balota & Chumbley, 1984; Frost et al., 1987) and of orthographic neighborhood density (Peereman & Content, 1995). However, unlike lexical decision tasks, naming tasks do not call for a decision component while informing us on the production component of the reading process. Consequently, lexical decision tasks and naming tasks have often been paired to disentangle purely lexical effects from confounding decision and/or production effects (Balota & Chumbley, 1984 for example). In this study, the aim in using the two tasks was to explore our three variables from different angles, each providing complementary information on the matter at hand, the lexical decision tasks giving a more direct view into the visual word recognition process whereas the naming task would allow a closer look at the sublexical processes involved in visual word processing and reading aloud.

In fact, as the naming task requires participants to either retrieve (following the direct/lexical pathway) or to compute (following the indirect/sublexical pathway) the phonological form associated with the printed word, the sublexical processing is rather encouraged in this task compared to the lexical decision task that can rely on mere lexical familiarity. This is supported by the exploration of the GPC consistency effect, which was commonly observed in several naming tasks, but not in lexical decision tasks, despite using the same material and participants (Jared, 1997; Jared et al., 1990). Therefore, it appears that the naming task is more suited to the exploration of GPC variations than lexical decision tasks are. Even more encouragingly for our purpose, naming tasks have been successfully used to explore the impact of GPC variations close to the consistency effect on L2 visual word processing (Jared et al., 2012; Jared & Kroll, 2001; Jared & Szucs, 2002).

Interestingly, the naming task could also be at an advantage when it comes to exhibiting an impact of the orthotactic distribution on visual word processing. This is the case for the bigram frequency in the monolingual literature, for which most of the studies using lexical decision tasks have failed to exhibit any significant effect, whereas those that used a naming task did manage to find some bigram frequency facilitative impact (although not consistently, see Chetail, 2015 for a review and Schmalz et al., 2017 for a recent exploration). Since, the orthographic markedness effect is also tied to the orthotactic distribution (across two languages), we could venture that naming tasks should be more prone to exhibit an orthographic markedness effect than lexical decisions are, especially since the orthographic markedness appear to be tied to its GPC component as we have seen in the first experiment.

So to sum up, the use of naming tasks in our study would allow us to explore our three variables from a different viewpoint, all the while providing us with complementary information to those collected in the lexical decisions described above. Moreover, the naming task appears to be more suited to the exploration of sublexical variables such as those we are investigating.

Experiment 2A: Investigating orthographic markedness

Method

Participants

The participants that took place in this experiment were the same individuals that participated to the lexical decisions described in the fourth chapter.

Material

The items used in this first naming task were the exact same items that were used for the first lexical decision task on orthographic markedness. In short, half of the items included a marked bigram (e.g., SHOW), the other half contained no such bigrams (e.g., HOUSE). They were all incongruent (i.e., the OU in the word HOUSE referred to the language-specific phoneme \aw\ not to the shared phoneme \u\) and contained a consistent rhyme (i.e., there is only one way to pronounce the rhyme EEP such as appears in SLEEP) or at least referred to the most frequent pronunciation of their rhyme.

A set of pseudowords was also created and presented to the participants, but they were not analysed in this experiment.

Procedure

Like for the lexical decision tasks described in the last chapter, participants were tested in an isolated room. After having performed two positioning test in English (DIALANG, University of Lancaster, and LEXTALE, Lemhöfer & Broersma, 2012), read a text in French ("le pollueur", Gola-Asmussen, et al., 2011) and after performing the three lexical decision tasks, they were instructed to take part to three naming tasks, the order of which was counterbalanced across participants.

Again, items appeared in white on a black background with a font size of 30. First a fixation cross was displayed for 500ms. Then the item was presented for 3000ms. Participants were asked to read aloud each word as soon as they appeared on screen. Their response was recorded using the DMDX program and both latency and accuracy were extracted at a later time through the Checkvocal extensions (Protopapas, 2007). No feedback was given as to the accuracy of the response.

Responses were then examined by the experimenter to retrieved the response times, and were later rated for accuracy by a French-native and an English-native speaker. However, concluding on the accuracy of an oral responses poses important difficulties, especially with L2-speakers. Therefore, the raters were instructed to indicate whether the productions were identifiable as English words. Consequently, we were prevented from analysing the pseudoword responses.

Results

Table 14: Mean Naming Latencies (in ms) and Accuracy (in %) to Marked and Non-Marked Words presented in Experiment 2A

Experiment 2A			
	Marked	Non-Marked	Total
Reaction times (in ms)	762(144)	752(135)	757(140)
Accuracy (in %) rated by a French speaker	85(0.36)	96(0.18)	91(0.27)
Accuracy (in %) rated by an English speaker	48(0.49)	50(0.50)	49(0.50)

Data trimming and analysis

First, the accuracy of the participants' production was assessed by two-raters, an English-native (naïve) speaker and a French-native (not naïve) speaker. The raters were asked to indicate whether the word was recognizable, but to allow for common French-speaker mistakes such as the omission of the aspired H and the R sound. Then the inter-rater agreement was measured using Cohen's Kappa. Since this inter-rater judgment was really low (k=0.11), probably due to the strictness of the English-native rater, two separate analyses were conducted for the accuracy data and the data trimming for the reaction times will not exclude incorrect responses. Thus, only timed-out responses and responses over 3 standard deviation from each participant's mean reaction times were excluded from the final data plot.

Reaction times (inverse transformed) and accuracy data on both words and pseudowords were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). In every analysis, the random effects only include by-participant and by-item random intercepts (i.e., model=Y~X(1|Participant)+(1|Item)) to ensure model convergence.

We used a stepwise model comparison approach in which we compared 4 models. Model 1 contained the random effects. Model 2 added the effect of L2 proficiency (as a continuous variable based on LEXTALE scores). Model 3 added the effect of orthographic markedness (Marked or Non-Marked). Model 4 added the interaction between L2 proficiency and orthographic markedness (details about model comparisons can be found in Appendix 7).

Reaction Times

Model 2 containing the effect of L2-proficiency was the best fit (AIC=-72336; $\chi^2(1,5)=4.76$; $p=.029$). A closer look at model 2 confirmed this effect of L2 proficiency ($b_2=-4.9^{e-06}$; SE= 2.2^{e-06}; t=-2.19), the increase in L2-proficiency is paired with faster reaction times. No model containing the orthographic markedness effect improved fit.

Accuracy – as rated by an English-native speaker

Model 2 was the best fit for accuracy data (AIC=4529; $\chi^2(1,4)=112.59$; $p<.001$). Looking closer, there was a main effect of L2-proficiency ($b_2=0.05$; SE= 0.004; z=10,49).

Accuracy – as rated by a French-native speaker

Model 3 was the best fit for accuracy data (AIC=1686; $\chi^2(1,5)=8.37$; $p=.004$). Looking closer (figure 15), there was a main effect of L2-proficiency ($b_2=0.02$; SE= 0.01; z=2.99), and a main effect of markedness (b^3= 1.41; SE= 0.47; z= 2.98), with Marked words being pronounced less accurately than Non-Marked words (Marked=85% vs. Non-Marked words=96%).

Table 15:Final model for the accuracy data as rated by a French-native speaker in Experiment 2A (1~ L2proficiency+Orthographicmarkedness).

Fixed Effects	Estimate	Standard Error	z
Intercept	2.17	0.73	2.99
L2 Proficiency	0.02	0.01	2.99
Orthographic Markedness	1.41	0.47	2.98
Random Effects	Variance	Standard Deviation	
Intercept Items	3.34	1.83	
Intercept Participants	3.69	1.92	

Discussion

Just like the first lexical decision task we presented, this first naming task failed to exhibit an orthographic markedness effect, despite it being argued to be more prone at revealing such an effect (see figure 16). In fact, the orthographic markedness effect was absent from the latency data and from the accuracy data when productions were rated by a Native English-speaker. The only analysis that revealed such an effect was the one on the accuracy data when productions were rated by a French-speaker that learned English as an L2, but since this person was not naïve to the manipulation, it is not safe to conclude on their rating. Therefore, the response pattern derived from this first naming task seems to indicate that adding an L2-specific orthographic pattern to the repertory does not impact reading aloud processes in school-taught late-learners of English.

This is not in line with the only other study conducting a naming task while exploring orthographic markedness (Oganian et al., 2016) that found an inhibitory effect of orthographic markedness. However, the authors used pseudowords rather than words since their aim was to uncover whether orthographic markers were used for language membership attribution, in the absence of lexico-semantic information. We on the other hand have only reported the data coming from word responses. Therefore, the two set of results can easily be reconciled by assuming that L2-readers are impacted by orthographic markers, which could be linked with weak GPC representations, only as long as a stronger lexical representation is not available.

Taken together with the first lexical decision task, this naming task suggests that orthographic markedness does not impact either the recognition nor the pronunciation component of L2-visual word processing. Still, in both studies we implemented a control of GPC congruency and GPC consistency that could be more prone than orthographic markedness to impact pronunciation.

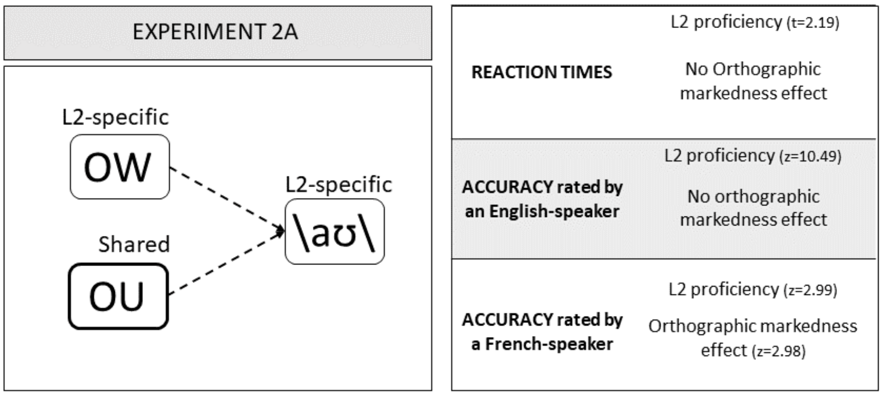


Figure 16: Summary of the results from experiment 2a

Experiment 2B: Investigating cross-language GPC congruency

Method

Participants

The participants that took place in this experiment were the same individuals that participated to the lexical decisions described in the fourth chapter.

Material

The items used in this first naming task were the exact same items that were used for the first lexical decision task on orthographic markedness. In short, half of the items were congruent (e.g., FISH), the other half were incongruent (e.g., TIME). None of them were Marked and they all contained a consistent rhyme (i.e., there is only one way to pronounce the rhyme EEP such as appears in SLEEP) or at least referred to the most frequent pronunciation of their rhyme.

A set of pseudowords was also created and presented to the participants, but they were not analysed in this experiment.

Procedure

Same as the first naming task.

Results

Table 16: Mean Naming Latencies (in ms) and Accuracy (in %) to Congruent and Incongruent Words presented in Experiment 2B

Experiment 2A			
	Congruent	Incongruent	Total
Reaction times (in ms)	753(149)	742(144)	748(147)
Accuracy (in %) rated by a French speaker	96(0.18)	94(0.25)	95(0.22)
Accuracy (in %) rated by an English speaker	65(0.48)	85(0.36)	75(0.42)

Data trimming and analysis

Again, the accuracy of the participants' production was assessed by two-raters, an English-native (naïve) speaker and a French-native (not naïve) speaker. Then the inter-rater agreement was measured using Cohen's Kappa. Since this inter-rater judgment was really low (k=0.16), probably due to the strictness of the English-native rater, two separate analysis will be conducted for the accuracy data and the data trimming for the reaction times will not exclude incorrect responses. Thus only timed-out responses and responses over 3 standard deviation from each participant's mean reaction times were excluded from the final data plot.

Reaction times (inverse transformed) and accuracy data on words were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). In every analysis, the random effects only includes by-participant and by-item random intercepts (i.e., model=Y~X(1|Participant)+(1|Item)).

We used a stepwise model comparison approach in which we compared 4 models. Model 1 contained the random effects. Model 2 added the effect of L2 proficiency (as a continuous variable based on LEXTALE scores). Model 3 added the effect of GPC congruency (Congruent or Incongruent). Model 4 added the interaction between L2 proficiency and GPC congruency (detail about model comparisons can be found in Appendix 8).

Reaction Times

Model 4 containing the interaction between L2-proficiency and GPC congruency was the best fit (AIC=-71182; $\chi^2(1,7)=3.57$; p=.059). A closer look at model 4 (see table 17) revealed a significant effect of L2-proficiency ($b_2=-6.4^{e-06}$; $SE=2.4^{e-06}$; t=2.63), with higher proficiency participants responding faster than lower-proficiency participants. The effect of GPC congruency was also significant ($b_3=-8.7^{e-05}$; $SE=3.8^{e-05}$; t=2.30), but surprisingly it was the incongruent items that lead to faster reaction times compared to congruent items (Congruent words=753ms vs. Incongruent words=742ms). Finally, the interaction between the two variables was trending ($b_4=-8.9^{e-07}$; $SE= 4.7^{e-07}$; t=-1.89), the difference between congruent and incongruent items being more important in lower-proficiency participants than in higher-proficiency participants (difference in LP participants=16ms vs. difference in HP participants=7ms; see figure 17).

*Table 17: Final model for the words' latency data in Experiment 2B (1~ L2proficiency+GPCcongruency+L2proficiency*GPCcongruency)*

Fixed Effects	Estimate	Standard Error	t
Intercept	8.9^{e-04}	1.8^{e-04}	4.99
L2 Proficiency	6.4^{e-06}	2.4^{e-06}	2.63
GPC Congruency	8.7^{e-05}	3.8^{e-05}	2.30
L2 Proficiency*GPC Congruency	-8.9^{e-07}	4.7^{e-07}	-1.89
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	4.43^{e-04}	6.66^{e-05}	
Intercept Participants	2.91^{e-08}	1.71^{e-04}	
Residuals	2.17^{e-08}	1.47^{e-04}	

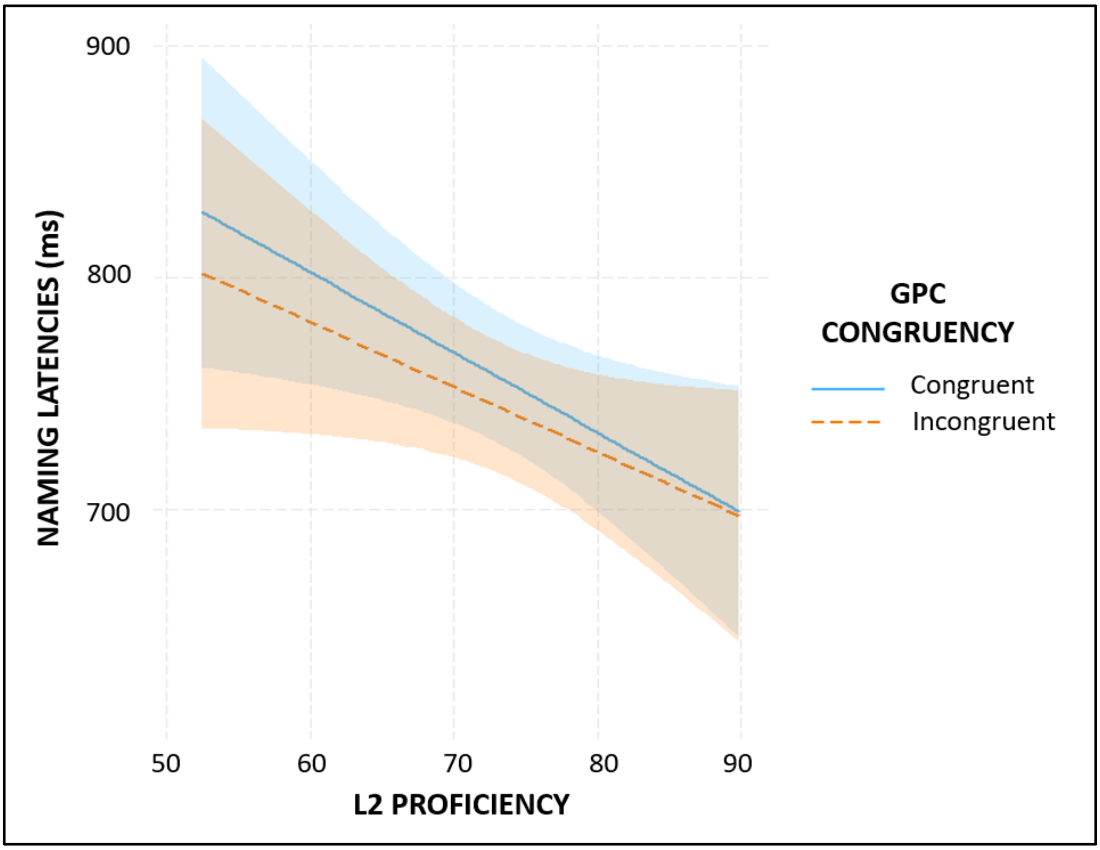


Figure 17: Interaction effect between L2 proficiency and GPC congruency on naming latencies (in ms) in Experiment 2B

Accuracy – as rated by an English native-speaker

Model 3 was the best fit for accuracy data (AIC=3844; $\chi^2(1,5)=21.12$; $p<.001$). Looking closer (see table 18), there was a main effect of L2-proficiency ($b_2=0.04$, SE= 0.02; z=2.70) with higher-proficiency participants pronouncing words with more accuracy than lower-proficiency participants. The GPC congruency effect was also significant ($b_3=1.96$; SE=0.41; z=4.83). Surprisingly, congruent words were pronounced less accurately than incongruent words (Congruent words=65% vs. Incongruent words=85%).

*Table 18: Final model for the words' accuracy data rated by an English-speaker in Experiment 2B (1~ L2proficiency+GPCcongruency+L2proficiency*GPCcongruency)*

Fixed Effects	Estimate	Standard Error	z
Intercept	-2.21	1.23	-1.80
L2 Proficiency	0.04	0.02	2.70
GPC Congruency	1.96	0.41	4.83
Random Effects	Variance	Standard Deviation	
Intercept Items	3.00	1.73	
Intercept Participants	1.17	1.08	

Accuracy – as rated by a French-native speaker.

Model 4 was the best fit for accuracy data (AIC=1462; $\chi^2(1,6)=4.13$; $p=.042$). Looking closer (table 19), there was a main effect of L2-proficiency ($b_2=0.06$; SE= 0.02; z=3.50) with higher-proficiency participants pronouncing words with more accuracy than lower-proficiency participants. The GPC congruency effect was trending ($b_3=2.41$; SE= 1.32; z=1.83), this time congruent words tended to be pronounced more accurately than incongruent words (Congruent words= 96% vs. Incongruent words=94%). The interaction between L2-proficiency and GPC congruency was also significant ($b_4=0.04$; SE= 0.02; z=2.01), however the difference between congruent and incongruent words was only slightly more important for higher-proficiency participants (Congruent items=98% vs. Incongruent items=95% accuracy) than for lower-proficiency participants (Congruent items=93% vs. Incongruent items=95% accuracy).

*Table 19: Final model for the words' accuracy data rated by a French-speaker in Experiment 2B (1~ L2proficiency+GPCcongruency+L2proficiency*GPCcongruency)*

Fixed Effects	Estimate	Standard Error	z
Intercept	0.17	1.30	-0.13
L2 Proficiency	0.06	0.02	3.50
GPC Congruency	2.41	1.32	1.83
L2 Proficiency*GPC Congruency	-0.04	0.02	-2.01
Random Effects	Variance	Standard Deviation	
Intercept Items	2.40	1.55	
Intercept Participants	0.58	0.76	

Discussion

Unlike the second lexical decision task, the present naming task revealed a GPC congruency effect, suggesting that while cross-language incongruencies did not impact L2 visual word recognition it impacted word pronunciation (see figure 18). However, to our surprise GPC congruency hindered naming performances both in the latency data and in the accuracy data when productions were rated by a Native English-speaker. The only analysis providing the expected facilitative GPC congruency effect was the analysis on accuracy data when productions were rated by a French-native English-speaker. Interestingly, this facilitative effect might provide part of the explanation for the inhibitory effect reported above, since it could indicate that French-native English-speakers do perceive the words from the congruent category as sounding like French-words, which might have encouraged them to use the actual French-GPC incorrectly when reading those words aloud, leading to a lot of pronunciation errors. When faced with incongruent words on the other hand, French-speakers would be aware of the "catch" and thus would make more effort to use L2 GPCs and in doing so would make less mistakes. Still this does not explain the GPC congruency effect on latency data since we could still have expected participants to respond fast and incorrectly to the so-called congruent words. Here we tentatively propose that in an L2-specific naming task, shared and congruent GPCs would be subjected to a more important control than L2-specific and incongruent GPCs, since in an L2-exclusive context an L2-learner might want to inhibit any representations that are too close to L1.

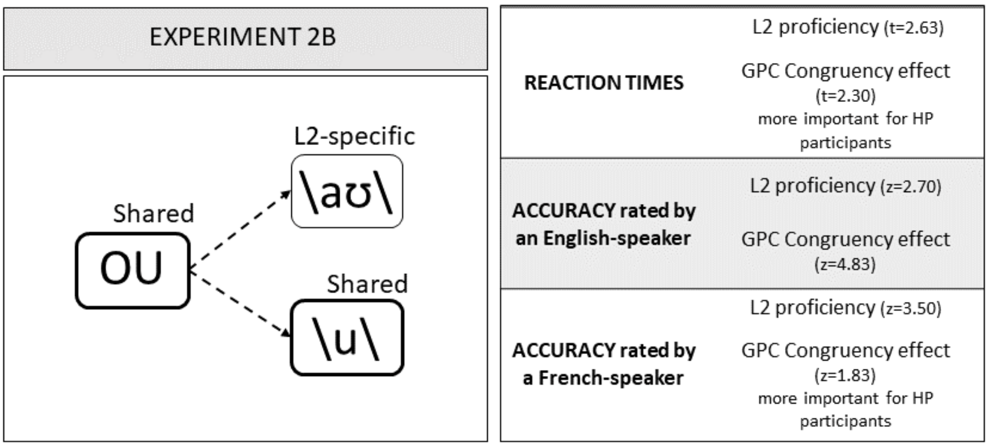


Figure 18: Summary of the results from experiment 2b

Moreover, we could be tempted to compare our naming task to the one described in Jared and Kroll's study (2001; experiment 4) in which they found that L2 words with L1 enemies, i.e., sharing their word-body with an L1 word but not their pronunciation (BAIT-LAIT), were named slower and led to more errors than words with no enemies (either in French or in English). However, while their L2 words with L1 enemies could be considered incongruent, they did not directly investigate GPC congruency, since they did not include a category in which an L2 word would have a L1 neighbor with a similar pronunciation, i.e., a congruent word. And while we manipulated GPC congruency by selecting words including a congruent GPC (the I in FISH) or an incongruent GPC (the I in TIME), we did not look at the existence of cross-language enemies. This could explain the discrepancy, and encourage future investigation of those two variables in tandem.

Experiment 2C: Investigation of L2 grapheme-to-phoneme consistency

Method

Participants

The participants that took place in this experiment were the same individuals that participated to the lexical decisions described in the fourth chapter.

Material

The items used in this first naming task were the exact same items that were used for the first lexical decision task on orthographic markedness. In short, half of the items were Marked (e.g., SLEEP), the other half were not (e.g., LOVE). Moreover, half of them contained a consistent rhyme (i.e., SHEEP and SAME) while the other half contained an inconsistent rhyme (e.g., BREAK and MOVE).

A set of pseudowords was also created and presented to the participants, but they were not analysed in this experiment.

Procedure

Same as the first two naming tasks.

Results

Table 20: Mean Naming Latencies (in ms) and Accuracy (in %) to Consistent and Inconsistent Words presented in Experiment 2C

Experiment 2C						
	Consistent			Inconsistent		
	Marked	Non Marked	Total	Marked	Non Marked	Total
Reaction times (in ms)	745(153)	754(161)	750(157)	802(207)	782(185)	748(147)
Accuracy (in %) rated by a French speaker	97(0.15)	97(0.16)	97(0.16)	98(0.13)	98(0.16)	98(0.15)
Accuracy (in %) rated by an English speaker	91(0.29)	96(0.20)	94(0.25)	78(0.42)	88(0.32)	83(0.37)

Data trimming and analysis

Again, the accuracy of the participants' production was assessed by two-raters, an English-native (naïve) speaker and a French-native (not naïve) speaker. The raters were asked to indicate whether the word was recognizable, but to allow for common French-speaker mistakes such as the omission of the aspired H and the R sound.

Then the inter-rater agreement was measured using Cohen's Kappa. Since this inter-rater judgment was really low (k=0.16), probably due to the strictness of the English-native rater, two separate analysis will be conducted for the accuracy data and the data trimming for the reaction times will not exclude incorrect responses. Thus, only timed-out responses and responses over 3 standard deviation from each participant's mean reaction times were excluded from the final data plot.

Reaction times (inverse transformed) and accuracy data on word responses were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). In every analysis, the random effects only includes by-participant and by-item random intercepts (i.e., model=Y~X(1|Participant)+(1|Item)) to ensure model convergence.

We used a stepwise model comparison approach in which we compared 8 models. Model 1 contained the random effects. Model 2 added the effect of L2 proficiency (based on LEXTALE scores). Model 3 added the effect of orthographic markedness (Marked or Non-Marked). Model 4 added the effect of GPC Consistency (Consistent vs. Inconsistent). Model 5 added the interaction effect between L2 proficiency and orthographic markedness, model 6, the interaction between L2 proficiency and GPC consistency and model 7 the interaction between orthographic markedness and GPC consistency. Finally, model 8 added the full interaction (L2 proficiency, Orthographic markedness, GPC consistency; details about model comparison can be found in Appendix 9).

Reaction Times

Model 6, containing the interaction between L2 proficiency and GPC consistency best fitted the latency data (AIC=-35033; $\chi^2(1,9)=2.82$; p=.093). A closer look at model 4 (see table 21) revealed a main effect of L2-proficiency ($b_2=4.9^{e-06}$; $SE=2.5^{e-06}$; $t=1.96$), with higher proficiency participants responding faster than lower-proficiency participants. On the other hand, the main effect of GPC consistency was not significant ($b_4=3.8^{e-05}$; $SE=6.3^{e-05}$; $t=0.60$). The interaction between the GPC consistency and L2 proficiency effects was also trending ($b_6=-1.3^{e-06}$; $SE=8.0^{e-07}$; $t=-1.68$), with the difference between consistent and inconsistent words being more pronounced for higher-proficiency participants (consistent words= 721ms vs. inconsistent words= 771ms) compared to lower-proficiency participants (consistent words= 778ms vs. inconsistent= 813ms).

*Table 21: Final model for the words' latency data in Experiment 2C (1~ L2proficiency+OrthographicMarkedness+GPCconsistency+L2proficiency*OrthographicMarkedness+L2 proficiency*GPCconsistency)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.0^{e-03}	1.9^{e-04}	5.41
L2 Proficiency	4.9^{e-06}	2.5^{e-06}	1.96
Markedness	-7.3^{e-05}	6.3^{e-05}	-1.15
GPC Consistency	3.8^{e-05}	6.3^{e-05}	0.60
L2 proficiency*Markedness	1.1^{e-06}	8.0^{e-07}	1.40
L2 proficiency*GPC Consistency	-1.3^{e-06}	8.0^{e-07}	-1.68
Random Effects	Variance	Standard Deviation	
Intercept Items	2.95^{e-08}	1.72^{e-04}	
Intercept Participants	4.74^{e-09}	6.89^{e-05}	
Residuals	3.16^{e-08}	1.77^{e-04}	

Going Further: analysis by subplot.

To better understand the potential dynamic between orthographic markedness and GPC consistency, we investigated these data further by analyzing subplots separated on GPC consistency (only consistent vs. only inconsistent words). For both subplots, we compared four models, including (1) random effects, (2) L2 proficiency effect, (3) orthographic markedness effect and (4) the interaction between L2 proficiency and orthographic markedness.

While model 4 containing the interaction between orthographic markedness and L2 proficiency was the best fit for the consistent words subplot (AIC=-17640; $\chi^2(1,7)=4.56$; $p=.033$), in the inconsistent plot, it was the second model containing only the effect of L2 proficiency that best fitted the data (AIC=-17266; $\chi^2(1,5)=2.74$; $p=.098$).

Looking closer at model 4 in the consistent subplot (see table 22), the interaction effect between L2-proficiency and orthographic markedness is indeed significant ($b_4=2.3^{e-06}$; $SE=1.1^{e-06}$; $t=2.14$), with virtually no difference between Marked and Non-Marked words in higher-proficiency participants but a bigger difference in lower-proficiency participants (difference in LP participants=18ms vs. difference in HP participants=1ms; see figure 19).

The same model 4 in the inconsistent subplot do not exhibit a significant effect of orthographic markedness (b_3=3.8^{e-05}; SE=9.4^{e-05}; t=0.40) nor any interaction between L2 proficiency and orthographic markedness effects (b_4=6.6^{e-08}; SE=1.2^{e-06}; t=-0.06).

*Table 22: Final model for the consitent words' latency data in Experiment 2C (1~ L2proficiency+OrthographicMarkedness+L2proficiency*OrthographicMarkedness)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.1^{e-03}	1.9^{e-04}	5.69
Lextale	4.3^{e-06}	2.5^{e-06}	1.72
Markedness	-1.8^{e-04}	8.4^{e-05}	-2.17
Lextale*Markedness	2.3^{e-06}	1.1^{e-06}	2.14
Random Effects	Variance	Standard Deviation	
Intercept Items	2.91^{e-08}	1.71^{e-04}	
Intercept Participants	3.82^{e-09}	6.18^{e-05}	
Residuals	2.88^{e-08}	1.70^{e-04}	

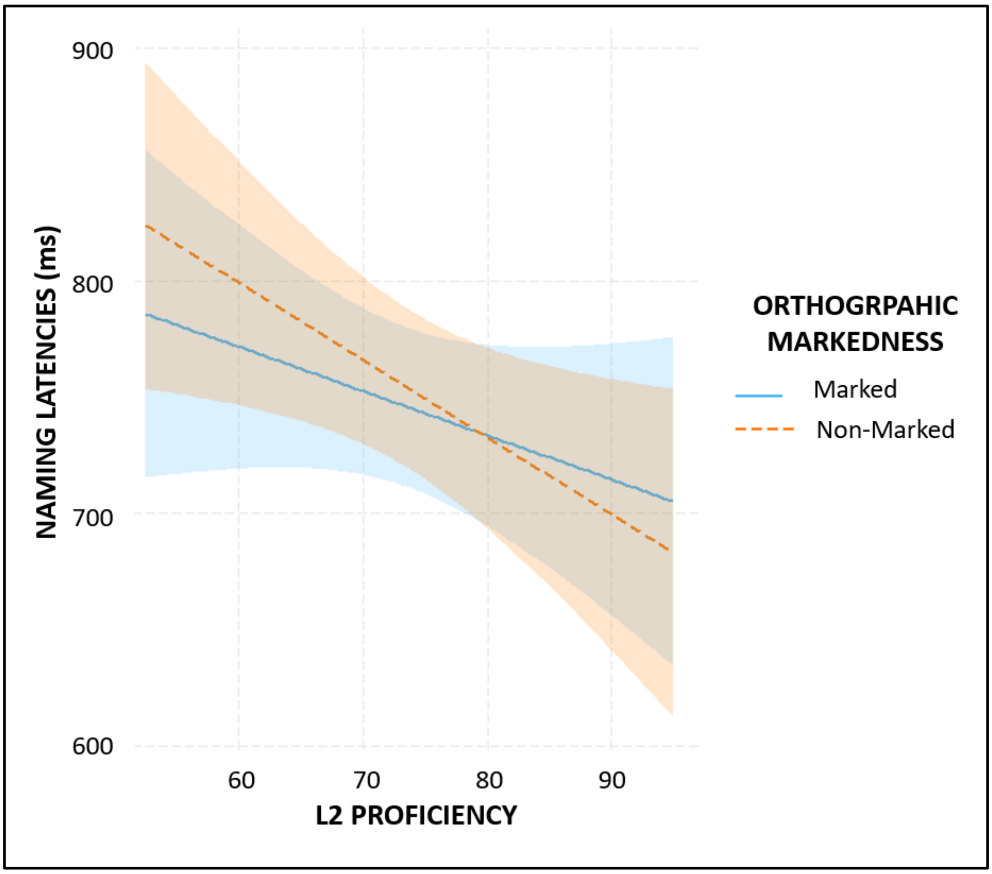


Figure 19: Interaction effect between L2 proficiency and Orthographic Markedness on naming latencies in the consistent word subplot of Experiment 2C

Accuracy- as rated by an English-native speaker

Model 4 was the best fit for accuracy data (AIC=1451; $\chi^2(1,6)=8.24$; $p=.004$). Looking closer (see table 23), there was a main effect of L2-proficiency ($b_2=0.03$, SE= 0.02; z=1.91) with a tendency for higher-proficiency participants to pronounce words with more accuracy than lower-proficiency participants. The Orthographic markedness effect is significant ($b_3= 0.92$; SE=0.46; z=2.02) with Marked words being pronounced less accurately than Non-Marked words (Marked words=85% vs. Non-Marked words=92% accuracy). Finally, the GPC consistency effect was also significant ($b_4=-1.37$; SE=0.46; z=-2.99), with consistent words being pronounced more accurately than inconsistent words (consistent words=94% vs. inconsistent words=83% accuracy).

Table 23: Final model for the words' accuracy data rated by an English-speaker in Experiment 2C (1~ L2proficiency+OrthographicMarkedness+GPCconsistency)

Fixed Effects	Estimate	Standard Error	z
Intercept	0.96	1.23	0.78
L2 Proficiency	0.03	0.02	1.91
Markedness	0.92	0.46	2.02
GPC Consistency	-1.37	0.46	-2.99
Random Effects	Variance	Standard Deviation	
Intercept Items	0.89	0.95	
Intercept Participants	1.72	1.31	

Accuracy- as rated by a French-native speaker.

Model 4 was the best fit for accuracy data (AIC=534; $\chi^2(1,6)=9.08$; $p=.003$). Looking closer (see table 24), there was an effect of L2-proficiency ($b_2=0.09$; SE=0.02; z=3.65) with a better accuracy for higher-proficiency participants than for lower-proficiency participants. The GPC consistency effect was also significant ($b_4=-2.41$; SE=0.79; z=-3.04), with consistent words being pronounced more accurately than inconsistent words (consistent words=99% vs. inconsistent words=89% accuracy).

Table 24: Final model for the words' accuracy data rated by a French-speaker in Experiment 2C (1~ L2proficiency+OrthographicMarkedness+GPCconsistency)

Fixed Effects	Estimate	Standard Error	z
Intercept	-0.98	1.51	-0.65
L2 Proficiency	0.09	0.02	4.30
Markedness	0.85	0.59	1.44
GPC Consistency	-2.36	0.62	-3.84
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	0.82	0.91	
Intercept Participants	2.29	1.51	

Discussion

The naming task presented here exhibited stronger GPC consistency effects than the third lexical decision task was able to reveal, even though it implied the same items and was performed by the same participants (see figure 20). This seems to indicate that the naming task was indeed more suited to the exploration of GPC- related effects than the lexical decision task was. More interestingly, this naming task suggests that L2-learners benefit from L2 within-language GPC consistency to accurately pronounce L2 words. Furthermore, the interaction between L2-proficiency and GPC consistency on the latency data seems to indicate that at first, lower-proficiency participants may not be aware that two pronunciations are available for inconsistent words resulting in an absence of a difference between consistent and inconsistent words. With increasing proficiency, L2 learners acquire new GPCs and with that become familiar with more GPC inconsistencies. They will later activate every GPCs associated with an item, resulting in a GPC consistency effect on naming latencies. Our results are in line with the study from Jared and Kroll., (2001, experiment 3 and 4) in which they found that L2 words with L2-enemies, i.e., words that share their word bodies but not their pronunciation (BEAD-DEAD), were recognized slower than L2 words without L2 enemies (BUMP-JUMP). However, unlike Jared & Kroll., we did not take into consideration the relative frequency between the different pronunciations associated with the word-bodies we selected, and so our manipulation of GPC consistency might have been less impactful (see Jared et al., 1990). Hence, it seems important to keep this factor in mind for a future exploration of L2 GPC consistency.

We also found an orthographic markedness effect in this naming task. The presence of a marker inhibited responses both on pronunciation's accuracy and on pronunciation's latency. Surprisingly, this markedness effect was restricted to the consistent word subplot, which is the opposite pattern to what we found in the corresponding lexical decision task in which a markedness effect only emerged in the consistent word subplot. To reconcile this discrepancy we might want to consider that inconsistent words prompt a stronger involvement of lexical processes than consistent words do in a naming task (Schmalz et al., 2016). The orthographic markedness effect which comes from a sublexical variable, may be found only when a sublexical processing of the target word is enabled, which was the case of the consistent word subplot but not of the inconsistent word subplot. In other word, the stronger involvement of lexical processes to read inconsistent words might have erased any orthographic markedness effect they might have harbored.

However, this finding is difficult to reconcile with the results coming from the first naming task reported here that revealed no such orthographic markedness effect while using only consistent words. This is quite puzzling, but a tentative explanation may come from the fact that in the first naming task we selected both monosyllabic and disyllabic words while in the present naming task we selected only monosyllabic words. And since GPC consistency rating have been computed on monosyllabic words (Ziegler et al., 1997), our control of the GPC consistency aspect might have been weaker in the first naming task than in the third one. Hence, future investigation into this issue might need to consider this factor.

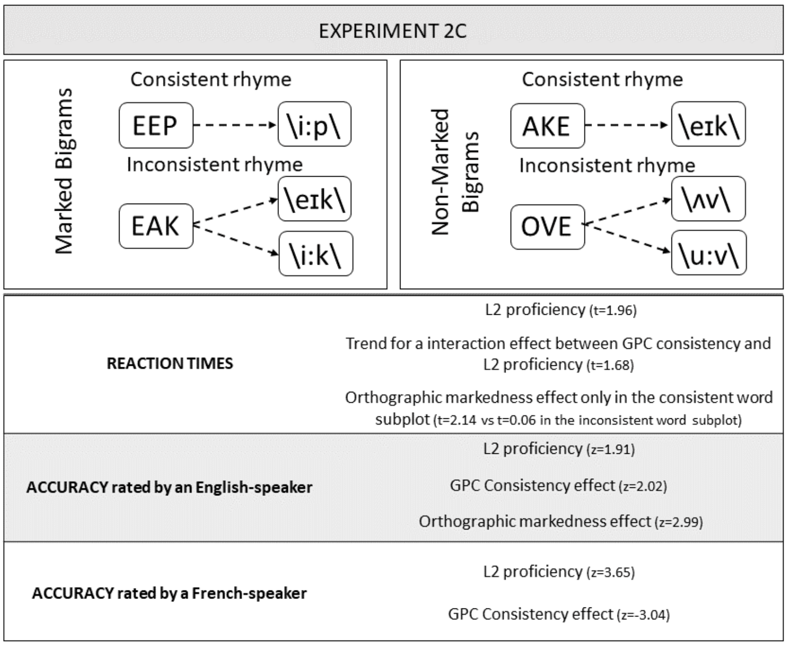


Figure 20: Summary of the results from experiment 2c

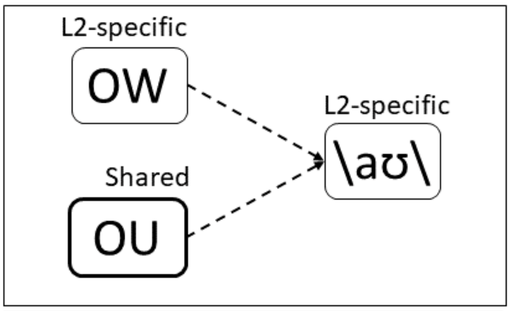
General Discussion for Experiment 2

In this second experiment we completed our investigation into three sublexical variables and their separate impact on L2-visual word processing by using three naming tasks each focusing on one of the variables (see figure 21).

Those naming tasks have been able to reveal two separate GPC congruency (experiment 2B) and GPC consistency (experiment 2C) effects. This suggest that late-learners of an L2 are indeed sensitive to cross-language GPC inconsistencies (e.g., the I in TIME does not sound like the I in French words), but also to L2 within-language inconsistencies (BEAD and DEAD are inconsistent for an English reader) even though some level of L2 proficiency might be required to develop such a sensitivity. Interestingly, while the GPC congruency effect was inhibitory with congruent items (I in FISH), leading to worse performances than with incongruent items (I in TIME), the GPC consistency effect was facilitative with consistent items (SHEEP) leading to better performances than inconsistent items (HEAR). The GPC consistency effect is easily reconciled with the broader literature and indicates that L2-learners, at least those that are proficient in their L2, activate every GPC associated with an L2 word upon its presentation. The inhibitory GPC congruency effect is more puzzling but might be explained in part by the poor perception of L2 phonological forms and partly by a stronger control of shared/congruent GPCs vs. L2-specific/incongruent GPCs when naming L2 words.

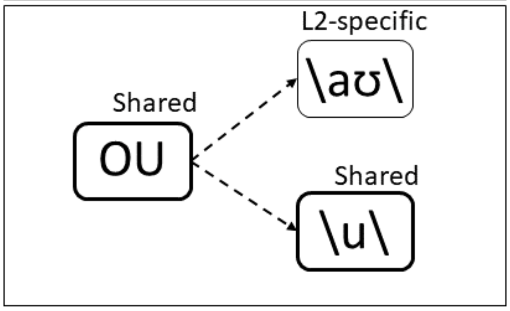
Results on the orthographic markedness variable on the other hand are rather mixed. When all other sublexical variables were controlled for (Experiment 2A), orthographic markedness did not have any impact on naming responses whereas an inhibitory effect of orthographic markedness was revealed in the consistent word subplot when manipulated in tandem with L2 GPC-consistency (Experiment 2C). Thus, it seems that overall, orthographic markedness does not hold any robust influence on L2-pronunciation processes. Only when words are strictly consistent, allowing L2-readers to rely on the sublexical pathway does orthographic markedness have any impact on naming responses. Besides, this impact is negative suggesting that new L2-GPC representations associated to orthographic markers (OW= \aʊ\) are weaker than those associated to shared orthographic patterns (OU= \aʊ\).

EXPERIMENT 2A	



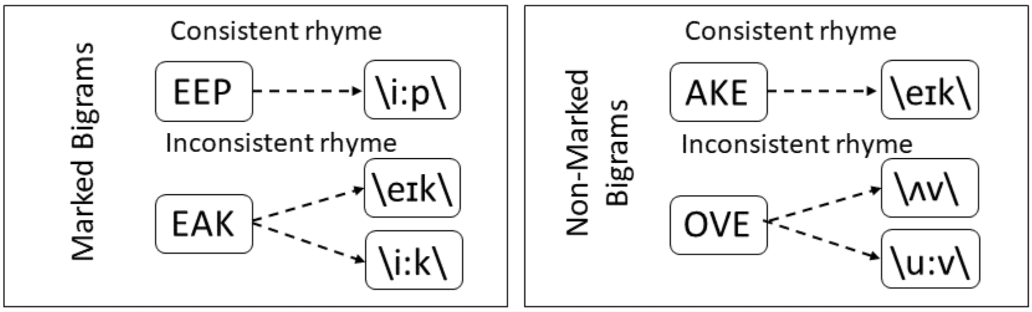
REACTION TIMES	L2 proficiency (t=2.19) No Orthographic markedness effect
ACCURACY rated by an English-speaker	L2 proficiency (z=10.49) No orthographic markedness effect
ACCURACY rated by a French-speaker	L2 proficiency (z=2.99) Orthographic markedness effect (z=2.98)

EXPERIMENT 2B	



REACTION TIMES	L2 proficiency (t=2.63) GPC Congruency effect (t=2.30) more important for HP participants
ACCURACY rated by an English-speaker	L2 proficiency (z=2.70) GPC Congruency effect (z=4.83)
ACCURACY rated by a French-speaker	L2 proficiency (z=3.50) GPC Congruency effect (z=1.83) more important for HP participants

EXPERIMENT 2C	



REACTION TIMES	L2 proficiency (t=1.96) Trend for a interaction effect between GPC consistency and L2 proficiency (t=1.68) Orthographic markedness effect only in the consistent word subplot (t=2.14 vs t=0.06 in the inconsistent word subplot)
ACCURACY rated by an English-speaker	L2 proficiency (z=1.91) GPC Consistency effect (z=2.02) Orthographic markedness effect (z=2.99)
ACCURACY rated by a French-speaker	L2 proficiency (z=3.65) GPC Consistency effect (z=-3.04)

Figure 21 : Summary of the results for experiments 2A to 2C.

General Discussion

To sum up, in this chapter we set out to investigate three sublexical variables involved in the acquisition of a second written language, which are orthographic markedness, GPC congruency and GPC consistency. We also wanted to disentangle their respective impact on L2 visual word processing from each other. To do that we chose to conduct three lexical decision tasks and three naming tasks with a group of French-native speakers that acquired English as an L2 through the school system. Each lexical decision was paired with a naming task and were designed to explore one of the variables while controlling for the others.

The exploration into orthographic markedness revealed that when GPC variables were controlled for, by selecting only incongruent and consistent words, orthographic markedness did not impact either the recognition or the pronunciation processes involved in L2 visual word processing. Furthermore, its investigation through its interaction with GPC consistency provided some interesting information on what it would take to exhibit such an orthographic markedness effect. Even more interestingly, those conditions depend on the task at hand. It seems that to find an (inhibitory) orthographic markedness effect on actual word recognition, this process needs to be slowed down by introducing inconsistent GPCs, whereas to exhibit an (inhibitory) orthographic markedness effect on pronunciation processes, the sublexical pathway of reading needs to be enabled by using strictly consistent words.

The second pair of experiments focused on GPC congruency and revealed that while it did not have any impact on word recognition processes, GPC congruency hindered L2 word pronunciation processes. This inhibitory impact was rather unexpected but the accuracy data hint at the L2 learners' misperception of the congruency of L2-words. The latency data on the other hand might reflect the stronger control L2-learners apply to shared and congruent GPCs (vs. L2-specific/ incongruent GPCs) that may sound too much like an L1-pronunciation when asked to name a list exclusively containing L2 words.

The third pair of experiments focused on GPC consistency and revealed that with increased L2-proficiency, L2-learners appear to become sensitive to L2 within-language GPC consistency. This effect seems to indicate that upon presentation of a written word, L2-learners activate every GPC appliable as long as they have acquired it. However, while GPC consistency clearly benefited pronunciation processes, its impact on actual visual word recognition was not significant. Still, a better GPC consistency manipulation might be required to explore this effect more precisely.

As a last consideration, the experiments described in this chapter have highlighted the importance of considering each one of the variables independently from each other but also the need for future studies to explore the relation they have with each other and/or to keep the others constant when investigating one of them. In fact, they allowed us to observe two separate GPC consistency effect coming from a cross-language source as well as a within-language source, but also to find evidence for the orthographic markedness dependance on GPC variations. Similarly, those experiments have underlined the importance of using several methods to gain perspective on the impact of those variables on the bilingual lexical access. In fact, the naming tasks appeared more suited to exhibit effects of GPC variations, suggesting that those may not have overly impacted L2 word recognition but seem to clearly influence L2 word pronunciation. On the other hand, neither of these tasks revealed a clear orthographic markedness effect. Yet, their joined use has managed to make light on the particular conditions required to exhibit an orthographic markedness either on the word recognition process or on the pronunciation process.

CHAPTER IV / Orthographic markedness and lexical access in context

In the previous chapter we have investigated the separate impact of three sublexical variables, i.e., orthographic markedness, cross-language GPC congruency and L2 within-language GPC consistency, on L2 visual word processing. Interestingly, orthographic markers are particularly relevant for the exploration of the bilingual lexical access, and as such, have been used to test and sometimes update the BIA+ model, especially when it comes to the language non-selectivity hypothesis. However, orthographic markers have not yet been used to test the BIA+ hypothesis following which the linguistic context may modulate the degree of language-selectivity applied to the bilingual lexical access. Thus, in this chapter we set out to explore the potential interaction between orthographic markedness and another source of language membership information, that is the larger linguistic context, and their impact on the bilingual lexical access.

As previously discussed, studies investigating this particular hypothesis of the BIA+ model have not been overly successful in exhibiting a clear modulation of the bilingual lexical access by the larger linguistic context (Gullifer et al., 2013; Lauro & Schwartz, 2017; Libben & Titone, 2009; Schwartz & Kroll, 2006; Titone et al., 2011; Van Assche et al., 2011). Studies on switch cost using sentences on the other hand, do suggest that bilinguals exert some control over their languages that may be tuned to the linguistic context (Bultena et al., 2015a, 2015b; Van Der Meij et al., 2011). Interestingly, a recent study from Hoversten and Traxler, (2020), advocate for an intermediate position, the partially selective access hypothesis, in which both languages are activated during target language processing, but at different degrees depending on the linguistic context.

The investigation into orthographic markedness has provided interesting results for understanding bilingual processing of visual words. While the advantage orthographic markers constitute for language membership detection is widely replicated (Borragan et al., 2020; Casaponsa et al., 2014; Jared et al., 2013), its impact on word recognition per se (Casaponsa & Duñabeitia, 2016; Commissaire, 2021; Commissaire, Audusseau, et al., 2019) is debated. Although some studies have found evidence that orthographic markedness only impacted word recognition when markers were associated with a decision advantage (van Kesteren et al., 2012), other studies found an automatic advantage of orthographic markedness even in tasks in which using orthographic markers was not useful to make decisions (Casaponsa et al., 2019). Orthographic markers therefore seem to constitute language membership cues that could be retrieved early enough to impact lexical access either through an inhibitory mechanism or by promoting within-language lexical activations (Casaponsa & Duñabeitia, 2016; Commissaire, 2021).

We argue that the linguistic context and the orthographic markedness variable could enhance each other's influence on the bilingual lexical access. More precisely we hypothesize that orthographic markers, by making the language membership information salient, could give an edge to bilinguals in determining the language at hand even when included in a linguistic context. Therefore, a Marked linguistic context, e.g., an English sentence containing English markers such as "These flowers grow in the jungle", would better restrain the bilingual lexical access than a Non-Marked linguistic context, e.g., an English sentence containing only

Non-Marked bigrams such as “These spiders live in the jungle”. To our knowledge this particular interaction has not yet been explored. We set out to do just that in this chapter.

Yet, the use of full sentences as a first foray into this issue seems overambitious considering we would need to closely manipulate the orthographic markedness of the linguistic context and keep a lot of other variables (e.g., lexical, semantic, sentence and word length etc…) constant, all the while preserving a meaningful sentence that would be predictable enough not to mask any potential effect of orthographic markedness.

An interesting alternative would be to use a derivation of the flanker task paired with a lexical decision task. This particular paradigm was introduced, to our knowledge, by Snell et al., (2017) to find out exactly which information do readers extract from the parafoveal vision, and whether those information impacted the response made on a target item. They focused mainly on monolingual reading and on semantic, syntactic and morphological information but critically for our purpose, those experiments also investigated whether language membership information could be extracted from the flanker items (Declerck et al., 2018). To do that they presented bilingual participants with an array of three items, one target item, e.g., DICE and two flankers which could come from the same language as the target, i.e., WOLF, or could come from the other language, i.e., LOUP, resulting in congruent trials, e.g., WOLF-DICE-WOLF and incongruent trials LOUP-DICE-LOUP. The results suggested that bilingual participants were able to extract language membership information from flanker items, and that this information impacted both language (Declerck et al., 2019; Eben & Declerck, 2019) and lexical (Declerck et al., 2018) decisions made on the target item.

Experiment 1: Language decision with a Flanker paradigm

To explore the interaction between orthographic markedness and the linguistic context, we propose to use a flanker paradigm, which provides a minimum linguistic context in which it is relatively easy to manipulate the orthographic markedness of both the target and the context (i.e., flankers) while controlling for confounding variables. This paradigm has already been used to show that bilinguals extract language membership information from the periphery, and that this information impacts the processing of a target word (Declerck et al., 2018, 2019; Eben & Declerck, 2019). However, in those studies, orthographic markedness was neither manipulated nor controlled for, leaving open the question of whether this paradigm can be used to exhibit an effect of orthographic markedness. Therefore, to address this concern, a first step in our exploration of the interaction between orthographic markedness and linguistic context was to pair the flanker paradigm with a language decision task, which has consistently exhibited an orthographic markedness effect in isolated word studies. This was done to ensure that we were in the best conditions for any orthographic markedness effect to arise in this new paradigm, thus if no orthographic markedness arise here it could mean that the flanker paradigm is not well suited to the investigation of orthographic markedness.

In this study we aimed to replicate the language congruency effect found by Declerck et al., (2019) and Eben & Declerck, (2019), in a language decision with a flanker paradigm and to investigate the impact orthographic markers may have on this effect. First, we predicted that congruent trials would lead to shorter reaction times and lower error rates than incongruent trials. Second, Orthographic markedness was not manipulated in the previous studies on the flanker task, but considering that orthographic markers facilitate language membership detection in classical language detection studies (Duñabeitia et al., 2020; van Kesteren et al., 2012), we predicted that Marked targets would lead to shorter reaction times than Non-Marked targets (Flanker-KNOW vs. Flanker-HOUSE). Third, Flanker markedness could modulate the trial's congruency effect. Marked flankers (KNIFE-Target) could give stronger language membership cues compared to Non-Marked flankers (GAME-Target), producing larger congruency effect for Marked flankers (KNIFE-HOUSE) vs. (KNIFE-JUPE) compared to Non-Marked flankers (GAME-HOUSE) vs. (GAME-JUPE). Additionally, since orthographic markedness effects on language detection tasks are asymmetrical with a greater sensitivity for L2 markers compared to L1 markers, we expect to find results along those lines for both Target and Flanker Markedness. Finally, we used two display latencies, a long display condition, i.e., 3000ms which allows for a full lexical processing of the items, and is close to the display latencies used in the previous experiments presented in this PhD work, and to the flanker tasks from Eben & Declerck, (2019), and a short display condition, i.e., 170ms, which is close to the condition in the flanker tasks from Declerck et al., (2018), and hinders the lexical processing. We expect to find some modulations of the different effect of trial congruency and orthographic markedness due to this display latency, but with no particular assumptions.

Method

Participants

Overall, 83 university students and 1 fifty-year old participated to this study via Psytoolkit (Stoet 2010, 2017). 41 participants were attributed the long display condition, in which the items were displayed during 3000ms or until a response was made. The other 43 participants were attributed the short display condition, in which the items were presented very briefly during 170ms on the screen.

For the long display condition subsample: 41 university students aged between 18 and 25 (mean age 21; sd=1.36) participated to this study. They all reported that their native and dominant language was French and they all started learning English as a second language between the ages of 5 and 12 (mean AoA= 10; sd=1.69); their L2 proficiency was measured using the LEXTALE score (Lemhöfer & Broersma, 2012; mean score= 69; sd=10.61). This sample contained 8 male and 32 female participants, 8 of them were left-handed.

For the short display condition subsample: 42 university students aged between 18 and 26 plus a fifty-year-old[33] (mean age 22; sd=4.70) participated to this study. They all reported that their native and dominant language was French and they all started learning English as a second language between the ages of 6 and 15 (mean AoA= 10; sd=2.26); their L2 proficiency was measured using the LEXTALE score (Lemhöfer & Broersma, 2012; mean score= 67; sd=7.76). This sample contained 8 male and 35 female participants, 2 of them were left-handed.

Material

English words

First, we selected a set of 120 English words from the British Lexicon Project database (BLP; Keuleers et al., 2012). Every word was monosyllabic and was 4 to 6 letter-long (mean Zipf=4.91; sd=0.58). Half of those items were marked for English, meaning that they contained

---

[33] *The fifty-year-old was attributed to the short display condition, data analysis run without this participant did not improve fit, therefore it was kept in the data plot.*

a bigram that was significatively more frequent in English than it was in French. For example, the word “FEAR” is marked for English because the ‘ea’ bigram is much less frequent in French (438 per million) than it is in English (2395 per million). To ensure that those words did indeed differ in terms of orthotactic probabilities in the two languages, we used the minimum bigram frequency measure coming from the CELEX and Lexique databases. The minimum bigram frequency (for Marked words) was significantly higher in English than it was in French (mean for English BF=909 per million, sd=586; mean for French BF=181 per million, sd=189; $p<.001$). The other half of those items were not marked for English, meaning that they contained only bigrams that were as plausible in French as they were in English. For example, the word “HARD” is not marked for English because it contains only bigrams that were legal and as frequent in English (SBF=6440) as they were in French (SBF=6498). We also used the minimum bigram frequency to ensure that these words were not marked for English. The minimum bigram frequency was similar in English than in French (mean English BF=905, sd=516; mean French BF=943, sd=596; $p=.710$). Both Marked and Non-Marked words, were matched on lexical frequency as measured by the BLP database (mean zipf for Marked words=4.99, sd=0.60; mean zipf for Non-Marked words=4.83, sd=0.54; $p=.127$), on length (mean number of letters for Marked words=4.4; sd=0.59, mean number of letters for Non-Marked words=4.3, sd=0.48; $p=.611$) and on Summed bigram frequency in English as measured by the CELEX database (mean for Marked words=6678, sd=2475; mean for Non-Marked words=6674, sd=3150; $p=.993$).

French words

Second, we selected a set of 120 French words from the Lexique database (New et al., 2001;mean Zipf= 4.83; sd=0.74). Every word was between 4 and 6 letter-long. Half of those items were marked for French, meaning that they contained a bigram that was significatively more frequent in French than it was in English. For example, the word “NEUF” is marked for French because the ‘eu’ bigram is much less frequent in English (148 per million) than it is in French (3153 per million). The minimum bigram frequency for those items (Marked for L1) was significantly higher in French than in English (mean for English BF= 561, sd=443; mean for French BF=1134, sd=818; *p*<.001). The other half of those items were not marked for French, meaning that they contained only bigrams that were as plausible in French as they were in English. For example, the word “FAIT” contained only bigrams that were legal and as frequent in English (SBF= 5848) and in French (SBF=5651). The minimum bigram frequency was similar in English than in French (mean English BF=747, sd=395; mean French BF=830, sd=383; *p*=.247). Both Marked and Non-Marked words, were matched on lexical frequency (mean zipf for Marked words=4.88, sd=0.77; mean zipf for Non-Marked words=4.78, sd=0.72; *p*=.495), on length (mean number of letters for Marked words=4.6, sd=0.78; mean number of letters for Non-Marked words=4.4, sd=0.59; *p*=.089) and on Summed bigram frequency in English (mean for Marked words=9783, sd=4105; mean for Non-Marked words=8647, sd=4089; *p*=.132).

Both sets

The two sets of French and English words were matched on lexical frequency for native speakers of each language (mean zipf for English words=4.91, sd=0.58; mean zipf for French words= 4.83, sd=0.74; *p*=.336), and on length (mean number of letters for English words=4.4, sd=0.53; mean number of letters for French words=4.5, sd=0.69; *p*=.098). Concerning bigram frequency, the two sets were matched on summed bigram frequency in English (mean for English words=6676, sd=2821; mean for French words=6627, sd=2632; *p*=.889), but not in

French (mean for English words=6651, sd=3259; mean for French words=9215, sd=4120; *p*<.001). Looking closer to the French bigram frequency, the difference seems to come from the Marked words, French Marked words used bigrams (not only the marked bigram) that were more likely to occur in French than English Marked words did (*p*<.001). Whereas both French and English Non-Marked words did not differ on this aspect (*p*=.122). This reflects the difficulty we encountered in order to find French markers. In fact, we used two different methods. To select English markers, we first referenced bigrams that were not frequent in French (e.g., SH, or OW) and then ensured those bigrams were more frequent in English. This allowed us to have Marked words with bigrams that were really rare in French (e.g., SHOW) and that lowered the summed bigram frequency (SBF in French=1347). To select French markers, we had to overcompensate for the fact that as French speakers we were less sensitive to French markers by finding bigrams that were highly frequent in French (e.g., EU or VR) and then ensure that those bigrams were not as frequent in English. This resulted in bigrams that were not so rare in English (e.g., BLEU; SBF in English=5933) but highly frequent in French, and increased the French bigram frequency (SBF in French=9713; the material used in experiment 1 and 2 and its pairing parameters can be found in Appendix 10A and 10B).

Target-Flanker pairs

In this experiment a target word will be presented alongside two flanker items (e.g., FEAR-FROG-FEAR). 40 English words from the set described above were used as flankers, the other 80 English words were used as targets. 40 French words from the set described above were used as flankers, the other 80 French words were used as targets. Meaning that the Flankers were constituted by 20 English Marked words, 20 English Non-Marked words, 20 French Marked words, and 20 French Non-Marked words. The Targets were constituted by 40

English Marked words, 40 English Non-Marked words, 40 French Marked words, and 40 French Non-Marked words. Flankers were repeated twice throughout the experiment in random association with two different targets.

We created 4 conditions (see table 25). (1) In the first condition, an English target word either Marked (i.e., FEAR) or Non-Marked (i.e., BIRD) was paired with another English word that was used as a Flanker. The flanker word could also be either Marked (i.e., FEED) or Non-Marked (i.e., SAME). This was referred to as the English-Congruent condition. (2) In the second condition, an English target word (Marked or Non-Marked) was paired with a French word that was used as a Flanker. Again the flanker could be either Marked (i.e., NEUF) or Non-Marked (i.e., VOLE). This was referred to as the English incongruent condition. (3) In the third condition, a French target word either Marked (i.e., PEUR) or Non-Marked (i.e., JUPE) was paired with another French word that was used as a Flanker. The flanker word could also be either Marked or Non-Marked (i.e., NEUF or VOLE). This was referred to as the French congruent condition. Finally, (4) in the fourth condition, a French target word (Marked or Non-Marked) was paired with an English word that was used as a Flanker. The flanker could either be Marked or Non-Marked. This was referred as the French incongruent condition.

Table 25: Summary of every trial congruency conditions

Trial Congruency	TARGET	FLANKER	Example
English-Congruent	Marked or Non-marked L2 item	Marked or non-marked L2 item	HOUSE-KNIFE-HOUSE
English-Incongruent	Marked or Non-marked L2 item	Marked or non-marked L1 item	JUPE-KNIFE-JUPE
French-Congruent	Marked or Non-marked L1 item	Marked or non-marked L1 item	JUPE-FLEUR-JUPE
French-Incongruent	Marked or Non-marked L1 item	Marked or non-marked L2 item	HOUSE-FLEUR-HOUSE

Every participant was presented with each target item only once, whether flanked by a Marked flanker or a Non-Marked flanker. Four lists were created in order to ensure that every target item was presented to the participants and that a fourth of these items was presented with a French Marked flanker, another fourth with a French Non-Marked flanker, a fourth with an English Marked flanker and another fourth with an English Non-Marked flanker.

Procedure

The flanker task was created using Psytoolkit (Stoet, 2010, 2017). Stimulus were presented in white font on a black background. There were two target presentation duration conditions: a long display condition and a short display condition. Participants were randomly assigned to one of the two conditions. In the long display condition, a fixation point appeared on the screen for 500ms and then was replaced by the target item and its two flankers that appeared on the screen until the participant gave a response or for a maximum of 3000ms. In the short display condition, a fixation point appeared on the screen for 500ms and then was replaced by the target item and its two flankers which appeared on the screen for 170ms. They were replaced by a dot that stayed present on the screen until the participant gave a response or for a maximum of 3000ms.

Participants were asked to indicate if the word appearing at the centre of the display was a French or an English word by pressing two different response keys, pressing a key with their dominant hand for the French words, and with their non-dominant hand for English words. They were explicitly asked to ignore the items at the periphery.

Background measures consisted of an English positioning task (LEXTALE; Lemhöfer & Broersma, 2012) a French positioning task (LEXTALE-Fr; Brysbaert, 2013) and a short language history questionnaire (about age of acquisition, language dominance and whether they had ever stayed in an English-speaking country).

Results of the long display condition

Table 26: Mean Language Decision Latencies (in ms) and Accuracy (in %) to Marked and Non-Marked words presented in the long display condition from Experiment 1

Reaction times (in ms)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total cible L1			total cible L1
M	759(222)	771(259)	765(242)	794(256)	766(250)	780(253)
NM	781(241)	787(274)	784(258)	816(261)	799(257)	807(260)
total flanker L1	770(232)	779(267)	775(250)	804(259)	782(254)	794(257)
Target L2						
M	757(221)	730(204)	744(213)	719(211)	718(202)	719(207)
NM	807(271)	809(273)	808(272)	770(261)	775(246)	773(253)
total flanker L2	781(248)	769(243)	776(243)	745(239)	747(226)	746(230)
Accuracy (in %)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total cible L1			total cible L2
M	97(0.16)	98(0.14)	99(0.15)	94(0.23)	96(0.21)	95(0.22)
NM	95(0.22)	95(0.22)	95(0.22)	89(0.31)	92(0.28)	91(0.30)
total flanker L1	96(0.19)	97(0.18)	97(0.19)	92(0.27)	94(0.25)	93(0.26)
Target L2						
M	99(0.10)	98(0.14)	99(0.12)	99(0.12)	98(0.13)	98(0.13)
NM	96(0.21)	94(0.24)	95(0.22)	98(0.14)	96(0.21)	97(0.18)
total flanker L2	97(0.16)	97(0.20)	97(0.18)	98(0.13)	97(0.17)	98(0.16)

Data trimming and analysis

Erroneous responses, timed out responses as well as responses above or below three times the standard deviation both by participants and by target item's type (L1-Marked, L1-Non-Marked, L2-Marked and L2-Non-Marked) were removed from the reaction time analyses. Responses under 300ms and timed-out responses were removed from the accuracy analyses.

Finally, the item FLOT that was responded with only 56% accuracy had to be removed, resulting in the exclusion of 2% of the data.

Reaction times (inverse transformed) and accuracy data were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). The models included by-participant and by-item random intercepts (i.e., model = Y ~ X (1|Participant) + (1|Target)).

Do we replicate the Trial Congruency Effect?

Model Comparison

First, we used a stepwise model comparison approach in which we compared 8 models, in order to replicate the congruency effect from Eben and Declerck. Model 1 contained the random effects. Model 2 added the effect of L2 proficiency (LEXTALE scores). Model 3 added the effect of target language (L1 or L2). Model 4 added the effect of flanker language (L1 or L2), model 5 to 7 added the interaction effect between each variable combination and model 8 added the full interaction. Since L2-proficiency prevented the models to converge on accuracy data, it was dropped from those models, resulting in a 4 model comparison for accuracy data (details about the model comparisons can be found in Appendix 11).

Reaction Times

Model 7, containing the interaction between target language and flanker language best fitted the latency data (AIC=-79887; $\chi^2(1,10)=28.66$; $p<.001$). A closer look at this model (see table 27) revealed an interaction effect between target and flanker languages ($b_7=8.2^{e-05}$; SE=1.5^{e-05}; t=5.36; see figure 23). Post-hoc comparisons exhibited a significant congruency effect for both L2 targets (L2 congruent= 745ms vs. L2 incongruent=775ms; $p<.001$) and L1 targets (L1 congruent= 774ms vs. L1 incongruent=793ms; $p=.023$). There also was an interaction effect between L2-proficiency and target language ($b_5=4.4^{e-05}$; SE=7.3^{e-07}; t=6.08; see figure 22). Overall, reaction times were longer for L1 targets than for L2 targets, but this seems to be the case only for higher-proficiency participants.

*Table 27: Final model for the congruency effect on reaction times in the long display condition (1~L2proficiency + TargetLanguage + FlankerLanguage + L2proficiency*TargetLanguage + L2proficiency*FlankerLanguage + TargetLanguage*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.6^{e-03}	1.8^{e-04}	8.97
L2 Proficiency	-3.4^{e-06}	2.6^{e-06}	-1.30
Target Language	-3.1^{e-04}	5.2^{e-05}	-5.88
Flanker Language	-1.7^{e-06}	5.1^{e-05}	-0.03
L2 Proficiency*Target Language	4.4^{e-06}	7.3^{e-07}	6.08
L2 Proficiency*Flanker Language	-4.2^{e-07}	7.3^{e-07}	-0.59
Target Language* Flanker Language	8.2^{e-05}	1.5^{e-05}	5.36
Random Effects	Variance	Standard Deviation	
Intercept Items	4.89^{e-09}	6.99^{e-05}	
Intercept Participants	2.85^{e-08}	1.69^{e-04}	
Residuals	8.62^{e-08}	2.94^{e-04}	

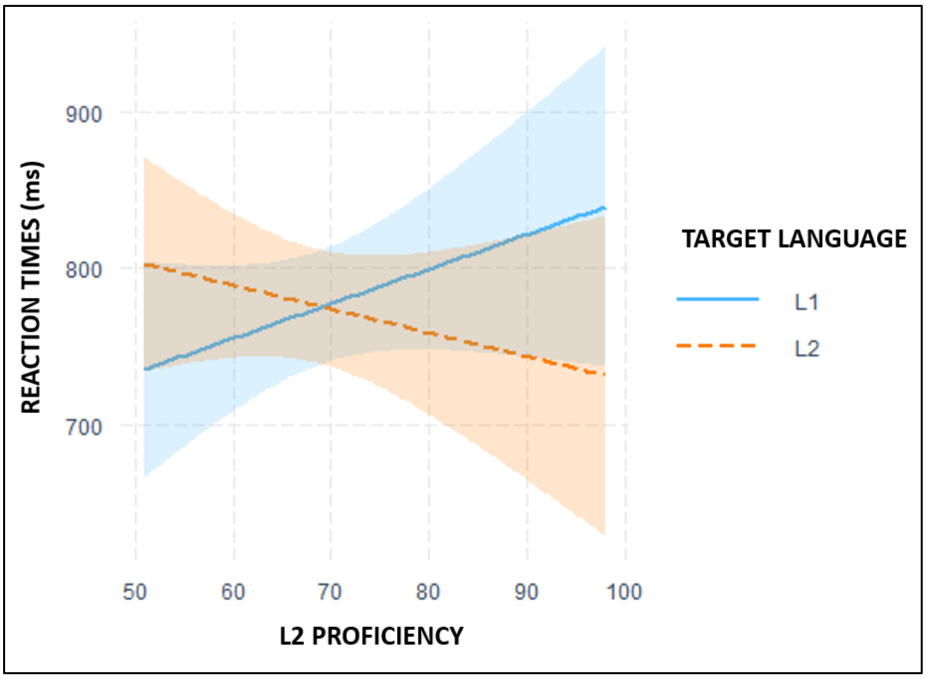


Figure 23: Interaction effect between Target language and L2 proficiency (using LEXTALE score) on reaction times in the long display condition from Experiment 1.

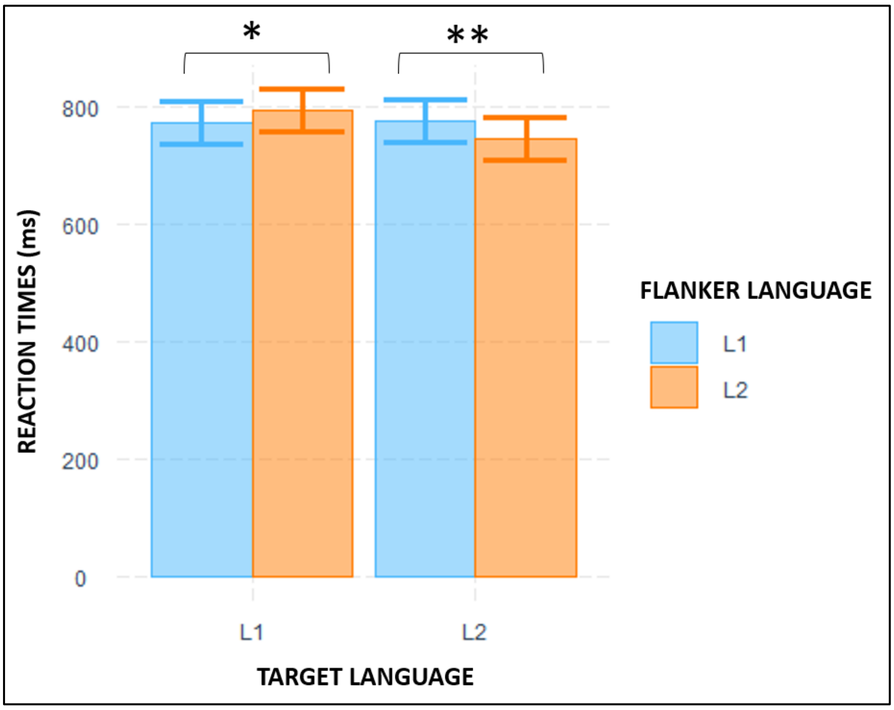


Figure 22: Interaction effect between Flanker language and Target language on reaction times in the long display condition from Experiment 1.

Accuracy

Model 4, containing the interaction effect between target and flanker languages best fitted the accuracy data (AIC=2071; $\chi^2(1,6)=15.96$; p <.001). A closer look at the model (see table 28) showed a main effect of flanker language ($b_3=-0.74$; SE=0.17; z=-4.30) with L1 flankers leading to better performances than L2 flankers (L1 Flanker=97% vs. L2 Flanker=96% accuracy). More importantly, the interaction effect between target language and flanker language was significant ($b_4=1.09$; SE=0.28; z=3.95, see figure 24), revealing higher accuracy for congruent trials compared to incongruent trials (congruent trials=98% vs. incongruent trials=95% accuracy). The post hoc comparisons revealed that the congruency effect was significant for L1 targets (L1 congruent=97% vs. L1 incongruent=93% accuracy; p<.001) but not for L2 targets (L2 congruent=98% vs. L2 incongruent=97% accuracy; p=.358).

*Table 28: Final model for the congruency effect on accuracy in the long display condition (1~ TargetLanguage + FlankerLanguage + TargetLanguage*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	z
Intercept	3.69	0.19	18.80
Target Language	0.07	0.24	0.28
Flanker Language	-0.74	0.17	-4.30
Target Language*Flanker Language	1.09	0.28	3.95
Random Effects	Variance	Standard Deviation	
Intercept Items	0.66	0.81	
Intercept Participants	0.24	0.49	

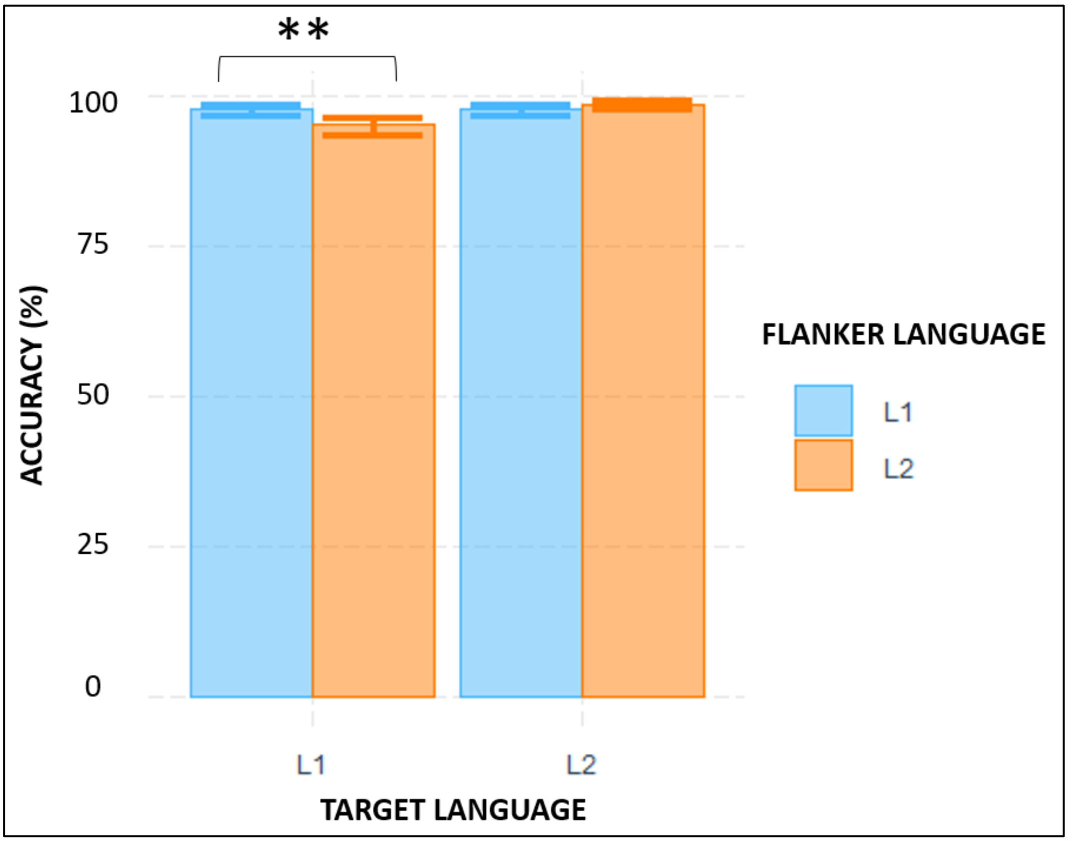


Figure 24: Interaction effect between Flanker language and Target language on accuracy in the long display condition from Experiment 1.

Does Target Markedness impact language decisions in a linguistic context?

Model Comparison

To answer this question, and considering the asymmetrical sensitivity to L1 and L2 markers (Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2002) we chose to use 2 stepwise model comparisons in which we compared 8 models on both L1 and L2 subplots.

Model 1 contained the random effects. Model 2 added the effect of L2 proficiency. Model 3 added the effect trial congruency (Congruent or Incongruent). Model 4 added the effect of target markedness (Marked or Non-Marked), model 5 to 7 added the different interaction between those factors, and model 8 added the full interaction. For accuracy data, L2-proficiency prevented the models from converging again and was therefore dropped, thus the final comparison contained only 4 models (details about the model comparisons can be found in Appendix 12).

On L1 Targets – Reaction Times. In the L1 target subplot, model 4 containing the main effect of trial congruency and of target markedness best fitted the latency data (AIC=-39263; $\chi^2(1,7)=3.11$; $p=.078$). A closer look to this model (see Table 29) revealed a main effect of trial congruency ($b_2=-3.2^{e-05}$; SE=1.1^{e-05}; t=-2.98), with congruent trials being responded to faster than incongruent trials (Congruent Trials=774ms vs. Incongruent Trials=793ms), and a trend toward a main effect of target markedness ($b_3=-3.6^{e-05}$; SE= 2.0^{e-05}; t=-1.77), with Marked targets leading to slightly faster reaction times than Non-Marked targets (Marked Targets=773ms vs. Non-Marked Targets=796ms).

Table 29: Final model for the target markedness effect on reaction times in the long display condition TL1 subplot (1~L2proficiency + TrialCongruency + TargetMarkedness)

Fixed Effects	Estimate	Standard Error	t
Intercept	1.7^{e-03}	1.8^{e-04}	9.22
L2 proficiency	-3.6^{e-06}	2.6^{e-06}	-1.40
Trial Congruency	-3.2^{e-05}	1.1^{e-05}	-2.98
Target Markedness	-3.6^{e-05}	2.0^{e-05}	-1.77
Random Effects	Variance	Standard Deviation	
Intercept Items	5.84^{e-09}	7.64^{e-05}	
Intercept Participants	2.85^{e-08}	1.69^{e-04}	
Residuals	8.40^{e-08}	2.90^{e-04}	

On L1 Targets – Accuracy. Model 3 containing the trial congruency and target markedness effect best fitted the accuracy data (AIC=1255; $\chi^2(1,6)=6.09$; $p=.014$). A closer look to this model (see table 30) revealed a main effect of trial congruency ($b_2=-0.73$; SE=0.17; z=-4.29), with congruent trials leading to better accuracy than incongruent trials (Congruent Trials=97% vs. Incongruent Trials=93% accuracy). The target markedness was also significant ($b_3=-0.63$; SE=0.25; z=-2.54), with Marked targets leading to better accuracy than Non-Marked targets (Marked targets=97% vs. Non-Marked targets=93% accuracy).

Table 30:Final model for the target markedness effect on accuracy in the long display condition TL1 subplot (1~TrialCongruency + TargetMarkedness)

Fixed Effects	Estimate	Standard Error	z
Intercept	3.97	0.24	19.83
Trial Congruency	-0.73	0.17	-4.29
Target Markedness	-0.63	0.25	-2.54
Random Effects	Variance	Standard Deviation	
Intercept Items	0.57	0.75	
Intercept Participants	0.16	0.40	

On L2 Targets – Reaction Times. In the L2 target subplot, model 4, which contained the effect of target markedness, best fitted the data (AIC=-40625; $\chi^2(1,7)=30.37$; $p<.001$). A closer look to this model (see table 31) revealed a main effect of trial congruency ($b_2=-5.1^{e-05}$; SE=1.1^{e-05};

t=-4.82), with congruent trials being responded to faster than incongruent trials (Congruent Trials=745ms vs. Incongruent Trials=775ms), and a main effect of target markedness (b_3=-8.7^{e-05}; SE=1.4^{e-05}; t=-6.05), with Marked target leading to faster reaction times than Non-Marked targets (Marked Targets= 732ms vs. Non-Marked Targets=791ms).

On L2 Targets – Accuracy. Model 3 containing the trial congruency and the target markedness effect best fitted the accuracy data (AIC=815; $\chi^2(1,6)=13.46$; $p<.001$). A closer look at this model (see table 32) revealed a trend toward a trial congruency effect (b_2=-0.35; SE=0.22; z=-1.61), with a tendency for congruent trials to lead to better accuracy than incongruent trials (Congruent Trials=98% vs. Incongruent Trials=97% accuracy). The target markedness on the other hand was significant (b_3=-1.04; SE= 0.28; t=-3.69), with Marked words leading to better accuracy than Non-Marked words (Marked words=99% vs. Non-Marked words= 96% accuracy).

Table 31: Final model for the target markedness effect on reaction times in the long display condition TL2 subplot (1~L2proficiency + TrialCongruency + TargetMarkedness)

Fixed Effects	Estimate	Standard Error	t
Intercept	1.4^{e-03}	1.9^{e-04}	7.65
L2 proficiency	7.9^{e-07}	2.7^{e-06}	0.30
Trial Congruency	-5.1^{e-05}	1.1^{e-05}	-4.82
Target Markedness	-8.7^{e-05}	1.4^{e-05}	-6.05
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	1.86^{e-09}	4.32^{e-05}	
Intercept Participants	3.09^{e-08}	1.76^{e-04}	
Residuals	8.60^{e-08}	2.93^{e-04}	

Table 32: Final model for the target markedness effect on accuracy in the long display condition TL2 subplot (1~TrialCongruency + TargetMarkedness)

Fixed Effects	Estimate	Standard Error	z
Intercept	4.63	0.30	15.63
Trial Congruency	-0.35	0.22	-1.61
Target Markedness	-1.04	0.28	-3.69
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	0.38	0.62	
Intercept Participants	0.25	0.50	

Does Flanker Markedness impact language membership decision on a target word?

Model comparison

To answer this question and considering the complexity of our design, we decided to use 2 stepwise model comparisons in which we compared 4 models on both Reaction Times and Accuracy for the Marked flanker subplot and for the Non-Marked flanker subplot separately.

Model 1 contained the random effects. Model 2 added the effect of trial congruency (congruent or incongruent). Model 3 added the effect of Flanker Language (L1 or L2), and Model 4 added the interaction between the two factors (details about the model comparisons can be found in Appendix 13).

When Flankers were MARKED – Reaction Times. In the Marked flankers subplot, model 4 containing the interaction effect best fitted the data (AIC=-39759; $\chi^2(1,7)=9.06$; $p=.003$). A closer look (see table 33) at this model revealed a main effect of flanker language ($b_3=4.93^{e-05}$; SE= 1.8^{e-05}; t=2.79), with L2 flankers leading to slightly faster reaction times than L1 flankers (L2 flankers=775ms vs. L1 flankers=776ms). The interaction between trial congruency and Flanker language was significant ($b_4=-8.6^{e-05}$; SE=2.8^{e-05}; t=-3.04; see figure 25) with the congruency effect being significant only when flankers came from L2 (Incongruent trials=804ms vs. Congruent trials=745ms; $p<.001$) and not when flankers came from L1 (Incongruent trials=781ms vs. Congruent trials=770ms; $p=.816$).

*Table 33:Final model for the target markedness effect on reaction times in the long display condition in the Marked Flankers subplot (1~TrialCongruency + FlankerLanguage+ TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.40^{e-03}	2.9^{e-05}	47.95
Trial Congruency	-1.56^{e-05}	1.8^{e-05}	-0.88
Flanker Language	4.93^{e-05}	1.8^{e-05}	2.79
Trial Congruency*Flanker Language	-8.55^{e-05}	2.8^{e-05}	-3.04
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	3.1^{e-09}	5.6^{e-05}	
Intercept Participants	2.6^{e-08}	1.7^{e-04}	
Residuals	8.8^{e-08}	2.9^{e-04}	

When Flankers were MARKED – Accuracy. Model 4 best fitted the accuracy data also (AIC=-1022; $\chi^2(1,7)=13.58$; *p*<.001). A closer look at this model (see table 34) revealed a main effect of trial congruency (b_3=0.84; SE=0.38; z=2.24), with congruent trials leading to better accuracy than incongruent trials (Congruent trials=97% vs. Incongruent trials=95% accuracy). The interaction between Flanker language and trial congruency was significant (b_4=-2.03; SE=0.55; z=-3.68; see figure 26) with the congruency effect being significant only when flankers came from L2 (Incongruent trials= 92% vs. Congruent trials=98%; *p*<.001) and not when flankers came from L1 (Incongruent trials= 96% vs. Congruent trials=97%; *p*=.714).

*Table 34: Final model for the trial congruency effect on accuracy in the long display condition marked flankers subplot (1~TrialCongruency + FlankerLanguage+TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	z
Intercept	3.72	0.26	14.11
Trial Congruency	0.36	0.34	1.06
Flanker Language	0.84	0.38	2.24
Trial congruency*Flanker Language	-2.03	0.55	-3.68
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	0.97	0.99	
Intercept Participants	0.20	0.45	

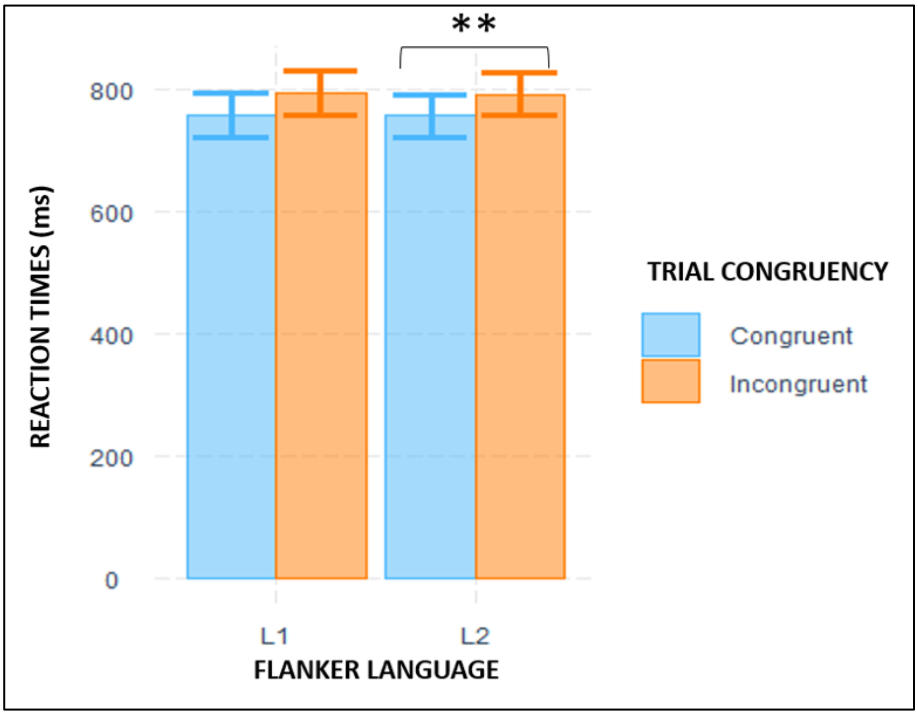


Figure 26: Interaction effect between Flanker language and Trial congruency on reaction times in the marked flankers subplot from the long display condition in Experiment 1.

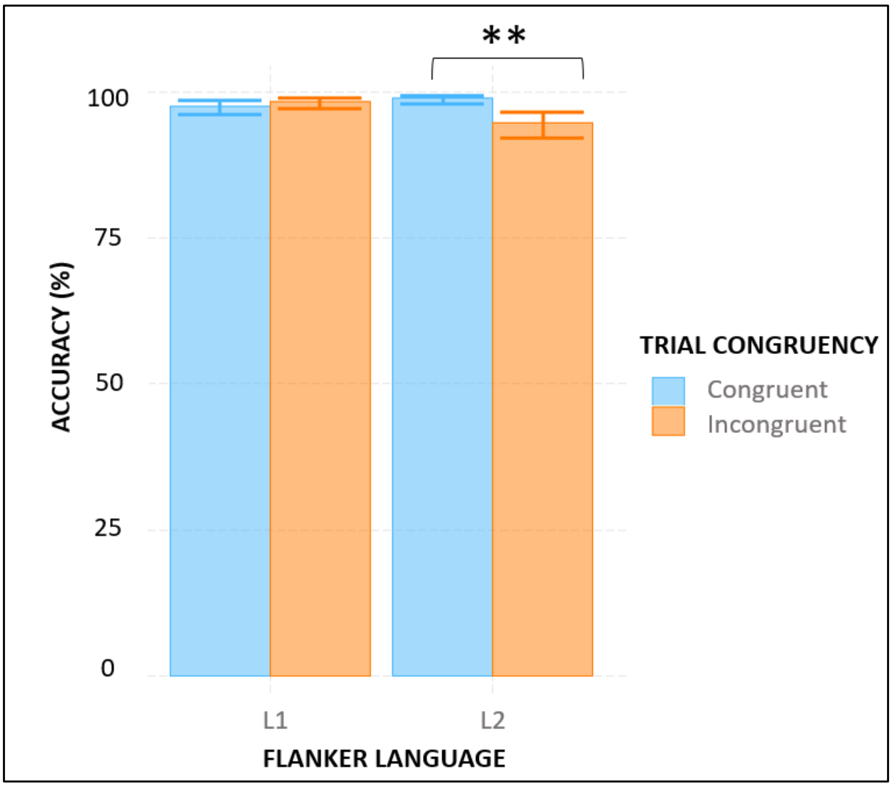


Figure 25: Interaction effect between Flanker language and Trial congruency on accuracy in the marked flankers subplot from the long display condition in Experiment 1.

When flankers were NON-MARKED – Reaction Times. In the Non-Marked flankers subplot, model 4 containing the interaction effect best fitted the data (AIC=-39941; $\chi^2(1,7)=3.64$; $p=.057$). A closer look at this model (see table 35) revealed a main effect of flanker language ($b_3=4.51^{e-05}$; $SE=1.9^{e-05}$; $t=2.27$), with L2 flankers leading to faster reaction times than L1 flankers (L2 flankers=765ms vs. L1 flankers=774ms). The interaction between trial congruency and Flanker language was significant ($b_4=-6.4^{e-05}$; $SE=3.3^{e-05}$; $t=-1.91$; see figure 27) with the congruency effect being significant only when flankers came from L2 (Incongruent trials=782ms vs. Congruent trials=747ms; $p=.037$) and not when flankers came from L1 (Incongruent trials=769ms vs. Congruent trials=779ms; $p=.957$).

*Table 35 : Final model for the target markedness effect on reaction times in the long display condition in the NonMarked Flankers subplot (1~TrialCongruency + FlankerLanguage+ TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.40^{e-03}	2.9^{e-05}	46.39
Trial Congruency	-1.01^{e-05}	1.9^{e-05}	0.51
Flanker Language	4.51^{e-05}	1.9^{e-05}	2.27
Trial Congruency*Flanker Language	-6.38^{e-05}	3.3^{e-05}	-1.91
Random Effects	Variance	Standard Deviation	
Intercept Items	6.5^{e-09}	8.1^{e-04}	
Intercept Participants	2.8^{e-08}	1.7^{e-04}	
Residuals	8.6^{e-08}	2.9^{e-04}	

When flankers were NON-MARKED – Accuracy. Model 2 best fitted the accuracy data (AIC=1078; $\chi^2(1,4)=6.53$; $p=.011$). A closer look at this model revealed a main effect of trial congruency ($b_2=-0.72$; $SE=0.20$; $z=-3.70$), with congruent trials leading to better accuracy than incongruent trials (Congruent trials=97% vs. Incongruent trials=96% accuracy).

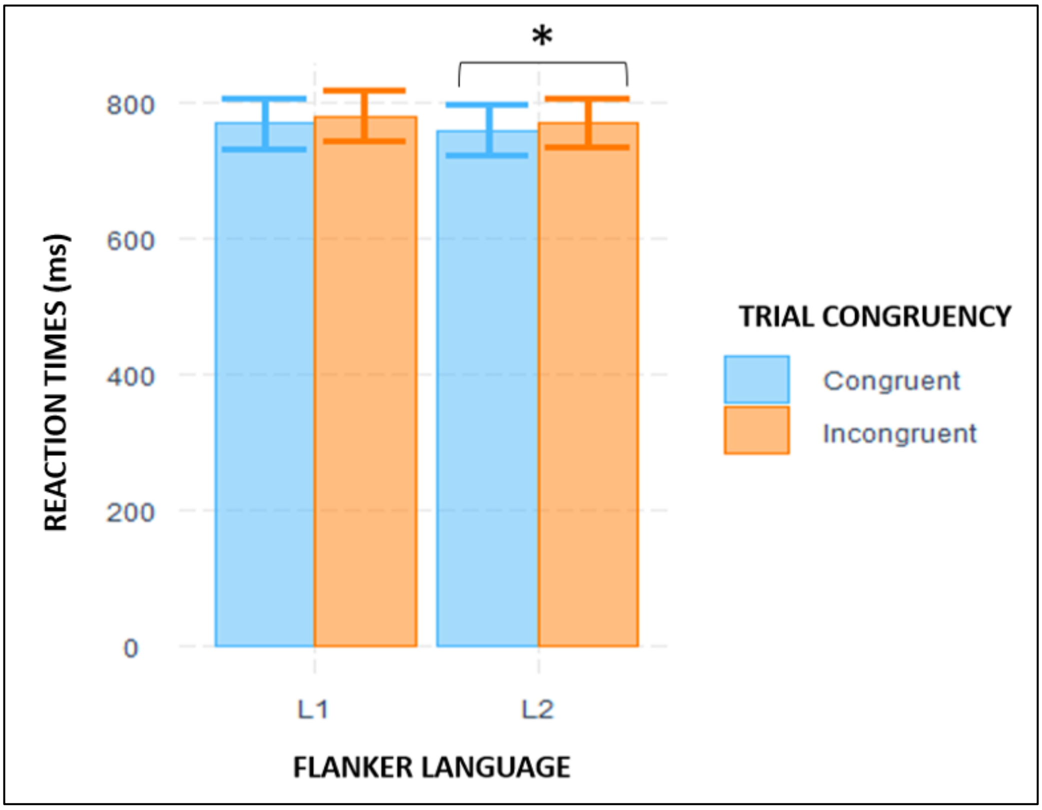


Figure 27: Interaction effect between Flanker language and Trial congruency on reaction time in the non-marked flankers subplot from the long display condition in Experiment 1

Long Display Condition – Intermediate Summary

Do we replicate the Trial Congruency Effect?

We found a Trial Congruency effect that was significant for both L1-Targets (L1-**L1**-L1 vs. L2-**L1**-L2) and L2-Targets (L2-**L2**-L2 vs. L1-**L2**-L1), indicating that participants were able to extract language membership information from the context and that those information impacted language decisions on target items.

Does Target Markedness impact language decision in a linguistic context?

We found a Target Markedness effect for both L2-Targets (Flanker-**KNOW**-Flanker vs. Flanker-**HOUSE**-Flanker) and L1-Targets (Flanker-**LIVRE**-Flanker vs. Flanker-**JUPE**-Flanker), suggesting that when enough time was given, participants could detect both L2 and L1 markers and use them to their benefit in a language decision task.

Does Flanker Markedness impact language decision on a target word?

The Trial Congruency effect was significant when the flankers were L2 words (**L2**-L2-**L2** vs. **L2**-L1-**L2**) but not when flankers were L1 words (**L1**-L1-**L1** vs. **L1**-L2-**L1**). Furthermore, this was not modulated by Flanker Markedness, suggesting that orthographic markers (sublexical information) coming from the context were not used to perform language decision tasks on a target word.

Results for the Short display condition

Table 36: Mean Language Decsision Latencies (in ms) and Accuracy (in %) to Marked and Non-Marked words presented in the short display condition from Experiment 1

Reaction times (in ms)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total target			total target
M	664(183)	683(219)	674(202)	687(219)	674(200)	680(210)
NM	670(176)	666(167)	668(167)	684(169)	687(198)	685(184)
total flanker	667(180)	674(192)	671(185)	685(197)	680(199)	683(198)
Target L2			total target			total target
M	651(178)	654(174)	652(176)	635(175)	646(173)	640(174)
NM	696(166)	689(184)	692(176)	671(184)	671(184)	671(184)
total flanker	673(174)	671(180)	672(176)	653(180)	658(179)	656(179)
Accuracy (in%)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total target			total target
M	94(0.24)	93(0.25)	94(0.24)	81(0.40)	85(0.36)	83(0.38)
NM	91(0.29)	90(0.30)	90(0.30)	77(0.42)	81(0.40)	78(0.41)
total flanker	92(0.27)	92(0.28)	92(0.28)	79(0.41)	82(0.38)	81(0.40)
Target L2			total target			total target
M	87(0.33)	88(0.32)	88(0.33)	96(0.20)	96(0.19)	96(0.20)
NM	83(0.37)	84(0.36)	84(0.37)	94(0.24)	94(0.24)	91(0.28)
total flanker	85(0.36)	86(0.34)	86(0.35)	95(0.22)	95(0.22)	95(0.24)

Data trimming and analysis

Erroneous responses, timed out responses as well as responses above or below three times the standard deviation both by participants and by target item's type (L1-Marked, L1-Non-Marked, L2-Marked and L2-Non-Marked) were removed from the reaction time analyses. Responses under 300ms and timed-out responses were removed from the accuracy analyses.

Finally, the item RUSE that was responded with only 50% accuracy had to be removed, resulting in the exclusion of 2% of the data.

Reaction times (inverse transformed) and accuracy data were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). The models included by-participant and by-item random intercepts (i.e., model = Y ~ X (1|Participant) + (1|Target)).

Do we replicate the Trial Congruency Effect?

Model comparison

First, we used a stepwise model comparison approach in which we compared 8 models, in order to replicate the congruency effect from Eben & Declerck, (2019). Model 1 contained the random effects. Model 2 contained the effect of L2 proficiency (LEXTALE scores). Model 3 contained the effect of target language (L1 or L2). Model 4 contained the effect of flanker language (L1 or L2), model 5 to 7 contained the interaction effect between each variable combination and model 8 the full interaction. The L2-proficiency factor was introduced in the models for accuracy data, but since it did not improve fit and that it prevented model convergence, it was then dropped from the final model comparison (details about model comparison can be found in Appendix 14).

Reaction Times

Model 7, containing the interaction between target language and flanker language best fitted the reaction times data (AIC=-78247; $\chi^2(1,10)=8.79$; p=.003). A closer look at this model (see table 37) revealed an interaction effect between target and flanker languages (b_7=4.7^{e-05}; SE= 1.6^{e-05}; t=2.96; see figure 28). Post-hoc comparisons exhibited a significant difference between reaction times for L2 congruent trials and L2 incongruent trials (L2 congruent=655ms vs. L2 incongruent=671ms; *p*=.005) all other relevant comparisons were not significant (*p*=.10). Reaction times were faster for L2-congruent trials (e.g., L2 target and L2 flankers) than for L2-incongruent trials (e.g., L2 target but L1 flankers). There was no significant difference between L1-congruent and L1-incongruent trials (L1 congruent=671ms vs. L1 incongruent=683ms; *p*>.10).

*Table 37: Final model for the congruency effect on reaction times in the short display condition (1~L2proficiency + TargetLanguage + FlankerLanguage + L2proficiency*TargetLanguage + L2proficiency*FlankerLanguage + TargetLanguage*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.2^{e-03}	2.5^{e-04}	4.83
L2 Proficiency	5.7^{e-06}	3.7^{e-06}	1.56
Target Language	-5.3^{e-05}	7.2^{e-05}	-0.74
Flanker Language	3.8^{e-06}	7.2^{e-05}	0.06
L2 Proficiency * Target Language	7.7^{e-07}	1.0^{e-06}	0.75
L2 Proficiency * Flanker Language	-2.2^{e-07}	1.1^{e-06}	-0.21
Target Language * Flanker Language	4.7^{e-05}	1.6^{e-05}	2.96
Random Effects	Variance	Standard Deviation	
Intercept Items	5.12^{e-09}	7.15^{e-05}	
Intercept Participants	3.16^{e-08}	1.78^{e-04}	
Residuals	8.87^{e-08}	2.98^{e-04}	

Accuracy

Model 4, containing the interaction effect between target and flanker languages best fitted the accuracy data (AIC=4098; χ^2(1,6)=184.32; *p*<.001). A closer look at the model (see table 38) showed a main effect of target language (b_2=-0.80; SE=0.16; z=-4.99) with L1 targets leading to worse performances than L2 targets (L1 target=86% vs. L2 target=90%). The main effect of flanker language (b_3=-1.26; SE=0.12; z=-10.50) with L1 flankers leading to better performances than L2 flankers (L1 Flanker=89% vs. L2 Flanker=87%). More importantly the interaction effect between those two factors was significant (b_4=2.37; SE=0.18; z=13.22; see figure 29), accuracy was higher for congruent trials than for incongruent trials (congruent trials= 93% vs. incongruent trials=83%). In this display condition, the congruency effect was significant for both L1 target (L1 congruent=92% vs. L1 incongruent=81% accuracy; *p*<.001) and L2 target (L2 congruent=94% vs. L2 incongruent=86%; *p*<.001).

*Table 38: Final model for the congruency effect on accuracy in the short display condition (1~ TargetLanguage + FlankerLanguage + TargetLanguage*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	z
Intercept	3.01	0.19	15.60
Target Language	-0.80	0.16	-4.99
Flanker Language	-1.26	0.12	-10.50
Target Language * Flanker Language	2.37	0.18	13.22
Random Effects	Variance	Standard Deviation	
Intercept Items	0.41	0.64	
Intercept Participants	0.91	0.96	

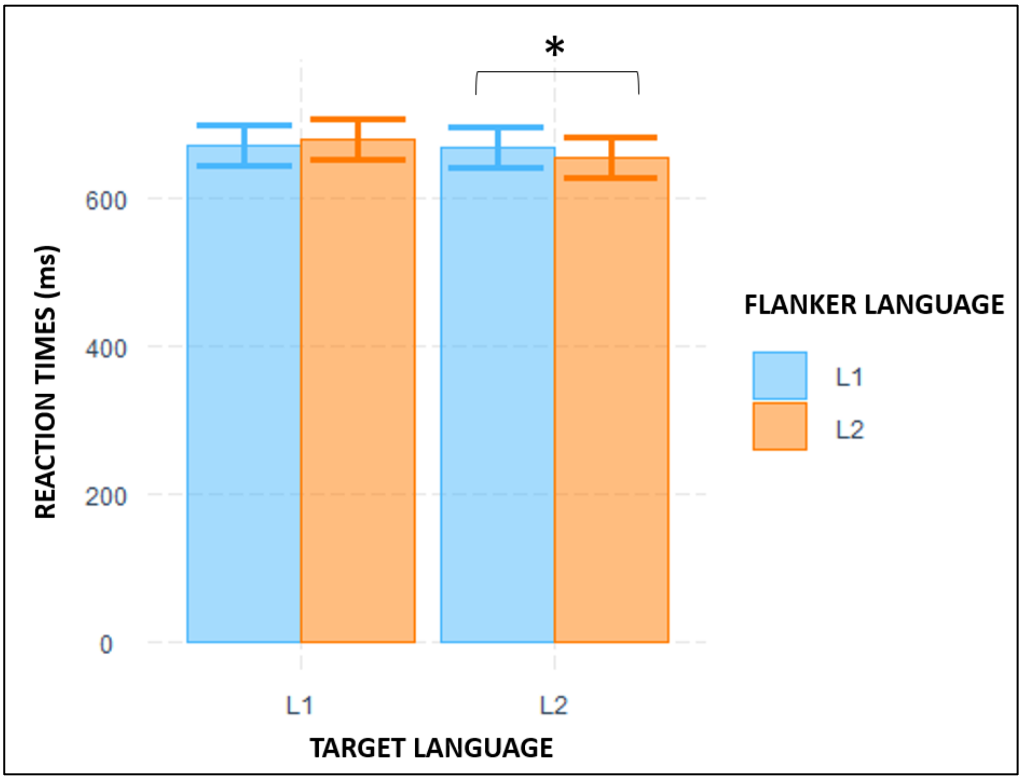


Figure 28: Interaction effect between Flanker language and Target language on reaction times in the short display condition from Experiment 1.

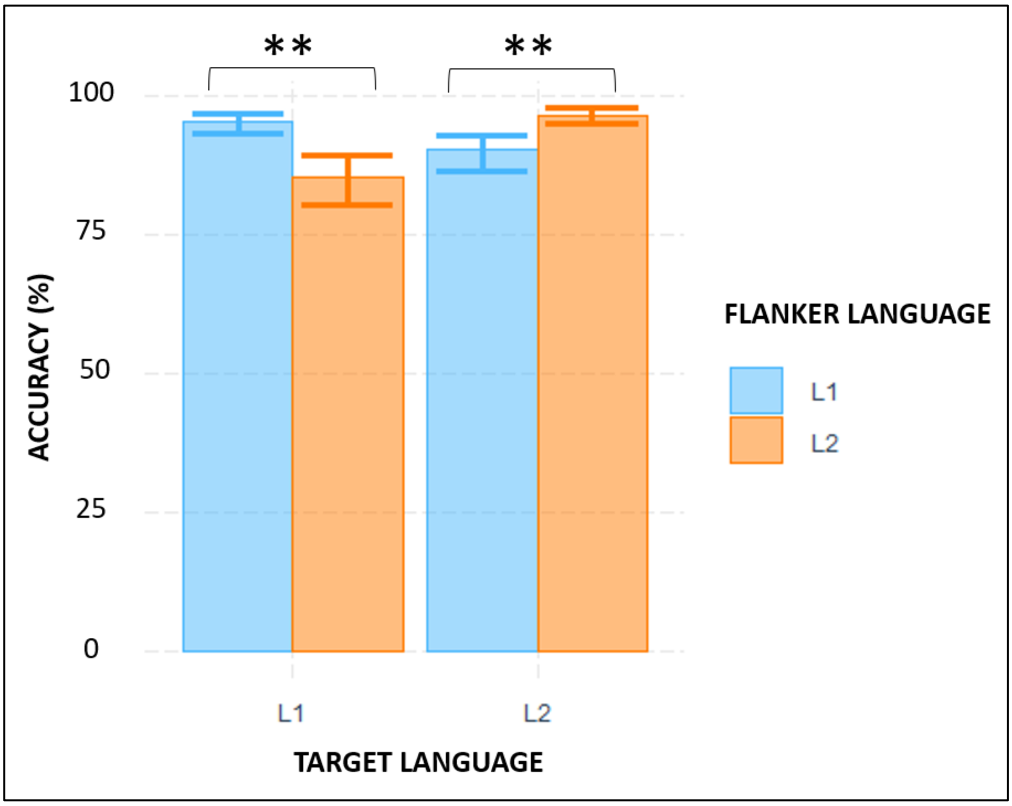


Figure 29:Interaction effect between Flanker language and Target language on accuracy in the short display condition from Experiment 1.

D***oes Target Markedness impact word recognition in a linguistic context?***

Model comparison

We used 2 stepwise model comparison in which we compared 8 models on both L1 and L2 subplots. Model 1 contained the random effects. Model 2 contained the effect of L2 proficiency. Model 3 contained the effect trial congruency (Congruent or Incongruent). Model 4 contained the effect of target markedness (Marked or Non-Marked), model 5 to 7 contained the different interaction between those factors, and model 8 contained the full interaction. For accuracy data and for the L1 subplot, the L2-proficiency prevented the models from converging and was therefore dropped, thus the final comparison contained only 4 models (details about model comparisons can be found in Appendix 15).

On L1 Targets – Reaction Times. In the subplot containing the L1 targets, none of those models improved fit for latency data (all *ps*>.10).

On L1 Targets – Accuracy. Model 3 containing the trial congruency and target markedness effect best fitted the accuracy data (AIC=2215; $\chi^2(1,5)=2.50$; $p=.113$). A closer look to this model revealed a main effect of trial congruency ($b_2=-1,28$; SE=0,12; z=-10,48), with congruent trials leading to better accuracy than incongruent trials (Congruent Trials=92% vs. Incongruent Trials=81% accuracy). The target markedness on the other hand, was not significant ($b_3=-0,32$; SE=0,19; z=-1,62).

On L2 Targets – Reaction Times. In the subplot containing the L2 targets, model 6 containing the interaction between L2-proficiency and target markedness best fitted the latency data (AIC=-40353; $\chi^2(1,9)=6,82$; $p=.009$). A closer look at this model (see table 39) showed a main effect of L2 proficiency ($b_2=8,3^{e-06}$; SE=$3,7^{e-06}$; t=2,25), with higher proficiency participants responding faster than lower proficiency participants. There also was a tendency toward a target markedness effect ($b_4=1,7^{e-04}$; SE= $9,6^{e-05}$; t=1,75), with Marked targets being responded to faster than Non-Marked targets (Marked targets=646ms vs. Non-Marked Targets= 682ms). The interaction between those two factors was significant ($b_6=-3,7^{e-06}$; SE=$1,4^{e-06}$; t=-2,61; see figure 30), the difference between Marked and Non-Marked targets being more important for higher proficiency participants.

On L2 Targets – Accuracy. Model 3 containing the trial congruency, the target markedness effect and their interaction best fitted the accuracy data (AIC=1932; $\chi^2(1,6)=3,24$; $p=.072$). A closer look to this model (see table 40) revealed a main effect of trial congruency ($b_2=-1,38$; SE=0,21; z=-6,54), with congruent trials leading to better accuracy than incongruent trials (Congruent Trials=94% vs. Incongruent Trials=86% accuracy). The target markedness was also significant (b_3=-0,86; SE=0,24; t=-3,52), with Marked words leading to better accuracy than Non-Marked words (Marked words=92% vs. Non-Marked words=88% accuracy). Finally, the interaction between trial congruency and target markedness was trending (b_4=0,49; SE= 0,27; t=1,83 see figure 31). The post-hoc test, revealed that although the congruency effect was significant for both Marked and Non-Marked targets (ps<.001), the target markedness effect was significant only in congruent trials (Marked words= 96% vs. Non-Marked words= 91%; *p*=.002) and not in incongruent trials (Marked words= 88% vs. Non-Marked words= 84%; *p*=.173).

*Table 39: Final model for the target markedness effect on reaction times in the short display condition for L2 targets (1~L2proficiency + TrialCongruency + TargetMarkedness + L2proficiency*TrialCongruency +L2proficiency*TargetMarkedness)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.1^{e-03}	2.5^{e-04}	4.38
L2 proficiency	8.3^{e-06}	3.7^{e-06}	2.25
Trial Congruency	-1.6^{e-05}	9.7^{e-05}	-0.17
Target Markedness	1.7^{e-04}	9.6^{e-05}	1.75
L2 proficiency*Trial Congruency	-2.7^{e-07}	1.4^{e-06}	-0.19
L2 proficiency*Target Markedness	-3.7^{e-06}	1.4^{e-06}	-2.61
Random Effects	Variance	Standard Deviation	
Intercept Items	3.0^{e-09}	5.5^{e-05}	
Intercept Participants	3.1^{e-08}	1.8^{e-04}	
Residuals	8.4^{e-08}	2.9^{e-04}	

*Table 40: Final model for the target markedness effect on accuracy in the short display condition for L2 targets (1~ TrialCongruency + TargetMarkedness +TrialCongruency*TargetMarkedness)*

Fixed Effects	Estimate	Standard Error	z
Intercept	3.74	0.25	14.99
Trial Congruency	-1.38	0.21	-6.54
Target Markedness	-0.86	0.24	-3.52
Trial Congruency*Target Markedness	0.49	0.27	1.83
Random Effects	Variance	Standard Deviation	
Intercept Items	0.21	0.45	
Intercept Participants	0.92	0.96	

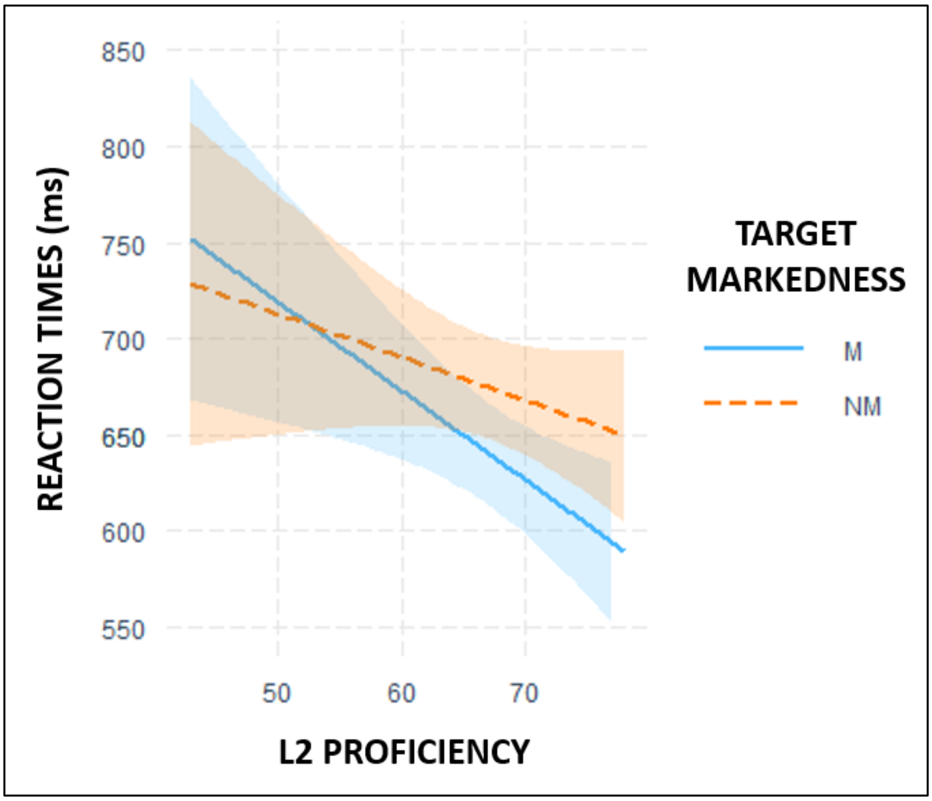


Figure 30: Interaction effect between Target markedness and L2 proficiency (using LEXTALE score) on reaction times in the L2 target subplot of the short display condition from Experiment 1

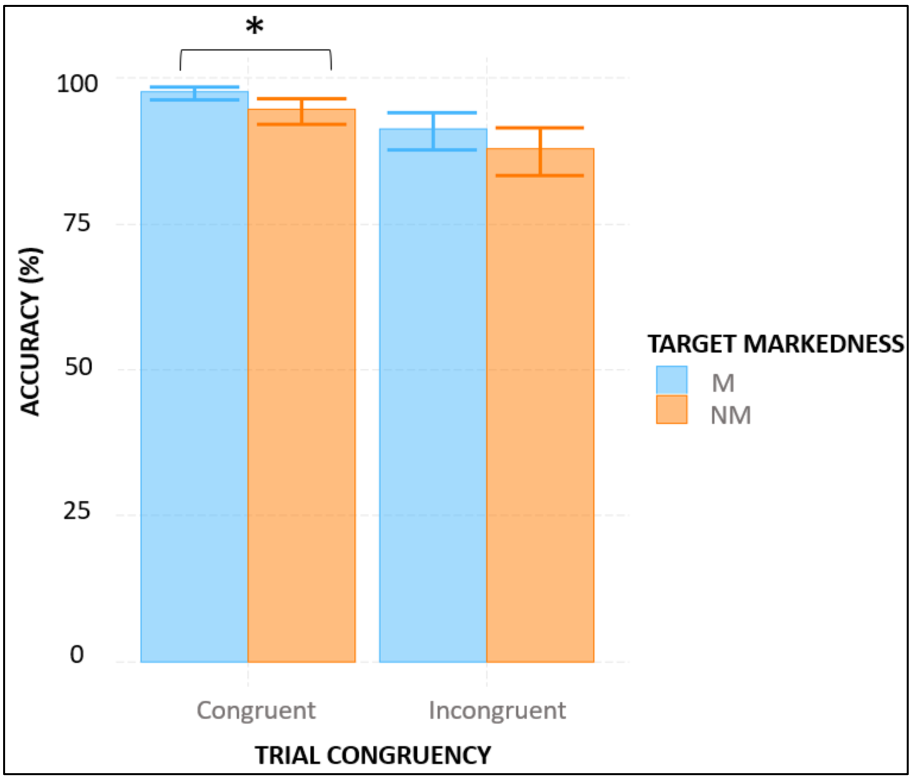


Figure 31: Interaction effect between Target Markedness and Trial congruency on accuracy in the L2 target subplot of the short display condition from Experiment

Does Flanker Markedness impact target recognition?

Model comparison

We used 2 stepwise model comparison in which we compared 4 models in both Marked and Non-Marked flankers.

Model 1 contained the random effects. Model 2 contained the effect of trial congruency (congruent or incongruent). Model 3 contained the effect of Flanker Language (L1 or L2), and Model 4 contained the interaction between the two factors (details about model comparisons can be found in Appendix 16).

When Flankers were MARKED – Reaction Times. In the Marked flankers subplot, the second model best fitted the data (AIC=-38896; $\chi^2(1,5)=9.54$; $p=.002$). A closer look at this model (see table 41) revealed a trial congruency effect ($b_2=-3.47^{e-05}$; $SE=1.1^{e-05}$; $t=-3.09$), with congruent trials leading to better performances than incongruent trials (congruent trials=660ms vs. incongruent trials=679ms). The model 4 containing the interaction tended to improve fit (AIC=-38896; $\chi^2(1,7)=3.01$; $p=.083$), the interaction between trial congruency and Flanker language was indeed trending ($b_4=5.5^{e-05}$; $SE=3.2^{e-05}$; $t=-1.74$; see figure 32) with the congruency effect being significant only when flankers came from L2 (Congruent trials=685ms vs. Incongruent trials=653ms; $p=.008$) and not when flankers came from L1 (Congruent trials= 667ms vs. Incongruent trials=673ms; $p>.10$).

*Table 41: Final model for the target markedness effect on reaction times in the short display condition for Marked flankers ((1~ TrialCongruency + FlankerLanguage + TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.58^{e-03}	3.14^{e-05}	50.47
Trial Congruency	-6.86^{e-06}	1.94^{e-05}	-0.35
Flanker Language	3.65^{e-05}	1.89^{e-05}	1.93
Trial Congruency*Flanker Language	-5.53^{e-05}	3.19^{e-05}	-1.74
Random Effects	Variance	Standard Deviation	
Intercept Items	5.26^{e-03}	7.25^{e-05}	
Intercpet Participants	3.45^{e-08}	1.86^{e-04}	
Intercept Participants	8.72^{e-08}	2.95^{e-04}	

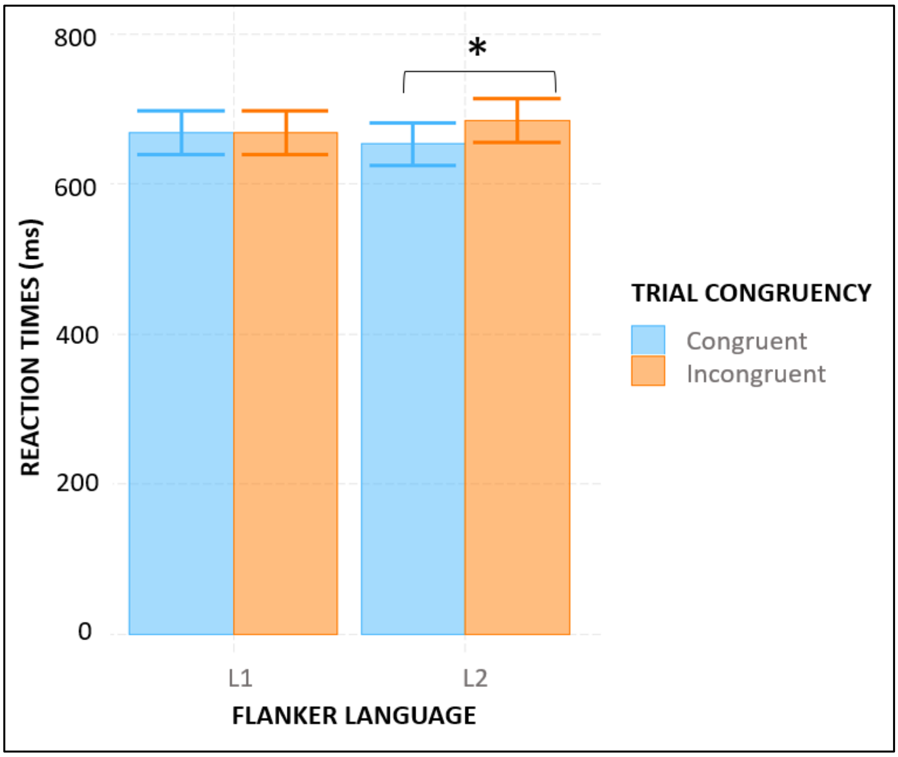


Figure 32: Interaction between Trial Congruency and Flanker language on reaction times in the Marked flankers subplot from the short display condition in Experiment 1

When Flankers were MARKED – Accuracy. Model 3, containing the effect of trial congruency and flanker language best fitted the accuracy data (AIC=2078; $\chi^2(1,5)=9.87$; p=.001). A closer look at this model revealed an effect of trial congruency (b_2=-1.45; SE=0.13; z=-11.14), with congruent trials leading to higher accuracy than incongruent trials (congruent trials=94% vs. incongruent trials=82% accuracy). The flanker language effect was also significant (b_3=0.51;

SE=0.16; z=3.25) with L2 flankers leading to worse performances than L1 flankers (L1 flankers=89% vs. L2 flankers=87% accuracy).

When Flankers were NON-MARKED – Reaction Times. In the Non-Marked flankers subplot, no model improved fit (ps >.10).

When Flankers were NON-MARKED – Accuracy. Model 3, containing the effect of trial congruency and flanker language best fitted the accuracy data (AIC=2107; $\chi^2(1,5)=3.82$; $p=.051$). A closer look at this model revealed an effect of trial congruency ($b_2=-0.96$; SE=0.12; z=-7.88), with congruent trials leading to better accuracy than incongruent trials (congruent trials=94% vs. incongruent trials=84% accuracy). The flanker language effect was also significant ($b_3=0.33$; SE=0.17; z=1.99) with L2 flankers leading to better performances than L1 flankers although they were numerically similar (L1 flankers= 89% vs. L2 flankers= 89% accuracy).

Short Display Condition – Intermediate Summary

Do we replicate the Trial Congruency Effect?

We found a Trial Congruency effect that was significant *only* for L2-Targets (L2-**L2**-L2 vs. L1-**L2**-L1) but not for L1-Targets (L1-**L1**-L1 vs. L2-**L1**-L2). This may indicate that participants were able to extract language membership information from the context, but it seems that those information impacted language decisions only on L2-target items, perhaps because L2 lexical representations were weak.

Does Target Markedness impact language decision in a linguistic context?

We found a Target Markedness effect *only* for L2-Targets (Flanker-**KNOW**-Flanker vs. Flanker-**HOUSE**-Flanker) and not for L1-Targets (Flanker-**LIVRE**-Flanker vs. Flanker-**JUPE**-Flanker). Moreover, the facilitative effect of L2-markers was more important for higher-proficiency participants than for lower-proficiency participants (on reaction times). This pattern suggests that, when proficient enough, participants were able to extract L2-markers from a target word and to use it to speed up a language decision task.

Does Flanker Markedness impact language decision on a target word?

The Trial Congruency effect was significant only when the flankers were marked for L2 (**KNIFE**-L2-**KNIFE** vs. **KNIFE**-L1-**KNIFE**) and not when the flankers were non-marked (GAME-L2-GAME vs. GAME-L1-GAME, or JOUR-L1-JOUR vs. JOUR-L2-JOUR) or when they were marked for L1 (NEUF-L1-NEUF vs. NEUF-L2-NEUF). This pattern suggests that the presence of L2-markers in the flanker words increased the saliency of the language membership information coming from the “linguistic context”.

Discussion for Experiment 1

For the sake of clarity, as this experiment called out a lot of variables such as language membership (L1 and L2) of the target and of the flankers, orthographic markedness (Marked vs. Non-Marked) of the target and of the flankers, and display condition (long vs. short), we will separate our discussion in three parts, corresponding to the three main effects we expected to find. Also, to ensure fluency we will rearrange their order starting by discussing the targets' orthographic markedness effect, then the trial congruency effect, and finally the dynamic between flankers' orthographic markedness effect and trial congruency (see figure 33).

The target markedness effect exhibited here replicated the orthographic markedness advantage for language detection largely found in the literature (Borragan et al., 2020; Casaponsa et al., 2014; Jared et al., 2013), and suggests that participants were able to use the orthographic markers contained in target items to sort them according to their language membership, even when additional information were provided. The effect of target markedness was different depending on the display duration condition. First, in the shorter duration condition, the target markedness effect was significant only in the L2-targets subplot, but not in the L1-targets subplot, replicating the asymmetry concerning orthographic markedness in language decision tasks, often described in the literature (Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2002), with a greater sensitivity to L2-markers compared to L1-markers. However, L1-markedness tended to impact responses in the longer display condition, suggesting that with longer display duration, participants could retrieve L1-markers, but that their impact on language membership decision was less reliable than the one coming from L2-markers.

Furthermore, in the short display condition, the target markedness effect interacted with L2-proficiency, the advantage for Marked targets being more important in Higher-Proficiency participants than it was for Lower-Proficiency participants. This is not completely in line with the literature which found that even monolinguals could use L2-markers to make language decisions, but may be attributed to the short display condition. The absence of such an interaction in the longer display condition, suggests that when the target item was available long enough, even lower-proficiency participants could detect orthographic markers and use them to ease their language decisions.

The target markedness effect on accuracy was also different in both display duration condition. In the shorter duration condition, only L2 flanker impacted accuracy and this only when the trial was congruent, indicating that an L2 Marked target could improve participants' performances but only if no conflict arose from the linguistic context. In longer display duration however, both L1 and L2 markers improved performances whether a conflict emerged or not from the linguistic context.

The Trial Congruency effect, which was reliably found across display conditions in this experiment, is in line with the results from Declerck et al., (2018) and from Eben & Declerck, (2019). This effect suggests that participants were able to extract language membership information from the context (via the flanker items) and that this impacted both the accuracy and the speed of the decisions they made on the target items' language membership. The present experiment also found evidence that the effect of trial congruency was modulated by the display duration. In fact, similarly to previous studies, the longer display condition (3000ms) revealed that trial congruency impacted reaction times for both L1 and L2 targets. On the other hand, the short display condition (170ms) we introduced revealed that trial congruency impacted reaction times only for L2 targets. At first glance, this could be interpreted in the following way; in the short display condition, L2 targets' language membership representation might be weakly activated and thus be impacted by information (most likely sublexical considering the task's constraints) coming from the flankers, whereas L1 targets' language membership might already be strong enough not to be impacted by information coming from the flankers. In the longer display condition, targets and flankers might interfere with each other on a lexical level (vs. sublexical) and consequently, both L1 and L2 targets might be impacted by trial congruency. However, this interpretation is not in line with the accuracy data since the trial congruency was significant on both L1 and L2 targets, for shorter display condition (170ms) and only on L1 targets for longer display condition (3000ms).

To better understand this modulation of the trial congruency by display duration, it might be useful to look at the interaction between trial congruency and flanker markedness. Interestingly at longer exposure, the trial congruency effect relied on L2 flankers, without any modulation of flanker markedness. These results might corroborate the hypothesis that at longer display duration, the conflict is on lexical representations (vs. through sublexical representations). When full lexical process is permitted, the presence of an L2 flanker will be able to speed up decision on L2 targets and/or to slow down decision on L1 targets, no matter if they (the flankers) are Marked or not.

Even more interestingly, in the short display duration the trial congruency effect is further modulated by the markedness of the context (i.e., Flanker markedness) since this effect was restricted to trials in which flankers were Marked for L2 (e.g., KNIFE), and was not significant when flankers were not Marked (e.g., HOUSE or JUPE). Suggesting that L2 Marked flankers made the language membership information coming from the context more salient, enabling the information on language membership coming from this context to impact the speed with which a decision will be reached as to the language membership of the target items, whereas conflicting information coming from a Non-Marked flanker would easily be ignored and therefore will not impact responses to the target item. Taken together with the overall analysis on the congruency effect discussed earlier, these results suggest that at short exposure, the congruency effect relies on L2 Marked flankers, that can help decisions on L2 targets but do not hinder decisions on L1 targets.

As for the accuracy data, looking closer at the interaction between trial congruency and flanker markedness did not clarify the mechanism behind the trial congruency effect, since it was not modulated by flanker markedness in any display duration condition. Maybe any modulation was erased by the high accuracy with which participants sorted the target items according to their language membership (as low as 79% in the short display condition and as high as 99% in the long display condition).

		LONG DISPLAY	SHORT DISPLAY
L1 TARGET MARKEDNESS	RT	Trend toward a markedness effect (t=1.77)	No markedness effect
	ACC	Facilitative markedness effect (z=-2.54)	No markedness effect
L2 TARGET MARKEDNESS	RT	Facilitative markedness effect (t=-6.05)	Markedness effect more important for HP participants (t=-2.61)
	ACC	Facilitative markedness effect (z=-3.69)	Facilitative markedness effect (z=-3.52)
TRIAL CONGRUENCY	RT	Trial congruency effect (t=5.36)	Trial congruency effect (t=2.96) L2 targets (*p*=.005) vs L1 targets (*p*>.10)
	ACC	Trial congruency effect (z=3.95) L1 targets (*p*<.001) vs L2 targets (*p*=.358)	Trial congruency effect (z=13.22)
TRIAL CONGRUENCY **Marked Flankers**	RT	Interaction with Flanker language (t=-3.04) L2 flankers (*p*<.001) vs L1 flankers (*p*=.816)	Trend for trial congruency effect (t=-1.74) L2 targets (*p*=.008) vs L1 targets (*p*>.10)
	ACC	Interaction with Flanker language (z=-3.68) L2 flankers (*p*<.001) vs L1 flankers (*p*=.714)	Trial congruency effect (z=-11.14)
TRIAL CONGRUENCY **Non-Marked Flankers**	RT	Interaction with Flanker language (t=-1.91) L2 flankers (*p*=.037) vs L1 flankers (*p*=.957)	No Trial Congruency effect
	ACC	Trial congruency effect (z=-3.70)	Trial congruency effect (z=-7.88)

Figure 33: Summary of the results from Experiment 1

Experiment 2: General Lexical Decision with a Flanker Paradigm

Introduction

In the first experiment described here, we showed that, under favourable circumstances (i.e., using a language decision task) the flanker task was able to exhibit an orthographic markedness effect on target word processing, and could also modulate the impact the linguistic context (flankers) had over the target word processing. Hence, it appears that the flanker paradigm is suited to our purpose. However, this first experiment does not allow us to conclude on the actual impact orthographic markers included in a linguistic context may have on visual word recognition *per se*. Besides, orthographic markers were tied to a decision advantage in the first experiment, which might explain the facilitative effect for orthographic markedness we found. Thus, to investigate the interaction between orthographic markedness and linguistic context and its impact on lexical access, we need to pair the flanker paradigm with a lexical decision task in which the markers were not associated to a decision advantage.

To do that, we conducted a non-blocked (both English and French items were mixed) general lexical decision task (in which participants should respond YES for both French and English words and NO for pseudowords), in which target words and pseudowords, but also flanker words and pseudowords could either be Marked or not. Importantly, a target word was always surrounded by flanker *words* coming from the same language or from the other language. On the other hand, target pseudowords were always surrounded by flanker *pseudowords* that resembled the same language it did or that resembled the other language. Hence, the conflict introduced between target and flanker items in our experiment was never on the lexical status, but on the language membership information. This was done to ensure that orthographic markers were not tied to any decision advantage.

Therefore, just as in the language decision task and the previous experiments using a lexical decision task (Chapter III), we predicted that congruent trials would lead to shorter reaction times and lower error rates than incongruent trials. Second, we looked for an effect of target markedness (Flanker-KNOW vs. Flanker-HOUSE) even though the results on the isolated word experiments presented in the previous chapter were not encouraging. Third, Flanker markedness could modulate the trial’s congruency effect. If orthographic markers introduced in the context restrict the bilingual lexical access toward their target language, we might expect Marked flankers to strengthen the congruency effect (KNIFE-HOUSE) vs. (KNIFE-JUPE) compared to Non-Marked flankers (GAME-HOUSE) vs. (GAME-JUPE). Finally, we used two display durations, a longer display condition and a short display condition which is close to the condition in the flanker tasks from Declerck et al., (2018), and restrict the lexical processing. We expected to find some modulations of the different effect of trial congruency and orthographic markedness due to this display latency, but with no particular assumptions.

Method

Participants

Overall, 97 university students and 1 forty-three year-old participated to this study via psytoolkit (Stoet 2010, 2017). 54 participants were attributed the long display condition, in which the items were displayed during 3000ms or until a response was made. The other 44 participants were attributed the short display condition, in which the items were presented very briefly during 170ms on the screen.

For the long display condition subsample: 54 university students aged between 18 and 33 (mean age= 22; sd=3.97) participated to this study. They all reported that their native and dominant language was French and they all started learning English as a second language between the age of 5 and 15 (mean AoA= 10; sd=2.15), their L2 proficiency was measured using the LEXTALE score (Lemhöfer & Broersma, 2012; mean score=75; sd=9.98). This sample contained 6 male and 41 female participants, 5 of them were left-handed.

For the long display condition subsample: 44 university students aged between 18 and 33 plus a forty-three year-old[34] (mean age= 23; sd=5.38) participated to this study. They all reported that their native and dominant language was French and they all started learning English as a second language between the age of 5 and 14 (mean AoA= 10; sd=1.85), their L2 proficiency was measured using the LEXTALE score (Lemhöfer & Broersma, 2012; mean score= 74; sd=10.88). This sample contained 9 male and 34 female participants, 2 of them were left-handed.

Material

Words. Same as experiment 1

Pseudowords. A set of pseudowords was created for the lexical decision task (the full list can be found in Appendix 17). Those pseudowords were created by replacing one letter from the selected words. They were distributed in the four categories used for the words. 120 pseudowords were Marked for English (e.g.,FEAM), 120 were Non-Marked and based on English words (e.g., VAME), 120 were Marked for French (e.g., FEUR) and the last 120 pseudowords were Non-Marked and based on French words (e.g., NOLE). Even though they

[34] *The forty-three year old was attributed to the short display condition, data analysis run without this participant did not improve fit, therefore it was kept in the data plot.*

were based on actual English and French words, Non-Marked pseudowords did not contain any orthographic pattern that could indicate its language membership. French-Marked pseudowords had a significantly higher minimum bigram frequency in French (mean=887; sd=642) than in English (mean=518; sd=431; $p<.001$). English-Marked pseudowords had significantly a higher minimum bigram frequency in English (mean=903; sd=589) than in French (mean=181; sd=190; $p<.001$). Both Non-Marked sets did not significantly differ as to their minimum bigram frequency between languages ($ps>.10$).

Every target word was flanked by another word, and every target pseudoword was flanked by another pseudoword. As in the first experiment there were 4 conditions: (1) the English-Congruent condition (e.g., SAME-FEAR-SAME or FEED-BIRD-FEED), (2) the English incongruent condition (e.g., VOLE-FEAR-VOLE or NEUF-BIRD-NEUF), (3) the French Congruent condition (e.g., VOLE-PEUR-VOLE or NEUF-JUPE-NEUF) and (4) the French Incongruent condition (e.g., SAME-PEUR-SAME or FEED-JUPE-FEED).

Procedure

Same as experiment 1, except that participants were asked to indicate if the word appearing at the centre of the display was a word, whether in French or in English by pressing a response key on their dominant side, or if the item did not exist in either language by pressing a response key on their non-dominant side.

Results for the Long Display condition

Table 42: Mean Lexical Decision Latencies (in ms) and Accuracy (in %) to Marked and Non-Marked words presented in the long display condition from Experiment 2

Reaction times (in ms)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total target			total target
M	712(181)	722(193)	717(187)	742(197)	730(197)	736(197)
NM	733(195)	720(202)	727(198)	752(198)	753(216)	752(207)
total flanker	722(188)	721(197)	722(193)	747(198)	741(207)	744(202)
Target L2			total target			total target
M	779(212)	767(217)	773(214)	773(205)	754(204)	764(205)
NM	783(226)	779(198)	781(213)	787(241)	776(206)	782(224)
total flanker	781(219)	773(208)	777(214)	780(224)	765(205)	773(215)
Accuracy (in%)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total target			total target
M	95(0.22)	94(0.23)	95(0.23)	93(0.26)	93(0.26)	93(0.26)
NM	93(0.25)	93(0.25)	93(0.25)	91(0.28)	89(0.31)	90(0.30)
total flanker	94(0.24)	94(0.24)	94(0.24)	92(0.26)	91(0.28)	92(0.28)
Target L2			total target			total target
M	94(0.24)	93(0.26)	93(0.25)	94(0.23)	94(0.25)	94(0.24)
NM	93(0.26)	92(0.27)	92(0.26)	95(0.21)	93(0.26)	94(0.24)
total flanker	93(0.25)	93(0.26)	93(0.26)	95(0.22)	93(0.25)	94(0.24)

Data trimming and analysis

Erroneous responses, timed out responses as well as responses above or below three times the standard deviation both by participants and by target item's type (L1-Marked, L1-Non-Marked, L2-Marked and L2-Non-Marked) were removed from the reaction time analyses. Responses under 300ms and timed-out responses were removed from the accuracy analyses. Finally, the words BROOM and ROPE, and the pseudowords DEAP, CROU and PRUME that were responded with only 52 to 74% accuracy, which was below three standard deviation from the mean, had to be removed, resulting in the exclusion of 1,89% of the data in the word plot and 1.86% from the pseudoword plot.

Reaction times (inverse transformed) and accuracy data were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). The models included by-participant and by-item random intercepts (i.e., model = Y ~ X (1|Participant) + (1|Target)).

Do we replicate the trial congruency effect?

Model comparison

First, we used a stepwise model comparison approach in which we compared 8 models, in order to replicate the congruency effect from Eben & Declerck, (2019).

Model 1 contained the random effects. Model 2 contained the effect of L2 proficiency (LEXTALE scores). Model 3 contained the effect of target language (L1 or L2). Model 4 contained the effect of flanker language (L1 or L2), model 5 to 7 contained the interaction effect between each variable combination and model 8 the full interaction. Since L2-proficiency prevented the models to converge on accuracy data, it was dropped from those models, resulting in a 4 model comparison for accuracy data (detail about model comparisons can be found in Appendix 18).

On Reaction Times. Model 7, containing the interaction between target language and flanker language best fitted the latency data (AIC=-104289; $\chi2(1,10)=17.33$; $p<.001$). A closer look at this model (see table 43) revealed an interaction effect between target and flanker languages (b7=4.9e-05; SE=1.2e-05; t=4.17; see figure 34). Post-hoc comparisons shows that the trial congruency effect was significant for L1 targets (L1 congruent= 722ms vs. L1 incongruent= 744ms; p<.001) but was not significant for L2 targets (L2 congruent= 773ms vs. L2 incongruent= 777ms; p=.873). There also was an interaction effect between L2-proficiency and target language ($b_5=1.4^{e-06}$; $SE=5.9^{e-07}$; $t=2.42$; see figure 35) overall, reaction times were longer for L2 targets than for L1 targets, but this difference seems to be more important for lower-proficiency participants (vs. higher-proficiency participants).

*Table 43: Final model for the congruency effect on reaction times in the long display condition for word items (1~L2proficiency + TargetLanguage + FlankerLanguage + L2proficiency*TargetLanguage + L2proficiency*FlankerLanguage + TargetLanguage*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.2^{e-03}	1.6^{e-04}	7.23
L2 Proficiency	3.7^{e-06}	2.1^{e-06}	1.73
Target Language	-2.0^{e-04}	4.7^{e-05}	-4.26
Flanker Language	-7.7^{e-05}	4.5^{e-05}	-1.70
L2 proficiency*Target Language	1.4^{e-06}	5.9^{e-07}	2.42
L2 proficiency*Flanker Language	4.6^{e-07}	5.9^{e-07}	0.77
Target Language*Flanker Language	4.9^{e-05}	1.2^{e-05}	4.17
Random Effects	Variance	Standard Deviation	
Intercept Items	8.22^{e-09}	9.07^{e-05}	
Intercept Participants	2.27^{e-08}	1.51^{e-04}	
Residuals	6.51^{e-08}	2.55^{e-04}	

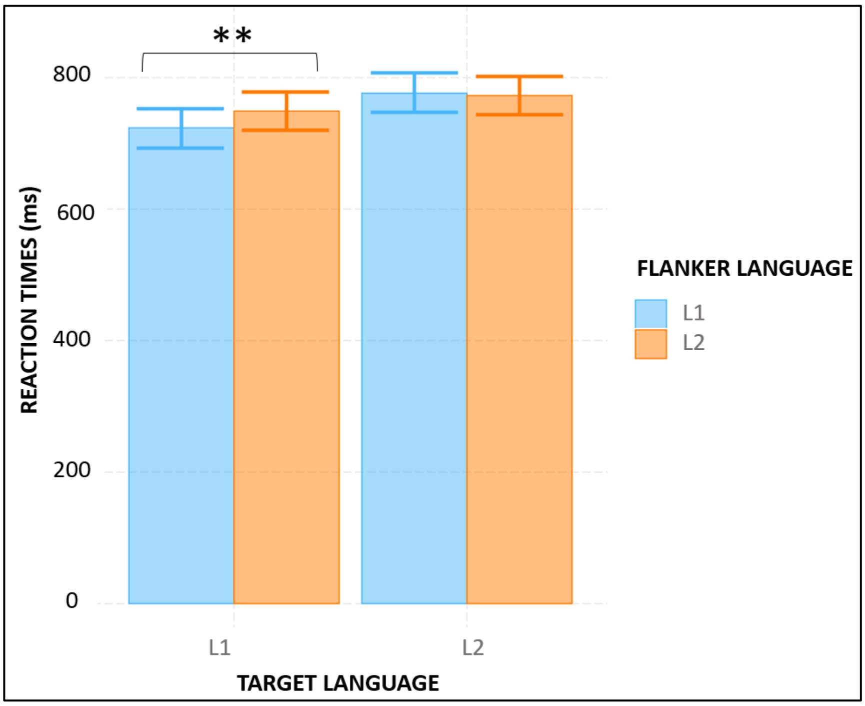


Figure 34 : Interaction effect between Flanker language and Target language on reaction times to WORD items in the long display condition from Experiment 2.

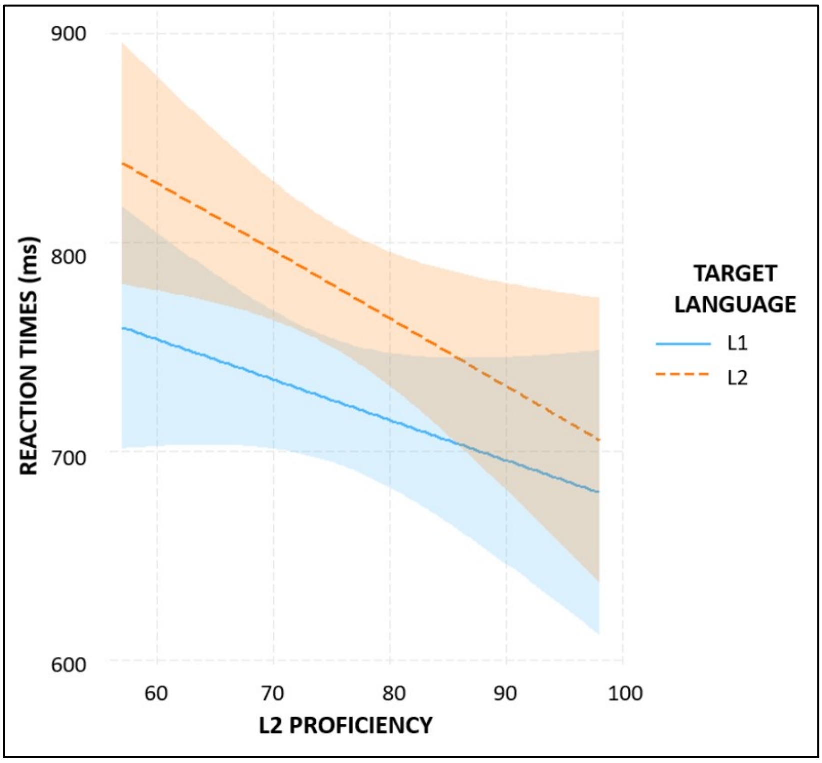


Figure 35: Interaction between target language and L2 proficiency (using LEXTALE scores) on reaction times to word items in the long display condition from Experiment 2.

On Accuracy. Model 4, containing the interaction effect between target and flanker languages best fitted the accuracy data (AIC=3793; $\chi^2(1,6)=12.80$; $p<.001$). A closer look at the model (see table 44) showed a main effect of flanker language ($b_3=-0.47$; SE=0.12; z=-3.74) with L1 flankers leading to better performances than L2 flankers (L1 Flanker= 94% vs. L2 Flanker= 93% accuracy). More importantly the interaction effect between target language and flanker language was significant ($b_4=0.66$; SE=0.18; z=3.71; see figure 36), accuracy was higher for congruent trials than for incongruent trials (congruent trials= 94% vs. incongruent trials= 93% accuracy). The post hoc comparisons revealed that the congruency effect was significant for L1 target (L1 congruent=94% vs. L1 incongruent=92% accuracy; $p<.001$) but not for L2 target (L2 congruent=94% vs. L2 incongruent=93%; $p=.419$).

*Table 44: Final model for the congruency effect on accuracy in the long display condition for word items (1~ TargetLanguage + FlankerLanguage + TargetLanguage*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	z
Intercept	3.44	0.18	18.40
Target Language	-0.29	0.22	-1.30
Flanker Language	-0.47	0.12	-3.74
Target Language*Flanker Language	0.66	0.18	3.71
Random Effects	Variance	Standard Deviation	
Intercept Items	1.12	1.06	
Intercept Participants	0.45	0.67	

Going Further – on Pseudowords.

No model improved fit for the latency data (all *ps*>.010).

Model 2, containing the target language effect improved fit for the accuracy data (AIC=3495; $\chi^2(1,4)=9.86$; $p=.002$). A closer look at the model showed a main effect of target language ($b_2=0.55$; SE=0.17; z=3.23) with L1-like pseudowords leading to better performances than L2-like pseudowords (L1-like pseudowords= 95% vs. L2-like pseudowords= 93% accuracy).

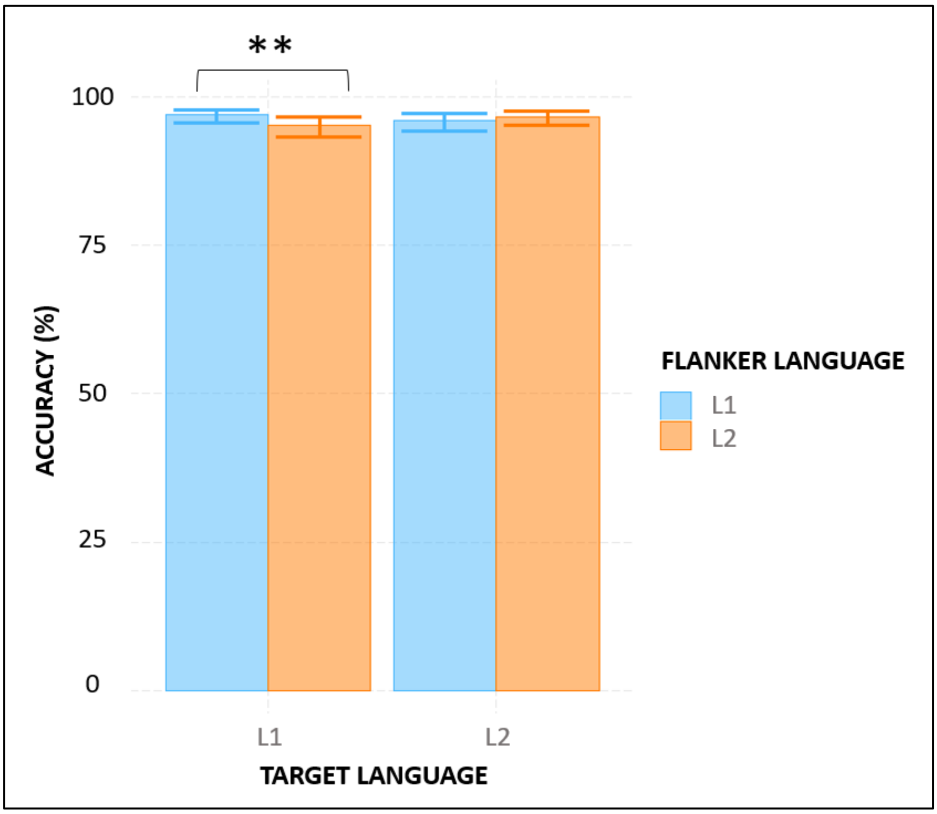


Figure 36: Interaction effect between Flanker language and Target language on accuracy to WORD items in the long display condition from Experiment 2.

Does target markedness impact word recognition in a linguistic context?

Model comparison

We used 2 stepwise model comparison in which we compared 8 models on both L1 and L2 subplots. Model 1 contained the random effects. Model 2 contained the effect of L2 proficiency. Model 3 contained the effect trial congruency (Congruent or Incongruent). Model 4 contained the effect of target markedness (Marked or Non-Marked), model 5 to 7 contained the different interactions between those factors, and model 8 contained the full interaction. For accuracy data L2-proficiency prevented the models from converging and was therefore dropped, thus the final comparison contained only 4 models (detail about model comparisons are available in Appendix 19.A).

On responses to L1-word targets – Reaction Times. In the L1 target subplot, model 3 containing the main effect of trial congruency best fitted the latency data (AIC=-52513; $\chi^2(1,6)=26.91$; p=.001). A closer look to this model confirmed the main effect of trial congruency ($b_3=-4.3^{e-05}$; SE=8.3^{e-06}; t=-5.20), with congruent trials being responded to faster than incongruent trials (Congruent Trials=722ms vs. Incongruent Trials=744ms).

On responses to L1-word targets – Accuracy. Model 2 containing the main effect of trial congruency effect best fitted the accuracy data (AIC=2006; $\chi^2(1,4)=13.78$; p<.001). A closer look to this model confirmed the main effect of trial congruency ($b_2=-0.47$; SE=1.12; z=-3.83), with congruent trials leading to better accuracy than incongruent trials (Congruent Trials=94% vs. Incongruent Trials=92% accuracy).

On responses to L2-word targets – Reaction Times. In the L2 target subplot, model 2, which contained the effect of L2 proficiency, best fitted the latency data (AIC=-51707; $\chi^2(1,5)=6.43$; p=.011). A closer look to this model confirmed the effect of L2 proficiency ($b_2=5.4^{e-06}$; SE=2.1^{e-06}; t=2.57), with higher-proficiency participants responding faster than lower-proficiency participants (Lower-proficiency participants= 799ms vs. Higher-proficiency participants= 753ms).

On responses to L2-word targets – Accuracy. No model improved fit for the accuracy data (all *ps*>.05).

Going Further – On Pseudowords

L1-like pseudoword targets. In the L1-like target subplot, no model improved fit for either the latency data or the accuracy data (all *ps*>.05).

L2- like pseudoword targets. No model improved fit for the latency data (all *ps*>.10).

Model 3 containing the main effect of target markedness best fitted the accuracy data (AIC=2406; $\chi^2(1,5)=4.89$; $p=.027$). A closer look at this model confirmed the main effect of target markedness ($b_3=0.53$; SE=0.23; z=2.27), with L2-Marked pseudowords being rejected less accurately than L2-like but Non-Marked pseudowords (L2-Marked= 91% vs. L2-Non-Marked= 94% accuracy).

Details about model comparisons are available in Appendix 19.B

Does flanker markedness impact a target word's recognition?

Model comparison

We used 4 stepwise model comparison in which we compared 4 models on both Reaction Times and Accuracy. Model 1 contained the random effects. Model 2 contained the effect of trial congruency (congruent or incongruent). Model 3 contained the effect of Flanker Language (L1 or L2), and Model 4 contained the interaction between the two factors (details about model comparisons are available in Appendix 20.A).

When flankers were MARKED – Reaction Times. In the Marked flankers subplot, model 4 containing the interaction effect best fitted the latency data (AIC=-52208; $\chi^2(1,7)=18.98$; $p<.001$). A closer look at this model (see table 45) revealed a main effect of trial congruency ($b_2=-9.52^{e-05}$; SE=1.8^{e-05}; t=-5.24), with congruent trials leading to faster reactions times than incongruent trials (congruent trials=751ms vs. Incongruent trials=764ms), and a main effect of flanker language ($b_3=-9.49^{e-05}$; SE=1.8^{e-05}; t=-5.23), with L2 flankers leading to longer reaction times than L1 flankers (L2 flankers=764ms vs. L1 flankers=752ms). The interaction between trial congruency and Flanker language was significant ($b_4=-1.5^{e-04}$;

SE=3.2^{e-05}; t=4.49; see figure 37). Every relevant post-hoc comparison was significant (all ps <.05), but while the congruency effect was facilitatory for L1 flankers (congruent trials=722ms vs. incongruent trials=781ms; *p*<.001) it was rather inhibitory for L2 flankers (congruent trials=780ms vs. incongruent trials=747ms; *p*=.033).

*Table 45: Final model for the trial congruency effect on reaction times in the long display condition for Marked flankers (1~TrialCongruency + FlankerLanguage + TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.46^{e-03}	2.49^{e-05}	58.45
Trial Congruency	-9.52^{e-05}	1.82^{e-05}	-5.24
Flanker Language	-9.49^{e-05}	1.82^{e-05}	-5.23
Trial Congruency*Flanker Language	1.45^{e-04}	3.23^{e-05}	4.49
Random Effects	**Variance**	**Standard Deviation**	
Intercept Items	7.41^{e-09}	861^{e-05}	
Intercept Participants	2.42^{e-08}	1.55^{e-04}	
Intercept Participants	6.59^{e-08}	2.57^{e-04}	

When flankers were MARKED – Accuracy. Model 2 best fitted the accuracy data (AIC=1865; χ^2(1,4)=8.07; *p*=.004). A closer look at this model revealed a main effect of trial congruency (b_2=-0.38; SE=0.14; z=-2.82), with congruent trials leading to better accuracy than incongruent trials (Congruent trials=95% vs. Incongruent trials=93% accuracy).

When flankers were NON-MARKED – Reaction Times. In the Non-Marked flankers subplot, model 4 containing the interaction effect between trial congruency and flanker language best fitted the latency data (AIC=-51838; χ^2(1,7)=14.23; *p*<.001). A closer look at this model (see table 46) revealed a main effect of trial congruency (b_2=-9.23^{e-05}; SE=1.9^{e-05}; t=-4.79), with congruent trials leading to faster responses than incongruent trials (congruent trials=743ms vs. incongruent trials=757ms), and a main effect of flanker language (b_3=-7.74^{e-05}; SE=1.9^{e-05}; t=-4.02), with L2 flankers leading to slower reaction times than L1 flankers (L2 flankers=753ms vs. L1 flankers=747ms). The interaction between trial congruency and Flanker language was significant (b_4=1.34^{e-04}; SE=3.5^{e-05}; t=3.86, see figure 38) with the congruency effect being significant only when flankers came from L1 (Incongruent trials=773ms vs. Congruent trials=721ms; *p*<.001) and not when flankers came from L2 (Incongruent trials=741ms vs. Congruent trials=765ms; *p*=.131).

When flankers were NON-MARKED – Accuracy. Model 2 best fitted the accuracy data (AIC=2032; χ^2(1,4)=5.51; *p*=.019). A closer look at this model revealed a main effect of trial congruency (b_2=-0.30; SE=0.13; z=-2.34), with congruent trials leading to better accuracy than incongruent trials (Congruent trials=94% vs. Incongruent trials=92% accuracy).

*Table 46: Final model for the trial congruency effect on reaction times in the long display condition for Non-Marked flankers (1~TrialCongruency + FlankerLanguage + TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.46^{e-03}	2.56^{e-05}	57.13
Trial Congruency	-9.23^{e-05}	1.93^{e-05}	-4.79
Flanker Language	-7.74^{e-05}	1.93^{e-05}	-4.02
Trial Congruency*Flanker Language	1.34^{e-04}	3.48^{e-05}	3.86
Random Effects	Variance	Standard Deviation	
Intercept Items	9.15^{e-09}	9.57^{e-05}	
Intercept Participants	2.48^{e-08}	1.58^{e-04}	
Intercept Participants	6.38^{e-08}	2.53^{e-04}	

Going Further – On Pseudowords

When flankers were MARKED. No model improved fit for the latency data (all *ps*>.10).

Model 4 best fitted the accuracy data (AIC=1774; χ^2(1,6)=4.01; *p*=.045). A closer look at this model revealed an interaction effect between trial congruency and flanker language (b_4=-0.85; SE=0.43; z=1.99). However, no post-hoc comparisons were significant (all *ps*>.10).

When flankers were NON-MARKED. No model improved fit for the latency data (all *ps*>.05).

Model 4 best fitted the accuracy data (AIC=1790; χ^2(1,6)=13.12; *p*<.001). A closer look at this model revealed a main effect of trial congruency (b_2=-0.95; SE=0.26; z=-3.66) with congruent trials leading to better accuracy (Congruent trials=95% vs. Incongruent trials=94% accuracy). The main effect of flanker language was also significant (b_3=-1.08; SE=0.26; z=-4.16) with L1-like flankers leading to better accuracy than L2-like flankers (L1 flankers=96% vs. L2 flankers=94% accuracy). The interaction effect between trial congruency and flanker language was also significant (b_4=1.47; SE=0.40; z=3.69). The post-hoc comparisons revealed that while the trial congruency effect was significant for L1-like flankers (Congruent trials=97% vs. Incongruent trials=95% accuracy; *p*=.002) it was not significant for L2-like flankers (Congruent trials=92% vs. Incongruent trials=92% accuracy; *p*=.116). The difference between L1-like flankers and L2-like flankers was also significant in congruent trials (L1-like flankers=97% vs. L2-like flankers=92%; *p*<.001) while it was not significant in incongruent trials (L1-like flankers=95% vs. L2-like flankers=92%; *p*=.331).

Details about model comparisons are available in Appendix 20.B.

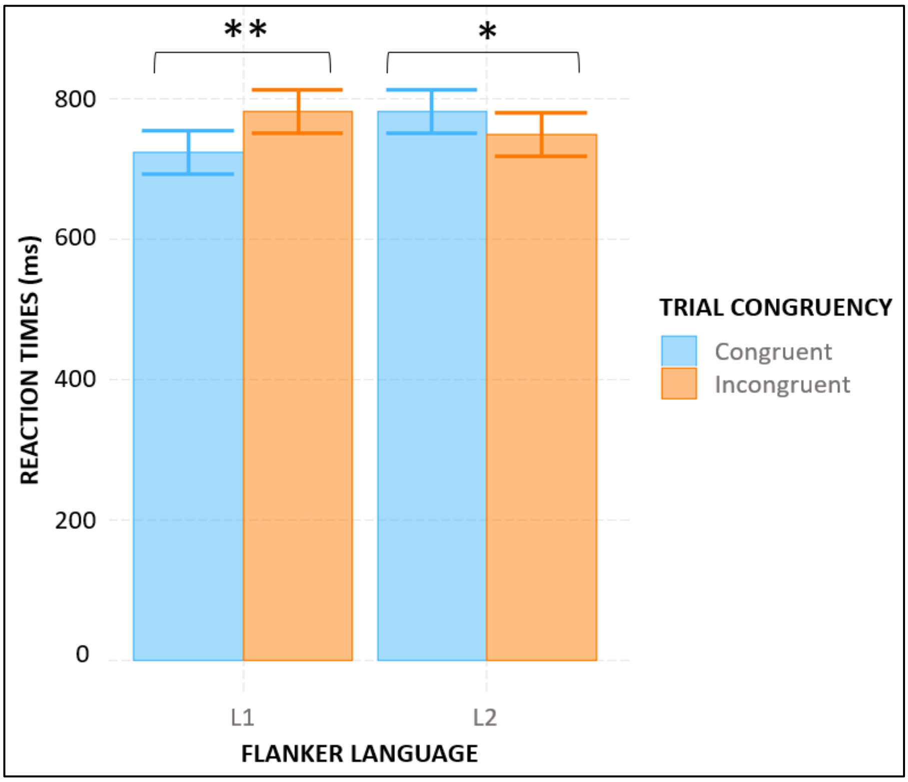


Figure 37: Interaction effect between trial congruency and flanker language on reaction times to WORD responses when flankers were MARKED in the long display condition of Experiment 2

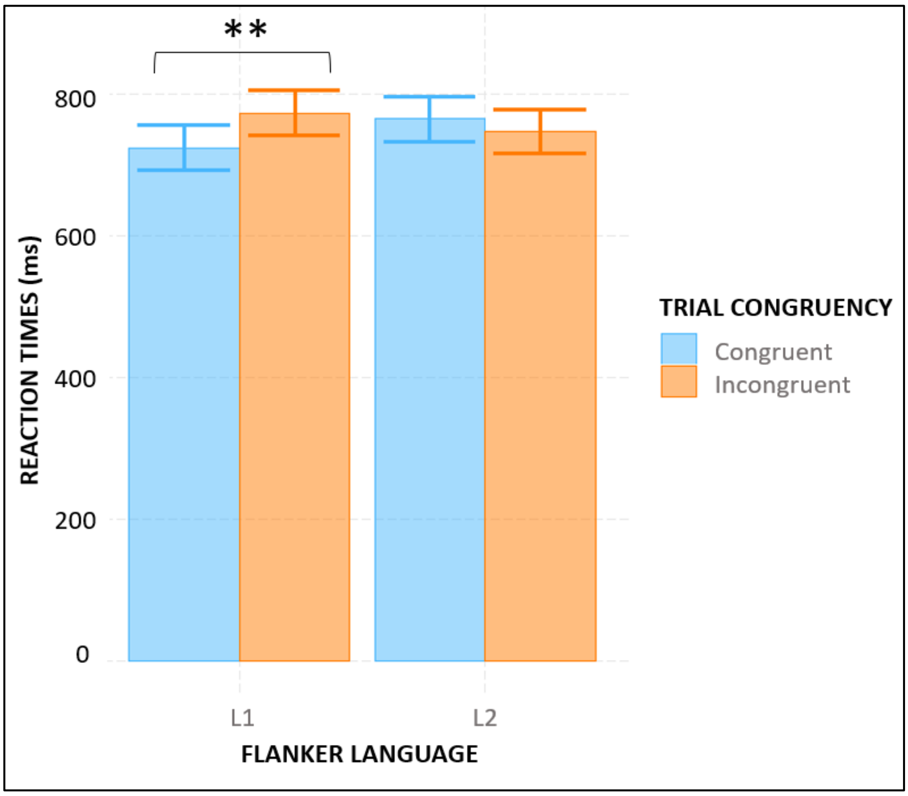


Figure 38: Interaction effect between trial congruency and flanker language on reaction times to WORD responses when flankers were NON-MARKED in the long display condition of Experiment 2

Long Display Condition – Intermediate Summary

Do we replicate the Trial Congruency Effect?

We found a Trial Congruency effect that was significant *only* for L1-Targets (L1-**L1**-L1 vs. L2-**L1**-L2) but not for L2-Targets (L2-**L2**-L2 vs. L1-**L2**-L1). When all the lexical representations involved are (supposedly) strong enough, the participants benefit from trial congruency. If not, language membership information coming from flanker items was not used to perform a lexical decision on a target item.

Does Target Markedness impact language decision in a linguistic context?

We did not find a Target Markedness effect *either* for L2-Targets (Flanker-**KNOW**-Flanker vs. Flanker-**HOUSE**-Flanker) or for L1-Targets (Flanker-**LIVRE**-Flanker vs. Flanker-**JUPE**-Flanker). It seems that even if markers are detected, they are not used to make lexical decisions on a target item.

Does Flanker Markedness impacts a target word's recognition?

The Trial Congruency effect was significant when the flankers were L1 words, were they marked (**NEUF**-L1-**NEUF** vs. **NEUF**-L2-**NEUF**) or not (**JOUR**-L1-**JOUR** vs. **JOUR**-L2-**JOUR**) and for L2-Marked-Flankers (**KNIFE**-L2-**KNIFE** vs. **KNIFE**-L1-**KNIFE**) *but not* for L2-Non-Marked-Flankers (**GAME**-L2-**GAME** vs. **GAME**-L1-**GAME**). This pattern suggests that only the lexical representations tied to L1-flankers and to L2-Marked-Flankers were strong enough (or activated enough) to impact a lexical decision on a target item.

Results for the Short display condition

Table 47: Mean Lexical Decision Latencies (in ms) and Accuracy (in %) to Marked and Non-Marked words presented in the short display condition from Experiment 2

Reaction times (in ms)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total target			total target
M	631(162)	637(175)	634(168)	647(165)	644(200)	646(171)
NM	631(162)	639(178)	635(170)	650(151)	641(151)	645(151)
total flanker	631(162)	638(176)	634(169)	649(159)	643(164)	646(161)
Target L2			total target			total target
M	666(173)	658(167)	662(170)	664(165)	675(165)	669(165)
NM	669(155)	666(163)	668(159)	659(157)	678(178)	668(168)
total flanker	667(164)	662(165)	665(164)	661(161)	676(171)	669(166)
Accuracy (in%)						
	Flanker L1			Flanker L2		
	M	NM		M	NM	
Target L1			total target			total target
M	93(0.25)	90(0.30)	92(0.28)	91(0.29)	93(0.24)	92(0.26)
NM	90(0.30)	90(0.30)	90(0.30)	87(0.34)	92(0.28)	89(0.31)
total flanker	92(0.28)	90(0.30)	91(0.29)	89(0.31)	93(0.26)	91(0.29)
Target L2			total target			total target
M	91(0.29)	92(0.28)	91(0.38)	92(0.28)	94(0.25)	93(0.26)
NM	86(0.34)	90(0.34)	88(0.32)	91(0.28)	91(0.28)	91(0.28)
total flanker	89(0.32)	91(0.29)	90(0.30)	91(0.29)	92(0.26)	92(0.27)

Data trimming and analysis

Erroneous responses, timed out responses as well as responses above or below three times the standard deviation both by participants and by target item's type (L1-Marked, L1-Non-Marked, L2-Marked and L2-Non-Marked) were removed from the reaction time analyses. Responses under 300ms and timed-out responses were removed from the accuracy analyses. Finally, the words BROOM and ROPE, and the pseudoword CREAR that were responded with only 43 to 52% accuracy, which was below three standard deviation from the mean, had to be removed, resulting in the exclusion of 1.79% of the data in the word plot and 1.54% from the pseudoword plot.

Reaction times (inverse transformed) and accuracy data were examined using linear and binomial mixed models respectively (by using the lme4 package on R, Bates et al., 2015). The models included by-participant and by-item random intercepts (i.e., model = Y ~ X (1|Participant) + (1|Target)).

Do we replicate the trial congruency effect?

Model comparison

First, we used a stepwise model comparison approach in which we compared 8 models, in order to replicate the congruency effect from Eben & Declerck, (2019). Model 1 contained the random effects. Model 2 contained the effect of L2 proficiency (LEXTALE scores). Model 3 contained the effect of target language (L1 or L2). Model 4 contained the effect of flanker language (L1 or L2), model 5 to 7 contained the interaction effect between each variable combination and model 8 the full interaction. Since L2-proficiency prevented the models to converge on accuracy data, it was dropped from those models, resulting in a 4 model comparison for accuracy data (details about model comparisons are available in Appendix 21).

On Reaction times. Model 4, containing both the effect of target language and the effect of flanker language best fitted the latency data (AIC=-82899; $\chi^2(1,7)=9.48$; *p*=.002). A closer look at this model (see table 48) revealed a main effect of target language ($b_3=6.2^{e-05}$; $SE=1.5^{e-05}$; t=-4.25), with L1 targets being responded to faster than L2 targets (L1 targets= 640ms vs. L2 targets=667ms), and a main effect of flanker language (b_4=-2.2^{e-05}; SE= 7.3^{e-05}; t= -3.08), with L1 flankers leading to faster responses than L2 flankers (L1 flankers=650ms vs. L2 flankers=657ms).

On Accuracy. No models improved fit for the accuracy data (all *ps*>.05).

Table 48: Final model for the congruency effect on reaction times in the short display condition for word items (1~L2proficiency + TargetLanguage + FlankerLanguage)

Fixed Effects	Estimate	Standard Error	t
Intercept	1.5^{e-03}	2.2^{e-04}	6.75
L2 Proficiency	2.5^{e-06}	2.9^{e-06}	0.85
Target Language	-6.2^{e-05}	1.5^{e-05}	-4.25
Flanker Language	-2.2^{e-05}	7.3^{e-05}	-3.08
Random Effects	Variance	Standard Deviation	
Intercept Items	6.41^{e-09}	8.01^{e-05}	
Intercept Participants	4.23^{e-08}	2.06^{e-04}	
Residuals	7.75^{e-08}	2.78^{e-04}	

Going Further – On Pseudoword

Model 5, containing the interaction between L2 proficiency and Target language best fitted the data (AIC=-80022; χ^2(1,8)=8.59; *p* =.003). A closer look at this model revealed a main effect of target language (b_3=1.2^{e-04}; SE=4.7^{e-05}; z=2.60) with L1-like pseudowords being rejected faster than L2-like pseudowords (L1-like pseudowords=772ms vs. L2-like pseudowords=781ms). Its interaction with L2-proficiency was also significant (b_5=-1.8^{e-06}; SE=6.1^{e-02}; z=-2.93), the difference between L1 and L2-like pseudowords being more important for higher-proficiency participants than for lower-proficiency participants.

No model significantly improved fit for accuracy data (all *ps*>.05).

Does target markedness impact word recognition in a linguistic context?

Model comparison

We used 2 stepwise model comparison in which we compared 8 models on both L1 and L2 subplots. Model 1 contained the random effects. Model 2 contained the effect of L2 proficiency. Model 3 contained the effect trial congruency (Congruent or Incongruent). Model 4 contained the effect of target markedness (Marked or Non-Marked), model 5 to 7 contained the different interaction between those factors, and model 8 contained the full interaction. For accuracy data L2-proficiency prevented the models from converging and was therefore dropped, thus the final comparison contained only 4 models (details about model comparisons are available in Appendix 22A).

On responses to L1-word targets – Reaction Times. In the L1 target subplot, model 3 containing the main effect of trial congruency best fitted the latency data (AIC=-41753; $\chi^2(1,6)=11.45$; $p<.001$). A closer look to this model confirmed the main effect of trial congruency ($b_3=-3.5^{e-05}$; $SE=1.0^{e-05}$; $t=-3.39$), with congruent trials being responded to faster than incongruent trials (Congruent Trials=634ms vs. Incongruent Trials=665ms).

On responses to L1-word targets – Accuracy. No models improved fit for the accuracy data (all *ps*>.05).

On responses to L2-word targets. No model improved fit for the latency data (all *ps*>.05) nor for the accuracy data (all *ps*>.05).

Going further – On pseudowords

L1-like pseudoword targets. In the L1-like target subplot, no model improved fit for either the latency data or the accuracy data (all *ps*>.05).

L2-like pseudoword targets. Model 4, containing the effect of target markedness best fitted the latency data (AIC=-38967; $\chi^2(1,7)=7.78$; p=.005). A closer look at this model confirmed the target markedness effect ($b_4=4.7^{e-05}$; SE=1.7^{e-05}; t=-2.84), with L2-Marked pseudowords being rejected slower than L2-like but Non-Marked pseudowords (L2-Marked=795ms vs. L2-not Marked=766ms).

Model 3, containing the effect of target markedness best fitted the accuracy data (AIC= 2730; $\chi^2(1,5)=7.34$; p=.007). A closer look at this model confirmed the target markedness effect (b_3=0.48; SE=0.17; z=2.79), with L2-Marked pseudowords leading to worse accuracy than L2-like but Non-Marked pseudowords (L2-Marked= 86% vs. L2-Non-Marked=81% accuracy).

Details about model comparisons are available in Appendix 22B.

Does flanker markedness impact target word recognition?

Model comparison

We used 4 stepwise model comparison in which we compared 4 models on both Reaction Times and Accuracy. Model 1 contained the random effects. Model 2 contained the effect of trial congruency (congruent or incongruent). Model 3 contained the effect of Flanker Language (L1 or L2), and Model 4 contained the interaction between the two factors (details about model comparisons are available in Appendix 23A).

When flankers were MARKED – Reaction Times. In the Marked flankers subplot, model 4 containing the interaction effect best fitted the latency data (AIC=-41182; $\chi^2(1,7)=11.44$; p<.001). A closer look at this model (see table 49) revealed a main effect of trial congruency

(b_2=-8.61^{e-05}; SE=1.8^{e-05}; t=-4.68), with congruent trials leading to faster reactions times than incongruent trials (congruent trials=646ms vs. Incongruent trials=658ms), and a main effect of flanker language (b_3=-7.19^{e-05}; SE=1.8^{e-05}; t=-3.92), with L2 flankers leading to longer reaction times than L1 flankers (L2 flankers=655ms vs. L1 flankers=649ms). The interaction between trial congruency and Flanker language was significant (b_4=-1.1^{e-04}; SE=3.1^{e-05}; t=3.45; figure 39). Post-hoc comparisons revealed a significant facilitatory effect of congruency for L1 flankers (congruent trials=631ms vs. incongruent trials=667ms; *p*<.001) that was not significant, and numerically reversed, for L2 flankers (congruent trials=661ms vs. incongruent trials=649ms; *p*=.719). Moreover, flanker language impacted congruent trials (L1 flankers=631ms vs. L2 flankers=661ms; *p*<.001) but not incongruent trials (L1 flankers=667ms vs. L2 flankers=649ms; *p*=.267).

*Table 49: Final model for the trial congruency effect on reaction times in the short display condition for Marked flankers (1~TrialCongruency + FlankerLanguage + TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.66^{e-03}	3.36^{e-05}	49.51
Trial Congruency	-8.61^{e-05}	1.84^{e-05}	-4.68
Flanker Language	-7.19^{e-05}	1.84^{e-05}	-3.92
Trial Congruency*Flanker Language	1.05^{e-04}	3.06^{e-05}	3.45
Random Effects	Variance	Standard Deviation	
Intercept Items	5.28^{e-09}	7.27^{e-05}	
Intercpet Participants	4.24^{e-08}	2.06^{e-04}	
Intercept Participants	7.47^{e-08}	2.73^{e-04}	

When flankers were MARKED – Accuracy. Model 2 best fitted the accuracy data (AIC=2032; χ^2(1,5)=10.64; *p*<.001). A closer look at this model revealed a main effect of trial congruency (b_2=-0.43; SE=0.14; z=-3.23), with congruent trials leading to better accuracy than incongruent trials (Congruent trials=92% vs. Incongruent trials=89% accuracy).

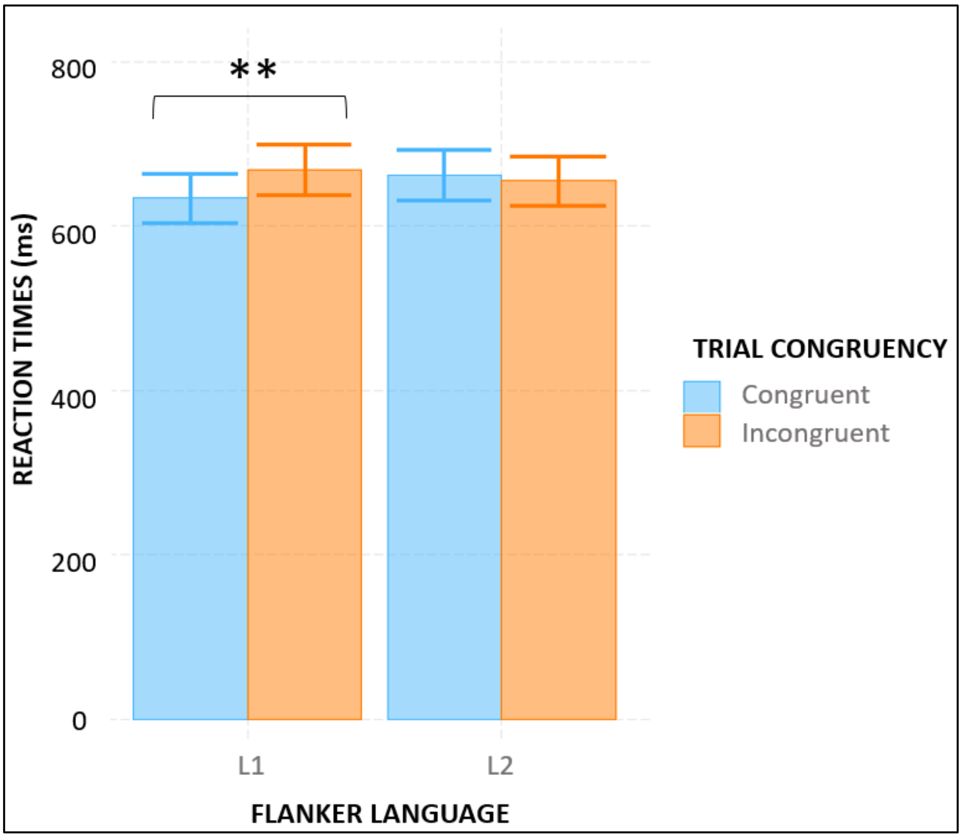


Figure 39: Interaction effect between trial congruency and flanker language on reaction times to word responses when flankers were marked in the short display condition of Experiment 2

When flankers were NON-MARKED – Reaction Times. In the Non-Marked flankers subplot, model 4 containing the interaction effect between trial congruency and flanker language best fitted the latency data (AIC=-41510; $\chi^2(1,7)=16.60$; *p*<.001). A closer look at this model (see table 50) revealed a main effect of trial congruency ($b_2=-6.03^{e-05}$; $SE=2^{e-05}$; t=-3.01), with congruent trials leading to faster responses than incongruent trials (congruent trials=657ms vs. incongruent trials=653ms), and a main effect of flanker language ($b_3=-9.40^{e-05}$; $SE=2^{e-05}$; t=-4.70), with L2 flankers leading to slower reaction times than L1 flankers (L2 flankers=660ms vs. L1 flankers=650ms). The interaction between trial congruency and Flanker language was significant ($b_4=1.41^{e-04}$; $SE=3.4^{e-05}$; t=4.18; see figure 40). Every relevant post-hoc comparison was significant (all ps <.05), but while the trial congruency effect for facilitatory for L1 flankers (congruent trials=638ms vs. incongruent trials=662ms; *p*=.014) that was reversed, for L2 flankers (congruent trials=676ms vs. incongruent trials=643ms; *p*<.001). Moreover, flanker language impacted congruent trials (L1 flankers=631ms vs. L2 flankers=661ms; *p*<.001) but not incongruent trials (L1 flankers=667ms vs. L2 flankers=649ms; *p*=.267).

When flankers were NON-MARKED – Accuracy. No model improved fit for accuracy data (all *ps*>.05).

*Table 50: Final model for the trial congruency effect on reaction times in the short display condition for Non-Marked flankers (1~TrialCongruency + FlankerLanguage + TrialCongruency*FlankerLanguage)*

Fixed Effects	Estimate	Standard Error	t
Intercept	1.65^{e-03}	3.38^{e-05}	48.56
Trial Congruency	-6.03^{e-05}	2.00^{e-05}	-3.01
Flanker Language	-9.40^{e-05}	2.00^{e-05}	-4.70
Trial Congruency*Flanker Language	1.41^{e-04}	3.38^{e-05}	4.18
Random Effects	Variance	Standard Deviation	
Intercept Items	7.14^{e-09}	8.45^{e-05}	
Intercpet Participants	4.16^{e-08}	2.04^{e-04}	
Intercept Participants	8.03^{e-08}	2.83^{e-04}	

Going Further – On Pseudowords

When flankers were MARKED. Model 3, containing the flanker language effect best fitted the latency data (AIC=-3845; $\chi^2(1,6)=6.05$; *p*=.014). A closer look at this model revealed a main effect of flanker language was significant ($b_3=-2.35^{e-05}$; $SE=9.56^{e-06}$; t=2.46) with L1-like flankers leading to longer reaction times than L2-like flankers (L1 flankers=783ms vs. L2 flankers=771ms). No model improved fit for the accuracy data (all *ps*>.10).

When flankers were NON-MARKED. No model improved fit for the latency data (all *ps*>.10) nor for the accuracy data (all *ps*>.05).

Details about model comparisons are available in Appendix 23B.

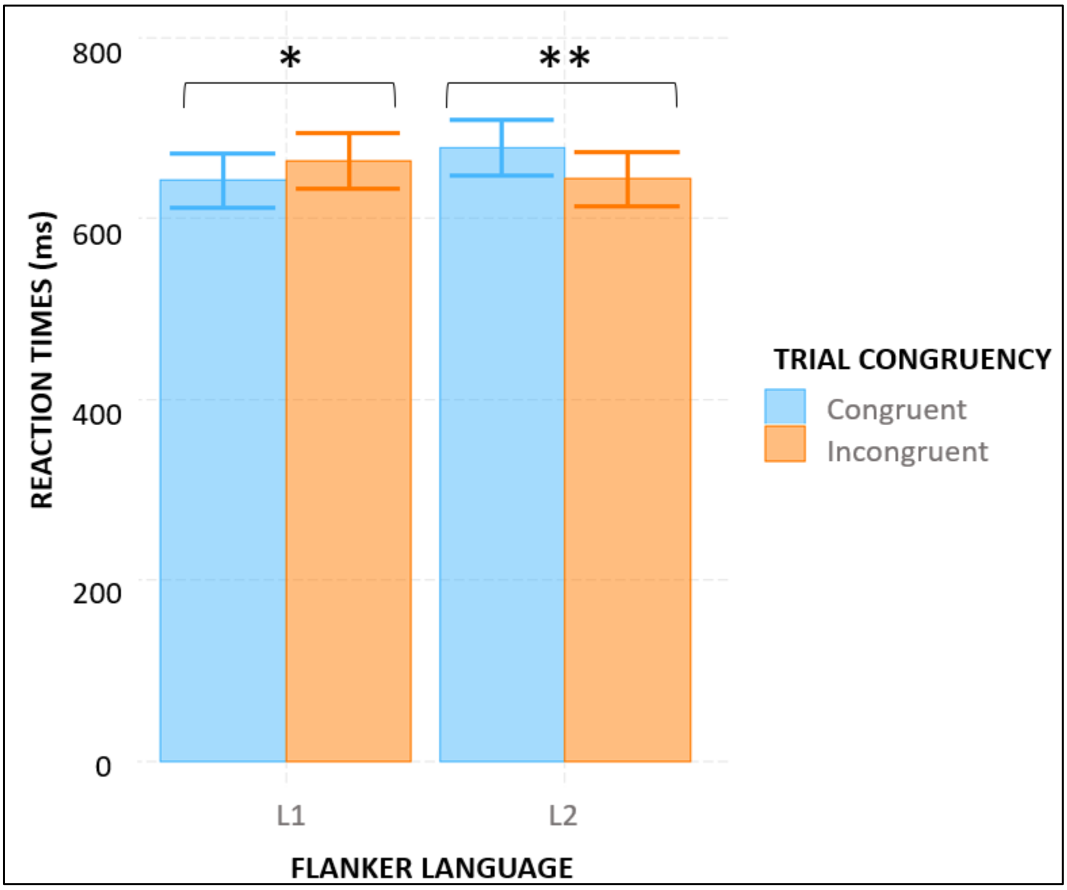


Figure 40: Interaction effect between trial congruency and flanker language on reaction times to word responses when flankers were non-marked in the short display condition of Experiment 2

Short Display Condition – Intermediate Summary

Do we replicate the Trial Congruency Effect?

We did not find a Trial Congruency effect *either* for L1-Targets (L1-**L1**-L1 vs. L2-**L1**-L2) or for L2-Targets (L2-**L2**-L2 vs. L1-**L2**-L1). It seems that even though language membership information coming from flanker items was retrieved by our participants (see experiment 1), this information were not used to perform a lexical decision on a target item.

Does Target Markedness impact language decision in a linguistic context?

We did not find a Target Markedness effect *either* for L2-Targets (Flanker-**KNOW**-Flanker vs. Flanker-**HOUSE**-Flanker) or for L1-Targets (Flanker-**LIVRE**-Flanker vs. Flanker-**JUPE**-Flanker). It seems that even if markers are detected, they are not used to make lexical decisions on a target item. However, L2-markers did slow down pseudoword rejection (Flanker-**LUVRE**-Flanker vs. Flanker-**VUPE**-Flanker), suggesting that L2-markers could be used for lexical decisions when the conditions allowed only for a superficial sublexical processing.

Does Flanker Markedness impacts a target word's recognition?

The Trial Congruency effect was significant when the flankers were L1 words, were they marked (**NEUF**-L1-**NEUF** vs. **NEUF**-L2-**NEUF**) or not (**JOUR**-L1-**JOUR** vs. **JOUR**-L2-**JOUR**) and for L2-Non-Marked-Flankers (**GAME**-L2-**GAME** vs. **GAME**-L1-**GAME**) *but not* for L2-Marked-Flankers (**KNIFE**-L2-**KNIFE** vs. **KNIFE**-L1-**KNIFE**). This pattern suggests that only the lexical representations tied to L1-flankers and to L2-Non-Marked-Flankers were strong enough (or activated enough) to impact a lexical decision on a target item.

Discussion for Experiment 2

The target markedness effect was more elusive in this second experiment compared to the first experiment we described in this chapter (see figure 41). For instance, when we look at the word responses, neither L1 nor L2 markers impacted either latency or accuracy no matter the display condition. Those results diverge from those coming from another generalized lexical decision task on isolated words by Thomas and Allport (2000), suggesting that our lack of an orthographic markedness effect might come from the presence of flanker items in our experiment which might have watered down the target markers' saliency. Furthermore, in line with the literature (Commissaire, Audusseau, et al., 2019; Lemhöfer & Radach, 2008) a negative orthographic markedness effect arose in the pseudoword responses. Interestingly, this effect was restricted to L2-like pseudowords, confirming the L2-readers' greater sensitivity toward L2-markers. Moreover, L2-markers delayed the rejection of target pseudowords only in the short display condition, suggesting that participants relied on markers for their decisions on the lexical status of the target item only when a more thorough processing was prevented from taking place.

At first glance, the overall lack of a trial congruency effect in this lexical decision task might suggest that participants did not retrieve the language membership information contained in flanker items, but this would be surprising since similar participants were able to do so under similar circumstances in the language decision task described above. It is rather more likely that this language information coming from the flankers (which might conflict with the language information coming from the target) did not overly impact decisions on the lexical status of the target item. Interestingly, this is also in line with the finding that the trial congruency in the longer display condition was restricted to L1 word targets which benefited from the presence of L1 word flankers. This seems to indicate that it was the only condition in which every items' lexical representation was strong enough to be reliably used by participants when making a decision on the lexical status of the target item. Finally it is important to consider that our results to the short display condition we shared with Declerck et al., (2018), differed from their own results since they found a clear effect of trial congruency. However, we might attribute this discrepancy to our use of a general lexical decision task rather than a language-blocked lexical decision task, which might have encouraged a bilingual mode in our participants in which both languages were expected to be encountered whereas Declerck et al., participants might have expected only language-specific words to be encountered, thus maximizing their chances to find a trial congruency effect (see language mode hypothesis; Grosjean, 2001).

		LONG DISPLAY	**SHORT DISPLAY**
L1 TARGET MARKEDNESS	RT	No markedness effect	No markedness effect
	ACC	No markedness effect	No markedness effect
L2 TARGET MARKEDNESS	RT	No markedness effect	No markedness effect
	ACC	No markedness effect	No markedness effect
TRIAL CONGRUENCY	RT	Trial congruency effect (t=4.17) L1 targets (*p*<.001) vs L2 targets (*p*=873)	No Trial Congruency effect
	ACC	Trial congruency effect (z=3.71) L1 targets (*p*<.001) vs L2 targets (*p*=.419)	No Trial Congruency effect
TRIAL CONGRUENCY **Marked Flankers**	RT	Interaction with Flanker language (t=-4.49)	Interaction with Flanker language (t=-3.86) L1 targets (*p*<.001) vs L2 targets (*p*=131)
	ACC	Trial congruency effect (z=-2.82)	Trial congruency effect (z=-2.34)
TRIAL CONGRUENCY **Non-Marked Flankers**	RT	Interaction with Flanker language (t=3.45) L1 targets (*p*<.001) vs L2 targets (*p*=719)	Interaction with Flanker language (t=4.18)
	ACC	Trial congruency effect (z=-3.23)	No Trial congruency effect

Figure 41: Summary of the results from experiment 2

To complete the discussion on the trial congruency effect, it is important to look at the analysis by subplot on flanker markedness. When we consider those analyses, we can see that while L1 word flankers facilitated L1 target decisions no matter their markedness, L2 word flankers on the other hand tended to impair L2 target word recognition. Suggesting that while L1 word flankers' lexical representation were strong enough for participants to use them reliably to make their decisions, L2 word flankers' lexical representations were weaker and less reliable thus slowing down target word processing. However, this effect could be attributed more simply to a difference between the processing time required for L1 targets compared to L2 targets. The most interesting finding here is that while L2 Marked flankers impacted responses in the longer display condition, L2 Non-Marked flankers impacted responses in the shorter condition. This is confusing, but here we propose a tentative interpretation that will require further exploration in future experiments. This interpretation goes as follow; maybe Non-Marked L2-flankers *lexical representations* were available early and also quickly discarded by participants and thus impacted only responses made on short display condition, whereas the Marked L2-flankers *lexical representations* were available later and thus impacted responses only on longer display condition.

General Discussion

To sum up, in this chapter we wanted to further explore the orthographic markedness effect but this time in interaction with the linguistic context (see figure 42 for a summary). In fact both orthographic markedness and linguistic context have been considered by some researchers to modulate somewhat the bilingual lexical access (Casaponsa et al., 2019; Hoversten & Traxler, 2020). However, actual evidence supporting these claims have been rather scarce (Lauro & Schwartz, 2017; van Kesteren et al., 2012). Still we hypothesized that the combination of orthographic markers and a linguistic context might strengthen each other's influence on lexical access. To do that, we chose to use a flanker paradigm as an alternative to the whole sentence presentation, in order to ease the control of confounding variables while manipulating orthographic markedness. This paradigm was paired first with a language decision task to maximize our chances of finding an orthographic markedness effect, and second with a lexical decision task to explore lexical access more directly. In each trial the target item was accompanied by two flankers that could conflict with the target item on language membership information, so that the conflict implied the relevant information for the language decision task but not for the lexical decision task. In line with previous experiment using a similar flanker paradigm (Declerck et al., 2018, 2019; Eben & Declerck, 2019), we expected to find a trial congruency effect that would indicate that participants retrieved language membership information from flanker items. Also, in line with previous studies on orthographic markedness we expected to find a facilitative effect of orthographic markers contained in target items, at least for the language decision task (van Kesteren et al., 2012). And finally, in line with our hypothesis, we expected flanker markedness to modulate the trial congruency effect. The results to these two experiments have already been discussed separately. In this section we will endeavour to discuss them in relation to each other.

First, taken together, the facilitative target markedness effect found in the language decision task, i.e., in which they were associated with a decision advantage, and the lack of this same effect in the lexical decision task, i.e., in which they were dissociated from any decision advantage, seem to support the strategic interpretation of orthographic markedness put forth by Van Kesteren et al., (2012). Furthermore, in line with the literature on orthographic markedness, the two experiments described in this chapter have revealed an asymmetry toward L2 markers. In fact, while L2 markers facilitated language decisions no matter the display condition and slowed down pseudoword rejection in a lexical decision task, L1 markers on the other hand tended to help language decision only in longer display condition. This pattern seems to indicate that L1 markers were extracted later than L2 markers, suggesting that L2 markers may be more salient since they diverged from the L1 orthotactic distribution. To further investigate those two issues, it might be interesting to replicate the lexical decision tasks from Van Kesteren et al., study[35] in which markers could be tied to a decision advantage (L2 markers in the L1 ldt) or not (L2 markers in the L2 ldt) this time using the flanker paradigm and using the stimuli list from the general lexical decision task described here. This would have the advantage of including non-target language pseudowords (L1-like pseudowords in the L2 ldt) meaning that participants would be exposed to the same amount of markers in each language in a particular lexical decision task.

Next, in this study, we were able to replicate the trial congruency effect on language decisions described by Eben and Declerck (2019), and to extend it to a short display condition. However, we did not manage to replicate the trial congruency effect described by Declerck et al., (2018) even in the short display condition we shared with their experiment. First, the present study highlights the advantage of using both long and short display conditions to better understand the temporal dynamics involved in word processing when flanker items are involved. Second, this pattern of results seems to indicate that participants were able to retrieve language membership information from the linguistic context / flanker items, and

[35] *A simple way to implement that would be to divide our general lexical decision task into several blocks each of which would be adapted into language-exclusive, either L1 or L2, lexical decision task.*

that this information would impact responses when the conflict implying the two language membership information, that of the target and that of the flanker, was relevant for the task at hand, i.e., in a language decision task but not in a lexical decision task. The linguistic context, here represented by the flanker items, may also exert somewhat of a strategic impact on bilingual lexical access. To investigate this issue, we would need to design an experiment in which the conflict between target and flankers would be about a relevant information for lexical decisions like the lexical status of both items. However, in doing that we would stray from our focus on orthographic markedness.

Finally, evidence for the interaction between flanker markedness and trial congruency and their impact on target processing is scarce, especially in the lexical decision task. Still, it looks like L2 markers included in the linguistic context modulated the trial congruency effect in the short display condition to the language decision, suggesting that (sublexical) language membership information coming from the context was detected early when orthographic markers were involved. Consequently, even though L2 Marked flankers lexical representations seemed to be available late, thus not impacting target lexical decisions, we might imagine that the language information given by L2 Marked flankers is sufficient to reduce cross-language interaction effects. Considering Interlingual homographs for example, a French linguistic context should encourage the corresponding CORNER interpretation of the word COIN, whereas presented as an isolated word, both French and English interpretations would be activated by the word COIN. As of today studies exploring this question have mainly manipulated the semantic context surrounding the target word, but it might be interesting to investigate the impact of the mere presence of a salient language membership information in the context. To do that we might want to conduct another lexical decision task paired with a flanker paradigm in which target items would either be language ambiguous (Interlingual homographs and Cognates) or not (controls) and in which flankers, which would always be pseudowords since we want to focus on language membership information, would either be marked or not. We would expect the interlingual homograph and the cognate effects to arise in trials for which the flankers were not marked but to be reduced or even non-existent in trials for which flankers were marked either for L1 or L2.

To conclude, the flanker paradigm used in the two experiments described in this chapter was able to exhibit both a linguistic context effect, i.e., the trial congruency effect, and an orthographic markedness effect. Thus, this paradigm seems to constitute an important first step in the investigation of the interaction of the two factors and their impact on bilingual lexical access. Yet, the flanker paradigm paired with the lexical decision task described in this chapter did not provide strong evidence supporting a stronger impact of Marked linguistic context compared to Non-Marked linguistic context. Still, other iterations of the flanker paradigm outlined in the discussion above might shed some much needed light on the issue before we set out to explore this interaction through whole-sentence processing.

Experiment 1: Language decision		LONG DISPLAY	SHORT DISPLAY
L1 TARGET MARKEDNESS	RT	Trend toward a markedness effect (t=1.77)	No markedness effect
	ACC	Facilitative markedness effect (z=-2.54)	No markedness effect
L2 TARGET MARKEDNESS	RT	Facilitative markedness effect (t=-6.05)	Markedness effect more important for HP participants (t=-2.61)
	ACC	Facilitative markedness effect (z=-3.69)	Facilitative markedness effect (z=-3.52)
TRIAL CONGRUENCY	RT	Trial congruency effect (t=5.36)	Trial congruency effect (t=2.96) L2 targets (*p*=.005) vs L1 targets (*p*>.10)
	ACC	Trial congruency effect (z=3.95) L1 targets (*p*<.001) vs L2 targets (*p*=.358)	Trial congruency effect (z=13.22)
TRIAL CONGRUENCY **Marked Flankers**	RT	Interaction with Flanker language (t=-3.04) L2 flankers (*p*<.001) vs L1 flankers (*p*=.816)	Trend for trial congruency effect (t=-1.74) L2 targets (*p*=.008) vs L1 targets (*p*>.10)
	ACC	Interaction with Flanker language (z=-3.68) L2 flankers (*p*<.001) vs L1 flankers (*p*=.714)	Trial congruency effect (z=-11.14)
TRIAL CONGRUENCY **Non-Marked Flankers**	RT	Interaction with Flanker language (t=-1.91) L2 flankers (*p*=.037) vs L1 flankers (*p*=.957)	No Trial Congruency effect
	ACC	Trial congruency effect (z=-3.70)	Trial congruency effect (z=-7.88)

Experiment 2: Lexical decision		LONG DISPLAY	SHORT DISPLAY
L1 TARGET MARKEDNESS	RT	No markedness effect	No markedness effect
	ACC	No markedness effect	No markedness effect
L2 TARGET MARKEDNESS	RT	No markedness effect	No markedness effect
	ACC	No markedness effect	No markedness effect
TRIAL CONGRUENCY	RT	Trial congruency effect (t=4.17) L1 targets (*p*<.001) vs L2 targets (*p*=873)	No Trial Congruency effect
	ACC	Trial congruency effect (z=3.71) L1 targets (*p*<.001) vs L2 targets (*p*=.419)	No Trial Congruency effect
TRIAL CONGRUENCY **Marked Flankers**	RT	Interaction with Flanker language (t=-4.49)	Interaction with Flanker language (t=-3.86) L1 targets (*p*<.001) vs L2 targets (*p*=131)
	ACC	Trial congruency effect (z=-2.82)	Trial congruency effect (z=-2.34)
TRIAL CONGRUENCY **Marked Flankers**	RT	Interaction with Flanker language (t=3.45) L1 targets (*p*<.001) vs L2 targets (*p*=719)	Interaction with Flanker language (t=4.18)
	ACC	Trial congruency effect (z=-3.23)	No Trial congruency effect

Figure 42 : Summary of the results from experiment 1 and 2

GENERAL DISCUSSION

In this PhD thesis, we aimed at investigating the impact three sublexical variables derived from the acquisition of a 2nd written language may exert on L2 visual word processing. The first variable refers to the language specificities (and similarities) that arise from the comparison of the two orthographic distributions (KN and VR vs. OU for a French-English speaker). It is referred to as orthographic markedness and is considered a reliable language membership cue that has been assumed to guide bilingual lexical access by some researchers (Casaponsa et al., 2019). The second variable refers to the discrepancies and to the similarities a second language reader may encounter when it comes to the Grapheme-to-Phoneme Correspondences each language will apply to a single (and shared) grapheme (the I in TIME vs. the I in FISH for a French-English Speaker). This is referred to as cross-language GPC congruency and was expected to provoke some cross-language interaction effects (FISH<TIME; Commissaire et al., 2014; Hevia-Tuero et al., 2021). The third variable is specific to the second language and refers to the consistency with which this written language will associate a grapheme to a phoneme, following which an L2-learner will have to map a

grapheme with one (EE= \i\) or several (EA= \i\ or \ɛ\) phonemes. This is referred to as L2 within-language GPC consistency and an impact of this variable on L2 word processing was thought to reflect the L2 reader sensitivity to L2-specific sublexical GPC variations.

To look into those variables' impact on L2 visual word recognition we conducted two studies. In our first study we endeavored to explore each sublexical variable separately by controlling for the other two variables. To do that we used three L2-specific lexical decision tasks paired with three naming tasks, which allowed us to get access to the recognition and to the pronunciation components of L2 visual word processing. In our second study, we attempted to investigate the interaction that might exist between one of these variables, namely orthographic markedness, and another source of language membership information that is, the linguistic context. To achieve that we paired a language decision task and a lexical decision task with a flanker paradigm adapted from Declerck et al., (2018, 2019) and Eben and Declerck, (2019).

Interestingly, the participants we recruited to perform the experiments described here all shared a comparable experience of L2-acquisition since they were all French-native adult speakers that learned English as a second language through the school system. Therefore, we could safely assume that they had previously acquired knowledge of French (L1) orthographic regularities and Grapheme-to-Phoneme correspondences when they were first exposed to the English written language. This decision was made to ensure that the L2 proficiency effects we were expecting to find were not confused with other variables tied to L2 acquisition experience such as Age of Acquisition (early acquisition vs. late acquisition) or acquisition environment (classical classroom; immersion classroom or at home etc.) with the added benefit that it better reflected the L2 acquisition experience of the French population.

I. Separate impact of each sublexical variables on L2 visual-word processing.

The first pair of experiments described in this study revealed no effect of orthographic markedness on either lexical decisions or naming responses. As previously discussed, this lack of an orthographic markedness effect on lexical decisions was likely due to the control of the GPC component in our study rather than any confounding effects related to task requirements, demographic specificities or material selection. In fact, we shared those parameters with an earlier study from Commissaire et al., (2019) in which they were able to exhibit a facilitative orthographic markedness effect. Moreover, our results on the naming task are also different from those described in an earlier experiment using a naming task and manipulating orthographic markedness (Oganian et al., 2016). However, this discrepancy is easily attributable to the fact that while this earlier study included only pseudowords, we on the other hand focused on word responses. Taken together, those two studies might indicate that orthographic markedness impacts naming responses only when no stronger lexical representations are available.

Therefore this first pair of experiments seems to indicate that the exclusive use of known words might have discouraged a sublexical processing and with it might have prevented us from exhibiting any orthographic markedness effect. Paired with the fact that we compared our Marked words to non-Marked words that were all incongruent when we considered L1 GPCs (e.g., HOUSE), and thus contained another type of L2-specificities, we might not have put ourselves in the ideal condition to find an orthographic markedness effect. Still, the lack of orthographic markedness we reported is informative, since this pattern of results seem to suggest that orthographic markedness' impact on word processing is not completely independent from the markers' association with the GPCs[36].

The second pair of experiments uncovered a surprising inhibitory GPC congruency effect on naming responses that was otherwise not impactful when considering lexical decisions. This pattern indicates that while the addition of a second GPC to a single grapheme brought on by second language acquisition did not impact an L2 word recognition, it did impact its pronunciation. Curiously, at first glance, the naming results seem to indicate that GPC incongruency benefited L2 visual word processing rather than hindering it. However, as previously discussed, it is important to consider how we selected and categorized our items on the congruency variable. In fact, the material was selected by a French native speaker that learned English as a second-language through the school system (late-learner). Similarly, once selected, the words used in this experiment were rated by a group of French-native, late English-learners, as following the French GPC rules or not. Therefore, our categories might not reflect a true congruency/incongruency dissociation but rather a perceived congruency variable that results from second-language learners phonological misperceptions. Nevertheless, this second pair of experiments seem to also indicate that naming tasks are better suited to the investigation of a GPC variable such as GPC congruency when compared to classical lexical decision tasks.

As noted previously, only a few studies have explored the impact cross-language GPC congruency may exert on L2 processing. Interestingly, our results to both lexical decision and

[36] *Besides, the inhibitory effect Oganian found on pseudowords are in line with this assumption, since it might be interpreted as showing that markers had weaker GPC representations than non-marked orthographic patterns.*

naming tasks did not align with any of the results described in those studies. In fact, cross-language GPC congruency was found to help processing in two letter detection tasks (Commissaire et al., 2014; Hevia-Tuero et al., 2021). However, this first discrepancy might merely be attributed to the difference in task requirements. If we take into consideration those two previous studies along with our own results, it looks like cross-language GPC congruency does not impact word recognition processes *per se*, but may benefit the sublexical processes involved in letter detection while it would rather hinder the sublexical processes involved when reading the word aloud, where it might reflect the stronger control L2-learners apply to shared and congruent GPCs (vs. L2specific/incongruent GPCs) that may sound too much like an L1 pronunciation when asked to name aloud a list exclusively containing L2 words.

Yet, an earlier study from Jared & Kroll, (2001) which took another approach to what we call the cross-language GPC congruency variable, found that L2 words with L1 enemies (e.g., BAIT-LAIT) that could be likened to our set of incongruent words, led to worse naming performances when compared to L2 words without L1 enemies (e.g., BUMP). This second discrepancy may be attributed, at least in part, to the unit each study chose to manipulate. In fact, while we considered simple grapheme units and the existence of a shared GPC association (I in FISH) and a language-specific GPC association (I in TIME) for each of them, Jared & Kroll, considered the existence of word body neighbours (related orthographically) that could be pronounced differently and compared them to word bodies that had a unique pronunciation, most of them because it was orthographically specific to one language (SWIFT and DUSK contains orthographic markers of English). Thus, on closer inspection, the two naming tasks may not test for the same variable. Still, it may be interesting to explore the separate impact of each manipulation on L2 word naming while controlling for orthographic markedness.

The third pair of experiments revealed a facilitative impact of L2 within-language GPC consistency that was trending in the lexical decision task and fully significant in the naming tasks. Again, the naming task seemed better suited to the investigation of a GPC variable when compared with the lexical decision task. Importantly, the results to this naming task indicates that L2-learners call on every L2 GPC/print to sound mapping they have acquired when presented with a grapheme in an L2 environment. Those results are in line with the previous experiments from Jared & Kroll, (2001), in which they found that L2 words with L2 enemies (e.g., BEAD-DEAD) led to poorer naming performances compared to L2 words without L2 enemies (e.g., BUMP-JUMP). Interestingly, Jared & Kroll, took care to select target words that had a more frequent orthographic neighbour so that the L2 enemies' pronunciation would be more frequent than the target word pronunciation actually is. This control may have strengthened their consistency manipulation (see Jared et al., 1990). We on the other hand chose to present our participants with the two pronunciations tied with a rhyme without considering their relative frequency, this was done to make the GPC inconsistency more salient but may have ended up weakening our GPC consistency manipulation.

Besides, this last pair of experiments also uncovered an interaction between the orthographic markedness effect and the L2 within-language consistency variable. Even more interestingly, this interaction was reversed in the lexical decision task and in the naming task. In fact, while the inhibitory orthographic markedness effect was significant only in the inconsistent word subplot in the lexical decision task, the same effect was restricted to the consistent word subplot in the naming task. Thus this pattern of results confirms that the orthographic markedness effect is intricately tied to the GPC component to which it is associated and complete the investigation of orthographic markedness effect undertaken in the first pair of experiments by giving an idea of what is required for this effect to arise in each task. In fact, it looks like the word recognition process needs to be slowed down, by adding some GPC inconsistency, for it to exhibit an orthographic markedness effect. On the other hand, the sublexical pathway need to be enabled, by using fully-consistent GPCs, for the naming task to reveal any orthographic markedness effect.

Overall, the experiments described in chapter III did allow us to explore the separate impact of orthographic markedness, cross-language GPC congruency, and L2 within-language GPC consistency on L2 visual word processing, from two point of view (see table 51). They also allowed us to explore the interaction between orthographic markedness and L2 within-language GPC consistency, although further investigation are required in order to explore the interaction between the cross-language GPC congruency and the within-language GPC consistency, and also to compare Marked words to Non-Marked and congruent words.

Table 51: Summary of the results from study 1 reported in Chapter III

		Lexical Decision Task		Naming Task
		Words	Pseudowords	Words
Orthographic Markedness	Reaction Times	No orthographic markedness effect	No orthographic markedness effect	No orthographic markedness effect
	Accuracy	orthographic markedness effect	Orthographic markedness effect	No orthographic markedness effect
GPC Congruency	Reaction Times	No GPC congruency effect	No GPC congruency effect	GPC Congruency effect (more important for higher-proficiency participants)
	Accuracy	No GPC congruency effect	No GPC congruency effect	GPC Congruency effect
GPC Consistency	Reaction Times	Trend for a GPC consistency effect Orthographic markedness effect (in the inconsistent word subplot)	No GPC consistency effect No Orthographic markedness effect	GPC Consistency effect (only for higher-proficiency participants) Orthographic markedness effect (in the consistent word subplot)
	Accuracy	No GPC consistency effect No orthographic markedness effect	No GPC consistency effect No Orthographic markedness effect	GPC Consistency effect Orthographic markedness effect

II. The interaction between Orthographic Markedness and the Linguistic Context

The two language decision tasks paired with a flanker paradigm revealed that only L2 orthographic markers contained in the target word were able to help language detection and that, even when a linguistic context was provided. This asymmetric facilitative effect is in line with the literature available on isolated word processing, and with the assumption that the benefit of orthographic markedness for language membership detection derives from its divergence from L1 orthographic patterns. Interestingly, the short display condition uncovered an interaction effect between L2 proficiency and orthographic markedness (of the target) suggesting that at such a display duration only higher-proficiency participants were able to benefit from the presence of an L2 marker. No such interaction effect was found in the longer display duration, indicating that with longer exposition even low-proficiency participants were able to extract and use orthographic markers to help detect the language membership of a known L2 word. This is interesting because even though previous studies were able to show that bilinguals and monolinguals alike were able to use orthographic markers in such a way (Borragan et al., 2020; Casaponsa et al., 2014; Jared et al., 2013), no other studies were able to show that the time course for its detection depended on L2-proficiency.

Moreover, those language decision tasks were able to give us some information on the detection of language membership cue provided by the linguistic context. In fact, they revealed that L2-learners were able to extract language membership information from the context and that this information impacted language membership detection of the target word. Again, the short display duration uncovered an interesting interaction effect, this time between trial congruency and flanker markedness that was not found in the longer display duration. This pattern of results seems to indicate that while language membership information coming from the context is extracted from the sublexical level in the short display condition, reliance on lexical information from the flanker item was enabled by the longer display condition only.

On the other hand, the two lexical decision tasks were not as successful in exhibiting an orthographic markedness effect. In fact, the presence of either an L1 or an L2 orthographic marker in the target item did not impact the lexical decisions made on the word items, no matter the display duration. Still, the L2 markers were found to hinder pseudoword rejection. This pattern of results seems to indicate that while participants were sensitive to the presence of L2 markers, they were not able to benefit from the language membership cue it provides for actual word recognition, when a lexical representation was available, in a generalized lexical deicion task. Moreover the absence of a clear trial congruency effect in the two lexical decision tasks show that the language membership information coming from the linguistic context, represented here by the flanker items, did not overly impact target processing when it was not required to make an accurate decision, i.e., when indicating the lexical status of the target item in a general lexical decision task.

The closer inspection of the trial congruency effect involving the orthographic markedness of the flanker items was again not as successful as we hoped, but did hint at some interesting findings. In fact, Marked and Non-Marked L1 flankers had a facilitatory impact on lexical decisions to both L1 and L2 target words, suggesting that the lexical representations associated to L1 words presented as flanker items were strong enough for participants to rely on them when making a lexical decision on target items, or more simply that L2-targets required longer processing times compared to L1 targets. On the other hand, L2 flankers tended to impair lexical decisions made on L2 target words. This pattern suggests that (1) L1 target lexical representations were strong enough not to be impacted by the presence of L2 words as flanker items, (2) L2 lexical representations, and especially those tied to the word presented as flanker items were weaker, leading to a slowing down of target word recognition. Most interestingly, but also most confusingly, L2 flankers' impact on target word recognition seemed to be modulated by its orthographic markedness. It looks like lexical representations of Non-Marked L2 words presented as flanker items were available early on, whereas the lexical representations of Marked L2 words presented as flanker items were available only later on, leading to a particular pattern of results and encouraging further investigation on this issue.

Overall, the experiments described in chapter IV provided us with some interesting findings regarding the reliance on orthographic markers contained in the target word but also in the linguistic context (see table 52).

As for the orthographic markers contained in the target words, this investigation revealed that they benefited language membership detection but did not impact actual word recognition when no decision advantage could be derived from them. This pattern is in line with van Kesteren et al.,'s position on the role played by orthographic markers in which they were assumed to be used strategically when an advantage could be gained from early language detection. Still, the presence of an inhibitory orthographic markedness effect on pseudoword rejection in the lexical decision task, albeit restricted to the short display condition, seems to indicate that actual sensitivity to orthographic markers was not modulated by task requirements. The impact orthographic markers may exert on L2 word processing on the other hand, looks like it depends on the task demands. Moreover, in line with the literature available on orthographic markedness (Casaponsa et al., 2014; Duñabeitia et al., 2020; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012), our participants appeared to be more sensitive toward L2 markers than toward L1 markers. This finding is even more interesting when we consider the dynamics revealed by the two display condition in interaction with the two tasks. In fact, while L2 markers impacted responses in short display condition in both language decision and lexical decision tasks, L1 markers tended to impact responses only in the longer display condition to the language decision task. This suggests that (1) L2-learners detect L2 markers more rapidly than they do with L1 markers, (2) and that L1 markers were used only when language membership information was relevant to perform the task at hand. Similarly, the fact that in the lexical decision task, the impact of L2 markers was restricted to pseudoword rejection suggests that L2 markers were used as long as a more thorough processing was prevented.

The trial congruency effect evidenced in the language decision task indicates that participants were able to extract language membership information form the context. However, they seemed to use this information only when it was relevant for the task at hand since the trial congruency effect was mostly absent from the lexical decision task results. This finding is interesting on its own right, yet we were more curious about the modulation of this trial congruency effect by the presence or absence of orthographic markers in the flanker items. In fact, we used this flanker paradigm in order to test the assumption that the presence of an orthographic marker in the linguistic context (here represented by the flanker items) would strengthen the impact the linguistic context may hold over the bilingual lexical access. Following this assumption, we expected the trial congruency effect which was considered to be an indicator of the linguistic context impact on target word processing, to be stronger when an orthographic marker was included in the flanker items compared to when they did not contain any orthographic marker. However, actual evidence supporting this assumption is scarce since only the short display condition to the language membership task revealed a modulation of the trial congruency effect by the flankers' orthographic markedness variable. Still, the exploration of this issue uncovered some interesting results suggesting that while L2 markers included in the linguistic context provided participants with language membership information early on in the processing, it (i.e., L2 markers) delayed access to the lexical representation they were tied to, preventing it from influencing target word recognition.

Table 52: Summary of the results from Study 2 reported in Chapter IV

		Language Decision	Lexical Decision	
		WORDS	WORDS	PSEUDOWORDS
Target Markedness	Long Display	Significant for L2 targets Trending for L1 targets	No orthographic markedness effect	No orthographic markedness effect
	Short Display	Only for L2 targets (more important for higher-proficiency participants)	No orthographic markedness effect	Only for L2 targets
Trial Congruency	Long Display	On both L1 and L2 targets	On L1 targets only	No trial congruency effect
	Short Display	On L2 targets only	No trial congruency effect	No trial congruency effect
Trial congruency by Flanker Markedness	Long Display	Only for L2 flankers, no matter their markedness	Facilitative for L1 flankers Inhibitory for L2 Marked Flankers	No trial congruency effect
	Short Display	Only for L2 marked flankers No trial congruency effect for Non-marked flankers	Facilitative for L1 flankers Inhibitory for L2 Non-Marked Flankers	No trial congruency effect

III. Additional theoretical implications and some methodological ones

The implications of this PhD thesis are twofold. From a theoretical point of view, it allowed us to get a better idea of the separate impact of each sublexical variable on several aspect of L2 visual word processing models. From a methodological point of view, our investigation into these sublexical variables allowed us to better determine the conditions required, both stimuli-wise and task-wise, to exhibit an effect of each of these variables. It also highlighted the importance of getting several points of view on a single issue to obtain a thorough picture on the topic. The theoretical implications of each study have already been thoroughly discussed above. Thus, in this section we set out to discuss first the theoretical, and then the methodological implications derived from the combination of the two studies.

1. Last theoretical implications for orthographic markedness' investigation

A. The orthographic markedness effect, strategic after all?

As stated previously in chapter III, following the absence of an orthographic markedness effect in a lexical decision and in a naming task, and in chapter IV, following the absence of the same effect in a general lexical decision task, our results contradict the position adopted by Casaponsa et al., 2019, about the automatic use of orthographic markers. On the contrary, even though we did not aim at settling this debate, the pattern of results we uncovered lean toward van Kesteren et al.,'s position on orthographic markedness strategical use, without being able to confirm it. To do that we would need to run further investigations where we would manipulate the decision advantage attributed to the orthographic markers.

Still, we wanted to propose a tentative reconciliation between our results and Casaponsa et al.,'s assumption. We could assume that the word recognition process may be too efficient in bilingual participants to reveal a sublexical effect such as the orthographic markedness effect, at least when it is tested directly through a classical lexical decision task. To find this effect, one may need to use a priming paradigm (Casaponsa et al., 2019; Commissaire, 2022) in which the lexical processing of the Marked words would be hindered (since they are presented as primes), or one may need to slow down the word recognition by manipulating variables such as GPC consistency (Experiment 1A) or orthographic neighborhood density (Commissaire et al., 2019).

B. Are orthographic markers a hindrance for the bilingual lexical access?

Furthermore, when considering the literature on the topic, we were encouraged, if any effect was to be found, to expect a facilitative orthographic markedness effect on the bilingual lexical access. Thus, the few inhibitory effects we uncovered in the lexical decision task (on inconsistent words), in the naming task (on consistent words), and on the flanker items (in the general lexical decision task), are surprising. Following the results obtained in this thesis, it looks like the presence of orthographic markers slows down the bilingual lexical access rather than promoting it. Two studies conducted by Bartolotti & Marian in 2017, might shed some light on the issue that is raised here. In fact, in those experiments the authors showed that English monolingual participants acquired native-language-like pseudowords[37] (e.g., NISH) with more ease than they did native-language-different pseudowords (e.g., GOPF), which could be considered as being marked for the foreign language. They also found a similar result for bilingual participants learning a third language.

---

[37] *To create their pseudowords the authors considered both bigram frequency and orthographic neighborhood density. For full disclosure, their categories were referred to as orthographically word-like and not orthographically word-like.*

Taken together all those results suggest that acquiring a word containing an orthographic pattern that is unfamiliar to the individual is more difficult than learning a word containing only orthographic patterns that are familiar to them. This hindrance for the acquisition would then lead to weaker lexical representations for Marked L2-words compared to Non-Marked L2-words, even in higher-proficiency L2-learners, effectively slowing down the L2-Marked word recognition. Yet, in fully balanced bilinguals, when the lexical representations associated to L2-Marked words become strong enough, participants may benefit from the presence of a marker to recognize the word just as Casaponsa et al. (2016;2019) suggested.

C. A word on measuring orthographic markedness

As a last consideration we want to bring up the issue of orthographic markedness measurement. In fact, even though clear definitions have been proposed for each type of markers (see Chapter II) no guidelines as to how one should categorize items on the markedness spectrum are available in the literature, which is particularly troublesome when working with frequency-based markers. Moreover, what constitutes an orthographic marker depends on the language pair under scrutiny. Some language pair will "produce" a lot of language-specific bigrams (e.g., Spanish-Basque) whereas other language pairs will only produce a few of those markers (e.g., French-English) encouraging the use of frequency-based markers.

Importantly, this methodological issue might have some far-reaching consequences since it might account for the discrepancy found in the literature when it comes to the orthographic markedness effect on the bilingual lexical access. In fact studies having reliably found an impact of orthographic markedness in the bilingual lexical access have used language pairs presenting a lot of language-specific bigrams (Casaponsa et al., 2014, 2019; Casaponsa

& Duñabeitia, 2016) whereas the ones that did not find such strong evidence, including our own studies, included frequency-based markers in their item lists (van Kesteren et al., 2012). We could therefore assume, in line with Casaponsa et al.’s and Commissaire's (2021), hypotheses, that language-specific bigrams are able to promote target-language lexical representations AND to inhibit non-target language lexical representations. However, frequency-based markers may not be able to inhibit non-target lexical representations since they are included in some of those representations (e.g., the EU bigram is more frequent in French but can still be found in some English words like NEUTRAL).

2. Methodological implications

A. A tight knot of sublexical variables to unravel

A first methodological implication of this thesis comes from the finding that the impact orthographic markedness exerts on L2 visual word processing appear to be tied to the GPC component it is associated with (see Study 1 described in chapter III). Thus, it appears that comparing Marked L2 words to Non-Marked L2 words that are all incongruent, i.e., L2 specific when considering the GPC component, is not suited to uncovering an orthographic markedness effect. Yet, comparing Marked L2 words (that are always tied to L2-specific GPCs) to Non-Marked L2 words that are all congruent, i.e., not language specific when considering either orthographic patterns or the GPC component, expose researchers to the possibility of a confounding GPC congruency effect since Marked L2 words are by design language specific on the GPC component. A first step to better understand the dynamic we uncovered here would be to compare a set of L2 Marked words to two sets of L2 Non-Marked words, one containing congruent words, the other containing incongruent words (but see Commissaire &

Demont, 2022 for that comparison with dyslexic readers). More importantly, this finding encourages future studies on orthographic markedness to at least take into account the cross-language GPC congruency of their Non-Marked words in order to ensure the accuracy of their ensuing interpretations. Similarly, following the pattern of results we uncovered about the modulation of the orthographic markedness effect by L2 GPC consistency depending on the task requirements, future experiments on the topic of orthographic markedness ought to carefully select their items by considering the L2 GPC consistency variable in accordance with their research questions and with the paradigm they plan to use.

B. The benefit of a multi-task approach to untangle this knot

The exploration of cross-language GPC congruency and of L2 within-language GPC consistency described in chapter III involved two different tasks aiming at exploring the lexical access of our participants, namely some classical L2 lexical decision tasks and some naming tasks. This approach revealed a theoretically interesting finding, since it shows that while our participants were sensitive to both cross-language and within-language GPC variations, as shown by the naming tasks, those variations did not impact L2-word recognition processes, as shown by the lexical decision tasks. This finding highlights the benefit of pairing a classical decision task with a naming task when setting out to investigating GPC variables, an approach that was successfully adopted by early studies on native-language GPC consistency (Jared, 1997; Jared et al., 1990)

Pairing a lexical decision task with a naming task was not as revealing for the orthographic markedness variable. In fact even though it was argued that the naming task could be more prone to show an orthographic markedness effect (see Schmalz et al., 2017 for a similar hypothesis on bigram frequency effects), we did not find a straightforward orthographic markedness effect in either tasks. Still, its combined investigation with the L2 within-language GPC consistency revealed an interesting pattern hinting at the condition required for each process to reveal an orthographic markedness effect. This further advocates for the benefit of pairing a lexical decision with a naming task to obtain a clearer picture of the issues raised by sublexical variables, but also for the importance of considering the full set of sublexical variables into account when exploring one of them.

C. And the benefit of multiplying the display conditions for the same purpose

The next study was designed to explore the impact of orthographic markedness in the linguistic context on target word processing. Even though it harbored little evidence supporting our hypothesis on the matter, the combined use of two different tasks and two different display duration conditions turned out to uncover a few interesting findings that have important methodological implications for future studies.

Again, the pairing of a lexical decision task with a language decision task was illuminating. Similarly to the first task comparison we discussed, the comparison of the lexical decision task to the language decision task used in the second study revealed that while our participants were sensitive to orthographic markedness, contained in the target items and possibly even those that were included in the linguistic context, they were not able to benefit from it when the language membership information was not at the center of the decision they had to make. Thus, the second study underline the benefit gained by the multi-task approach to a single issue.

Besides, the comparison between the two display duration conditions we implemented in each task provided us with some information about the time course of L2 visual word processing, and particularly for the flanker items. In the lexical decision task for instance, its seems that in the short display duration condition, the interaction between the target's and the flanker's language membership information relied on a sublexical processing of the flanker item combined with a weaker L2 lexical representation for the target item (vs. L1 lexical representations). In the long display duration condition on the other hand, the same interaction seemed to rely on a lexical processing of every item on display, including the flanker items. Moreover, when we look at the lexical decision task, the overall trial congruency effect was not significant no matter the display duration condition. Yet, a closer inspection, taking into consideration the markedness of the flanker items, suggest that the lexical processing of an L2 Marked word is delayed when compared to that of a Non-Marked L2 word. Taken together, all those comparisons seem to indicate that the language membership tied to an L2 Marked word could be retrieved earlier than the one that is tied to an L2 Non-Marked word, whereas it appears that the reverse might be true of the lexical representations. Hence, these findings emphasize the benefit of combining the flanker paradigm with several tasks, but also with several display durations to establish a full picture of the issue under consideration.

D. A quick note on the participants' profile

As a side note, even if our study did not directly investigate the impact the participants' second language history may have on the appearance of each and every variable introduced in this document, it did highlight the importance of broadening the research on bilingual processing to a larger specter of bilinguals and L2-speakers (see de Bruin, 2019). This has already led to some interesting findings especially when it comes to explaining the origin of the orthographic markedness effect and when determining what it is used for (Borragan et al., 2020; Casaponsa et al., 2014; Commissaire, Audusseau, et al., 2019). Moreover, our first study also underlined the importance of selecting words that are suited to the participants we are planning to test, i.e., that reflects their experience with the L2-vocabulary.

E. Conclusion on the methodological implications

Thus a close inspection of the results evidenced in this PhD thesis encourages future studies aiming at the exploration of a sublexical variable in L2 word processing, to carefully consider every other sublexical variables mentioned here and the way they may impact each other, to improve L2 visual word recognition models. It also urges those studies to multiply the tasks and the parameters attached to them in order to obtain a full picture of the issue at hand. Finally, this research also incites future studies to pay close attention to their participants' particular second language history, particularly when it comes to the way they acquired this L2, and to select the material in a way that reflect the actual L2-learners experience with L2 "visual" vocabulary.

3. The last pages of a beginning

Until now, we have been able to avoid mentioning the impact the COVID-19 pandemic has had on this research project, but it seems to us necessary to say a word on the context surrounding this thesis in order to better understand the choices that were made during its realization.

In fact, when we started out on this project we had a completely different plan in mind. Slight changes were not unexpected of course, but about a year and a half into this PhD thesis we had to decide on some major redirections in order to adapt to the shift in the larger societal context we were experiencing. The original aim of this thesis was not merely to explore the impact of orthographic markedness, cross-language GPC congruency and L2 within-language

GPC consistency on L2 word recognition, but to explore their impact (on L2 word recognition) at different stages of L2-acquisition. Therefore, the plan was to test teenagers that just began to learn English as an L2 from their 1st year to their 4th year in middle-school, in order to obtain a developmental trajectory of those effects.

However, about a week before we were scheduled to run our first real "batch" of experiments in a middle-school, we all went into full-lockdown. And even when the schools re-opened, we had a hard time convincing teachers and principals to let us in. As a consequence, during those 4 to 5 years, we only managed to test 20 teenagers. Therefore, the first study we presented here, which was designed as a control experiment for this developmental study, became our main focus.

Similarly, we planned an eye-tracking experiment on whole-sentences for the second study, but with the multiple lockdowns and sanitary measures, and considering that we needed some training courses on how to run those experiments, we weren't able to conduct this study. Thus, for this thesis we decided to focus on the flanker paradigms which were already conceived as a first step toward an eye-tracking experiment.

Yet, the pandemic and the many challenges it brought to the table had also some more positive consequences for this research project. For instance, the difficulty we encountered when recruiting participants for in person testing encouraged us to learn how to implement and conduct online experiments via platforms such as the psytoolkit platform. In doing so we were also able to find out which experiments were possible to realize online, those for which it was impossible and those that needed adapting, and to determine what were the advantages and limitations of such experiments.

Likewise, the long days spent away from the lab without any occasion to do testing were also a good opportunity to stop and take a second more thorough look at the literature available on our subject, but also to read more in depth about topics surrounding our main area of interest, topics that might have been overlooked or barely looked at in more normal circumstances. In our case it even led to the redaction of a review of literature about orthographic markedness that hopefully will soon be published.

Besides, even though we weren't able to complete the experiments we planned in a timely manner, we still intend to do so. As you read this for instance, the developmental study is currently underway, and making some progress. As for the eye-tracking experiment, sentences have already been selected and now need to be tested for confounding variables before we begin testing. Moreover, as you may have noticed scattered throughout this document we also have many ideas about what to test next, from the dynamic between orthographic markedness and cross-language GPC congruency to the nature of the orthographic markedness effect when a linguistic context is provided. For some of these issues we already have an idea as to how we may explore our hypotheses, for others we still need to ponder on the matter. Some questions that are intriguing to us have been only alluded to in this document, this is the case of the relation between orthographic markedness and orthographic redundancies, (i.e., are orthographic markedness effects merely a special case of bigram frequency effects?, does orthographic markedness play the same role that orthographic redundancy does in monolingual processing?), but also of the best way to compute orthographic markedness (on a frequency continuum? via illegal vs. legal bigram dissociation? Do we need to take into account the bigram frequency difference in each language separately on top of the difference between languages?) and how to implement sublexical units in L2 visual word recognition models? Those issue deserves further investigation in the future. All this to say that we consider the research presented in this document as only the beginning of a more thorough examination of the issue raised by the exploration of L2 visual word recognition.

REFERENCES

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals : Neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689-698. https://doi.org/10.1017/S1366728916000225

Antzaka, A., Martin, C., Caffarra, S., Schlöffel, S., Carreiras, M., & Lallier, M. (2018). The effect of orthographic depth on letter string processing : The case of visual attention span and rapid automatized naming. *Reading and Writing*, *31*(3), 583-605. https://doi.org/10.1007/s11145-017-9799-0

Arana, S. L., Oliveira, H. M., Fernandes, A. I., Soares, A. P., & Comesaña, M. (2022). The cognate facilitation effect depends on the presence of identical cognates. *Bilingualism: Language and Cognition*, 1-19. https://doi.org/10.1017/S1366728922000062

Baayen, R H., Piepenbrock, R, & Gulikers, L. (1995). *CELEX2* (p. 287744 KB) [Data set]. Linguistic Data Consortium. https://doi.org/10.35111/GS6S-GM48

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340-357. https://doi.org/10.1037/0096-1523.10.3.340

Bassetti, B. (2008). Orthographic input and second language phonology. In *Input matters in SLA* (p. 191-206).

Baten, K., Hofman, F., & Loeys, T. (2011). Cross-linguistic activation in bilingual sentence processing : The role of word class meaning. *Bilingualism: Language and Cognition*, *14*(3), 351-359. https://doi.org/10.1017/S1366728910000246

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bijeljac-babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, *25*(4), 447-457. https://doi.org/10.3758/BF03201121

Borragan, M., Casaponsa, A., Antón, E., & Duñabeitia, J. A. (2020). Incidental changes in orthographic processing in the native language as a function of learning a new language late in life. *Language, Cognition and Neuroscience*, 1-10. https://doi.org/10.1080/23273798.2020.1784446

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners : Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, *109*(4), 383-396. https://doi.org/10.1016/j.jecp.2011.03.012

Brysbaert, M. (2001). Prelexical phonological coding of visual words in Dutch : Automatic after all. *Memory & Cognition*, *29*(5), 765-773. https://doi.org/10.3758/BF03200479

Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals : Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 137-148. https://doi.org/10.1037/0096-1523.25.1.137

Buetler, K. A., de León Rodríguez, D., Laganaro, M., Müri, R., Nyffeler, T., Spierer, L., & Annoni, J.-M. (2015). Balanced bilinguals favor lexical processing in their opaque language and conversion system in their shallow language. *Brain and Language*, *150*, 166-176. https://doi.org/10.1016/j.bandl.2015.10.001

Buetler, K. A., de León Rodríguez, D., Laganaro, M., Müri, R., Spierer, L., & Annoni, J.-M. (2014). Language context modulates reading route : An electrical neuroimaging study. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00083

Bultena, S., Dijkstra, T., & Van Hell, J. G. (2015a). Language switch costs in sentence comprehension depend on language dominance : Evidence from self-paced reading. *Bilingualism: Language and Cognition*, *18*(3), 453-469. https://doi.org/10.1017/S1366728914000145

Bultena, S., Dijkstra, T., & Van Hell, J. G. (2015b). Switch cost modulations in bilingual sentence processing : Evidence from shadowing. *Language, Cognition and Neuroscience*, *30*(5), 586-605. https://doi.org/10.1080/23273798.2014.964268

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate Effects in Sentence Context Depend on Word Class, L2 Proficiency, and Task. *Quarterly Journal of Experimental Psychology*, *67*(6), 1214-1241. https://doi.org/10.1080/17470218.2013.853090

Carrasco-Ortiz, H., Midgley, K. J., & Frenck-Mestre, C. (2012). Are phonological representations in bilinguals language specific? An ERP study on interlingual homophones: Phonological representations in bilingual word recognition. *Psychophysiology*, *49*(4), 531-543. https://doi.org/10.1111/j.1469-8986.2011.01333.x

Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts : The impact of orthographic markedness. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00424

Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2015). How do bilinguals identify the language of the words they read? *Brain Research*, *1624*, 153-166. https://doi.org/10.1016/j.brainres.2015.07.035

Casaponsa, A., & Duñabeitia, J. A. (2016). Lexical organization of language-ambiguous and language-specific words in bilinguals. *The Quarterly Journal of Experimental Psychology*, *69*(3), 589-604. https://doi.org/10.1080/17470218.2015.1064977

Casaponsa, A., Thierry, G., & Duñabeitia, J. A. (2019a). The Role of Orthotactics in Language Switching : An ERP Investigation Using Masked Language Priming. *Brain Sciences*, *10*(1), 22. https://doi.org/10.3390/brainsci10010022

Casaponsa, A., Thierry, G., & Duñabeitia, J. A. (2019b). The Role of Orthotactics in Language Switching : An ERP Investigation Using Masked Language Priming. *Brain Sciences*, *10*(1), 22. https://doi.org/10.3390/brainsci10010022

Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00645

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103-120. https://doi.org/10.1016/j.cognition.2017.02.015

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. https://doi.org/10.1037/0033-295X.108.1.204

Commissaire, E. (2021). Do both WRAP and TRAP inhibit the recognition of the French word DRAP? Impact of orthographic markedness on cross-language orthographic priming. *Quarterly Journal of Experimental Psychology*.

Commissaire, E., Audusseau, J., & Casalis, S. (2019). Disentangling cross-language orthographic neighborhood from markedness effects in L2 visual word recognition. *Psychonomic Bulletin & Review*, *26*(1), 353-359. https://doi.org/10.3758/s13423-018-1518-6

Commissaire, E., & Demont, E. (2022). Investigating L2 reading aloud and silent reading in typically developing readers and dyslexic adolescents from grades 6 to 9. *Dyslexia*, *28*(1), 40-59. https://doi.org/10.1002/dys.1693

Commissaire, E., Duncan, L. G., & Casalis, S. (2014). Grapheme coding in L2 : How do L2 learners process new graphemes? *Journal of Cognitive Psychology*, *26*(7), 725-739. https://doi.org/10.1080/20445911.2014.951363

Commissaire, E., Duncan, L. G., & Casalis, S. (2019). Investigating pseudohomophone interference effects in young second-language learners. *Journal of Experimental Child Psychology*, *180*, 1-18. https://doi.org/10.1016/j.jecp.2018.11.010

Cornut, C. (2021). *L2 word recognition among late bilinguals : Does modality matter? : investigation of the modality effect in English second language among typical and dyslexic-readers* [These de doctorat, Université de Lille (2018-2021)]. https://www.theses.fr/2021LILUH039

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English : Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 668-687. https://doi.org/10.1037/0096-1523.32.3.668

De Moor, W., & Brysbaert, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research*, *63*(2), 159-162. https://doi.org/10.1007/PL00008174

De Moor, W., & Verguts, T. (2006). When are neighbours hostile? Inhibitory neighbour effects in visual word recognition. *Psychological Research*, *70*(5), 359-366. https://doi.org/10.1007/s00426-005-0225-x de Bruin, A. (2019). Not All Bilinguals Are the Same : A Call for More Detailed Assessments and Descriptions of Bilingual Experiences. *Behavioral Sciences*, *9*(3), 33. https://doi.org/10.3390/bs9030033

Declerck, M., Eben, C., & Grainger, J. (2019). A different perspective on domain-general language control using the flanker task. *Acta Psychologica*, *198*, 102884. https://doi.org/10.1016/j.actpsy.2019.102884

Declerck, M., & Grainger, J. (2017). Inducing asymmetrical switch costs in bilingual language comprehension by language practice. *Acta Psychologica*, *178*, 100-106. https://doi.org/10.1016/j.actpsy.2017.06.002

Declerck, M., Snell, J., & Grainger, J. (2018). On the role of language membership information during word recognition in bilinguals : Evidence from flanker-language congruency effects. *Psychonomic Bulletin & Review*, *25*(2), 704-709. https://doi.org/10.3758/s13423-017-1374-9

de León Rodríguez, D., Buetler, K. A., Eggenberger, N., Laganaro, M., Nyffeler, T., Annoni, J.-M., & Müri, R. M. (2016). The Impact of Language Opacity and Proficiency on Reading Strategies in Bilinguals : An Eye Movement Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00649

Dijkstra, T., Bruijn, E. D., Schriefers, H., & Brinke, S. T. (2000). More on interlingual homograph recognition : Language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition*, *3*(1), 69-78. https://doi.org/10.1017/S1366728900000146

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs : The Neglected Role of Phonology. *Journal of Memory and Language*, *41*(4), 496-518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., Hilberink-Schulpen, B., & Van Heuven, W. J. B. (2010). Repetition and masked form priming within and between languages using word and nonword neighbors. *Bilingualism: Language and Cognition*, *13*(3), 341-357. https://doi.org/10.1017/S1366728909990575

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284-301. https://doi.org/10.1016/j.jml.2009.12.003

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On Being Blinded by Your Other Language : Effects of Task Demands on Interlingual Homograph Recognition. *Journal of Memory and Language*, *42*(4), 445-464. https://doi.org/10.1006/jmla.1999.2697

Dijkstra, T., Van Hell, J. G., & Brenders, P. (2015). Sentence context effects in bilingual word recognition : Cognate status, sentence language, and semantic constraint. *Bilingualism: Language and Cognition*, *18*(4), 597-613. https://doi.org/10.1017/S1366728914000388

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system : From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175-197. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink : A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*(04), 657-679. https://doi.org/10.1017/S1366728918000287

Dirix, N., Cop, U., Drieghe, D., & Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *43*(6), 887-915. https://doi.org/10.1037/xlm0000352

Duñabeitia, J. A., Borragán, M., de Bruin, A., & Casaponsa, A. (2020a). Changes in the Sensitivity to Language-Specific Orthographic Patterns With Age. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01691

Duñabeitia, J. A., Borragán, M., de Bruin, A., & Casaponsa, A. (2020b). Changes in the Sensitivity to Language-Specific Orthographic Patterns With Age. *Frontiers in Psychology*, *11*, 1691. https://doi.org/10.3389/fpsyg.2020.01691

Duyck, W. (2005). Translation and Associative Priming With Cross-Lingual Pseudohomophones : Evidence for Nonselective Phonological Activation in Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1340-1359. https://doi.org/10.1037/0278-7393.31.6.1340

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context : Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663-679. https://doi.org/10.1037/0278-7393.33.4.663

Eben, C., & Declerck, M. (2019). Conflict monitoring in bilingual language comprehension? Evidence from a bilingual flanker task. *Language, Cognition and Neuroscience*, *34*(3), 320-325. https://doi.org/10.1080/23273798.2018.1537499

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2 : Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research*, *25*(1), 57-70. https://doi.org/10.1016/j.cogbrainres.2005.04.007

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. https://doi.org/10.1080/02724989208250619

Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society*, *31*(2), 119-122. https://doi.org/10.3758/BF03334157

Ferrand, L., & Grainger, J. (2003). Homophone interference effects in visual word recognition. *The Quarterly Journal of Experimental Psychology: Section A*, *56*(3), 403-419. https://doi.org/10.1080/02724980244000422

Forster, K. I., & Forster, J. C. (2003). DMDX : A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116-124. https://doi.org/10.3758/BF03195503

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461-1469. https://doi.org/10.1037/a0036756

Frost, R. (1994). Prelexical and postlexical strategies in reading : Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 116-129. https://doi.org/10.1037/0278-7393.20.1.116

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth : A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 104-115. https://doi.org/10.1037/0096-1523.13.1.104

Gade, M., Declerck, M., Philipp, A. M., Rey-Mermet, A., & Koch, I. (2021). Assessing the Evidence for Asymmetrical Switch Costs and Reversed Language Dominance Effects – A Meta-Analysis. *Journal of Cognition*, *4*(1), 55. https://doi.org/10.5334/joc.186

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency : Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, *19*(1), 19-52. https://doi.org/10.1017/S0142716400010560

Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology Section A*, *39*(2), 295-319. https://doi.org/10.1080/14640748708401788

Grainger, J., & Ferrand, L. (1994). Phonology and Orthography in Visual Word Recognition : Effects of Masked Homophone Primes. *Journal of Memory and Language*, *33*(2), 218-233. https://doi.org/10.1006/jmla.1994.1011

Grainger, J., & Ferrand, L. (1996). Masked Orthographic and Phonological Priming in Visual Word Recognition and Naming : Cross-Task Comparisons. *Journal of Memory and Language*, *35*(5), 623-647. https://doi.org/10.1006/jmla.1996.0033

Grainger, J., & Holcomb, P. J. (2009). Watching the Word Go by : On the Time-course of Component Processes in Visual Word Recognition: Time-course of Word Recognition. *Language and Linguistics Compass*, *3*(1), 128-156. https://doi.org/10.1111/j.1749-818X.2008.00121.x

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

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In *Language acquisition across linguistic and cognitive systems* (Vol. 52, p. 267-283).

Green, D. W. (1997). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67-81. https://doi.org/10.1017/S1366728998000133

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals : The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515-530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (2001). *The bilingual's language modes*. 27.

Grossi, G., Savill, N., Thomas, E., & Thierry, G. (2012). Electrophysiological Cross-Language Neighborhood Density Effects in Late and Early English-Welsh Bilinguals. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00408

Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When Language Switching has No Apparent Cost : Lexical Access in Sentence Context. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00278

Haigh, C. A., & Jared, D. (2007). The activation of phonological representations by bilinguals while reading silently : Evidence from interlingual homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 623-644. https://doi.org/10.1037/0278-7393.33.4.623

Hevia-Tuero, C., Incera, S., & Suárez-Coalla, P. (2021). Does English orthography influence bilingual Spanish readers? The effect of grapheme crosslinguistic congruency and complexity on letter detection. *Cognitive Development*, *59*, 101074. https://doi.org/10.1016/j.cogdev.2021.101074

Hevia-Tuero, C., Incera, S., & Suárez-Coalla, P. (2022). Influences of First and Second Language Phonology on Spanish Children Learning to Read in English. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/article/10.3389/fpsyg.2022.803518

Hoversten, L. J., Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Language Membership Identification Precedes Semantic Access : Suppression during Bilingual Word Recognition. *Journal of Cognitive Neuroscience*, *27*(11), 2108-2116. https://doi.org/10.1162/jocn_a_00844

Hoversten, L. J., Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Early processing of orthographic language membership information in bilingual visual word recognition : Evidence from ERPs. *Neuropsychologia*, *103*, 183-190. https://doi.org/10.1016/j.neuropsychologia.2017.07.026

Hoversten, L. J., & Traxler, M. J. (2020). Zooming in on zooming out : Partial selectivity and dynamic tuning of bilingual language control during reading. *Cognition*, *195*, 104118. https://doi.org/10.1016/j.cognition.2019.104118

Jared, D. (1997). Spelling–Sound Consistency Affects the Naming of High-Frequency Words. *Journal of Memory and Language*, *36*(4), 505-529. https://doi.org/10.1006/jmla.1997.2496

Jared, D., Cormier, P., Levy, B. A., & Wade-Woolley, L. (2012). Cross-language activation of phonology in young bilingual readers. *Reading and Writing*, *25*(6), 1327-1343. https://doi.org/10.1007/s11145-011-9320-0

Jared, D., Cormier, P., Levy, B. A., & Wade-Woolley, L. (2013). Discrimination of English and French orthographic patterns by biliterate children. *Journal of Experimental Child Psychology*, *114*(4), 469-488. https://doi.org/10.1016/j.jecp.2012.11.001

Jared, D., & Kroll, J. F. (2001). Do Bilinguals Activate Phonological Representations in One or Both of Their Languages When Naming Words? *Journal of Memory and Language*, *44*(1), 2-31. https://doi.org/10.1006/jmla.2000.2747

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. https://doi.org/10.1016/0749-596X(90)90044-Z

Jared, D., & Szucs, C. (2002a). Phonological activation in bilinguals : Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition*, *5*(3), 225-239. https://doi.org/10.1017/S1366728902003024

Jared, D., & Szucs, C. (2002b). Phonological activation in bilinguals : Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition*, *5*(3), 225-239. https://doi.org/10.1017/S1366728902003024

Katz, L., & Frost, R. (1992). Chapter 4 The Reading Process is Different for Different Orthographies : The Orthographic Depth Hypothesis. In *Advances in Psychology* (Vol. 94, p. 67-84). Elsevier. https://doi.org/10.1016/S0166-4115(08)62789-2

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project : Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287-304. https://doi.org/10.3758/s13428-011-0118-4

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming : Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, *33*(2), 149-174. https://doi.org/10.1006/jmla.1994.1008

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2015). Do semantic sentence constraint and L2 proficiency influence language selectivity of lexical access in native language listening? *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1524-1538. https://doi.org/10.1037/a0039782

Lallier, M., & Carreiras, M. (2018). Cross-linguistic transfer in bilinguals reading in two alphabetic orthographies : The grain size accommodation hypothesis. *Psychonomic Bulletin & Review*, *25*(1), 386-401. https://doi.org/10.3758/s13423-017-1273-0

Lallier, M., Carreiras, M., Tainturier, M.-J., Savill, N., & Thierry, G. (2013). Orthographic transparency modulates the grain size of orthographic processing : Behavioral and ERP evidence from bilingualism. *Brain Research*, *1505*, 47-60. https://doi.org/10.1016/j.brainres.2013.02.018

Lallier, M., Martin, C. D., Acha, J., & Carreiras, M. (2021). Cross-linguistic transfer in bilingual reading is item specific. *Bilingualism: Language and Cognition*, *24*(5), 891-901. https://doi.org/10.1017/S1366728921000183

Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts : A meta-analytic review. *Journal of Memory and Language*, *92*, 217-233. https://doi.org/10.1016/j.jml.2016.06.010

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE : A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325-343. https://doi.org/10.3758/s13428-011-0146-0

Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO : Cognate effects in trilingual word recognition. *Language and Cognitive Processes*, *19*(5), 585-611. https://doi.org/10.1080/01690960444000007

Lemhöfer, K., & Radach, R. (2008). Task Context Effects in Bilingual Nonword Processing. *Experimental Psychology*, *56*(1), 41-47. https://doi.org/10.1027/1618-3169.56.1.41

Lemhöfer, K., & Radach, R. (2009). Task Context Effects in Bilingual Nonword Processing. *Experimental Psychology*, *56*(1), 41-47. https://doi.org/10.1027/1618-3169.56.1.41

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context : Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381-390. https://doi.org/10.1037/a0014875

Litcofsky, K. A., & Van Hell, J. G. (2017). Switching direction affects switching costs : Behavioral, ERP and time-frequency analyses of intra-sentential codeswitching. *Neuropsychologia*, *97*, 112-139. https://doi.org/10.1016/j.neuropsychologia.2017.02.002

Macizo, P., Bajo, T., & Paolieri, D. (2012). Language switching and language competition. *Second Language Research*, *28*(2), 131-149. https://doi.org/10.1177/0267658311434893

Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals : The input switch. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 480-487. https://doi.org/10.1016/S0022-5371(71)80018-X

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND : Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLOS ONE*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574-588. https://doi.org/10.1016/j.jml.2013.08.001

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. https://doi.org/10.1037/0033-295X.88.5.375

Mcquade, D. V. (1981). Variable Reliance on Phonological Information in Visual Word Recognition. *Language and Speech*, *24*(1), 99-109. https://doi.org/10.1177/002383098102400107

Meuter, R. F. I., & Allport, A. (1999). Bilingual Language Switching in Naming : Asymmetrical Costs of Language Selection. *Journal of Memory and Language*, *40*(1), 25-40. https://doi.org/10.1006/jmla.1998.2602

Midgley, K. J., Holcomb, P. J., vanHeuven, W. J. B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research*, *1246*, 123-135. https://doi.org/10.1016/j.brainres.2008.09.078

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching Languages, Switching Palabras (Words) : An Electrophysiological Study of Code Switching. *Brain and Language*, *80*(2), 188-207. https://doi.org/10.1006/brln.2001.2588

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165-178. https://doi.org/10.1037/h0027366

Mulder, K., van Heuven, W. J. B., & Dijkstra, T. (2018). Revisiting the Neighborhood : How L2 Proficiency and Neighborhood Manipulation Affect Bilingual Processing. *Frontiers in Psychology*, *9*, 1860. https://doi.org/10.3389/fpsyg.2018.01860

Nas, G. (1983). Visual word recognition in bilinguals : Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of*

Verbal Learning and Verbal Behavior, *22*(5), 526-534. https://doi.org/10.1016/S0022-5371(83)90319-5

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE™//A lexical database for contemporary french : LEXIQUE™. *L'année psychologique*, *101*(3), 447-462. https://doi.org/10.3406/psy.2001.1341

Oganian, Y., Conrad, M., Aryani, A., Heekeren, H. R., & Spalek, K. (2016). Interplay of bigram frequency and orthographic neighborhood statistics in language membership decision. *Bilingualism: Language and Cognition*, *19*(3), 578-596. https://doi.org/10.1017/S1366728915000292

Oganian, Y., Conrad, M., Aryani, A., Spalek, K., & Heekeren, H. R. (2015). Activation Patterns throughout the Word Processing Network of L1-dominant Bilinguals Reflect Language Similarity and Language Decisions. *Journal of Cognitive Neuroscience*, *27*(11), 2197-2214. https://doi.org/10.1162/jocn_a_00853

Olson, D. J. (2016). The gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics*, *37*(3), 725-756. https://doi.org/10.1017/S0142716415000223

Orfanidou, E., & Sumner, P. (2005). Language switching and the effects of orthographic specificity and response repetition. *Memory & Cognition*, *33*(2), 355-369. https://doi.org/10.3758/BF03195323

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition : The word-superiority effect. *Psychological Review*, *89*, 573-594. https://doi.org/10.1037/0033-295X.89.5.573

Paulesu, E., Bonandrini, R., Zapparoli, L., Rupani, C., Mapelli, C., Tassini, F., Schenone, P., Bottini, G., Perry, C., & Zorzi, M. (2021). Effects of Orthographic Consistency on Bilingual Reading : Human and Computer Simulation Data. *Brain Sciences*, *11*(7), 878. https://doi.org/10.3390/brainsci11070878

Peereman, R., & Content, A. (1995). Neighborhood size effect in naming : Lexical activation or sublexical correspondences? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 409-421. https://doi.org/10.1037/0278-7393.21.2.409

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals : RT and ERP effects. *Journal of Memory and Language*, *68*(4), 315-332. https://doi.org/10.1016/j.jml.2012.12.003

Peeters, D., Vanlangendonck, F., Rueschemeyer, S.-A., & Dijkstra, T. (2019). Activation of the language control network in bilingual visual word recognition. *Cortex*, *111*, 63-73. https://doi.org/10.1016/j.cortex.2018.10.012

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification : Evidence from backward masking and priming. *Journal of Memory and Language*, *30*(4), 473-485. https://doi.org/10.1016/0749-596X(91)90017-E

Perfetti, C. A., & Dunlap, S. (2008). 2 Learning to read. In *Learning to Read Across Languages* (p. 13).

Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading : Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 787-796. https://doi.org/10.1037/a0035583

Poort, E. D., & Rodd, J. M. (2017). The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition. *Acta Psychologica*, *180*, 52-63. https://doi.org/10.1016/j.actpsy.2017.08.008

Poort, E. D., Warren, J. E., & Rodd, J. M. (2016). Recent experience with cognates and interlingual homographs in one language affects subsequent processing in another language. *Bilingualism: Language and Cognition*, *19*(1), 206-212. https://doi.org/10.1017/S1366728915000395

Protopapas, A. (2007). CheckVocal : A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859-862. https://doi.org/10.3758/BF03192979

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English : Are they real? Do they matter? *Cognitive Psychology*, *53*(2), 97-145. https://doi.org/10.1016/j.cogpsych.2006.01.002

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, *10*(6), 645-657. https://doi.org/10.1016/S0022-5371(71)80071-3

Sauval, K., Perre, L., Duncan, L. G., Marinus, E., & Casalis, S. (2017). Automatic phonological activation during visual word recognition in bilingual children : A cross-language masked priming study in grades 3 and 5. *Journal of Experimental Child Psychology*, *154*, 64-77. https://doi.org/10.1016/j.jecp.2016.10.003

Schmalz, X., Beyersmann, E., Cavalli, E., & Marinus, E. (2016). Unpredictability and complexity of print-to-speech correspondences increase reliance on lexical processes : More evidence for the orthographic depth hypothesis. *Journal of Cognitive Psychology*, *28*(6), 658-672. https://doi.org/10.1080/20445911.2016.1182172

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review*, *22*(6), 1614-1629. https://doi.org/10.3758/s13423-015-0835-2

Schmalz, X., Robidoux, S., Castles, A., Coltheart, M., & Marinus, E. (2017). German and English Bodies : No Evidence for Cross-Linguistic Differences in Preferred Orthographic Grain Size. *Collabra: Psychology*, *3*(1), 5. https://doi.org/10.1525/collabra.72

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197-212. https://doi.org/10.1016/j.jml.2006.03.004

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English : Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, *22*(1), 106-129. https://doi.org/10.1080/01690960500463920

Segui, J., & Grainger, J. (1990). Priming Word Recognition With OrthographicNeighbors : Effects of Relative Prime-Target Frequency. *Journal of Experimental Psychology: Human Perception and Performanc*, *16*(1), 65-76.

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, *30*(6), 627-643. https://doi.org/10.1016/0749-596X(91)90029-J

Snell, J., Meeter, M., & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words during reading. *PLOS ONE*, *12*(3), e0173720. https://doi.org/10.1371/journal.pone.0173720

Starreveld, P. A., De Groot, A. M. B., Rossmark, B. M. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals : Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, *17*(2), 258-276. https://doi.org/10.1017/S1366728913000308

Thomas, M. S. C., & Allport, A. (2000). Language Switching Costs in Bilingual Visual Word Recognition. *Journal of Memory and Language*, *43*(1), 44-66. https://doi.org/10.1006/jmla.1999.2700

Timmer, K., Christoffels, I. K., & Costa, A. (2019). On the flexibility of bilingual language control : The effect of language context. *Bilingualism: Language and Cognition*, *22*(3), 555-568. https://doi.org/10.1017/S1366728918000329

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading : The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1412-1431. https://doi.org/10.1037/a0024492

Vaid, J., & Frenck-Mestre, C. (2002). Do orthographic cues aid language recognition? A laterality study with French–English bilinguals. *Brain and Language*, *82*(1), 47-53. https://doi.org/10.1016/S0093-934X(02)00008-1

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*(1), 88-107. https://doi.org/10.1016/j.jml.2010.08.006

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does Bilingualism Change Native-Language Reading? : Cognate Effects in a Sentence Context.

Psychological Science, *20*(8), 923-927. https://doi.org/10.1111/j.1467-9280.2009.02389.x

Van Der Meij, M., Cuetos, F., Carreiras, M., & Barber, H. A. (2011). Electrophysiological correlates of language switching in second language learners : Electrophysiological correlates of language switching. *Psychophysiology*, *48*(1), 44-54. https://doi.org/10.1111/j.1469-8986.2010.01039.x

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals : Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 616-627. https://doi.org/10.1037/0096-1523.28.3.616

van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*(3), 431-451. https://doi.org/10.1016/j.actpsy.2008.03.010

van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780-789. https://doi.org/10.3758/BF03196335

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998a). Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language*, *39*(3), 458-483. https://doi.org/10.1006/jmla.1998.2584

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998b). Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language*, *39*(3), 458-483. https://doi.org/10.1006/jmla.1998.2584 van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language Conflict in the Bilingual Brain. *Cerebral Cortex*, *18*(11), 2706-2716. https://doi.org/10.1093/cercor/bhn030

van Kesteren, R., Dijkstra, T., & Smedt, K. de. (2012). Markedness effects in Norwegian–English bilinguals : Task-dependent use of language-specific letters and bigrams. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2129-2154. https://doi.org/10.1080/17470218.2012.679946

Vanlangendonck, F., Peeters, D., Rueschemeyer, S.-A., & Dijkstra, T. (2019). Mixing the stimulus list in bilingual lexical decision turns cognate facilitation effects into mirrored inhibition effects. *Bilingualism: Language and Cognition*, 1-9. https://doi.org/10.1017/S1366728919000531

von Studnitz, R. E., & Green, D. W. (1997). Lexical Decision and Language Switching. *International Journal of Bilingualism*, *1*(1), 3-24. https://doi.org/10.1177/136700699700100102

von Studnitz, R. E., & Green, D. W. (2002). The cost of switching language in a semantic categorization task. *Bilingualism: Language and Cognition*, *5*(3), 241-251. https://doi.org/10.1017/S1366728902003036

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. https://doi.org/10.1037/0033-2909.131.1.3

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision : Still a challenge for current word recognition models. *Journal of Experimental*

Psychology: Human Perception and Performance, *27*, 547-559. https://doi.org/10.1037/0096-1523.27.3.547

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical Words are Read Differently in Different Languages. *Psychological Science*, *12*(5), 379-384. https://doi.org/10.1111/1467-9280.00370

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). *What is the pronunciation for -ough and the spelling for lui ? A database for computing feedforward and feedback consistency in English*. 19.

Zirnstein, M., van Hell, J. G., & Kroll, J. F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, *176*, 87-106. https://doi.org/10.1016/j.cognition.2018.03.001

APPENDICES

Appendix 1.A – Chapter III, Experiment 1a – Material used in experiment 1A

Marked Words: Clean – Mean – Beach – Pool – Cook – Draw – Took – Feet – Week – Sleep – School – Show – Grow – Look – Town – Door – Floor – Saw – Teach – Book – Away – Awful – Between – Fifteen – Ready – Follow – Yellow – Easy – Shower – Window – Heavy – Agree – Career – Freedom – Nineteen – Allow – Teacher – Below – Flower – Power.

Non-Marked Words: Mouse – Proud – Sound – Found – Voice – Choice – Hand – Land – Pink – Drink – House – Spend – Lunch – Since – Fun – Ground – Hair – Rain – Link – Jump – Clever – Older – Again – Given – Golden – Iron – Picture – Broken – Pencil – Forget – Outfit – Outside – Daily – Behind – Inside – Open – Seven – Later – Sister – Corner.

Marked Pseudowords: Apree – Awab – Awdul – Azlow – Baneer – Bepow – Ceach – Clower – Daw – Dook – Dool – Eapy – Feacher – Feek – Flean – Fook – Fower – Fraw – Freelom – Meach – Meady – Metween – Noor – Peet – Pook – Pown – Scrool – Skower – Sloor – Smow – Sneep – Teavy – Tellow – Tifteen – Tindow – Tollow – Trow – Vean – Vook – Wolden.

Non-Marked Pseudowords: Anain – Bezind – Bink – Cloice – Dain – Dand – Dound – Droud – Fouse – Frink – Froken – Hink – Inbide – Inon – Laily – Naneteen – Niven – Noice – Nump – Oken – Outnide – Outnit – Peven – Plever – Rair – Rencil – Rister – Sater – Sicture – Slend – Tand – Tince – Torget – Torner – Tound – Touse - Tround – Tunch – Ulder – Vun.

Appendix 2.B – Chapter III, Experiment 1a – Pairing Parameters for the words used in Experiment 1A

Word items			
Marked vs. Non-marked items			
Measure	**M**	**NM**	***p***
Lexical frequency in zipf (1)	5(0.42)	5(0.39)	.372
Distance of Levenshtein (1)	3(0.50)	3(0.41)	.057
Orthographic neighbourood frequency in English (2)	66(102)	86(221)	.605
Orthographic neighbourood frequency in French (2)	64(288)	74(410)	.899
Mean Bigram Frequency (1)	10025(3788)	11978(5848)	.080
RT for monolinguals (4)	529(32)	534 (24)	.405
Orthographic Markedness Calculation			
Measure	**In French**	**In English**	***p***
Minimum Bigram Frequency for Non-Marked words (3)	843(805)	860(571)	.911
Minimum Bigram Frequency for Marked words (3)	91(144)	682(497)	<.001
	<.001	.141	

1: APPREL2, 2: Clearpond database, 3: Celex database, 4: BLP database

Appendix 3.A – Chapter III, Experiment 1a - Model comparison on word data in experiment 1A

On reaction times					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-59894			
Model 2	L2 proficiency	-59900	7.92	1,7	.004
Model 3	Orthographic markedness	-59898	0.57	1,8	.451
Model 4	L2 proficiency * Orthographic markedness	-59898	1.46	1,9	.227
On accuracy					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	936			
Model 2	L2 proficiency	934	3.74	1,4	.053
Model 3	Orthographic markedness	931	5.16	1,5	.023
Model 4	L2 proficiency * Orthographic markedness	933	0.02	1,6	.555

Appendix 4.B – Chapter III, Experiment 1a - Model comparison on pseudoword data in experiment 1A

On reaction times					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-58761			
Model 2	L2 proficiency	-58771	12.12	1,7	<.001
Model 3	Orthographic markedness	-58771	1.31	1,8	.251
Model 4	L2 proficiency * Orthographic markedness	-58769	0.16	1,9	.692
On accuracy					
		AIC	**χ²**	**ddl**	**P**
Model 1	Random	1631			
Model 2	L2 proficiency	1620	12.80	1,4	.001
Model 3	Orthographic markedness	1619	2.89	1,5	.089
Model 4	L2 proficiency * Orthographic markedness	1621	0.25	1,6	.614

Appendix 5.A – Chapter III, Experiment 1b – Material used in experiment 1B

Congruent Words: Biggest – Body – Box – Bridge – Carry – Catch – Cold – Dark – Drop – Every – Expect – Extract – Fifty – Fish – Flag – Focus – Gold – Grid – Kid – Lovely – Map – Messy – Milk – Mom – Most – Past – Perhaps – Practice – Pretty – Sad – Selfish – Sixty – Spirit – Story – Task – Tip – Topic – Travel – Trip - Very

Incongruent Words: Advice – Almost – Alone – Baby – Became – Bike – Busy – Careful – Child – Classmate – Crazy – Cup – Curly – Cut – Dirty – Drive - Five – Game – Girl – Ice – Lady – Lake - Late – Movie – Much – Myself – Nine – Same – Save – Scary – Side – State – Study – Such – Surname – Talk – Tidy – Turn – Ugly – Useful

Congruent Pseudowords: Crip – Dap – Dast – Datch – Dody – Dox – Dretty – Eltract – Extect – Filk – Fom – Fost – Gid – Gridge – Grop – Klag – Liggest – Mip – Nad – Nery – Nold – Overy – Pervaps – Pessy – Pravel – Rask – Rixty – Rold – Ropic – Sarry – Selpish – Sish – Skirit – Smory – Tark – Tifty – Tocus – Tovely – Tractice – Trid.

Incongruent Pseudowords: Apmost – Apone - Admice – Agly – Bepame – Caleful – Chird – Cusy – Dake – Dirl – Dut – Fide – Fidy – Fuch – Furly – Furn – Gady – Ike – Luch – Mame – Mirty – Mive – Mybelf – Naby – Palk – Rike – Rine – Rovie – Rup – Slassmate – Smary – Smate – Spudy – Surnime – Tate – Tave – Trazy – Trive – Uleful – Vame.

Appendix 6.B – Chapter III, Experiment 1b – Pairing Parameters for the words used in Experiment 1B

Word items			
Congruent vs. Icongruent items			
Measure	**Congruent**	**Incongruent**	***p***
Lexical frequency in zipf (1)	5(0.37)	5(0.32)	.926
Distance of Levenshtein (1)	3(0.43)	3(0.43)	.589
Orthographic neighbourood frequency in English (2)	73(134)	90(128)	.559
Orthographic neighbourood frequency in French (2)	328(1482)	66(328)	.281
Total Bigram Frequency in English (3)	6772(3767)	6468(4167)	.733
Mean Bigram Frequency (1)	7633(3014)	6801(3430)	.236
RT for monolinguals (4)	549(35)	543 (31)	.383

1: APPREL2, 2: Clearpond database, 3: Celex database, 4: BLP database

Appendix 7.A – Chapter III, Experiment 1b - Model comparison on word data in experiment 1B

On reaction times					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	-59710			
Model 2	L2 proficiency	-59715	7.26	1,7	.007
Model 3	GPC Congruency	-59715	1.49	1,8	.222
Model 4	L2 proficiency * GPC Congruency	-59713	0.29	1,9	.592
On accuracy					
		AIC	**χ^2**	**ddl**	**P**
Model 1	Random	1003			
Model 2	L2 proficiency	991	13.37	1,4	<.001
Model 3	GPC Congruency	993	0.22	1,5	.636
Model 4	L2 proficiency * GPC Congruency	992	2.73	1,6	.098

Appendix 8.B – Chapter III, Experiment 1b - Model comparison on pseudoword data in experiment 1B

On reaction times					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	-59681			
Model 2	L2 proficiency	-59681	2.01	1,5	.157
Model 3	GPC Congruency	-59679	0.25	1,6	.618
Model 4	L2 proficiency * GPC Congruency	-59679	1.78	1,7	.182
On accuracy					
		AIC	**χ^2**	**ddl**	**P**
Model 1	Random	1794			
Model 2	L2 proficiency	1793	3.63	1,4	.057
Model 3	GPC Congruency	1795	0.18	1,5	.673
Model 4	L2 proficiency * GPC Congruency	1797	0.09	1,6	.766

Appendix 9.A – Chapter III, Experiment 1c – Material used in experiment 1C

Consistent and Marked Words : Boat – Coat – Dream – Keep – Moon – Scream – Sheep – Soon – Street – Sweet. **Consistent and Non-Marked Words :** Bed – Blue – Care – Cake – Dare – Make – Red – Run – Sun - True

Inconsistent and Marked Words : Bear – Bread – Break – Dear – Down – Head – Near – Own – Read – Speak. **Inconsistent and Non-Marked Words :** Cost – Find – Kind – Lose – Lost – Love – Mind – Move – Nose – Wind.

Consistent and Marked Pseudowords : Doon – Feep – Pream – Roat – Roon – Sneet – Soat – Speep – Spream – Spreet. **Consistent and Non-Marked Pseudowords :** Gake – Lare – Mun – Pake – Ped – Prue – Sare – Sed – Slue – Wun.

Inconsistent and Marked Pseudowords : Awn – Cear – Cread – Mear – Pead – Preak – Rown – Sead – Sheak – Vear. **Inconsistent and Non-Marked Pseudowords :** Bost – Bove – Fose – Lind – Mose – Nind – Pind – Rost – Tind – Tove.

Appendix 10.B – Chapter III, Experiment 1c – Pairing Parameters for the words used in Experiment 1C

Word items			
Consistent vs. Inconsistent items			
Measure	**Consistent**	**Inconsistent**	***p***
Lexical frequency in zipf (1)	5(0.40)	5(0.69)	.752
Distance of Levenshtein (1)	3(0.13)	3(0.38)	.198
Orthographic neighbourood frequency in English (2)	91(108)	178(255)	.168
Orthographic neighbourood frequency in French (2)	473(1045)	458(1915)	.976
Total Bigram Frequency in English (3)	7149(2984)	6347(3082)	.204
Minimum Bigram Frequency (1)	4605(4271)	4632(2547)	.490
RT for monolinguals (4)	524 (20)	540 (26)	.032
Orthographic Markedness Calculation			
Measure	**In French**	**In English**	***p***
Minimum Bigram Frequency for Non-Marked words (3)	759(705)	959(567)	.403
Minimum Bigram Frequency for Marked words (3)	208(195)	913(604)	.001
	.001	.403	

1: APPREL2, 2: Clearpond database, 3: Celex database, 4: BLP database

Appendix 11.A – Chapter III, Experiment 1c - Model comparison on word data in experiment 1C

On reaction times					
		AIC	χ^2	ddl	p
Model 1	Random	-29945			
Model 2	L2 proficiency	-29951	8.12	1,5	.004
Model 3	Orthographic markedness	-29954	5.15	1,6	.023
Model 4	GPC Consistency	-29955	3.42	1,7	.064
Model 5	L2 proficiency * Orthographic markedness	-29954	0.10	1,8	.755
Model 6	L2 proficiency * GPC Consistency	-29952	0.33	1,9	.564
Model 7	Orthographic markedness * GPC Consistency	-29950	0.31	1,10	.580
Model 8	Full Interaction	-29949	0.91	1,11	.341
On accuracy					
		AIC	χ^2	ddl	p
Model 1	Random	438			
Model 2	L2 proficiency	438	1.62	1,4	.203
Model 3	Orthographic markedness	440	0.72	1,5	.396
Model 4	GPC Consistency	442	0.03	1,6	.871
Model 5	Orthographic markedness * GPC Consistency	444	0.003	1,7	.959

Appendix 12.B – Chapter III, Experiment 1c - Model comparison on word data in experiment 1C by consistency subplot

On Consistent words					
		AIC	χ^2	ddl	p
Model 1	Random	-15021			
Model 2	L2 proficiency	-15027	7.60	1,5	.006
Model 3	Orthographic markedness	-15026	1.45	1,6	.228
Model 4	L2 proficiency * Orthographic markedness	-15025	0.88	1,7	.349
On Inconsistent words					
		AIC	χ^2	ddl	P
Model 1	Random	-14817			
Model 2	L2 proficiency	-14823	7.98	1,4	.004
Model 3	Orthographic markedness	-14825	3.97	1,5	.046
Model 4	L2 proficiency * Orthographic markedness	-14823	0.18	1,7	.674

Appendix 13.C – Chapter III, Experiment 1c - Model comparison on pseudoword data in experiment 1C

On reaction times					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	-29974			
Model 2	L2 proficiency	-29975	3,02	1,5	.082
Model 3	Orthographic Markedness	-29976	2,25	1,6	.134
Model 4	GPC Consistency	-29975	1,00	1,7	.317
Model 5	L2 proficiency * Orthographic Markedness	-29973	0,48	1,8	.489
Model 6	L2 proficiency * GPC Consistency	-29972	0,56	1,9	.455
Model 7	Orthographic Markedness * GPC Consistency	-29970	0,19	1,10	.662
Model 8	Full Interaction	-29970	2,12	1,11	.146
On accuracy					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	1486			
Model 2	L2 proficiency	1480	7,42	1,4	.006
Model 3	Orthographic Markedness	1481	0,81	1,5	.368
Model 4	GPC Consistency	1481	2,07	1,6	.150
Model 5	Orthographic Markedness * GPC Consistency	1483	0,69	1,7	.407

Appendix 14 – Chapter III, Experiment 2a - Model comparison on word data in experiment 2A

On reaction times					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-72333			
Model 2	L2 proficiency	-72336	4.76	1,5	.029
Model 3	Orthographic Markedness	-72335	1.45	1,6	.229
Model 4	L2 proficiency * Orthographic Markedness	-72334	0.59	1,7	.442
On accuracy as rated by an English-speaker					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	4639			
Model 2	L2 proficiency	4529	112.59	1,4	<.001
Model 3	Orthographic Markedness	4530	0.15	1,5	.699
Model 4	L2 proficiency * Orthographic Markedness	4532	0.51	1,6	.821
On accuracy as rated by a French-speaker					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1700			
Model 2	L2 proficiency	1693	9.23	1,4	.002
Model 3	Orthographic Markedness	1686	8.37	1,5	.004
Model 4	L2 proficiency * Orthographic Markedness	1688	0.30	1,6	.586

Appendix 15 – Chapter III, Experiment 2b - Model comparison on word data in experiment 2B

On reaction times					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-71177			
Model 2	L2 proficiency	-71181	5.94	1,5	.015
Model 3	GPC Congruency	-71181	1.98	1,6	.159
Model 4	L2 proficiency * GPC Congruency	-71182	3.57	1,7	.059
On accuracy as rated by an English-speaker					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	3869			
Model 2	L2 proficiency	3864	7.02	1,4	.008
Model 3	GPC Congruency	3844	21.12	1,5	<.001
Model 4	L2 proficiency * GPC Congruency	3847	0.04	1,6	.846
On accuracy as rated by a French-speaker					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	1468			
Model 2	L2 proficiency	1462	8.12	1,4	.004
Model 3	GPC Congruency	1464	0.08	1,5	.783
Model 4	L2 proficiency * GPC Congruency	1462	4.13	1,6	.042

Appendix 16 – Chapter III, Experiment 2c - Model comparison on word data in experiment 2C

On reaction times					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	-35027			
Model 2	L2 proficiency	-35029	3.87	1,5	.049
Model 3	Orthographic Markedness	-35027	0.15	1,6	.699
Model 4	GPC Consistency	-35032	6.82	1,7	.009
Model 5	L2 proficiency * Orthographic Markedness	-35032	1.95	1,8	.163
Model 6	L2 proficiency * GPC Consistency	-35033	2.82	1,9	.093
Model 7	Orthographic Markedness * GPC Consistency	-35032	1.13	1,10	.289
Model 8	Full Interaction	-35032	2.21	1,11	.137
On accuracy as rated by an English-speaker					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	1460			
Model 2	L2 proficiency	1458	3.53	1,4	.060
Model 3	Orthographic Markedness	1457	3.15	1,5	.076
Model 4	GPC Consistency	1451	8.24	1,6	.004
Model 5	L2 proficiency * Orthographic Markedness	1453	0.01	1,7	.906
Model 6	L2 proficiency * GPC Consistency	1452	3.01	1,8	.083
Model 7	Orthographic Markedness * GPC Consistency	1454	0.02	1,9	.884
Model 8	Full Interaction	1456	0.24	1,10	.627
On accuracy as rated by a French-speaker					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	1460			
Model 2	L2 proficiency	774	687.64	1,4	<.001
Model 3	Orthographic Markedness	775	1.23	1,5	.268
Model 4	GPC Consistency	763	13.86	1,6	<.001
Model 5	L2 proficiency * Orthographic Markedness	764	1.31	1,7	.253
Model 6	L2 proficiency * GPC Consistency	765	0.38	1,8	.540
Model 7	Orthographic Markedness * GPC Consistency	767	0.14	1,9	.704
Model 8	Full Interaction	766	3.29	1,10	.070

Appendix 17.A – Chapter IV, Experiment 1 and 2 – Material used in experiment 1 and 2 - Words

L1 – Marked Target Words: Bien – Broche – Chien – Ciel – Cuir – Douche – Feux – Fier – Fleur – Fleuve – Gauche – Gras – Gros – Hier – Ivre – Jeune – Jeux – Leur – Livre – Meuble – Miel – Mien – Moche – Mouche – Niche – Peau – Peur – Poil – Poire – Preuve – Proche – Rien – Seul – Soif – Soir – Sous – Toile – Vache – Vivre – Vrai. **L1 – Non-Marked Target Words:** Bain – Barbe – Bise – Bord – Boucle – Bout – Bruit – Chou – Cible – Clou – Dort – Drap – Flot – Huit – Jaune – Jour – Jupe – Lune – Muet – Neige – Nord – Nuit – Plume – Pomme – Pont – Pouce – Rouge – Ruse – Sans – Sept – Sourd – Tard – Tige – Tort – Trop – Ventre – Vert – Vite – Vive – Voile.

L2 – Marked Target Words: Beach – Book – Broom – Clear – Coat – Cook – Deaf – Deal – Dear – Door – Down – Dream – Fear – Feel – Feet – Food – Free – Green – Head – Hear – Knife – Light – Meal – Mean – Moon – Near – Need – Night – School – Scream – Sheep – Show – Snow – Soon – Speak – Speed – Street – Team – Town – Wear. **L2 – Non-Marked Target Words:** Bird – Blade – Blind – Brain – Burn – Child – Cute – Fast – Fate – Found – Five – Frog – Game – Girl – Give – Gold – Hair – Have – Home – Hope – Horse – House – Loud – Mind – Most – Nose – Nurse – Proud – Purse – Rope – Save – Side – Smart – Smile – Soft – Stone – Storm – Taste – Trust – Turn.

L1 – Marked Flanker Words: Bouche – Bras – Cache – Deux – Heure – Lien – Noir – Noix – Nous – Poche – Seau – Sien – Sont – Triste – Trois – Trouve – Boire – Gris – Neuf – Pneu. **L1 – Non-Marked Flanker Words:** Bouge – Dans – Douce – Dure – Fait – Fille – Flou – Foin – Foire – Fraise – Mise – Mort – Plein – Sucre – Tout – Vole – Lourd – Perd – Pire – Vide.

L2 – Marked Flanker Words: Bean – Blood – Boat – Bread – Deep – Draw – Feed – Fight – Lead – Leaf – Meat – Read – Right – Seat – Seen – Sleep – Slow – Sweet – Twin – Twist. **L2 – Non-Marked Flanker Words:** Dice – Drop – Dust – Fail – Frame – Glad – Hand – Hide – Jump – Late – Must – Paste – Price – Pride – Same – Sand – Scar – Told – Trade – Paid.

Appendix 18.B – Chapter IV, Experiment 1 – Pairing Parameters for the words used in Experiment 1

	English words			French words		
Measure	**M**	**NM**	***p***	**M**	**NM**	***p***
Lexical frequency in zipf	5(0.60)	5(0.54)	.127	5(0.77)	5(0.72)	.495
Length	4(0.59)	4(0.48)	.611	5(0.78)	4.40(0.59)	.089
Summed Bigram Frequency	6678(2475)	6674(3150)	.993	9783(4105)	8647(4089)	.132

	Both sets		
Measure	**French**	**English**	***p***
Lexical frequency in zipf	5(0.74)	5(0.58)	.336
Length	5(0.69)	4(0.53)	.098
English Summed Bigram Frequency	6627(2632)	6676(2821)	.889
French Summed Bigram Frequency	9215(4120)	6651(3259)	<.001

Appendix 19 – Chapter IV, Experiment 1 -Long Display Duration- Model comparison- Do we replicate the trial congruency effect?

On reaction times					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-79824			
Model 2	L2 proficiency	-79822	0.30	1,5	.584
Model 3	Target Language	-79828	7.28	1,6	.007
Model 4	Flanker Language	-79827	1.86	1,7	.172
Model 5	L2 proficiency * Target Language	-79862	36.38	1,8	<.001
Model 6	L2 proficiency * Flanker Language	-79860	0.39	1,9	.535
Model 7	Target Language * Flanker Language	-79887	28.66	1,10	<.001
Model 8	Full interaction	-79887	2.04	1,11	.153
On accuracy					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	2099			
Model 2	Target Language	2089	11.20	1,4	<.001
Model 3	Flanker Language	2985	6.22	1,5	.013
Model 4	Target Language * Flanker Language	2071	15.96	1,6	<.001

Appendix 20 – Chapter IV, Experiment 1 -Long Display Duration- Model comparison- Does Target Markedness impact language decisions in a linguistic context?

On L1 target subplot (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-39255			
Model 2	L2 Proficiency	-39255	1.99	1,5	.158
Model 3	Trial Congruency	-39262	8.85	1,6	.003
Model 4	Target Markedness	-39263	3.11	1,7	.078
Model 5	L2 proficiency * Trial Congruency	-39263	2.21	1,8	.137
Model 6	L2 proficiency * Target Markedness	-39263	1.56	1,9	.211
Model 7	Trial Congruency * Target Markedness	-39261	0.41	1,10	.522
Model 8	Full interaction	-39259	0.14	1,11	.707
On L1 target subplot (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1277			
Model 2	Trial Congruency	1259	19.31	1,4	<.001
Model 3	Target Markedness	1255	6.09	1,5	.014
Model 4	Trial Congruency * Target Markedness	1257	0.44	1,6	.507
On L2 target subplot (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-40577			
Model 2	L2 Proficiency	-40575	0.09	1,5	.762
Model 3	Trial Congruency	-40596	22.99	1,6	<.001
Model 4	Target Markedness	-40625	30.37	1,7	<.001
Model 5	L2 proficiency * Trial Congruency	-40623	0.36	1,8	.550
Model 6	L2 proficiency * Target Markedness	-40621	0.14	1,9	.712
Model 7	Trial Congruency * Target Markedness	-40619	0.01	1,10	.917
Model 8	Full interaction	-40617	0.08	1,11	.772
On L2 target subplot (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	827			
Model 2	Trial Congruency	827	2.64	1,5	.104
Model 3	Target Markedness	815	13.46	1,6	<.001
Model 4	Trial Congruency * Target Markedness	815	1.59	1,7	.207

Appendix 21 – Chapter IV, Experiment 1 -Long Display Duration- Model comparison- Does Flanker Markedness impact language identification of a target word?

On Marked Flankers subplot (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-39727			
Model 2	Trial Congruency	-39753	28,53	1,5	<.001
Model 3	Flanker Language	-39751	0,40	1,6	.527
Model 4	Trial Congruency*Flanker Language	-39759	9,06	1,7	.003
On Marked Flankers subplot (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1047			
Model 2	Trial Congruency	1034	14,40	1,5	<.001
Model 3	Flanker Language	1034	2,42	1,6	.120
Model 4	Trial Congruency*Flanker Language	1022	13,58	1,7	<.001
On Non-Marked Flankers subplot (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-39937			
Model 2	Trial Congruency	-39939	4,11	1,5	.043
Model 3	Flanker Language	-39939	1,52	1,6	.217
Model 4	Trial Congruency*Flanker Language	-39941	3,64	1,7	.057
On Non-Marked Flankers subplot (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1082			
Model 2	Trial Congruency	1078	6,53	1,5	.011
Model 3	Flanker Language	1079	1,33	1,6	.248
Model 4	Trial Congruency*Flanker Language	1079	1,26	1,7	.262

Appendix 22 – Chapter IV, Experiment 1 - Short Display Duration- Model comparison- Do we replicate the trial congruency effect?

On reaction times					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-78241			
Model 2	L2 proficiency	-78242	2.81	1,5	.094
Model 3	Target Language	-78242	2.80	1,6	.094
Model 4	Flanker Language	-78243	2.96	1,7	.085
Model 5	L2 proficiency * Target Language	-78242	0.56	1,8	.453
Model 6	Lextale * Flanker Language	-78240	0.05	1,9	.824
Model 7	Target Language * Flanker Language	-78247	8.79	1,10	.003
Model 8	Full interaction	-78245	0.56	1,11	.454
On accuracy					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	4290			
Model 2	Target Language	4283	8.87	1,4	.002
Model 3	Flanker Language	4281	4.70	1,5	.030
Model 4	Target Language * Flanker Language	4098	184.32	1,6	<.001

Appendix 23 – Chapter IV, Experiment 1 -Short Display Duration- Model comparison- Does Target Markedness impact language decisions in a linguistic context?

		AIC	χ²	ddl	p
On L1 targets subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-37872			
Model 2	L2 Proficiency	-37872	2.13	1,5	.144
Model 3	Trial Congruency	-37871	1.30	1,6	.254
Model 4	Target Markedness	-37870	0.58	1,7	.445
Model 5	L2 proficiency * Trial Congruency	-37869	0.79	1,8	.374
Model 6	L2 proficiency* Target Markedness	-37870	3.68	1,9	.055
Model 7	Trial Congruency* Target Markedness	-37869	0.67	1,10	.412
Model 8	Full interaction	-37867	0.02	1,11	.883
On L1 targets subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	2330			
Model 2	Trial Congruency	2216	116.61	1,4	.001
Model 3	Target Markedness	2215	2.50	1,5	.113
Model 4	Trial Congruency* Target Markedness	2216	0.90	1,6	.341
On L2 targets subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-40322			
Model 2	L2 Proficiency	-40323	3.22	1,5	.073
Model 3	Trial Congruency	-40331	10.37	1,6	.001
Model 4	Target Markedness	-40350	20.87	1,7	<.001
Model 5	L2 proficiency * Trial Congruency	-40348	0.05	1,8	.827
Model 6	L2 proficiency* Target Markedness	-40353	6.82	1,9	.009
Model 7	Trial Congruency* Target Markedness	-40352	1.23	1,10	.267
Model 8	Full interaction	-40351	0.44	1,11	.509
On L2 targets subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	2011			
Model 2	Trial Congruency	1941	72.10	1,4	.001
Model 3	Target Markedness	1933	9.85	1,5	.001
Model 4	Trial Congruency* Target Markedness	1932	3.24	1,6	.072

Appendix 24 – Chapter IV, Experiment 1 -Short Display Duration- Model comparison- Does Flanker Markedness impact language identification of a target item?

On Marked Flankers subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-38889			
Model 2	Trial Congruency	-38896	9,54	1,5	.002
Model 3	Flanker Language	-38895	0,78	1,6	.378
Model 4	Trial Congruency*Flanker Language	-38896	3,01	1,7	.083
On Marked Flankers subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	2218			
Model 2	Trial Congruency	2085	135,14	1,4	<.001
Model 3	Flanker Language	2078	0,78	1,5	.001
Model 4	Trial Congruency*Flanker Language	2079	3,01	1,6	.723
On Non-Marked Flankers subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-39178			
Model 2	Trial Congruency	-39177	1,37	1,5	.242
Model 3	Flanker Language	-39178	2,43	1,6	.119
Model 4	Trial Congruency*Flanker Language	-39177	1,22	1,7	.270
On Non-Marked Flankers subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	2169			
Model 2	Trial Congruency	2109	61,92	1,4	.001
Model 3	Flanker Language	2107	3,82	1,5	.051
Model 4	Trial Congruency*Flanker Language	2109	0,28	1,6	.599

Appendix 25 – Chapter IV, Experiment 1 and 2 – Material used in experiment 2 - Pseudowords

L1 – Marked Target Pseudowords: Bruche – Buir – Cleur – Crien – Dauche – Dien – Dier – Dros – Feur – Fien – Fleune – Gouche – Groche – Jeuse – Luvre – Meuche – Miep – Mier – Neur – Noche – Noil – Ovre – Peuble – Piel – Pien – Poine – Pras – Preule – Reau – Reux – Roir – Saul – Souf – Teux – Toif – Toine – Vavre – Voche – Vroi – Vuche. **L1 – Non-Marked Target Pseudowords:** Beucle – Blot – Brop – Buin – Burbe – Clon – Crou – Dord – Fuet – Gord – Joune – Louge – Mert – Mourd – Muit – Mune – Nans – Nard – Nour – Nuige – Pimme – Pout – Prap – Pruce – Pruit – Prume – Punt – Pupe – Rort – Ruit – Rupe – Tept – Tible – Tise – Tite – Vife – Vige – Voine – Vort – Vuntre.

L2 – Marked Target Pseudowords: Cear – Coot – Crear – Cree – Deap – Dight – Feal – Feam – Gight – Hean – Knipe – Leel – Lown – Meap – Mear – Moop – Neap – Neeb – Nood – Noor – Peet – Pook – Poon – Pream – Preen – Proom – Rown – Scrool – Seach – Sheak – Smeep – Snaw – Sneed – Soat – Spow – Spream – Streen – Tean – Vear – Wead. **L2 – Non-Marked Target Pseudowords:** Blain – Blane – Bline – Bope – Chold – Crog – Dast – Daste – Dound – Dute – Fird – Fome – Frust – Goft – Hirl – Houne – Jair – Mame – Mave – Murn – Murse – Nost – Nurst – Pide – Pive – Ploud – Poud – Rold – Sind – Sive – Smirt – Spile – Storn – Stote – Surn – Tate – Tave – Vorse – Vose – Wope.

L1 – Marked Flanker Pseudowords: Ciche – Deau – Doir – Heune – Jien – Neux – Nien – Noif – Noun – Puche – Sunt – Tras – Trispe – Troip – Troune – Vouche – Dris – Peuf – Pseu – Toire. **L1 – Non-Marked Flanker Pseudowords:** Bluge – Dand – Dille – Druce – Faid – Feire – Flon – Foid – Fraipe – Fucre – Mase – Morp – Nole – Nure – Ploin – Toud – Lound – Mide – Pime – Terd.

L2 – Marked Flanker Pseudowords: Bream – Clow – Deel – Deen – Fean – Feen – Joat – Meaf – Peat – Pight – Praw – Reaf – Sead – Sear – Slood – Sneep – Sween – Twast – Twon – Vight. **L2 – Non-Marked Flanker Pseudowords:** Blad – Brice – Dron – Dush – Fride – Frume – Hice – Hine – Lold – Pail – Rand – Saste – Sate – Slar – Tand – Trane – Tump – Tust – Vame – Baid.

Appendix 26 – Chapter IV, Experiment 2 - Long Display Duration- Model comparison- Do we replicate the trial congruency effect?

On WORD items (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-104243			
Model 2	L2 proficiency	-104246	4.81	1,5	.028
Model 3	Target Language	-104264	19.57	1,6	<.001
Model 4	Flanker Language	-104272	9.83	1,7	.002
Model 5	L2 proficiency * Target Language	-104275	5.81	1,8	.016
Model 6	L2 proficiency * Flanker Language	-104274	0.66	1,9	.417
Model 7	Target Language * Flanker Language	-104289	17.33	1,10	<.001
Model 8	Full interaction	-104288	0.26	1,11	.613
On WORD items (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	3802			
Model 2	Target Language	3804	0.09	1,4	.764
Model 3	Flanker Language	3804	2.53	1,5	.112
Model 4	Target Language * Flanker Language	3793	12.80	1,6	<.001
On PSEUDOWORD items (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-105127			
Model 2	L2 proficiency	-105128	2.70	1,5	.100
Model 3	Target Language	-105127	0.80	1,6	.371
Model 4	Flanker Language	-105125	0.03	1,7	.870
Model 5	L2 proficiency * Target Language	-105124	0.70	1,8	.404
Model 6	L2 proficiency * Flanker Language	-105122	0.53	1,9	.465
Model 7	Target Language * Flanker Language	-105120	0.00	1,10	.996
Model 8	Full interaction	-105119	0.72	1,11	.396
On PSEUDOWORD items (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	3502			
Model 2	Target Language	3495	9.86	1,4	.002
Model 3	Flanker Language	3496	1.41	1,5	.235
Model 4	Target Language * Flanker Language	3498	0.08	1,6	.780

Appendix 27.A – Chapter IV, Experiment 2 -Long Display Duration- WORD RESPONSES - Model comparison- Does Target Markedness impact language decisions in a linguistic context?

On L1-WORD target subplot (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-105127			
Model 2	L2 proficiency	-105128	2.70	1,5	.100
Model 3	Target Language	-105127	0.80	1,6	.371
Model 4	Flanker Language	-105125	0.03	1,7	.870
Model 5	L2 proficiency * Target Language	-105124	0.70	1,8	.404
Model 6	L2 proficiency * Flanker Language	-105122	0.53	1,9	.465
Model 7	Target Language * Flanker Language	-105120	0.00	1,10	.996
Model 8	Full interaction	-105119	0.72	1,11	.396
On L1-WORD target subplot (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	2018			
Model 2	Target Markedness	2006	13.78	1,4	<.001
Model 3	Trial Congruency	2007	1.60	1,5	.205
Model 4	Target Markedness* Trial Congruency	2009	0.19	1,6	.664
On L2-WORD target subplot (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-51702			
Model 2	L2 Proficiency	-51707	6,43	1,5	.011
Model 3	Trial Congruency	-51705	0,56	1,6	.455
Model 4	Target Markedness	-51705	1,60	1,7	.206
Model 5	L2 proficiency * Trial Congruency	-51704	0,94	1,8	.333
Model 6	L2 proficiency* Target Markedness	-51703	1,55	1,9	.214
Model 7	Trial Congruency* Target Markedness	-51702	0,48	1,10	.490
Model 8	Full interaction	-51700	0,01	1,11	.937
On L2-WORD target subplot (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	2018			
Model 2	Target Markedness	1758	2,32	1,4	.127
Model 3	Trial Congruency	1760	0,02	1,5	.879
Model 4	Target Markedness* Trial Congruency	1762	0,15	1,6	.694

Appendix 28.B – Chapter IV, Experiment 2 -Long Display Duration- PSEUDOWORD RESPONSES - Model comparison- Does Target Markedness impact language decisions in a linguistic context?

On L1-like PSEUDOWORD target subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-53020			
Model 2	L2 Proficiency	-53021	3.13	1,5	.077
Model 3	Trial Congruency	-53019	0.02	1,6	.889
Model 4	Target Markedness	-53017	0.05	1,7	.831
Model 5	L2 proficiency * Trial Congruency	-53015	0.01	1,8	.932
Model 6	L2 proficiency* Target Markedness	-53013	0.41	1,9	.523
Model 7	Trial Congruency* Target Markedness	-53013	2.20	1,10	.138
Model 8	Full interaction	-53011	0.03	1,11	.863
On L1-like PSEUDOWORD target subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	1474			
Model 2	Target Markedness	1475	0.89	1,4	.346
Model 3	Trial Congruency	1477	0.09	1,5	.758
Model 4	Target Markedness* Trial Congruency	1477	2.01	1,6	.157
On L2-like PSEUDOWORD target subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-51974			
Model 2	L2 Proficiency	-51975	2,25	1,5	.134
Model 3	Trial Congruency	-51973	0,01	1,6	.935
Model 4	Target Markedness	-51971	0,71	1,7	.399
Model 5	L2 proficiency * Trial Congruency	-51971	1,26	1,8	.261
Model 6	L2 proficiency* Target Markedness	-51972	2,99	1,9	.084
Model 7	Trial Congruency* Target Markedness	-51970	0,11	1,10	.740
Model 8	Full interaction	-51968	0,50	1,11	.481
On L2-like PSEUDOWORD target subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	2048			
Model 2	Target Markedness	2049	0,59	1,4	.444
Model 3	Trial Congruency	2046	4,89	1,5	.027
Model 4	Target Markedness* Trial Congruency	2046	2,15	1,6	.142

Appendix 29.A – Chapter IV, Experiment 1 - Long Display Duration - WORD RESPONSES - Model comparison - Does Flanker Markedness impact target word recognition?

On WORD responses when flankers were MARKED (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-52180			
Model 2	Trial Congruency	-52185	7.56	1,5	.006
Model 3	Flanker Language	-52191	7.31	1,6	.007
Model 4	Trial Congruency*Flanker Language	-52208	18.98	1,7	<.001
On WORD responses when flankers were MARKED (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1870			
Model 2	Trial Congruency	1865	8.07	1,4	.004
Model 3	Flanker Language	1867	0.21	1,5	.644
Model 4	Trial Congruency*Flanker Language	1868	0.58	1,6	.447
On WORD responses when flankers were NON-MARKED (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-51819			
Model 2	Trial Congruency	-51826	9,32	1,5	.002
Model 3	Flanker Language	-51825	1,57	1,6	.211
Model 4	Trial Congruency*Flanker Language	-51838	14,23	1,7	<.001
On WORD responses when flankers were NON-MARKED (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	2036			
Model 2	Trial Congruency	2032	5,51	1,4	.019
Model 3	Flanker Language	2032	2,63	1,5	.105
Model 4	Trial Congruency*Flanker Language	2034	0,06	1,6	.810

Appendix 30.B – Chapter IV, Experiment 1 - Long Display Duration - PSEUDOWORD RESPONSES - Model comparison - Does Flanker Markedness impact target word recognition?

On PSEUDOWORD responses when flankers were MARKED (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-52495			
Model 2	Trial Congruency	-52496	2.46	1,5	.117
Model 3	Flanker Language	-52495	0.84	1,6	.359
Model 4	Trial Congruency*Flanker Language	-52493	0.01	1,7	.921
On PSEUDOWORD responses when flankers were MARKED (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1773			
Model 2	Trial Congruency	1774	0.74	1,4	.388
Model 3	Flanker Language	1776	0.06	1,5	.813
Model 4	Trial Congruency*Flanker Language	1774	4.01	1,6	.045
On PSEUDOWORD responses when flankers were NON-MARKED (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-52441			
Model 2	Trial Congruency	-52443	3,32	1,5	.068
Model 3	Flanker Language	-52442	1,62	1,6	.204
Model 4	Trial Congruency*Flanker Language	-51442	1,67	1,7	.196
On PSEUDOWORD responses when flankers were NON-MARKED (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1803			
Model 2	Trial Congruency	1804	0,67	1,4	.413
Model 3	Flanker Language	1802	4,65	1,5	.031
Model 4	Trial Congruency*Flanker Language	1790	13,12	1,6	<.001

Appendix 31 – Chapter IV, Experiment 2 - Short Display Duration- Model comparison- Do we replicate the trial congruency effect?

On WORD responses (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-82878			
Model 2	L2 proficiency	-82876	0.74	1,5	.389
Model 3	Target Language	-82891	17.16	1,6	<.001
Model 4	Flanker Language	-82899	9.48	1,7	.002
Model 5	L2 proficiency * Target Language	-82898	0.70	1,8	.401
Model 6	L2 proficiency * Flanker Language	-82896	0.03	1,9	.857
Model 7	Target Language * Flanker Language	-82897	2.80	1,10	.094
Model 8	Full interaction	-82897	2.34	1,11	.126
On WORD responses (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	3833			
Model 2	Target Language	3835	0.30	1,4	.583
Model 3	Flanker Language	3837	0.60	1,5	.437
Model 4	Target Language * Flanker Language	3736	2.82	1,6	.093
On PSEUDOWORD responses (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-80018			
Model 2	L2 proficiency	-80017	1.18	1,5	.277
Model 3	Target Language	-80016	1.09	1,6	.297
Model 4	Flanker Language	-80015	0.99	1,7	.319
Model 5	L2 proficiency * Target Language	-80022	8.59	1,8	.003
Model 6	L2 proficiency * Flanker Language	-80021	0.89	1,9	.345
Model 7	Target Language * Flanker Language	-80019	0.22	1,10	.637
Model 8	Full interaction	-80017	0.46	1,11	.496
On PSEUDOWORD responses (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	5183			
Model 2	Target Language	5181	3.64	1,4	.056
Model 3	Flanker Language	5182	0.89	1,5	.345
Model 4	Target Language * Flanker Language	5184	0.24	1,6	.626

Appendix 32.A – Chapter IV, Experiment 2 -Short Display Duration- WORD RESPONSES - Model comparison- Does Target Markedness impact language decisions in a linguistic context?

	On L1-WORD target subplot (RT)				
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-41745			
Model 2	L2 Proficiency	-41744	0,57	1,5	.450
Model 3	Trial Congruency	-41753	11,45	1,6	<.001
Model 4	Target Markedness	-41751	0,36	1,7	.547
Model 5	L2 proficiency * Trial Congruency	-41751	1,37	1,8	.242
Model 6	L2 proficiency* Target Markedness	-41749	0,05	1,9	.829
Model 7	Trial Congruency* Target Markedness	-41747	0,17	1,10	.681
Model 8	Full interaction	-41745	0,29	1,11	.588
	On L1-WORD target subplot (Accuracy)				
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	1964			
Model 2	Target Markedness	1965	0,33	1,4	.564
Model 3	Trial Congruency	1966	1,58	1,5	.209
Model 4	Target Markedness* Trial Congruency	1967	0,70	1,6	.404
	On L2-WORD target subplot (RT)				
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-41072			
Model 2	L2 Proficiency	-41071	0,94	1,5	.331
Model 3	Trial Congruency	-41070	0,99	1,6	.319
Model 4	Target Markedness	-41068	0,15	1,7	.699
Model 5	L2 proficiency * Trial Congruency	-41067	0,95	1,8	.329
Model 6	L2 proficiency* Target Markedness	-41068	2,97	1,9	.085
Model 7	Trial Congruency* Target Markedness	-41067	1,11	1,10	.291
Model 8	Full interaction	-41065	0,10	1,11	.749
	On L2-WORD target subplot (Accuracy)				
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	1878			
Model 2	Target Markedness	1871	3,07	1,4	.080
Model 3	Trial Congruency	1872	1,78	1,5	.182
Model 4	Target Markedness* Trial Congruency	1873	0,13	1,6	.715

Appendix 33.B – Chapter IV, Experiment 2 -Short Display Duration- PSEUDOWORD RESPONSES - Model comparison- Does Target Markedness impact language decisions in a linguistic context?

		AIC	χ²	ddl	p
On L1-PSEUDOWORD target subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-40960			
Model 2	L2 Proficiency	-40960	2,04	1,5	.154
Model 3	Trial Congruency	-40958	0,08	1,6	.784
Model 4	Target Markedness	-40956	0,15	1,7	.703
Model 5	L2 proficiency * Trial Congruency	-40956	1,54	1,8	.214
Model 6	L2 proficiency* Target Markedness	-40955	0,95	1,9	.330
Model 7	Trial Congruency* Target Markedness	-40953	0,15	1,10	.699
Model 8	Full interaction	-40953	1,65	1,11	.199
On L1-PSEUDOWORD target subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	2509			
Model 2	Target Markedness	2510	0,98	1,4	.322
Model 3	Trial Congruency	2509	3,33	1,5	.068
Model 4	Target Markedness* Trial Congruency	2511	0,04	1,6	.832
On L2-PSEUDOWORD target subplot (RT)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	-38964			
Model 2	L2 Proficiency	-38963	0,54	1,5	.460
Model 3	Trial Congruency	-38962	0,88	1,6	.347
Model 4	Target Markedness	-38967	7,78	1,7	.005
Model 5	L2 proficiency * Trial Congruency	-38965	0,06	1,8	.815
Model 6	L2 proficiency* Target Markedness	-38964	0,37	1,9	.544
Model 7	Trial Congruency* Target Markedness	-38962	0,30	1,10	.583
Model 8	Full interaction	-38961	1,42	1,11	.233
On L2-PSEUDOWORD target subplot (Accuracy)					
		AIC	**χ²**	**ddl**	**p**
Model 1	Random	2733			
Model 2	Target Markedness	2735	0,11	1,4	.745
Model 3	Trial Congruency	2730	7,34	1,5	.007
Model 4	Target Markedness* Trial Congruency	2732	0,04	1,6	.838

Appendix 34.A – Chapter IV, Experiment 2 - Short Display Duration - WORD RESPONSES - Model comparison - Does Flanker Markedness impact target word recognition?

On WORD responses when flankers were MARKED (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-41162			
Model 2	Trial Congruency	-41171	10.76	1,5	.001
Model 3	Flanker Language	-41173	3.62	1,6	.057
Model 4	Trial Congruency*Flanker Language	-41182	11.44	1,7	<.001
On WORD responses when flankers were MARKED (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	2040			
Model 2	Trial Congruency	2032	10.64	1,4	.001
Model 3	Flanker Language	2033	0.72	1,5	.395
Model 4	Trial Congruency*Flanker Language	2034	0.57	1,6	.452
On WORD responses when flankers were NON-MARKED (RT)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	-41494			
Model 2	Trial Congruency	-41493	1.03	1,5	.311
Model 3	Flanker Language	-41496	4.72	1,6	.030
Model 4	Trial Congruency*Flanker Language	-41510	16.60	1,7	<.001
On WORD responses when flankers were NON-MARKED (Accuracy)					
		AIC	χ^2	**ddl**	**p**
Model 1	Random	1891			
Model 2	Trial Congruency	1893	0.39	1,4	.531
Model 3	Flanker Language	1891	3.41	1,5	.065
Model 4	Trial Congruency*Flanker Language	1893	0.01	1,6	.939

Appendix 35.B – Chapter IV, Experiment 2 - Short Display Duration - PSEUDOWORD RESPONSES - Model comparison - Does Flanker Markedness impact target word recognition?

On PSEUDOWORD responses when flankers were MARKED (RT)					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	-39842			
Model 2	Trial Congruency	-39841	0.55	1,5	.457
Model 3	Flanker Language	-39845	6.05	1,6	.014
Model 4	Trial Congruency*Flanker Language	-39843	0.23	1,7	.632
On PSEUDOWORD responses when flankers were MARKED (Accuracy)					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	2650			
Model 2	Trial Congruency	2651	1.17	1,4	.278
Model 3	Flanker Language	2652	0.50	1,5	.479
Model 4	Trial Congruency*Flanker Language	2652	2.38	1,6	.123
On PSEUDOWORD responses when flankers were NON-MARKED (RT)					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	-40014			
Model 2	Trial Congruency	-40012	0.27	1,5	.601
Model 3	Flanker Language	-40011	0.91	1,6	.339
Model 4	Trial Congruency*Flanker Language	-40011	1.83	1,7	.176
On PSEUDOWORD responses when flankers were NON-MARKED (Accuracy)					
		AIC	**χ^2**	**ddl**	**p**
Model 1	Random	2651			
Model 2	Trial Congruency	2652	0.40	1,4	.525
Model 3	Flanker Language	2654	0.26	1,5	.608
Model 4	Trial Congruency*Flanker Language	2653	3.16	1,6	.076

Université de Strasbourg

Marie-Ange LECERF

An investigation of three sublexical variables and their influence on L2 visual word recognition.

LPC Laboratoire de psychologie des cognitions | LPC Université de Strasbourg

Résumé

Dans un contexte scolaire, l'apprenant de langue seconde (L2) est massivement exposé à l'orthographe de sa L2, ce qui implique l'acquisition de nouveau pattern orthographiques (i.e., le bigramme KN pour un français apprenant l'anglais), et la mise à jour de leur répertoire de correspondances Graphèmes-Phonèmes (CGP) qui peuvent varier entre les langues mais aussi au sein de la L2 (i.e., le OU peut se prononcer \u\ en français mais aussi \ow\ en anglais). L'objectif de cette thèse consiste donc à explorer les conséquences d'un tel apprentissage sublexical sur le traitement des mots L2, et ce en (1) identifiant l'impact respectif de trois variables, i.e., le marquage orthographique, la congruence et la consistance des CGP, et en (2) étudiant la dynamique existant entre le marquage orthographique et le contexte linguistique. Les résultats de notre première expérience ont révélé que bien que les variations intra- et inter-langues des CGP aient un impact sur la prononciation des mots de L2, elles n'en avaient pas sur l'identification de ces mêmes mots. En revanche, l'apparition de l'effet du marquage orthographique semble étroitement lié aux propriétés des CGPs mobilisées. De plus, notre seconde étude a mis en évidence un effet significatif du marquage orthographique sur l'identification de la langue d'un item qui est absent lors de l'identification du mot per se (en accord avec Van Kesteren et al., 2012). Pour finir, les lecteurs d'une L2 semble être capables de détecter les marqueurs orthographiques présents dans un contexte linguistique bien que leur impact sur la réponse à un item cible ne soit pas équivoque. (Mots Clefs : Marquage Orthographique ; Congruence des CGP ; Consistance des CGP ; Acquisition d'une L2 ; Apprentissage Sublexical ; Traitement visuel de mots L2)

Abstract

In a school setting, second-language (L2) acquisition is often imparted through visual medias. As a result, L2-learners are heavily exposed to their L2-orthography, which implies the acquisition of new orthographic patterns (i.e., the KN bigram for a French native learner of English), and the revision of their set of Grapheme-to-Phoneme Correspondences (GPCs) which can vary across-languages but also within the L2 (i.e., the OU bigram is pronounced \u\ in French but also \aʊ\ in English). The aim of this PhD-thesis is to explore the consequences of such a sublexical learning for later L2 word processing by, 1) identifying the separate impact of three variables, namely orthographic markedness, GPC congruency and GPC consistency, and by 2) investigating the dynamic between orthographic markedness and another source of language membership cue, i.e., the linguistic context. The results of our first experiment revealed that while GPC variations, both within L2 and across languages impacted L2 word pronunciation they did not impact L2 visual word identification. On the other hand, the effect of orthographic markedness was elusive and seemed to be tied to the GPC properties of the items selected. Moreover, our second experiment showed that while orthographic markers were useful for language membership identification it was not used for word identification per se (in line with results from Van Kesteren et al., 2012). This second experiment also indicated that L2-readers were able to pick out on orthographic markers presented in a context although there was little evidence of them impacting responses on a target item. (Keywords= Orthographic Markedness ; GPC congruency ; GPC consistency ; L2 acquisition; Sublexical Learning ; L2 visual word processing)