Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz
https://creativecommons.org/licenses/by/4.0/deed.de

LA MAITRISE DES CONNECTEURS DISCURSIFS EN LANGUE SECONDE

ÉVALUATION EXPÉRIMENTALE DES FACTEURS LINGUISTIQUES ET COGNITIFS

THÈSE INAUGURALE DE LA FACULTÉ DES LETTRES DE L'UNIVERSITÉ DE BERNE
POUR L'OBTENTION DU TITRE DOCTEUR,
PRÉSENTÉE PAR

MATHIS WETZEL

SOUMISE À

PROF. DR. SANDRINE ZUFFEREY,
INSTITUT DE LANGUE ET DE LITTÉRATURE FRANÇAISES DE L'UNIVERSITÉ DE BERNE

ET

PROF. DR. LIESBETH DEGAND
INSTITUT LANGAGE ET COMMUNICATION DE L'UNIVERSITÉ CATHOLIQUE LOUVAIN

PRÉSENTÉE ET SOUTENUE LE 26.09.2023

WETZEL, MATHIS
CHEMIN DU CRÊT 6
2533 EVILARD
SUISSE
17-117-516

BERNE, LE 31.03.23

source: https://doi.org/10.48549/4618 | downloaded: 3.7.2025

REMERCIEMENTS **5**

1 INTRODUCTION **7**

1.1 Comment expliquer les difficultés à bien maitriser les connecteurs dans une langue seconde ? 8
1.2 Les questions clés 10
1.3 Structure du travail 11
1.4 Considérations théoriques 14
1.5 Considérations méthodologiques 15
1.6 Remarques préliminaires 18

2 LES DIFFÉRENTS MODÈLES DE DISCOURS ET LE RÔLE DES CONNECTEURS **19**

2.1 Le point commun des modèles de discours 20
2.2 Les différences entre les modèles de discours 21

2.2.1 Aperçu de quelques modèles de discours 23
2.2.2 L'approche cognitive des relations de cohérence 28

2.3 Conclusion intermédiaire 33
2.4 Moyens linguistiques pour établir la cohésion et cohérence textuelle 33

2.4.1 Les marques cohésives selon Halliday et Hasan (1976) 34
2.4.2 Les marques relationnelles 36

2.5 Les connecteurs discursifs 37

2.5.1 Les instructions procédurales des connecteurs 39
2.5.2 L'utilité des connecteurs pour la lecture 43
2.5.3 Dans quels cas convient-il d'utiliser un connecteur ? 46
2.5.4 Les fonctions pragmatiques des connecteurs 51

2.6 Conclusion 55

3 LA MAITRISE DES CONNECTEURS DISCURSIFS EN LANGUE SECONDE **57**

3.1 Qu'est-ce la maitrise ? 57

3.1.1 Quelques considérations théoriques préliminaires 58
3.1.2 Les méthodes courantes utilisées 60

3.2 L'utilité des connecteurs et des signaux alternatifs pour le traitement et la compréhension d'un texte en L2 69
3.3 La maitrise insuffisante des connecteurs par les personnes non-natives 73
3.4 Le rôle de l'enseignement 77
3.5 La variation interpersonnelle 79
3.6 L'hétérogénéité des connecteurs 83
3.7 Les questions non-résolues 86
3.8 Conclusion 90

4 L'IMPACT DU TYPE DU CONNECTEUR SUR LE TRAITEMENT MENTAL EN L1 91

4.1 Quelles propriétés des connecteurs peuvent influencer la lecture d'un discours en L1 ? 92

4.1.1 La fréquence du connecteur 93
4.1.2 La polyfonctionnalité du connecteur 94
4.1.3 Y-a-t-il un impact différent de la fréquence et de la polyfonctionnalité d'un connecteur en fonction de la relation de cohérence ? 96

4.2 La série d'expériences présente 98
4.3 Expérience 1 : la réaction à l'incohérence pour des phrases contenant les connecteurs *donc* et *mais* 99

4.3.1 Les participant-e-s 100
4.3.2 Le design et la procédure 100
4.3.3 L'analyse 103
4.3.4 Discussion 107

4.4 Expérience 2 : la réaction à l'incohérence pour des phrases contenant les connecteurs *ainsi* et *néanmoins* 109

4.4.1 Les participant-e-s 109
4.4.2 Le design et la procédure 109
4.4.3 L'analyse 109
4.4.4 Discussion 113

4.5 Expérience 3 : la réaction à l'incohérence pour des phrases contenant les connecteurs *aussi* et *or* 114

4.5.1 Les participant-e-s 116
4.5.2 Le design et la procédure 116
4.5.3 L'analyse 116
4.5.4 Discussion 119

4.6 Conclusion 121

5 L'IMPACT DU TYPE DU CONNECTEUR SUR LA CONNAISSANCE EXPLICITE EN L2 125

5.1 Quels facteurs peuvent rendre les connecteurs plus difficiles à utiliser dans une langue seconde? 126

5.1.1 Le rôle de la relation de cohérence 127
5.1.2 La fréquence du connecteur 128
5.1.3 Quelles compétences prédisent une meilleure maitrise des connecteurs en L2 ? 130

5.2 Tâche d'insertion de connecteurs 132

5.2.1 Pré-test 133
5.2.2 Les participant-e-s 136
5.2.3 Le design et la procédure 137
5.2.4 Les résultats et l'analyse 140
5.2.5 Discussion 156

5.3 Conclusion 161

6 LA CAPACITÉ DES PERSONNES NON-NATIVES À INTÉGRER LES SIGNIFICATIONS DES CONNECTEURS LORS DE LA LECTURE 163

6.1 Le traitement mental en L2 164

6.1.1 Les caractéristiques de la lecture en L2 : théories et modèles sur le traitement mental 165

6.1.2 Les connecteurs peuvent-ils faciliter le traitement en ligne d'un discours dans une langue seconde ? ... 167

6.2 Expérience 1 : Mauvaises utilisations des connecteurs et perturbation de la lecture ... 168

6.2.1 Les participant-e-s ... 170
6.2.2 Le design et la procédure ... 171
6.2.3 L'analyse ... 171
6.2.4 Discussion ... 175

6.3 Expérience 2: la réaction aux connecteurs manquants ... 177

6.1.1 Les participant-e-s ... 178
6.1.2 Le design et la procédure ... 179
6.3.1 L'analyse ... 179
6.3.2 Discussion ... 187

6.4 Conclusion ... 188

7 L'IMPACT DU TYPE DE CONNECTEUR SUR LE TRAITEMENT MENTAL EN L2 ... 191

7.1 Est-ce qu'un connecteur moins accessible aux personnes non-natives réduit leur sensibilité à l'incohérence ? ... 193
7.2 Analyse de la complexité des connecteurs *mais* et *cependant* ... 194
7.3 Expérience 1 : la réaction à l'incohérence en fonction du type du connecteur lors du traitement mental ... 199

7.3.1 Les participant-e-s ... 199
7.3.2 Le design et la procédure ... 200
7.3.3 L'analyse ... 203
7.3.4 Discussion ... 209

7.4 Expérience 2 : la réaction à l'incohérence en fonction du type de connecteur lors d'une tâche d'évaluation ... 211

7.4.1 Les participant-e-s ... 212
7.4.2 Le design et la procédure ... 213
7.4.3 Les résultats et l'analyse ... 214

7.5 Discussion ... 222
7.6 Conclusion ... 224

8 LES INFLUENCES CROSSLINGUISTIQUES ... 227

8.1 Les influences de la L1 pour la maitrise des connecteurs en L2 ... 229

8.1.1 Les différents types d'influence de la L1 pour la maitrise des connecteurs en L2 ... 229
8.1.2 Les apprenant-e-s sont-ils capables de détecter un mauvais emploi d'un connecteur dû au transfert négatif ? ... 232
8.1.3 Les caractéristiques des spécifications et leur rôle dans le transfert négatif ... 233

8.2 Expérience 1 : la détection de l'emploi erroné de *alors* lors de l'accès aux connaissances explicites ... 237

8.2.1 Les participant-e-s ... 238
8.2.2 Le design et la procédure ... 239
8.2.3 L'analyse ... 240
8.2.4 Discussion ... 244

8.3 Expérience 2 : la détection de l'emploi erroné de *alors* lors de l'accès aux connaissances implicites ... 246

8.3.1 Les participant-e-s ... 247

8.3.2 La procédure 248
8.3.3 L'analyse 248
8.3.4 Discussion 252

8.4 Conclusion 253

9 LA SENSIBILITÉ AUX CONTRAINTES SYNTAXIQUES ET PRAGMATIQUES LORS DE L'ACCÈS AUX CONNAISSANCES EXPLICITES 257

9.1 Les contraintes qui conduisent à utiliser un connecteur spécifique 259

9.1.1 Les contraintes syntaxiques des connecteurs 259
9.1.2 Les contraintes pragmatiques des connecteurs 261
9.1.3 Les contraintes pragmatiques et syntaxiques de *parce que* et *puisque* 262

9.2 Expérience : La sensibilité aux contraintes syntaxiques et pragmatiques de *parce que* et *puisque* 267

9.2.1 Les participant-e-s 268
9.2.2 Le design et la procédure 269
9.2.3 Les résultats et l'analyse 271
9.2.4 Analyse exploratoire des connaissances métalinguistiques 280
9.2.5 Discussion 282

9.3 Conclusion 285

10 CONCLUSION 288

10.1 Les principaux résultats 288

10.1.1 « Les connecteurs ont des degrés de complexité différents pour les apprenant-e-s » 289
10.1.2 « Les causes à la difficulté des connecteurs en L2 sont multifactorielles » 290
10.1.3 « La maitrise des connecteurs en L2 peut être prédite par l'exposition à l'écrit et par la compétence linguistique de l'apprenant-e» 301
10.1.4 « La maitrise des connecteurs dépend du type de connaissances auxquelles l'apprenant-e peut accéder » 304

10.2 Les observations sur la maitrise des connecteurs en L1 306
10.3 Limites 308
10.4 Futures directions de recherches 318
10.5 Le mot de la fin 322

11 RÉFÉRENCES 323

Remerciements

Cette thèse de doctorat n'aurait pas été possible sans le soutien de nombreuses personnes exceptionnelles, que je tiens à remercier chaleureusement. Ces dernières années au cours desquelles cette thèse a été rédigée ont été mouvementées, que ce soit au niveau global ou privé, et je suis d'autant plus reconnaissant pour tout le soutien que j'ai pu recevoir durant cette période.

Tout d'abord j'aimerais remercier chaleureusement Sandrine Zufferey pour m'avoir initié aux questions du discours, pour son engagement, pour avoir guidé et amélioré mon travail, pour avoir cru en mes capacités, pour son temps et sa disponibilité pour mes questions et difficultés… la liste ne saurait s'arrêter là. Je suis très reconnaissant pour ses conseils, tant sur les questions techniques que sur les questions plus générales. C'est sans exagération que je peux dire que je me suis senti, à tout moment de ces dernières années, très privilégié de sa supervision.

Ensuite, je tiens à remercier chaleureusement Pascal Gygax pour son aide sur les questions statistiques, ainsi que pour son support et son aide dans toute la réalisation des expériences. J'ai énormément appris de lui et j'ai particulièrement apprécié sa disponibilité et sa volonté à répondre à toutes mes questions.

Un très grand merci également à ma collègue, Katya Tskhovrebova : d'une part pour les discussions techniques et les échanges sur les connecteurs, les références, etc., mais aussi, bien plus souvent, pour les petites pauses café et les bavardages sur tout et rien. Ainsi, j'ai trouvé que l'ambiance de travail dans notre projet a rendu ma tâche extrêmement agréable.

Outre Sandrine et Katya, j'aimerais également remercier d'autres personnes qui font – ou ont fait – partie de l'équipe de la section de linguistique à l'Institut au cours de ces dernières années. Ce sont Baptiste Bersier, Sandra Schwab, Jennifer Schumann et Ruedi Rohrbach. Il m'a été un grand plaisir de travailler avec eux et faire partie de ce groupe agréable.

J'aimerais également remercier Liesbeth Degand, non seulement pour avoir accepté d'évaluer ce travail mais aussi pour ses remarques qui m'ont aidé considérablement pour la mise en forme du travail et représentaient ainsi une grande influence sur la présente thèse.

Un grand merci également à Ludivine Crible pour toute son aide et soutien. Sa volonté et serviabilité m’ont beaucoup impressionné et j’ai beaucoup appris lors de notre collaboration.

Par ailleurs, je tiens aussi à remercier tous les participant-e-s de mes expériences ainsi que les enseignant-e-s qui ont accepté de passer ces expériences dans ses classes et qui m’ont ainsi permis de réaliser cette thèse. Je tiens également à remercier le Fonds National Suisse (FNS) pour le soutien financier du projet.

J’aimerais enfin remercier tous mes ami-e-s, lointains ou proches, qui m'ont soutenu pendant toute cette période. La contribution qu'ils ont apportée à la rédaction de cette thèse par leur soutien, leur distraction et leur divertissement ne peut être sous-estimée. En me permettant d’écrire ces quelques lignes en allemand, *möchte ich schließlich auch meinen Eltern und meiner Schwester für die Unterstützung und Teilnahme in den letzten Jahren danken.*

Enfin, un merci tout particulier à Maude et Smilla - vous savez quelle contribution vous avez apportée à cette thèse. Je vous remercie donc pour votre soutien au cours de ces dernières années et, je l'espère, pour de nombreuses années encore à venir.

Berne, le 31.03.2023

1 Introduction

En 1990, le linguiste William Crewe fait une constatation quelque peu accablante :

> « Le mauvais usage des connecteurs logiques est une caractéristique presque universelle de l'écriture des étudiant-e-s d'ALS [anglais langue seconde, note de l'auteur] » (1990 : 317, ma traduction).

Comment le chercheur en arrive-t-il à ce constat ? Et de quels éléments linguistiques parle-t-il exactement ?

D'abord, ce qu'il entend par *connecteurs logiques*, ce sont des mots comme *donc*, *tandis que* et *si*, donc des mots qui servent à établir une relation entre deux segments d'un texte (voir exemples 1-3).

1) Je n'aime pas la coriandre, *donc* je ne mangerai pas ce sandwich.
2) *Tandis que* certaines personnes aiment la coriandre, moi je ne l'aime pas.
3) *Si* vous mettez de la coriandre sur ce sandwich, je ne le mangerai pas.

En analysant des productions de textes des apprenant-e-s d'anglais, Crewe (1990) constate que ces mots sont dans la plupart des cas incorrectement utilisés. Il remarque, par exemple, que les apprenant-e-s confondent souvent les connecteurs anglais *on the contrary* ('au contraire') et *however* ('cependant') et que cela conduit à des problèmes de cohérence.

Un même constat peut être fait en regardant des productions d'apprenant-e-s dans d'autres langues que l'anglais. Considérons l'exemple suivant, issu d'une rédaction en français d'un gymnasien-ne suisse de langue maternelle allemande :

> « Pour moi, cette photo a quelquechose très intérieure. La lumière comme un auréole de l'arbre, *alors* de la nature, n'est pas éxplicapable pour moi. »

Dans cet exemple, nous voyons que l'élève utilise le connecteur *alors* pour spécifier le mot *arbre*, à savoir en disant que celui-ci représente la nature en général. Cependant, le connecteur *alors* ne convient pas dans ce cas. Une explication de cette erreur est que l'élève a supposé, par erreur, que le connecteur français *alors* correspond au connecteur allemand *also*, donc un connecteur qui est utilisé dans

ces situations.[1] Mais les difficultés ne semblent pas émerger seulement à cause d'une confusion entre les connecteurs de différentes langues. Considérons encore un exemple, cette fois tiré d'un corpus de productions de textes par des apprenant-e-s du français (LIDILEM, Jacques & Rinck, 2017).

> « *Ainsi* il serait possible de construire un modèle pour l'investissement textuel permettant de faire émerger des objectifs pédagogiques clairs. *Pourtant* les chercheurs émettent de nombreuses réserves face à l'hétérogénéité des textes.
>
> *Cependant* les travaux réalisés amènent à constituer cinq catégories de types de textes à savoir les textes descriptifs, narratifs, explicatifs, programmatifs et argumentatifs. *Pourtant* cette catégorisation laisse en suspens la place des textes informatifs et dialogiques, montrant les incertitudes épistémologiques. »

Ici, l'utilisation excessive des connecteurs rend le paragraphe non seulement très dense, il n'est pas non plus tout à fait clair comment les idées mentionnées se relient. De plus, l'exemple illustre un autre phénomène de l'utilisation des connecteurs dans une langue seconde : la tendance à placer les connecteurs en position initiale de la phrase.

Partant du principe que les connecteurs sont difficiles à maitriser dans une langue seconde, la question se pose inévitablement de savoir si tous les connecteurs sont également difficiles à maitriser ou s'il existe certains connecteurs (ou circonstances) qui rendent leur utilisation plus difficile. La présente thèse vise à répondre à cette question. Plus précisément, notre travail cherche à explorer les facteurs linguistiques et cognitifs qui permettent d'expliquer pourquoi certains connecteurs semblent être plus difficiles à maitriser dans une langue seconde (ci-après L2) que d'autres. Pour pouvoir retracer les répercussions de l'hétérogénéité des connecteurs sur leur maitrise en langue seconde, une approche psycholinguistique et expérimentale a été adoptée.

Tout d'abord, nous allons examiner plus en détail comment les difficultés des personnes non-natives avec les connecteurs sont documentées dans la littérature, ainsi que les raisons qui sont avancées pour cela.

1.1 Comment expliquer les difficultés à bien maitriser les connecteurs dans une langue seconde ?

Bien que les connecteurs représentent des outils linguistiques extrêmement importants pour assurer la cohérence d'un discours, la recherche a montré que les

[1] Nous allons revenir sur ce type d'erreur au chapitre 8.

personnes non-natives, ayant même une compétence linguistique très élevée, éprouvent d'énormes difficultés à les maitriser de manière adéquate dans leur langue cible (p.ex., Zufferey & Gygax, 2017). Les analyses des productions de texte des apprenant-e-s font par exemple état de sur- ou sous-utilisations des connecteurs (p.ex., Field & Yip, 1992; Granger & Tyson, 1996). D'autres mettent en évidence des cas où les connecteurs sont utilisés de manière incorrecte (p.ex., Milton & Tsang, 1993; Lamiroy, 1994; Tapper, 2005). Bien que de nombreux chercheurs-e-s discutent le rôle de l'enseignement à cet égard (p.ex., Crewe, 1990), la recherche n'a pas encore pu identifier de manière satisfaisante tous les facteurs qui rendent les connecteurs difficiles à apprendre en langue seconde.

Pour commencer, il convient de rappeler les facteurs qui influencent l'apprentissage d'une langue seconde de manière générale. Les chercheurs VanPatten et Benati (2010) distinguent trois domaines différents.

Le premier de ces domaines est *l'enseignement*. Bien qu'il soit fortement supposé que l'enseignement influence l'acquisition de la L2 et donc aussi la manière dont les apprenant-e-s maitrisent les connecteurs (Crewe, 1990; voir aussi chapitre 3.4), un enseignement insuffisant ne peut pas expliquer entièrement pourquoi l'apprentissage des connecteurs semble être si compliqué dans une L2. En effet, on a pu observer des différences dans la maitrise des connecteurs même au sein d'un groupe ayant reçu le même enseignement (p.ex., Geva, 1986).

Le deuxième domaine évoqué par VanPatten et Benati (2010) est la *variation interpersonnelle*, donc l'idée que certaines caractéristiques et compétences personnelles prédisent une meilleure maitrise d'une L2. Concernant les connecteurs de discours, la littérature a montré l'existence de variations interpersonnelles considérables, à la fois en L1 et en L2 (Zufferey, Tskhovrebova, Wetzel & Gygax, 2022; voir aussi chapitre 3.5). Cependant, les compétences individuelles ne peuvent pas expliquer pourquoi certains connecteurs sont moins bien maitrisés que d'autres. Pour comprendre pourquoi certains connecteurs sont plus difficiles que d'autres, il doit donc y avoir encore d'autres facteurs.

A cet égard, il convient de considérer le troisième domaine évoqué par VanPatten et Benati (2010) : les *facteurs linguistiques* et *cognitifs*. Il semble que ces facteurs-là sont particulièrement importants pour une bonne maitrise de la L2 :

> « SLA [angl. *Second Language Acquisition*, « acquisition de langue seconde »] est complexe et une variété de facteurs linguistiques, psycholinguistiques et contextuels interagissent pour façonner et limiter le cours de l'acquisition. Selon nous, les facteurs linguistiques et psycholinguistiques (traitement) jouent un rôle majeur dans le développement des systèmes linguistiques. » (Van Patten & Benati, 2010 : 57, ma traduction) »

Un exemple de facteur linguistique est la structure syntaxique d'un énoncé. Il est possible de faire l'hypothèse, par exemple, qu'une clause subordonnée est plus difficile à produire et à comprendre qu'une structure linguistiquement plus simple, comme par exemple un temps composé (Ellis, 2006; voir aussi Pienemann, 1998). Ainsi, le fait que certains éléments de la langue soient plus complexes (en termes de propriétés linguistiques et cognitives) que d'autres les rend également plus difficiles à apprendre (Pienemann, 2015).

Dans la présente thèse, nous allons étudier le rôle de ces facteurs pour l'acquisition des connecteurs de discours. Plus précisément, nous allons examiner quels facteurs influencent la maitrise des connecteurs du français par des personnes non-natives.

1.2 Les questions clés

L'enjeu principal de la présente thèse est d'examiner plus en détail la maitrise des connecteurs du français par les personnes non-natives. A cet égard, nous nous intéresserons à la question de savoir si la maitrise des connecteurs dépend du type de connecteur, c'est-à-dire si certains connecteurs sont plus difficiles à maitriser que d'autres. Si tel est bien le cas, la question se pose de savoir pourquoi. Dans ce contexte, nous mettrons particulièrement l'accent sur les *facteurs linguistiques et cognitifs,* comme par exemple le type de relation qu'un connecteur indique, sa fréquence ou la présence d'un connecteur dans la L1 de l'apprenant-e qui partage la même fonction qu'un connecteur de la L2. Bien que ces facteurs semblent être des facteurs *fondamentaux* pour la maitrise d'une L2 en général (VanPatten & Benati, 2010), il n'y a pas, à notre connaissance, d'études expérimentales qui évaluent l'impact de ces facteurs sur les connecteurs de discours. La présente thèse vise à combler ce vide.

Il est évident qu'il existe – à part ces facteurs qui sont directement liés aux connecteurs – également d'autres variables qui influencent la maitrise des connecteurs en L2. Cette maitrise doit être considérée, par exemple, à la lumière

des facteurs individuels, c'est-à-dire le fait que certaines caractéristiques propres aux individus, comme leur niveau de compétences linguistiques, puisse influencer la maitrise des connecteurs.

A cela s'ajoute le fait que la maitrise des connecteurs n'est pas constante, même pour une personne donnée. Par exemple, si une personne se concentre consciemment sur la question de savoir si un connecteur est correct ou incorrect, elle évaluera mieux les connecteurs que si elle lisait les mêmes phrases pour obtenir des informations sur le contenu du texte (voir aussi Degand et al., 1999; Zufferey et al., 2015 pour des observations en L1). Une distinction importante à cet égard est la différence entre la connaissance implicite et explicite (selon Ellis, 2009a) que nous allons élaborer plus en détail au chapitre 3.1.1.

Afin de prendre en compte des variables qui pourraient influencer la maitrise des connecteurs en L2, nous aborderons également la question de savoir comment les caractéristiques individuelles des personnes, ainsi que le type de connaissances auxquelles elles peuvent accéder, influencent leur maitrise des connecteurs.

1.3 Structure du travail

Afin d'aborder les questions mentionnées dans la section précédente, la thèse suivra la structure suivante :

Au chapitre 2, nous aborderons les notions clés de la thèse, comme par exemple la *cohérence*, le *discours*, les *relations de cohérence* ainsi que les *connecteurs de discours*. Plus précisément, nous discuterons dans la première partie du chapitre des modèles de discours, tel que le modèle CCR (C*ognitive approach to Coherence Relations*) de Sanders, Spooren et Noordman (1992). Dans la deuxième partie, nous aborderons le rôle des connecteurs, et plus particulièrement la nécessité de les employer dans certains contextes, ainsi que leurs fonctions pragmatiques.

Sur la base de ces descriptions théoriques, nous nous concentrerons ensuite au chapitre 3 sur la maitrise des connecteurs dans une L2. Après avoir clarifié ce que la *maitrise d'un connecteur* veut dire dans le cadre d'une langue seconde, nous passerons en revue les études antérieures sur ce sujet, et constaterons que les connecteurs semblent être très difficiles à maitriser dans une L2. La question se

posera de savoir pourquoi il en est ainsi. Comme les explications d'un enseignement insuffisant et de la variation individuelle n'expliquent pas toutes les difficultés observées, nous ferons l'hypothèse que la maitrise des connecteurs dépend aussi du connecteur en question. L'idée est que la catégorie des connecteurs est très hétérogène et que certains connecteurs sont beaucoup plus difficiles que d'autres.

Au cours de ce chapitre, nous aborderons les questions de savoir si, d'une part, l'hétérogénéité des connecteurs se répercute sur leur maitrise en L2, et si oui, quels facteurs linguistiques et cognitifs rendent certains connecteurs plus difficiles que d'autres. Pour trouver des réponses à ces questions, nous présenterons à partir du chapitre 4 des expériences qui visent à tester les facteurs qui influencent la maitrise des connecteurs.

Dans le premier chapitre expérimental, le chapitre 4, nous testerons l'impact du type de connecteur sur la lecture en L1. En effet, avant de savoir si et comment le type de connecteur influence la maitrise en L2, il convient de déterminer comment l'hétérogénéité influence la lecture dans la L1.

Ensuite, au chapitre 5, nous tenterons de déterminer si l'hétérogénéité des connecteurs se répercute sur la maitrise des personnes non-natives. En utilisant une tâche d'insertion, nous testerons la maitrise des apprenant-e-s du français dans une expérience incluant douze connecteurs différents. Cette première tâche est de type hors-ligne, et aucune limite de temps ne sera fixée. Les participant-e-s peuvent donc prendre le temps de choisir la bonne réponse, ce qui leur permettra d'accéder à leurs connaissances explicites.

Au chapitre 6, nous examinerons si les personnes non-natives réagissent à la présence de connecteurs durant la lecture, c'est-à-dire lorsqu'elles n'ont éventuellement pas le temps de repérer des erreurs parce qu'elles essaient de construire une représentation mentale de la phrase. Nous allons faire cela en utilisant une tâche de type en-ligne sous forme de lecture segmentée, afin de tester les connaissances implicites. Dans une première expérience, nous tenterons de déterminer si les personnes non natives remarquent l'absence d'un connecteur qui est nécessaire pour que la phrase soit cohérente. Ensuite, nous testerons si les personnes non-natives remarquent également l'incohérence lorsque celle-ci est due à l'utilisation incorrecte d'un connecteur. Ensemble, ces deux expériences nous

donneront un aperçu important de la maitrise des connecteurs dans une langue seconde lorsqu'on fait appel à des connaissances implicites.

La question se pose ensuite, au chapitre 7, de savoir si le *type* de connecteur a également un impact sur la lecture en L2. Plus précisément, nous comparerons les temps de lecture pour des phrases qui contiennent des connecteurs qui indiquent la même relation de cohérence, mais qui se distinguent par leur fréquence et du point de vue du registre de langue dans lequel ils sont typiquement utilisés. Dans ce chapitre, nous testerons à la fois les connaissances implicites et explicites des personnes interrogées.

Au chapitre 8, nous examinerons plus en détail si une utilisation incorrecte d'un connecteur est détectée lorsqu'il y a un connecteur dans la L1 qui est apparemment équivalent et correct. En effet, les théories qui prévoient un transfert négatif de la langue maternelle nous permettent de prédire que ces erreurs devraient être plus difficiles à détecter. Dans ce chapitre, nous allons examiner si la capacité à détecter les emplois incorrects dépend aussi de la tâche utilisée : dans une première tâche, les participant-e-s liront des phrases afin de répondre à des questions de compréhension et dans la deuxième tâche, les participant-e-s évalueront consciemment les phrases. Mises ensemble, ces deux tâches nous permettront de déterminer si les personnes non-natives arrivent à détecter des erreurs malgré un transfert négatif lorsqu'elles accèdent à leurs connaissances implicites et à leurs connaissances explicites.

Dans le dernier chapitre expérimental, le chapitre 9, nous nous pencherons sur la question de savoir si les personnes non-natives sont sensibles à des contraintes différentes entre des connecteurs de sens proche. Nous partons du constat que bien que certains connecteurs indiquent la même relation, ils ne sont pas toujours interchangeables, car chacun d'entre eux a un profil spécifique qui détermine quand il peut être utilisé. Par exemple, le connecteur *puisque*, contrairement à *car* et *parce que*, est utilisé pour introduire des informations qui sont déjà connues du destinataire (Groupe λ-l, 1975).

Parmi les propriétés qui font d'un certain connecteur la seule option possible dans une phrase, on peut distinguer entre des contraintes *syntaxiques* (p.ex., utilisation dans une structure clivée avec *c'est*) et *pragmatiques* (p.ex., la propriété mentionnée de *puisque* d'indiquer une information déjà connue). Dans notre chapitre, nous nous demandons si les personnes non-natives arrivent à identifier

ces caractéristiques spécifiques des connecteurs. En utilisant une tâche d'insertion, nous tenterons de déterminer si les personnes non-natives identifient mieux les contraintes syntaxiques que les contraintes pragmatiques, et s'il existe des contraintes pragmatiques plus faciles à reconnaitre que d'autres.

En conclusion, nous synthétiserons et discuterons les observations faites dans les six chapitres expérimentaux et évoquerons des pistes de recherche pour des travaux futurs.

1.4 Considérations théoriques

Il convient de noter que le cadre théorique choisi se composera de plusieurs approches. Par exemple, nous concevrons le discours et les relations de cohérence en suivant l'approche cognitive des relations de cohérence, la CCR (Sanders et al., 1992; voir 2.2.2). Le cadre de la CCR représente selon nous le meilleur cadre pour une étude qui investigue les facteurs linguistiques et leur impact sur le traitement mental, car les relations y sont définies par des entités cognitives qui diffèrent en complexité. La CCR tient également compte du fait qu'une relation de cohérence peut être marquée par différents connecteurs. La question de savoir quel connecteur choisir constitue une difficulté supplémentaire pour les apprenant-e-s, comme nous allons le voir au chapitre 9.

Quant aux relations de cohérence, nous allons utiliser pour désigner des étiquettes fréquentes, telle qu'une *concession* au lieu d'une *conséquence-cause-contrastive* comme le prévoit la CCR (selon Sanders et al., 1992) ou une *conséquence* au lieu d'une *cause-conséquence*. Indépendamment de la terminologie, les exemples choisis suffiront pour illustrer clairement la relation visée. De plus, les différences subtiles dans la conception des relations de cohérence, qui conduisent les différents cadres théoriques à différencier des sous-types de relations très fins, sont négligeables pour une étude sur les personnes non-natives. Nous nous intéressons plutôt aux différences plus importantes entre les relations majeures comme une *concession,* une *cause* ou une *conséquence* (voir aussi Halliday & Hasan, 1976).

Pour pouvoir classifier les connecteurs, nous utiliserons les classifications du dictionnaire LexConn (voir chapitre 1.5.1) qui se basent sur un autre modèle du discours que la CCR, à savoir la SDRT (voir 2.2.1). Cela n'est toutefois pas

problématique, car les différences théoriques entre la CCR et la SDRT ne sont pas pertinentes pour notre projet.[2] De plus, nous ne disposons pas tout simplement d'une classification des connecteurs qui se base sur la CCR, ce qui nous a conduit à choisir LexConn comme dictionnaire de référence.

Bien que le cadre théorique de notre projet ne paraisse donc en détail pas forcément uniforme concernant la taxonomie des relations de cohérence, il est tout de même globalement cohérent et plus important, permet une évaluation de la thématique en question.

1.5 Considérations méthodologiques

Le sujet des connecteurs de discours a souvent été abordé par le biais d'analyses de corpus (p. ex., Field & Yip, 1992; Granger & Tyson, 1996; Bolton, Nelson & Hung 2002). Cependant, dans la présente thèse, un paradigme expérimental est adopté (comme dans p.ex., Degand & Sanders, 2002; Zufferey et al. 2015).[3]

Il est certes vrai que les études de corpus ont pour avantage d'analyser des occurrences authentiques d'un phénomène de langue alors que la recherche expérimentale a, pour des raisons de contrôle, besoin des phrases artificiellement construites. Mais il est vrai aussi que les approches expérimentales ont des avantages importants. Zufferey (2020) élabore les avantages de la recherche expérimentale par rapport à la recherche de corpus.

Tout d'abord, une approche expérimentale permet, grâce à une manipulation de variables, de tester une hypothèse concernant la causalité entre deux variables.

Ensuite, une expérience permet également une investigation d'un phénomène linguistique même si celui-ci est rare. Par exemple, si nous nous intéressons à la question de savoir comment un connecteur tel que *quand bien même que* est maitrisé par les apprenant-e-s, nous ne pourrons pas savoir si l'absence de ce connecteur dans les productions de texte est dû à sa rareté en général ou parce que les apprenant-e-s ne le maitrisent pas. En revanche, en

[2] En revanche, ce qui sera plus important c'est la distinction concernant la *continuité* des relations (Segal et al., 1991). Cette distinction peut être faite aussi bien dans la CCR que dans la SDRT.

[3] Cependant, des brèves études de corpus seront de temps en temps réalisées, par exemple pour déterminer la fréquence d'un connecteur ou pour étudier et discuter certaines utilisations d'un connecteur.

testant explicitement les connaissances pour ce connecteur, nous pouvons déterminer s'il est maitrisé ou non.

Et finalement, une approche expérimentale permet d'observer des corrélations entre les différentes mesures. Par exemple, une expérience peut détecter une corrélation entre la compétence linguistique en général et la maitrise des connecteurs.

Il convient aussi de noter qu'une partie des expériences n'examine pas la maitrise des connecteurs en termes de *production* ou de *compréhension* mais plutôt en termes de détection d'une mauvaise utilisation d'un connecteur. Cela peut surprendre vu que les connecteurs sont correctement utilisés dans la plupart des discours que nous rencontrons dans la vie quotidienne. Cependant, une investigation de la sensibilité à l'incohérence due à une mauvaise utilisation (ou absence) d'un connecteur a plusieurs avantages.

Premièrement, la capacité à détecter une mauvaise utilisation d'un connecteur reflète une partie importante de ce qui constitue la maitrise des connecteurs. En examinant cette capacité, nous pouvons tirer des conclusions sur la sensibilité aux connecteurs dans plusieurs contextes. Nous allons voir, par exemple, si les personnes non-natives sont sensibles aux absences de connecteurs qui rendent le discours incohérent (chapitre 6). Cette sensibilité est également importante pour la rédaction d'un texte, car une personne non-native doit décider quand une relation doit être marquée explicitement et quand la relation peut être reconstruite par inférence. Nous allons également examiner si la capacité à détecter des emplois incorrects des connecteurs dépend aussi du type de connaissances auxquelles les apprenants peuvent avoir accès, à savoir leurs connaissances explicites ou implicites (chapitres 8 et 9).

Deuxièmement, notre approche nous permettra de faire de nouvelles observations sur les aspects cognitifs de la lecture dans une L2. En utilisant une tâche de lecture segmentée, nous pourrons observer, par exemple, quand les personnes non-natives activent la signification d'un connecteur et quand elles réalisent que l'instruction donnée par le connecteur crée de l'incohérence.

Troisièmement, la question de savoir comment les personnes non-natives détectent des utilisations incorrectes des connecteurs est aussi intéressante dans une perspective pédagogique. Dans beaucoup de cas, les enseignant-e-s d'une L2 sont également des personnes non-natives (p.ex., Agudo, 2017). Comprendre

comment et dans quels cas une personne non-native arrive mieux à détecter un faux emploi d'un connecteur aide ainsi à déterminer les circonstances qui entravent ou favorisent la correction des enseignant-e-s non-natifs d'une L2.

1.6 Remarques préliminaires

- Ce travail a été soutenu par le *Fonds National Suisse de la Recherche Scientifique* (FNS) pour le projet *Discovering Discourse: The acquisition of connectives in L1 and L2* (subsides N° 100012_184882 et 10012_184882_2) de la Professeure Dr. Sandrine Zufferey (Université de Berne).

- Certains des chapitres contiennent des données et résultats qui ont été publiés sous forme différente dans des journaux scientifiques. Toutes les informations concernant ces publications sont indiquées au début de chaque chapitre. Comme ces articles et chapitres de livre contiennent des approches et angles différents par rapport aux phénomènes examinés, les lecteur-trice-s sont également invité-e-s à consulter ces publications. Bien que ces publications impliquent des co-auteur-trice-s, la réalisation des études, l'analyse des données et la rédaction des articles a été principalement effectuée par moi-même.

- Pour les chapitres expérimentaux, tous les participant-e-s ont donné leur consentement à participer. Les participant-e-s ont été informé-e-s de leurs tâches, des avantages et des inconvénients de leur participation, de leurs droits (par exemple, demander la suppression de toutes les données sans justification) ainsi que de la protection de leurs données.

- Les études menées ont été menées conformément à la Déclaration d'Helsinki (et ses amendements ultérieurs ou à des normes éthiques comparables), et le protocole éthique a été approuvé par le *Fonds National Suisse de la Recherche Scientifique* (100012_184882).

2 Les différents modèles de discours et le rôle des connecteurs

À la fin des années 1980, les linguistes Mann et Thompson ont reçu une lettre de demande de collecte de fonds d'une société privée concernée par la surpopulation aux Etats-Unis. Les chercheur-se-s ont décidé de la transmettre à leurs collègues en leur demandant d'analyser linguistiquement son contenu. Les réponses, rassemblées dans un livre (Mann & Thompson, 1992), illustrent la variété des approches que l'on peut appliquer pour analyser un texte.

Dans un premier temps, tous les auteur-trice-s s'accordent à dire que la lettre (et tout discours en général) n'est pas une simple suite de phrases ou paragraphes. Ce qui définit le texte comme un tout cohérent par opposition à une suite aléatoire de phrases semble être le fait qu'il possède une structure interne et logique. Cependant, les différences émergent lorsqu'il est question d'identifier cette structure. Callow et Callow (1992), par exemple, se concentrent dans leur analyse principalement sur l'objectif du texte plutôt que sur sa forme et son organisation linguistique. Ainsi, ils définissent la structure comme une « distillation sous forme verbale des pensées et des perspectives de la personne communicatrice » (Callow & Callow, 1992 : 5, ma traduction). Meyer (1992), en revanche, s'intéresse plutôt à la hiérarchie du contenu en analysant l'organisation logique de l'argumentation d'un niveau local jusqu'à un niveau global.

Prises ensemble, les différentes approches recensées dans le livre de Mann et Thompson (1992) montrent à quel point l'analyse d'un discours représente un champ « vaste et ambigu » (Schiffrin, 1987 : 1, ma traduction). Une raison en est que les conceptions de ce qui constitue un discours et les modèles sous-jacents varient considérablement d'une approche à l'autre. La littérature connait ainsi des modèles de discours qui se caractérisent par une vision mathématique et statique (Montague, 1974), une vision purement pragmatique (Grosz & Sidner, 1986), une vision lexico-sémantique (Kamp, 1981) ou encore par une vision cognitive (Sanders et al., 1992) du discours.

Dans la première partie de ce chapitre, nous allons brièvement parcourir quelques-unes de ces approches théoriques en mettant l'accent sur les cadres

théoriques les plus pertinents pour notre thèse. Ce faisant, nous allons discuter quelques concepts clés du discours tels que la *cohérence*, la *cohésion* et les *relations de cohérence*.

Partant des modèles de discours présentés dans la première partie, nous présenterons dans une deuxième partie les moyens linguistiques qui assurent la cohérence d'un texte. A cet égard, nous mettrons spécialement l'accent sur l'importance et les caractéristiques des connecteurs discursifs. Pour finir, nous discuterons également de leurs fonctions procédurales et pragmatiques.

2.1 Le point commun des modèles de discours

Avant de nous focaliser sur les différences entre les modèles de discours, il convient d'aborder leurs points communs. Pour ce faire, revenons au livre de Mann et Thompson (1992), mentionné dans l'introduction.

Dans toutes les analyses qui y sont rassemblées, il est frappant de constater qu'il existe des points communs sur la question de savoir ce qui constitue un discours. Selon Mann et Thompson (voir aussi Busquets, Vieu & Asher, 2001), il s'agit de la prémisse selon laquelle la structure d'un texte « a un rapport avec les relations entre les parties constituantes » (Mann & Thompson, 1992 : 3, ma traduction), en d'autres termes qu'il existe des liens logiques s'articulant entre les différents segments d'un discours. A titre d'illustration, considérons l'exemple (1) (traduit de Halliday & Hasan, 1976 : 8).

1) Marie a promis d'envoyer une photo de ses enfants, mais elle ne l'a pas fait.

On peut y distinguer deux propositions (*la promesse de Marie* et *son inactivité*) qui sont logiquement liées entre elles. Ce rapport entre les deux propositions – leur *cohérence* (Blakemore, 1987) – structure l'énoncé et lui donne un sens global.

A cet égard, il convient de noter que deux niveaux de cohérence peuvent être distingués. Le premier, exemplifié en (1), est celui de la cohérence *locale*. Elle se produit « lorsque l'information textuelle est liée sans ambiguïté aux segments de texte qui la précèdent et la suivent » (Lehmann & Schraw, 2002, ma traduction). La cohérence *globale*, en revanche, concerne l'ensemble d'un texte et permet aux personnes lectrices d'intégrer les idées proposées dans une représentation mentale complète (Kendeou, van den Broek, Helder & Karlsson, 2014). Ce modèle

mental (Khemlani, Byrne & Johnson-Laird, 2018) ou modèle de situation (Kintsch, 1998) d'un texte permet, par exemple, de « raconter l'histoire, appliquer les connaissances acquises dans le texte ou d'identifier le thème » (Sterpin et al., 2021 : 169, ma traduction).

Dans plusieurs modèles théoriques, les liens logiques qui établissent la cohérence (locale ou globale) d'un texte sont appelés les *relations de cohérence* (p.ex., Halliday & Hasan, 1976; Hobbs, 1979; Sanders et al., 1992; Callow & Callow, 1992). Pour plusieurs chercheur-se-s (p.ex., Sanders & Spooren, 2001; Cain & Nash, 2011), ces relations représentent « les pierres angulaires de la compréhension » (Graesser, Singer & Trabasso, 2003 : 82, ma traduction).

Comme le montrent les exemples (2) à (5), il existe différents types de relations de cohérence qui peuvent être établies entre deux segments d'un discours.

2) Marie va à l'école mais Pierre n'est pas là.
3) Marie va à l'école parce que Pierre n'est pas là.
4) Marie va à l'école même si Pierre n'est pas là.
5) Marie va à l'école tandis que Pierre n'est pas là.

La littérature n'est pas univoque lorsqu'il s'agit de déterminer le nombre exact des relations de cohérence (Hovy, 1990). Ainsi, Grimes (1975) indique que « chaque linguiste qui a écrit sur les relations a dressé une liste différente de ce que sont les relations [de cohérence] » (1975 : 11, ma traduction) – avant de donner son propre classement qui « diffère légèrement de tous les autres que je connais » (Grimes, 1975 : 119, ma traduction).

Le fait qu'il n'y ait pas d'unanimité entre des différents modèles de discours n'est toutefois pas surprenant. Selon Halliday et Hasan (1976), il ne peut y avoir un inventaire unique, car « différentes classifications sont possibles, chacune d'entre elles mettant en évidence différents aspects des faits » (Halliday & Hasan, 1976 : 238, ma traduction). Par conséquent, il existe de grandes différences entre les différents modèles de discours, comme nous allons le voir maintenant.

2.2 Les différences entre les modèles de discours

Depuis les approches rhétoriques classiques de Socrate ou d'Aristote (voir Grimes, 1975; Hovy, 1990), beaucoup de modèles de discours ont émergé notamment dans

la deuxième moitié du xxe siècle (entre autres : Jackendoff, 1972; Grimes 1975; Heim, 1982; Mckeown, 1982, Longacre, 1983; Dahlgren, 1988; Polanyi, 1988; Webber, 1988). Dans ces modèles, les relations de cohérence ont été abordées à partir de cadres théoriques différents, comme par exemple la sémantique formelle (p.ex., Montague, 1974), la philosophie du langage (p.ex., Toulmin, 1958), la pragmatique (p.ex., Grosz & Snider), l'analyse du discours (p.ex., Mann & Thompson, 1988), la psycholinguistique (p.ex., Sanders et al., 1992) ou encore la pragmatique (p.ex., Blakemore, 1978). C'est pourquoi le nombre de relations proposées entre les différents modèles du discours varie de deux à plus de 100 (Knott & Sanders, 1998). En plus du fait que le nombre de relations de cohérence ne soit pas standardisé dans la littérature, la terminologie n'est pas non plus uniforme. Par exemple, tandis que Grimes (1975) nomme les relations s'articulant dans un discours des *prédicats rhétoriques* (voir aussi Meyer, 1975), Grosz & Sidner (1986) et Asher & Lasacrides (2003) utilisent le terme de *relations rhétoriques*. Dans d'autres modèles, ces relations sont nommées des *relations séquentielles* (Fillmore, 1974), *logiques* (Halliday & Hasan, 1974) ou encore des *propositions relationnelles* (Mann & Thompson, 1986; Grimes, 1975).

La variété des termes utilisés indique que les relations de cohérence ne sont pas définies de la même manière dans tous ces modèles. Par exemple, tandis que Longacre (1979) discute différents *types de paragraphes*, les *prédicats rhétoriques* de Meyer (1975) incluent également les liens qui peuvent exister au sein d'une clause simple – ce qui les différencient également des *relations de cohérence* de Sanders et al. (1992, voir aussi Hobbs, 1979; Callow & Callow, 1992).

Selon Busquets et al. (2001) les différents modèles proposés se distinguent sur trois questions principales :

1. Quel est le nombre de relations de cohérence (et ce nombre peut-il être établi avec précision ?)
2. Existe-t-il des types de discours différents ?
3. Comment un discours peut-il être segmenté ?

Afin d'exemplifier ces différences, il convient de présenter quelques-uns de ces modèles de manière détaillée. L'accent sera mis ensuite sur l'approche cognitive de Sanders et al. (1992), car c'est ce modèle que nous utiliserons dans les chapitres empiriques de notre travail.

2.2.1 Aperçu de quelques modèles de discours

L'approche de Montague (1970, voir aussi 1974) représente un modèle influent (Lamillar & Godart-Wendling, 2002) de représentation du discours. Ce modèle adopte une vision mathématique et statique du langage. Bien qu'il s'agisse d'une approche formelle (c'est-à-dire une approche qui traduit les éléments du discours en formules algébriques et en suivant les règles de la logique), elle tente également de tenir compte du contexte et des inférences que les personnes lectrices doivent faire afin d'interpréter un discours. Ainsi, le sens d'une phrase est compris de manière *compositionnelle*, c'est-à-dire en tant qu'une « fonction des significations de ses parties et de la façon dont elles sont combinées syntaxiquement » (Partee & Hendriks, 1997 : 21, ma traduction). En d'autres termes, il se compose d'un continuum entre des entités sémantiques minimales (c'est-à-dire des mots) jusqu'à un système contextuel. Lamillar et Godart-Wendling (2002) résument :

> « (…) il s'agit donc d'établir un modèle de l'interprétation du langage naturel qui mette en œuvre des tâches de raisonnement à la fois déductives et naturelles, afin de pouvoir expliquer certains phénomènes syntaxiques et sémantiques (tels que l'anaphore) en termes de processus logique déductif » (Lamillar & Godart-Wendling, 2002 : 82)

Cependant, cette conception « se heurtait à certains problèmes d'interprétation des pronoms au-delà des limites de la phrase, et en particulier, au problème des relations anaphoriques entre les pronoms et les descriptions définies » (Busquets et al., 2001 : 5). En effet, l'approche de Montague (1974) s'intéressait plutôt à l'interprétation au niveau phrastique qu'au niveau discursif.

Par la suite, d'autres modèles sémantiques ont été développés (p.ex., Seuren, 1985; Seuren & Wekker, 1986; Groenendijk & Stokhof, 1991) dans le but d'analyser la cohérence des suites de phrases. L'une d'entre elle est la *Théorie des Représentations Discursives,* ou DRT (Discursive Representation Theory), de Kamp (1981, voir aussi Kamp, Van Genabith & Reyle, 2011). Cette approche met l'accent sur le rôle du contexte dans l'interprétation. Ce faisant, la DRT adopte, contrairement à l'approche de Montague (1974), une vision *dynamique* entre la sémantique et le contexte pour arriver à l'interprétation d'un discours (Busquets et al., 2001) :

> « L'interaction entre le contexte et l'énoncé est réciproque. Chaque énoncé contribue (via l'interprétation qui lui est donnée) au contexte dans lequel il est émis. Elle modifie le contexte en un nouveau contexte, dans lequel cette contribution est reflétée; et c'est ce nouveau contexte qui informe l'interprétation de l'énoncé suivant » (Kamp et al., 2011 : 125, ma traduction)

Ainsi, selon Gamut (1991), la DRT « réconcilie (…) la vision statique ou déclarative [du langage] avec une vision dynamique et procédurale » (Gamut, 1991 : 265, ma traduction).

Tandis que Kamp (1981) et Montague (1974) développent des modèles de discours *sémantiques* en essayant d'intégrer le contexte comme une variable supplémentaire à l'interprétation, d'autres approches mettent encore plus en avant les intentions de la personne locutrice. Comme ces modèles s'intéressent moins « aux conditions de vérité d'un discours » (Busquets et al., 2001 : 7) mais plutôt aux intentions des personnes, *l'analyse du discours* représente un outil important pour eux. La distinction entre ces deux pôles (un modèle purement sémantique vs. un modèle basé uniquement sur les intentions) n'est toutefois pas nette et il existe une gradation dans les modèles, de telle sorte que certains mettent plutôt l'accent sur l'un que sur l'autre.

Un exemple d'un tel modèle est celui de Hobbs (1985). Alors que ce modèle s'inscrit toujours plutôt dans la ligne des modèles sémantiques (Knott & Sanders, 1988), il se sert de l'analyse du discours et tient compte des intentions des personnes locutrices. Pour analyser la cohérence d'un texte, Hobbs (1985) propose quatre étapes.

Dans un premier temps, il faut découper le texte de manière intuitive en segments. Alors qu'une première segmentation divise le texte en seulement deux ou trois sections, les segmentations suivantes se poursuivent jusqu'au niveau de la phrase. Dans un deuxième temps, les segments sont étiquetés avec des relations de cohérence. Hobbs (1985) distingue 10 relations de cohérence différentes, comme par exemple une *élaboration* ou une *violation d'attente discursive*. Dans la troisième étape, on précise explicitement les inférences qui sont nécessaires pour comprendre les relations. Pour Hobbs (1985), ce sont justement les inférences qui sont importantes pour saisir les relations d'un texte :

> « Les relations de cohérence qui donnent une structure à un texte font partie de ce qu'est une interprétation; elles sont définies en termes d'inférences qui doivent être tirées pour les reconnaitre, et spécifient ainsi une connexion qui doit exister entre les interprétations et la connaissance. » (Hobbs, 1985 : 33, ma traduction)

Dans la quatrième et dernière étape de cette analyse, on valide les conclusions tirées dans la troisième étape, autrement dit on vérifie si les inférences faites dans la troisième étape représentent réellement la base du texte. Cela est réalisé en

construisant une base de connaissance commune pour tous les textes issus du même cadre culturel et ciblant le même groupe de lecteurs-trices.

Au cours de cette analyse, la complexité de la tâche augmente à chaque étape :

> « La première étape est une question de minutes pour un texte de la longueur d'un paragraphe, l'étape 2 est une question d'une heure ou deux, et l'étape 3 est une question de jours, puis l'étape 4 est une question de mois ou d'années. » (Hobbs, 1985 : 33, ma traduction)

Bien que les inférences et les intentions des personnes prennent une place importante dans la conception de Hobbs (1985), il conçoit les relations de cohérence, tout comme les unités minimales discursives, toujours comme des unités *sémantiques* et non intentionnelles (Busquets et al., 2001).

Ceci est contraire à d'autres modèles qui mettent davantage l'accent sur les intentions des personnes locutrices. L'un d'entre eux est la *Théorie de la Structure Rhétorique,* ou RST (Rhetorical Structure Theory), proposée par Mann & Thompson (1986, 1987, 1988). Cette approche met l'accent spécifiquement sur les intentions des personnes locutrices en se basant notamment sur les théories de la communication, telles que la théorie des actes de langage de Austin (1962) et Searle (1969). Très brièvement, la RST distingue les rapports entre des segments textuels définis comme des *noyaux* et des *satellites*. Un noyau est un segment discursif privilégié sur lequel les satellites, les informations plus interchangeables, s'alignent.

En ce qui concerne les relations de discours, la RST distingue entre les relations telle que la *motivation*, la *justification* ou la *preuve*, en se basant sur les intentions des personnes locutrices. Mann et Thompson notent que ces relations sont fondamentales, c'est-à-dire qu'elles ne sont pas interchangeables car chacune est unique. La RST implique également une dimension cognitive, car les personnes lectrices activent cognitivement l'interprétation d'une certaine relation lorsqu'ils rencontrent des indications dans un texte.

Cependant, Sanders et al. (1992) critiquent le fait que les relations proposées par Mann et Thompson (1986) sont, d'un point de vue cognitif, descriptives et non explicatives, autrement dit elles ne sont pas réellement fondées cognitivement. Selon Sanders et collègues, la structure d'un discours ne peut pas seulement être analysée en fonction des intentions, mais reflète des entités cognitives. Nous reviendrons sur leur approche dans la suite de ce chapitre.

Une deuxième critique concerne le fait que la RST implique l'unicité de chaque relation. Moore et Pollack (1992, cité dans Asher & Lascarides, 2003 : 103) donnent les exemples (6) et (7).

6) *George Bush supports big business.*
'George Bush soutient les grandes entreprises'

7) *He's sure to vote House Bill (1711).*
'Il est certain qu'il votera le projet de loi de la Chambre (1711).'

Si l'on comprend, selon ce que prévoit la RST, (6) comme le noyau de (7), la relation est une *cause volitionnelle*. Par contre, si l'on interprète (7) comme le noyau de (6) la relation serait de type *preuve*. Cependant, « la contrainte d'unicité dans la RST empêche les deux relations d'exister simultanément » (Asher & Lascarides, 2003 : 103).

Selon Busquets et al. (2001), cette contrainte pose aussi des problèmes lorsqu'on peut identifier deux relations de cohérence simultanément dans un énoncé. Considérons (8), tiré de Busquets et al. (2001 : 11) qui contient à la fois une relation de *narration* et de *contraste*.

8) Jean a donné un livre à Marie, mais il le lui a ensuite repris.

Selon la RST, il faudrait « inférer la relation intentionnelle à partir de la relation sémantique, ou *vice versa* » ce qui n'est pas possible car « il n'y a pas de correspondance entre elles » (Busquets et al., 2001 : 11; Moore & Pollack, 1992).

Tandis que la RST comprend les relations de cohérence à la fois comme sémantiques et intentionnelles, la théorie de Grosz & Sidner (1986) adopte une approche purement pragmatique du discours. De manière générale, ils pensent que les intentions d'une personne produisant un discours peuvent être si diverses qu'une simple liste de relations de cohérence telle que celle présentée par Hobbs (1979) ne suffit pas à expliquer la cohérence d'un discours.

Grosz et Sidner (1986) définissent le discours comme « un comportement langagier qui implique généralement plusieurs énoncés et plusieurs participant-e-s » (Grosz & Sidner 1986 : 176, ma traduction). Ainsi, contrairement aux conceptions de Hobbs (1979) ou de la RST qui impliquent également les intentions des locuteurs-trice-s, l'approche de Grosz et Sidner consiste à définir les segments textuels en fonction de l'intention de la personne qui les a conçus (Busquets et al., 2001).

Selon Grosz et Sidner (1986), il ne peut y avoir de classification exhaustive de relations de cohérence (l'approche ne distingue que deux types de relations), car plus on examine les relations de cohérence en détail, plus des différences entre les relations singulières émergent, « jusqu'à ce que vous vous retrouviez sur le point d'atteindre la complexité totale de la sémantique proprement dite « (Hovy, 1993 : 19, ma traduction).

Tout comme d'autres modèles de discours, cette approche a également fait l'objet de critiques. Par exemple, les linguistes travaillant sur l'intelligence artificielle - c'est-à-dire, entre autres, l'idée de planifier et de générer du discours de manière computationnelle - soutiennent qu'il ne suffirait pas de constituer un discours uniquement par des intentions d'un-e locuteur-trice:

> « Du point de vue de l'analyse de texte, l'approche parcimonieuse semble satisfaisante. (...) Cependant, d'un point de vue de la génération d'un texte cette approche n'est pas suffisant. L'expérience pratique a montré que les planificateurs-trices de textes ne peuvent pas se contenter de considérations intentionnelles, mais qu'ils ont besoin de beaucoup plus d'informations rhétoriques et sémantiques pour construire un texte cohérent » (Hovy, 1990 : 129, ma traduction)

Pour générer un texte de manière computationnelle, il faudrait plutôt adopter une vision du discours qui distingue entre les différentes relations de cohérence.

Finalement, un dernier modèle est celui de la *Théorie de la Représentation du Discours Segmenté*, ou SDRT (Segmented Discourse Representation Theory) de Lascarides & Asher (1993, voir aussi Asher, 1993). Cette approche cherche à étendre l'approche sémantique de la DRT de Kamp (1981) en ajoutant le niveau pragmatique du discours (Busquets et al., 2001). Les segments textuels se constituent de manière sémantique et peuvent contenir, contrairement à la RST, plusieurs relations de cohérence à la fois. La SDRT met spécialement l'accent sur la structure d'un discours, c'est-à-dire sur la hiérarchie entre les différentes relations de cohérence. Le modèle distingue entre les relations *subordonnées* (relations entre deux segments dont un est plus important que l'autre) et les relations *coordonnées* (relations entre deux segments ayant la même importance). Par exemple, tandis que les relations d'une *élaboration* (9) ou *explication* (10) sont des relations subordonnées, une *narration* (11) est une relation coordonnée (Roze, Danlos & Muller, 2012) (tous les exemples tirés et traduits de Asher et Lascarides, 1993 : 159, 162).

9) Max est tombé. John l'a poussé.

10) Alexis a très bien réussi à l'école cette année. Elle a eu les meilleures notes dans toutes les matières.
11) Max est tombé. John l'a aidé à se relever.

Ces relations sont interprétées à partir d'inférences. Ainsi, la SDRT tient également compte des informations non linguistiques, telles que les connaissances du monde, comme points de repère pour une interprétation d'un texte.

En conclusion, nous avons vu à travers de brèves présentations de plusieurs modèles discursifs que ceux-ci se distinguent les uns des autres selon plusieurs critères. Tandis que certains modèles mettent l'accent davantage sur la sémantique, en développant par exemple un cadre de sémantique formelle (Montague, 1974) ou de sémantique dynamique (Hobbs, 1985), d'autres mettent en avant les intentions des personnes (Grosz & Sidner, 1986) pour pouvoir décrire un discours. A la section suivante, nous allons présenter une autre approche encore un peu plus en détail, à savoir la taxonomie des relations de cohérence en primitives cognitives, proposée par Sanders et al. (1992).

A notre avis, cette approche représente le meilleur cadre théorique pour une approche psycholinguistique sur la lecture dans une langue seconde. En effet, il existe des théories selon lesquelles la lecture dans une deuxième langue est sur le plan cognitif plus exigeante que la lecture en L1 (Clahsen & Felser, 2006a, 2006b). Par conséquent, un cadre théorique qui comprend les relations de cohérence comme des entités cognitives et qui constate des complexités différentes entre elles permet de mieux investiguer la difficulté de la lecture en L2. Ainsi, l'approche cognitive de Sanders et al., (1992) est hautement compatible avec notre analyse des difficultés des apprenant-e-s avec les connecteurs en L2.

2.2.2 L'approche cognitive des relations de cohérence

Nous avons vu que différents modèles du discours ont émergé au cours des années, souvent en renégociant la part de contenu sémantique et les intentions des locuteur-rice-s comme éléments d'interprétation des discours. Nous allons présenter dans cette section une autre approche, celle proposée par Sanders, Spooren et Noordman (1992, 1993). Il s'agit d'une approche purement cognitive des relations de cohérence qui à l'origine, contrairement à d'autres modèles de

discours, n'a pas été conçue comme une approche d'annotation de discours (Hoek, Evers-Vermeul & Sanders, 2019).

Selon la CCR (C*ognitive approach to Coherence Relations*) de Sanders et al. (1992), les relations de cohérence sont constituées d'entités cognitives. Les principes organisant un discours, les relations de cohérence, reflètent ainsi des catégories de base de la cognition humaine (Sanders, 2005). Il en résulte que les relations de cohérence diffèrent en termes de complexité cognitive. En effet, nous savons que les enfants acquérant le langage produisent certaines relations de cohérence avant d'autres (p.ex., Nippold, Schwarz & Undlin, 1992; Cain & Nash, 2011). Il semble, par exemple, que les relations additives et temporelles sont maitrisées avant les relations causales et adversatives (Bloom, Lahey, Hood, Lifter, & Fiess, 1980). Cet ordre d'acquisition s'explique par le fait que les enfants développent successivement des capacités cognitives qui leur permettent, avec le temps, de comprendre et d'exprimer des relations de plus en plus complexes. L'idée que certaines relations de cohérence sont cognitivement plus complexes que d'autres est également soutenue par les résultats d'études sur la lecture chez les adultes. Par exemple, les résultats de et Xu, Chen, Panther & Wu (2018) démontrent que les relations plus complexes, telles que les concessions, sont lues plus lentement que les relations plus simples, telle que les causes (voir aussi Black & Bern, 1981; Trabasso, Secco & Van den Broek, 1985; Trabasso & Sperry, 1985).

A l'origine, la CCR a proposé une taxonomie en quatre primitives binaires qui permettent de définir et de regrouper plus précisément les relations de cohérence. Nous allons présenter brièvement ces quatre primitives au moyen de l'exemple (12) (traduit de Sanders et al., 1992 : 16) :

12) En raison de la présence d'une zone de basse pression au-dessus de l'Irlande, le mauvais temps se dirigera vers nous.

La première primitive consiste à distinguer les relations additives ou causales, concepts de base de toute relation de cohérence. La relation entre les deux segments d'un énoncé est *additive* lorsque les deux segments se lient sans une implication entre eux. Il s'agit d'une relation *causale* si une implication peut être établie entre les deux segments. A cet égard, la phrase (12) peut être classifiée comme *causale* car une zone de basse pression implique une prévision du mauvais temps. En revanche, l'exemple (13) illustre le cas d'une relation *additive* :

13) Il y a une zone de basse pression au-dessus de l'Irlande et une zone de haute pression au-dessus de la Floride.

Dans cet exemple, deux faits (les positions d'une basse pression et d'une haute pression) sont présentés l'un après l'autre sans établir une implication entre les deux.

La seconde primitive est la distinction entre la dimension sémantique et pragmatique d'une relation. Dans une relation sémantique, les deux segments d'un discours sont liés par leur contenu propositionnel, tandis que dans une relation pragmatique, les deux segments sont liés par l'intention illocutoire (cette distinction sera discutée plus en détail à la section 2.5)

La phrase (12), par exemple, peut être classée comme sémantique, car la raison (i.e., *présence d'une zone de basse pression au-dessus de l'Irlande*) de la prévision d'un mauvais temps n'est pas établie par un point de vue subjectif ou une intention de la personne qui parle. De même pour (13), qui contient deux faits établis dans le monde. En revanche, l'exemple (14) peut être considéré comme étant pragmatique, car il nécessite que l'interlocuteur-trice active une inférence à laquelle la personne parlante fait allusion (à savoir que les nuages gris correspondent à des nuages de pluie) pour arriver à l'interprétation.

14) Le mauvais temps se dirigera vers nous, parce que je vois des nuages gris.

La troisième primitive est l'ordre des segments, c'est-à-dire le fait que l'ordre d'implication de la relation soit constitué successivement (*basique*) ou rétroactivement (*non-basique*) (Sanders, Demberg, Hoek, Scholman, Asr, Zufferey & Evers-Vermeul, 2018). Seules les relations causales (qui contiennent une implication entre deux segments) peuvent être distinguées selon cette primitive, car les relations additives « sont symétriques par définition » (Sanders et al., 2018 : 13, ma traduction).

Par exemple, la phrase (15) est basique car le déroulement temporel ainsi que la progression logique suivent la forme : $p \rightarrow q$. En revanche, (16) est non-basique car les faits sont représentés selon la formule $p \leftarrow q$.

15) Il y a une zone de basse pression au-dessus de l'Irlande, donc du mauvais temps se dirigera vers nous.

16) Du mauvais temps se dirigera vers nous, car il y a une zone de basse pression au-dessus de l'Irlande.

La quatrième primitive est la polarité, qui sépare les relations avec un caractère négatif ou positif. Dans les exemples (15) et (16), la polarité est positive. Si la phrase contenait une négation, par exemple sous forme d'une concession (17), la relation serait négative.

17) Il y a une zone de basse pression au-dessus de l'Irlande mais le mauvais temps ne se dirigera pas vers nous.

Il est à noter que le concept de polarité négative ne s'applique pas seulement aux négations explicites mais aussi aux négations implicites, c'est-à-dire aux phrases sans marques de négation comme (18), dans lesquelles on peut inférer la polarité négative de la clause grâce à la sémantique du mot *épargner*.

18) Il y a une zone de basse pression au-dessus de l'Irlande, mais nous serons épargnés par ce mauvais temps.

Ainsi, la relation de (12) (citée à nouveau en 19), peut-être classifiée comme étant *causale*, *sémantique*, *basique* et *positive* selon la taxonomie établie par Sanders et al. (1992) et correspond donc à une relation qui peut être labelisée comme *cause-conséquence*. En changeant l'ordre basique, la relation devient une *conséquence-cause,* comme en (20). En changeant la polarité, la relation de cohérence change d'une *cause-conséquence* à une *cause-conséquence contrastive*, comme en (21).

19) En raison de la présence d'une zone de basse pression au-dessus de l'Irlande, le mauvais temps se dirigera vers nous.
20) Du mauvais temps se dirigera vers nous car il y a une zone de basse pression au-dessus de l'Irlande.
21) Il y a une zone de basse pression au-dessus de l'Irlande mais le mauvais temps ne se dirigera pas vers nous.

Grace à ces quatre primitives, Sanders et al. (1992) répertorient au total douze types de relations de cohérence (12 et non 16 car l'ordre dans les relations additives n'importe pas). L'ensemble de toutes les relations de cohérence établies par Sanders et al. (1992) est reporté dans le Tableau 1.1.

Tableau 1.1. Aperçu des relations de cohérence selon la taxonomie de Sanders et al. (1992, ma traduction)

	Opération Basique	*Source de Cohérence*	*Ordre*	*Polarité*	*Relation*
1.	Causal	Sémantique	Basique	Positif	*Cause - conséquence*
2.	Causal	Sémantique	Basique	Négatif	*Cause – conséquence contrastif*
3.	Causal	Sémantique	Non-basique	Positif	*Conséquence - cause*
4.	Causal	Sémantique	Non-basique	Négatif	*Conséquence – cause contrastif*
5a.	Causal	Pragmatique	Basique	Positif	*Argument – claim*
5b.					*Instrument – objectif*
5c.					*Condition – conséquence*
6.	Causal	Pragmatique	Basique	Négatif	*Argument – claim contrastif*
7a.	Causal	Pragmatique	Non-basique	Positif	*Claim – argument*
7b.					*Goal – instrument*
7c.					*Conséquence – condition*
8.	Causal	Pragmatique	Non-basique	Négatif	*Claim – argument contrastif*
9.	Additive	Sémantique	-	Positif	*Liste*
10a.	Additive	Sémantique	-	Négatif	*Exception*
10b.					*Opposition*
11.	Additive	Pragmatique	-	Positif	*Énumération*
12.	Additive	Pragmatique	-	Négatif	*Concession*

L'approche cognitive implique également que les relations de cohérence ne correspondent pas forcément aux instructions explicites d'un connecteur. Ainsi le connecteur *et* peut indiquer à la fois une relation de *cause-conséquence* (1., voir Tableau 1.1) et une *opposition* (10b.) - et donc non seulement une *addition* (9.) « comme le suggère le connecteur » (Hoek, Evers-Vermeul & Sanders, 2019 : 3, voir aussi Crible & Pickering, 2020). Le modèle CCR permet ainsi d'observer les similitudes et les différences entre les relations de cohérence : le fait que *et* puisse indiquer soit des relations d'addition (9., 10b.), soit une relation de causalité (1.) Mais pas une relation concessive (2.) Indique une certaine proximité entre l'addition et la causalité (Sanders et al., 1992).

Plus récemment, le modèle CCR a été retravaillé par de nombreux chercheur-se-s qui ont proposé des nouvelles distinctions aux primitives de base. Evers-Vermeul, Hoek et Scholman (2017), par exemple, ont proposé une primitive de temporalité (reprise ensuite par Sanders et al., 2018). Cette nouvelle primitive distingue les relations *temporelles* et *non-temporelles*. Parmi les relations temporelles, on peut encore distinguer entre les relations chronologiques, antichronologiques ou synchrones.

A part l'ajout de la temporalité, d'autres raffinements au modèle CCR ont été proposés, comme par exemple la distinction entre relations *volitionnelles* et *non-volitionnelles* au sein des relations causales (Stukker, Sanders & Verhagen, 2008), entre relations *directes* et *indirectes* au sein des relations négatives (Pander Maat,

1998) entre relations *objectives* et *positives* au sein des relations causales (Sanders et al., 2018) ou, au sein des relations additives, la distinction entre les relations *alternatives* et *non-alternatives* (Hoek et al., 2019).

2.3 Conclusion intermédiaire

Dans la première partie de ce chapitre, nous avons vu jusqu'il existe de nombreux modèles de discours qui établissent différents critères pour la structuration logique d'un texte. Malgré toutes les différences entre les modèles, les classifications théoriques s'accordent à dire qu'un discours se caractérise par une structure logique, la cohérence. A cet égard, un élément central sont les relations de cohérence, donc des rapports logiques entre les segments textuels.

Nous allons maintenant examiner avec quels moyens linguistiques la structuration d'un texte, notamment les relations de cohérence, peut être marquée explicitement.

2.4 Moyens linguistiques pour établir la cohésion et cohérence textuelle

Pour arriver à un texte cohérent, il convient de signaler la nature des relations entre segments, en utilisant des marques de *cohésion*.

En reconsidérant l'exemple de Halliday et Hasan (1976) présenté en (22), nous observons l'emploi d'une marque de cohésion : le pronom *elle,* qui indique qu'il s'agit de la même personne dans les deux clauses.

22) Marie a promis d'envoyer une photo des enfants, mais elle ne l'a pas fait.

Il convient de noter que l'utilisation des marques de cohésion n'est pas toujours nécessaire pour établir la cohérence, car il est aussi possible de comprendre le rapport entre des phrases sans marquage explicite. Par exemple, c'est grâce aux inférences il est possible de relier les deux propositions (23) et (24), même si aucune marque de cohésion n'est présente (Mann & Thompson, 1988; Hobbs, 1990).

23) Marc a oublié sa veste.
24) Il fait froid.

Il n'est donc pas surprenant que le contexte d'un énoncé puisse faciliter l'identification des relations de cohérence (Song, 2010). Nous reviendrons sur cette question à la section 1.5.3.

Malgré le fait que les marques de cohésion ne sont pas nécessaires dans certains cas, la question se pose de savoir quelles sont ces marques. Une liste des marques de cohésion est fournie par Halliday et Hasan (1976), comme nous allons le voir dans la section suivante.

2.4.1 Les marques cohésives selon Halliday et Hasan (1976)

Dans leur livre, Halliday et Hasan (1976) distinguent cinq catégories d'éléments qui contribuent à la cohésion d'un texte.

La première catégorie est la cohésion par *référence*. Elle comprend les références personnelles (c'est-à-dire les pronoms possessifs et relatifs comme *son*), démonstratives (c'est-à-dire les mots démonstratifs comme *celle-ci*) et comparatives (c'est-à-dire les mots comparatifs comme *plus*). La cohésion s'établit en faisant référence à un concept déjà mentionné au moyen d'un élément linguistique spécifique (voir aussi : Hobbs, 1979; Grimes, 1975; Johnson & Arnold, 2022).

La deuxième catégorie, la cohésion par *substitution*, consiste à remplacer un élément par un autre. Contrairement à la *référence*, la *substitution* constitue « une relation plutôt dans les mots que dans le sens » (Halliday & Hasan, 1976 : 88), en d'autres termes, un remplacement plutôt grammatical que sémantique. La substitution comprend également trois sous-catégories, à savoir la substitution nominale (25), verbale (26) et phrastique (27) (tous les exemples sont traduits de Halliday & Hasan, 1976 : 92, 114, 130).

25) A : Ces cookies sont périmés.
B : Achètes-<u>en des frais.</u>

26) A : Ont-ils enlevé leurs meubles ?
B : Ils <u>ont fait</u> les tables, mais c'est tout pour l'instant.

27) A : Est-ce qu'il va y avoir un tremblement de terre ?
B : <u>C'est annoncé</u>.

La troisième catégorie est constituée d'ellipses qui peuvent être soit nominales, soit verbales, soit phrastiques. Plus précisément, ce sont les phrases « dont la structure

est telle qu'elle présuppose un élément précédent, qui sert de source à l'information manquante » (Halliday & Hasan, 1976; 143, ma traduction).

La quatrième catégorie est celle des *conjonctions* ou connecteurs de discours, une catégorie que nous allons approfondir davantage dans la section suivante.

La dernière façon d'assurer ou d'améliorer la cohésion d'un texte selon Halliday et Hasan (1976) se produit au niveau lexical, par exemple en répétant les mots ou en utilisant des synonymes (p. ex. *Maison* - *logement*), des antonymes (p. ex. *grand* - *petit*), des hyponymes (p. ex. *arbre* - *boulot*) ou des collocations (*tomber malade*).

La classification proposée par Halliday et Hasan (1976) représente une contribution importante dans le champ de la linguistique de discours mais elle a également suscité des critiques (voir p.ex. Huddleston, 1978; Mosenthal & Tierney, 1984). Par exemple, Doyle (1982) argumente que la conception proposée par Hasan et Halliday (1976) reste à un niveau superficiel de la cohésion, car elle ne fournit pas d'explications suffisantes concernant les processus cognitifs sous-jacents à la rédaction et à la lecture de textes. En d'autres termes, Halliday et Hasan (1976) définissent un discours non en fonction de sa structure cognitive, mais plutôt comme une *unité d'organisation situationnelle-sémantique* (Dressler, 1978 : 677). Doyle explique (1982) :

> « En limitant leur préoccupation aux textes après leur production et en marquant la phrase comme une unité dans la cohésion textuelle, Halliday et Hasan se limitent à une discussion du sens tel qu'il apparait dans la structure de surface; les questions de cohérence, des relations entre les propositions dans le monde textuel créé par l'auteur-e et recréé par la personne lectrice, restent non examinées. » (Doyle, 1982:390, ma traduction)

Comme Halliday et Hasan se limitent au niveau textuel, leur théorie n'apporte que très peu aux questions cognitives de la création du discours et n'est utile que pour décrire la structure cohésive d'un texte une fois qu'elle est établie. Ainsi, la théorie a un caractère descriptif plutôt qu'explicatif (Doyle, 1982 : 393).

Depuis la catégorisation des différentes marques de cohésion produite par Halliday et Hasan (1976), la littérature a examiné encore d'autres moyens de signaler des rapports logiques entre des segments de texte. Il convient de noter à cet égard la distinction faite plus récemment par Spooren et Sanders (2008), à savoir la distinction entre la cohérence *référentielle* et *relationnelle*. Tandis que la cohérence référentielle se caractérise par la référence au même référent par des

unités linguistiques simples (comme les catégories de la *référence* ou *substitution* de Halliday & Hasan, 1976), les marques d'une cohérence relationnelle indiquent des relations de cohérence (comme les *conjonctions* dans la conception de Halliday & Hasan, 1976). Nous allons maintenant présenter quelques marques *relationnelles*.

2.4.2 Les marques relationnelles

D'abord, il convient de noter que certaines indications des relations de cohérence existent uniquement à l'écrit, tel que les titres (Christofalos, Raney, Daniel & Demos, 2020), alors que d'autres se limitent à la langue parlée, tels que la prosodie (Tyler, 2014) ou les gestes (Laparle, 2021). Mais il y a aussi des cas où ces signaux peuvent apparaitre dans les deux modes. Bien qu'étudiée seulement à l'écrit jusqu'à présent, un tel cas pourrait être la structure de parallélisme, c'est-à-dire une répétition d'un élément de la phrase, comme en (28).

28) Nick <u>fait les</u> gâteaux les plus délicieux. Grace <u>fait les</u> desserts les plus savoureux.

Il a été démontré qu'au niveau inter-phrastique, les structures parallèles facilitent l'identification d'une relation contrastive (Crible & Pickering, 2020; Crible, Wetzel & Zufferey, 2021; exemple traduit de Crible & Pickering, 2020 : 3).

Les relations de cohérence peuvent également être indiquées à un niveau plus restreint. En utilisant des mots comme *either* en anglais 'soit' (Staub & Clifton, 2006) ou encore *on the one hand* 'd'un côté' (Scholman, Rhode & Demberg, 2017), le locuteur-trice guide ses lecteur-trice-s à attendre une certaine suite du discours. Un autre exemple provient des mots comme *quelques* ou *certains* qui activent l'attente que la suite du discours contiendra une liste ou une énumération, comme en (29) (Scholman, Demberg & Sanders, 2020, ma traduction).

29) L'adolescent a eu des désaccords avec <u>quelques</u> personnes hier. Il s'est disputé avec sa mère…

Outre ces signaux de listes, il existe aussi certains verbes qui créent des attentes spécifiques concernant la suite du discours (Kehler, Kertz, Rohde & Elman, 2008; Rhode & Horton, 2014; Koornef & Sanders, 2013). La polarité du verbe semble également jouer un rôle à cet égard : lorsque le verbe a une polarité négative (c'est-à-dire qu'il contient des marqueurs de négation, voir exemple 30) au sein des

relations concessives, le traitement mental semble être plus facile que lorsque le verbe a une polarité positive (voir exemple 31) (Crible, 2021).

30) Christine n'est pas en bonne santé, mais elle va aller travailler.
31) Christine est malade mais elle va aller travailler.

Finalement, des travaux ont montré qu'une relation de cohérence est souvent marquée par plusieurs signaux à la fois, en particulier pour certains types de relation. En annotant les marques et signaux indiquant des relations de cohérence dans le *Discourse Treebank Corpus* (Prasad, Dinesh, Lee, Miltsakaki, Robaldo, Joshi & Webber, 2008). Das et Taboada (2018) ont trouvé, par exemple, que les relations additives sont plus souvent marquées explicitement par plusieurs marques que les relations causales. Parmi les signaux examinés se trouvent par exemple des signaux de ponctuation (comme des virgules, points virgules ou les parenthèses) ou encore des marques syntaxiques.

Bien que tous ces exemples démontrent l'éventail des moyens permettant d'établir la cohésion textuelle, des travaux expérimentaux manquent encore afin de déterminer comment ces marques influencent la lecture (Crible & Pickering, 2020).

Dans la section suivante, nous présenterons plus en détail la catégorie des *connecteurs discursifs*. Cette catégorie est considérée comme étant *prototypique* (Crible, 2022 : 2) pour marquer des relations de cohérence, à tel point que tous les autres signaux de cohésion sont appelés des signaux *alternatifs* ou « des signaux de lexicalisation *alternative* » (Crible & Demberg, 2020, ma traduction) et qu'une phrase qui ne contient pas de connecteur – bien qu'elle soit marquée d'une autre manière – est considérée comme étant *implicite* (Crible & Demberg, 2020).

2.5 Les connecteurs discursifs

Les connecteurs de discours sont des éléments linguistiques qui indiquent des informations comment on doit interpréter la relation de cohérence entre deux segments d'un discours. Ce faisant, ils contribuent à la cohésion d'un texte et assurent sa cohérence. Considérons les exemples (32) – (34).

32) Nadia sort souvent courir, *alors* elle reste en bonne santé.
33) Fabien n'aime pas le sport, *mais* il en fait quand même pour être en bonne santé.
34) Matthieu va sûrement venir, *pourvu que* son train soit à l'heure.

Dans ces exemples, les connecteurs utilisés (*alors*, *mais*, *pourvu que*) indiquent des relations de cohérence d'une *cause* (32), d'une *concession* (33) et d'une *condition* (34). Comme les connecteurs ne font pas partie du contenu propositionnel d'un énoncé, cette catégorie est proche de celle des *marqueurs de discours*. Selon Cuenca (2013) une différence entre ces deux catégories consiste dans le fait que les connecteurs établissent une connexion entre deux segments, tandis que les marqueurs du discours (ou marqueurs modaux) modifient ou influencent l'illocution[4]. Considérons les mots *alors* et *donc* qui, d'une part, peuvent servir comme des connecteurs de discours lorsqu'ils indiquent une relation de conséquence, comme illustré dans les exemples (32) et (33).

35) Je veux faire du camping *alors* j'ai besoin d'une tente.
36) Je veux faire du camping, j'ai *donc* besoin d'une tente.

D'autre part, ces deux mots peuvent également être utilisés comme marqueurs de discours lorsqu'ils indiquent une prise-parole (32) ou une conclusion (33).

37) *Alors*, qu'est-ce qu'on fait ce soir ?
38) *Donc*, tu ne viens pas finalement ?

Nous voyons que *alors* et *donc* ne servent pas comme connecteurs de discours dans les exemples (37) et (38) car ils n'y rendent pas une relation de cohérence explicite. Cependant, comme précise Cuenca (2003) la frontière entre ce qui est un *connecteur* (c'est-à-dire un élément indiquant une relation de cohérence) et un *marqueur de discours*[5] (c'est-à-dire un élément influençant l'énoncé d'une manière ou d'une autre) n'est pas toujours aussi nette. Par exemple, il est connu que le connecteur français *puisque* ne peut non seulement indiquer une relation d'une *conséquence-cause*, mais également une attitude dissociative envers le contenu (Schumann, Zufferey & Oswald, 2020). Nous allons revenir sur ce point au chapitre 1.5.4.

Il est à noter que nous avons choisi d'utiliser le terme de *connecteur discursif* (selon, entre autres, Blakemore, 1987). Tout comme pour les relations de cohérence, on trouve également pour les connecteurs une variété terminologique

[4] Il convient de noter que Cuenca (2003) inclut les connecteurs dans la catégorie des *marqueurs de discours* et discute plutôt leur différence aux *marqueurs modaux*. Dans notre terminologie, les *connecteurs de discours* sont des éléments indiquant les relations de cohérence et les *marqueurs de discours* des éléments modifiant ou influençant l'énoncé.
[5] Selon notre terminologie

considérable dans la littérature. Ainsi, ces éléments (et leurs équivalents) sont aussi nommés des *connecteurs logiques* ou *adverbiaux* (Altenberg, 2006; Crewe 1990), *signaux de liaison* (Leech & Svartvik 1975), *connecteurs logico-sémantiques* (Crothers 1979), *marqueurs lexicaux* ou *discursifs* (Roze et al. 2012; Stede & Umbach, 1998), *référents discursifs* (Kamp, Van Genabith & Reyle, 2011), *phrases repères* (Grosz & Sidner, 1986), *conjonctions* (Halliday & Hasan, 1976), *marqueurs* ou encore *marqueurs de structuration du discours* (Traugott, 2022). De nouveau, la variété de tous ces termes, issus des travaux divers, indique que la conception et définition des connecteurs est loin d'être univoque dans la littérature. Ce qui réunit la plupart des définitions est le fait de concevoir les connecteurs comme des éléments grammaticalisés qui, à partir d'un sens conceptuel original, ont obtenu un sens *procédural* (Moeschler, 2005). Nous allons élucider ce sens procédural maintenant plus en détail.

2.5.1 Les instructions procédurales des connecteurs

Afin de définir et discuter les fonctions procédurales des connecteurs, il convient de les envisager à travers de la *Théorie de la Pertinence* proposée par Sperber & Wilson (1986). Selon Blakemore (1978) la notion de cohérence textuelle « peut être dérivée de la notion de pertinence, et c'est donc la théorie de la pertinence (...) qui fournit le cadre le plus approprié pour l'explication des liens cohésifs dans le discours » (Blakemore, 1976 : 111, ma traduction). Pour faire cela, il convient de rappeler les notions les plus importantes de cette conception.

Tout comme la conception de Sanders et al., (1992), la théorie de la pertinence se base sur les processus cognitifs de la communication. Elle suppose que la base de tous les processus cognitifs humains est le principe que les personnes essaient d'obtenir autant *d'effets contextuels*[6] que possible et cela avec un minimum d'effort cognitif. Afin d'atteindre un bon équilibre entre l'effort et le gain dans la communication, chaque individu concentre son attention sur les informations qui lui semblent les plus pertinentes. Inversement, cela signifie que chaque personne essaie de diriger l'attention des personnes lectrices de manière la plus efficace. Sperber et Wilson (1987) conçoivent la communication comme « un

[6] Un *effet contextuel* peut être la contextualisation d'une nouvelle information, renforcement d'une idée déjà prononcée ou son élimination (Recio Fernández, 2020)

comportement rationnel orienté vers un but » (Hirst, 1989 : 3, ma traduction) et établissent « une sorte d'analyse coût/bénéfice » (Sperber & Wilson, 1987 : 703, ma traduction) du traitement d'un énoncé.

A cet égard, Blakemore (1987 : 144) fait la distinction entre des informations *conceptuelles* et *procédurales* (voir aussi Sperber & Wilson 1986 : 258). Pour encoder ces deux types d'informations, une personne doit activer deux processus cognitifs différents. L'encodage d'une information conceptuelle évoque un concept (resp. sa représentation mentale) et peut activer des inférences pragmatiques, tandis que l'encodage d'une information procédurale consiste dans le regroupement des informations conceptuelles et contraint les inférences pragmatiques (Blakemore, 1987, 2002 : 4f). Moeschler (1998) résume :

> « L'information conceptuelle est l'information qui détermine le contenu d'un concept (plus généralement, d'une représentation mentale), tandis que l'information procédurale détermine comment le système de traitement (par exemple, la cognition humaine) doit traiter l'information conceptuelle » (Moeschler, 1998 : 5, ma traduction).

Les mots possédant des informations procédurales donnent donc des instructions sur la manière de lier logiquement les idées évoquées par les mots conceptuels. Parmi les éléments linguistiques qui fournissent des instructions procédurales, outre les temps verbaux par exemple, se trouvent les *connecteurs discursifs* (Moeschler, 2005).

Pour donner un exemple de la distinction[7] entre les informations provenant des mots conceptuels et les instructions procédurales des connecteurs, considérons l'exemple (39), tiré de Moeschler (2005 : 37).

39) L'avion atterrit et les passagers descendirent.

Les mots conceptuels tels qu'*avion*, *passagers*, *atterrir* et *descendre* évoquent deux représentations mentales différentes (notées selon Moeschler, 2005) : [@avion atterrit] et [@passagers descendirent]. Ces deux représentations mentales sont reliées par le connecteur *et*. Le connecteur n'évoque pas un concept précis mais donne plutôt l'instruction de regrouper ces mots conceptuels dans *une seule* représentation mentale : [@avion atterrit] > [@passagers descendirent]. Selon Moeschler, cette nouvelle représentation « contient une nouvelle information, non

[7] Notons que Moeschler (2005) ainsi que Stede et Umbach (1998) contredisent une distinction nette entre les mots conceptuels et procéduraux. Selon les chercheur-se-s, des connecteurs peuvent posséder à la fois une fonction procédurale que conceptuelle.

présente dans les RM [représentations mentales] qui la composent, la relation d'ordre temporel » (Moeschler, 2005 : 37, voir aussi l'idée d'un *surplus relationnel* de Sanders et al., 1992).[8] Ainsi les instructions procédurales des connecteurs contraignent « le processus interprétatif du traitement des énoncés en pointant le contexte le plus pertinent dans une situation communicative donnée » (Recio Fernández, 2020 : 24, ma traduction).

Nous concluons que les connecteurs de discours peuvent être résumés dans une catégorie linguistique qui encode explicitement les relations de cohérence entre deux segments d'un discours. Provenant de catégories linguistiques différentes, ils partagent la caractéristique d'être morphologiquement invariants (Moeschler, 2005). Conformément à la *Théorie de la Pertinence* (Sperber & Wilson, 1987), une personne locutrice peut utiliser les connecteurs « afin de contraindre la pertinence de la proposition qu'elle introduit » (Blakemore, 2002 : 130, ma traduction). En d'autres termes, pour qu'une personne puisse agir dans un échange communicatif avec le moins d'effort cognitif possible et, inversement, pour qu'elle puisse construire son énoncé de manière à ce qu'il puisse être facilement traité par les autres, elle peut donc faire appel aux fonctions procédurales des connecteurs. À cet égard, il convient également de noter que ces éléments linguistiques ne *créent* pas les relations de cohérence (Hartnett, 1986 : 143; Moeschler, 2005 : 42) mais qu'ils rendent les relations qui préexistent dans la conception cognitive de la personne locutrice entre plusieurs représentations mentales plus « visibles », en donnant l'instruction de décoder le contenu d'une certaine manière.

Il convient toutefois de noter que si la *Théorie de la Pertinence* peut décrire la signification procédurale des connecteurs, elle ne permet pas donner une description précise du sens de différents connecteurs. A cet égard, plusieurs chercheur-se-s distinguent les connecteurs en fonction de la manière dont ils indiquent les relations de cohérence. Par exemple, Hartnett (1986) précise que, de manière générale, les marques de cohésion peuvent être *statiques* ou *dynamiques*. Par dynamique, la chercheuse entend le fait « d'exprimer une manipulation rhétorique du sujet » (1986 : 145, ma traduction). En d'autres termes, les marques

[8] Cependant, Moeschler (2005) note également que les connecteurs « ne conduisent pas automatiquement au regroupement des représentations mentales » (2005 : 42), par exemple dans la phrase suivante : « Marie vient ce soir, parce que ça devrait t'intéresser. » Ici, le connecteur se réfère à un acte de langage, un phénomène sur lequel nous allons revenir au chapitre 2.5.4.

de cohésion dynamiques sont utilisées pour indiquer le développement d'une idée, par exemple en indiquant les changements argumentatifs ou les liens avec d'autres concepts et idées. En revanche, les marques de cohésion statiques servent uniquement à indiquer la répétition d'une idée. Il s'agit plutôt d'éléments tels que des synonymes, des quasi-synonymes, des antonymes ou des collocations. Quant aux connecteurs discursifs, Harnett (1986) précise que ceux qui transmettent une relation de continuité ou d'addition seraient plutôt statiques alors que les connecteurs temporels, causaux ou adversatifs sont dynamiques.

Dans la classification de Halliday et Hasan (1976) les connecteurs sont également classifiés en fonction de la relation de cohérence qu'ils indiquent. Les chercheur-se-s distinguent quatre relations de cohérence différentes :

- Relations additives (p.ex., indiquées par *et*)
- Relations adversatives (p.ex., indiquées par *mais*)
- Relations causales (p.ex., indiquées par *car*)
- Relations temporelles (p.ex., indiquées par *puis*)

Le nombre de relations de cohérence proposé par ces chercheur-se-s représente ainsi une différence importante par rapport à d'autres types de théories qui comptent plus de relations (p.ex. Sanders et al., 1992; Mann & Thompson, 1986).

Bien qu'une classification des connecteurs selon la relation de cohérence qu'ils indiquent soient pertinente, il convient également de noter qu'il existe une multitude de connecteurs qui indiquent la même relation. Par exemple, le dictionnaire des connecteurs *Lexconn* (Roze et al., 2012) répertorie 32 connecteurs indiquant une *concession* (selon le modèle SDRT) :

> alors même que / bien que / cependant que / dire que / et dire que / tout en / en dépit du fait que / en dépit que / en même temps que / encore que / excepté que / hormis le fait que / hormis que / malgré le fait que / malgré que / mis à part le fait que / mis à part que / même si / nonobstant que / pourtant / quand bien même / quand bien même que / quoique / si / si ce n'est que / sinon que / soit dit en passant / à ça près que / à ceci près que / à cela près que / à défaut de / à part que

Étant donné que tous ces connecteurs peuvent indiquer la même relation, la question se pose de savoir en quoi ils diffèrent et comment classer les différences.

Un facteur permettant de distinguer les connecteurs est, par exemple, leur modalité. Tandis que certains connecteurs sont liés plutôt au mode écrit (comme *notwithstanding* en anglais) et presque absents dans la langue orale, d'autres sont plus typiques à l'oral (p.ex., Biber 2006; Crible & Cuenca, 2017). Nous reviendrons au chapitre 3.6 encore plus en détail sur la question de savoir en quoi et selon

quelles caractéristiques les connecteurs spécifiques se distinguent les uns des autres.

Enfin, d'autres chercheur-se-s ont également essayé d'établir des dictionnaires de connecteurs dans différentes langues. Dans le cas de la langue française, le dictionnaire le plus important pour le présent travail (mais voir aussi Afantenos, Asher, Benamara, Bras, Fabre, Ho-Dac, Le Draoulec, Muller, Péry-Woodley, Prevot, Rebeyrolle, Tanguy, Vergez-Couret & Vieu 2012; Laali & Kosseim, 2017) est le dictionnaire déjà mentionné *LexConn* (Roze et al., 2012) qui reprend les catégories de la SDRT de Asher & Lascarides (2003) et classifie 328 connecteurs français différents. Dans celui-ci, les connecteurs sont répertoriés avec leurs définitions et en donnant des exemples.

Après avoir défini les connecteurs, nous allons maintenant aborder la question de savoir comment, et dans quelle mesure, les connecteurs affectent la lecture et la compréhension d'un discours.

2.5.2 L'utilité des connecteurs pour la lecture

Nous savons qu'un bon niveau de cohérence locale « influence positivement le traitement syntaxique et la compréhension des idées principales en permettant aux personnes lectrices de construire des liens référentiels et causaux entre des segments de texte adjacents » (Lehman & Schraw, 2002 : 738, ma traduction). Inversement, si la cohérence locale est réduite, elle engendre un traitement plus approfondi, et a pour conséquence que la personne lectrice doit activer ses connaissances du monde pour générer des inférences.

Étant donné que les connecteurs représentent un moyen important pour assurer la cohérence locale, est-ce qu'il en résulte qu'une utilisation des connecteurs permet une meilleure compréhension ?

Il semble effectivement que parmi tous les moyens cohésifs disponibles pour marquer et garantir la cohérence d'un discours, les connecteurs de discours constituent un moyen très efficace facilitant la compréhension. Cela est particulièrement le cas dans le mode écrit car d'autres indices, notamment prosodiques, n'y sont pas disponibles (McClure & Geva, 1983). De nombreuses études ont constaté ainsi que la présence des connecteurs améliore la compréhension locale et globale d'un texte (p.ex., Spyridakis & Standal, 1987; Millis & Just, 1994; Degand, Lefèvre & Bestgen, 1999; Van Silfhout, Evers-Vermeul

& Sanders, 2014) et permettent un meilleur rappel de son contenu (Caron, Micko, & Thüring, 1988; mais voir Sanders & Noordman, 2000). De plus, les connecteurs semblent aider le traitement mental d'un contenu textuel, c'est-à-dire l'intégration des propriétés syntaxiques, sémantiques et grammaticales pendant la lecture (VanPatten & Benati, 2010). Cet effet est visible dans le fait que les phrases reliées par des connecteurs sont lues plus rapidement que les phrases qui ne sont pas explicitement marquées par des connecteurs (e.g., Britton, Mayer & Penland, 1982; Sanders & Noordman, 2000; Degand & Sanders, 2002; Zufferey, 2014; Kleijn, Mak, & Sanders, 2020).

Il faut noter cependant que l'avantage produit par la présence de connecteurs dépend de la personne lectrice. Plusieurs études ont en effet tenté de déterminer si les personnes lectrices très compétentes profitent plus ou moins des connecteurs que des personnes peu compétentes (p.ex., Marshall & Glock, 1978; Lorch & Lorch, 1986; McNamara, Kintsch, Songer & Kintsch, 1996). Meyer, Brand et Bluth (1980), par exemple, ont observé que les signaux de structure textuelle (entre autres, les connecteurs) facilitent la mémoire du contenu notamment pour des jeunes élèves ayant des difficultés de compréhension. De même, Goldsmith (1982) a observé que les personnes avec un faible niveau de lecture bénéficient de la présence des connecteurs adversatifs (voir aussi Marshall & Glock, 1978). Il semble donc y avoir un lien entre la performance en lecture et le fait de bénéficier des connecteurs : les personnes lectrices moins compétentes semblent bénéficier plus des connecteurs.

Cependant, Johnston & Pearson (1982) ont fait une observation tout à fait différente. Contrairement à Meyer et al. (1980), Goldsmith (1982) et Marshall et Glock (1978) ont trouvé que seules les personnes lectrices les plus compétentes bénéficiaient des connecteurs. Quelques années plus tard, Zinar (1990) fait la même observation.

Pour expliquer cela, Johnston et Pearson (1982) ainsi que Zinar (1990) proposent l'*Hypothèse de développement*. Selon cette hypothèse, on distingue entre trois phases par lesquelles les jeunes personnes lectrices passent durant leurs vies. La première est marquée par l'ignorance ou la négligence des connecteurs explicites dans un texte. Dans cette phase, les élèves « digèrent le texte en unités de mots ou de propositions et font peu d'efforts pour intégrer les unités » (Johnston & Pearson, 1982 : 25, ma traduction). La deuxième étape est

marquée par la prise de conscience des connecteurs et de leurs fonctions, dont résulte une dépendance aux connecteurs. La troisième et dernière étape est décrite comme une lecture 'active' (Nishimito, 1997) dans laquelle les relations logiques dans un texte sont déduites sans l'aide des connecteurs. Ainsi, Zinar (1990) conclut :

> « Pour toute personne lectrice, quel que soit son âge ou ses capacités, il peut exister un niveau optimal de difficulté conceptuelle pour lequel l'attention portée aux signaux de la structure du texte est une stratégie utile : ni si facile que les relations sont déduites spontanément, ni si difficile que la personne lectrice ne dispose pas de la base cognitive essentielle pour comprendre. » (1990 : 198, ma traduction).

Une fois cet équilibre atteint, les connecteurs peuvent grandement faciliter la compréhension (voir aussi Degand & Sanders, 2002, pour une argumentation similaire).

Cela a pour conséquence que des mesures quantitatives de cohésion seules ne permettent pas de déduire la qualité d'un texte (Hartnett, 1986 : 33). De la même manière, Halliday et Hasan (1976) indiquent qu'une analyse de la cohésion d'un texte ne peut servir ni à interpréter, ni à évaluer la qualité d'un texte.[9] Les chercheur-se-s précisent :

> « L'analyse de la cohésion ne vous dira pas que tel ou tel texte est un bon ou un mauvais texte ou un texte efficace ou inefficace dans un contexte donné. Mais elle vous dira quelque chose de POURQUOI VOUS PENSEZ que c'est un bon texte ou un mauvais texte. » (1976 : 328, ma traduction)

Hartnett (1986) propose plutôt de compléter l'analyse quantitative par une analyse qualitative, en réorganisant la taxonomie proposée par Halliday & Hasan (1976) en fonction de la question de recherche.

Bien que les connecteurs soient des éléments dont on peut énormément profiter dans certains cas pour identifier et décoder des relations de cohérence, Crewe (1990) soutient qu'une utilisation excessive peut entraver la compréhension (voir aussi Mosenthal & Tierney, 1984; Crewe, Wright & Leung, 1985). Hartnett (1986) s'accorde à dire que les connecteurs « peuvent contribuer à déformer un sujet de trop de façons différentes en même temps, ce qui fait que l'écriture semble dense, opaque, ou même incohérente pour la personne lectrice » (Hartnett, 1986 : 146, ma traduction). Similairement, Degand et al. (1999) notent que les connecteurs

[9] Ce point devient particulièrement important à retenir dans le contexte d'analyses de corpus qui comparent le nombre de connecteurs utilisés dans les textes de personnes non-natives et natives et en concluent la qualité (subjective) du texte (voir chapitre 3.3).

« ne doivent pas être considérés comme des éléments linguistiques qui peuvent être 'branchés' entre deux phrases, quel que soit leur contenu » (1999 : 5, ma traduction). L'une des raisons pour lesquelles l'utilisation excessive de connecteurs n'améliore pas dans tous les cas la lisibilité et la compréhensibilité d'un texte est que, dans certains cas, la relation logique peut être inférée en leur absence. Prenons le dialogue (40) et (41) (traduit de Mann & Thompson, 1986 : 12) :

40) A : J'ai faim.
41) B : Allons au *Fuji Gardens* !

Dans le chapitre suivant, nous nous intéresserons à la question de savoir pourquoi cette relation est intuitivement et instantanément identifiée correctement (il s'agit d'une relation de 'solution' selon Mann & Thompson, 1986 : 12) alors qu'aucun connecteur n'est présent pour indiquer cette relation de cohérence.

2.5.3 Dans quels cas convient-il d'utiliser un connecteur ?

Nous avons vu qu'un discours peut rester cohérent même si les relations de cohérence ne sont pas marquées explicitement (e.g., Hartnett, 1986; Crewe, 1990). Considérons encore l'exemple (42), tiré de Moeschler (2005).

42) L'avion atterrit. Les passagers descendirent.

Alors que la relation de cohérence peut être indiquée explicitement avec un connecteur temporel (comme p.ex. *ensuite*), une personne lectrice n'a pas de mal à comprendre le lien logique entre les deux phrases et à construire une seule représentation mentale de la phrase. Cela soulève inévitablement la question suivante : quand convient-il d'utiliser un connecteur pour indiquer une relation de cohérence pour qu'il ait un impact positif sur la lecture ? Degand et al. (1999) résument à cet égard:

> « La principale difficulté de l'auteur-e consiste (…) à décider comment et quelle information fournir à la personne lectrice : suffisamment mais pas trop, pour que la personne lectrice puisse se construire une représentation mentale cohérente sans être encombré d'informations déjà connues » (Degand et al., 1999 : 12, ma traduction)

Tout d'abord, il semble que le facteur le plus important soit la relation de cohérence, car toutes les relations ne peuvent pas être laissées implicites. Selon Mann & Thompson (1986), par exemple, ce sont notamment des relations comme les *présuppositions*, les *implications* ainsi que *les actes de langage indirects* (selon Searle, 1975) qui sont laissées les plus souvent implicites (voir aussi Das &

Taboada, 2018). Pourquoi ces relations sont-elles moins marquées que d'autres ? Il existe plusieurs théories qui proposent des réponses à cette question.

La première est *l'Hypothèse de la Continuité* de Segal, Duchan et Scott (1991). Ces chercheur-se-s font la distinction entre les relations de cohérence *continues* et *discontinues*. Une relation continue est présente si une information nouvellement introduite est intégrée dans la « construction en cours d'une composante intégrée de la signification du récit » (Segal et al., 1992 : 31, ma traduction), tandis que les relations discontinues représentent une rupture avec la ligne continue de la narration. Comme les relations continues représentent un déroulement continu du texte, elles se caractérisent par une réalisation des attentes (43). En revanche, une phrase discontinue représente des changements soudains de sujet ou les violations générales d'une attente (44).

43) J'ai pris mon parapluie parce qu'il pleut.
44) Il pleut à Paris, mais j'ai loupé mon train.

Dans deux séries d'expériences, Murray (1994, 1997) a testé comment la lecture de ces deux types de relations est influencée par la présence de connecteurs. D'abord, en 1994, Murray a testé si l'absence de connecteurs additifs et causaux (c'est-à-dire des connecteurs indiquant les relations continues) et adversatifs (c'est-à-dire des relations discontinues) influence le temps de lecture. Plus précisément, le chercheur demandait à des participant-e-s natifs de l'anglais de lire 30 paires de phrases comprenant soit une relation adversative, soit une relation causale ou additive. Une moitié des phrases était explicitement marquée par un connecteur, tandis que dans l'autre moitié, le connecteur était omis. Les résultats ont montré que l'omission du connecteur n'a pas influencé le temps de lecture des phrases qui contenaient une relation continue. Pour les phrases discontinues en revanche, la présence d'un connecteur a entrainé une diminution du temps de lecture par rapport à la phrase dans laquelle le connecteur n'était pas explicitement marqué. Murray (1994) en a conclu que les connecteurs adversatifs ont un impact plus important sur la cohérence inter-phrastique que les connecteurs indiquant des relations continues. Les résultats indiquent ainsi que l'utilisation d'un connecteur pour une relation discontinue semble être plus importante pour assurer la cohérence : comme les personnes lectrices s'attendent à ce que le discours se déroule de

manière continue, une relation discontinue représente un changement qui doit être signalé.

Dans une étude ultérieure, Murray (1997) a montré que l'absence de connecteur adversatif entrainait un temps de lecture plus élevé. Dans cette étude, le chercheur a testé l'impact des connecteurs continus et discontinus véhiculant soit une relation continue, c'est-à-dire l'addition et la cause, soit une relation discontinue, c'est-à-dire une relation adversative. Pour faire cela, Murray (1997) a demandé à 45 participant-e-s natifs de l'anglais de lire des paires de phrases et a mesuré leurs temps de lecture. Les paires de phrases contenaient soit un connecteur inapproprié (p.ex. pour une relation additive un connecteur adversatif), soit aucun connecteur. Les résultats montrent que les temps de lecture deviennent plus lents par rapport aux phrases sans connecteur lorsque la phrase est introduite par un connecteur inapproprié. Cet effet s'est manifesté même lorsque le connecteur était un marqueur (inapproprié) de continuité dans une phrase de continuité (p.ex. un connecteur causal pour une relation additive). Par ailleurs, Murray (1997) a observé que les connecteurs adversatifs inappropriés perturbaient davantage la lecture que les connecteurs causaux ou additifs. Le chercheur en a conclu, en d'accord avec ses conclusions de l'étude précédente, que ce résultat était probablement dû au fait que le 'signal' d'un connecteur adversatif est plus fort et important que celui d'un connecteur additif ou causal.

Les résultats obtenus dans ces deux études s'expliquent par le principe de continuité (Segal et al., 1991). Comme les personnes lectrices ont l'attente que le discours se déroule de manière continue, le marquage d'une relation discontinue est crucial pour assurer la cohérence, alors que les relations continues peuvent plus facilement être laissées implicites. Ainsi, on considère que les connecteurs discontinus fournissent des signaux plus forts, ce qui entraine des problèmes de lecture plus importants lorsqu'ils sont absents (Murray, 1994) ou incorrects (Murray, 1997).

Quant aux relations continues, *l'Hypothèse de la causalité-par-défaut* de Sanders (2005) permet de faire encore une distinction plus fine entre les relations additives et les relations causales. En bref, cette hypothèse suggère que les personnes lectrices supposent par défaut un déroulement causal. Ainsi, l'absence de connecteurs causaux entre deux segments reliés de manière causale ne les rend pas nécessairement incohérents. Comme illustré en (45) et (46) (traduits de

Sanders, 2005), les relations de cause-conséquence et de conséquence-cause peuvent être comprises même si elles ne sont pas explicitement marquées.

45) Bill est entré dans la pièce. Bob est parti.
46) Bob est parti. Bill est entré dans la pièce.

Cette hypothèse est soutenue par les résultats de Noordman et Sanders (2000) qui ont démontré que les connecteurs causaux sont traités plus rapidement que les connecteurs additifs, alors que conceptuellement la causalité représente une relation plus complexe que l'addition. En conséquence, les relations causales sont laissées plus souvent implicites que les additions (Asr & Demberg, 2012).

Il convient cependant de noter qu'il y a des restrictions à cette hypothèse. Bien qu'il soit, de manière générale, plus facile de comprendre les relations causales de manière implicite que d'autres relations de cohérence, cela ne veut pas dire que toutes les relations causales peuvent être laissées implicites. En effet, un rapport causal entre les phrases (47) et (48) (traduites de Moeschler, 2005 : 46) ne peut pas être établi sans marquage explicite, respectivement sans un contexte plus élaboré. Dans cet exemple, l'insertion d'un connecteur causal comme *parce que* indiquerait clairement le concept de causalité qui n'aurait pas été activé autrement.

47) Abi a arrêté de manger le gâteau aux pommes.
48) Axel a commencé de manger une salade.

De même, l'hypothèse de Sanders (2005) n'implique pas non plus que toutes les autres relations doivent toujours être marqués avec un connecteur. Considérons la phrase (49), dans laquelle la relation de cohérence discontinue peut être comprise malgré l'absence de connecteur.

49) Pierre a coupé les oignons. Il n'a pas pleuré.

Bien que les hypothèses de *Causalité-par-défaut* (Sanders, 2005) et de la *Continuité* (Segal et al., 1991; Murray, 1994, 1997) permettent d'identifier certains cas dans lesquels une relation peut être communiquée sans connecteur, elles n'expliquent pas tout.

A cet égard, une autre hypothèse permet d'ajouter des éléments explicatifs supplémentaires. Il s'agit de *l'Hypothèse de la densité d'Information uniforme* de Levy & Jaeger (2006). Cette hypothèse suppose qu'une personne locutrice essaie toujours de produire un flux d'information d'une densité régulière autant que

possible, c'est-à-dire sans passages plus ou moins informatifs que d'autres. Ainsi, « la quantité de surprise d'un mot à l'autre » (Asr & Demberg, 2015 : 119, ma traduction) devrait être à peu près égale tout au long d'un texte. Cette hypothèse est soutenue dans plusieurs domaines de la linguistique, comme par exemple la phonétique (Aylett & Turk, 2004). Quant au marquage des relations de cohérence, Asr et Demberg (2015) ont démontré que l'emploi de certaines marques cohésives peuvent également être prédits par *l'Hypothèse de la densité d'Information uniforme*. Dans leur étude, elles ont constaté que les relations inter-phrastiques qui étaient moins prévisibles (compte tenu de leur contexte linguistique) étaient plus souvent marqués par un connecteur optionnel.

Outre les théories linguistiques plus générales, il convient également de noter l'aspect individuel et subjectif lors de l'établissement d'un discours. Le travail de Holtgrave (1997), par exemple, démontre que les personnes varient lorsqu'il s'agit d'exprimer et de comprendre des énoncés implicites et indirects. Le chercheur a demandé aux participant-e-s d'indiquer leurs propres usages et compréhension d'une communication indirecte. Il a constaté que les personnes qui indiquaient être conscientes qu'il existe des significations importantes qui sont implicites et qui les recherchaient activement étaient beaucoup plus susceptibles d'accepter les interprétations indirectes que les autres. Ce travail indique que la tendance à laisser certaines relations de cohérence implicite pourrait également être lié à la personne qui conçoit le discours. Cela ne vaut évidemment que pour les cas où le marquage d'un connecteur est hautement facultatif.

Parallèlement, cela soulève la question de savoir dans quelle mesure le marquage des relations de cohérence est influencé par le cadre culturel. Selon Kanno (1989), les relations de cohérence sont par exemple moins souvent marquées explicitement en japonais qu'en anglais. Nous reviendrons sur la question de l'influence de la stylistique de la L1 sur l'utilisation des connecteurs au chapitre 8.4, ainsi que les différences individuelles dans la maitrise des connecteurs au chapitre 3.5.

En conclusion, nous avons vu que plusieurs facteurs entrent en jeu pour expliquer pourquoi certaines relations sont plus souvent implicites que d'autres. Selon l'*Hypothèse de la continuité*, les relations discontinues telles qu'une concession doivent être marquées explicitement plus souvent que les autres (Segal et al., 1991;

Murray, 1994, 1997). Lorsqu'il est question de marquer une relation continue qui ne nécessite pas forcément de connecteur (p.ex. une phrase causale, Sanders, 2005), l'*Hypothèse de la Densité d'Information Uniforme* de Levy & Jaeger (2006) peut fournir des explications sur les raisons d'utiliser ou non un connecteur (Asr & Demberg, 2015).

Finalement, nous avons vu que le marquage des relations de cohérence peut dépendre de la personne qui conçoit un discours. Ainsi, le cadre culturel dans lequel un discours est conçu pourrait avoir un impact sur la manière dont les relations de cohérence sont marquées (Kanno, 1989). Pourtant, de plus amples recherches sont encore nécessaires afin de pouvoir répondre à cette question de manière satisfaisante. A cet égard, il sera également important de prendre en compte des facteurs stylistiques tels que le sujet abordé, les connaissances des auteur-e-s ainsi que le public cible.

2.5.4 Les fonctions pragmatiques des connecteurs

L'identification des liens logiques d'un texte n'est pas toujours facile et nous avons vu que les connecteurs représentent un outil important, assurant que les relations de cohérence sont facilement identifiables. Outre leurs fonctions procédurales, il convient de noter que les connecteurs possèdent également des fonctions *pragmatiques* (p.ex., Fraser, 1998; Moeschler 2005, Zufferey 2012, 2014). Stede et Umbach (1998) indiquent ainsi que les connecteurs peuvent porter « une grande variété de connotations sémantiques et pragmatiques, qui rendent le choix d'un marqueur orienté vers le sens, par opposition à une simple conséquence de décisions structurelles » (Stede & Umbach, 1998 : 1239, ma traduction).

Pour donner un premier exemple, considérons les phrases (50) et (51). Selon Fraser (1998), le connecteur concessif utilisé en (51) doit impliquer une forme de protestation, une fonction qui n'est pas remplie par l'anglais *however* ('cependant'). Seul *but* ('mais') convient dans ce cas, même si les deux connecteurs indiquent exactement la même relation de cohérence (selon p.ex. la taxonomie de Sanders et al., 1992).

50) *A : I thought you would come.*
'A : Je pensais que tu allais venir.'
51) *B : But /* However I did !*
'B : Mais / *Cependant je l'ai fait !'

Évidemment, les fonctions pragmatiques ne se limitent pas seulement aux connecteurs concessifs et un grand nombre d'études a investigué en particulier les fonctions pragmatiques des connecteurs causaux.

La recherche a démontré, par exemple, que le connecteur causal du français *puisque* possède des charges pragmatiques très nuancées, de sorte qu'il peut avoir une signification échoïque (Zufferey, 2012, 2014) ou impliquer même une attitude dissociative envers le contenu de la personne locutrice (Schumann et al., 2020).[10]

Les caractéristiques de *puisque* ont aussi un impact sur la lecture. Par exemple, la recherche a démontré que des arguments fallacieux (p.ex. les détournements des idées par l'interlocuteur-rice) sont mieux détectés lorsqu'ils sont introduits par *puisque* en comparaison avec d'autres connecteurs tels que *vu que*, *comme* ou *étant donné que* (Schumann et al., 2020).

Un autre exemple des fonctions pragmatiques concerne les différents types de causalité qu'un connecteur peut indiquer. Comme nous voyons dans les exemples suivants (traduits de Degand & Pander Maat, 2003 : 175), une relation causale peut être plutôt *objective* (52) ou *subjective* (53).

52) John est revenu parce qu'il l'aime.

53) John l'aime, parce qu'il est revenu.

Étant donné que la causalité dans (53) est tirée d'une conclusion de la personne locutrice (i.e., *j'en conclus* que le fait que John l'aime *doit être* la raison pour laquelle il est revenu), elle ne peut être établie que lorsque la personne lectrice adopte un point de vue subjectif. En revanche, la causalité dans (52) peut être établie sans tenir compte du 'raisonnement' de la personne locutrice et peut donc être considéré plutôt *objective*. Selon Canestrelli, Mak et Sanders (2013), cette distinction se retrouve dans des nombreux modèles de discours sous différents termes.[11]

Le degré d'objectivité d'une relation causale a également un impact sur la lecture. Il a été démontré que les relations causales subjectives prennent plus de

[10] Nous allons revenir sur ce connecteur plus en détail au chapitre 9.

[11] On trouve comme autre dénomination par exemple les relations *sémantiques* ou *pragmatiques* (Sanders et al., 1992; Moeschler, 2005), *sujet abordé* ou *présentiel* (*subject matter / presential*, Mann & Thompson, 1988) ou encore les *sens interpersonnels* et *idéationnels* (*interpersonal / ideational meaning*, Halliday & Hasan, 1976). En outre, des raffinements supplémentaires ont été apportés à ces deux types de causalité. Par exemple, Kerbrat-Oreccioni (1991) distingue encore au sein de la causalité subjective la subjectivité *déictique* et la subjectivité *affective-évaluative*. Et tandis que plusieurs modèles conçoivent une distinction nette entre les relations subjectives et objectives, Degand et Pander Maat (2003) le conçoivent de manière scalaire.

temps à traiter que les relations causales objectives (Canestrelli et al., 2013; Kleijn et al., 2020). Les travaux de Traxler, Bybee et Pickering (1997) ont également montré que les relations subjectives nécessitent un effort mental élevé que les relations objectives, car les personnes lectrices doivent réviser leurs représentations mentales (Kleijn et al., 2020 : 42).

Comme nous l'avons indiqué, le degré d'objectivité d'une relation causale n'a pas seulement un effet sur la lecture, il influence également le choix du connecteur. Dans de nombreuses langues, il existe une préférence pour indiquer une relation causale subjective avec un connecteur spécifique. Parmi ces langues se trouvent par exemple le néerlandais (Sanders & Spooren 2015, Canestrelli et al., 2013) l'allemand (Waltereit, 2001; Stukker & Sanders, 2012) ou le chinois mandarin (Li, Evers-Vermeul & Sanders, 2013). En français moderne, la situation n'est pas aussi tranchée. Comme l'ont démontré Zufferey, Mak, Verbrugge & Sanders (2018), même si la causalité subjective été historiquement plutôt marquée par le connecteur *car* au lieu du connecteur plus objectif *parce que,* il semble que la sensibilité des personnes francophones natives leur permettant de distinguer ces usages ait diminué et que les différences entre ces deux connecteurs soient désormais perçues plutôt en termes de registre (voir aussi les travaux de Degand & Pander Maat, 2003; Fagard & Degand, 2008; Degand & Fagard, 2018; Lambrecht Bordeaux & Reichle, 2006).

Une petite digression sur ce sujet s'impose, car la question se pose inévitablement de savoir si et comment les gens peuvent anticiper une lecture subjective dans des langues qui n'ont pas de préférence pour un connecteur subjectif particulier (comme l'anglais, Traxler et al., 1997). Une possibilité serait que cela soit fait par des signaux alternatifs. En effet, nous avons vu à la section 2.4.2 qu'il existe, outre les connecteurs, une multitude d'autres signaux qui peuvent indiquer des relations de cohérence. Considérons par exemple (54), dans lequel le fait d'énoncer explicitement un raisonnement subjectif par l'insertion d'une phrase d'introduction (*John pensait que)* pourrait faciliter le traitement de la causalité subjective dans la clause enchâssée.

54) *John pensait que* les voisins n'étaient pas là car leur voiture n'était pas dans le parking.

Cependant, les résultats de Kleijn et al. (2020) ne montrent pas un tel effet. En mesurant les temps de lecture des participant-e-s, les chercheur-se-s ont observé que la présence d'une phrase d'introduction telle que (54), censée être une indication de subjectivité, ne facilitait pas le traitement mental de la causalité subjective. Les chercheur-se-s ont ainsi conclu que la présence d'une telle phrase n'a pas évoqué « une configuration subjective de l'espace mental nécessaire à l'interprétation des relations subjectives » (Kleijn et al., 2020 : 52, ma traduction).

L'étude de Canestrelli, Mak et Sanders (2016) démontre également qu'une lecture subjective ne s'active pas si facilement par des ajouts lexicaux. Dans cette étude, les chercheur-se-s ont testé l'impact de la présence d'adverbes évaluatifs tels que *fantastique* et *horriblement*, donc des mots indiquant un point de vue subjectif. Malgré le fait que ces mots impliquaient également une forte implication émotionnelle de la personne locutrice, leur présence ne facilitait pas non plus le traitement mental des relations causales subjectives.

Alors que des ajouts lexicaux ne semblent donc pas faciliter le traitement d'une relation causale subjective, l'étude de Kleijn et collègues (2020) démontre qu'un *style* subjectif peut, quant à lui, guider les personnes lectrices vers une lecture subjective. Dans la deuxième expérience de leur étude mentionnée ci-dessous, l'équipe de recherche a observé que des textes écrits dans le style d'un flux continu de conscience, tel que (55) (traduit de Kleijn et al., 2020 : 22), ont conduit les participant-e-s à adopter une lecture subjective.

55) Ruben a conduit jusqu'au magasin d'électronique. *Wow ! Cette forme, cette couleur, cet écran, c'était ça ! Il y avait là la télé de ses rêves. Et quelle chance,* le magasin offrait un rabais incroyable. Ainsi, il a économisé beaucoup d'argent.

Par conséquent, la relation causale subjective présentée à la fin (*il a économisé beaucoup d'argent*), a été lue plus rapidement non seulement lorsqu'un connecteur subjectif était utilisé, mais également lorsque les textes contenaient des marques subjectives (représentées dans l'exemple par des italiques).

Enfin, pour compléter les discussions sur ces deux types de causalité, il convient encore de noter que certaines conceptions dans la littérature ont identifié également un troisième type de causalité. Selon Sweetser (1990), un lien causal

peut aussi relier des propositions pour établir un acte de langage, tel que représenté dans (56) (traduit de Degand & Pander Maat, 2002 : 175).

56) Qu'est-ce que tu fais ce soir car il y a un bon film au cinéma.

Dans cet exemple, le connecteur n'établit pas de la causalité entre les contenus proposés en soi, mais justifie l'acte de langage en question, la proposition implicite d'aller au cinéma ensemble. Ainsi, la causalité indiquée par le connecteur ne peut être comprise correctement que si l'on tient compte du contexte et, surtout, des intentions de la personne locutrice. Il semble que les connecteurs subjectifs soient également plus susceptibles d'être utilisés pour les actes de langage (Zufferey, 2012; voir aussi selon une conception scalaire de l'implication de la personne locutrice: Degand & Pander Maat, 1999; Pander Maat & Degand, 2002).

Pour conclure, nous venons de voir que les connecteurs ne sont pas seulement des éléments qui servent à structurer un discours, mais qu'ils peuvent également communiquer une multitude de nuances et sous-entendus au niveau pragmatique. Tandis que certains connecteurs causaux indiquent un point de vue subjectif dissociatif envers le contenu (Schumann et al., 2020), d'autres peuvent conduire à une lecture subjective (p.ex., Sanders & Spooren, 2015). Il en résulte que les connecteurs ne peuvent pas être utilisés de manière interchangeable dans tous les cas et que les fonctions pragmatiques représentent dans certains cas des contraintes à utiliser certains connecteurs plutôt que d'autres (p.ex., Fraser, 1998; Zufferey, 2012).[12]

2.6 Conclusion

Dans ce chapitre, nous avons vu qu'un discours se caractérise par une logique interne, c'est-à-dire par sa cohérence. Les rapports logiques entre les différentes propositions ou segments d'un discours sont appelés les *relations de cohérence*. Ces relations peuvent être classifiées de plusieurs façons, par exemple en mettant l'accent sur les intentions de la personne locutrice.

Le large éventail de modèles du discours proposés classifie ces relations selon différentes perspectives. La littérature inclut ainsi des modèles marqués par des visions statiques et mathématiques (Montague, 1974), sémantiques dynamiques

[12] Nous reviendrons sur ce point plus en détail au chapitre 9.

(Kamp, 1981), pragmatiques et intentionnelles (Grosz & Snider, 1986) ou inférentielles (Mann & Thompson, 1988). Finalement, nous avons également discuté l'idée que les relations de cohérence représentent des entités cognitives comme base de toute cognition humaine (Sanders et al., 1992).

Quel que soit le modèle théorique, les relations de cohérence peuvent être signalées à l'aide de moyens différents, tels que les structures parallèles ou des moyens lexicaux (Das & Taobada, 2018). L'un d'entre eux est la catégorie des *connecteurs discursifs*. De manière générale, nous avons vu que les connecteurs facilitent l'identification des relations de cohérence et permettent ainsi une meilleure compréhension du texte (e.g., Millis & Just, 1994). Cependant, l'effet de facilitation d'un connecteur dépend aussi de la relation de cohérence en question : certaines relations de cohérence nécessitent un marquage explicite par un connecteur, tandis que d'autres peuvent être plus facilement laissées implicite (p.ex., Segal et al., 1991; Murray, 1994, 1997; Sanders, 2005).

Finalement, nous avons vu que les connecteurs de discours peuvent posséder des caractéristiques pragmatiques très nuancés (p.ex., Zufferey, 2012, 2014). Par exemple, nous avons vu que de nombreuses langues possèdent des connecteurs différents pour indiquer différents types de causalité, à savoir la causalité subjective ou objective. Au travers de ce chapitre, nous avons donc vu que les connecteurs représentent - malgré les différences qui existent entre les différentes conceptualisations du discours - une catégorie d'éléments linguistiques qui, grâce à leurs fonctions pragmatiques, sont extrêmement utiles, non seulement pour structurer un discours et assurer sa cohérence, mais aussi pour l'enrichir de manière pragmatique. Issus de catégories linguistiques diverses, les connecteurs se trouvent à l'interface de plusieurs domaines linguistiques, comme la syntaxe, le lexique, la pragmatique et la sémantique. Par conséquent, il est légitime de faire l'hypothèse que les connecteurs sont des éléments difficilement acquis et maitrisés dans une deuxième langue. En effet, lorsqu'une personne non-native a des déficits dans les domaines linguistiques dans lesquels les connecteurs se trouvent (par exemple la pragmatique), elle maitrisera peut-être moins bien les connecteurs. La question se pose aussi de savoir si les personnes non-natives peuvent bénéficier des connecteurs lors de la lecture. Nous aborderons ces questions dans le chapitre suivant.

3 La maitrise des connecteurs discursifs en langue seconde

Dans le chapitre précédent, nous avons vu que les connecteurs de discours sont des éléments linguistiques qui, en encodant des significations procédurales entre deux segments d'un discours (p.ex., Halliday & Hasan, 1976; Sanders et al., 1992), assurent la cohérence d'un texte et contribuent ainsi à sa compréhensibilité (p.ex., Degand et al., 1999).

Dans ce chapitre, nous allons aborder la question de savoir à quel égard les personnes non-natives, quant à elles, profitent également des connecteurs pour mieux comprendre un texte. Nous examinerons également si les locuteur-trice-s non-natifs maitrisent les connecteurs, par exemple dans leurs productions écrites. Comme cela ne semble pas vraiment être le cas, nous discuterons des facteurs qui en sont responsables, comme par exemple des facteurs interpersonnels ou un enseignement insuffisant.

En particulier, nous nous concentrons sur un facteur qui a malheureusement été négligé jusqu'à présent par la recherche : le fait que la catégorie des connecteurs est extrêmement hétérogène. Nous poserons la question de savoir si et comment cette hétérogénéité se répercute sur l'acquisition et la maitrise des connecteurs en L2.

Avant d'aborder ces questions, il est nécessaire de définir ce que signifie de *maitriser* un connecteur. À cette fin, nous discuterons tout d'abord deux dimensions de la maitrise, à savoir la *connaissance explicite* et *implicite* selon la conception de Ellis (2004, 2005, 2009b) ainsi que leurs opérationnalisations dans la recherche linguistique, ce qui permet une meilleure mise en contexte des études menées dans ce domaine.

3.1 Qu'est-ce la maitrise ?

Avant d'évaluer si les personnes non-natives ont des problèmes à bien maitriser les connecteurs dans leur langue seconde, il faut d'abord s'interroger ce qui signifie

maitriser un connecteur. En effet, vue de près, cette notion soulève de nombreuses questions, telles que :

- La maitrise comprend-elle seulement la production des connecteurs dans la langue écrite ou aussi leur compréhension en lisant ?
- Le terme implique-t-il également le traitement mental des fonctions des connecteurs « en ligne », c'est-à-dire lors de la lecture?
- La maitrise d'un connecteur par une personne native est-elle la norme et le point de référence de la maitrise des connecteurs dans une langue donnée?

Afin de pouvoir répondre à ces questions et de définir plus précisément la notion de *maitrise*, nous allons maintenant discuter les formes que peut prendre la maitrise des connecteurs dans une deuxième langue.

3.1.1 Quelques considérations théoriques préliminaires

Dans la présente thèse, la notion de maitrise des connecteurs englobe à la fois leur compréhension, leur production et leur traitement mental. Il est à noter que le terme de compréhension se réfère ici à la compréhension de la *fonction du connecteur* et non à la compréhension du *type de la relation* ou du *discours* en général. Évidemment, l'un ne va pas sans l'autre et il a été démontré qu'une bonne compréhension d'un texte permet de conclure que la personne comprend (c'est-à-dire *maitrise*) les connecteurs (Geva, 1986, 1992; Ozono & Ito, 2003). Il est toutefois impératif de questionner la mise en équivalence de la compréhension d'un texte avec la compréhension des connecteurs qui y sont présents et cela pour deux raisons. La première raison est que d'autres facteurs influencent également la compréhension d'un texte indépendamment des connecteurs, comme par exemple une connaissance préalable du sujet (Brantmeier, 2005). La deuxième raison est que cette distinction permet de tirer des conclusions sur l'acquisition de connecteurs en tant qu'éléments particuliers d'une L2 et sur la manière dont leurs différents facteurs linguistiques y contribuent. Considérons par exemple l'étude de Ozono et Ito (2003), dans laquelle l'équipe de recherche a testé trois connecteurs de relations de cohérence de complexité différente. Les participant-e-s ont dû remplir à l'aide de ces connecteurs les trous dans des phrases données. Les résultats montrent que les apprenant-e-s ayant une compétence linguistique faible

obtenaient de meilleurs résultats avec les connecteurs encodant une relation de cohérence simple. L'équipe de recherche en a conclu qu'une charge cognitive élevée (issue de la relation de cohérence) influençait la lecture et la compréhension des connecteurs en L2. Ces observations sont certes cohérentes et instructives, mais on ne sait pas si ces résultats sont vraiment dus à la complexité de la relation de cohérence ou plutôt à la (non-)maitrise des connecteurs testés. Comme seulement un connecteur par relation de cohérence a été testé, il n'est pas clair si les propriétés linguistiques des connecteurs (comme leurs registres ou leurs modalités dans laquelle ils sont utilisés) ont influencé les résultats - un aspect qui a aussi été discuté par les auteurs (Ozono & Ito, 2003 : 294).

Après avoir défini la maitrise des connecteurs en termes de compréhension, de traitement mental et de production, la question se pose maintenant de savoir quels types de connaissances peuvent refléter ces différentes dimensions. De manière générale, on peut distinguer entre deux types de connaissances juxtaposées autour desquelles une compétence linguistique dans une deuxième langue peut s'articuler : *les connaissances explicites* et *les connaissances implicites* (p.ex., Ellis, 2004, 2005, 2009a, 2009b). Par une connaissance *implicite* on entend des connaissances tacites, intuitives et intériorisées qui restent souvent incomplètes. Par exemple, si une personne non-native d'une connaissance implicite élevée lirait la phrase (57), elle détecterait l'erreur mais sans pouvoir dire pourquoi et en quoi l'erreur consiste.

57) *Sandra est une personne très ouvert.

En revanche, une connaissance est *explicite* lorsque celle-ci est acquise consciemment et qu'elle se présente sous forme déclarative ou encyclopédique. Ainsi, ce type de connaissance peut être, contrairement à la connaissance implicite, potentiellement verbalisée et être, théoriquement, acquises entièrement. Une personne ayant une connaissance explicite élevée qui lisait (57) saurait donc indiquer qu'il s'agit d'une erreur d'accord de l'adjectif masculin *ouvert* qui se réfère au nom féminin *personne*. Alors que la distinction entre ces deux dimensions est généralement acceptée, il existe aucune unanimité sur la question de savoir si ces deux celles-ci se juxtaposent ou s'il faudrait les concevoir plutôt comme deux pôles d'un continuum. Selon Paradis (1994) les deux pôles sont opposés et non reliés, en d'autres termes d'un transfert d'une connaissance à l'autre ne serait pas

possible (voir aussi Ellis, 2006), tandis que d'autres chercheur-se-s envisagent un continuum entre ces deux pôles de maitrise (Dienes & Perner, 1999). Il est à noter que même si on concevait les deux dimensions comme deux pôles juxtaposés, elles pourraient toujours s'articuler entre elles, de façon qu'une connaissance explicite pourrait faciliter l'acquisition d'une connaissance implicite (Ellis, 1993).

En conclusion, la distinction entre ces deux types de connaissances permet de mieux identifier et de séparer les facteurs problématiques des apprenant-e-s en langue seconde. Cela est avantageux car il se peut, par exemple, que les connecteurs posent certains problèmes seulement à la connaissance implicite tandis que la connaissance explicite n'est pas touchée (ou *vice-versa*). Avant de thématiser ceci, il convient de s'interroger d'abord sur l'opérationnalisation de ces deux types de connaissances (c'est-à-dire la question de savoir comment on peut les mesurer) afin de mieux pouvoir mettre en contexte les études menées sur la maitrise des connecteurs.

3.1.2 Les méthodes courantes utilisées

En ce qui concerne les connaissances implicites et explicites, Ellis (2009b : 38) distingue sept critères qui se reflètent dans la manière dont les personnes montrent leurs connaissances et qui permettent donc d'opérationnaliser les deux types de connaissances. Ces critères sont résumés ci-dessous :

- Le degré de conscience (angl. *degree of awareness*) : lorsque les participant-e-s répondent à une tâche plutôt en fonction de leur « sentiment » (Ellis, 2009b : 40, ma traduction) et non des règles acquises, la connaissance implicite est mesurée.
- La disponibilité en temps (angl. *time available*) : comme l'accès aux connaissances explicites nécessite un certain temps, l'imposition d'une limite de temps aux participant-e-s les force à répondre plutôt selon leurs connaissances implicites.
- Le centre d'attention (angl. *focus on attention*) : si, dans une tâche, l'accent est mis sur la signification de la réponse, on mesure les connaissances implicites, alors les tâches visant la forme indiquent plutôt des connaissances explicites.
- La systématicité (angl. *systematicity*) : tandis que les réponses systématiques indiquent une bonne connaissance implicite, la connaissance explicite évoque plutôt des réponses variables
- La confiance (angl. *certainty*) : on suppose que les apprenant-e-s sont plus sûr-e-s de leurs réponses lorsqu'ils accèdent à leurs connaissances implicites, alors que les connaissances explicites laissent une plus grande place au doute.

- L'utilité des connaissances du métalangage (angl. *utility of knowledge of metalanguage*) : tout commentaire métalinguistique ne peut pas se référer exclusivement à des connaissances explicites.
- La capacité d'apprentissage (angl. *learnability*) : les apprenant-e-s qui ont appris une langue tôt dans leur vie ont souvent de meilleures connaissances implicites, tandis que l'acquisition tardive est plutôt caractérisée par des connaissances explicites.

Quant à la maitrise des connecteurs en L2, les critères de Ellis (2009b) permettent donc le développement des tâches qui visent des connaissances différentes. Par exemple, en ce qui concerne les connaissances explicites, les productions de textes en L2 sont censées de permettre aux participant-e-s d'accéder à leur connaissances explicites, car elles impliquent généralement que les personnes peuvent choisir et utiliser les connecteurs de manière consciente, sans contraintes de temps. Cette approche permet donc de bien documenter la production des connecteurs à travers des analyses quantitatives et qualitatives. Pourtant, l'évaluation de la maitrise des connecteurs peu fréquents est, par manque d'occurrences, très difficile et elle ne permet pas non plus de savoir si les apprenant-e-s évitent certains connecteurs (et si oui, lesquels et pourquoi). De même, elle ne permet pas d'inférer la façon dont les lecteur-trice-s en L2 réagissent lorsqu'ils rencontrent des connecteurs et comment ils les comprennent et intègrent pendant la lecture. De plus, les interprétations restent plus ou moins spéculatives, l'analyse des cas où les connecteurs sont mal utilisés ne permet pas souvent de déterminer avec certitude *pourquoi* ils sont mal utilisés.

Pour mieux évaluer la connaissance explicite d'un connecteur plus rare ou d'un connecteur qui se présente dans une condition spécifique il est plus avantageux d'adopter une approche expérimentale qui permet d'isoler certaines variables et d'observer avec beaucoup plus de contrôle les phénomènes en question. Il convient de réaliser, par exemple, des tâches d'elicitation ou d'évaluation. Un grand avantage de ces tâches est qu'elles permettent de contrôler le niveau de discours auquel on souhaite évaluer les connaissances des connecteurs : *intra-phrastique*, *inter-phrastique* ou au *niveau du discours* (Geva, 1986). Selon Geva (1986), le contexte restreint du niveau *intra-phrastique* devrait permettre aux apprenant-e-s de mieux utiliser le connecteur, c'est-à-dire que l'identification et l'insertion de la relation de cohérence est facilitée. D'autre part, au niveau *intra-phrastique* et encore plus au niveau du discours, il devient plus difficile

de faire les bonnes inférences. Voici quelques exemples des tâches conçues pour mesurer ces différents niveaux.

Un exemple d'une tâche d'elicitation intra-phrastique est illustré en (1). Dans cette tâche, (exemple tiré de Geva, 1986 : 88, voir aussi Geva & Ryan, 1993), les participant-e-s choisissent la suite cohérente d'une phrase contenant un connecteur.

1) We could not see the man, although_____
Nous ne pouvions pas voir l'homme, même si _____

(a) He could have missed the car ride.
Il aurait pu manquer le trajet en voiture.

(b) We will have seen the car there.
Nous aurons vu le chat là-bas.

(c) He was hidden behind a tree.
Il était caché derrière un arbre.

(d) We watched the old house.
Nous avons observé la vieille maison.

L'exemple (1) mesure donc si les participant-e-s possèdent des connaissances explicites de la fonction du connecteur anglais *although* (« bien que ») sur un niveau intra-phrastique. Afin de s'assurer que les participant-e-s puissent accéder à leurs compétences explicites, aucune limite de temps n'est fixée pour l'accomplissement de cette tâche. Afin de tester la connaissance explicite des connecteurs de manière inter-phrastique, on peut utiliser une *tâche de continuation de phrase* (McClure & Geva, 1983), telle qu'illustrée en (2), exemple tiré de Geva (1986 : 88).

2) It was cold outside, although it was sunny.
Il faisait froid dehors, bien qu'il y ait du soleil.

(a) So, it was a good day for skiing.
C'était donc une bonne journée pour faire du ski.

(b) So, Johnny's mother made him wear a sweater.
La mère de Johnny lui a donc fait porter un pull.

Finalement, on peut tester la même compétence avec les connecteurs également au niveau du discours en utilisant un *texte à trous*. Dans une telle tâche, on présente des textes avec des connecteurs omis pour lesquels les participant-e-s

choisissent le bon connecteur parmi quatre options données (Geva, 1986). Une autre mesure de la connaissance explicite consiste dans *les tâches de jugement grammatical* (telles que celles utilisées par Zufferey, Mak, Degand & Sanders, 2015; Zufferey et Gygax, 2017; voir aussi Cho, 2020 : 22). Dans cette approche, on présente des phrases telle que (3) et on demande aux participant-e-s d'évaluer la phrase à l'aide une échelle allant de, par exemple, *correct* à *incorrect*.

3) Marie va à l'école, alors que Peter reste à la maison.

En utilisant cette tâche, on peut présenter également des phrases incorrectes ou incohérentes pour observer si les apprenant-e-s sont sensibles à la détection d'erreurs causées par le connecteur.

En général, les inconvénients de ces tâches sont 1) qu'on ne sait pas si les lecteur-trice-s comprennent les phrases de la manière prévue, 2) que la lecture des phrases est peu naturelle et 3) que le jugement métalinguistique pose une contrainte supplémentaire sur les ressources de traitement ce qui biaiserait éventuellement les résultats (Ortifelli & Polinsky, 2017).

En ce qui concerne la connaissance implicite, on peut distinguer également entre différents types de mesure. Une première possibilité est d'observer le traitement mental d'une phrase et son éventuelle perturbation par l'utilisation incorrecte de connecteurs, en utilisant une tâche de lecture segmentée (par ex. Jiang, 2011; Cho 2020). Dans cette tâche, on présente les segments d'une phrase de manière isolée, comme représenté par les barres obliques dans (4).

4) Ceci // est // un // exemple.

Les segments sont affichés sur un écran d'ordinateur, l'un après l'autre, la personne lectrice doit appuyer sur un bouton (par exemple la barre d'espace) pour que le prochain segment s'affiche. Chaque segment s'efface après que le nouveau segment s'affiche ce qui évite que les participant-e-s reviennent en arrière pour relire certains mots sans que cela soit mesuré (ou même remarqué). Ainsi, on peut retracer le temps de lecture pour chaque segment en mesurant le temps écoulé entre deux activations du bouton. Il existe différentes versions de cette tâche : les *fenêtres mouvantes* (en l'anglais *moving windows*) et des *versions centrées* (en anglais *centered versions*). Tandis que les segments d'une phrase s'affichent dans une *version centrée* à la même position (souvent au milieu de l'écran), les segments

dans une version avec *fenêtres mouvantes* s'affichent l'un après l'autre horizontalement (donc sur la position où ils seraient dans un texte normal). En comparaison avec une version centrée, une version avec des fenêtres mouvantes imite donc plus une lecture normale et naturelle (Just, Carpenter & Woolley, 1982).

Pour donner un exemple comment cette tâche peut être utilisé dans le champ de la linguistique, considérons les phrases (5 et 6)

5) Marie avait // très faim // alors // elle a commandé // une grande pizza // au restaurant.
6) ?Marie avait // très faim // mais // elle a commandé // une grande pizza // au restaurant.

Dans le troisième segment, on trouve une différence entre les deux phrases : nous avons échangé dans (6) le connecteur *alors* contre *mais*, ce qui rend la clause suivante, normalement une conséquence de la première clause, incohérente avec la première. En comparant les temps de lecture pour les segments après le connecteur entre les deux phrases, nous pouvons observer si les personnes intérogées étaient affectées par l'incohérence de (6) et liraient en conséquence cette phrase plus lentement. Comme ces segments sont exactement les mêmes dans les deux phrases, nous pouvons en outre nous assurer que tout ralentissement observé n'est pas dû aux différences entre les phrases.

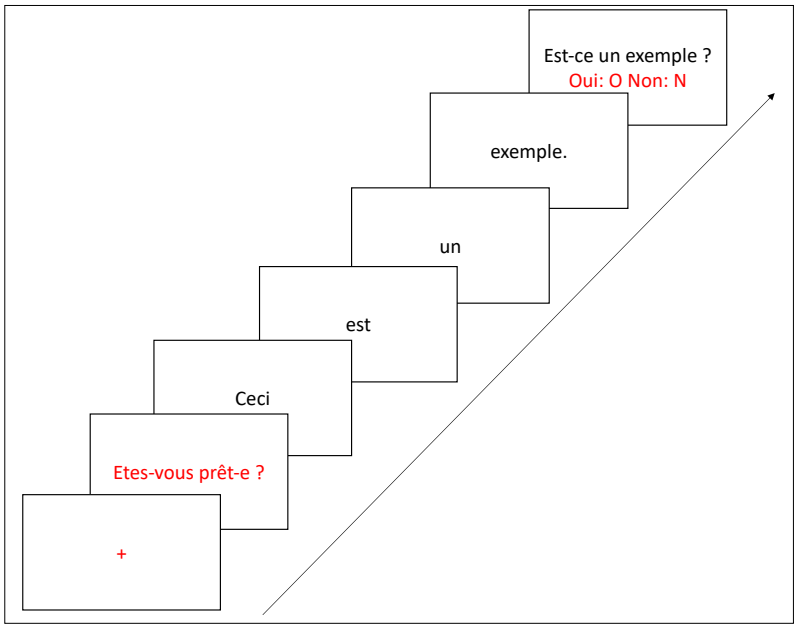
Les personnes lectrices produisent une représentation mentale de la phrase plutôt à la fin ou après avoir lu une phrase (Kintsch & Van Dijk. 1978; Kintsch, 1988; Lyu, Tu & Lin, 2020). Ainsi, poser une question visant le contenu de la phrase tout à la fin, comme dans (7), permet également de mesurer les effets de la perturbation du traitement qui se prolongeraient à ces régions-là. De plus, l'utilisation des questions de vérification permet de s'assurer que les participant-e-s aient bien lu la phrase car si on n'avait pas lu (5) ou (6) on ne saurait pas répondre à la question (7).

7) Est-ce que Marie a commandé une pizza?

Un exemple d'une version centrée de cette tâche, incluant une question de vérification, est visuellement représenté dans la Figure 3.1. On y voit également que les segments de la phrase cible sont introduits par deux écrans : premièrement, le symbole + s'affichant pendant quelques secondes à l'endroit où le contenu sera présent, ainsi que la phrase *Êtes-vous prêt-e ?* à laquelle les participant-e-s doivent

répondre de manière affirmative. Ces deux écrans permettent de focaliser la concentration de la personne lectrice sur ce qui suit.

Figure 3.1. Exemple d'une tâche utilisant la lecture segmentée (version centrée)



L'avantage de la lecture segmentée c'est qu'elle permet de bien retracer le temps de lecture et – grâce aux manipulations effectuées dans la phrase – de documenter dans quels segments les perturbations du traitement mental émergent et combien de temps celles-ci durent. Un autre avantage consiste dans le fait que la lecture segmentée permet de tester par Internet, permettant donc d'avoir accès à un nombre de participant-e-s plus grand, recruté dans un temps court. Évidemment, cela peut également représenter un désavantage car le recrutement manque de contrôle – on ne sait pas qui on teste ou dans quelles conditions les personnes se trouvent. Le seul moyen pour pouvoir tirer des conclusions valables des données c'est donc un nombre de participant-e-s assez grand qui élimine (ou, au moins diminue) ce biais potentiel. Un autre désavantage est que les participant-e-s ne peuvent pas aller en arrière pour relire et vérifier certains passages qu'ils viennent de lire. Pour observer cela, des mesures plus fines sont nécessaires, telle que les appareils d'oculométrie.

Les appareils d'oculométrie représentent des mesures encore plus fines pour mesurer le traitement mental car ils permettent de retracer la lecture et permettent de regarder en arrière, simulant donc à une plus grande extension (en comparaison avec la lecture segmentée) une lecture naturelle. Leur principe est simple : on mesure le point de fixation de chaque œil de la personne participante à l'aide des

petites caméras (voir Holmqvist, Nyström, Andersson, Dewhurst, Halszka & van de Weijer, 2011, pour un résumé). Ces caméras peuvent placées, par exemple, directement sous les yeux avec des lunettes. Afin de mesurer comment une personne traite un discours, un texte est affiché sur un écran et les participant-e-s portant un appareil d'oculométrie sont demandé-e-s de le lire. En mesurant les saccades des points de fixations de leur lecture, on arrive à identifier les zones de texte qui sont éventuellement problématique pour le traitement mental, car les personnes y ralentissent leur vitesse de lecture et reviennent plusieurs fois. L'enregistrement de la lecture avec un appareil d'oculométrie permet donc la récupération des différents types de données (Rayner, 2009; Holmqvist et al., 2011): le temps de lecture au premier passage (ou FPRT d'après l'angl. *first-pass-reading time*), le temps de lecteur au deuxième passage (ou SPRT d'après l'angl. *second pass reading time*) ou bien – similaire aux mesures de la lecture segmentée - la durée total de la lecture (TRT d'après l'angl. *total reading time*). Un des plus grands désavantages de ces mesures du traitement mental est que la lecture des personnes se déroule dans une situation très peu naturelle. Il n'est donc pas exclu que d'autres biais apparaissent en raison du caractère artificiel de la situation ou que les personnes lectrices adoptent un comportement un peu différent de leur lecture habituelle.

Une autre méthode qui permet de retracer le traitement mental lors de la lecture est celle du *potentiel évoqué* (ou ERG, en anglais *Event-Related Potential*). Dans cette méthode, on mesure les processus neuronaux qui émergent directement après la lecture d'un stimuli. Notamment, on différencie dans le cas du langage entre deux réactions : l'une qui émerge à 400ms et l'autre qui émerge à 600ms après l'avoir lu (Kutas & Federmeier, 2011; Gouvea, Phillips, Kazanina & Poeppel, 2010). Il semble que l'effet observé à 400ms (nommé *N400*) soit lié à l'activation de l'information lexicale d'un mot tandis que l'effet émergeant à 600ms (nommé *P600*) soit lié à son intégration sémantique (Brower, Fitz & Hoeks, 2012). Ensemble, ces deux réactions permettent d'investiguer les effets qui sont liés à la cohérence ou incohérence d'un discours (p.ex., Drenhaus, Demberg, Köhne & Delogu, 2014; Xiang & Kuperberg, 2015).

Outre la mesure du traitement mental, les connaissances implicites peuvent également être mesurées d'autres manières. En effet, en imposant une pression temporelle, l'accès aux connaissances explicites est restreint et les personnes non-

natives doivent se fier à leurs connaissances implicites. Ainsi, la connaissance implicite peut être également testée en utilisant les questions à trous (comme dans Geva, 1986) ou tâches de reconnaissance de mots (comme dans Dufva & Voeten, 1999) si les participant-e-s sont soumis-e-s à une forte pression temporelle. Cela s'applique même aux productions de texte : pendant la rédaction d'un texte sous pression temporelle, les auteur-trice-s sont obligé-e-s de concevoir un texte sans avoir la possibilité de le lire, de le réévaluer et de le corriger, et ils doivent à la place se servir de leurs connaissances implicites (Yuan & Ellis, 2003; voir aussi Ellis, 2006, 2009a).

Après avoir défini les différents domaines de maitrise et présenté leur opérationnalisation, il est important de se demander également à quoi une maitrise des connecteurs en L2 peut être comparée, ou en d'autres termes, ce qui constitue un bon usage des connecteurs. Il est vrai que les personnes natives[13] sont souvent considérées comme des expertes de leur langue et leur utilisation de la L1 est l'optimum que les apprenant-e-s de cette langue devraient atteindre. Toutefois, la recherche en linguistique a démontré que la maitrise des connecteurs des personnes natives se présente sous forme très hétérogène (p.ex., Lamiroy, 1994). Par exemple, dans l'étude de Tskhovrebova, Zufferey & Gygax (2021, voir aussi p.ex., Nippold et al., 1992, Cain & Nash, 2011), des jeunes adolescent-e-s maitrisent moins bien les connecteurs moins fréquents - un effet qui se prolonge même jusqu'à l'âge adulte (Zufferey & Gygax, 2020b). Évidemment, cela pose un problème pour la recherche en acquisition d'une langue seconde, car il est nécessaire de disposer d'une valeur de référence stable permettant de comparer la maitrise des connecteurs en L2. Ceci soulève ainsi la question suivante: quelle est la norme pour une bonne maitrise des connecteurs ? Et par conséquent, à quoi peut-on donc comparer un usage d'une personne non-native ? Pour répondre à ces questions, différentes approches ont été trouvées, qui permettent tout de même de qualifier et d'évaluer la compétence des personnes non-natives avec des connecteurs.

[13] Rappelons que la distinction entre *natif et non-natif* n'est pas claire et que même, comme démontré par Cook (1999), la notion d'une personne *native* n'est pas non plus forcément univoque (Cook, 1999:186, voir aussi Tsehaye, Pashkova, Tracy & Allen, 2021). Dans le travail présent, une personne est considérée étant native lorsqu'elle a grandi avec cette langue dès sa toute première enfance tandis qu'une personne non-native l'a apprise à un moment ultérieur de sa vie.

Premièrement, on peut tester les compétences des personnes avec les connecteurs à la fois dans leur L1 ainsi que dans leur L2. Cela aurait comme avantage qu'on tiendrait compte du fait qu'il y a des différences considérables parmi les personnes natives (voir chapitre 2.5.2). Toutefois, un biais important de cette approche sont les différences entre les langues. Comme discuté auparavant, les connecteurs possèdent souvent un profil pragmatique spécifique qui dépend fortement de la langue en question (voir chapitre 2.5.4) ce qui peut rendre une comparaison crosslinguistique difficile.

Deuxièmement, comme le recommandent Bolton et al. (2002), on peut établir une bonne norme d'utilisation des connecteurs en considérant seulement les textes des auteur-trice-s natifs qui publient au niveau académique. Dans leur étude, les auteurs ont analysé l'utilisation des connecteurs dans 40 échantillons d'articles et de livres universitaires rédigés par des universitaires professionnels natifs. Selon Bolton et al., (2002) ces cas représentent la meilleure utilisation possible des connecteurs dans une langue donnée, et donc une bonne valeur de comparaison. Cette approche est pourtant désavantageuse pour l'approche expérimentale et ceci pour des nombreuses raisons (comme par exemple la difficulté de recruter un grand nombre des participant-e-s).

Lors d'adoption d'une méthodologie expérimentale, il convient plutôt de comparer la maitrise des connecteurs en L2 à la maitrise des connecteurs par un public natif, supposé représenter en moyenne la norme d'une variété de langue (comme dans Crible et al., 2021). Moyennent un nombre de participant-e-s assez important, cela a pour avantage de permettre de refléter une norme linguistique que la plupart des locuteur-trice-s non-natifs souhaitent atteindre et qui tient également compte de la variance au sein de la population native. En outre, il est possible de transposer certaines caractéristiques du groupe de personnes non-natives testé au groupe de personnes natives. Cela signifie, par exemple, que l'on compare un groupe de jeunes élèves non-natifs à un groupe de jeunes élèves natifs (comme p.ex., dans Crosson & Lesaux, 2013) ou bien un groupe d'étudiant-e-s non-natifs et natifs (comme p.ex., dans Field & Yip, 1992). Cela est particulièrement utile dans les cas où les caractéristiques du groupe sont soupçonnées d'avoir une influence sur la performance dans la tâche.

En résumé, nous avons vu que de nombreuses techniques ont été développées pour mesurer la maitrise - conceptualisée en termes de

connaissances explicites et implicites - des connecteurs discursifs dans une langue seconde. Nous revenons maintenant à la question de savoir si et comment les personnes non natives peuvent bénéficier des connecteurs.

3.2 L'utilité des connecteurs et des signaux alternatifs pour le traitement et la compréhension d'un texte en L2

En tenant compte du fait que les connecteurs sont des éléments qui peuvent faciliter la lecture et la compréhension d'un texte pour les personnes natives (p.ex., Degand et al. 1999; van Silfhout et al., 2014), la question se pose de savoir si ces mots aident également les personnes qui sont linguistiquement plus faibles et qui ont généralement plus de difficultés à comprendre les liens logiques d'un texte (Cohen, Ferrar & Fine 1979), à savoir les personnes non-natives.

De manière générale, certains travaux ont mentionné le fait que les apprenant-e-s d'une L2 ne profitent pas autant que les personnes natives des connecteurs (Berman 1979; Cooper 1984). Geva (1986), par exemple, a documenté que les personnes non-natives inféraient mieux les relations de cohérence dans des phrases qui ne contenaient pas de connecteurs que dans des phrases qui en contenaient. Pourtant, lorsqu'elle les phrases testées contenaient des connecteurs qui étaient marqués en gras, les personnes plus compétentes en L2 profitaient des connecteurs et infèreraient mieux les relations de cohérence en général. Le lien entre le bénéfice des connecteurs pour la compréhension d'un texte et la compétence linguistique a été également démontré par Degand et Sanders (2002). Dans leur étude, ils ont présenté des textes dans lesquels les relations causales contenaient soit un marquage explicite ou implicite des connecteurs. Après avoir lu ces phrases, des personnes natives et non-natives ont dû répondre à des questions de compréhension. Les résultats ont montré que les deux groupes ont profité de la présence des connecteurs, même si aucune différence n'a été trouvée entre les groupes linguistiques, c'est-à-dire que les personnes non-natives ne bénéficieraient pas plus que les personnes natives des connecteurs. Ceci est surprenant car il a été démontré que les personnes natives moins compétentes bénéficient plus des connecteurs (p.ex., Marshall & Glock, 1978; Meyer et al., 1980). Même si les lecteur-trice-s encore moins compétent-e-s, à savoir les personnes non-natives, ne bénéficient pas davantage des connecteurs que les

personnes natives, on peut constater, contrairement aux observations de Geva (1986), qu'ils en bénéficient tout de même. Selon Degand et Sanders (2002), cela a été rendu possible par le fait que les participant-e-s avaient une bonne compétence linguistique dans leur langue cible, un effet qui s'était donc déjà manifesté faiblement chez Geva (1986, voir aussi Geva & Ryan, 1993). Toutefois, il reste surprenant que les participant-e-s compétent-e-s de Geva (1986) ne comprenaient pas des textes qui contenaient des connecteurs mieux que les textes qui n'en contenaient pas, alors que cet effet a été trouvé par Degand et Sanders (2002).

A part les connecteurs, nous avons discuté dans le chapitre 2.4.2 que les lecteur-trice-s peuvent profiter également d'autres moyens pour l'insertion des relations de cohérence, comme par exemple les signaux alternatifs. Scholman et al. (2017), par exemple, ont montré que les personnes natives s'attendaient à des relations de cohérence spécifiques lorsque celles-ci étaient amorcées par des indices spécifiques, c'est-à-dire qu'ils anticipaient un contraste après avoir lu le connecteur *on the one hand* (« d'une part »). Dans une étude ultérieure, Scholman et al. (2020) ont également observé que les lecteur-e-s natives ont été sensibles aux indices lexicaux et qu'ils basent leurs prédictions sur les relations de cohérence à venir sur ces indices. Dans cette étude, les participant-e-s ont produit, dans une tâche de complétion de phrases, beaucoup plus de listes lorsque les phrases contenaient des signaux de listes (c'est-à-dire des mots tels que *plusieurs*, *quelques*). Quant à l'acquisition d'une L2, la question se pose de savoir si les personnes non-natives peuvent également profiter de ces signaux alternatifs pour inférer des relations de cohérence. Malheureusement, jusqu'à présent, on manque des données pour répondre à cette question avec certitude, car, à notre connaissance, seule l'étude de Crible et al., (2021) fournit des indications à cet égard. Dans cette étude, l'équipe de recherche a investigué si les structures parallèles facilitent l'intégration des relations contrastives. On entend par une structure parallèle une suite d'éléments avec une structure syntaxique similaire et une répétition lexicale comme illustré dans (8), exemple tiré de Crible et al. (2021)

8) Benjamin veut habiter en ville. Par contre, Alex veut habiter à la campagne.

Ici, outre l'emploi du connecteur *par contre*, la répétition des verbes *veut habiter* pourrait faciliter l'inférence d'une relation de contraste. Comme l'effet de facilitation de ces structures a été démontré pour les personnes natives (Crible & Pickering, 2020), Crible et collègues (2021) ont formulé l'hypothèse que les personnes non-natives pourraient également profiter de ces signaux. La raison en est que les personnes non-natives sont moins sensibles à l'identification des relations de cohérence et des indices discursifs en général, car ils doivent activer et décoder les sens de mots avant de les aligner dans un tout cohérent. Comme cela représente un effort considérable dans une L2 (Clahsen & Felser, 2006a, 2006b), tout signal de facilitation, tels que les structures parallèles, seraient une aide pour atteindre la compréhension d'un texte. De plus, on peut supposer que les structures parallèles indiquant un contraste ne sont pas un phénomène spécifique à une langue, mais sont également accessibles aux locuteur-trice-s de différentes langues. Cela ferait de ces structures un signal simple pour les personnes non-natives les aidant à identifier les relations de cohérence.

Pour tester donc cette hypothèse, Crible et al. (2021) ont mesuré le traitement mental des personnes natives et non-natives du français en utilisant la méthode de la lecture segmentée. Les participant-e-s lisaient des phrases contrastives qui ont été modifiées au cours de trois expériences.

Dans la première expérience, les phrases ont été modifiées selon design factoriel 2x2 dont les facteurs étaient la présence des connecteurs (*absents*, *présents*) et la présence d'une structure parallèle (*absente*, *présente*). Les résultats de la première expérience montrent que les personnes non-natives bénéficient des structures parallèles lorsqu'elles n'ont pas accès au signal du connecteur. Si le connecteur est disponible, son signal est suffisant pour inférer la relation. Dans la deuxième expérience, l'équipe a de nouveau testé les effets des structures parallèles (à nouveau absentes ou présentes dans les phrases) en interaction avec le type de connecteur : les phrases contenaient en plus soit le connecteur *en revanche*, soit le connecteur *par contre*. Les résultats de cette deuxième expérience ont montré à nouveau l'effet de facilitation des structures parallèles pour la lecture en L2, tandis qu'une différence entre la lecture pour les deux connecteurs émergeait seulement temporairement au milieu de la phrase. Finalement, dans la troisième expérience, les connecteurs utilisés dans la deuxième expérience ont été échangés contre les deux connecteurs *et* et *mais*. Tandis que le connecteur

polyfonctionnel *mais* peut indiquer, entre autres, une relation contrastive, le connecteur sous-spécifié *et* est sémantiquement additif. Ainsi, les lecteur-trice-s devaient insérer un sens augmenté pragmatiquement du connecteur pour établir la relation contrastive. Les résultats de cette troisième expérience ont montré que les personnes non-natives bénéficiaient de la structure parallèle pour établir la relation contrastive, tandis que le connecteur ambigu *et* posait un grand problème pour l'inférence de la relation. Seulement lorsque la phrase contrastive était renforcée par la structure parallèle, son traitement a été facilité.

En résumé, deux conclusions majeures pour la lecture non-native peuvent être tirées des observations faites par Crible et al. (2021). Premièrement, les personnes non-natives profitent des structures parallèles pour inférer des relations contrastives. Comme discuté par Crible et al. (2021), il est cognitivement exigeant de lire dans une L2 et tout signal qui aide à déduire la relation de cohérence, à condition qu'il soit accessible, est donc le bienvenu. Deuxièmement, les connecteurs sont également des bons repères pour les personnes non-natives pour insérer la relation de cohérence. Comme pour la structure parallèle, cela ne s'applique que lorsque le connecteur est *accessible*. Alors que le connecteur issu d'un registre élevé *en revanche* ne provoquait qu'une perturbation temporaire du traitement, le connecteur sous-spécifié *et* rendait l'inférence de la relation beaucoup plus difficile.

En conclusion, les indices discursifs tels que les connecteurs ou les signaux alternatifs peuvent aider les personnes non-natives à mieux identifier les relations de cohérence et donc à mieux comprendre un texte (Degand & Sanders, 2002; Crible et al., 2021). Pour être en mesure de décoder le sens d'un connecteur et d'accéder à sa fonction procédurale, il semble toutefois important que la personne ait un niveau linguistique élevé dans sa L2 (nous allons revenir sur ce point dans la section 3.5), ce qui lui permettrait d'accéder au sens d'un connecteur et d'activer sa fonction procédurale. Cependant, comme nous allons le voir maintenant, il semble qu'il soit extrêmement difficile pour les personnes non-natives de bien maitriser les connecteurs dans une langue étrangère.

3.3 La maitrise insuffisante des connecteurs par les personnes non-natives

Bien que les personnes non-natives puissent bénéficier des connecteurs (tant qu'elles ont une bonne compétence linguistique), il existe des forts indices qu'à l'inverse une bonne utilisation des connecteurs en L2 n'est que très difficilement atteinte. Pour en donner un exemple simple, il suffit de regarder les trente mots le plus recherchés dans la version française du dictionnaire de langue en ligne *Linguee* (*https://www.linguee.com*, consulté le 01.06.22). Comme présenté dans la Figure 3.2, on trouve un total de *quinze* connecteurs[14] parmi ces trente mots ce qui indique les grands problèmes (ou au moins des incertitudes) des apprenant-e-s de français par rapport aux connecteurs.

Figure 3.2. Capture d'écran de *Linguee* (https://www.linguee.com/french-english/topfrench/1-200.html, consulté le 01.06.22) : les mots le plus recherchés sur le site anglais pour le français. Marquage des connecteurs en rouge par l'auteur.

1.: en effet	2.: ainsi	3.: notamment	4.: de plus	5.: alors que
6.: par ailleurs	7.: cependant	8.: mise en place	9.: surtout	10.: dont
11.: dans le cadre de	12.: selon	13.: enfin	14.: mettre en place	15.: ainsi que
16.: formation	17.: donc	18.: soit	19.: en fonction de	20.: pourtant
21.: bilan	22.: d'ailleurs	23.: par contre	24.: cordialement	25.: bien que
26.: proposer	27.: cahier des charges	28.: prise en charge	29.: fonctionnement	30.: location

La difficulté des personnes non-natives avec les connecteurs est également étayée par de nombreuses observations de différentes études de corpus. Par exemple, il a été documenté que les apprenant-e-s d'une L2 utilisent une plus petite variété de connecteurs que les personnes natives (p.ex., Leedham & Cai, 2013; Park, 2013; Anderson, 2014; Chen, 2014; Adeyemi, 2017; Shi, 2017; Don & Sriniwass, 2017) et plusieurs études rapportent que les apprenant-e-s utilisent fréquemment les connecteurs de manière déviante par rapport aux personnes natives, comme par exemple une utilisation plus importante des connecteurs familiers (p.ex., Field & Yip, 1992; Zhang, 2000; Casteele & Collewaert, 2013; Leedham & Cai, 2013; Lee, 2013; Anderson, 2014; Chen, 2014; Ha, 2016; Shi, 2017; Don & Sriniwass, 2017) ou la tendance à placer les connecteurs en position initiale de la phrase (p. ex., Leedham & Cai, 2013; Park, 2013; Don & Sriniwass, 2017; Shi, 2017). Les études montrent par ailleurs que les personnes non-natives surutilisent certains

[14] Classification des connecteurs selon le dictionnaire de connecteurs LexConn (Roze et al., 2012)

connecteurs (Field & Yip, 1992; Milton & Tsang, 1993; Granger & Tyson, 1996; Zhang, 2000; Lee, 2013 Leedham & Cai, 2013; Park, 2013; Chen, 2014; Ha, 2016; Adeyemi, 2017; AlAttar & Abu-Ayyash, 2020) tandis que d'autres connecteurs semblent être sous-utilisés (par ex, Field & Yip, 1992; Granger & Tyson, 1996; Lee, 2013; Park, 2013; Chen, 2014; Ha, 2016; Gao, 2016; Shi, 2017). Cela peut s'expliquer par l'interprétation selon laquelle les apprenant-e-s auraient tendance à s'appuyer sur un ensemble limité de *mots de confort* (d'après anglais 'teddy bear words'; Hasselgård, 2019, 2019), c'est-à-dire des mots qu'ils sont à l'aise d'utiliser (comme observé et discuté dans Casteele & Collewaert, 2013; Leedham & Cai, 2013; Don & Sriniwass, 2017). Toutefois, on ne sait que peu de choses sur les mots qui semblent les plus appropriés pour devenir de tels mots de confort. A part les sous- et surutilisations des connecteurs dans les productions de texte en L2, de nombreuses études montrent également des usages incorrects des connecteurs (p.ex., Field & Yip, 1992; Milton & Tsang, 1993; Granger & Tyson, 1996; Zhang, 2000; Lee, 2013; Park, 2013; Hamed, 2014; Ha, 2016; Gao, 2016; Don & Sriniwass, 2017; AlAttar & Abu-Ayyash, 2020). Crewe (1990), par exemple, note que les connecteurs sont souvent utilisés comme des « exhausteurs stylistiques » (1990 : 316, ma traduction), utilisés n'importe où dans le texte afin de lui donner une impression superficielle de cohérence.

Tandis que les études citées documentent la maitrise clairement insuffisante des connecteurs de discours dans les productions textuelles en L2, les résultats des études expérimentales montrent des résultats en peu plus nuancés. Par exemple, en ce qui concerne les connaissances explicites, Steffani & Nippold (1997) ont analysé comment des étudiant-e-s d'origine japonaise aux États-Unis maitrisaient les connecteurs de l'écrit. Dans une première tâche, les personnes non-natives ont dû compléter un petit texte en écrivant une phrase finale introduite par un connecteur rare et utilisé que dans le mode écrit. Dans une deuxième tache, les participant-e-s ont choisi le connecteur approprié pour un petit texte à trous parmi quatre options données. Comme on pouvait s'y attendre, les résultats de ces deux taches montrent les problèmes majeurs des apprenant-e-s non-natifs dans la maitrise des connecteurs : 90% des étudiant-e-s japonais n'ont pas obtenu un score de plus que 90% dans la première tâche (10% pour les natifs). En outre, la première tâche a été particulièrement difficile, car celle-ci demandait, selon les chercheuses une connaissance plutôt *active* (i.e., production explicite), tandis que la

connaissance de la deuxième tâche était plus *passive* (i.e., choix entre plusieurs options).

Les difficultés éprouvées par les apprenant-e-s d'une L2 avec les connecteurs ont été également documentées dans une autre étude expérimentale. En 2015, Zufferey et collègues ont testé dans quelle mesure les apprenant-e-s d'une L2 étaient capables de détecter des emplois erronés des connecteurs dans une langue seconde. A cette fin, l'équipe de recherche a présenté des phrases en anglais aux apprenant-e-s de langue maternelle néerlandaise et française qui contenaient des erreurs due à une confusion crosslinguistique des fonctions des connecteurs. Plus précisément, les chercheurs ont présenté des phrases comme (9), donc des phrases dans lesquelles le connecteur *if* (« si » ou « quand ») était utilisé pour indiquer un contraste.

9) * Many products are exported from Switzerland at various scales. If Swiss chocolate is universally known, Swiss wines are known only by the specialists.
** De nombreux produits sont exportés de Suisse à différentes échelles. CONNECTEUR le chocolat suisse est universellement connu, les vins suisses ne sont connus que des spécialistes.*

On a supposé que cette utilisation erronée de *if*, créerait une confusion chez les francophones, car on a supposé qu'ils feraient une mauvaise mise en équivalence du connecteur *if* avec le connecteur français *si* (ou *transfert négatif*, voir chapitre 8). Contrairement au connecteur anglais *if*, le connecteur français *si* peut être utilisé pour indiquer un contraste. Les chercheur-se-s ont donc supposé qu'il était plus difficile pour les francophones de détecter l'incohérence. Et en effet, les participant-e-s n'ont pas été capables de détecter les utilisations erronées des connecteurs dans une tâche d'évaluation sans contrainte de temps, c'est-à-dire dans une tâche qui leur permettait d'accéder à leurs connaissances explicites. Pourtant, lors de la lecture avec un appareil d'oculométrie, les même participant-e-s francophones étaient affectés par la fausse utilisation de *if*. Ainsi, il semble que la détection des erreurs de connecteurs malgré les influences crosslinguistiques dépende fortement du type de connaissances auxquelles les apprenant-e-s accèdent. Par ailleurs, ces résultats sont surprenants car d'autres études ont montré que les personnes non-natives ont tendance à avoir de meilleures performances lorsqu'elles accèdent à leurs connaissances explicites (p.ex., Roberts, Gullberg & Indefrey, 2008). Pour expliquer le fait que les participant-e-s étaient moins sensibles aux emplois erronés

des connecteurs durant dans les tâches d'évaluation, tandis qu'ils et elles les détectaient lors de la lecture, Zufferey et al., (2015) ont envisagé que la charge du traitement mental des connecteurs puisse être différente de celle d'autres phénomènes linguistiques pour lesquels l'effet inverse a été trouvé. Une autre explication pour ces résultats provient de l'hypothèse selon laquelle les tâches d'évaluation sont plus difficiles que les tâches de lecture car elles impliquent l'activation des connaissances méta et leur verbalisation (Ortfielli & Polinsyk, 2017). Cependant, ceci n'explique pas pourquoi d'autres études, comme Roberts et al. (2008), ont trouvé que les personnes non-natives arrivaient mieux à résoudre des pronoms durant les tâches qui leur permettaient d'accéder à leurs connaissances explicites. Même si la généralisation des résultats de Zufferey et al. (2015) n'est donc pas entièrement assurée, on peut conclure que leurs observations montrent à la fois les difficultés des apprenant-e-s avec les connecteurs de la L2 ainsi que l'importance de tenir compte de la distinction entre les connaissances implicites et explicites.

Dans une autre étude, Zufferey et Gygax (2017) ont examiné si le traitement mental ainsi que l'évaluation des phrases par des apprenant-e-s de français étaient affectés par l'utilisation du connecteur *en effet.* Dans deux tâches, à savoir une tâche de lecture segmentée et une tâche d'évaluation, ils ont présenté des phrases qui contenaient soit une relation de conformation, soit une relation causale. Ces deux relations ont été indiquées soit explicitement par le connecteur *en effet,* soit elles ont été laissées implicites. Malgré le fait qu'une confirmation nécessite un marquage explicite, les participant-e-s non-natifs ont évalué les phrases contenant une confirmation sans connecteur plus cohérente que celles qui ont été laissées implicites. De même, lors de la tâche de la lecture segmentée, les participant-e-s non-natifs n'ont pas réagi à l'absence du connecteur dans cette condition comme le groupe natif. Ainsi, malgré le fait que *en effet* soit assez fréquent, les participant-e-s non-natives ne maitrisaient pas ce connecteur de manière satisfaisante. Les auteur-trice-s expliquent cette observation par la complexe mise en correspondance entre la forme et les fonctions de *en effet*. A cet égard, « les futures études expérimentales devraient chercher à déterminer si les mises en correspondance de la forme et fonction plus complexes causent plus de difficultés aux apprenant-e-s » (Zufferey & Gygax, 2017 : 9, ma traduction).

Prises ensemble, toutes les études citées indiquent que les apprenant-e-s d'une langue ont beaucoup de difficulté de bien maitriser les connecteurs de discours. Cela se montre à la fois dans la production de textes ainsi que leur connaissances implicites et explicites. Il est intéressant de noter à cet égard que ces difficultés persistent apparemment même pour les apprenant-e-s qui sont hautement compétent-e-s dans leur langue seconde. Des emplois déviants de connecteurs ont été trouvés par exemple dans des écrits universitaires publiés (p.ex., Carrió-Pastor, 2013; Gao, 2016; Uçar & Yükselir, 2017).

Inévitablement, toutes les observations mentionnées dans cette section soulèvent la question suivante : pourquoi les apprenant-e-s d'une langue seconde ont-ils autant des difficultés à bien maitriser les connecteurs ? Ceci est autant plus curieux que les connecteurs existent souvent également dans la L1 des apprenant-e-s (cf. Zufferey et al., 2015). Ainsi, le seul fait que les connecteurs soient des éléments procéduraux et non conceptuels ne devrait pas rendre leur maitrise difficile en soi. Afin de pouvoir élucider pourquoi les apprenant-e-s éprouvent des problèmes avec les connecteurs, il convient de s'interroger d'abord sur la manière dont ces éléments sont enseignés.

3.4 Le rôle de l'enseignement

En général, il est bien connu que l'enseignement et les facteurs pédagogiques qui y sont liés influencent fortement l'acquisition d'une langue seconde. La recherche a examiné à cet égard l'impact des facteurs comme la motivation (p.ex., Gardner & Lambert, 1959; Mahmoodi & Yousefi, 2021), l'ennui pendant l'apprentissage d'une langue (p.ex., Li, Dewaele & Hu; 2021) ou bien la méthode d'acquisition (p.ex., Oberg & Daniels, 2013).

Quant aux connecteurs, on peut assumer que l'enseignement insuffisant représente une raison importante pourquoi les connecteurs sont si faiblement maitrisés. Ainsi, de nombreuses études de corpus expliquent la mauvaise maitrise documentée par un enseignement insuffisant et proposent des recommandations pour améliorer les cours (p. ex., Milton & Tsang 1993; Lee, 2013; Park 2013; Anderson, 2014; Hamed, 2014; Ha, 2016; Don & Sriniwass 2017). Selon Zamel (1983), la méthode la plus répandue pour enseigner les connecteurs est la présentation des listes. La chercheuse élabore :

> « (…) malgré le rôle essentiel que jouent les connecteurs dans la rédaction, les étudiant-e-s de langue anglaise ne sont pas toujours en mesure d'en tirer parti. Cela peut être dû principalement au fait qu'on ne leur a pas appris à les identifier pendant la lecture ou à les utiliser correctement dans leurs productions d'écrits. Ce qu'on leur a proposé, si les textes de composition sont le reflet de nos stratégies d'enseignement, ce sont des listes d'éléments cohésifs, classés par fonction » (1983 : 23, ma traduction)

Pourtant, ces listes seraient « trompeuses », « absurdes » et « inefficace comme aide à l'enseignement » (Zamel, 1983:24, mes traductions) car elles posent de nombreux problèmes pour les apprenant-e-s. Elles ne tiendraient par exemple pas compte des connecteurs qui peuvent indiquer des relations de cohérences différentes (comme *en effet* qui indique une confirmation et causalité), ni du contexte discursif. Ainsi, plusieurs chercheur-se-s soutiennent que les listes qui présentent des connecteurs comme des éléments interchangeables auxquelles on peut facilement attribuer des fonctions spécifiques devraient être évitées (Granger & Tyson, 1996; Mohamed, 2016). Par ailleurs, nous avons vu dans le chapitre précédent que les apprenant-e-s font une sur- et sous-utilisation des connecteurs ce qui ne montre pas seulement à quel point les connecteurs ne sont pas maitrisés, mais aussi qu'il existe une insécurité concernant le marquage des relations de cohérence en général. En conséquence, l'enseignement sur le marquage des relations de cohérence ne devrait pas se concentrer seulement sur les connecteurs, mais traiter également d'autres moyens, tels que des paraphrases et pronoms démonstratifs (Cho & Shin, 2014). De plus, l'enseignement devrait également aborder quand et pourquoi une relation de cohérence peut être laissée implicite et l'utilisation d'un connecteur devient superflue (Zamel, 1983; Granger & Tyson, 1996; Anderson, 2014).

Tenant compte de l'importance de l'enseignement pour la maitrise des connecteurs en L2, peut-on en déduire inversement qu'un meilleur enseignement assure automatiquement une meilleure compétence des apprenant-e-s avec les connecteurs? La réponse n'est pas si évidente. Si l'enseignement était le seul facteur qui permettait une bonne maitrise des connecteurs, il devrait influencer l'ensemble des élèves de la même manière. Or, la recherche a pu montrer qu'il existe des différences dans la maitrise des connecteurs au sein du groupe de personnes non-natives (p.ex., Geva, 1986; Kiany & Khezri Nejad, 2001; Cho & Shin 2014). Il doit y avoir donc des facteurs interpersonnels qui rendent les connecteurs plus difficiles pour certaines personnes, et inversement, qui facilitent la maitrise des connecteurs pour certaines personnes comme nous allons le voir maintenant.

3.5 La variation interpersonnelle

En général, les différences individuelles entre les personnes ont été négligées depuis longtemps dans la recherche psycholinguistique et ont été décrites comme étant « une vérité qui dérange » (Kidd, Donnelly & Christiansen, 2018 : 1, ma traduction). Pourtant, les facteurs interpersonnels sont supposés pouvoir expliquer les résultats hétérogènes au sein d'un groupe de participant-e-s et ceci dans des dimensions linguistiques différentes :

> « Les différences individuelles sont évidentes dans toutes les composantes du système linguistique, de l'acquisition et du traitement des propriétés de la parole aux structures grammaticales complexes et au discours » (Kidd et al., 2018 : 1, ma traduction).

La question se pose donc de savoir quels facteurs interpersonnels pourraient prédire une meilleure compétence en matière de connecteurs et expliquer la variation observée dans leur utilisation au sein des classes. Malheureusement, la recherche n'en a guère étudié jusqu'à présent, bien que de nombreux facteurs puissent être en question à cet égard. Dans ce qui suit, nous allons discuter brièvement deux facteurs à titre d'exemple, qui pourraient influencer la maitrise des connecteurs en L2. Il est bien entendu qu'outre ces exemples, de nombreux autres facteurs peuvent aussi influencer la manière dont les apprenant-e-s de L2 acquièrent des connecteurs, comme par exemple le statut socio-économique (Kidd et al. 2018), les capacités cognitifs (Skehan & Foster, 2001) ou l'exposition à l'écrit (Zufferey & Gygax, 2020b, Zufferey et al., 2022).

Le premier facteur interpersonnel qui pourrait influencer la maitrise des connecteurs est l'âge. On peut supposer que non seulement l'âge d'acquisition (le moment de la vie où les gens apprennent les connecteurs) ait une influence sur l'acquisition et la maitrise des connecteurs en L2, mais également l'âge de la personne apprenante. Comme les apprenant-e-s sont parfois encore en développement cognitif lors (ou même après) de l'acquisition de leur L2, on peut assumer que la maitrise des connecteurs en est influencée : de nombreuses études spécialisées dans l'acquisition des connecteurs en L1 ont montré que l'acquisition des connecteurs par les personnes natives est liée à leur développement cognitif (p.ex., Cain & Nash, 2011; Vermeul & Sanders, 2009, 2011; Pyykkönen & Järvikivi, 2012; Tskhovrebova et al., 2021; voir aussi *l'hypothèse de développement,* p.33). Cela signifierait par exemple, que les jeunes apprenant-e-s auraient - tout comme leurs pairs natifs (Nippold et al., 1992) - plutôt des problèmes avec les connecteurs

plus complexes tandis que les connecteurs plus simples (et ceux qui sont acquis tôt dans la vie par les personnes natives), ne devraient pas représenter un défi en particulier. A cet égard, il est à noter que l'âge devrait être considéré en interaction avec le développement cognitif en général et que la corrélation entre les deux dépend également de la personne en question : il est tout à fait possible qu'une jeune personne aux capacités mentales élevées puisse, malgré son âge, maitriser des connecteurs plus difficiles, alors que ses pairs ont encore des difficultés. Cependant, l'influence de l'âge des apprenant-e-s sur l'acquisition et la maitrise des connecteurs en L2 est encore très spéculative, car il existe peu de littérature sur l'acquisition des connecteurs d'une L2 à différents stades de la vie.

Le deuxième facteur (et probablement même un des plus importants) qui pourrait prédire la facilité avec laquelle un-e apprenant-e est capable de maitriser les connecteurs dans une L2 est sa compétence linguistique dans cette langue. Évidemment, la maitrise des connecteurs fait déjà partie intégrante de la compétence linguistique et même si on considère ces deux compétences comme étant deux compétences distinctes, une corrélation entre elles est en soi peu surprenante. La recherche dans ce domaine a examiné donc plutôt la maitrise des connecteurs du point de vue des différents types de connaissances (implicites ou explicites), de la question du niveau de langue requis pour atteindre un niveau de langue maternelle dans les deux dimensions ou encore (comme nous l'avons déjà vu) l'avantage des connecteurs pour la lecture en L2 en fonction de la compétence linguistique.

De manière plus générale, il est connu que la compétence linguistique influence la manière dont les apprenant-e-s lisent et comprennent un texte. Plusieurs études ont montré par exemple l'existence d'une corrélation entre une compétence linguistique élevée et une meilleure capacité à utiliser des stratégies de lecture (Barnett, 1988; Zhang & Seepho, 2013; Feller, Kopatich, Lech & Higgs, 2020). Concernant les connecteurs plus concrètement, l'étude mentionnée de Geva (1986) a démontré que les participant-e-s ayant une compétence orale plus élevée étaient plus susceptibles de montrer une meilleure connaissance explicite des connecteurs.[15] Ses résultats montrent que le groupe linguistiquement plus

[15] Il est à noter cependant que Geva (1986) a traité la compétence linguistique de ses participant-e-s non pas comme une variable continue, mais comme un indicateur permettant de diviser ce groupe en deux sous-groupes : un groupe intermédiaire et un groupe avancé. Toutefois, pour des

compétent a bénéficié de la présence des connecteurs présentés en en caractère gras, tandis que le groupe moins avancé n'en ont pas bénéficié. La chercheuse explique ce résultat quelque peu surprenant par l'effet de distraction de cette condition pour le groupe moins compétent et suppose que les personnes moins compétentes avaient moins de ressources cognitives libres pour inférer une relation logique implicite (Geva, 1986 : 95). Dans une expérience ultérieure, Geva (1992) a observé de nouveau cette différence en fonction du niveau de compétence linguistique et a montré par ailleurs qu'une bonne compétence linguistique à l'oral prédisait une meilleure compétence avec les connecteurs. De manière générale, la chercheuse résume le lien entre la compétence linguistique et la capacité à utiliser les moyens discursifs:

> « Les apprenant-e-s qui maitrisent mieux la L2, en termes de lexique et de divers aspects de la connaissance syntaxique, sont mieux à même de traiter et d'intégrer des informations à des niveaux plus globaux dans les tâches de lecture. » (Geva, 1992 : 743, ma traduction)

Ainsi, non seulement la compréhension générale d'un texte est facilitée par une meilleure compétence linguistique, on peut également supposer qu'une compétence linguistique plus élevée se traduit par des meilleures connaissances explicites et implicites (et donc une meilleure *maitrise*) des connecteurs. Ainsi, le rôle de la compétence linguistique influencerait également la rédaction des textes: plusieurs études ont montré que les personnes linguistiquement très compétentes utilisent un plus grand nombre de connecteurs ainsi que plus de connecteurs différents que les locuteur-trice-s moins compétent-e-s (p.ex., Kiany & Nejad, 2001; Cho & Shin 2014).

En ce qui concerne le traitement mental, on sait encore peu de choses sur le niveau auquel une capacité de traitement des connecteurs semblable à celle des personnes natives est atteinte. En utilisant des appareils d'oculométrie, Recio Fernández (2020) a examiné le traitement mental de relations de cohérence marquées correctement ou incorrectement avec des connecteurs en espagnol par des personnes non-natives de différents niveaux linguistiques. Ses résultats indiquent que les apprenant-e-s au niveau C1 ont atteint dans une large mesure une automaticité de traitement semblable à celle du groupe natif alors que le groupe

---

raisons théoriques et empiriques, il est fortement déconseillé de traiter la compétence linguistique autrement que comme une variable continue (Leal, 2018).

non-natif moins compétent semblait traiter de manière moins différenciée les phrases qui contenaient une contradiction entre la relation de cohérence et l'indication du connecteur. Recio Fernández est partie de l'idée que les personnes moins compétentes utiliseraient des stratégies de traitement différentes afin de résoudre ces contradictions. Tandis que ces résultats indiquent que les personnes d'un niveau linguistique élevé peuvent atteindre un niveau de traitement qui ressemble à celui du groupe natif, Zufferey & Gygax (2017) ont constaté que les personnes non-natives linguistiquement très compétentes ne remarquaient pas durant leur traitement mental l'absence du connecteur français *en effet* dans une phrase qui devrait être marquée par ce connecteur. Ainsi, il reste encore à explorer à quel niveau linguistique le traitement des connecteurs devient comparable au traitement mental d'une personne native.

Nous constatons que l'enseignement – bien qu'il soit un facteur très important – n'est pas le seul factor qui explique la maitrise insuffisante des connecteurs dans une langue seconde et que la variation interpersonnelle a manifestement une grande influence sur l'acquisition d'une langue seconde également. Il semble par exemple que les personnes ayant une meilleure connaissance d'une L2 comprennent mieux les connecteurs, les utilisent de manière plus appropriée et sont plus sensibles aux faux emplois. Pourtant, il reste des points à éclaircir. Si les participant-e-s de Recio Fernández (2020) traitaient des connecteurs espagnols tels que *por lo tanto* (« donc ») et *sin embargo* (« cependant ») à un niveau proche du groupe natif, les participant-e-s dans Zufferey & Gygax (2017), également d'un niveau linguistique très avancé, ne semblaient maitriser aucune des fonctions du connecteur français *en effet*. Comment expliquer le fait qu'une meilleure compétence linguistique prédisait une meilleure maitrise des connecteurs dans Recio Fernández (2020) tandis que les participant-e-s très compétent-e-s de Zufferey & Gygax (2020, voir aussi Lei, 2012; Carrió-Pastor, 2013; Gao, 2016; Uçar & Yükselir, 2017) ont eu tant de difficultés avec les connecteurs? Une explication pour les observations unanimes est que les connecteurs, quant à eux, varient aussi dans leur complexité, à savoir que les connecteurs *por lo tanto* (« donc ») et *sin embargo* (« cependant ») seraient plus faciles à maitriser dans une langue seconde que *en effet*.

3.6 L'hétérogénéité des connecteurs

Il est vrai que la recherche linguistique a souvent traité les connecteurs comme une catégorie plutôt homogène, ou au moins, a négligé les différences entre les connecteurs. Ainsi, la recherche expérimentale a souvent examiné un groupe relativement restreint de connecteurs comprenant principalement des connecteurs fréquents, comme *and* (« et ») ou *but* (« mais ») (p.ex., Cain & Nash, 2011; Canestrelli et al., 2013; Crible and Pickering, 2020). Comme nous allons le démontrer maintenant, il y a toutefois des fortes indications qu'il existe des différences importantes entre les connecteurs et que ces différences influencent la manière dont les personnes lisent un discours.

Le constat selon lequel la catégorie des connecteurs est hétérogène peut être justifié déjà au niveau théorique. Dans les travaux précédents, plusieurs chercheur-es-s ont essayé de rassembler tous les facteurs qui rendent un trait grammatical compliqué pour les apprenant-e-s d'une langue seconde. Par exemple, Ellis (2006) a établi à cette fin cinq dimensions générales, susceptibles d'expliquer pourquoi certains éléments grammaticaux sont plus complexes à acquérir que d'autres.

La première dimension est celle de la *fréquence*, donc l'idée qu'un trait grammatical est plus difficile si on ne le rencontre que très rarement. Deuxièmement, la *saillance* d'un trait grammatical, c'est-à-dire si un élément est facile à remarquer dans un texte ou énoncé. Si un élément est facile à remarquer, il est plus facilement acquis. La troisième dimension évoquée par Ellis (2006) est la *valeur fonctionnelle*, ou en d'autres termes, si un élément grammatical remplit une fonction claire et distincte ou s'il est polyfonctionnel. Des éléments polyfonctionnels sont considérés comme plus difficiles à acquérir car ils nécessitent d'apprendre chaque fonction et être capable de faire la distinction entre les fonctions. Quatrièmement, la *régularité*, donc si un élément peut être expliqué selon des règles stables et bien définies qui faciliteraient l'acquisition, contrairement à un élément qui ne suit pas de règles fixes. Finalement, la *facilité de traitement*, c'est-à-dire si l'élément ou trait grammatical est facile à traiter. Si c'est le cas, ce trait est tendanciellement plus facile à apprendre qu'un trait difficile à traiter.

Si on applique cette conception aux connecteurs, on remarque que les connecteurs diffèrent considérablement selon ces cinq dimensions. Il y a des

connecteurs fréquents, tel que *en effet* tandis que d'autres sont beaucoup plus rares tels que *en outre* (et il a été montré même que cette différence de fréquence influence la manière dont les adolescent-e-s natifs acquièrent ces connecteurs, p.ex., Tskhovrebova et al., 2021). De même, comme les connecteurs proviennent des catégories linguistiques distinctes, on pourrait argumenter que certains connecteurs sont plus *saillants* (tels que *néanmoins*) tandis que d'autres, notamment ceux qui ne sont pas encore hautement grammaticalisés, restent plus inaperçus (comme *tout en*). Les différences entre les connecteurs s'observent également concernant la dimension de la *polyfonctionnalité*. Certains connecteurs (p.ex., comme *or,* d'après LexConn, Roze et al., 2012) remplissent plusieurs fonctions alors que d'autres sont monosémiques (p.ex., *tandis que*) et n'indiquent qu'une seule relation de cohérence. Finalement, les connecteurs diffèrent également quant à la dimension de la *régularité* : même si morphologiquement stable, la compréhension et désambiguïsation des connecteurs dépend de plusieurs choses, qui sont toutefois différentes pour chaque connecteur. Par exemple, si le connecteur polyfonctionnel *aussi* peut être désambiguïsé en fonction du positionnement syntaxique (LexConn, Roze et al., 2012), cela n'est pas le cas pour *or*. De même, *en effet* peut être désambiguïsé s'il est introduit avec un *et* (dans ce cas le connecteur indique une confirmation, Zufferey & Gygax, 2017) tandis que en ajoutant un *et* avant *autrement* n'indiquerait rien de plus sur la fonction du connecteur.

Nous constatons donc qu'il existe sur le plan théorique des dimensions de Ellis (2006) des différences considérables entre les connecteurs. En plus, il n'est pas difficile de concevoir d'autres dimensions encore. Considérons, par exemple, les deux connecteurs français *du coup* et *en effet* qui ne diffèrent pas seulement quant aux dimensions de Ellis (2006), mais également en termes de leurs degrés de grammaticalisation, le type de causalité qu'ils indiquent, leurs 'charges' pragmatiques (voir chapitre 2.5.4), leur modalité, leur registre ou encore leurs positionnements syntaxiques possibles. Ainsi, on doit conclure que d'un point de vue théorique, il est relativement facile de constituer les connecteurs comme une catégorie très *hétérogène*.

Existe-t-il également des preuves que cette hétérogénéité des connecteurs influence effectivement la manière dont les gens lisent, comprennent et produisent les discours ? En ce qui concerne les personnes natives, de nombreuses études

ont effectivement montré que les connecteurs influencent le traitement mental des phrases de différentes manières (p.ex., Goldman & Murray, 1992; Nippold et al., 1992; Millis, Graesser & Haberlandt, 1993; Cain et al., 2005; Pretorious, 2006; Crosson et al., 2013). De manière très générale, on a constaté que les connecteurs indiquant une relation *discontinue*, telle qu'une concession (et qui est donc plus complexe que la relation causale qui est *continue*) a une influence négative sur le traitement mental (Murray, 1994, 1997, voir aussi chapitre 2.5.3). Ainsi, nous constatons que les connecteurs diffèrent également selon la cinquième dimension de Ellis (2006), celle de la *facilité du traitement mental*. A cet égard, il est important, comme discuté auparavant, de ne pas équivaloir la complexité de la relation de cohérence automatiquement avec la complexité d'un connecteur. On ne sait pas, par exemple, si tous les connecteurs au sein d'une relation de cohérence sont maitrisés de la même manière par des apprenant-e-s. Ainsi, la mise en équivalence entre la complexité de la relation et la complexité du connecteur (bien qu'il y ait bien sûr un lien) doit être nuancée : il faudrait plutôt concevoir la relation de cohérence comme un facteur parmi d'autres dont on peut supposer qu'il influence la maitrise d'un connecteur en L2.

Les connecteurs au sein d'une même relation de cohérence peuvent également varier. Pour rappel, Schumann et al. (2020) ont testé l'impact de quatre connecteurs français qui indiquent tous la même relation de cohérence sur la détection des arguments fallacieux. Ils ont pu observer qu'un argument de l'homme de paille (i.e., reprise et détournement d'une idée proposée par l'interlocuteur) a été mieux détecté lorsqu'il a été introduit avec *puisque* en comparaison avec d'autres connecteurs comme *comme* ou *étant donné que*. De même, Crible et al. (2021) ont également trouvé que le connecteur *en revanche* a suscité un traitement mental temporairement différent du connecteur *par contre*, et cela à la fois pour les personnes natives et non-non-natives. L'hétérogénéité des connecteurs se manifeste également quant à l'acquisition des connecteurs dans la L1. Tskhovrebova, Zufferey & Gygax (2021) ont constaté, par exemple, que des adolescent-e-s maîtrisaient mieux les connecteurs qui étaient plus fréquents que des connecteurs plus rares.

Finalement, Yung et al. (2021) ont rapporté que les connecteurs ne sont pas interchangeables dans tous les cas et que certains connecteurs sont préférés dans certains contextes. Dans leur étude, les chercheur-se-s ont demandé aux

participant-e-s de combler les trous de phrases données avec des connecteurs. Ce faisant, ils ont constaté que si certains connecteurs semblaient être totalement interchangeables (comme *for example* au lieu *for instance* 'par exemple'), les participant-e-s préféraient dans d'autres cas un connecteur à un autre (*however* vs. *still* 'cependant'). Ils ont conclu que cette préférence pouvait s'expliquer par le fait que les connecteurs n'indiquent pas tout à fait la même relation de cohérence. Alors que ces différences subtiles ne sont pas prises en compte par la classification des relations, elles amènent les participant-e-s à préférer un certain connecteur à un autre

En conclusion, nous constatons qu'il existe de très fortes indications selon lesquelles la catégorie des connecteurs peut effectivement être considérée comme étant fortement hétérogène. Les conceptions théoriques (p.ex., Ellis, 2005) ainsi que des observations expérimentales montrant l'impact aux personnes natives (p.ex., Schumann et al., 2020) font preuve des différences énormes parmi les connecteurs. Toutefois, on manque encore des données, par exemple comment les différents types de connecteurs influencent le traitement mental d'un texte. Comme nous avons vu que les connecteurs ne sont souvent pas bien maitrisés dans une deuxième langue, la question se pose inévitablement de savoir si l'hétérogénéité des connecteurs ne se répercute pas également sur la manière comment ils sont acquis et utilisés en L2. Comme nous allons le voir maintenant, cette question n'a pas encore trouvé de réponse satisfaisante.

3.7 Les questions non-résolues

Dans les chapitres précédents, il a été montré que les apprenant-e-s d'une langue seconde éprouvent d'énormes difficultés à utiliser les connecteurs de discours. Partant du constat que les connecteurs constituent une catégorie linguistique extrêmement variée, la question se pose de savoir si et comment l'hétérogénéité des connecteurs se répercute sur leur maitrise dans la langue seconde. En d'autres termes, certains connecteurs sont-ils plus difficiles à maitriser que d'autres ?

Bien que la recherche en linguistique expérimentale ait de plus en plus commencé à s'intéresser à l'impact des profils uniques des connecteurs sur la lecture d'un texte en L1 (et cela n'a été que récemment, voir p.ex., Zufferey, 2012; Santana et al., 2021), l'impact de cette hétérogénéité pour l'acquisition d'une

deuxième langue reste encore entièrement à découvrir. Parmi les très rares études expérimentales qui ont pu documenter le fait que la maitrise des connecteurs dépend effectivement du type de connecteur en question, Crosson et al. (2008) ont constaté que les connecteurs fréquents indiquant des relations simples (comme une *addition*), ont été mieux maitrisés que les connecteurs rares de la même relation. Cependant, les participant-e-s de cette étude étaient des enfants des minorités linguistiques d'une langue seconde, ce qui met en question la généralisation de ces observations pour des apprenant-e-s qui acquièrent une langue après leur L1. Bien que cette étude indique que les connecteurs diffèrent les uns des autres en termes de complexité, de nombreux points restent encore à éclaircir. Par exemple, on ne sait pas si les personnes non-natives maitrisent (tout comme les personnes natives, Zufferey & Gygax, 2020a) les connecteurs plus fréquents mieux que les connecteurs rares (nous allons aborder cette question dans le chapitre 5).

Il n'est pas clair non plus si les connecteurs qui indiquent des relations de cohérence plus complexes sont moins bien maitrisés que les connecteurs qui indiquent des relations simples et si cela dépend aussi du type de connaissance auxquelles les apprenant-e-s accèdent (nous allons aborder ces questions dans les chapitres 5 et 6). De même, on peut s'interroger sur l'impact des caractéristiques syntaxiques et pragmatiques sur la maitrise des connecteurs en L2. Est-ce que les contraintes syntaxiques d'un connecteur, par exemple, sont plus saillantes pour une personne non-native que les contraintes pragmatiques ? Et est-ce que les connecteurs qui sont hautement chargés de fonctions pragmatiques sont plus difficiles à maitriser dans langue seconde (nous allons aborder ces points dans les chapitres 5 et 9) ? Finalement, la question se pose aussi de savoir si, de manière plus générale, les personnes non-natives recourent autant aux connecteurs que les personnes natives lors de la lecture et si ceci est influencé par le type de la relation de cohérence (nous discuterons ces questions dans les chapitres 7 et 8).

Jusqu'à présent, la littérature est malheureusement peu abondante sur ces questions. Ceci est étonnant, car les études qui démontrent des difficultés chez les apprenant-e-s ont insisté déjà très tôt sur l'hétérogénéité des connecteurs et se sont opposées à opposées à l'utilisation de listes qui présentent des connecteurs de manière homogène et interchangeable dans l'enseignement (p.ex., Zamel, 1983; Granger & Tyson, 1996; Anderson, 2014). Bien que les analyses des

productions textuelles en L2 donnent un bon aperçu des difficultés des apprenant-e-s à utiliser les connecteurs en documentant un statut quo, ces observations ne permettent guère d'expliquer avec certitude *pourquoi* les apprenant-e-s éprouvent des difficultés à maitriser les connecteurs. A cet égard, il est impératif d'opter pour une approche expérimentale afin de pouvoir identifier, isoler et manipuler les facteurs en question.

Afin de combler le vide dans la littérature, la présente thèse a pour but d'examiner la maitrise des connecteurs en L2 de manière expérimentale. A cette fin, deux questions de recherche sont formulées :

I. L'hétérogénéité des connecteurs se répercute-t-elle également sur leur maitrise en L2 ?
II. Quels sont les facteurs linguistiques et cognitifs qui rendent certains connecteurs plus difficiles que d'autres ?

Si la maitrise des connecteurs en L2 dépendait effectivement du type de connecteur, cela donnerait une explication supplémentaire aux observations contradictoires (ou du moins, ambiguës) concernant le bénéfice qu'apportent (ou non) les connecteurs en L2. Pour rappel, nous avons vu que les connecteurs peuvent faciliter la compréhension d'un texte, à condition que les personnes aient une compétence linguistique élevée (Degand & Sanders, 2002), ce qui explique pourquoi un effet de facilitation des connecteurs n'a pas toujours été observé (Geva, 1986). Toutefois, il est également possible que ces résultats s'expliquent par le fait que les connecteurs testés dans ces études n'étaient pas les mêmes. Si les connecteurs diffèrent quant à leur complexité, l'impact de la compétence linguistique doit être évalué en interaction avec le connecteur en question. Par exemple, un connecteur simple et facile peut être maitrisé très tôt même par les apprenant-e-s moins compétentes, alors que la maitrise d'un connecteur plus complexe est seulement atteinte par des personnes plus compétentes. En outre, certaines études de corpus ont trouvé que les personnes non-natives utilisent beaucoup moins des connecteurs que les personnes natives (p.ex., Ha, 2016; Shi, 2017) tandis que d'autres ont trouvé une tendance inverse pour d'autres connecteurs (p.ex., Granger & Tyson, 1996; Leedham & Cai, 2013; Chen, 2014). Cela pourrait à nouveau être lié à l'hétérogénéité des connecteurs : un connecteur simple pourrait devenir plus facilement un mot de confort avec lequel les apprenant-e-s se sentent à l'aise, tandis qu'un connecteur plus complexe serait plutôt évité.

Si l'hétérogénéité des connecteurs influence leur acquisition et leur maitrise en L2, il devient légitime de questionner l'idée selon laquelle les apprenant-e-s ont des difficultés avec la catégorie des connecteurs en général (ou, en tout cas, pas plus que les personnes natives). La maitrise insuffisante des connecteurs par les personnes non-natives ne serait donc pas due à leurs fonctions procédurales, mais plutôt à leurs caractéristiques individuelles et le fait que les fonctions des connecteurs de la L1 ne correspondent pas à celles des connecteurs de la L2. Il est vrai aussi que les connecteurs existent souvent également dans la première langue des personnes non-natives. Ainsi, le simple fait que les connecteurs structurent un discours en indiquant des relations de cohérence ne devrait pas être un facteur de difficulté en soi, puisque ce fonctionnement est connu en L1. Toutefois, comme la littérature n'a pas encore investigué l'influence de différentes caractéristiques des connecteurs sur leur maitrise en langue seconde, cette idée reste pour le moment spéculative.

Finalement, une investigation expérimentale sur la maitrise des connecteurs permet aussi d'investiguer en parallèle d'autres variables qui influencent la manière dont les personnes non-natives maitrisent les connecteurs. Comme nous venons de voir, il semble que certaines personnes arrivent mieux à maitriser les connecteurs que d'autres. Ainsi, il convient d'examiner plus en détail quelles compétences individuelles, comme par exemple des connaissances en grammaire, permettent une meilleure maitrise des connecteurs (nous allons aborder cette question dans les chapitres 5, 7, 8 et 9).

De plus, il est aussi probable que la maitrise d'un connecteur dépende aussi de la manière dont une personne lit un discours. Par exemple, il est possible qu'il lui soit plus facile à détecter une mauvaise utilisation d'un connecteur lorsqu'elle accède à ses connaissances explicites (par exemple en évaluant consciemment une phrase) qu'en accédant à ses connaissances implicites. Les données manquent toutefois pour tirer une conclusion certaine à ce sujet (nous allons aborder cette question notamment dans les chapitres 7 et 8).

Ainsi, pour tenir compte de ces deux enjeux, nous complétons nos questions de recherches avec deux questions supplémentaires :

III. Quel est l'impact des compétences individuelles sur la maitrise des connecteurs en L2 ?

IV. Quel est l'impact du type de connaissances sur la maitrise des connecteurs en L2 ?

Cela nous permet non seulement d'examiner les facteurs linguistiques des connecteurs individuels, mais aussi de tirer des conclusions plus générales sur la maitrise des connecteurs dans une langue seconde.

3.8 Conclusion

Dans ce chapitre, nous avons vu que la maitrise des connecteurs d'une langue seconde peut s'articuler au sein de deux types de connaissances : les connaissances explicites et les connaissances implicites (Ellis, 2004, 2005, 2009b). La recherche a trouvé différentes manières de mesurer ces deux types de connaissance et a découvert que, bien que les connecteurs puissent en principe faciliter la compréhension d'un texte en L2 (Degand & Sanders, 2002), leur maitrise dans une langue seconde est très difficilement atteinte (Zufferey & Gygax, 2017). Nous avons également discuté les facteurs qui pourraient prédire une meilleure maitrise des connecteurs, comme par exemple des facteurs interpersonnels (tels qu'une compétence linguistique élevée) ou un enseignement adapté qui prendrait en compte les différents moyens à marquer des relations de cohérence. Nous avons particulièrement souligné le fait que la catégorie des connecteurs est très hétérogène, ce qui fait que certains connecteurs sont plus difficiles à apprendre que d'autres. Toutefois, bien qu'il soit probable que les connecteurs diffèrent en termes de complexité, la littérature n'a pas encore investigué en détail comment cette hétérogénéité pourrait se répercuter sur la maitrise des connecteurs dans une langue seconde.

Pour combler ce vide dans la littérature, plusieurs séries d'expériences sont présentées dans les chapitres suivants. On y examine non seulement si les caractéristiques individuelles des connecteurs influencent la manière dont ils sont utilisés par les personnes non-natives, mais aussi quels facteurs rendent un connecteur plus difficile qu'un autre.

A cet égard, une première piste des facteurs qui pourraient différencier les connecteurs en termes de complexité sont leur fréquence ainsi que la relation de cohérence qu'ils servent à communiquer.

4 L’impact du type du connecteur sur le traitement mental en L1

Dans les chapitres précédents, nous avons mis en évidence que les connecteurs appartiennent à une catégorie d'éléments linguistiques qui, malgré son importance pour la structuration d'un discours, semble difficile à apprendre et à maitriser en L2. Toutefois, la difficulté éprouvée des apprenant-e-s ne peut pas nécessairement être généralisée pour tous les connecteurs, car il s’agit d’une catégorie extrêmement hétérogène. On peut même supposer que la catégorie des connecteurs n'est pas difficile en soi pour un-e apprenant-e, mais que ce sont plutôt les propriétés individuelles d’un connecteur qui engendrent sa difficulté. Cela soulève la question de savoir de quels facteurs il s’agit.

Avant d'explorer ce point, il est important d’examiner si la maitrise des connecteurs varie également pour les personnes natives : ce n'est qu'en sachant si et comment les personnes natives sont affectées par l'hétérogénéité des connecteurs que nous pouvons faire des comparaisons avec les personnes non-natives. De plus, si nous constatons que la lecture des personnes natives est influencée par certaines caractéristiques d’un connecteur, cela nous indiquerait que les apprenant-e-s devraient être également affectés.

Est-il possible que les connecteurs soient également maitrisés de manière différente en L1 ? Bien qu'il soit vrai que les personnes natives sont souvent perçues comme des experts de leur L1, la recherche a constaté de fortes différences dans leurs compétences linguistiques (Kidd et al., 2018). Si nous considérons les personnes natives comme un groupe hétérogène dans lequel il y a des personnes avec de meilleures et de moins bonnes compétences linguistiques, cela signifierait qu'il devrait y avoir également des connecteurs plus difficiles et plus faciles pour les personnes natives. Et en effet, on trouve des indications dans la littérature que certains connecteurs sont mieux maitrisés que d’autres par les personnes natives. Par exemple, il a été montré que les adolescent-e-s francophones maitrisaient moins bien les connecteurs rares dans une tâche d’insertion que les connecteurs fréquents (Zufferey & Gygax, 2020b).

Nous ne savons toutefois pas dans quelle mesure cet effet se manifeste également durant la lecture, c'est-à-dire si un connecteur rare complique le traitement mental d'un discours. Par exemple, il se peut que l'incohérence soit détectée plus tardivement dans une phrase qui contient un connecteur rare que dans une phrase contenant un connecteur fréquent.

Pour tester cette hypothèse et combler le vide de la littérature, nous menons trois expériences dans lesquelles nous examinons le traitement mental des personnes natives pour des phrases cohérentes et incohérentes qui contiennent soit une concession, soit une conséquence. Comme nous nous intéressons à la question de savoir si et quand l'incohérence est détectée et si le type de connecteur utilisé peut faciliter ou freiner la détection, nous changeons dans chaque expérience les connecteurs qui indiquent ces relations. D'abord, on teste des connecteurs fréquents (Expérience 1), ensuite des connecteurs rares et monofonctionnels (Expérience 2) et finalement des connecteurs rares et polyfonctionnels (Expérience 3). Nous supposons que les connecteurs rares devraient réduire la sensibilité à l'incohérence car les personnes interrogées n'y sont pas habituées et n'ont pas forcément intériorisé les instructions procédurales au même degré que pour un connecteur fréquent. La polyfonctionnalité d'un connecteur devrait également retarder la détection de l'incohérence, car le connecteur doit d'abord être désambiguïsé pour comprendre la phrase.

En testant ces hypothèses, nous clarifions l'impact des connecteurs différents sur le traitement mental des personnes natives et fournissons simultanément des premières indications pour les facteurs qui pourraient rendre un connecteur plus difficile pour un-e apprenant-e.

4.1 Quelles propriétés des connecteurs peuvent influencer la lecture d'un discours en L1 ?

Nous avons discuté dans le chapitre 3.6 que la catégorie des connecteurs ne devrait pas être constituée comme une catégorie homogène, car elle comprend une grande variété d'éléments divers qui sont utilisés pour indiquer des relations de cohérence. Dans ce qui suit, nous allons examiner les différentes propriétés des connecteurs qui pourraient engendrer une lecture différente pour les personnes natives. Nous nous concentrons sur deux facteurs pour lesquels nous faisons

l'hypothèse qu'ils influencent la manière dont on traite l'incohérence dans sa L1: la *fréquence* du connecteur, ainsi que sa *polyfonctionnalité*. Nous allons maintenant expliquer pourquoi nous pensons que ces facteurs peuvent influencer la lecture des personnes natives.

4.1.1 La fréquence du connecteur

De manière générale, il est fortement supposé que la fréquence d'un élément linguistique ait un impact sur la manière dont on l'acquière, à la fois en L2 (Ellis, 2006) et en L1 (Kidd et al., 2018). Pour les personnes natives, Kidd et al. (2018) indiquent que :

> « L'effet environnemental le plus immédiat sur le langage concerne la fréquence d'utilisation de la langue, qui affecte à la fois l'acquisition et le traitement du langage à l'âge adulte (…). Toutes choses étant égales par ailleurs, les mots et les structures plus fréquents sont généralement acquis plus tôt et sont traités plus facilement » (Kidd et al., 2018 : 160, ma traduction).

Quant aux connecteurs du discours, cela signifierait-il que les connecteurs plus fréquents sont mieux compris ou utilisés par les personnes natives? Dans la littérature, on trouve en effet des indications pour un tel effet de fréquence, et cela pour différentes tranches d'âge. Par exemple, Crosson, Lesaux et Martiniello (2008) ont démontré que les enfants de minorités linguistiques comprenaient mieux les relations plus simples (par exemple, les relations additives) lorsque les connecteurs étaient plus fréquents. En effet, les participant-e-s ont obtenu des meilleurs résultats dans une tâche de complétion de phrases lorsque les phrases contenaient des connecteurs fréquents. De même, Zufferey & Gygax (2020b) ont trouvé que des adolescent-e-s francophones maitrisaient dans une tâche d'insertion les connecteurs moins fréquents dans la langue orale moins bien que les connecteurs plus fréquents.

Il semble même que la fréquence d'un connecteur influence sa maitrise à l'âge adulte. Zufferey et al. (2015), par exemple, ont observé que les personnes natives détectaient mieux des faux emplois des connecteurs dans les relations conditionnelles que dans des relations contrastives. Cette observation peut être expliqué avec un effet de fréquence : le connecteur testé pour la relation conditionnelle est très fréquemment utilisé pour indiquer cette relation. En revanche, les connecteurs testés pour la relation contrastive indiquent plus rarement cette relation.

En conclusion, plusieurs études indiquent que les connecteurs moins fréquents sont moins bien maitrisés par les personnes natives alors que les connecteurs plus fréquents semblent en quelque sorte plus accessibles. Comme cet effet n'a été documenté que pour la connaissance explicite, on peut se demander si la fréquence d'un connecteur a également un effet sur la connaissance implicite. On peut supposer, par exemple, que la détection de l'incohérence est facilitée lorsque des connecteurs fréquents sont utilisés, et freinée lorsque des connecteurs rares sont présents. La sensibilité moindre aux connecteurs rares pourrait s'expliquer par deux facteurs : l'activation de la signification des connecteurs rares prend un peu plus de temps et, par rapport aux connecteurs fréquents, on a peut-être un peu moins confiance d'avoir bien (ou mal) compris la phrase. En conséquence, la détection d'incohérence devrait être plus difficile. Malheureusement, on manque pour l'instant de données en faveur de cette hypothèse.

Alors que la fréquence d'un connecteur est un facteur qui influence probablement la lecture d'un texte, les connecteurs varient également selon d'autres propriétés. L'une d'entre elles est la polyfonctionnalité, c'est-à-dire le fait qu'un connecteur puisse indiquer plus qu'une seule relation de cohérence selon le contexte. Comme nous allons discuter maintenant, nous supposons que cette caractéristique peut également affecter la lecture en L1.

4.1.2 La polyfonctionnalité du connecteur

En entend par la polyfonctionnalité d'un connecteur sa capacité d'indiquer plus qu'une relation de cohérence. Selon le dictionnaire des connecteurs *LexConn* (Roze et al., 2012), le connecteur français *si*, par exemple, peut être utilisé pour indiquer une condition (1), ou un contraste (2).

1) S'il pleut, nous n'allons pas faire la randonnée.
2) Si certaines personnes aiment la pluie, d'autres la détestent.

Zawel (1983) indique que dans certains cas, la polyfonctionnalité d'un connecteur peut rendre l'interprétation de la phrase plus difficile. Elle donne l'exemple du connecteur anglais *since* ('depuis que' / 'comme'), qui peut indiquer soit une relation causale, soit une relation temporelle. Dans certains cas, la distinction entre ces

deux fonctions n'est pas très claire sans contexte. La chercheuse donne l'exemple de la phrase suivante (Zamel, 1983 : 24).

3) *Since you went away, the days went cold*
‘CONNECTEUR tu es parti-e, les jours sont devenus froids.

Dans cette phrase, la relation du connecteur peut varier selon l'interprétation subjective, la seule façon d'établir une relation temporelle ou causale serait de prendre en compte le contexte et les intentions de la personne locutrice. Compte tenu de cette ambiguïté du connecteur qui doit être résolue pour comprendre la phrase, en résulte-t-il que la polyfonctionnalité d’un connecteur influence la façon dont on lit un discours ?

On trouve des indications en faveur de cette hypothèse dans la littérature, notamment pour les connecteurs sous-spécifiés français *et* ainsi que pour son équivalent anglais *and*. Le phénomène de la sous-spécificité d’un connecteur est similaire à la polyfonctionnalité et on peut supposer que ces deux caractéristiques ont une influence similaire sur la lecture. La raison en est que le processus de désambiguïsation nécessite un certain temps et la solution doit être revérifiée et réévaluée tout au long de la lecture des phrases. Ainsi, de nombreuses études ont montré qu’un connecteur sous-spécifié (tel que le connecteur *et* ou son équivalent anglais *and*) peut rendre l’insertion de la bonne relation de cohérence plus difficile (Koornneef & Sanders, 2013; Crible & Pickering, 2020; Crible et al., 2021). De même, l’étude de Cain et Nash (2011) a montré que le processus de désambiguïsation d’un connecteur influence la lecture des enfants à l’âge de 8-10 ans. Dans leur expérience, les chercheuses ont constaté que lorsque les phrases temporelles, causales ou adversatives étaient marquées avec le connecteur sous-spécifié *and* (« et »), leurs participant-e-s ralentissaient leur lecture par rapport aux phrases qui étaient marquées avec des connecteurs plus spécifiques dans leurs instructions procédurales, tels que *because* (« parce que ») ou *but* (« mais »).

Comme ces travaux montrent que la sous-spécificité d’un connecteur peut être considéré comme un facteur qui affecte la fluidité de la lecture, il est légitime de supposer que la polyfonctionnalité d’un connecteur influence également la manière dont on détecte, traite et essaie de résoudre l’incohérence d’une phrase. Il convient de noter à cet égard qu’un connecteur polyfonctionnel ne devrait pas rendre la lecture plus *difficile* pour une personne native (car on peut supposer

qu'elle maitrise les fonctions différentes). Plutôt, un tel connecteur pourrait nécessiter des capacités mentales plus importantes pour créer une représentation mentale de la phrase et détecter ainsi une incohérence potentielle.

Enfin, si on veut tester l'impact de la polyfonctionnalité ainsi que celui de la fréquence d'un connecteur sur la lecture, il est impératif de tenir compte des différences entre les relations de cohérence. Il est connu que le traitement mental varie en fonction de la complexité de la relation de cohérence, ce qui pourrait peut-être avoir pour effet que les facteurs de la polyfonctionnalité et la fréquence ont un impact différent selon la relation de cohérence.

4.1.3 Y-a-t-il un impact différent de la fréquence et de la polyfonctionnalité d'un connecteur en fonction de la relation de cohérence ?

Nous avons vu au chapitre 2.2.2 que les relations de cohérence varient en termes de complexité cognitive (Sanders et al., 1992). En conséquence, les relations de cohérence différentes provoquent des lectures différentes.

Haberlandt et Bingham (1978) ont constaté, par exemple, qu'on lit deux phrases plus vite lorsqu'elles ont un lien causal. Cet effet de facilitation de la lecture pour les relations causales est même scalaire : dans une étude plus nuancée, Keenen, Baillet et Brown (1984) ont trouvé que le degré d'évidence de la causalité affecte le temps de lecture. Dans leur expérience, les participant-e-s lisaient des phrases plus rapidement lorsque le lien de causalité était plus évident.

Nous avons déjà discuté dans le chapitre 2.5.3 les différentes théories qui proposent des explications pour ces observations. *L'hypothèse de continuité*, par exemple, est fondée sur l'idée que les personnes lectrices font par défaut l'hypothèse qu'un texte est logiquement et chronologiquement continu (Segal et al., 1991). Par conséquent, les connecteurs qui indiquent un changement de cette hypothèse ont un impact plus important sur la lecture (Murray, 1997). *L'hypothèse de causalité-par-défaut* propose qu'une personne lectrice s'attend, lorsqu'aucun autre indice n'est disponible, à une relation causale (Sanders, 2005). Cette hypothèse a trouvé confirmation dans l'étude de Sanders et Noordman (2000) qui démontre que les personnes lectrices traitent les liens causaux plus rapidement que des liens additifs. Cela pourrait surprendre vu qu'une addition (en particulier si celle-ci est sémantique, van Dijk, 1979 : 450) peut être considérée comme étant plus simple qu'une cause (Halliday & Hasan, 1976 : 233; Sanders & Nordmann,

2000 : 39). Pourtant, les chercheur-se-s suggèrent que « la préférence pour la détection d'une explication causale » (Sanders & Noordman, 2000 : 53, ma traduction) découle de l'attente sous-jacente d'une structure causale dans le texte. Cela a pour conséquence qu'une structure causale est traitée avec plus d'aisance qu'une relation additive, même si celle-ci est cognitivement plus simple.

Finalement, Lyu et al. (2020) ont constaté que leurs participant-e-s traitaient l'incohérence dans une phrase concessive différemment que dans une phrase causale. Plus précisément, les chercheur-se-s ont testé la lecture des phrases contenant différents types de plausibilité (non-plausible, moins plausible et hautement plausible) de relations concessives et causales. Ils ont constaté que les phrases causales non-plausibles perturbaient le traitement mental plus longuement que les phrases concessives. Selon Lyu et al. (2020), les relations concessives impliquent le point de vue subjectif dans une plus large mesure alors que les relations causales peuvent être plus facilement alignées sur les connaissances du monde de la personne lectrice :

> « Dans une relation causale, les personnes lectrices s'appuient davantage sur leur propre connaissance du monde et montrent un effet soutenu lorsque leur attente n'est confirmée par la suite du discours. Au cours du traitement de la concession, un changement de perspective est impliqué de sorte que les participant-e-s prennent en compte le point de vue de la personne locutrice et tolèrent l'incohérence, ce qui nécessite un effort cognitif plus important. » (Lyu et al. 2020 : 339, ma traduction)

L'observation qu'une relation causale est plus simple à traiter a été trouvée déjà en 1978. Les chercheurs Haberlandt et Bingham (1978) ont constaté qu'on lit deux phrases plus vite lorsqu'elles se lient de manière causale (voir aussi Keenan et al., 1984). En outre, il a été également montré que les relations causales facilitent le rappel du contenu que d'autres relations (Trabasso, Secco, & Van Den Broek, 1982). Similairement, Morera, León, Escudero et Vega (2017) ont montré qu'un connecteur causal induit une continuité émotionnelle alors qu'un connecteur concessif ne le fait pas. Dans leur expérience, le connecteur concessif était moins rapidement traité qu'un connecteur causal (voir aussi Köhne & Demberg, 2013; Xu et al., 2018).

En conclusion, on peut dire que les relations de cohérence varient dans leur complexité cognitive et cela influence la manière dont on lit un discours. Si on veut examiner l'impact des connecteurs sur la lecture, respectivement la réaction à l'incohérence, il est indispensable de prendre en compte des relations de

cohérence différentes : il est probable que la réaction à l'incohérence soit différente pour une phrase causale que pour une phrase concessive (p.ex., Lyu et al., 2020). Ainsi, il convient de tester deux connecteurs de deux relations de cohérence différentes.

4.2 La série d'expériences présente

Pour examiner l'influence du type du connecteur sur la lecture en L1, nous allons mener trois expériences[16], au cours desquelles nous mesurerons le temps de lecture pour des phrases qui contiennent des connecteurs différents. Dans la première expérience, nous testerons deux connecteurs qui sont très fréquents à la fois à l'oral et à l'écrit (*donc* et *mais*). Ensuite, dans la deuxième expérience, nous allons tester des connecteurs beaucoup moins fréquents (*néanmoins* et *ainsi*). Finalement, dans la troisième expérience, nous allons tester des connecteurs rares qui sont, en outre, polyfonctionnels (*or* et *aussi*).

Nous pensons que les différences entre les connecteurs testés peuvent influencer la manière dont les personnes natives détectent et traitent l'incohérence. Nous supposons que la sensibilité à l'incohérence diminue lorsque la phrase contient un connecteur moins fréquent et polyfonctionnel. Par contre, il est moins probable que la lecture des phrases cohérentes soit affectée par le type de connecteur, car on peut supposer que les personnes natives arrivent à structurer un discours même à l'aide de connecteurs rares. En mesurant donc non seulement le temps de lecture des phrases cohérentes mais également incohérentes, nous arriverons à détecter des nuances subtiles entre la manière dont on récupère et traite les instructions procédurales de connecteurs différents.

Finalement, nous allons tester la sensibilité à l'incohérence pour des relations de cohérence différentes. A cet égard, deux effets seraient possibles : il est possible que les gens réagissent plus rapidement à l'incohérence pour une phrase concessive, car cette relation nécessite un traitement plus approfondi et attentif. En revanche, il se peut également que la complexité d'une concession fasse que les gens ne perçoivent l'incohérence qu'à un stade ultérieur de la phrase,

[16] Les résultats de ces expériences ont été publiés sous une autre forme dans l'article « How robust is discourse processing for native readers? The role of connectives and the coherence relations they convey » dans le journal *Frontiers in Psychology* (Wetzel, Zufferey & Gygax, 2022a).

car ils doivent d'abord traiter mentalement la phrase (ce qui nécessite un temps plus long).

4.3 Expérience 1 : la réaction à l'incohérence pour des phrases contenant les connecteurs *donc* et *mais*

Dans la première expérience, nous examinons deux connecteurs qui sont extrêmement fréquents, à la fois dans la langue écrite et orale : *donc* pour indiquer une conséquence et *mais* pour indiquer une concession.

Pour justifier la classification de ces connecteurs comme fréquents, nous avons comparé leur fréquence avec les deux connecteurs testés dans la deuxième expérience, *néanmoins* et *ainsi*. Les fréquences d'occurrences dans les corpus oraux et écrits dans la collection de corpus ORFEO (> 10 millions de mots, Benzitoun, Debaisieux & Deulofeu, 2016) de ces connecteurs sont reportés dans le Tableau 4.1.

Tableau 4.1. Les nombres d'occurrences de *mais*, *donc*, *néanmoins* et *ainsi* dans ORFEO

Expérience	Connecteur	*À l'oral*		*A l'écrit*	
		Nombre d'occurrences	%	Nombre d'occurrences	%
1	*mais*	882	.0022	1143	.0191
1	*donc*	882	.0221	676	.0113
2	*néanmoins*	22	.0006	121	.0020
2	*ainsi*	138	.0035	580	.0097

Pour déterminer si *mais* et *donc* (Expérience 1) sont plus fréquents que les connecteurs *néanmoins* et *ainsi* (Expérience 2) nous avons effectué un test de χ^2 des occurrences totales des connecteurs en utilisant la fonction *chisq.test()* du *stats-package* (v.3.6.2.) dans *R* (R Core Team, 2020). Les résultats de ce test démontrent que les connecteurs *mais* et *donc* sont effectivement plus fréquents que *néanmoins* et *ainsi* ($\chi^2(1) = 76{,}93$, $p = 2.2^{e\text{-}16}$)

A part leur fréquence, nous testons ces deux connecteurs parce qu'ils indiquent des relations de cohérence différentes qui diffèrent considérablement dans leur complexité. En comparaison avec une conséquence, une concession est plus complexe car elle implique une polarité négative (voir Sanders et al., 1992). Cela nous permet d'observer si une phrase incohérente qui est concessive provoque des perturbations du traitement mental différents d'une phrase causale.

Nous supposons qu'en raison de leur fréquence, les fonctions de ces connecteurs devraient être hautement intériorisées par les personnes natives. En conséquence, ces dernières devraient réagir immédiatement et fortement en cas d'incohérence.

4.3.1 Les participant-e-s

Nous avons recruté 122 personnes natives du français via la plateforme *Prolific* (Prolific, London, 2020). Deux personnes ont été exclues de l'analyse des données car elles ont répondu à un taux trop faible aux contrôles d'attention (réponses inférieures de 75%, voir ci-dessous pour la tâche). Au total, les données des 120 personnes ont été analysées (âge moyen = 29.7, écart type = 9.3, 46 femmes). Tous les participant-e-s montraient une participation satisfaisante dans des études précédentes sur la plateforme (minimum de 95% des participations approuvées) et ont été rémunéré-e-s dans notre expérience par £3.15. Tous les participant-e-s ont donné leur consentement de participer à cette expérience.

4.3.2 Le design et la procédure

Pour pouvoir tester la réaction à l'(in)cohérence lors de la lecture des phrases qui sont introduits par des connecteurs différents, nous avons utilisé un design factoriel 2^x2. Les variables étaient le *Type de connecteur* (*donc* où *mais*) et le *Type de phrase* (*cohérente* ou *incohérente*).

Nous avons établi 40 phrases expérimentales contenant toutes une première et une deuxième clause qui étaient liées avec un connecteur. La première clause contenait toujours un sujet animé et la deuxième clause un pronom qui s'y référait. Un exemple de phrase expérimentale est présenté en (4).

4) Jean veut connaitre les nouvelles du jour, donc il lit le journal tous les matins.

Nous avons segmenté chaque phrase expérimentale en sept segments dont le quatrième comportait toujours le connecteur. Cette segmentation est illustrée en (5). Les barres obliques représentent la démarcation entre deux segments.

5) Jean veut $_{(1)}$ // connaitre les nouvelles $_{(2)}$ // du jour $_{(3)}$ // donc $_{(4)}$ // il lit $_{(5)}$ // le journal $_{(6)}$ // tous les matins $_{(7)}$ //

Pour obtenir une phrase incohérente qui ne différait pas dans les segments 4 à 7 de ceux de (5), nous avons fait des modifications dans les segments 1 à 3. Par exemple, pour la phrase (5), nous avons utilisé le verbe *vouloir* avec une polarité négative[17] dans le premier segment, comme illustré en (6).

6) ? Jean ne veut pas (1) // connaitre les nouvelles (2) // du jour (3) // donc (4) // il lit (5) // le journal (6) // tous les matins (7) //

Comme la phrase résultante contient maintenant une relation concessive, l'utilisation du connecteur *donc* rend la phrase incohérente.

Pour obtenir une phrase cohérente contenant *mais*, nous avons donc pu utiliser (6) simplement en échangeant *donc* et *mais*. Inversement, pour les phrases incohérentes contenant *mais*, nous avons échangé le connecteur *donc* contre *mais* dans les phrases cohérentes de conséquence (telle que 4). De cette manière, nous avons obtenu quatre versions d'une phrase, présentées dans le Tableau 4.2.

Tableau 4.2. Exemple d'une phrase expérimentale utilisée dans ses quatre conditions

Connecteur	*Cohérence*	*Exemple d'une phrase*
mais	cohérent	Jean ne veut pas connaitre les nouvelles du jour, mais il lit les nouvelles tous les matins.
mais	incohérent	Jean veut connaitre les nouvelles du jour, mais il lit les nouvelles tous les matins.
donc	cohérent	Jean ne veut pas connaitre les nouvelles du jours, donc il lit les nouvelles tous les matins.
donc	incohérent	Jean ne veut pas connaitre les nouvelles du jours, donc il lit les nouvelles tous les matins.

Comme les modifications ont été effectuées dans les segments 1, 2 ou 3, les segments 5, 6 et 7 restaient toujours les mêmes dans toutes les phrases. Cela nous a permis, en conséquence, de comparer les temps de lecture mesurés pour ces segments.

Pour les phrases incohérentes, il convient de noter que l'incohérence ne peut pas être anticipée dans le cinquième segment. En revanche, le sixième segment

[17] Crible (2021) a trouvé que le traitement d'une concession est facilité lorsque le verbe contient une polarité négative plutôt que positive. Toutefois, un biais dû à la polarité du verbe (quel qu'il soit) dans notre expérience est presque exclu pour plusieurs raisons. Tout d'abord, dans nos phrases expérimentales modifiées se trouvent des verbes de polarité négative et positive. De plus, la polarité du verbe apparaît dans le premier des quatre segments non pertinents (car non analysés). Troisièmement, nous utilisons quatre listes avec un quarré latin qui baisse l'influence des phrases individuelles (voir ci-dessous). Et quatrièmement, nous avons défini les *Phrases* comme effet aléatoire dans tous nos modèles statistiques (voir ci-dessous).

est le premier segment dans lequel l'incohérence devient théoriquement anticipable, même si d'autres aboutissements des phrases sont encore possibles à ce stade. Dans le septième segment, le dernier segment de la phrase, l'incohérence devient évidente et ne peut pas être résolue par d'autres moyens.

En suivant le modèle d'un carré latin, quatre listes ont été établies dans lesquelles chaque phrase expérimentale a été présentée dans les quatre conditions testées (i.e., *donc / cohérent*, *mais / cohérent*, *donc / incohérent*, *mais / incohérent*). Cela nous a permis d'éviter que d'autres caractéristiques de la phrase, qui auraient pu influencer le traitement mental, biaisent nos résultats. Chaque personne interrogée a été attribuée aléatoirement à une seule liste.

A part les phrases expérimentales, nous avons également utilisé 48 phrases de distraction qui contenaient un pronom relatif au lieu d'un connecteur au segment 4. L'utilisation de ces phrases nous a permis de détourner l'attention des participant-e-s des connecteurs. Un exemple d'une phrase de distraction segmentée est présenté dans (7).

7) Matthieu lisait (1) // attentivement (2) // le texte (3) // auquel (4) // son père (5) // se référait (6) // sans cesse. (7)

La moitié des phrases de distraction contenaient de faux emplois d'un pronom relatif, comme présenté en (8).

8) ? Martin lisait (1) // le même (2) // livre (3) // auquel (4) // il avait gardé (5) // un exemplaire (6) // depuis sa jeunesse. (7)

Au total, chaque participant-e lisait donc 88 phrases. Après la lecture de chaque phrase, une question de vérification correspondante a été posée qui se référait à la phrase qui venait d'être lue. La réponse possible à la question était soit affirmative, soit négative. Par exemple, pour la phrase expérimentale (5), la question de vérification est présentée en (9).

9) Jean lit toujours le journal. Vrai ou faux ?

Le premier but de cette question était de vérifier si les personnes avaient bien lu la phrase. En effet, sans avoir lu la phrase (5) on ne saurait pas répondre à la question (9). Tous les participant-e-s qui donnaient moins de 75% de bonnes réponses ont été exclus de l'analyse des données. Le deuxième avantage de ces questions était qu'elles représentent une région postcritique, nous permettant d'observer des effets de débordement (voir p.ex., Lyu et al., 2020). Concrètement, nous supposions par

exemple qu'une personne qui réfléchit sur la phrase pour trouver la réponse à (9) allait avoir un temps de réponse plus lent si la phrase subséquente était incohérente.

Pour que chaque réponse soit associée à une phrase indépendamment de sa cohérence ou de son incohérence, il était indispensable que les questions se réfèrent à la deuxième clause (car nous avons modifié les segments dans la première). Cependant, cela comportait le risque que les participant-e-s ne lisent – par effet d'entrainement – que la deuxième clause pour répondre à la question de vérification. Pour éviter cela, les questions de vérification pour les phrases de distraction se réfèreraient à la première clause. Comme les participant-e-s ne savaient pas s'ils allaient lire une phrase de distraction ou une phrase expérimentale, ils devaient lire toutes les clauses de manière attentive pour pouvoir répondre à la question.

L'expérience a été programmée à l'aide de *Psychopy* (version 2020.20) et hébergée sur les serveurs de *Pavlovia* (*www.pavlovia.org*). Au début de l'expérience, deux items d'entrainement ont été présentés pour familiariser les participant-e-s avec la tâche. Ensuite, les phrases expérimentales segmentées d'une liste ont été présentées, en ordre aléatoire et différent pour chaque participant-e. Avant chaque dernier segment d'une phrase, la phrase « Appuyez sur la BARRE D'ESPACE pour voir apparaître la phrase suivante» a été affichée. Cela a permis d'éviter les effets de débordement potentiels d'une phrase à l'autre. Ensuite, le symbole « + » a été présenté pendant une seconde à l'endroit où le premier segment de phrase allait être présenté. Chaque segment a été présenté de manière isolée et les participant-e-s ont appuyé sur la barre d'espace pour accéder au segment suivant.

Les segments ont été présentés en noir sur blanc dans une police lisible au milieu gauche de l'écran. La souris ne s'affichait pas. Après la lecture des segments d'une phrase, la question de vérification correspondante a été présentée et les participant-e-s devaient juger si la question était correcte en cliquant sur la touche « c », ou incorrecte en cliquant sur la touche « i ».

4.3.3 L'analyse

Pour analyser les temps de lecture, nous avons construit des modèles linéaires à effets mixtes à l'aide du logiciel *R* (R Core team, 2020). Nos facteurs étaient le *Type*

de phrase (cohérente ou incohérente) et le *Type de Connecteur* (concessif ou causal).

Nous avons monté les modèles en suivant les recommandations de Baayen (2008): nous avons ajouté successivement un effet fixe à la fois et avons comparé chaque fois le modèle résultant avec un modèle qui ne contenait pas cet effet fixe. Nous avons comparé les modèles en effectuant les tests du rapport de vraisemblance (angl. *likelihood-ratio tests*) à l'aide de la fonction *anova()* du package *stats* (R Core Team, 2020). Les modèles ont été montés en utilisant la fonction *lmer()* du package *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2017). A partir du tout premier modèle, tous les modèles contenaient toujours les *Phrases expérimentales* et les *Participant-e-s* comme effets aléatoires. Nous avons obtenu le niveau significatif et les tailles d'effets estimées en utilisant la fonction *summary()* du *base R* (R Core Team, 2020). Si le modèle final contenait une interaction entre deux effets fixes nous avons effectué des comparaisons *post-hoc* en utilisant la fonction *glht()* du package *multcomp* (Hothorn, Bretz & Westfall, 2008).

Tous les temps de lecture inférieurs à 50ms et supérieurs à 3s ont été exclus de l'analyse. Pour les questions de vérification, les seuils étaient à 50ms minimum et à 10s maximum. Ensuite, nous avons fait une transformation logarithmique (comme dans Crible et al., 2021). L'asymétrie de nos données a été contrôlée à l'aide de la fonction *skewness()* du package *stats* (R Core Team, 2020).

Le temps de lecture pour les segments 5, 6 et 7

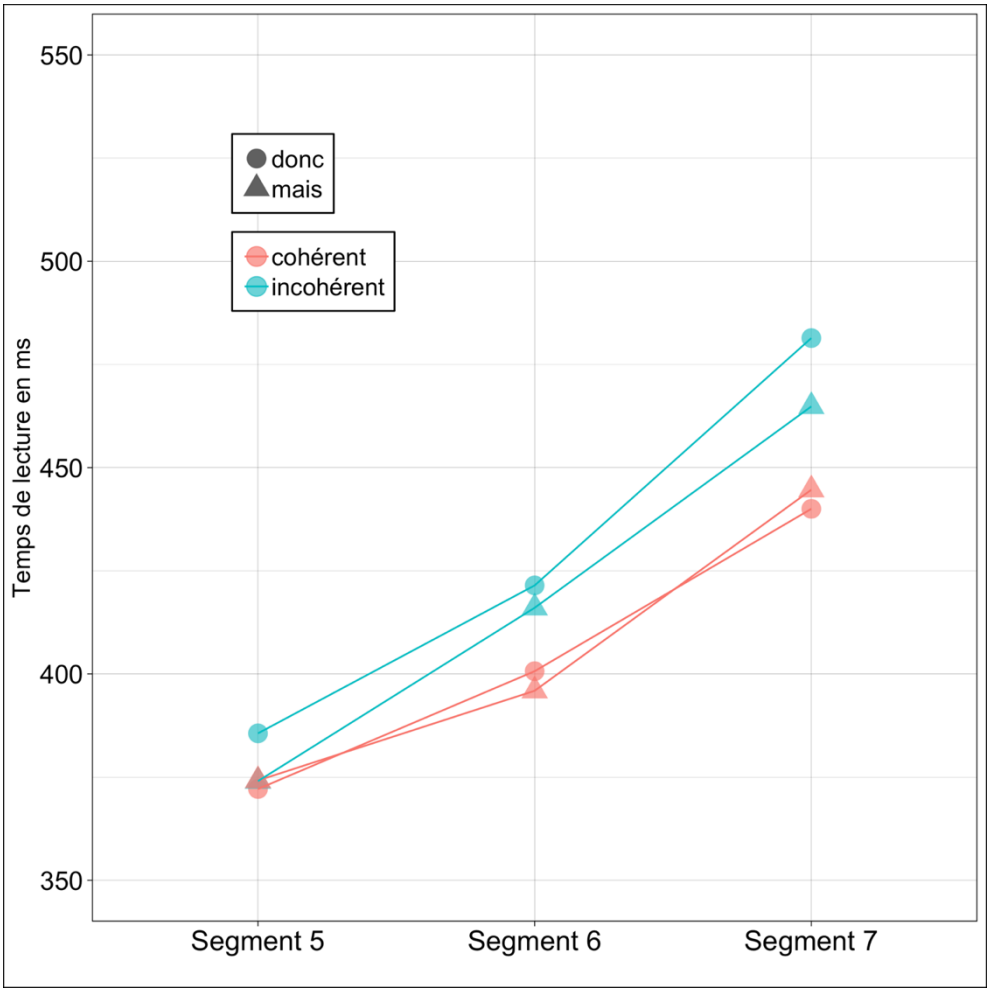
Pour les temps de lecture du segment 5, ni l'ajout du *Type de phrase* (cohérente ou incohérente, $\chi(1) = 1.99$, $p = .16$), ni du *Type de connecteur* (*donc* ou *mais,* $\chi(1) = 2.99$, $p = .09$) comme effets fixes ont amélioré le modèle. Pour le segment 6, le modèle s'est amélioré en ajoutant le *Type de phrase* comme effet fixe ($\chi(1) = 10.65$, $p = 9.28^{e-06}$). L'interaction avec le *Type de connecteur* n'a pas amélioré davantage le modèle ($\chi(1) = 1.49$, $p = .48$). Pour le dernier segment de la phrase, le segment 7, l'ajout du *Type de phrase* comme effet fixe a amélioré le modèle ($\chi(1) = 13.24$, $p < .0005$) ainsi que l'ajout de l'interaction avec le *Type de connecteur* l'a amélioré davantage ($\chi(2) = 13.89$, $p < .001$). Les outputs des modèles des segments 6 et 7 sont reportés dans le Tableau 4.3.

Tableau 4.3. Expérience 1. L'output des modèles pour les segments 6 et 7.

	β	*ET*[18]	*ddl*	*t*	*Pr(>|t|)*
Segment 6					
(Intercept)	-1.13	.05	141.13	-24.55	$< 2.00^{e-16}$
Type (*incohérent*)	.05	.01	4526.17	4.44	9.29^{e-06}
Segment 7					
(Intercept)	-0.98	.04	157.00	-25.01	$< 2.00^{e-16}$
Type (*incohérent*)	.07	.02	4554.00	4.31	< .00005
Connecteur (mais)	.00	.02	4556.00	-0.26	.79
Type (*incohérent*) : Connecteur (*mais*)	-0.06	.02	4557.00	-2.44	.01

Comme visualisé dans la Figure 4.1, nous observons un ralentissement pour les phrases incohérentes à partir du Segment 6.

Figure 4.1. Expérience 1, les temps de lecture moyens pour les segments 5, 6 et 7



[18] Nous utilisons l'abréviation *ET* pour *erreur type* et non pour *écart type*. Les écarts types ne seront pas abrégés.

De plus, nous observons une interaction avec le type de connecteur dans le dernier segment de la phrase. Comme le montre déjà visuellement la Figure 4.1, les comparaisons *post-hoc* indiquent que les phrases incohérentes ont été lues plus lentement lorsqu'elles contenaient *donc* (β = -0.06, *ET* = .02, z = -3.72, p < .005). En outre, les comparaisons indiquent que les phrases qui contenaient *mais* n'ont pas été lues significativement plus lentement lorsqu'elles étaient incohérentes (β = .01, *ET* = .02, z = .84, p = .83).

Le temps de réponse à la question de vérification

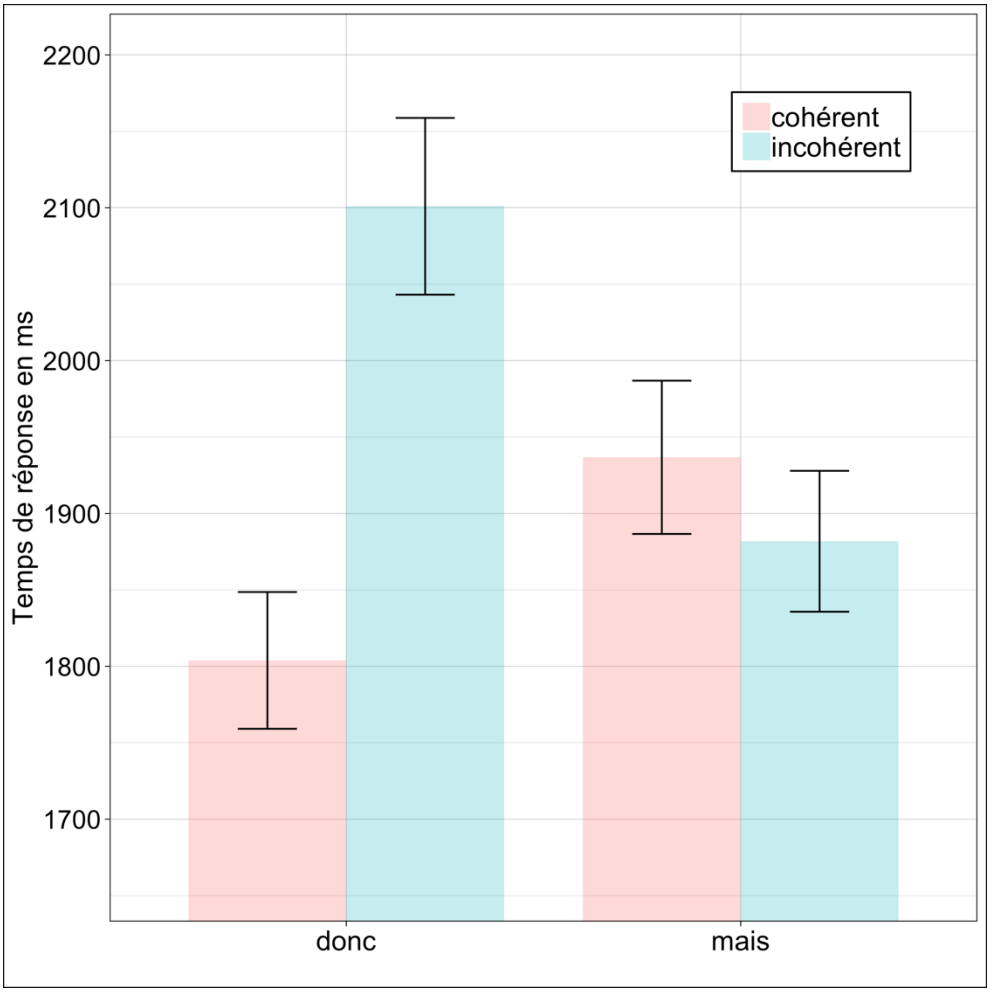
L'ajout du *Type de la phrase* comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 46.97$, $p = 7.22^{e-12}$). Lors de l'ajout de l'interaction avec le *Type de connecteur*, le modèle s'est encore amélioré ($\chi(2) = 8.35$, $p < 2.2^{e-16}$). L'output du modèle final est reporté dans le Tableau 4.4.

Tableau 4.4. Expérience 1, output du modèle final pour la question de vérification

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	.51	.03	154.64	15.81	2.00^{e-16}
Type (*cohérent*)	.14	.01	4555.62	11.12	2.00^{e-16}
Connecteur (*mais*)	.06	.01	4556.10	5.19	2.19^{e-07}
Type (cohérent) : Connecteur (*mais*)	-0.15	.02	4556.05	-8.78	2.00^{e-16}

Comme le montre également la visualisation dans la Figure 4.2, les comparaisons *post-hoc* indiquent que les participant-e-s répondaient pour une phrase cohérente plus rapidement lorsqu'elle contenait *donc* plutôt que *mais* (β = .07, *ET* = .01, z = 5.19, $p < 1^{e-04}$). Pour les phrases incohérentes nous observons l'effet inverse, c'est-à-dire qu'une phrase incohérente suscitait un temps de réponse plus rapide lorsqu'elle contenait *mais* (β = .01, *ET* = .01, z = -7.26, $p < 1^{e-04}$).

Figure 4.2. Expérience 1, le temps de réponse à la question de vérification



Nous observons également que les phrases incohérentes suscitaient un temps de réponse plus long que les phrases cohérentes. Cependant, cet effet émergeait seulement pour les phrases cohérentes et incohérentes qui contenaient *donc* (β = .14, *ET* = .01, *z* = 11.12, $p < 1^{e-04}$), et non pour les phrases cohérentes et incohérentes qui contenaient *mais* (β = -0.02, *ET* = .01, *z* = -1.28, *p* = .58).

4.3.4 Discussion

Dans cette expérience, nous avons mesuré le temps de lecture des personnes natives pour des phrases cohérentes et incohérentes qui ont été marquées soit avec le connecteur concessif *mais*, soit avec le connecteur de conséquence *donc*. En particulier, nous nous sommes intéressés à la question de savoir comment les personnes natives réagissent et traitent l'incohérence.

Les résultats obtenus démontrent clairement que les participant-e-s de cette expérience étaient hautement sensibles à l'incohérence. Nous avons observé des

effets de ralentissement à partir du segment 6. Ces ralentissements peuvent être attribués aux perturbations du traitement mental suscités par l'incohérence. Ils se manifestaient également dans la région postcritique, c'est-à-dire la réponse à la question de vérification.

Nous avons observé des effets différents selon le type de connecteur utilisé : la perturbation était plus prononcée et constante pour le connecteur *donc* que pour le connecteur *mais*. Pour le connecteur concessif, nous n'avons observé qu'un ralentissement dans le segment 6 pour les phrases incohérentes. Ensuite, cet effet s'est affaibli pour les phrases contenant *mais* de telle sorte n'a plus été constaté statistiquement dans le segment suivant (ni dans la région postcritique). Cela pourrait refléter le fait que l'incohérence est traitée différemment entre les relations concessives et les relations de conséquence. Ainsi, cette observation confirme les résultats de Lyu et al. (2020) qui n'avaient montré également qu'un effet temporaire pour la résolution de l'incohérence dans une phrase concessive. Nous allons discuter les raisons qui en sont responsables dans la conclusion.

Finalement, nous observons également une différence entre les relations de cohérence, lorsqu'elles étaient correctement marquées : une phrase causale (cohérente) a suscité des temps de réponses plus courtes aux questions de vérification qu'une phrase concessive (cohérente). Cela peut s'expliquer par la différence de complexité des relations de cohérence (Sanders et al., 1992), les phrases concessives étant cognitivement plus complexes que les phrases causales.

En résumé, nous avons pu observer une réaction immédiate à l'incohérence pour des phrases qui sont marquées avec deux connecteurs extrêmement fréquents. Cela soulève la question de savoir si la réaction à l'incohérence se manifesterait également si rapidement lorsque des connecteurs moins fréquents sont utilisés. Dans ce qui suit, nous allons refaire la même expérience avec deux connecteurs différents, qui sont beaucoup moins fréquents que ceux utilisés dans l'expérience actuelle : les connecteurs *ainsi* et *néanmoins*. A cause de leur fréquence basse, nous supposons que les personnes lectrices n'ont pas forcément la même vitesse de réaction à l'incohérence pour ces connecteurs.

4.4 Expérience 2 : la réaction à l'incohérence pour des phrases contenant les connecteurs *ainsi* et *néanmoins*

Dans cette expérience, nous allons refaire l'expérience précédente seulement cette fois, avec des connecteurs moins fréquents. Les connecteurs utilisés sont *ainsi* pour indiquer une conséquence et *néanmoins* pour indiquer une concession. Comme nous avons déjà vu au chapitre 4.3, les connecteurs *ainsi* et *néanmoins* sont significativement moins fréquents que *donc* et *mais*. Ainsi, nous supposons que les personnes natives ne bénéficient pas d'une facilitation du traitement mental pour ces connecteurs (telle qu'observée dans la première expérience). Si cela est le cas, nous devrions observer des réactions à l'incohérence un peu plus tardives en comparaison avec les réactions observées dans la première expérience.

4.4.1 Les participant-e-s

Nous avons recruté 128 participant-e-s via *Prolific* (âge moyen = 28.5, écart type = 8.67, 35 femmes). Tous les participant-e-s étaient de langue maternelle française et avaient montré une bonne participation dans des études précédentes sur la plateforme *Prolific* (minimum de 95% de participations approuvées). Tous les participant-e-s ont donné leur consentement de participer dans cette expérience et ont été rémunéré-e-s par £3.15. Aucun-e des participant-e-s n'avait participé à l'Expérience 1.

4.4.2 Le design et la procédure

Nous avons refait l'expérience 4.3 en modifiant les phrases expérimentales : dans les phrases concessives, nous avons utilisé le connecteur *néanmoins* au lieu de *mais* et dans les phrases de conséquence, *donc* a été échangé contre *ainsi*. La procédure était la même que dans l'expérience précédente.

4.4.3 L'analyse

Nous avons analysé les données de cette expérience de la même manière que les données de l'expérience précédente, comme décrit au chapitre 4.4.3.

Le temps de lecture pour les segments 5, 6 et 7

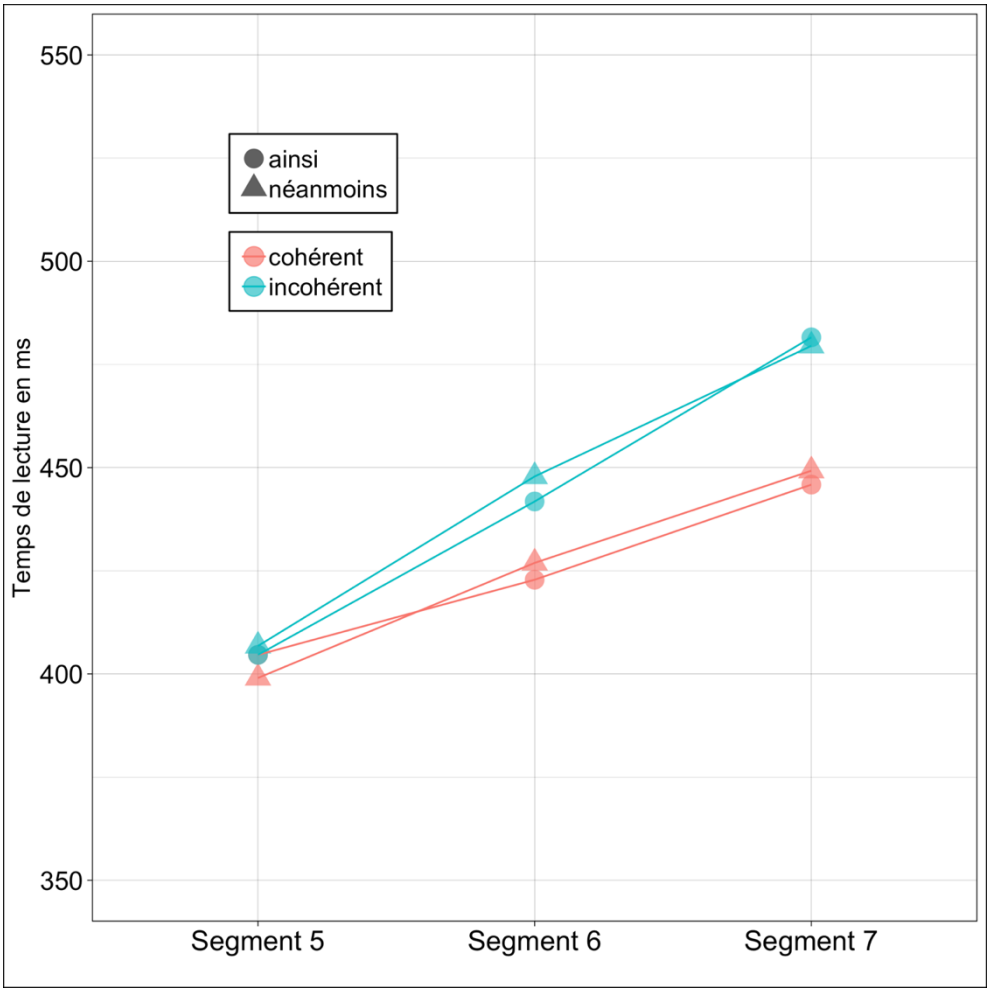
Pour le segment 5, ni l'ajout du *Type de phrase* ($\chi(1) = 1^{e-04}$, $p = .99$), ni le *Type de connecteur* ($\chi(1) = 37$, $p = .37$) comme effets fixes n'ont amélioré le modèle. Dans le segment suivant (le segment 6), à nouveau, ni l'ajout du *Type de phrase* ($\chi(1) = .29$, $p = .59$), ni le *Type de connecteur* ($\chi(1) = 66$, $p = .42$) comme effets fixes ont amélioré le modèle. Dans le dernier segment de la phrase, à savoir le segment 7, l'ajout du *Type de phrase* comme effet fixe a amélioré le modèle ($\chi(1) = 12.89$, $p < .0005$). En ajoutant l'interaction avec le *Type de connecteur* le modèle ne s'est pas amélioré davantage ($\chi(2) = 2.19$, $p = .36$). L'output du modèle final est reporté dans le Tableau 4.5.

Tableau 4.5. Expérience 2, l'output du modèle final du segment 7

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.96	.05	132.23	-21.20	2.00^{E-16}
Type (*incohérent*)	.05	.01	2921.65	3.60	< .0005

Nous concluons donc que les personnes ont ralenti de manière significative à l'incohérence lors de la lecture du dernier segment de la phrase. Comme on le voit également dans la Figure 4.3, on n'observe pas d'interaction avec le type de connecteur.

Figure 4.3. Expérience 2, les temps de lecture moyens pour les segments 5, 6 et 7



Le temps de réponse à la question de vérification

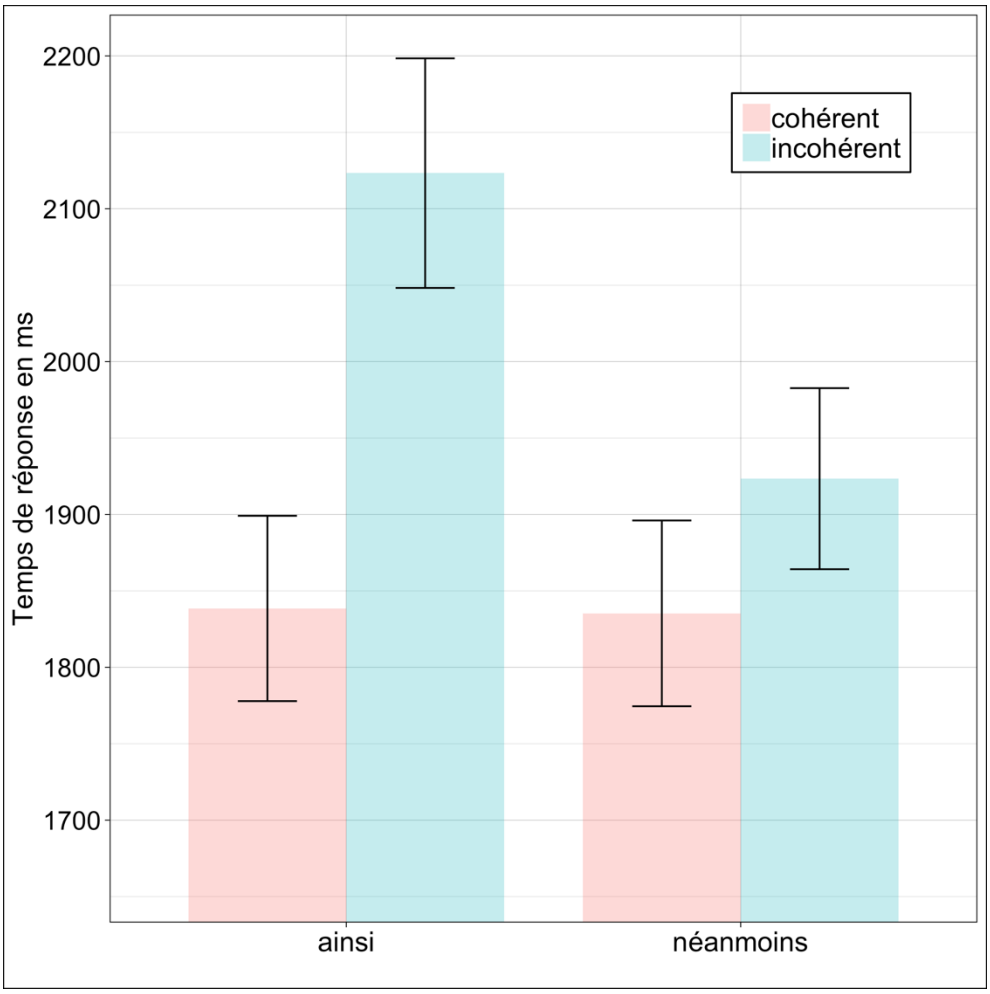
Pour le temps de réponse à la question de vérification, l'ajout du *Type de phrase* comme effet fixe ($\chi(1) = 50.32$, $p = 1.31^{e-12}$) ainsi que son interaction avec le *Type de connecteur* ($\chi(2) = 23.04$, $p = 9.94^{e-06}$) ont successivement amélioré le modèle. L'output de ce modèle est reporté dans le Tableau 4.6.

Tableau 4.6. Expérience 2. L'output du modèle final du temps de réponse à la question de vérification

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	.53	.03	172.28	15.65	2.00^{e-16}
Type (*incohérent*)	.14	.02	2857.11	7.79	9.63^{e-15}
Connecteur (*néanmoins*)	.02	.02	2756.71	1.06	.29
Type (*incohérent*) : Connecteur (*néanmoins*)	-0.10	.03	2820.62	-4.03	5.65^{e-05}

Comme on peut le voir dans la Figure 4.4, les temps de réponse aux questions de vérification pour les phrases cohérentes ne différeraient pas en fonction du type de connecteur (β = .02, *ET* = .02, z = 1.06, p = .71).

Figure 4.4. Expérience 2. Les temps de réponses aux questions de vérification



En revanche, on observe un effet entre les phrases incohérentes selon le connecteur utilisé : les phrases incohérentes qui contenaient *ainsi* suscitaient des temps de réponse significativement plus longues que toutes les autres phrases (en comparaison avec *ainsi / cohérent* : β = .14, *ET* = .02, z = 7.79, p < .001; en comparaison avec *néanmoins / cohérent* : β = .12, *ET* = .02, z = 6.91, p < .001; en comparaison avec *néanmoins / incohérent* : β = -0.08, *ET* = .02, z = -4.71, p < .001). Finalement, nous n'observons pas de différence entre les temps de réponse aux questions pour les phrases contenant *néanmoins*, qu'elles soient cohérentes ou incohérentes (β = .04, *ET* = .02, z = 2.15, p = .14).

4.4.4 Discussion

Dans cette expérience, nous avons refait la première expérience en échangeant les connecteurs : au lieu des connecteurs *donc* et *mais* (Expérience 1), nous avons utilisé les connecteurs *ainsi* et *néanmoins*.

Comme dans l'expérience précédente, nous avons observé que les participant-e-s étaient sensibles à l'incohérence. Ils montraient des ralentissements de lecture pour les phrases incohérentes, ce qu'on peut attribuer aux perturbations du traitement mental. Cependant, contrairement à la première expérience, ces perturbations n'émergeaient que dans le dernier segment de la phrase. Comme nous avons utilisé exactement les mêmes phrases que dans la première expérience, nous concluons que cet effet doit être dû aux connecteurs testés : il semble que les personnes réagissent plus lentement à l'incohérence lorsque la phrase est marquée par un connecteur moins fréquent.

Comme dans la première expérience, nous n'avons observé dans l'expérience actuelle qu'une perturbation temporaire pour les phrases incohérentes marquées avec un connecteur concessif. Tandis que nous avons observé cet effet dans la première expérience seulement dans le sixième segment, dans l'expérience actuelle l'effet n'émergeait qu'au septième segment. Cela pourrait indiquer une résolution différente de l'incohérence entre les phrases concessives et causales. Nous allons revenir sur cette question dans la conclusion.

Finalement, nous n'avons pas observé pour les phrases cohérentes que les phrases concessives suscitaient des réponses moins rapides à la question de vérification que les phrases de conséquence. Cet effet a été trouvé dans la première expérience et nous l'avons attribué à la complexité élevée de traiter cognitivement une concession. L'absence de cet effet dans l'expérience actuelle pourrait être dû à la fréquence basse des connecteurs : des connecteurs moins fréquents pourraient rendre le traitement mental de la phrase un peu plus difficile, de sorte que les différences subtiles entre la complexité des différentes relations seraient gommées. En testant d'autres connecteurs plus rares, nous allons examiner si cette interprétation est valide.

En résumé, nous avons vu dans cette deuxième expérience que les personnes réagissaient de la même manière à l'incohérence que dans la première expérience, et nous en concluons que la lecture dans une langue maternelle est

assez robuste. Cependant, dans l'expérience actuelle, les réactions à l'incohérence n'ont atteint le seuil de significativité statistique qu'à un stade ultérieur de la phrase. Cela signifie que les connecteurs moins fréquents ne facilitent pas la détection de l'incohérence autant que les connecteurs testés dans Expérience 1 (pour lesquels nous avons observé une réaction à l'incohérence déjà dans le sixième segment).

Cependant, les connecteurs ne se distinguent pas seulement par leur fréquence : les connecteurs qui indiquent plusieurs relations peuvent également avoir une influence sur le traitement, car ils doivent être désambiguïsés. Comme ce processus de désambiguïsation prend un certain temps et nécessite une charge mentale supplémentaire, la sensibilité à l'incohérence pourrait être réduite. Ainsi, on peut supposer que l'incohérence dans les phrases contenant des connecteurs polyfonctionnels provoque des réactions encore plus tardives que dans l'expérience actuelle. A cet égard, il est vrai que le connecteur testé dans la première expérience, *alors,* peut être considéré comme étant polyfonctionnel. Pourtant, ces résultats obtenus ne sont pas concluants pour le facteur de la polyfonctionnalité étant donné que *alors* est un connecteur extrêmement fréquent et pour lequel nous avons observé un effet de facilitation pour la détection de l'incohérence dans la première expérience. Pour tester indubitablement l'hypothèse selon laquelle que la polyfonctionnalité peut diminuer la sensibilité à l'incohérence, il convient donc de tester des connecteurs rares qui sont, en outre, polyfonctionnels.

4.5 Expérience 3 : la réaction à l'incohérence pour des phrases contenant les connecteurs *aussi* et *or*

Dans la troisième expérience, nous allons examiner l'impact des connecteurs rares et polyfonctionnels. Peu de données indiquent comment les connecteurs polyfonctionnels peu fréquents sont traités, c'est-à-dire si la complexité de la polyfonctionnalité s'ajoute à la complexité supposée des connecteurs rares. Étant donné que les connecteurs polyfonctionnels doivent être désambiguïsés avant que l'on puisse accéder à leur instruction procédurale, nous supposons que l'identification correcte d'une relation implique un effort de traitement plus important lorsqu'elle est indiquée par un connecteur polyfonctionnel par rapport à un connecteur monofonctionnel. Si la polyfonctionnalité s'ajoute à la difficulté de traiter

un connecteur rare, la réaction à l'incohérence devrait émerger encore plus tardivement que dans l'expérience précédente.

Pour tester cette hypothèse, nous allons refaire les expériences précédentes, cette fois-ci avec les connecteurs *aussi* et *or* qui, sont selon le dictionnaire des connecteurs *LexConn* (Roze et al., 2012), polyfonctionnels. *Aussi* est utilisé soit pour indiquer une conséquence (la fonction testée dans notre expérience, voir exemple (10)), soit pour mettre en évidence la concomitance ou parallélisme des actions, comme dans (11) (tous les exemples sont tirés de *LexConn*).

10) Eux-mêmes n'osent se parler librement devant elle. Aussi continuent-ils à passer une partie des nuits dans ma chambre.

11) Marie adore aller au cinéma. Julie aussi y va de temps en temps.

Le connecteur *or* est soit utilisé pour indiquer une concession (la fonction testée dans notre expérience, voir exemple (12)), soit une continuation, comme en (13) :

12) Tout, dans son comportement, semblait dire : puisqu'il ne m'aime plus, rien ne m'importe. Or je l'aimais encore, et même je ne l'avais jamais tant aimée; mais le lui prouver ne m'était plus possible.

13) Nous sommes attirés par qui nous flatte, de quelque façon que ce soit. Or François admirait le comte.

Outre leur polyfonctionnalité, ces deux connecteurs peuvent être, selon notre analyse de corpus, considérés comme plutôt rares.

Nous avons obtenu les fréquences des connecteurs à l'aide du corpus ORFEO, une collection des différents sous-corpus oraux et écrits (~10 millions de mots, Benzitoun et al., 2016). Comme nous étions intéressés par les utilisations précises des connecteurs *or* et *aussi*, (à savoir leurs fonctions d'indiquer une concession et une conséquence) nous avons analysé pour les deux connecteurs 100 occurrences, tirées au hasard du corpus. Le nombre d'occurrence résultant de soit *aussi* indiquant une conséquence, soit *or* indiquant une concession, nous a permis de calculer les fréquences normalisées par million de mots.

Le Tableau 4.7 reporte les fréquences normalisées de *aussi* et *or*, ainsi que, par convenance, celles des connecteurs testés dans les expériences précédentes. Il en ressort que *aussi* et *or* sont significativement moins fréquents que les connecteurs utilisés dans la première expérience, à savoir *donc* et *mais* (χ^2 = 27.21,

ddl = 1, *p* = 1.83^{e-07}, calculée à l'aide de la fonction chisq.test() - du paquet *stats v3.6.2* dans *R,* R Core Team, 2020).

Tableau 4.7. Les fréquences normalisées par un million de mots des connecteurs testés

Expérience	Connecteur	*À l'écrit*	*À l'oral*	*total*
1	mais	191	221	412
1	donc	113	221	334
2	néanmoins	20	6	26
2	ainsi	97	35	132
3	aussi (*conséquence*)	91 (estimé)	0 (estimé)	91
3	or (*concession*)	28 (estimé)	11 (estimé)	39

4.5.1 Les participant-e-s

Nous avons recruté 114 participant-e-s via la plateforme *Prolific* (âge moyen= 28, écart type = 8.97, 54 femmes). Tous les participant-e-s avaient montré une bonne participation dans des études précédentes sur la plateforme *Prolific* (minimum de 95% de participations approuvées) et ont été rémunéré-e-s avec £3.15. Tous les participant-e-s ont donné leur consentement de participer dans cette expérience. Aucun-e des participant-e-s n'avait participé à l'Expérience 1 ou 2.

4.5.2 Le design et la procédure

Nous avons refait la même expérience que dans les deux premières. De nouveau, nous avons modifié les phrases expérimentales : dans cette expérience, nous avons utilisé le connecteur *or* dans les phrases concessives et *aussi* dans les phrases de conséquence. La procédure de l'expérience était la même que dans les expériences précédentes, telle que décrite au point 4.3.2.

4.5.3 L'analyse

Nous avons effectué les analyses pour cette expérience de la même manière que dans les expériences précédentes, telle que décrites au chapitre 4.3.3.

Le temps de lecture pour les segments 5, 6 et 7

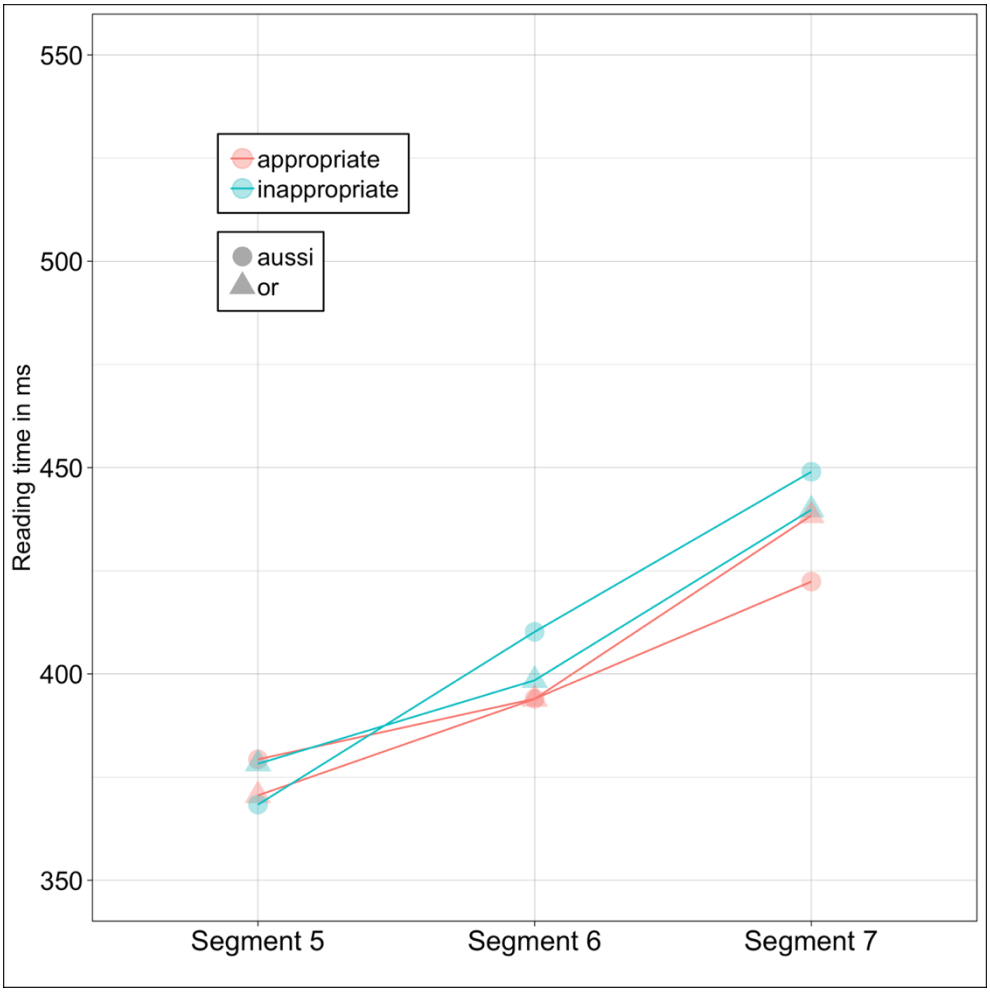
L'ajout du *Type de phrase* comme effet fixe n'a pas amélioré le modèle pour le segment 5 ($\chi(1) = .09$, $p = .77$). De même, l'ajout du *Type de connecteur* comme effet fixe n'a pas non plus amélioré le modèle en comparaison avec le modèle zéro qui ne contenait pas d'effet fixe ($\chi(1) = 1.36$, $p = .24$). Pour le segment 6, ni l'ajout du facteur *Type de phrase* ($\chi(1) = .96$, $p = .34$), ni le *Type de connecteur* ($\chi(1) = .73$, $p = .39$) a amélioré le modèle. En revanche, pour le dernier segment de la phrase, l'ajout du *Type de phrase* comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 11.34$, $p < .001$). En ajoutant l'interaction avec le *Type de connecteur* le modèle ne s'est pas amélioré davantage ($\chi(2) = 3.43$, $p = .18$). L'output du modèle final qui contenait donc seul le *Type de phrase* comme effet fixe (et *Participant-e-s* et *Phrases* comme effets aléatoires) est reporté dans le Tableau 4.8.

Tableau 4.8. Expérience 3, l'output du modèle final du segment 7

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.99	.03	146.22	-28.42	2.00^{e-16}
Type (*incohérent*)	.04	.01	4361.60	3.37	< .001

Comme on peut le voir également dans la Figure 4.5, nous observons un ralentissement pour les phrases incohérentes seulement dans le segment 7.

Figure 4.5. Expérience 3, les temps de lecture pour les segments 5, 6 et 7



Les temps de réponses aux questions de vérification

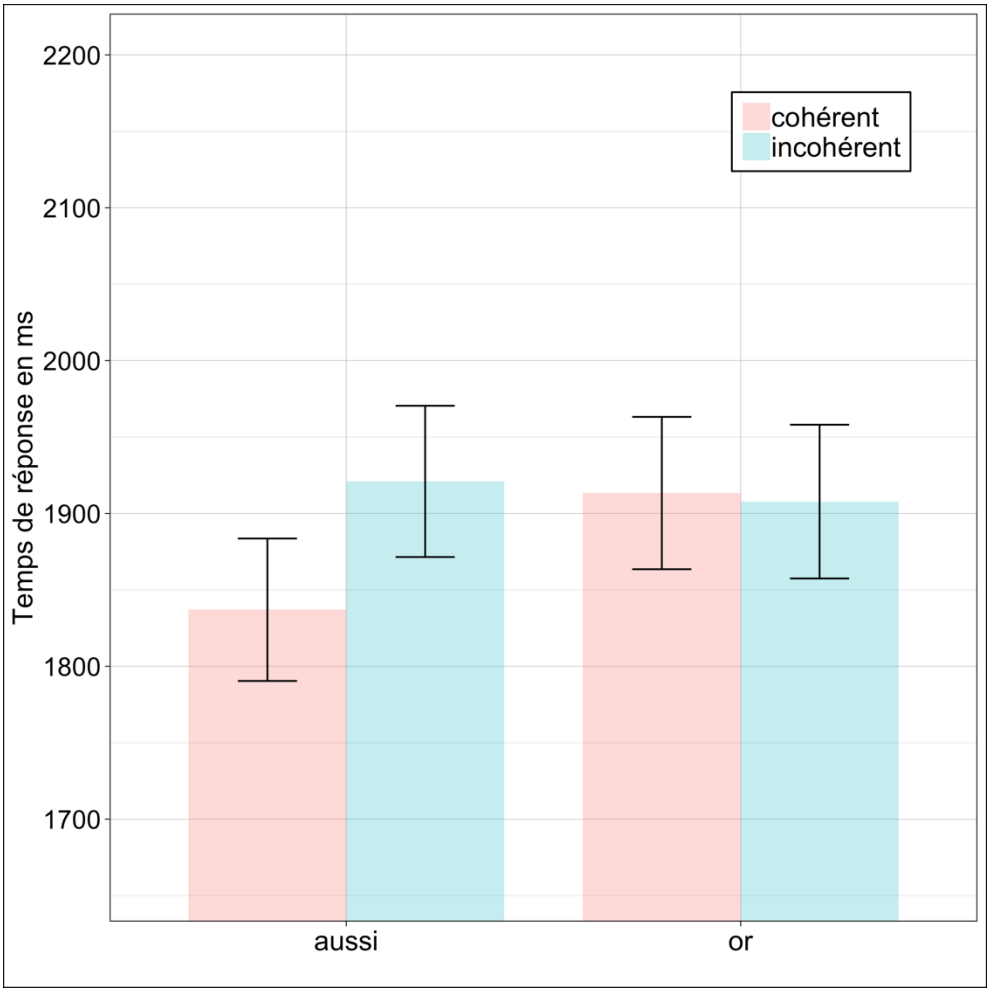
Pour les temps de réponses aux questions de vérification, l'ajout successif du *Type de phrase* comme effet fixe ($\chi(1) = 9.73$, $p < .005$), ainsi que son interaction avec le *Type de connecteur* ($\chi(2) = 2.40$, $p < .01$) a amélioré le modèle de manière significative. L'output du modèle est reporté dans le Tableau 4.10.

Tableau 4.10. Expérience 2. L'output du modèle final du temps de réponse à la question de vérification

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	.53	.03	161.52	17.46	$2\text{-}00^{e\text{-}16}$
Type (*incohérent*)	.06	.01	4312.67	4.59	$4.57^{e\text{-}06}$
Connecteur (*or*)	.04	.01	4311.12	3.11	< .005
Type (*incohérent*) : Connecteur (*or*)	-0.06	.02	4311.20	-3.37	< .001

Les temps de réponse sont visualisés dans la Figure 4.6. Pour les questions qui précédaient les phrases cohérentes, les comparaisons *post-hoc* indiquent que les phrases qui contenaient *aussi* ont suscité les réponses plus rapides que celles qui contenaient *or* (β = .04, *ET* = .01, z = 3.11, p =.01). Les temps de réponse ne différaient pas pour les phrases qui contenaient *or*, qu'elles soient cohérentes ou non (β < .01, *ET* = .01, z = .16, p > .99), tandis que les phrases cohérentes qui contenaient *aussi* ont suscité des réponses plus rapides que les phrases incohérentes (β = .06, *ET* = .01, z = 4.59, p < .001).

Figure 4.6. Expérience 3, les temps de réponses moyens aux questions de vérification en moyennes



4.5.4 Discussion

Dans cette expérience, nous avons mesuré le temps de lecture pour des phrases cohérentes ou incohérentes qui ont été marqués avec *aussi* et *or*. Contrairement

aux deux premières expériences, les connecteurs testés étaient donc des connecteurs polyfonctionnels et rares.

Nous avons observé dans cette expérience que les participant-e-s étaient sensibles à l'incohérence et ralentissaient leur lecture dans des phrases incohérentes. Il est important de noter à cet égard que la réaction à l'incohérence n'émergeait que dans le dernier segment de la phrase. Comme cela était le cas également pour les connecteurs *néanmoins* et *ainsi* de la deuxième expérience, nous concluons que cette réaction quelque peu retardée à l'incohérence est due à la fréquence plus basse des connecteurs. Ainsi, nous n'avons pas pu observer un effet spécifiquement attribuable à la polyfonctionnalité : l'incohérence a provoqué une réaction dans le dernier segment de la phrase que la phrase soit marquée par des connecteurs rares monofonctionnels (Expérience 2) ou rares et polyfonctionnels (Expérience actuelle). Il semble donc que la polyfonctionnalité ne représente pas un obstacle pour la détection de l'incohérence, respectivement que le processus de désambiguïsation du connecteur est rapidement accompli.

Nous avons également observé des effets entre les deux types de connecteurs testés. Premièrement, nous avons vu pour les phrases cohérentes que le connecteur concessif provoquait des temps de réponse à la question de vérification plus lents que le connecteur de conséquence – ce qui s'explique de nouveau par le degré de complexité différent entre ces relations (voir aussi p.ex., Köhne & Demberg, 2013). Au vu de l'expérience précédente, nous pouvons donc exclure l'interprétation selon laquelle l'utilisation d'un connecteur rare fait disparaitre la différence subtile entre les relations.

Par ailleurs, il est curieux que nous n'ayons observé qu'un effet temporaire pour l'incohérence dans les phrases marquées avec le connecteur concessif (comme dans les deux expériences précédentes). Nous discuterons l'absence d'effet d'incohérence dans les derniers segments des phrases contenant les connecteurs concessifs à travers les trois expériences plus en détail dans la conclusion.

En résumé, nous avons observé dans l'expérience actuelle une réaction à l'incohérence comparable à celle de la deuxième expérience. Nous en concluons que l'effet de fréquence est relativement stable tandis que la polyfonctionnalité des connecteurs testés ne représente pas une difficulté supplémentaire à la lecture en L1.

4.6 Conclusion

Au cours de trois expériences, nous avons mesuré le temps de lecture des personnes natives pour des phrases concessives et causales. Comme certaines de ces phrases étaient incohérentes, nous avons pu retracer la durée et l'intensité des perturbations du traitement mental qui en résultaient. Tandis que les phrases de la première expérience contenaient des connecteurs extrêmement fréquents, les connecteurs utilisés dans la deuxième expérience étaient plus rares. En comparant les réactions à l'incohérence observées dans les trois expériences, nous pouvons tirer des conclusions sur la manière dont certains connecteurs renforcent ou atténuent la sensibilité à l'incohérence.

Comme attendu, le traitement mental des participant-e-s a été fortement perturbé par l'incohérence dans toutes les expériences, ce qui s'est manifesté par des temps de lecture plus longs. En plus, nous avons également vu que la réaction à l'incohérence était liée au connecteur utilisé : pour les relations causales, nous avons observé que les personnes réagissaient déjà à l'incohérence à partir du segment 6 lorsque la phrase était introduite par un connecteur fréquent, telle que *donc* (Expérience 1). Lorsqu'elle était introduite par un connecteur rare (Expériences 2 et 3), les réactions à l'incohérence n'émergeaient seulement au dernier segment de la phrase. Cela démontre que la sensibilité à l'incohérence en L1 varie en fonction du connecteur en question. A cet égard, il est probable aussi que la lecture des personnes natives n'était pas *perturbée* ou *compliquée* par la fréquence ou la polyfonctionnalité des connecteurs, mais plutôt *facilitée* par le connecteur fréquent. Ainsi, nous avons observé plutôt un effet de facilitation dans l'Expérience 1 plutôt que des effets de perturbation dans les Expériences 2 et 3.

En revanche, nous n'avons pas observé des réactions différentes à l'incohérence entre les expériences 2 et 3. Nous concluons que la polyfonctionnalité ne rend pas un connecteur rare encore plus difficile à traiter. Il semble plutôt qu'un connecteur polyfonctionnel peut être désambiguïsé très rapidement. Cette conclusion est également étayée par les observations de la première expérience. Malgré le fait que le connecteur concessif utilisé, *alors,* soit également polyfonctionnel, l'incohérence a été détectée très tôt dans la phrase.

Nous avons également observé dans les Expériences 1 et 2 qu'une relation causale (cohérente) a suscité des temps de réponse à la question de vérification

plus rapides qu'une relation concessive (cohérente). Cela reflète la complexité différente entre les relations de cohérence (Sanders et al., 1992). Il est curieux que nous n'ayons pas observé un tel effet de complexité cognitive de la relation de cohérence dans la deuxième expérience alors que cet effet était présent dans la première et troisième expérience. Comme nous n'avons pas d'explication pertinente pour l'absence de l'effet dans l'Expérience 2, nous concluons que plus de recherches sont nécessaires afin de déterminer quels connecteurs renforcent ou atténuent la complexité cognitive d'une concession.

Finalement, il est intéressant de noter que nous n'avons observé dans toutes les expériences qu'un effet temporaire des perturbations du traitement mental pour les phrases incohérentes qui contenaient les connecteurs concessifs (*mais, néanmoins* et *or*). Pour le connecteur *mais* un effet clair émergeait seulement dans les segments 6 et 7, pour *néanmoins* et *or* seulement dans le segment 7. On peut donner plusieurs explications à cette observation.

Premièrement, l'effet temporaire d'incohérence pour les phrases marquées avec un connecteur concessif pourrait s'expliquer par le fait que l'incohérence de ces phrases a pu être résolue. Considérons l'exemple (14), qui peut être interprété de manière suivante : il se peut que Jean veuille connaitre *toutes* les nouvelles durant *toute la journée*, ce qu'il n'arrivera pas en ne lisant que le journal du matin.

14) Jean veut connaitre les nouvelles du jour [mais] il lit le journal tous les matins.

Selon cette interprétation, l'incohérence de la phrase serait résolue. Cette explication soutiendrait également les résultats de Lyu et al (2020) qui ont montré que les phrases causales incohérentes entrainaient des perturbations plus longues que les phrases concessives. L'équipe de recherche en a conclu que « tandis que l'incohérence dans la causalité a été considérée comme complètement fausse (…) celle dans la concession a pu être résolue par des stratégies différentes » (Lyu et al., 2020 : 337, ma traduction).

Toutefois, il convient de noter que les phrases incohérentes contenant un connecteur concessif étaient, en soi, causales dans notre expérience. Considérez par exemple encore l'exemple (14), dans lequel la relation est causale si on ne tient pas compte du connecteur *mais*. Ainsi, une autre explication serait que la fonction

procédurale du connecteur concessif aurait pu être plus facilement négligée car on s'attend par défaut à une structure causale dans le discours (Sanders, 2005).

Troisièmement, l'effet temporaire pour le marquage cohérent ou incohérent avec un connecteur concessif pourrait être expliqué également par le fait que les connecteurs concessifs étaient plus saillants que les connecteurs causaux. Murray (1997), par exemple, a montré qu'un connecteur concessif mal utilisé provoque plus de perturbations mentales qu'une mauvaise utilisation d'un connecteur causal. Cela aurait pu mener des participant-e-s dans notre expérience à identifier plus rapidement une phrase incohérente marquée avec un connecteur concessif comme *incorrecte*. En conséquence, la phrase aurait été rejetée plus facilement et les participant-e-s n'auraient pas essayé de résoudre l'incohérence.

Limites et futures directions

Comme mentionné ci-dessus, nous avons conclu que les connecteurs fréquents peuvent fonctionner comme des catalyseurs de lecture qui facilitent la détection de l'incohérence en comparaison avec des connecteurs moins fréquents. Il convient, pourtant de noter que les connecteurs diffèrent également sur d'autres caractéristiques, ce qui implique potentiellement que la fréquence n'est pas le seul facteur déterminant. Une différence existe par exemple en ce qui concerne leurs occurrences dans la langue écrite ou orale. Les connecteurs de la première expérience sont fréquemment utilisés à la fois l'oral et à l'écrit tandis que les connecteurs des Expériences 2 et 3 sont notamment utilisés à l'écrit. En conséquence, les connecteurs diffèrent également quant au registre de langue qu'ils évoquent. Toutefois, il est difficile de distinguer ces différents facteurs les uns des autres, car la modalité, le registre et la fréquence dépendent fortement les uns des autres.

Un autre biais pourrait être le fait que les connecteurs fréquents et rares n'ont pas été directement comparés entre eux. Il convient toutefois de noter ici que nous avons testé un nombre assez important de participant-e-s dans les trois expériences, ce qui est censé annuler une potentielle différence entre les groupes de participant-e-s. Cependant, si des recherches futures testent des personnes locutrices dont la maitrise peut varier de manière encore plus individuelle, telles

que les personnes non-natives, il serait approprié de comparer directement les temps de lecture pour les connecteurs (voir chapitre 7).

Finalement, il convient de noter que le connecteur *aussi* aurait pu être mieux identifié grâce à sa position syntaxique, car ce connecteur indique toujours une conséquence lorsqu'il se trouve en position initiale de la phrase. Bien que cela ait pu représenter un biais dans notre étude, nous ne savons pas à quel degré les personnes lectrices sont vraiment sensibles aux contraintes syntaxiques des connecteurs. Nous concluons donc que plus de recherches sont nécessaires à cet égard (voir chapitre 9).

En conclusion, les expériences menées nous indiquent que le type de connecteur est déterminant pour l'aisance avec laquelle l'incohérence d'une phrase est détectée. Plus précisément, les connecteurs fréquents agissent en tant que catalyseurs de lecture. Ainsi, notre étude montre des différences considérables entre les différents connecteurs et ces dernières influencent la façon dont on lit un discours dans sa langue maternelle.

Inévitablement, ces observations soulèvent la question de savoir si l'hétérogénéité des connecteurs se refléterait également dans la maitrise des personnes non-natives. Compte tenu des différences entre les connecteurs et de leur impact sur la lecture, on peut supposer qu'elles rendraient également certains connecteurs plus difficiles à apprendre et à utiliser en L2 que d'autres. Pour examiner cette question plus en détail, nous allons présenter dans le chapitre suivant une expérience visant à étudier la maitrise des connecteurs par des apprenant-e-s non-natifs du français.

5 L'impact du type du connecteur sur la connaissance explicite en L2

Nous avons discuté au chapitre 3 que la maitrise des connecteurs dans une langue seconde soulève plusieurs questions et que les difficultés auxquelles les apprenant-e-s sont confronté-e-s sont loin d'être documentées et analysées de manière satisfaisante. Ainsi, la présente thèse vise à examiner plus en détail ces difficultés et à identifier les facteurs linguistiques et cognitifs qui en sont responsables. En nous basant sur les observations de la littérature résumées au chapitre 3 ainsi que sur les résultats obtenus au chapitre précédent, nous pouvons faire trois hypothèses.

Premièrement, nous supposons que la maitrise des connecteurs en L2 dépend hautement du connecteur en question, en d'autres termes, il y a des connecteurs qui sont plus complexes à apprendre et à utiliser que d'autres. Comme discuté au chapitre 3, plusieurs études de corpus montrent que les apprenant-e-s de langue sur- ou sous-utilisent certains types de connecteurs et tendent à s'appuyer sur un ensemble limité de *mots de confort* (e.g., Field & Yip, 1992; Granger & Tyson, 1996; Leedham & Cai, 2013). Bien que cela suggère que certains connecteurs sont plus difficiles à utiliser que d'autres, ces observations ne permettent de tirer que des conclusions limitées pour déterminer quels sont les connecteurs difficiles. Comme nous l'avons déjà discuté, les études de corpus ne permettent pas de savoir, par exemple, quels connecteurs les apprenant-e-s évitent consciemment d'utiliser.

Deuxièmement, nous supposons que des facteurs tels que la fréquence d'un connecteur ainsi que la relation de cohérence qu'il indique influencent la maitrise des connecteurs en L2. Nous avons vu au chapitre 4 que ces deux facteurs ont influencé la manière dont on lit dans sa L1. Pour rappel, lorsqu'une phrase était marquée avec un connecteur fréquent, les personnes interrogées ralentissaient leur lecture plus tôt par rapport aux phrases qui étaient marquées avec un connecteur plus rare. De plus, une phrase concessive (cohérente) provoquait des temps de réponses aux questions de vérification plus lents qu'une phrase causale

(cohérente). Ces observations nous laissent supposer que ces facteurs devraient également influencer la maitrise des connecteurs en L2.

Troisièmement, comme discuté au chapitre 3.5, il a été démontré qu'il existe des grandes différences individuelles parmi les apprenant-e-s, c'est-à-dire que certaines personnes arrivent mieux à maitriser les connecteurs que d'autres. Cependant, le types de compétences nécessaires pour bien maitriser les connecteurs n'ont pas encore été établies avec certitudes. Il est possible, par exemple, que les connaissances en vocabulaire et en grammaire ainsi qu'une exposition régulière avec la langue écrite soient corrélées avec la maitrise des connecteurs.

En conclusion, nos hypothèses se résument à trois questions:

- Est-ce qu'il y a des connecteurs français qui sont plus difficiles que d'autres pour les apprenant-e-s de langue ?
- Est-ce que ces difficultés pourraient être expliquées par la complexité cognitive de la relation de cohérence et la fréquence du connecteur ?
- Est-ce qu'il existe des caractéristiques individuelles des apprenant-e-s qui prédisent une meilleure compétence avec les connecteurs ?

Dans ce chapitre, nous allons présenter une expérience qui vise à répondre à ces questions. En particulier, nous nous intéressons aux connaissances explicites que les apprenant-e-s germanophones ont des connecteurs du français. Nous allons les examiner en utilisant une tâche à choix multiple d'insertion de connecteur réalisée de manière non-chronométrée.

Avant de présenter l'expérience, nous allons expliquer pourquoi nous pensons que la relation de cohérence et la fréquence peuvent influencer la maitrise des connecteurs en L2. Nous discuterons ensuite l'impact potentiel des compétences linguistiques et de l'exposition à l'écrit sur la maitrise des connecteurs en L2.

5.1 Quels facteurs peuvent rendre les connecteurs plus difficiles à utiliser dans une langue seconde?

Nous avons discuté dans les chapitres précédents que la maitrise des connecteurs en langue seconde pourrait dépendre du connecteur en question, de sorte qu'il y a certaines caractéristiques linguistiques et cognitives qui rendent un connecteur plus

difficile. Évidemment, ceci soulève la question de savoir quels sont ces facteurs et comment ils se répercutent sur la maitrise des connecteurs dans une L2. Dans cette étude, nous allons examiner deux facteurs à cet égard : celui de la *fréquence* du connecteur et celui de la complexité de la *relation de cohérence.* Dans ce qui suit, nous expliquerons pourquoi nous pensons que ces deux facteurs peuvent affecter la maitrise des connecteurs en L2.

5.1.1 Le rôle de la relation de cohérence

Les relations de cohérence diffèrent en termes de complexité cognitive (Sanders et al., 1992) et il a été démontré que ces différences influencent la manière dont les personnes lisent un discours dans leur L1 (p.ex., Köhne & Demberg, 2013; chapitre 4). Plusieurs théories expliquent que cela est dû au fait que les personnes lectrices traitent plus facilement les relations continues (Segal et al., 1991, Murray, 1994, 1997) et qu'elles infèrent par défaut et en l'absence d'autre indication une relation causale (Sanders, 2005). Comme ces théories se fondent sur la cognition en général, on peut se poser la question de savoir si ces observations sont aussi valides dans une deuxième langue – en d'autres termes, si la difficulté causée par la complexité cognitive d'une relation de cohérence affecte également les personnes qui écrivent et lisent dans une deuxième langue.

Il y a des indices qui tendent à indiquer que c'est bien le cas. Plusieurs études de corpus ont constaté que les apprenant-e-s en L2 sur- ou sous-utilisent les connecteurs en fonction de la relation de cohérence qu'ils indiquent (p.ex., Granger & Tyson, 1996; Tapper, 2005). Il semble que les connecteurs indiquant une relation additive soient utilisés de manière excessive, tandis que les relations contrastives et adversatives sont généralement mal indiquées, ce qui pourrait refléter leur complexité cognitive (p.ex., Milton & Tsang 1993; Park 2013; Ha, 2016; Shi, 2017). Ces observations peuvent être considérées comme une première indication du fait que la complexité de la relation de cohérence influence la maitrise des connecteurs en L2.

Dans une approche expérimentale, l'étude Ozono et Ito (2003) soutient cette idée. Dans leur travail, les chercheur-se-s ont fait l'hypothèse que la complexité de la relation de cohérence pourrait avoir un effet sur la maitrise des connecteurs en L2. En utilisant une tâche d'insertion, leurs résultats indiquaient que les personnes non-natives moins compétentes dans la langue cible, dans ce cas l'anglais,

montraient la tendance à avoir plus de difficulté avec le connecteur *however* (‘cependant’) qu’avec le connecteur *for example* (‘par exemple’). Pour expliquer cette observation, les chercheur-se-s avancent que les relations de cohérence plus complexes impliquent une charge cognitive élevée qui rend certains connecteurs plus difficiles. Cependant, un seul connecteur par relation de cohérence a été testé, ce qui ne permet pas de savoir si les différences observées sont vraiment dues à la différence entre relations de cohérence ou aux propriétés des connecteurs testés.

Toutes les observations mentionnées justifient donc l'hypothèse selon laquelle la relation de cohérence est un facteur qui influence l'utilisation et la maitrise d'un connecteur dans une langue seconde. Plus précisément, on peut supposer qu'un connecteur indiquant une relation de cohérence complexe (telle qu’une *concession*) est plus difficile à utiliser qu'un connecteur indiquant une relation plus simple (telle qu’une *conséquence*). Malheureusement, jusqu’à présent des données documentant sans ambiguïté l’influence de la complexité de la relation de cohérence sur les connaissances explicites des apprenant-e-s manquent encore.

Nous avons observé dans le chapitre précèdent qu’il existe, à part la relation de cohérence, d’autres propriétés des connecteurs qui pourraient influencer la manière dont les apprenant-e-s les maitrisent. Un de ces facteur est la fréquence du connecteur.

5.1.2 La fréquence du connecteur

Nous avons constaté au chapitre 4 que la fréquence d’un connecteur peut affecter la réaction à l’incohérence des personnes natives. Les participant-e-s réagissaient plus rapidement à l’incohérence lorsque la phrase était marquée avec un connecteur fréquent par rapport à un connecteur rare. De même, la fréquence élevée d’un connecteur facilite son acquisition dans la L1 (Zufferey & Gygax, 2020a). Cette observation soulève la question de savoir si un connecteur fréquent est également plus facile à maitriser pour les personnes non-natives.

En général, il est bien connu que le facteur de la fréquence peut être une variable qui influence la difficulté d’un élément linguistique en L2. Ellis (2006), par exemple, classifie le facteur de la *fréquence* comme un facteur parmi cinq autres qui rend un élément grammatical plus ou moins complexe (Ellis, 2006, voir aussi

chapitre 3.6). L'idée est que si un élément est plus fréquent, les apprenant-e-s d'une langue le rencontreront plus souvent et auront beaucoup plus de possibilités de l'intégrer. Par conséquent, ils maitriseront mieux cet élément qu'un élément qu'ils ont eu moins de temps pour apprendre (voir aussi Van Patten & Benati, 2010).

Quant aux connecteurs du discours, il y a des indications dans la littérature qui indiquent que les connecteurs fréquents pourraient être plus faciles à appendre que des connecteurs rares. Par exemple, de nombreuses études sur l'acquisition de la L1 s'accordent à dire que les jeunes, notamment à partir de l'adolescence, maitrisent les connecteurs fréquents mieux que les connecteurs plus rares (e.g., Nippold et al., 1992; Zufferey & Gygax, 2020a; Tskhovrebova et al., 2022). Si les connecteurs rares sont moins bien maitrisés même par les personnes natives, on peut supposer que ceux-ci sont également plus difficiles pour les personnes non-natives. Néanmoins, l'étude de Zufferey & Gygax (2017) démontre que la fréquence seule ne peut pas expliquer pourquoi les apprenant-e-s ont du mal à maitriser un connecteur. Dans leur étude, l'équipe de recherche a trouvé que les apprenant-e-s – malgré une compétence linguistique très élevée - maitrisaient mal le connecteur français *en effet,* un connecteur qui peut être considéré comme fréquent (p.ex., 923 occurrences dans les corpus de ORFEO, Benzitoun et al., 2013).

En conclusion, la mesure dans laquelle la fréquence peut être un facteur de prédiction de la maitrise des connecteurs de la L2 n'est pas établie de manière concluante. Ainsi, plus d'études sont nécessaires pour démontrer l'influence de la fréquence sur la maitrise des connecteurs en L2.

Il convient de noter à cet égard qu'un effet de fréquence pourrait s'appliquer également d'une autre manière, à savoir individuellement. Si une personne rencontre plus souvent un connecteur (par exemple parce qu'elle lit beaucoup), elle le maitrisera probablement mieux qu'une personne qui ne lit pas beaucoup et qui ne l'a donc pas souvent rencontré. Ainsi, outre la fréquence globale d'un connecteur, la fréquence avec laquelle une personne est exposée à sa L2 et ses connecteurs pourrait également résulter dans une meilleure maitrise de ces derniers. Si on veut examiner les facteurs qui facilitent la maitrise des connecteurs en L2, il convient de tenir compte de la variation individuelle des apprenant-e-s.

5.1.3 Quelles compétences prédisent une meilleure maitrise des connecteurs en L2 ?

Comme nous l'avons discuté au chapitre 3.5, il est probable qu'il existe des différences individuelles importantes qui pourraient expliquer et prédire la maitrise des connecteurs en L2. Dans ce qui suit, nous allons discuter deux facteurs dont nous supposons qu'ils prédisent une meilleure compétence avec les connecteurs en L2, à savoir la compétence linguistique et l'exposition à l'écrit.

Comme discuté au chapitre 3.5, il est fortement probable qu'une meilleure compétence linguistique corrèle avec une bonne maitrise des connecteurs en L2. Geva (1986, 1992), par exemple, a trouvé un lien entre une compétence linguistique élevée dans la langue cible et une meilleure connaissance explicite des connecteurs. La question qui se pose ici est de savoir *quel type* de compétence linguistique est nécessaire pour bien maitriser les connecteurs.

Considérons par exemple, le vocabulaire. Selon Sterpin, Ortiz, Formoso et Barreyro (2021), les connaissances de vocabulaire d'une langue se divisent en connaissances *profondes* et connaissances *larges*. On entend par une connaissance large le fait qu'une personne connaisse beaucoup de mots, alors qu'une connaissance profonde du vocabulaire se réfère au fait de connaitre leurs aspects pragmatiques, phonétiques, grammaticaux, syntaxiques, orthographiques et sémantiques. (Sterpin et al., 2021; voir aussi Li & Kirby, 2015). Comme les connecteurs sont en partie des indices lexicaux, une bonne compréhension du vocabulaire (profonde ou large) pourrait également prédire une meilleure maitrise des connecteurs. Et en effet, en testant les facteurs qui peuvent contribuer à une bonne compréhension de texte, Crosson et Lesaux (2013) ont constaté une corrélation entre la taille du vocabulaire et la connaissance des connecteurs des jeunes apprenant-e-s bilingues. Comme il s'agissait d'apprenant-e-s bilingues, la généralisation de ces résultats pour tous les apprenant-e-s de langue n'est toutefois pas garantie. Bien que les participant-e-s soient issu-e-s de familles hispanophones et que leurs compétences en anglais aient été testées, ils ont grandi aux États-Unis et fréquentaient des écoles anglophones. Il faut donc les considérer plutôt comme bilingues.

Au-delà du fait qu'ils sont des éléments lexicaux, les connecteurs ont également des fonctions de structuration du discours et parfois encodent parfois des contenus pragmatiques nuancés. Par conséquent, des compétences dans

d'autres domaines linguistiques pourraient également prédire une meilleure maitrise des connecteurs en L2. L'une d'entre elles pourrait être la compétence grammaticale. Zufferey & Gygax (2020b) ont constaté que la compétence grammaticale de leurs participant-e-s natifs était corrélée à une exposition à l'écrit plus élevée, ce qui prédisait à son tour une meilleure compétence avec les connecteurs. Ainsi, cette étude montre non seulement que les compétences grammaticales pourraient prédire la maitrise des connecteurs, mais aussi l'importance de l'exposition à la langue écrite.

En général, l'exposition à la langue écrite dans le contexte de l'apprentissage d'une langue est connue pour faciliter l'acquisition de nombreux éléments, prédisant par exemple une meilleure compréhension ou rédaction en L2 (p.ex., Sparks, Patten, Ganschow & Humbach, 2012). De manière générale, il a été prouvé qu'une exposition élevée à la langue cible facilite l'acquisition des éléments de cette langue (p.ex., Nippold et al., 1996, voir aussi des travaux concernant l'exposition hors de l'école: De Wilde, Brysbaert & Eyckmans, 2020, 2022). Ainsi, il est plausible que ce facteur soit également très pertinent pour l'acquisition des connecteurs en L2. Steffani et Nippold (1997), par exemple, ont pu montrer qu'une exposition élevée à la langue anglaise prédisait une meilleure maitrise des connecteurs anglais pour des apprenant-e-s japonais-e-s. Sur la base de leurs résultats, les chercheuses argumentent qu'une meilleure compétence avec les connecteurs pourrait être acquise si on « expose systématiquement les élèves à des mots spécifiques qui posent problème » (Steffani & Nippold, 1997 : 1053, ma traduction).

Concernant l'exposition à *l'écrit*, Zufferey & Gygax (2020a) ont également constaté que les participant-e-s qui faisaient preuve d'une exposition à l'écrit élevée se montraient plus capables de détecter une mauvaise utilisation d'un connecteur dans leur L1. Dans la même ligne que les résultats de Zufferey & Gygax (2020b), cela semble donc indiquer qu'une personne qui a beaucoup lu dans sa vie a pu développer grâce à ses lectures une certaine sensibilité pour la structuration d'un texte ainsi qu'une meilleure maitrise des connecteurs. Pour mesurer l'exposition à l'écrit, l'équipe de recherche a développé une version française de la tâche de reconnaissance d'auteur-e-s (ART, d'après Anglais *Author Recognition Task*), initialement développée pour l'anglais par Stanovich et West (1995). Dans cette tâche, les participant-e-s indiquent tous les noms d'auteur-e-s qu'ils reconnaissent parmi une liste donnée qui comprend des noms des personnes existantes ainsi que

des noms inventés. Afin d'éviter que les participant-e-s choisissent par désirabilité sociale (Cronbach, 1946) beaucoup plus des noms qu'ils ne connaissent en réalité, ils sont avertis que le choix d'un nom inventé résulte en une déduction de point.

En résumé, il est légitime de supposer que les apprenant-e-s diffèrent les uns des autres en ce qui concerne la maitrise des connecteurs dans une deuxième langue. Parmi les facteurs qui peuvent prédire si une personne maitrise bien un connecteur se trouve sa compétence linguistique sous plusieurs formes (p.ex., vocabulaire ou grammaire) ou encore son exposition à la langue écrite.

5.2 Tâche d'insertion de connecteurs

Dans cette étude[19], nous allons examiner les connaissances explicites des apprenant-e-s germanophones du français pour plusieurs connecteurs discursifs en français. Pour faire cela, nous allons utiliser une tâche d'insertion de connecteurs dans laquelle les participant-e-s doivent choisir pami plusieurs options le connecteur le plus approprié.

Cette expérience vise à répondre à trois questions. D'abord, nous allons examiner si certains connecteurs du français sont plus difficiles à utiliser que d'autres. Nous supposons que si tel est le cas, nous observerons des scores très variants entre les connecteurs.

Deuxièmement, nous allons investiguer si la fréquence du connecteur ainsi que la relation de cohérence qu'il indique ont une influence sur la façon dont il est maitrisé par les apprenant-e-s. Nous supposons qu'un connecteur fréquent est mieux maitrisé qu'un connecteur rare et les connecteurs qui indiquent une relation de cohérence plus complexe (telle qu'une concession) sont moins bien maitrisés que les connecteurs qui indiquent les relations de cohérence plus simples.

Troisièmement, nous allons voir si les différences individuelles quant aux compétences de grammaire et vocabulaire ainsi que l'exposition à l'écrit prédisent une bonne maitrise des connecteurs. Nous supposons qu'une personne ayant un vocabulaire large et des connaissances grammaticales élevées en L2 maitrisera mieux les connecteurs de la L2. Par ailleurs, on peut s'attendre à un effet similaire

[19] Les résultats de cette expérience ont été publiés sous une autre forme dans l'article «Second Language Acquisition and the Mastery of Discourse Connectives : Assessing the Factors that Hinder L2 – Learners from Mastering French Connectives », dans le journal *Languages* (Wetzel, Zufferey & Gygax, 2020).

pour l'exposition à l'écrit : plus qu'une personne a lu dans sa vie, plus elle sera capable de bien utiliser les connecteurs dans sa L2. Dans ce contexte, il ne devrait pas y avoir de différence si elle a beaucoup lu dans sa L1 ou sa L2, car dans les deux cas, il en résulte une meilleure sensibilité aux fonctions des connecteurs.

5.2.1 Pré-test

Dans l'expérience principale, nous allons tester des phrases qui contiennent différentes relations de cohérence. Pour cette raison, il n'est pas possible de présenter les phrases suivant un design de type carré latin, comme par exemple au chapitre 4. Cela signifie que chaque phrase testée peut représenter un biais si la relation de cohérence ne peut pas être insérée sans ambiguïté. Ainsi, on ne saura pas si un score bas pour un certain connecteur est dû au connecteur ou aux phrases utilisées. Pour éviter ce biais, il convenait donc de mener un pré-test afin de tester les phrases qui vont être utilisées lors de l'expérience.

Pour tester les phrases qui allaient être utilisées dans l'expérience principale, 50 participant-e-s de langue maternelle française ont été recrutés à l'aide de la plateforme *Prolific* (Prolific, London, 2020). Comme une participante n'a pas indiqué explicitement dans le questionnaire que le français était sa langue maternelle, ses réponses ont été exclues et les données de quarante-neuf participant-e-s ont été analysées (17 femmes, avec un âge moyen de 30.6 ans avec un écart-type de 11.4). Chaque participant-e a reçu une rémunération de 4.18 £ pour sa participation à cette expérience.

Design et procédure

Les participant-e-s de ce pré-test ont fait l'expérience qui va être présentée plus en détail à la section 5.2.3, à l'exception de quelques phrases testées comme nous allons voir dans un instant. En bref, les participant-e-s lisaient 60 phrases qui contenaient un blanc au lieu d'un connecteur et choisissaient le connecteur approprié parmi 6 options données.

Résultats descriptifs

En moyenne et pour chaque phrase, 39 participants sur 49 ont été capables d'évaluer la relation de manière appropriée et ont choisi le connecteur

correspondant. Cela signifie que le taux de réponses appropriées était en moyenne de 80% (voir Tableau 5.1).

Tableau 5.1. Pré-expérience, taux de réponses pour tous les connecteurs

Relation	Connecteur	*n* des réponses correctes	*n* des réponses incorrectes	%
Addition	*par ailleurs*	167	78	68
	en outre	179	66	73
Conséquence	*ainsi*	206	39	84
	c'est pourquoi	228	17	93
Concession	*cependant*	184	61	75
	néanmoins	170	75	69
Contraste	*par contre*	192	53	78
	en revanche	181	64	74
Condition	*dans le cas où*	180	65	74
	pourvu que	229	16	94
Cause	*car*	228	17	93
	puisque	208	37	85
			total	80

Quant aux phrases utilisées, il est à noter que quatre phrases ont suscité un taux de réponse déviant. Indépendamment de ces quatre phrases, chaque relation a été identifiée de manière appropriée par une moyenne de 40 participants sur 49 (écart-type = 5.2, taux de réponse appropriée : 82%). Les phrases problématiques sont les suivantes (entre parenthèses, le connecteur approprié) :

1) Camille adore aller au théâtre, ____ elle aime voir des vieux films au cinéma. (*en outre*)
2) Juliette est toujours prête à aider, ____ elle est très ouverte. (*par ailleurs*)
3) Emma boit beaucoup de thé noir, ____ elle adore le café. (*par ailleurs*)
4) Jacqueline parle bien plusieurs langues, ____ elle est traductrice (*puisque*)

La phrase (1) était censée évoquer une relation additive et le connecteur correspondant était *en outre*. Cependant, 10 personnes ont supposé que la relation entre les deux énoncés était causale et ont choisi le connecteur causal *car*. 11 personnes ont également supposé que la relation devait être concessive et ont choisi le connecteur *cependant*. Au total, sur l'ensemble des réponses données pour cet item, 24 réponses étaient donc ‘inappropriées’, ce qui fait que seulement 51% des participant-e-s ont donné la réponse ‘appropriée’. En conséquence, le taux de réponse pour cette phrase diffère significativement ($\chi^2(1) = 8.83$, $p < .005$) du taux de réponse moyen pour les quatre phrases de la même catégorie (i.e., une addition indiquée par *en outre*).

Pour (2), 26 des 49 participant-e-s ont supposé une relation causale entre le fait que Juliette soit toujours prête à aider et le fait qu'elle soit ouverte. 9 participant-e-s ont choisi un connecteur indiquant une relation de conséquence. Au total, 71% de l'ensemble du groupe de participant-e-s ont estimé que la relation devait être différente de l'addition qui était censée indiquer la cohérence pour cette phrase. Le taux de réponse pour cet item diffère ainsi également de manière significative ($\chi 2(1) = 33.61$, $p < .005$) du taux de réponse moyen pour les deux items non problématiques de la même catégorie (i.e., *addition* indiquée par *par ailleurs*).

Pour (3), 15 participant-e-s ont choisi un connecteur concessif, alors que la relation 'appropriée' était celle d'addition. Au total, cette phrase a reçu un score de 47% de réponses 'fausses'. Le taux de réponse de cette phrase diffère ainsi significativement ($\chi 2(1) = 12.98$, $p < .001$) du taux de réponse moyen de deux autres phrases non-problématiques de la même catégorie (*addition* indiquée par *par ailleurs*).

Pour (4), 19 participant-e-s ont cliqué sur un connecteur qui indiquait une relation de conséquence, 4 participant-e-s ont choisi le connecteur additif. Cependant, relation 'appropriée' a été supposée être une relation causale, indiquée par *puisque*. Au total, le pourcentage de réponses 'fausses' pour cette phrase était de 53%. Le taux de réponse pour cette phrase diffère ainsi significativement ($\chi 2(1) = 28.23$, $p > .0001$) du taux de réponse moyen pour les quatre autres phrases de la même catégorie (relation causale indiquée par *puisque*).

Modification des phrases ambiguës

Les résultats ont montré que parmi tous les phrases testées, quatre ont obtenu des scores significativement différents par rapport aux autres scores de la même relation et connecteur. Ces phrases étaient donc ambiguës dans leurs relations et, par conséquent, n'auraient pas été appropriés pour l'expérience principale. Les phrases (1) – (4) ont donc été modifiées en (5) – (8):

5) Camille adore aller au stade de foot, __ elle aime voir des vieux films au cinéma. (*en outre*)
6) Juliette est toujours prête à aider, __ elle est très intelligente. (*par ailleurs*)
7) Emma boit beaucoup de thé noir, __ elle adore le chocolat. (par ailleurs)

8) Jacqueline parle bien l'anglais, __ elle a grandi aux États-Unis. (*puisque*)

L'item (1) a été transformé en (5). Du fait qu'une causalité évidente entre les stades de foot et les vieux films au cinéma peut être écartée, un connecteur causal rendrait la phrase incohérente. Il en va de même pour (2) qui a été transformé en (6) : le simple fait d'être intelligent ne prédit rien sur la volonté à aider. Ainsi, les connecteurs de conséquence et de causalité représentent également des choix incohérents pour cet item. L'analyse a révélé en outre que plusieurs participant-e-s supposaient une relation concessive pour (3). Bien sûr, il est possible de boire beaucoup de thé noir et d'apprécier le café en même temps, mais comme il s'agit de deux boissons chaudes du matin, on peut également argumenter que si l'on préfère l'une, on ne boit pas l'autre. Dans ce cas, la relation en (3) serait effectivement concessive. Pour éviter cette ambiguïté, le mot « café » a été remplacé par le mot « chocolat » (7), la seule relation qui rende l'item cohérent est donc l'addition. Pour (4), selon le contexte, une relation de conséquence ou d'addition était possible alors que la relation appropriée était censée être causale. Afin que seul un connecteur causal rende la phrase cohérente, l'item a été remplacé par (8), qui serait incohérent si l'un des autres connecteurs proposés avait été sélectionné.

Comme le pré-test n'a pas montré d'autres problèmes en dehors de ces quatre phrases, l'expérience principale a été réalisée une fois que les quatre phrases ont été modifiées.

5.2.2 Les participant-e-s

Pour l'expérience principale, nous avons recruté 151 personnes germanophones et apprenant-e-s de français dans des gymnases suisses et allemands (l'âge moyen = 18 ans; écart type = 1.13, 100 femmes). Toutes les personnes de ce groupe ont été scolarisées dans un programme scolaire monolingue germanophone. Le groupe ne contenait pas des participant-e-s bilingues français-allemand.

Pour le groupe de contrôle, nous avons recruté 63 personnes natives (23.1 ans, écart type = 6.7, 40 femmes) sur la plateforme *Prolific* (n = 39), ainsi que parmi des étudiant-e-s au *Département de Psychologie* à l'Université de Fribourg (n = 24). Tandis que les participant-e-s de *Prolific* ont été rémunérés avec 3.15 GBP pour

leur participation, les étudiant-e-s de l'université de Fribourg ont reçu un crédit de cours.

Tous les participant-e-s ont été inclus-e-s dans l'expérience après avoir validé un formulaire de consentement.

5.2.3 Le design et la procédure

Pour tester si la maitrise des connecteurs dépend du type de connecteur, nous avons testé douze connecteurs, issus de six relations différentes (deux par relation de cohérence). Les variables dépendantes de cette expérience étaient la relation de cohérence et la fréquence du connecteur (la variable indépendante étant le score pour chaque connecteur).

Pour observer si une éventuelle différence pourrait s'observer du fait de la différente de complexité entre relations de cohérence, nous avons testé des relations de cohérence qui différeraient en termes de complexité cognitive. Nous avons testé les relations suivantes, par ordre de complexité :

addition < *cause* < *conséquence* < *contraste* < *condition* < *concession*

Cette classification des relations de cohérence est toutefois plutôt approximative, car elle a été établie en considérant la littérature de référence (notamment : Sanders et al., 1992; Murray, 1997; Sanders, 2005; Morera et al., 2017).

Pour chacune de ces relations, nous avons testé deux connecteurs : un connecteur fréquent et un qui était moins fréquent. Cela nous a permis de comparer directement les connecteurs fréquents et rares ainsi que d'observer une potentielle interaction avec la relation de cohérence. Nous avons établi la fréquence à l'aide du corpus *FrenchWeb 2017* (*frTenTen17*, (Jakubíček, Kilgarriff, Kovář, Rychlý & Suchomel, 2013), consulté via *SketchEngine* (Kilgarriff, Baisa, Jakubíček, Kovář, Michelfeit, Rychlý & Suchomel, 2014). Ainsi, nous avons testé les douze connecteurs inclus dans le Tableau 5.2.

Tableau 5.2. Variables testées dans la tâche principale, complexité de la relation supposée indiquée avec des croix (plus de croix, plus de complexité)

	Relation	Complexité de la relation	Fréquence	Connecteur testé
1	addition	-	Haute	*par ailleurs*
2	addition	-	Basse	*en outre*
3	cause	+	Haute	*car*
4	cause	+	Basse	*puisque*
5	conséquence	++	Haute	*ainsi*
6	conséquence	++	Basse	*c'est pourquoi*
7	contraste	+++	Haute	*par contre*
8	contraste	+++	Basse	*en revanche*
9	condition	++++	Haute	*dans le cas ou*
10	condition	++++	Basse	*pourvu que*
11	concession	+++++	Haute	*cependant*
12	concession	+++++	Basse	*néanmoins*

Pour mesurer la maitrise de ces douze connecteurs, nous avons conçu une tâche d'insertion sans contrainte de temps, ce qui permettait aux participant-e-s d'accéder à leurs connaissances explicites. Dans cette tâche, des phrases à trous étaient présentées et les participant-e-s devaient les remplir en choisissant le connecteur approprié. Comme illustré en (9), ils ont choisi le connecteur approprié parmi une liste de six options. L'ordre des options était présenté de manière aléatoire.

9) Jacqueline parle bien l'anglais ____ elle a grandi aux Etats-Unis.

- *c'est pourquoi*
- *par contre*
- *puisqu'*
- *néanmoins*
- *par ailleurs*
- *dans le cas où*

Les options inappropriées pour chaque phrase expérimentale étaient toujours les autres connecteurs testés. Les autres options étaient toujours incorrectes. Pour chaque connecteur nous avons utilisé cinq phrases, ce qui donnait un total de 60 phrases à évaluer (5 phrases par connecteur multipliées par 12 connecteurs). Ces phrases ont été présentées de manière aléatoire. Pour permettre aux gens d'accéder à leurs connaissances explicites, aucune limite de temps a été imposée.

L'expérience a été programmée à l'aide du site *Qualtrics* (Qualtrics LLC, Provo [UT], Etats-Unis) et les participant-e-s y ont accédé via un lien Internet. Après avoir complété la tâche principale (c'est-à-dire la tâche d'insertion de connecteur), les participant-e-s ont fait les tâches supplémentaires qui visaient à mesurer la taille de leur vocabulaire, leur compétences grammaticales et leur exposition à l'écrit.

Pour mesurer la taille du vocabulaire, nous avons utilisé la version française de le tâche *Lextale* (Brysbaert, 2013). Dans cette tâche, les participant-e-s indiquent les mots français qu'ils reconnaissent parmi une liste donnée. La liste contient d'une part des mots réels du français et d'autre part des mots inventés, ce qui évite que les participant-e-s indiquent reconnaitre tous les mots. De plus, ils ont été avertis que le fait de cliquer sur un mot inventé entrainerait une déduction de point. Au total, la liste comportait 64 vrais mots existants en français, le score maximum était donc de 64.

Pour mesurer la compétence grammaticale, nous avons utilisé la tâche de grammaire de Zufferey et Gygax (2020a). Dans cette tâche, les participant-e-s évaluent 40 phrases françaises dont la moitié contient des erreurs typiques de l'écrit. Les participant-e-s évaluent ces phrases en déplaçant un curseur d'une échelle allant de « Je suis sûr-e que c'est correct » à « Je suis sûr-e que c'est incorrect ». Le positionnement du curseur est ensuite transformé en valeur numérique de 0 à 100. Par exemple, si un-e participant-e positionne le curseur à « Je suis sûr-e que c'est incorrect » pour une phrase incorrecte, 100 points sont attribués. Ainsi, le score maximal pour cette tâche est de 100 (cela signifie que toutes les phrases correctes et incorrectes étaient identifiées comme telles et cela sans aucun doute).

Ensuite, pour mesurer l'exposition à l'écrit, nous avons utilisé deux versions d'une tâche de reconnaissance d'auteur-e-s (ou *ART* ci-après, après angl. *author recogntion task*). La première visait à mesurer l'exposition à l'écrit en allemand (Grolig, Tiffin-Richards & Schroeder, 2020), tandis que la deuxième mesurait l'exposition à l'écrit en français (Zufferey & Gygax, 2020b). Dans ces tâches, une liste d'auteur-trice-s est présentée et les participant-e-s doivent indiquer tous les noms qu'ils connaissent. Comme pour la tâche de *Lextale*, la liste contient, à part les noms existants, également des noms inventés. Cela évite également que les participant-e-s n'indiquent beaucoup plus de noms pour des raisons de désirabilité sociale. Au total, les listes contiennent 50 (version allemande) et 56 (version françaises) noms existants, ce qui représente en conséquence les scores maximums.

Après avoir fait ces trois tâches, les participant-e-s répondaient à des questionnaires démographiques. Cette partie contenait également des questions sur leur propre histoire linguistique. Entre autres, les participant-e-s devaient

évaluer l'importance de sept facteurs sur une échelle de 0 à 10 pour leur propre acquisition de français. Les facteurs étaient : *école, amis, famille, travail, lecture, musique* et *télévision*.

5.2.4 Les résultats et l'analyse

Nous avons fait les analyses en trois étapes, en suivant nos trois questions de recherche. D'abord, nous avons examiné les différences entre les connecteurs, indépendamment de la relation ou de la fréquence. Ensuite, nous avons analysé l'influence de nos facteurs, à savoir la relation de cohérence et la fréquence du connecteur. Finalement, nous avons analysé la variation individuelle et examiné si les scores des mesures de variations individuelles allaient prédire les scores obtenus dans la tâche principale.

Nous avons fait des modèles logistiques généralisés qui ont été montés selon la procédure de Baayen (2008) à l'aide de *R* (R Core Team 2020). Cela veut dire que nous avons ajouté successivement les effets fixes et avons comparé chaque fois le modèle résultant avec le modèle qui ne contenait pas cet effet fixe. Pour tester l'amélioration lors de l'ajout d'un effet fixe (i.e., pour calculer la valeur $\chi 2$ du test du rapport de vraisemblance) nous avons utilisé la fonction *anova()* de *base-R* (R Core Team 2020). Pour monter et exécuter les modèles, nous avons utilisé la fonction *glmer()* du package *lm4* (Bates et al., 2015). Afin d'obtenir les valeurs *p* et les tailles d'effet estimées, nous avons utilisé la fonction *summary()* du package *car* (Fox & Weisberg, 2019). Pour les tests *post-hoc*, nous avons utilisé la fonction *lsmeans()* du package *emmeans* (Lenth, 2020). Tous les modèles contenaient toujours les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires.

L'analyse des données de la tâche principale

Dans un premier temps, il convenait d'analyser si les scores obtenus dépendaient des deux groupes testés (*natifs* vs. *non-natifs*) ainsi que des connecteurs différents.

En ajoutant le *Groupe de langue* (*natif*, *non-natif*) comme effet fixe, le modèle s'est amélioré de manière significative ($\chi(1) = 113.6$, $p < 2.2^{e-16}$). En ajoutant l'interaction avec le *Type de connecteur* (indépendamment de la relation qu'il

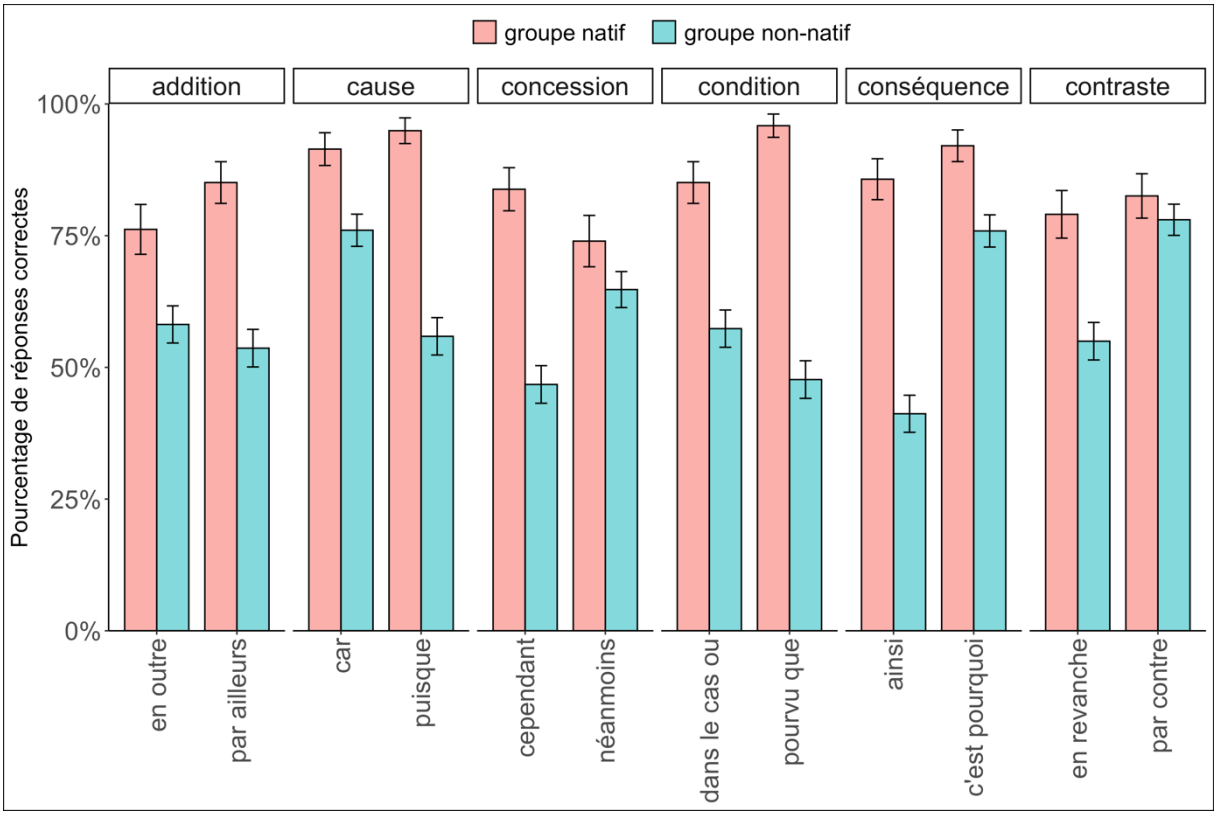
indique ou sa fréquence), le modèle s'est amélioré à nouveau ($\chi(22) = 302.54$, $p < 2.2^{e-16}$). L'output du modèle final est reporté dans le Tableau 5.3

Tableau 5.3. Output du modèle final (contenant l'interaction du *Groupe de langue* et *Type de connecteur* ainsi que les *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires)

	β	*ET*	*z*	*Pr(>|z|)*
(Intercepte)	2.08	.27	7.84	< .0001
Groupe (*non-natif*)	-2.52	.23	-11.03	< 2.00^{e-16}
Connecteur (*c'est pourquoi*)	.74	.36	2.02	.04
Connecteur (*car*)	.59	.36	1.65	.10
Connecteur (*cependant*)	-.16	.34	-.46	.64
Connecteur (*dans le cas où*)	-.01	.34	-.04	.97
Connecteur (*en outre*)	-.68	.33	-2.08	.04
Connecteur (*en revanche*)	-.54	.33	-1.64	.10
Connecteur (*néanmoins*)	-.85	.33	-2.60	.01
Connecteur (*par ailleurs*)	-.06	.34	-.17	.87
Connecteur (*par contre*)	-.28	.33	-.83	.41
Connecteur (*pourvu que*)	1.46	.41	3.55	< .0001
Connecteur (*puisque*)	1.18	.39	3.01	< .0001
Groupe (*non-natif*) : Connecteur (*c'est pourquoi*)	1.07	.30	3.59	< .0001
Groupe (*non-natif*) : Connecteur (*car*)	1.19	.29	4.10	< .0001
Groupe (*non-natif*) : Connecteur (*cependant*)	.42	.26	1.64	.10
Groupe (*non-natif*) : Connecteur (*dans le cas où*)	.80	.26	3.06	< .0001
Groupe (*non-natif*) : Connecteur (*en outre*)	1.51	.25	6.09	< .0001
Groupe (*non-natif*) : Connecteur (*en revanche*)	1.20	.25	4.82	< .0001
Groupe (*non-natif*) : Connecteur (*néanmoins*)	1.99	.25	8.14	< .0001
Groupe (*non-natif*) : Connecteur (*par ailleurs*)	.65	.26	2.51	.01
Groupe (*non-natif*) : Connecteur (*par contre*)	2.20	.26	8.47	< 2e-16
Groupe (*non-natif*) : Connecteur (*pourvu que*)	-1.15	.35	-3.28	< .0001
Groupe (*non-natif*) : Connecteur (*puisque*)	-.48	.33	-1.46	.14

Comme visuellement représenté à la Figure 5.1 on constate des différences importantes entre les deux groupes linguistiques. Les comparaisons *post-hoc* montrent que le groupe non-natif a scoré significativement moins haut que le groupe natif pour tous les connecteurs, à l'exception des connecteurs *par contre* ($\beta(1.46) = .33$, $ET = .22$, $p > .99$) et *néanmoins* ($\beta(2.56) = .53$, $ET = .21$, $p = .63$) pour lesquels les personnes non-natives ont scoré à un niveau natif.

Figure 5.1. Pourcentage de réponses correctes des personnes natives et non-natives pour tous les connecteurs testés. Les intervalles des confiances (95%) sont indiqués par des barres d'erreur.



On remarque également que le groupe non-natif a obtenu un score différent en fonction du connecteur. Toutes les différences significatives entre les scores de connecteurs sont reportées dans le Tableau 5.4.

Tableau 5.4. Comparaisons *post-hoc* des scores du groupe non-natif pour les connecteurs (seulement résultats significatifs avec α < .05 reportés)

Comparaison			*β*	*ET*	*z*	*Pr(>|z|)*
ainsi	-	*c'est pourquoi*	-1.80	.27	-6.59	< .0001
ainsi	-	*car*	-1.78	.27	-6.52	< .0001
ainsi	-	*néanmoins*	-1.15	.27	-4.24	< .01
ainsi	-	*par contre*	-1.92	.27	-7.00	< .0001
c'est pourquoi	-	*cependant*	1.54	.27	5.62	< .0001
c'est pourquoi	-	*dans le cas où*	1.01	.27	3.70	< .05
c'est pourquoi	-	*en revanche*	1.15	.27	4.20	< .05
c'est pourquoi	-	*pourvu que*	1.49	.27	5.46	< .0001
car	-	*cependant*	1.51	.27	5.54	< .0001
car	-	*en revanche*	1.12	.27	4.12	< .05
car	-	*par ailleurs*	1.18	.27	4.34	< .01
car	-	*pourvu que*	1.47	.27	5.38	< .0001
car	-	*puisque*	1.08	.27	3.96	< .05
cependant	-	*par contre*	-1.65	.27	-6.03	< .0001
dans le cas ou	-	*par contre*	-1.13	.28	-4.12	< .01
en outre	-	*par contre*	-1.10	.28	-4.00	< .05
par ailleurs	-	*par contre*	-1.32	.27	-4.83	< .0001
en revanche		*par contre*	-1.27	.27	-4.62	< .001
par contre	-	*pourvu que*	1.61	.27	5.87	< .0001
par contre	-	*puisque*	1.22	.27	4.45	< .01

Quant au groupe natif, on remarque également des différences entre les scores pour les différents connecteurs. Toutes les différences significatives entre les connecteurs sont reportées dans le Tableau 5.5.

Tableau 5.5. Comparaisons *post-hoc* des scores du groupe natif pour les connecteurs (seulement résultats significatifs avec α < .05 reportés)

Comparaison			*β*	*ET*	*z*	*Pr(>|z|)*
c'est pourquoi	-	*en outre*	1.42	.35	4.01	< .05
c'est pourquoi	-	*néanmoins*	1.58	.35	4.51	< .0001
car	-	*néanmoins*	1.44	.35	4.16	< .05
cependant	-	*pourvu que*	-1.62	.41	-3.95	< .05
en outre	-	*pourvu que*	-2.14	.40	-5.33	< .0001
en outre	-	*puisque*	-1.86	.38	-4.87	< .0001
en revanche	-	*puisque*	-1.72	.38	-4.48	< .0001
néanmoins	-	*pourvu que*	-2.31	.40	-5.76	< .0001
néanmoins	-	*puisque*	-2.03	.38	-5.32	< .0001
par ailleurs	-	*pourvu que*	-1.52	.41	-3.69	< .05
par contre	-	*pourvu que*	-1.74	.41	-4.27	< .0001
par contre	-	*puisque*	-1.46	.39	-3.76	< .03

Ces résultats montrent donc que les connecteurs ont engendré des scores très différents, à la fois pour le groupe natif et le groupe non-natif.

La relation de cohérence et la fréquence – des facteurs explicatifs ?

Comment peut-on expliquer ces différences de scores entre les connecteurs? Afin de répondre à cette question, nous procédons avec l'analyse des facteurs qui pourraient prédire les résultats observés, à savoir le type de *Relation de cohérence* (*cause*, *concession*, *condition*, *conséquence*, *contraste* ou *addition*) et la *Fréquence* (*haute* ou *basse*) du connecteur.

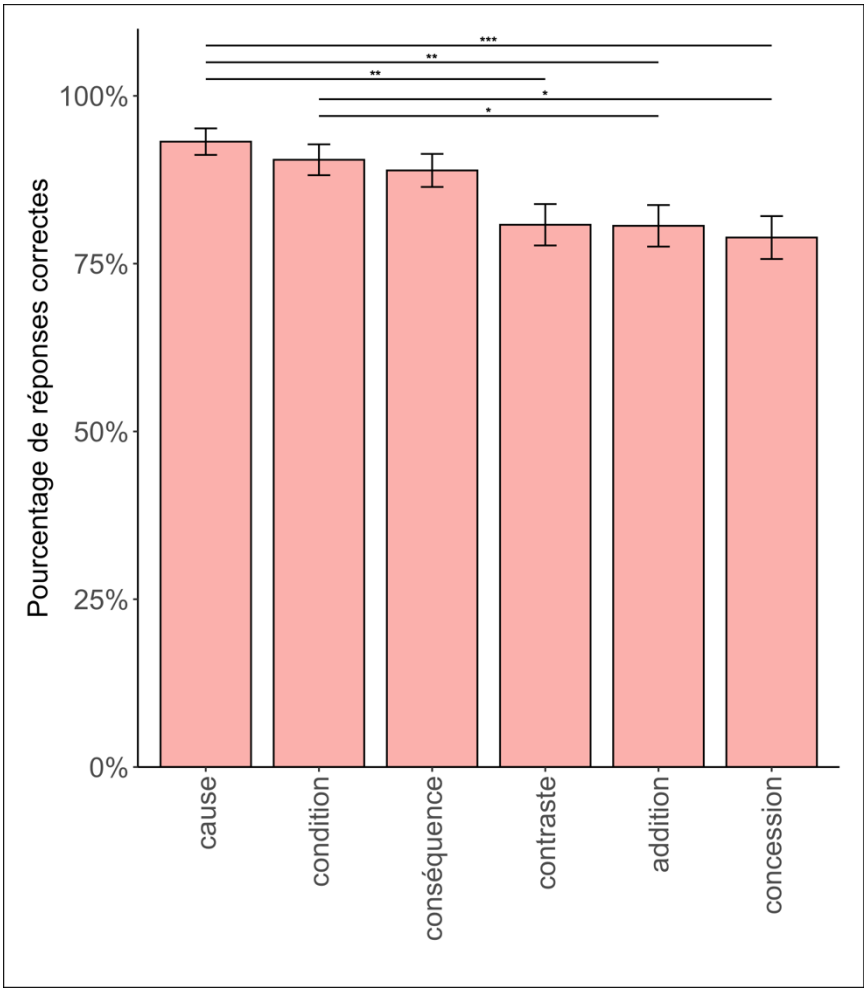
Après avoir ajouté le *Groupe de langue* des participant-e-s (pour rappel, amélioration après avoir ajouté le groupe linguistique comme effet fixe : $\chi(1) = 113.6$, $p < 2.2^{e-16}$). L'ajout successif de l'interaction avec la *Relation de cohérence* a amélioré le modèle de manière significative ($\chi(10) = 115.13$, $p < 2.2^{e-16}$). L'ajout de l'interaction avec la *Fréquence* a amélioré le modèle davantage ($\chi(12) = 187.4$, $p < 2.2^{e-16}$). L'output du modèle résultant - contenant donc la triple interaction entre la *Langue*, la *Relation de cohérence* et la *Fréquence* ainsi que les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires - est reporté dans le Tableau 5.6.

Tableau 5.6. Output du modèle final (contenant l'interaction entre le *Groupe de langue*, la *Relation de cohérence* et la *Fréquence,* ainsi que les *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires)

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	1.40	.25	5.61	< .0001
Groupe (*non-natif*)	-1.02	.21	-4.86	< .0001
Relation (*cause*)	1.86	.38	4.88	< .0001
Relation (*concession*)	-.16	.31	-.53	.60
Relation (*condition*)	2.14	.40	5.34	< .0001
Relation (*conséquence*)	1.42	.35	4.02	< .0001
Relation (*contraste*)	.14	.32	.45	.66
Fréquence (*haute*)	.63	.33	1.92	.06
Groupe (*non-natif*) : Relation (*cause*)	-1.98	.31	-6.30	< .0001
Groupe (*non-natif*) : Relation (*concession*)	.49	.23	2.14	.03
Groupe (non-natif) : Relation (*condition*)	-2.66	.34	-7.84	< .0001
Groupe (non-natif) : Relation (*conséquence*)	-.44	.28	-1.56	.12
Groupe (non-natif) : Relation (*contraste*)	-.31	.23	-1.34	.18
Groupe (non-natif) : Fréquence (*haute*)	-.85	.24	-3.49	< .001
Relation (*cause*) : Fréquence (*haute*)	-1.21	.52	-2.33	.02
Relation (*concession*) : Fréquence (*haute*)	.07	.46	.15	.89
Relation (*condition*) : Fréquence (*haute*)	-2.10	.52	-4.00	< .0001
Relation (*conséquence*) : Fréquence (*haute*)	-1.36	.49	-2.79	<.01
Relation (*contraste*) : Fréquence (*haute*)	-.36	.46	-.79	.43
Groupe (*non-natif*) : Relation (*cause*) : Fréquence (*haute*)	2.52	.43	5.92	< .0001
Groupe (*non-natif*) : Relation (*concession*) : Fréquence (*haute*)	-.72	.34	-2.10	<.05
Groupe (*non-natif*) : Relation (*condition*) : Fréquence (*haute*)	2.80	.43	6.58	< .0001
Groupe (*non-natif*) : Relation (*conséquence*) : Fréquence (*haute*)	-.21	.38	-.56	.58
Groupe (*non-natif*) : Relation (*contraste*) : Fréquence (*haute*)	1.85	.34	5.37	< .0001

Pour le groupe non-natif, les comparaisons *post-hoc* n'indiquent pas des différences significatives en fonction de la relation de cohérence. En revanche, pour le groupe natif, la relation de cohérence était un facteur prédictif. Nous observons des différences entre les scores pour la *cause* et la *concession* (β = 1.43, *ET* = .33, z = 4.38 p < .001), entre la *cause* et l'*addition* (β = 1.30, *ET* = .33, z = −3.97, p < .005), entre la *cause* et le *contraste* (β = 1.27, ET = .33, z = 3.88, p < .01), entre la *concession* et la *condition* (β −1.06, *ET* = .32, z = −3.36, p < .05) ainsi qu'entre la *concession* et la *conséquence* (β = −1.03, *ET* = .32, z = −3.27, p < .05). Ces résultats sont illustrés par la Figure 5.2.

Figure 5.2. Les pourcentages de réponses correctes selon les relations de cohérence (groupe natif). Les intervalles des confiances (95%) sont indiqués par des barres d'erreur.

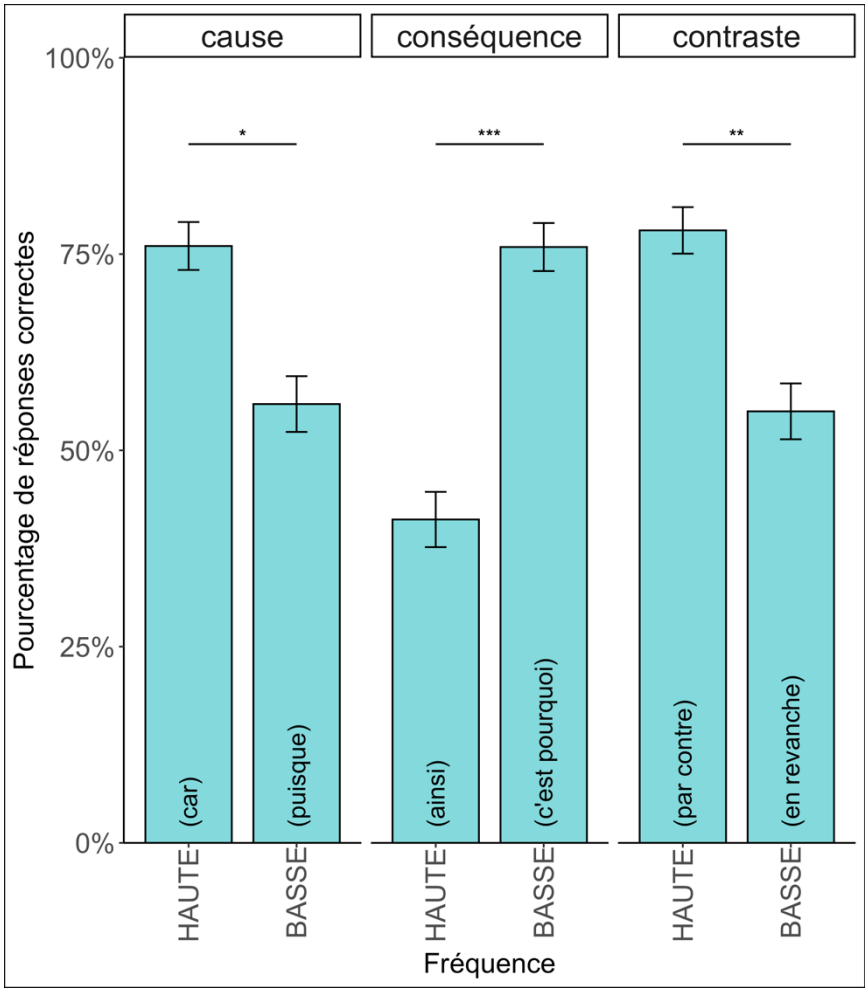


Nous constatons donc que la relation de cohérence n'était un facteur prédictif pour la maitrise des connecteurs que pour le groupe natif. Pour ce groupe, une *cause* et une *condition* ont notamment suscité des scores hauts, alors qu'une *concession* et une *addition* ont engendré des scores bas.

Quant au facteur de la fréquence du connecteur, les réponses du groupe natif ne montraient aucune différence significative en fonction de la fréquence du connecteur. Pour le groupe non-natif, cet effet n'était pas non plus très important. Les comparaisons *post-hoc* pour le groupe non-natif ont uniquement indiqué des différences au sein de la relation de *cause* (*car* vs. *puisque*; β = -1.08, *ET* = .27, *z* = -3.96, $p < .05$), de la relation *contraste* (*par contre* vs. *en revanche* β = -1.27, *ET* = .27, *z* = -4.62, $p < .001$) ainsi que de la relation de *conséquence* (*c'est pourquoi* vs. *ainsi*; β = 1.80, *ET* = .27, *z* = 6.59, $p < .0001$). En plus, les effets de fréquence (si ces différences les reflètent) ne semblaient pas être réguliers: tandis que les

connecteurs fréquents étaient mieux utilisés au sein des relations causales et contrastives, nous observons l'effet inverse pour la relation de conséquence, dans la mesure où le connecteur rare *c'est pourquoi* a suscité plus de réponses correctes que le connecteur plus fréquent *ainsi*. Ces différences sont illustrées par la Figure 5.3.

Figure 5.3. Les différences significatives pour les scores dans la tâche principale selon la fréquence du connecteur (groupe non-natif)



La variation interpersonnelle : les scores des tâches supplémentaires

Concernant les tâches supplémentaires, leurs résultats ont été ajoutés dans des modèles séparés pour chaque groupe testé. Cela permettait de respecter l'absence d'autocorrélation entre les effets fixes, c'est-à-dire le lien, par exemple, entre les scores obtenus dans les mesures de compétences linguistiques et le fait d'être une

personne native ou non-native (Schreiber-Gregory, 2018). Les corrélations entre les différentes mesures sont reportées et illustrées par la Figure 5.4.

Figure 5.4. Les corrélations entre les mesures supplémentaires (toutes les corrélations avec $p < .0001$)

Pour ce qui est des scores à la tâche de vocabulaire *Lextale*, nous observons une différence considérable entre les deux groupes. Les personnes natives ont obtenu en moyenne 81.01% (95% IC [80.71; 81.31]). En revanche, les personnes non-natives ont obtenu un score moyen de 24.76% (95% IC [24.46; 25.07]). Pour les personnes non-natives, l'ajout des *Scores obtenus dans la tâche de Lextale* comme effet fixe a significativement amélioré le modèle ($\chi(1) = 29.32$, $p = 6.13^{e-08}$). L'output de ce modèle est reporté dans le Tableau 5.7.

Tableau 5.7. Output du modèle final pour les scores obtenus dans la tâche *Lextale* (personnes non-natives)

	β	ET	z	*Pr(>|z|)*
(Intercept)	-0.19	.16	-1.17	.24
Score Lextale	.05	.01	5.67	1.41^{e-08}

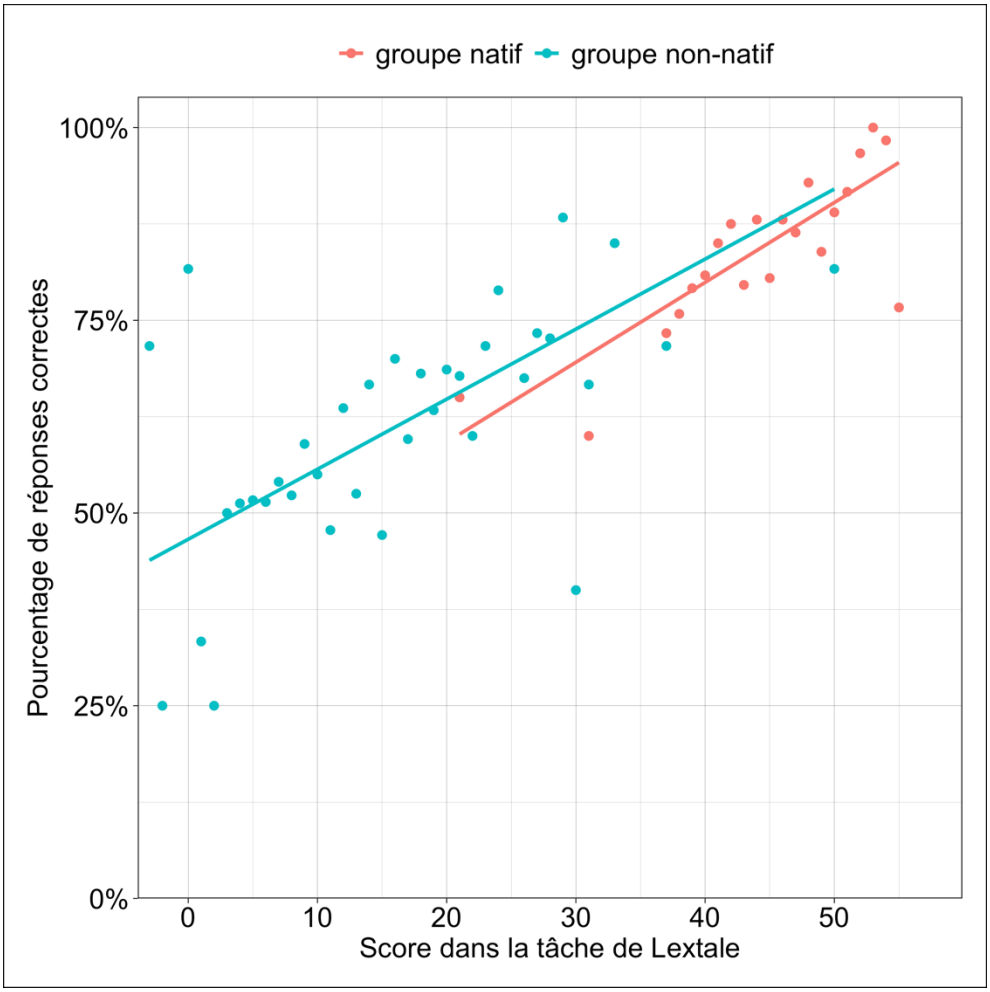
Pour les personnes natives, l'ajout des *Scores obtenus dans la tâche de Lextale* comme effet fixe a également amélioré le modèle ($\chi(1) = 21.08$, $p = 4.42^{e-06}$). L'output de ce modèle est reporté dans le Tableau 5.8.

Tableau 5.8. Output du modèle final pour les scores obtenus dans la tâche *Lextale* (personnes natives)

	β	ET	z	*Pr(>|z|)*
(Intercept)	-2.00	.86	-2.33	.02
Score *Lextale*	.09	.02	4.87	1.09^{e-06}

Nous concluons donc que les scores de la tâche Lextale prédisaient pour les deux groupes linguistiques les scores obtenus dans la tâche principale. Ces effets sont visualisés dans la Figure 5.5.

Figure 5.5. Pourcentage des réponses correctes dans la tâche principale et les scores obtenus dans la tâche Lextale



Concernant la mesure de compétences grammaticales, l'ajout des *Scores obtenus dans la tâche de grammaire* comme effet fixe a amélioré le modèle des personnes non-natives ($\chi(1) = 19.51$, $p = 9.99^{e-06}$). L'output de ce modèle est reporté dans le Tableau 5.9.

Tableau 5.9. Output du modèle final pour les scores obtenus dans *la tâche de grammaire* (personnes non-natives)

	β	ET	z	*Pr(>|z|)*
(Intercept)	-1.76	.95	-1.85	.06
Score *Tâche de grammaire*	.06	.01	4.16	3.20^{e-05}

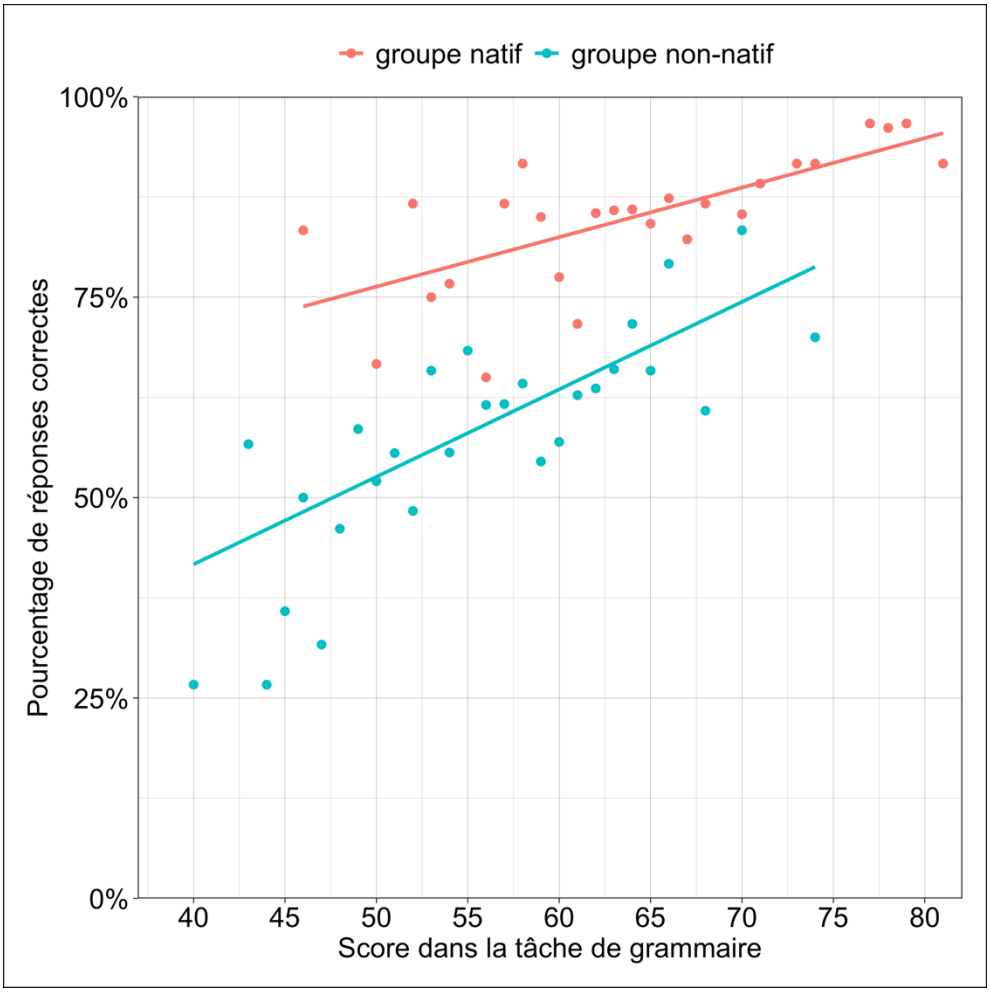
L'ajout des scores de la tâche de grammaire comme effet fixe ont également amélioré le modèle de manière significative pour le groupe natif ($\chi(1) = 15.34$, $p = 9.00^{e-05}$). L'output de ce modèle est reporté dans le Tableau 5.10.

Tableau 5.10. Output du modèle final pour les scores obtenus dans la tâche de grammaire (personnes natives)

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	-2.65	.69	-3.83	< .0005
Score *Tâche de grammaire*	.06	.01	4.55	5.28^{e-06}

Nous concluons donc que des meilleurs scores obtenus dans la tâche de grammaire étaient prédictifs d'un meilleur score dans la tâche principale et cela pour les deux groupes linguistiques (*natif* ou *non-natif*). Ces résultats sont illustrés par la Figure 5.6.

Figure 5.6. Pourcentage des réponses correctes dans la tâche principale et les scores obtenus dans la tâche de grammaire



Concernant l'exposition à l'écrit, l'ajout des scores du *ART (version française)* comme effet fixe a amélioré le modèle de manière significative pour le groupe non-natif ($\chi(1) = 7.25$, $p < .01$). L'output de ce modèle est reporté dans le Tableau 5.11.

Tableau 5.11. Output du modèle final pour les scores obtenus dans la tâche de *ART* (version française, personnes non-natives)

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	.26	.14	1.90	.06
Score *ART français*	.07	.03	2.72	< .05

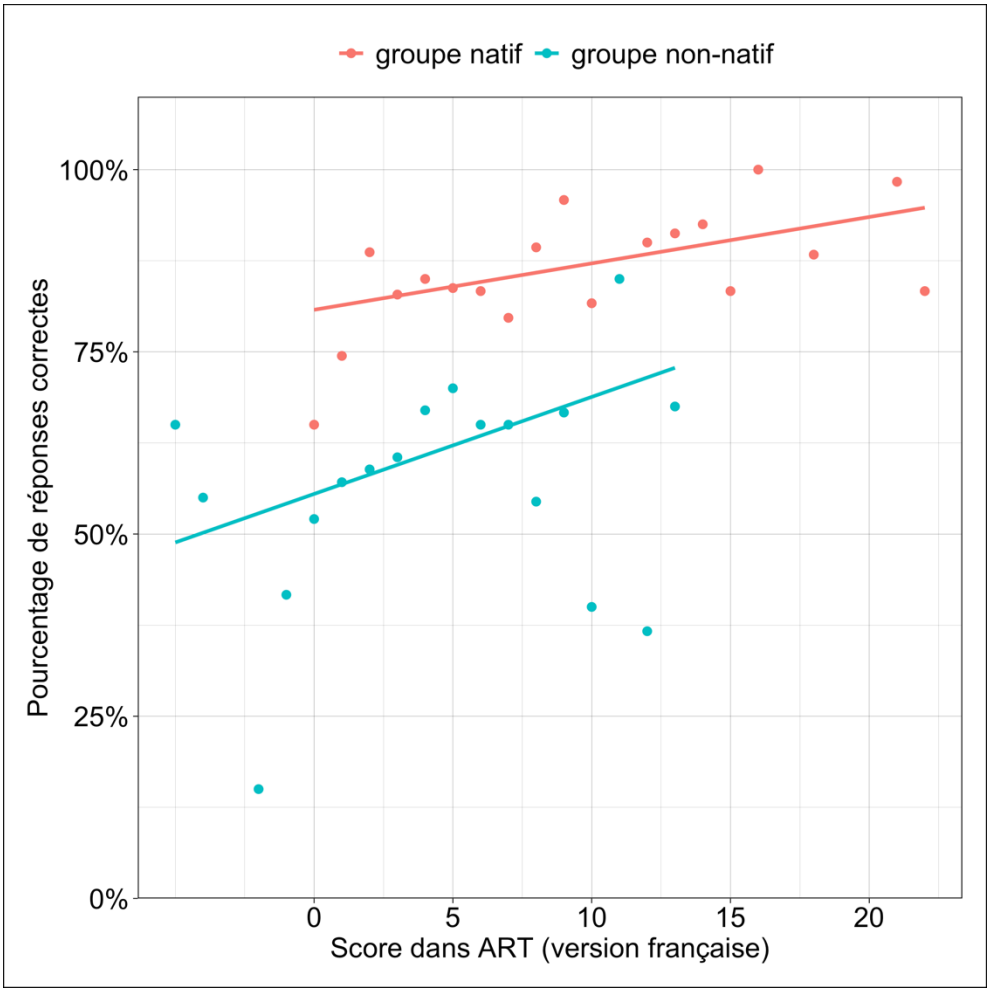
Pour le group natif, l'ajout des scores du *ART (version française)* a également amélioré le modèle ($\chi(1) = 6.61$, $p < .05$). Le modèle résultant est reporté dans le Tableau 5.12.

Tableau 5.12. Output du modèle final pour les scores obtenus dans la tâche de *ART* (version française, personnes natives)

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	1.73	.23	7.48	7.64^{e-14}
Score *ART français*	.06	.02	2.64	< .01

Nous observons donc qu'un meilleur score dans la tâche de *ART (version française)* prédisait un meilleur score dans la tâche principale, et cela pour les deux groupes linguistiques. Ces résultats sont illustrés par la Figure 5.7.

Figure 5.7. Pourcentage des réponses correctes dans la tâche principale et les scores obtenus dans la tâche de *ART (version française)*



La pertinence de ces effets doit toutefois être considérée avec prudence. D'abord, nous voyons que l'effet prédictif est plutôt subtil pour les personnes natives. De plus, les scores du *ART* (version française) du groupe non-natif étaient extrêmement bas. En moyenne, ils ont obtenu un score de seulement 2.78 (écart type = 2.89, médiane = 2), ce qui correspond à une reconnaissance de 5% des auteur-e-s existant-e-s. Ceci, ainsi que la présence de nombreuses données aberrantes (comme on peut le voir sur la Figure 5.7), doit être pris en compte lors de l'interprétation.

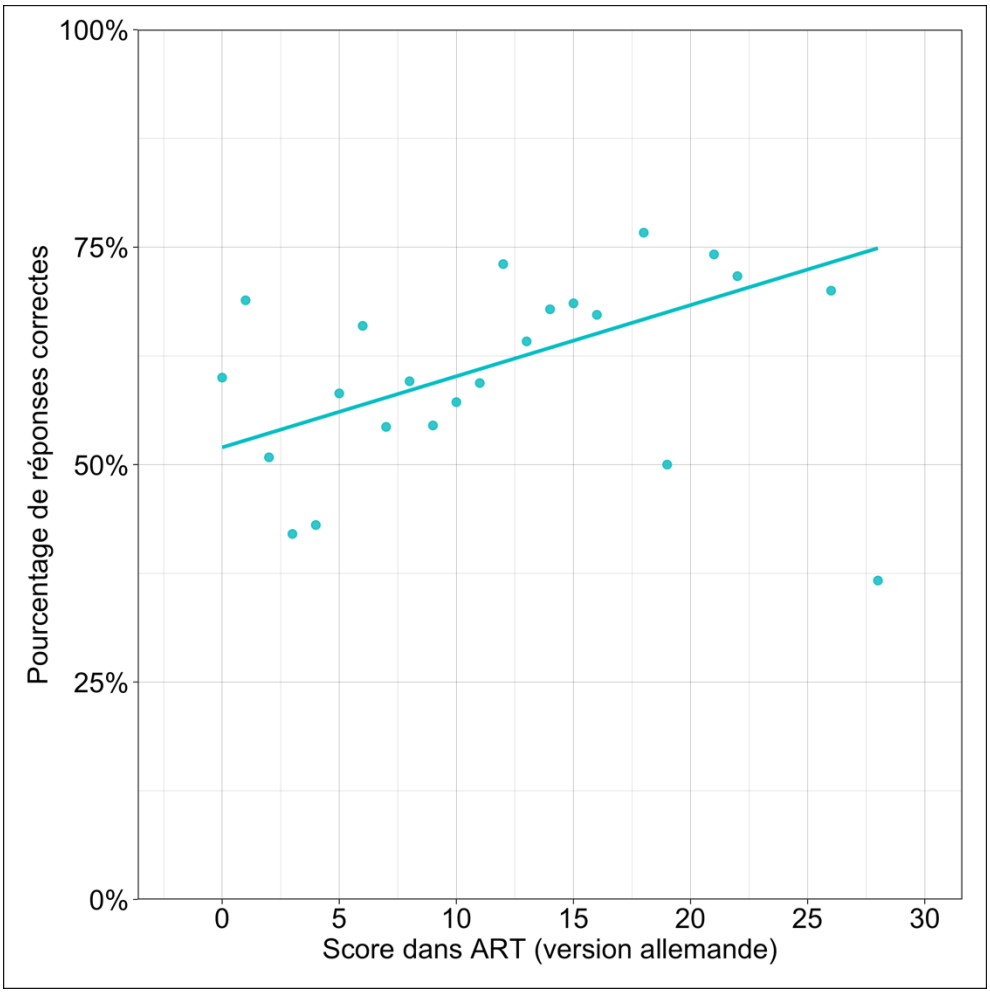
Pour les scores obtenus dans la version allemande de cette tâche, leur ajout comme effet fixe a amélioré de manière significative ($\chi(1) = 9.73$, $p = < .005$) le modèle pour les personnes non-natives (les personnes natives n'ont pas fait cette tâche). Le modèle résultant est reporté dans le Tableau 5.13.

Tableau 5.13. Output du modèle final pour les scores obtenus dans la tâche de *ART* (version allemande, personnes non-natives)

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	.09	.17	.51	.61
Score *ART allemand*	.04	.01	3.17	< .005

Nous concluons donc qu'un meilleur score dans la version allemande de *ART* prédisait un meilleur score dans la tâche principale, la tâche d'insertion de connecteurs français. Ces résultats sont illustrés par la Figure 5.8.

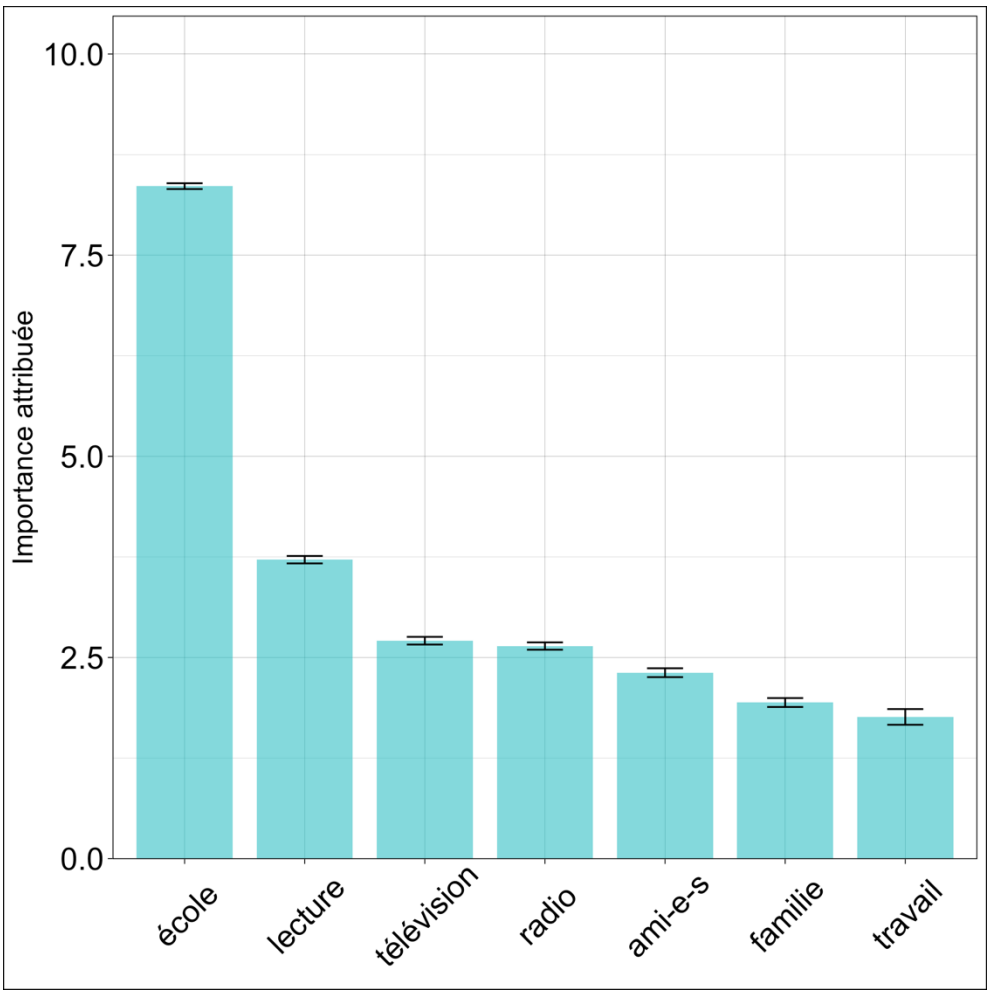
Figure 5.8. Pourcentage des réponses correctes dans la tâche principale et les scores obtenus dans la tâche de ART (version allemande, seulement des personnes non-natives)



Évaluation de l'importance des facteurs pour l'apprentissage de langue

Comme visuellement représenté dans la Figure 5.9, les personnes non-natives ont attribué la plus grande importance pour leur propre acquisition du français au facteur de *l'école*.

Figure 5.9. Évaluation subjective des participant-e-s de l'importance de sept facteurs pour leur propre acquisition du français (10 correspond à l'importance la plus grande, les intervalles de confiance [95%] sont indiqués par des barres d'erreur)



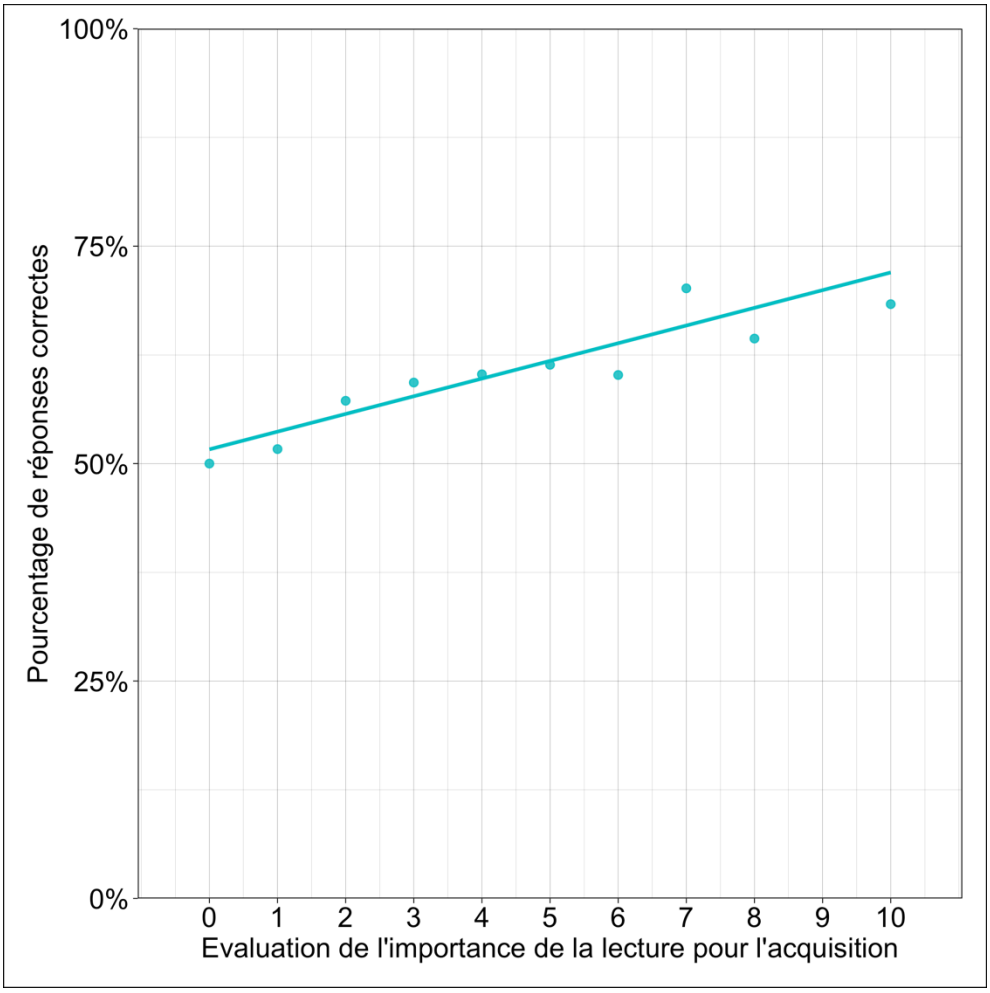
Cependant, ce facteur n'était pas prédictif d'une meilleure maitrise des connecteurs dans la tâche principale. Lors de l'ajout de ce facteur comme effet fixe, le modèle ne s'est pas amélioré ($\chi(1) = .04$, $p = .83$).

Parmi tous les facteurs, seul le facteur de la *Lecture* a pu prédire des meilleurs scores dans la tâche principale, car son ajout comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 10.06$, $p < .005$). L'output du modèle résultant est reporté dans le Tableau 5.14, l'effet est également visualisé dans la Figure 5.10.

Tableau 5.14. Output du modèle final pour l'évaluation de l'importance de la lecture pour l'acquisition du français par les personnes non-natives

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	.07	.17	.40	.69
L'importance de la lecture	.10	.03	3.22	< .005

Figure 5.10. Évaluation de la lecture pour l'acquisition du français par les personnes non-natives (la valeur de 10 correspond à l'importance la plus haute)



Nous concluons donc que plus une personne estimait que la lecture était importante pour sa propre acquisition du français, plus elle avait des scores élevés dans la tâche principale.

5.2.5 Discussion

Dans cette expérience, nous avons tenté de déterminer si certains connecteurs sont plus difficiles que d'autres et si leur fréquence ainsi que la relation qu'ils indiquent pourraient en être la cause. Pour ce faire, nous avons examiné la

connaissance explicite des apprenant-e-s du français pour 12 connecteurs en utilisant une tâche à choix multiple d'insertion de connecteur. En mesurant également les compétences linguistiques et l'exposition à l'écrit de nos participant-e-s, nous avons pu déterminer s'il existe des facteurs interpersonnels influençant la maitrise des connecteurs.

Certains connecteurs sont-ils plus difficiles que d'autres pour les apprenant-e-s ?

Premièrement, l'hypothèse selon laquelle certains connecteurs sont plus difficiles à maitriser que d'autres dans une L2 est renforcée par nos données. Nous avons observé des scores très bas pour des connecteurs comme *ainsi* alors que d'autres tels que *néanmoins* ont été maitrisé au même niveau que le groupe natif. Nous avons également vu qu'il existe des différences entre des connecteurs qui indiquent la même relation de cohérence. Il semble par exemple que *ainsi* soit beaucoup plus difficile à maitriser que *c'est pourquoi*. Cela montre l'importance de tenir compte des propriétés exactes d'un connecteur dans les expériences linguistiques avec des apprenant-e-s de langue seconde (voir aussi chapitre 3.6). Une expérience qui s'intéresse par exemple à la compréhension de la relation de cohérence d'une conséquence obtiendra probablement des résultats différents selon que *ainsi* ou *c'est pourquoi* est utilisé.

La question se pose donc de savoir pourquoi certains connecteurs semblent être plus difficiles ou faciles que d'autres.

La fréquence et la relation de cohérence : des facteurs prédictifs ?

Dans notre expérience, nous n'avions pas été en mesure d'attribuer incontestablement la maitrise des connecteurs aux deux facteurs testés, à savoir la fréquence du connecteur ou la relation de cohérence qu'il indique.

Concernant les effets de fréquence – l'idée qu'un connecteur plus fréquent serait plus facile à maitriser qu'un connecteur rare – nous avons observé des différences entre la maitrise des connecteurs seulement au sein de trois relations de cohérence sur les six testées. De plus, dans le cas de la relation de *conséquence*, le connecteur moins fréquent (*c'est pourquoi*) était mieux maitrisé que le connecteur fréquent (*ainsi*). Ces résultats suggèrent donc plutôt qu'il doit y

avoir d'autres caractéristiques que la fréquence d'un connecteur qui permettent d'expliquer les différences observées.

Quant au groupe natif, il n'est pas surprenant qu'il n'ait pas non plus été influencé par la fréquence des connecteurs, car il s'agissait d'un groupe adulte. Ces participant-e-s ont ainsi pu accéder à leurs connaissances explicites sur leur propre L1 sans contrainte de temps. Il est, en revanche, peut-être plus surprenant que nous ayons constaté des différences entre les scores pour les connecteurs en général (et non des effets de plafond). Cette observation confirme que même les personnes natives ont une maitrise très variable des connecteurs (voir aussi p.ex., Lamiroy, 1994; Zufferey & Gygax, 2020a, 2020b).

A cet égard, nos résultats suggèrent qu'une des raisons est la complexité cognitive de la relation de cohérence. Comme les personnes natives ont obtenu de meilleurs scores pour les connecteurs causaux que pour les connecteurs concessifs, nos résultats confirment que la difficulté à traiter une relation plus complexe est un facteur qui prédit la maitrise des connecteurs et la lecture en L1 (Murray, 1994, 1997; Sanders 2005).

Nous avons également observé que la relation d'une *condition* était très souvent correctement identifiée. Une raison pour cela pourrait être les scores très hauts pour *pourvu que* qui s'expliquent par le fait que ce connecteur nécessite d'être suivi par le subjonctif. Comme certains verbes des phrases expérimentales ont une forme qui varie entre l'indicatif et le subjonctif (e.g., *pouvoir*, *avoir*), il se peut que les personnes natives aient mieux identifié le connecteur approprié dans ces cas.[20]

Pour le groupe non-natif, nous n'avons pas observé qu'une relation de complexité élevée (telle qu'une *concession*) rendrait les connecteurs plus difficiles. Étant donné que ni la fréquence ni la relation de cohérence ne semblent avoir influencé les résultats des personnes non-natives, la question se pose de savoir pourquoi certains connecteurs semblent être plus difficiles pour le groupe non-natif que d'autres. Quelques explications peuvent être avancées.

[20] Par contre, on ne sait pas si cela représentait également une aide pour les personnes non-natives. Leurs scores n'étaient pas très hauts pour ce connecteur (48% de bonnes réponses, écart type = .50). Ainsi, il est également probable que le subjonctif, élément difficile pour les apprenant-t-e-s du français (McManus & Mitchell, 2015), rende la maitrise de *pourvu que,* en fait, plus difficile. L'influence du subjonctif sur la maitrise des connecteurs en L2 reste un thème à investiguer.

Premièrement, nous avons observé que le connecteur *c'est pourquoi* a été significativement mieux maitrisé que son équivalent *ainsi*. Un facteur qui pourrait expliquer cette différence est la transparence sémantique de *c'est pourquoi* en opposition avec l'opacité sémantique de *ainsi*. Cela voudrait dire que la fonction d'indiquer une conséquence de *c'est pourquoi* peut être comprise même si l'apprenant-e n'a pas appris *c'est pourquoi* en tant que connecteur. Il suffit simplement de comprendre les mots *c'est* et *pourquoi* pour pouvoir comprendre la conséquence. Cela s'oppose à l'acquisition du connecteur *ainsi* qui – si la personne ne l'a pas appris en tant que connecteur – reste, du fait d'une grammaticalisation plus avancée, sémantiquement opaque. Cela peut avoir également une influence sur la manière dont les personnes non-natives acquièrent les connecteurs : un connecteur sémantiquement plus transparent, c'est-à-dire avec un degré de grammaticalisation plus bas, pourrait être plus facile à apprendre.

Deuxièmement, nous avons observé que le connecteur *puisque* a suscité des scores significativement plus bas que *car*. Une raison pourrait être la charge pragmatique de *puisque*. Il est bien connu que ce connecteur est utilisé pour indiquer une information déjà connue ou une attitude dissociative envers le contenu (p.ex., Zufferey, 2012, 2014; Schumann et al., 2020). Étant donné ce profil pragmatique élaboré, on peut se poser la question de savoir si ces charges pragmatiques ont compliqué la maitrise de *puisque* dans notre expérience, respectivement le fait que les personnes non-natives ont hésité à le choisir. Nous allons revenir sur la charge pragmatique de *puisque* et les problèmes qu'elle pourrait poser aux apprenant-e-s d'une L2 dans le chapitre 9.

Troisièmement, nous avons observé que *néanmoins* a été maitrisé au même niveau que le groupe natif, contrairement à *cependant.* Une explication pour cette observation pourrait être un transfert positif de l'allemand *nichtsdestotrotz* ('néanmoins', ou encore lexicalement plus proche mais beaucoup plus rare : *nichtsdestoweniger*) qui aurait facilité la maitrise de *néanmoins*. Nous allons revenir sur les effets crosslinguistiques entre les connecteurs de la L1 et d'une L2 dans le chapitre 8.

Un dernier facteur qui pourrait expliquer nos résultats est celui du registre linguistique. Dans notre expérience, *par contre* a été beaucoup mieux maitrisé par les personnes natives que le connecteur contrastif *en revanche*. Étant donné que

par contre est en comparaison avec *en revanche* plutôt utilisé dans la langue orale, cela pourrait avoir conduit à une meilleure maitrise de ce connecteur.

Finalement, on peut se demander si les différences observées ne sont pas simplement dues à l'enseignement. Même si cela pourrait expliquer une partie des résultats, une anecdote provenant de la mise-en-œuvre de l'expérience mérite d'être mentionnée à ce sujet. Lors d'un entretien individuel, une enseignante de plusieurs classes testées a indiqué qu'elle utilisait uniquement *cependant* (donc le connecteur moins bien maitrisé dans notre expérience) et jamais *néanmoins*. Il convient également de noter que les élèves qui ont considéré l'école comme importante pour leur propre apprentissage du français n'ont pas mieux maitrisé les connecteurs. Dans les mesures d'auto-évaluation, un autre facteur semblait beaucoup plus important et prédictif pour une bonne maitrise des connecteurs : le fait de lire en L2. Cela confirme également nos résultats concernant les différences individuelles.

Les différences individuelles

Les résultats montrent non seulement qu'il existe de grandes différences individuelles parmi les apprenant-e-s quant à la maitrise des connecteurs, mais aussi que celle-ci peut être prédite par différentes compétences. Par exemple, nous avons vu qu'un meilleur vocabulaire ainsi qu'une meilleure compréhension grammaticale prédisaient une meilleure maitrise des connecteurs. Cela indique que la maitrise des connecteurs se situe à l'intersection de ces deux domaines linguistiques, et nous en concluons que leur mémorisation en tant que partie du vocabulaire d'une L2 ne devrait pas suffire à elle seule pour permettre leur bonne maitrise. D'autres compétences sont également nécessaires, notamment une bonne connaissance de la grammaire.

A cet égard, nous avons également vu qu'une exposition à l'écrit élevée, même en L1, prédisait une meilleure maitrise des connecteurs en L2. Bien que cette observation puisse surprendre, le fait qu'une compétence acquise en L1 puisse être transférée dans une L2 n'est pas inconnu dans la recherche. Par exemple, Dufva et Voeten (1999) ont observé dans une étude longitudinale sur des élèves finlandais d'une école primaire qu'un meilleur score dans des tâches de reconnaissance de mots et de compréhension, testées dans la L1, prédisaient une meilleure acquisition

d’une L2. L’équipe de recherche a donc conclu, selon l’*hypothèse de l’interdépendance* de Cummins (1979), qu’une compétence stratégique élevée en matière de traitement de texte dans la L1 pouvait être transférée à la seconde langue.

Quant à nos résultats et aux connecteurs de discours, cela voudrait dire qu’une personne qui a beaucoup lu dans sa L1 a déjà pu atteindre, de manière générale, une bonne compréhension des connecteurs et est consciente de leur importance pour structurer et enrichir un discours. Cette compétence générale des connecteurs pourrait être un avantage lors de l’acquisition des connecteurs d’une L2.

Les résultats indiquant des liens entre les compétences diverses et la maitrise des connecteurs doivent bien sûr être considérés avec précaution. Il est bien connu qu’une corrélation de deux variables ne signifie pas automatiquement une causalité. Il se peut en outre qu'il y ait des facteurs supplémentaires, voir confondants (tels que des facteurs socio-économiques) qui pourraient avoir une influence. Par exemple, il se peut qu’une personne issue d'un milieu proche de l'éducation ait été éduquée à lire beaucoup et à mieux s'impliquer dans sa scolarité. En conséquence, les liens observés entre la maitrise des connecteurs, la compétence de grammaire, la richesse du vocabulaire et l’exposition à l’écrit dans notre expérience s’expliquerait plutôt par des variables socio-économiques et/ou pédagogiques. Des recherches supplémentaires sont donc nécessaires pour distinguer davantage les facteurs individuels qui mènent à une meilleure maitrise des connecteurs en L2.

5.3 Conclusion

En conclusion, nous avons vu dans cette expérience que les apprenant-e-s maitrisent les connecteurs du français de manière très hétérogène et que des variables comme le type de relation et la fréquence du connecteur ne sont que partiellement explicatives. Pour certains connecteurs (p.ex., *néanmoins*), les apprenant-e-s ont montré une maitrise à un niveau natif. Pour d’autres (p.ex., *puisque* ou *ainsi*), les scores étaient beaucoup plus bas. Du fait que la relation de cohérence et la fréquence du connecteur ne semblaient pas être des facteurs prédictifs pour les personnes non-natives, nous avons supposé que ces derniers

étaient plus influencés par le registre et la transparence sémantique d'un connecteur.

Dans les chapitres suivants, nous investiguerons plus en détail quels facteurs linguistiques et cognitifs influencent la maitrise des connecteurs en L2. Dans ce contexte, il convient également de différencier deux types de connaissances, à savoir la connaissance explicite et implicite. Il se peut, par exemple, que les connaissances implicites des connecteurs se développent différemment des connaissances explicites. Au chapitre suivant, nous allons donc examiner la connaissance implicite des connecteurs par les personnes non-natives, d'abord en testant à quel degré les personnes non-natives sont capables d'intégrer les significations des connecteurs lors de la lecture (ce qui implique d'accéder à leurs connaissances implicites).

La question qui se pose ensuite est de savoir si les connaissances implicites des personnes non-natives sont plus ou moins développées par rapport aux connaissances explicites. Nous allons comparer ces deux types de connaissances pour différents connecteurs dans les chapitres 7 et 8.

6 La capacité des personnes non-natives à intégrer les significations des connecteurs lors de la lecture

Jusqu'à présent, nous avons observé que la maitrise des connecteurs dans une L2 varie considérablement en fonction du type de connecteur (chapitre 5). Par exemple, les connecteurs tels que *c'est pourquoi* ou *par contre* semblent, pour des raisons diverses, beaucoup plus facile à maitriser que des connecteurs comme *cependant* ou *puisque*. Nous avons également observé que le type de connecteur influence la réaction à l'incohérence dans la L1 (chapitre 4). Plus concrètement, nous avons vu que l'incohérence dans une phrase provoquait des perturbations du traitement mental très tôt lorsqu'elle était marquée avec un connecteur fréquent. Prises ensemble, ces deux observations indiquent que l'hétérogénéité des connecteurs pourrait se répercuter également sur le traitement mental des personnes non-natives, c'est-à-dire qu'un connecteur plus complexe pourrait rendre la détection de l'incohérence plus difficile pour une personne non-native.

Avant d'examiner cette hypothèse (voir chapitre 7), il convient de déterminer d'abord si (et si oui, dans quelle mesure) les personnes non-natives intègrent les significations des connecteurs lors du traitement mental d'un texte. Ce n'est que lorsque nous saurons à quel degré les personnes non-natives sont sensibles aux connecteurs que nous pourrons évaluer plus précisément l'influence des différents types de connecteurs. La *sensibilité aux connecteurs* signifie ici une grande facilité d'activer leurs instructions procédurales lors de la lecture et une mise en correspondance immédiate des propositions du discours. Une bonne sensibilité aux connecteurs permet donc également de détecter rapidement l'incohérence lorsqu'un connecteur est mal utilisé.

Le fait que les personnes non-natives soient sensibles aux connecteurs n'est pas une certitude. Il est discuté dans des nombreux travaux que la lecture dans une deuxième langue est très différente de celle dans une L1 (Clahsen & Felser, 2006b; Recio Fernández, 2020). Ainsi, les effets de facilitation des connecteurs pour les personnes natives (e.g., Britton et al., 1982; Haberlandt, 1982) ne peuvent pas être forcément transférés et appliqués au traitement mental dans une L2. Bien que

certaines études indiquent que la présence des connecteurs améliore la compréhension d'un texte (Degand & Sanders, 2002), Papadopoulou & Clahsen (2003) ont trouvé que les personnes non-natives sont moins fortement guidées par des indices fonctionnels (que par des indices lexicaux et conceptuels). Cela pourrait signifier que les personnes non-natives sont, d'une certaine manière, moins sensibles aux connecteurs durant la lecture.

Jusqu'à présent, il n'existe malheureusement pas de preuves concluantes à ce sujet. Certes, plusieurs études ont déjà examiné le traitement mental des connecteurs dans une L2, mais ces études ont mis en évidence les problèmes liés aux connecteurs compliqués, atypiques ou dans des contextes difficiles (e.g., Zufferey et al., 2015; Zufferey & Gygax, 2017; Crible et al., 2021). Il manque donc des études montrant l'influence de connecteurs qui ne sont pas particulièrement complexes sur le traitement mental dans une deuxième langue.

Pour combler ce vide, nous nous intéressons dans ce chapitre au traitement mental des personnes non-natives pour des phrases correctement ou incorrectement marquées par des connecteurs (Expérience 1). L'idée est que si les personnes non-natives sont sensibles aux connecteurs durant le traitement mental, elles doivent être immédiatement affectées par les phrases incorrectement marquées et ralentir par conséquent leur lecture. Dans un deuxième temps, nous nous intéresserons à la question de savoir si les personnes non-natives sont également sensibles à l'incohérence lorsqu'elle est causée par une absence de connecteur (Expérience 2). Nous supposons que l'incohérence qui résulte de l'absence de connecteur est peut-être moins saillante que celle qui est causée par une indication erronée donnée par un connecteur mal utilisé. Ainsi, elle devrait être plus difficile à détecter.

Ensemble, ces deux expériences nous aident à clarifier si les personnes non-natives peuvent efficacement intégrer le sens d'un connecteur lors de la lecture ou si elles se fient plutôt au contenu lexical d'une phrase en négligeant des instructions du connecteur.

6.1 Le traitement mental en L2

Dans ce chapitre, nous nous interrogeons sur la question de savoir si les personnes non-natives sont sensibles aux connecteurs. Par rapport au chapitre précédent, ce

chapitre ne s'intéresse donc pas aux connaissances explicites des personnes non-natives, mais examine plus en détail ce qui se passe *pendant* la lecture en L2. Nous supposons que les connecteurs peuvent avoir un impact différent sur la lecture en L2 par rapport à la lecture en L1, car il existe de grandes différences entre la lecture en L1 et en L2, comme nous allons maintenant le voir.

6.1.1 Les caractéristiques de la lecture en L2 : théories et modèles sur le traitement mental

Lors de la lecture d'un texte, qu'il soit en L1 ou en L2, il y a deux processus successifs qui se déroulent à des niveaux différents (Kendeou et al., 2014). Dans une première étape, une personne lectrice doit activer les sens des mots, c'est-à-dire convertir les mots écrits dans des unités de sens. Ensuite, ces unités de sens doivent être combinées et reliées dans un tout cohérent. Ce n'est que lorsque ces deux processus fonctionnent et interagissent qu'une personne arrive à comprendre un texte. Il existe des indices que ces deux processus sont plus difficiles dans une L2.

D'abord, l'activation des sens de mots est plus difficile dans une L2 car ceux-ci peuvent être inconnus, seulement partiellement connus ou même mal compris. Et même s'ils sont connus en principe, leur activation automatisée[21] par les personnes natives (Segalowitz 2000, 2001) ne fait que se développer chez les personnes non-natives (Segalowitz & Segalowitz, 2008; Tavakoli, 2019). Évidemment, cela a également des conséquences pour la compréhension du discours. Taguchi (2008), par exemple, a pu montrer que moins les apprenant-e-s accédaient rapidement aux sens des mots, moins ils comprenaient les implicatures conversationnelles. D'autres études ont également mis en évidence que les personnes non-natives éprouvent des difficultés à coordonner les informations syntaxiques, sémantiques et pragmatiques, et cela même pour des opérations apparemment simples, comme la résolution des pronoms (Felser, Roberts, Marinis & Gross, 2003; Marinis, Roberts, Felser & Clahsen, 2005; Roberts, Gullberg & Indefrey, 2008).

[21] Définie ici comme « la capacité d'accomplir des choses sans effort conscient, surveillance ou l'occupation de la capacité mentale et les ressources attentionnelles qui sont normalement disponibles pour traiter l'information » (Tavakoli, 2019:4, ma traduction)

Une des raisons est que nous disposons de moins de ressources cognitives lorsque nous lisons dans une langue seconde (Clahsen & Felser, 2006a). Selon Recio Fernández (2020), le traitement mental dans une langue seconde est, par rapport au traitement dans la L1, plus superficiel, moins automatique et manque de vigilance épistémique (Sperber, Clément, Heintz, Mascaro, Mercier, Origgi, Wilson, 2010), c'est-à-dire implique une capacité moindre à évaluer la fiabilité d'un énoncé. Ainsi, plusieurs études ont démontré en utilisant les méthodes des *potentiels évoqués* que les personnes non-natives traitent des anomalies syntaxiques et sémantiques différemment que les personnes natives (Hahne & Friederici, 2001; Weber-Fox & Neville, 1996; Ardal, Donald, Meuter, Muldrew & Luce, 1990; Felser et al., 2003). Il semble en outre que l'effort mental nécessaire pour lire un texte dans une langue seconde favorise inconsciemment certaines stratégies de lecture. Papadopoulou & Clahsen (2003), par exemple, décrivent que les personnes non-natives bénéficient plus des mots conceptuels que des mots fonctionnels. Sorace (2011) affirme également que la superficialité du traitement mental d'une personne non-native conduit à un recours excessif aux informations lexicales qui semblent apprises.

En conclusion, le traitement mental en L2 nécessite plus de ressources cognitives que le traitement mental en L1 (Clahsen & Felser, 2006a). Il en résulte, comparé avec la lecture en L1, que le traitement en L2 est plus superficiel (Recio Fernández, 2020), moins automatisé (Segalowitz & Segalowitz, 2008) et caractérisé par une 'sur-dépendance' aux formes lexicales (Papadopoulou & Clahsen, 2003). Tenant compte de ces difficultés, il n'est pas surprenant que, de manière générale, de nombreux auteur-trice-s ont observé que la lecture des personnes non-natives est généralement plus lente que celle des personnes natives (e.g., Crible et al., 2021).

Considérant le fait qu'il est difficile de lire dans une L2 et que spécialement les connecteurs sont complexes pour les apprenant-e-s (e.g., Zufferey & Gygax, 2017, chapitre 5), pourrait-on donc conclure que les connecteurs sont tout simplement trop difficiles à traiter dans une langue seconde ? La réponse n'est pas si évidente. Étant donné que les personnes non-natives ont besoin de davantage de ressources mentales pour décoder le contenu d'un discours (p.ex., Clahsen & Felser, 2006; Recio Fernández, 2020), on pourrait également argumenter que les personnes non-natives devraient profiter de la présence des connecteurs car ceux-

ci fournissent des instructions claires sur la façon d'interpréter les liens logiques d'un texte. Degand & Sanders (2002:743) expliquent :

> « Les lecteur-trice-s devraient profiter d'un marquage linguistique de la structure d'un texte car les marqueurs leur donnent une aide supplémentaire dans la construction de la représentation du texte. (...) En fait, l'hypothèse est que les personnes non-natives peuvent utiliser leur connaissance des marqueurs de structure pour pallier leur faible connaissance linguistique générale et implicite dans la langue seconde. » (Degand & Sanders, 2002 : 743, ma traduction)

Et en effet, leur étude a montré que la présence des connecteurs a facilité la compréhension d'un texte en L2, comme nous allons maintenant le voir.

6.1.2 Les connecteurs peuvent-ils faciliter le traitement en ligne d'un discours dans une langue seconde ?

Comme les personnes non-natives sont moins sensibles aux mots fonctionnels (Papadopoulou & Clahsen, 2003), on peut supposer qu'elles ne sont pas sensibles aux connecteurs lors de la lecture d'un texte. Par exemple, Zufferey et Gygax (2017) ont trouvé que des apprenant-e-s de français n'étaient pas affectés par l'absence du connecteur *en effet* dans des relations de confirmation, alors que cette absence rendait la relation incohérente pour des personnes natives.

Les connecteurs sont des mots qui servent à *faciliter* la lecture et il a été également démontré que les apprenant-e-s comprennent mieux des textes marqués avec des connecteurs. Degand et Sanders (2002), par exemple, ont montré que les apprenant-e-s L2 de français et de néerlandais qui lisaient des textes répondaient plus correctement aux questions de compréhension lorsque ces textes contenaient des connecteurs causaux (tels que *parce que* ou son équivalant néerlandais *omdat*). Cela veut dire que la présence de connecteurs causaux, bien qu'elle ne soit pas obligatoire, les a aidés à identifier et à décoder les relations causales et leur a permis, par conséquent, de mieux comprendre les textes. Les auteur-trice-s surlignent que cela a été rendu possible par le fait que les connecteurs étaient bien maitrisés par les personnes interrogées.

Comme d'autres études ont également démontré que les personnes non-natives bénéficient de la présence des connecteurs (par exemple, Crosson & Lesaux, 2013; van den Bosch, Segers & Verhoeven, 2018), on peut supposer que les connecteurs facilitent, d'une manière ou d'une autre, la lecture en L2. Pourtant, il est difficile de tirer de ces études sur la compréhension des conclusions pertinentes pour la capacité des personnes non-natives à intégrer les significations

des connecteurs durant la lecture. Seulement très peu d'études ont montré leur influence directe et cela seulement de manière « négative », par exemple un manque de sensibilité lorsque le connecteur est atypique ou complexe (Zufferey & Gygax, 2017; Crible et al., 2021). Ainsi, nous ne savons que très peu de choses sur la manière dont les connecteurs qui sont bien maitrisés affectent la lecture en L2. Même si les connecteurs ne facilitent en tout cas pas toujours la lecture en L2 (Zufferey & Gygax, 2017; Crible et al., 2021), cela ne signifie pas pour autant que les locuteur-trice-s ne peuvent jamais en bénéficier.

Pour tester dans quelle mesure les connecteurs influencent l'intégration des significations des connecteurs lors de la lecture des personnes non-natives, nous allons maintenant présenter une expérience de lecture segmentée.

6.2 Expérience 1 : Mauvaises utilisations des connecteurs et perturbation de la lecture

Dans cette expérience[22], nous allons examiner la capacité des personnes non-natives à intégrer la signification des connecteurs lors de la lecture. Pour faire cela, nous nous servirons d'une méthodologie contenant deux variables qui permet d'investiguer en détail le rôle des connecteurs lors de la lecture d'une personne non-native.

Premièrement, nous nous intéressons spécifiquement à déterminer dans quelle mesure les personnes non-natives sont capables de détecter de l'incohérence causée par une utilisation incorrecte d'un connecteur. En mesurant les temps de lecture pour différentes parties d'une phrase, nous allons retracer si et à quel moment l'incohérence provoque des perturbations de la lecture. Cette approche nous a déjà servi au chapitre 4 et a permis de démontrer que le type de connecteur utilisé peut influencer la détection de l'incohérence chez les personnes natives.

Deuxièmement, nous nous intéressons aux connecteurs qui indiquent des relations de cohérence différentes. Nous savons que les relations de cohérence varient dans leur complexité cognitive (Sanders et al., 1992) et que, pour les

[22] Les résultats des deux expériences ont été publiés, sous une autre forme, dans le chapitre de livre intitulé « Do non-native readers rely on connectives? The processing of coherence relations in a second language » (Wetzel, Zufferey & Gygax, 2022b) dans le livre *Discourse Markers in Interaction: From Production to Comprehension* (Cuenca & Degand, 2022).

personnes natives, les relations plus complexes sont plus difficiles à traiter que les relations plus simples (e.g., Köhne & Demberg, 2013). Nous avons d'ailleurs observé nous-mêmes que les relations plus complexes ont influencé la lecture chez les personnes natives (chapitre 4) et ont suscité des taux de bonnes réponses plus bas dans la tâche d'insertion (chapitre 5). En testant également des relations de cohérence de complexités différentes, nous pouvons examiner si une concession affecte le traitement mental d'une personne non-native dans une plus large mesure qu'une conséquence.

Étant donné que nous allons tester des connecteurs bien maitrisés (voir ci-dessous), nous nous attendons à ce que le traitement mental des personnes non-natives soit affecté par leur mauvaise utilisation. En conséquence, les personnes non-natives devraient ralentir leur lecture pour des phrases incohérentes. Nous nous attendons également à un effet de relation de cohérence : comme la concession est plus complexe à traiter que la cause, la réaction à l'incohérence devrait être un peu retardée pour les phrases concessives qui sont incorrectement marquées (p.ex., Köhne & Demberg, 2013; Lyu et al., 2020).

Les connecteurs testés

Contrairement aux expériences du chapitre 4, ce chapitre ne traite pas des différences entre connecteurs mais examine plutôt la sensibilité aux connecteurs en général (nous examinerons l'impact du type de connecteur sur la lecture en L2 au chapitre 7). Évidemment, cela implique le risque de confondre un manque de sensibilité avec la non-maitrise d'un connecteur. Pour nous assurer que la complexité des connecteurs n' influence pas négativement l'activation de leurs significations, nous avons utilisé des connecteurs dont on peut estimer qu'ils sont bien maitrisés par les personnes non-natives. Ce n'est qu'en utilisant des connecteurs auxquels les personnes non-natives sont habituées que nous pouvons déterminer leur sensibilité aux connecteurs. Nous avons choisi les connecteurs *donc* (conséquence) et *mais* (concession), donc les mêmes connecteurs utilisés au chapitre 4.

Nous pensons que ces connecteurs permettent d'examiner la sensibilité des personnes non-natives aux instructions procédurales des connecteurs pour deux raisons.

D'abord, comme nous avons déjà vu au chapitre 4.5, ces connecteurs sont extrêmement fréquents à la fois à l'oral et à l'écrit en comparaison avec d'autres connecteurs qui indiquent la même relation de cohérence. Les fréquences des différents connecteurs dans la collection des différents sous-corpus oraux et écrits ORFEO (Benzitoun et al., 2016) testés au chapitre 4 sont à nouveau reportées dans le Tableau 6.1 pour rappel.

Tableau 6.1. Les nombre d'occurrences des connecteurs mais, donc, néanmoins, ainsi, aussi et or dans la collection de corpus ORFEO (voir également chapitre 4.5)

Connecteur	*À l'écrit*	*À l'oral*	*total*
mais	191	221	412
donc	113	221	334
néanmoins	20	6	26
ainsi	97	35	132
aussi (*conséquence*)	91 (estimé)	0 (estimé)	91
or (*concession*)	28 (estimé)	11 (estimé)	39

Deuxièmement, nous avons vu au chapitre 4 que seuls les connecteurs *donc* et *mais* ont exercé un effet de facilitation pour la détection de l'incohérence chez les personnes natives. Ainsi, on peut supposer qu'il existe un effet similaire chez les personnes non-natives.

6.2.1 Les participant-e-s

Nous avons recruté 57 participant-e-s anglophones via la plateforme *Prolific* (Prolific, London, 2020) qui ont tout-e-s indiqué parler le français comme langue seconde. Nous avons dû exclure 10 personnes de l'analyse, à cause de taux insatisfaisants de bonnes réponses dans les questions de vérification (< 75% voir ci-dessous pour la tâche). Ainsi, les données de 47 personnes ont été analysées (âge moyen = 42 ans, écart type = 15.8, 35 femmes). Il n'a pas été nécessaire de recruter un groupe de contrôle natif, car nous possédons déjà des valeurs de comparaisons, à savoir les résultats du chapitre 4.2.

Tous les participant-e-s ont donné leur consentement de participer à cette expérience et ont été rémunéré-e-s avec 3.15 GBP. Tous les participant-e-s ont montré dans d'autres études sur la plateforme *Prolific* des participations satisfaisantes dans au moins 95% des cas.

Pour permettre une comparaison avec d'autres expériences, il convenait de contrôler la compétence linguistique des participant-e-s. Nous avons fait cela en utilisant la tâche *Lextale* (version française, Brysbaert, 2013) qui a montré des

corrélations avec la compétence de grammaire et la maitrise des connecteurs des apprenant-e-s au chapitre 5.

Dans cette tâche, les participante-e-s de cette expérience ont obtenu en moyenne un score de 46%, (95% IC [45.25, 46.89]) avec un écart type de 18.23, ce qui correspond à un score moyen de 25.76 dans cette tâche. Veuillez noter que les scores *Lextale* n'étaient à aucun moment prédictifs pour les temps de lecture.

6.2.2 Le design et la procédure

Dans cette expérience, nous avons refait l'expérience du chapitre 4.3, cette fois avec des personnes non-natives. Ainsi, nous avons mesuré les temps de lecture des participant-e-s suivant un design factoriel de 2^x2, dont les variables étaient le *Type de phrase* (*cohérente* ou *incohérente*) et le *Type de relation de cohérence* (*conséquence* ou *concession*). Pour rappel, les phrases incohérentes étaient marquées par un connecteur incorrect (*donc* dans une concession et *mais* dans une conséquence).

Nous avons utilisé les mêmes phrases et la même procédure que celle décrite dans le chapitre 4.3.2.

6.2.3 L'analyse

Nous avons réalisé la même analyse qu'au chapitre 4 (voir 4.3.3.) c'est-à-dire en faisant des modèles linéaires à effets mixtes à l'aide du logiciel *R* (R Core Team, 2020), montés selon la procédure de Baayen (2008). Nos facteurs étaient le *Type de phrase* (*cohérente* ou *incohérente*) et le *Type de relation de cohérence* (*conséquence* ou *concession*). Les modèles comprenaient toujours les *Participant-e-s* et les *Phrases expérimentales* commet effets aléatoires. Pour tester l'amélioration des modèles lors de l'ajout successif des effets fixes, les tests du rapport de vraisemblance ont été effectués à l'aide de la fonction *anova()* du package *stats* (R Core Team, 2020). Les modèles ont été exécutés à l'aide de la fonction *lmer()* du package *lmerTest* (Kuznetsova et al., 2017). Pour obtenir les tailles d'effet estimées et le niveau de signification, nous avons appliqué la fonction *summary()* du package *stats* (R Core Team, 2020) aux modèles finaux. Pour les tests *post-hoc*, nous avons effectué des tests simultanés pour les hypothèses

linéaires générales (comparaisons multiples de moyennes, contrastes de Tukey) en utilisant la fonction *glht()* du package *multcomp* (Hothorn et al., 2008).

Comme dans les expériences du chapitre 4, nous avons exclu les temps de lecture inférieurs à 50ms et supérieurs à 3s pour les segments de la phrase et pour les réponses aux questions de compréhension, nous avons autorisé des réponses inférieures à 50ms et supérieurs à 10s. Ces mesures ont déjà réduit considérablement l'asymétrie de nos données (telle qu'examinée avec la fonction *skewness()* du package *stats*, R Core Team, 2020). De plus, nous avons fait une transformation logarithmique (Crible et al., 2021) qui a réduit encore l'asymétrie de nos données.

Les temps de lecture pour les segments 5, 6 et 7

Pour le segment 5, ni l'ajout du *Type de phrase* (qu'elle soit cohérente ou incohérente, $\chi(1) = 1.15$, $p = .28$), ni l'ajout du *Type de relation de cohérence* (qu'elle soit de conséquence ou concessive, $\chi(1) = 2.86$, $p = .09$) comme effets fixes ont amélioré le modèle de manière significative. En revanche, nous avons observé un effet de cohérence dans le segment 6 : le modèle s'est amélioré en ajoutant le *Type de phrase* comme effet fixe ($\chi(1) = 17.59$, $p = 2.75^{e-05}$). L'ajout de l'interaction avec le *Type de relation de cohérence* n'a pas amélioré le modèle davantage ($\chi(2) = .32$, $p = .85$). L'output du modèle final, contenant donc le *Type de phrase* comme seul effet fixe (ainsi que les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires), est reporté dans Tableau 6.2.

Tableau 16.2. Expérience 1. L'output du modèle final pour le segment 6

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	- 1.11	.06	67.22	-17.61	$< 2^{e-16}$
Phrase (*incohérent*)	.04	.02	2293.21	2.20	.03

Comme nous le voyons, les phrases incohérentes ont été lues plus lentement dans le segment 6. Cet effet semblait se prolonger également jusqu'au segment 7 : lors de l'ajout du *Type de phrase* (cohérente, incohérente) comme effet fixe, le modèle pour le segment 7 s'est également amélioré ($\chi(1) = 30.03$, $p = 4.27^{e-08}$). L'ajout de l'interaction avec le *Type de relation de cohérence* n'a pas amélioré le modèle davantage ($\chi(1) = 2.71$, $p = .23$). L'output du modèle du segment 7 (contenant donc

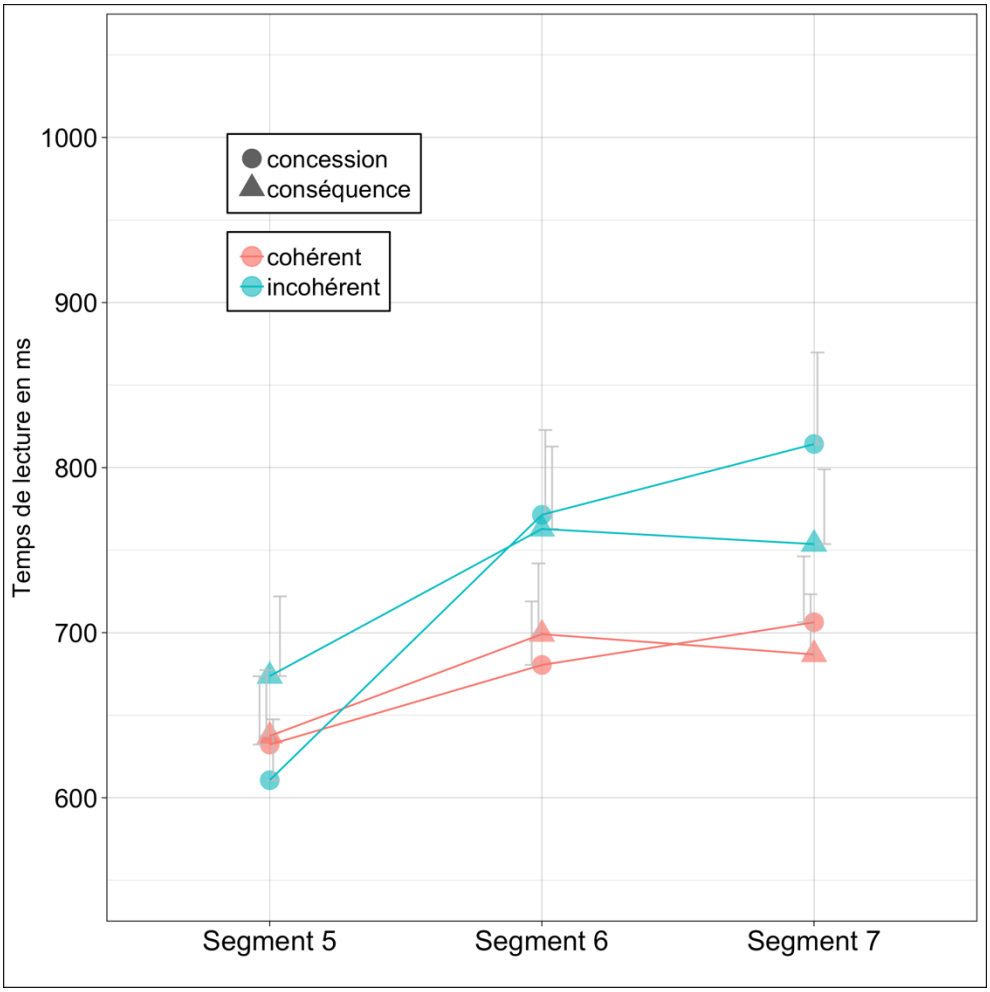
seulement le *Type de phrase* comme effet fixe et les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires), est reporté dans le Tableau 6.3.

Tableau 6.3. Expérience 1. L'output du modèle final pour le segment 7

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.48	.06	62.65	-8.60	3.29^{e-12}
Phrase (*incohérent*)	.10	.02	1761.95	5.50	4.27^{e-08}

Nous observons donc un ralentissement pour les phrases incohérentes à partir du segment 6, mais cela indépendamment de la relation de cohérence en question. Les temps de lecture sont également illustrés à la Figure 6.1.

Figure 6.1. Expérience 1. Les temps de lectures en moyennes pour les segments 5, 6 et 7 (limites supérieures de l'IC 95% indiquées par des barres d'erreur)



Les réponses aux questions de vérification

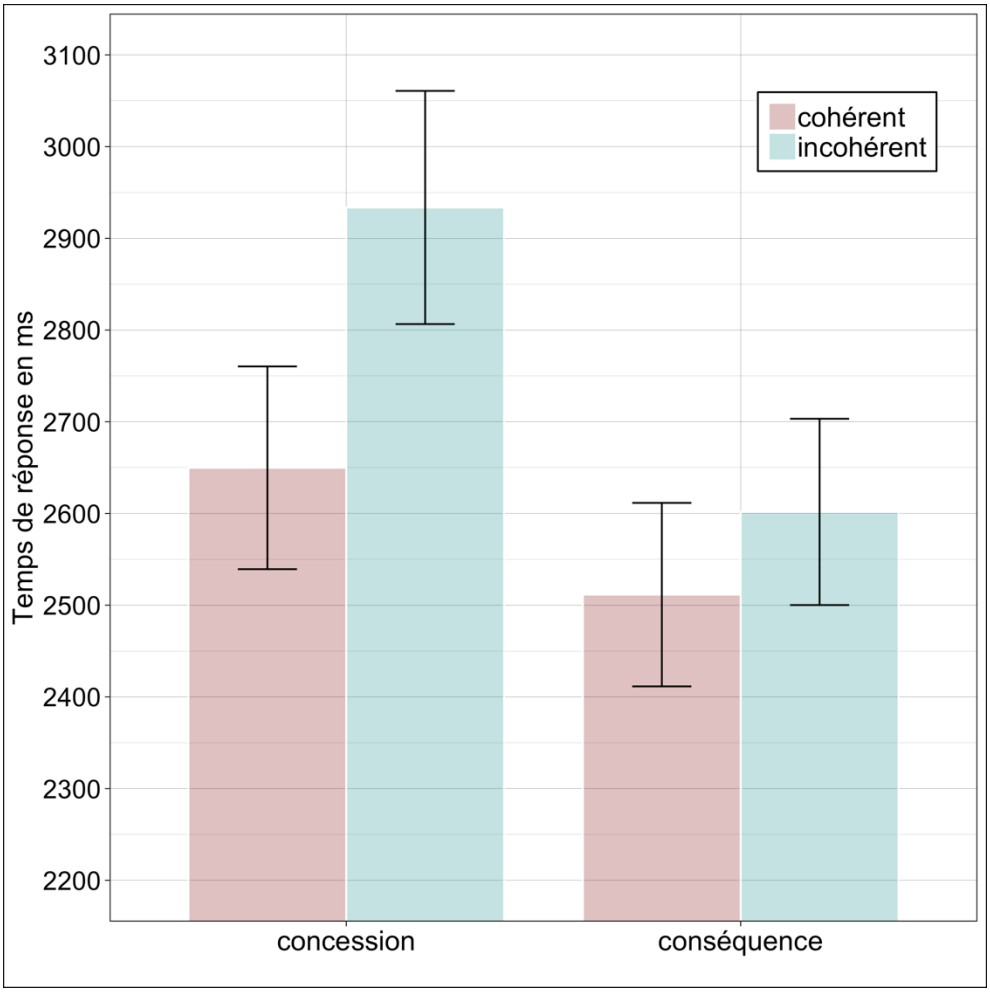
Pour le temps de réponses aux questions de vérification, l'ajout du *Type de phrase* comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 25.24$, $p = 5.06^{e-07}$). En plus, l'ajout de l'interaction avec l'effet fixe du *Type de relation de cohérence* a également amélioré le modèle ($\chi(1) = 34.21$, $p = 3.73^{e-08}$). L'output du modèle final, contenant donc l'interaction entre le *Type de phrase* et le *Type de relation de cohérence* (ainsi que les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires), est reporté dans Tableau 6.4.

Tableau 6.4. Expérience 1. L'output du modèle final pour les temps de réponses aux questions de vérification

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	.90	.05	70.53	18.95	< .001
Type (*incohérent*)	.10	.02	1736.23	5.19	2.39^{e-07}
Relation (*conséquence*)	-0.04	.02	1736.04	-2.29	.02
Type (*incohérent*) : Relation (*conséquence*)	-0.06	.03	1736.23	-2.25	.03

Les tests *post-hoc* ont indiqué que les temps de réponse aux questions étaient plus lents lorsque la phrase était concessive et incohérente (par rapport aux phrases concessives cohérentes : $\beta = .10$, $ET = .02$, $z = 5.19$, $p < .001$; par rapport aux phrases de conséquence incohérentes : $\beta = .12$; $ET = .02$, $z = -5.42$, $p < .001$ et par rapport aux phrases de conséquence cohérentes : $\beta = .15$, $ET = .02$, $z = 7.5$, $p < .001$). En outre, les temps de réponses ne différaient pas significativement pour les phrases de conséquence, qu'elles soient cohérentes ou incohérentes ($\beta = .04$, $ET = .02$, $z = 2.02$, $p = .18$). Ces résultats sont également illustrés à la Figure 6.2.

Figure 6.2. Expérience 1. Les temps de réponses aux questions de vérification



6.2.4 Discussion

Dans cette expérience, nous nous sommes intéressés à la question de savoir si et dans quelle mesure les personnes non-natives sont sensibles aux connecteurs durant la lecture. Nous avons examiné cela en mesurant les temps de lecture pour des phrases qui étaient indiquées correctement par un connecteur (et qui étaient donc *cohérentes*) et des phrases qui étaient indiquées incorrectement par un connecteur (et qui étaient donc *incohérentes*). Les connecteurs utilisés étaient *donc* et *mais*, deux connecteurs dont on peut supposer qu'ils sont bien maitrisés par les personnes non-natives. En testant ces deux connecteurs différents, nous avons également pu tester les temps de lecture pour deux relations de cohérence qui diffèrent dans leur complexité.

Les résultats obtenus reflètent une réaction rapide aux mauvais usages des connecteurs. Les phrases qui contenaient un connecteur incorrect ont été lues significativement plus lentement que les phrases marquées correctement. Il est

même remarquable que nous ayons observé un effet déjà dans le segment 6, donc un segment ou l'incohérence de la phrase peut théoriquement être encore résolue. Cette réaction rapide est même comparable aux observations faites au chapitre 4. Pour rappel, nous y avons observé que les personnes natives ont été également affectées par l'incohérence (pourvu que les connecteurs soient fréquents) à partir du segment 6.

Malgré le fait que la lecture dans une L2 nécessite plus de ressources cognitives et soit plus superficielle que dans la L1 (Clahsen & Felser, 2006; Recio Fernández, 2020), nous avons vu que les personnes non-natives peuvent être sensibles aux fonctions procédurales des connecteurs, reflétée par une perturbation de la lecture lorsque les connecteurs sont utilisés incorrectement. Toutefois, nous avons testé des connecteurs très fréquents et ces réactions immédiates ne s'appliquent probablement pas à tous les connecteurs (voir chapitre 7). Pour l'expérience actuelle, les réactions observées à l'incohérence justifient l'hypothèse que *mais* et *donc* sont généralement maitrisés par les personnes non-natives.

Nous avons également observé que la réaction à l'incohérence dépendait de la relation de cohérence. Il semble qu'il était spécialement difficile de répondre aux questions de vérification lorsque la phrase précédente était incohérente et concessive. En effet, une phrase concessive et incohérente suscitait des temps de réponses plus longs qu'une phrase de conséquence incohérente. Nous pouvons donner deux explications à cela, qui ne sont d'ailleurs pas contradictoires.

Premièrement, il est possible que les temps de réponses étaient plus longs pour une phrase concessive (incohérente) que pour une phrase de conséquence (incohérente) parce que ces deux relations de cohérence diffèrent dans leur niveau de complexité cognitive (Sanders et al., 1992). Ainsi, l'intégration des informations fourni par le connecteur concessif nécessite un effort mental élevé ce qui, en plus de l'incohérence de la phrase, provoquait plus de perturbation de la lecture que pour une phrase causale.

Deuxièmement, il est possible qu'à l'inverse, l'incohérence dans une phrase de conséquence (incohérente), telle que (1), était plus facile à résoudre grâce au lien causal qui existe entre la première et deuxième clause. Cela voudrait dire que les personnes auraient négligé plus facilement l'indication du connecteur *mais* car

elles s'attendaient à une continuité causale dans la deuxième clause (Sanders, 2005).

1) Jean veut connaitre les nouvelles du jours [mais] il lit le journal tous les matins.

Nous avons d'ailleurs observé et discuté un effet similaire pour les personnes natives au chapitre 4. Nous concluons que les personnes non-natives étaient dans une certaine mesure capable de négliger les indications perturbantes fournies par les connecteurs afin de construire une représentation mentale cohérente de la phrase.

Prises ensemble, nos observations indiquent que les personnes non-natives sont sensibles aux indications fournies par les connecteurs lors du de la lecture. Cependant, il faut noter qu'une contradiction entre l'indication du connecteur et les propositions de la phrase est plutôt saillante. On peut donc se demander si l'incohérence provoque des réactions tout aussi fortes lorsque l'incohérence est plus subtile, par exemple lorsque celle-ci est due à l'absence de connecteur. Si l'utilisation d'un connecteur est obligatoire mais que ce dernier est absent, il se peut que l'incohérence résultante soit perçue beaucoup plus tard (ou pas du tout, cf. Zufferey & Gygax, 2017). Réaliser une expérience sur la réaction aux connecteurs manquants en reprenant le matériel linguistique de cette expérience nous permettra d'examiner de manière plus nuancée la sensibilité aux connecteurs des personnes non-natives lors de la lecture.

6.3 Expérience 2: la réaction aux connecteurs manquants

Même si les personnes non-natives semblent être sensibles aux instructions des connecteurs (Expérience 1), nous supposons qu'elles auront plus de problèmes à détecter les cas où le connecteur manque. Cela s'applique spécialement aux relations concessives, car celles-ci ne peuvent pas[23] être laissée implicites (Murray, 1997).

Dans cette expérience, nous allons donc refaire l'expérience précédente, avec des phrases implicites au lieu de phrases incohérentes.

[23] Ou, si c'est le cas, seulement dans de très rares cas (lorsque les inférences sont tellement pervasives qu'un marquage explicite n'est pas nécessaire).

Nous nous attendons à ce que les phrases concessives qui ne sont pas marquées par un connecteur (et qui sont donc implicites) provoquent des perturbations de la lecture. Cependant, des effets de ralentissement pourraient émerger un peu plus tard, car l'incohérence est moins saillante que dans la première expérience ce qui fait que l'incohérence est moins rapidement détectée.

6.1.1 Les participant-e-s

Nous avons recruté 53 personnes anglophones via la plateforme *Prolific* (Prolific, London, 2020) qui ont déclaré parler le français comme langue seconde. Quatre personnes ont été exclues de l'analyse à cause des taux insatisfaisants dans leurs réponses aux questions de vérification (<75% de bonnes réponses). Par conséquent, les données de 49 personnes non-natives (âge moyen = 36.6 ans, avec un écart type de 12.8, 44 femmes) ont été analysées.

Contrairement à la première expérience, nous n'avions pas cette fois de données empiriques pour des personnes natives, qui nous permettraient de mieux interpréter les résultats des personnes non-natives. Ainsi, nous avons inclus un groupe de contrôle natif. Ce groupe comprenait 65 personnes natives (âge moyen = 27.5 ans avec un écart type de 8.2, 33 femmes) qui ont également été recrutées via la plateforme *Prolific* (Prolific, London, 2020)

Comme dans la première expérience, tous les participant-e-s ont donné leur consentement de participer à cette expérience et ont été rémunéré-e-s avec 3.15 GBP. De plus, tous les participant-e-s ont montré dans d'autres études sur la plateforme *Prolific* des participations satisfaisantes dans au moins 95% des cas.

Concernant leur compétence linguistique, les personnes non-natives ont eu un score de 41.55% (95% IC [40.82; 42.29]) en moyenne avec un écart type de 16.61 dans la tâche *Lextale*. Les personnes natives ont obtenu en moyenne un score de 81.06% (95% IC [80.59; 81.54]) avec un écart type de 12.24. Ainsi, les scores obtenus dans la tâche *Lextale* pour les participant-e-s de cette expérience étaient plus hauts que ceux des participant-e-s dans la première expérience. Nous allons revenir sur cette différence dans la conclusion. Toutefois, les scores *Lextale* n'étaient pour aucun segment prédictifs pour les temps de lecture.

6.1.2 Le design et la procédure

Nous avons utilisé le même design et matériel linguistique que dans la première expérience. La seule différence était que nous avons utilisé des phrases implicites (i.e., sans marquage explicite par un connecteur) au lieu de phrases incohérentes. Cela avait des conséquences différentes pour les deux relations de cohérence. L'omission du connecteur dans une phrase concessive telle que (2) rend la phrase incohérente. En revanche, l'omission du connecteur dans une phrase de conséquence telle que (3) ne produit pas d'incohérence car le lien causal peut être inféré en l'absence de connecteur (Sanders, 2005).

2) ?Jean ne veut pas connaitre les nouvelles du jour. Il lit le journal tous les matins.
3) ?Jean veut connaitre les nouvelles du jour. Il lit le journal tous les matins

Ainsi, matériel linguistique utilisé contenait, selon un design *2x2*, deux différents *Types de phrase* (soit *implicites, soit explicites*) et deux différents *Types de relation de cohérence* (soit *concessive*, soit *conséquence*).

La procédure était la même que dans l'expérience précédente.

6.3.1 L'analyse

Les analyses ont été effectués exactement comme dans la première expérience, telle que décrites au chapitre 6.2.3.

Les temps de lecture des segments 5, 6 et 7 (personnes natives)

L'ajout du *Type de phrase* (implicite, explicite) comme effet fixe a amélioré le modèle de manière significative pour les temps de lecture du segment 5 ($\chi(1) = 40.82$, $p = 1.67^{e-10}$). L'interaction avec le *Type de relation de cohérence* n'a pas amélioré le modèle davantage ($\chi(2) = .53$, $p = .77$). L'output du modèle final qui contenait le *Type de phrase* comme seul effet fixe (ainsi que les *Phrases expérimentales* et *Participant-e-s* comme effets aléatoires) est reporté dans le Tableau 6.5.

Tableau 6.5. Expérience 2. L'output du modèle final pour le segment 5 (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	- 1.12	.06	70.61	-20.44	$< 2^{e-16}$
Type (*implicite*)	.11	.02	2453.16	6.42	1.67^{e-10}

Au segment 5, les phrases implicites étaient donc lues plus lentement que les phrases explicites.

Pour les temps de lecture du segment 6, l'ajout du *Type de phrase* (implicite, explicite) comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 8.48$, $p < .005$). L'ajout de l'interaction avec le *Type de relation de cohérence* n'a pas amélioré le modèle davantage ($\chi(2) = 1.53$, $p = .47$). L'output du modèle final (contenant donc le *Type de phrase* comme seul effet fixe et les *Phrases expérimentales* et Participant-e-s comme effets aléatoires) est reporté dans le Tableau 6.6.

Tableau 6.6. Expérience 2. L'output du modèle final pour le segment 6 (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	- 1.09	.06	71.28	-18.29	$< 2^{e-16}$
Type (*implicite*)	-0.05	.02	2442.28	-2.91	< .005

Contrairement au segment 5, les phrases implicites étaient donc lues plus rapidement que les phrases explicites au segment 6.

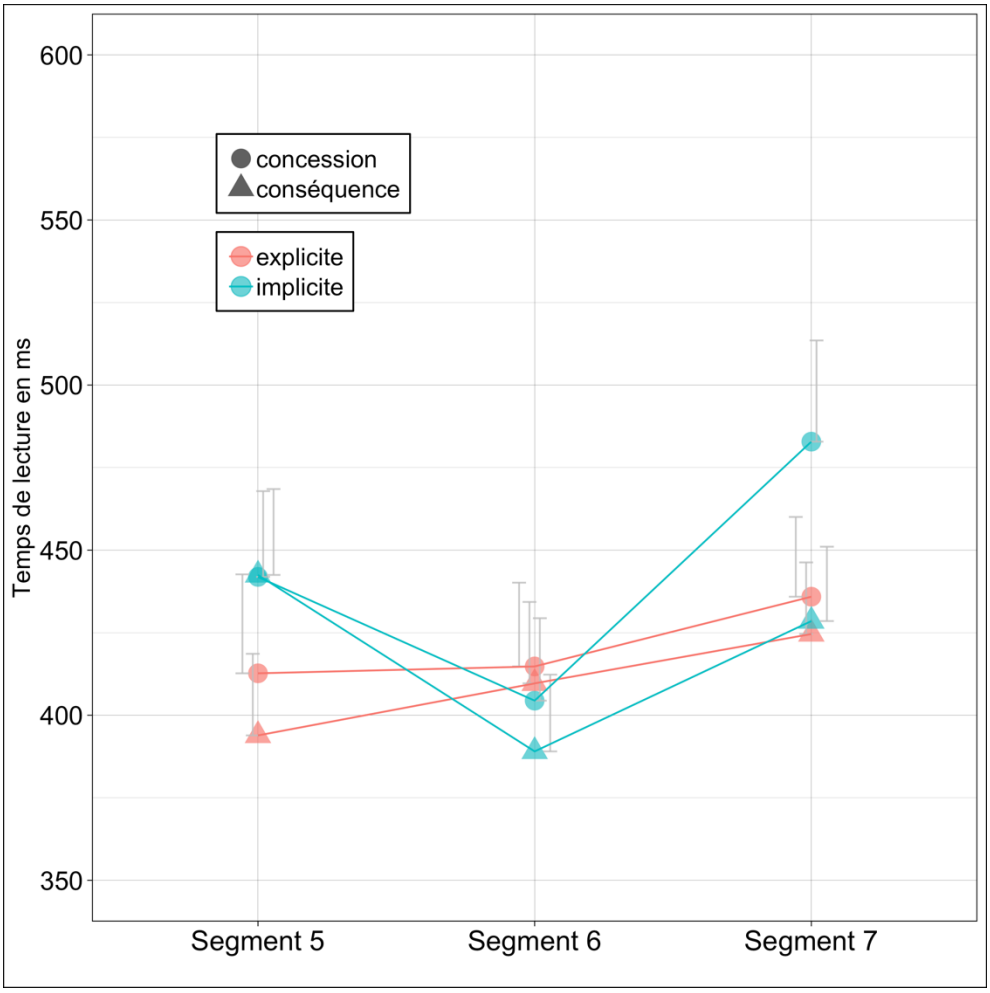
Pour le dernier segment de la phrase, le segment 7, l'ajout du *Type de phrase* comme seul effet fixe n'a pas amélioré le modèle de manière significative ($\chi(1) = 2.44$, $p = .12$). Par contre, le modèle s'est amélioré en ajoutant le *Type de relation de cohérence* comme seul effet fixe ($\chi(1) = 6.37$, $p < .05$). De plus, l'ajout de l'interaction avec le Type de phrase a amélioré le modèle davantage ($\chi(2) = 8.63$, $p < .05$). L'output du modèle final qui contenait donc l'interaction entre le *Type de phrase* et le *Type de relation de cohérence* (ainsi que la structure aléatoire) est reporté dans le Tableau 6.7.

Tableau 6.7. Expérience 2. L'output du modèle final pour le segment 7 (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-1.00^{e+00}	5.40^{e-02}	7.96^{e+01}	-18.58	$< 2^{e-16}$
Relation (*conséquence*)	-1.08^{e-03}	2.15^{e-02}	2.45^{e+03}	-0.05	.96
Type (*implicite*)	6.20^{e-02}	2.16^{e-02}	2.44^{e+03}	2.87	< .01
Relation (*conséquence*) : Type (*implicite*)	-7.56^{e-02}	3.05^{e-02}	2.44^{e+03}	-2.48	< .01

Les comparaisons *post-hoc* ont relevé que les phrases concessives implicites provoquaient des temps de lecture plus longs que toutes les autres phrases (par rapport aux concessions explicites : $\beta = .06$, $ET = .02$, $z = 2.87$, $p < .05$; par rapport aux conséquences implicites : $\beta = .01$, $ET = .02$, $z = -0.6$, $p < .01$; et par rapport aux conséquences explicites : $\beta = .06$, $ET = .02$; $z = -2.91$, $p < .05$). Ces effets sont également visualisés dans la Figure 6.3.

Figure 6.3. Expérience 2. Les temps de lecture des segments 5, 6 et 7 (personnes natives, limites supérieures de l'IC 95% indiquées par des barres d'erreur)



Nous observons donc une accélération pour les phrases qui ont été marquées avec un connecteur dans le segment 5, ainsi qu'un ralentissement pour les phrases concessives implicites dans le segment 7.

Les temps de réponses aux questions de vérification (personnes natives)

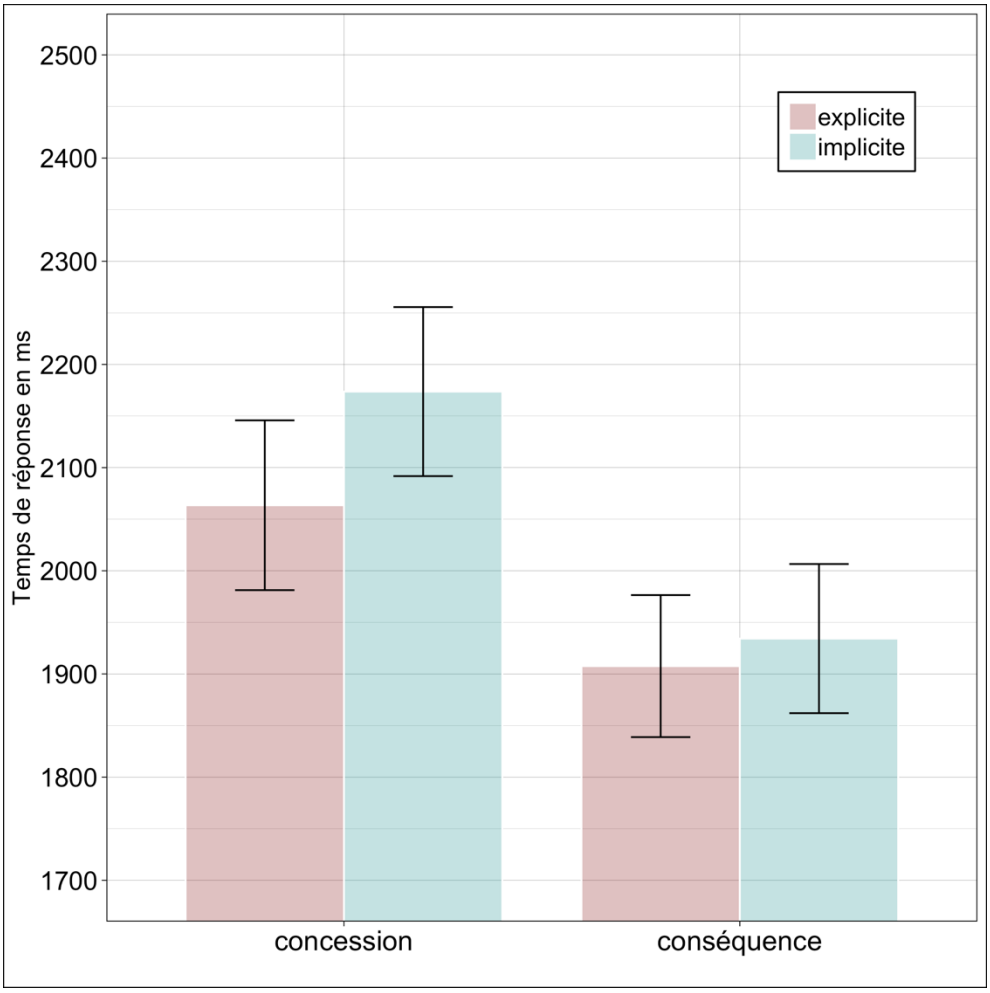
Pour les temps de réponses aux questions de vérification, l'ajout du *Type de phrase* (implicite, explicite) comme effet fixe a amélioré le modèle ($\chi(1) = 7.55$, $p < .01$). De plus, l'interaction avec le *Type de relation de cohérence* a amélioré le modèle davantage ($\chi(2) = 50.27$, $p = 1.22^{e-11}$). L'output du modèle final contenant l'interaction entre le *Type de phrase* et le *Type de relation de cohérence* (ainsi que la structure aléatoire) est reporté dans le Tableau 6.8.

Tableau 6.8. Expérience 2. L'output du modèle final pour les temps de réponse aux questions de vérification (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	1.07	.02	111.50	43.95	$< 2^{e-16}$
Relation (*conséquence*)	-0.04	.01	2461.88	-3.48	< .001
Type (*implicite*)	.04	.01	2463.13	3.36	< .001
Relation (*conséquence*) : Type (*implicite*)	-0.03	.02	2461.96	-1.95	.05

Comme pour les temps de lecture du segment 7, les comparaisons *post-hoc* indiquent que les phrases concessives implicites provoquaient des temps de réponses plus longs en comparaison avec toutes les autres phrases (par rapport aux concessions explicites : $\beta = .04$, $ET = .01$, $z = 3.36$, $p < .01$; par rapport aux conséquences implicites : $\beta = .08$, $ET = .01$, $z = - 6.2$, $p < .001$; et par rapport aux conséquences explicites : $\beta = .08$, $ET = .01$, $z = -6{,}82$, $p < .001$). De plus, en comparant les temps de réponses pour les phrases explicites, les comparaisons *post-hoc* indiquent une différence significative en fonction de la relation de cohérence, de sorte que les phrases concessives suscitaient un temps de réponse plus long que les phrases de conséquence ($\beta = -0.04$, $ET = .01$, $z = -3.48$, $p < .01$). Ces effets sont également illustrés dans la Figure 6.4.

Figure 6.4. Expérience 2. Les temps de réponses aux questions de vérification (personnes natives)



Les temps de lecture des segments 5, 6 et 7 (personnes non-natives)

Pour les temps de lecture du segment 5, l'ajout du *Type de phrase* (*implicite*, *explicite*) comme effet fixe a significativement amélioré le modèle ($\chi(1) = 5.24$, $p = .02$). L'ajout de l'interaction avec le *Type de relation de cohérence* (*concession, conséquence*) n'a pas amélioré le modèle davantage ($\chi(3) = 6.32$, $p = .10$).[24] L'output de ce modèle est reporté dans le Tableau 6.9.

[24] Les moyennes visualisées dans la Figure 6.5 pourraient faire croire à un effet pour le *Type de relation* au segment 5. Pourtant, cela n'est pas le cas : l'ajout du *Type de relation* comme seul effet fixe n'a pas amélioré le modèle ($\chi(1) = .31$, $p = .58$).

Tableau 6.9. Expérience 1. L'output du modèle final pour segment 5 (personnes non-natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.66	.06	75.93	-11.67	$< 2^{e-16}$
Type (*implicite*)	.04	.02	1860.28	2.29	.02

Comme pour les personnes natives, nous constatons donc que les phrases implicites ont été lues plus lentement que les phrases explicites.

Pour le segment 6, le modèle s'est amélioré lors de l'ajout du *Type de phrase* comme effet fixe ($\chi(1) = 15.61$, $p = 7.79^{e-05}$), tandis que l'ajout de l'interaction avec le *Type de relation de cohérence* n'a pas amélioré le modèle davantage ($\chi(2) = 1.05$, $p = .59$). L'output du modèle final du segment 6 (contenant donc le *Type de phrase* comme seul effet fixe, ainsi que les *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires) est reporté dans le Tableau 6.10.

Tableau 6.10. Expérience 1. L'output du modèle final pour segment 6 (personnes non-natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.56	.06	70.77	-9.19	1.07^{e-13}
Type (*implicite*)	-0.07	.92	1865.31	-3.96	7.81^{e-05}

De même que chez les personnes natives, nous observons que les phrases implicites ont été lues plus rapidement que les phrases explicites au segment 6 par les personnes non-natives.

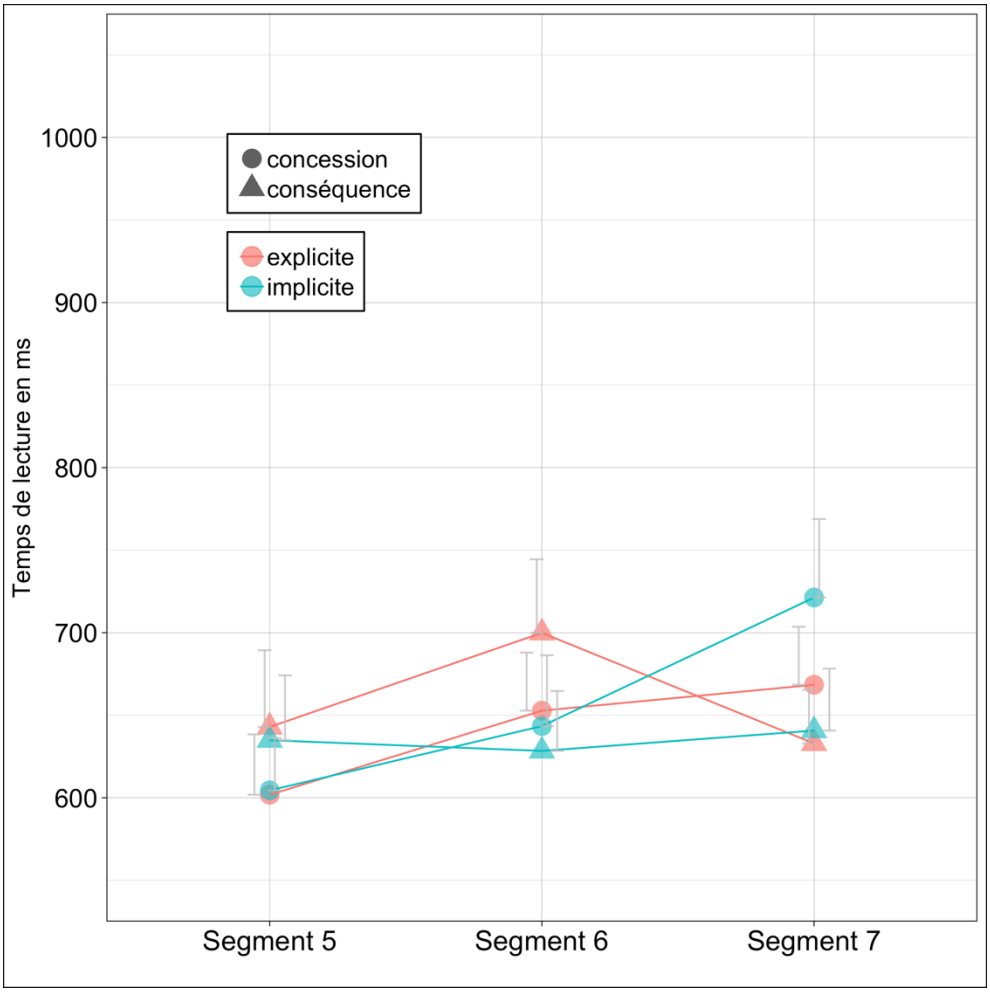
Pour le dernier segment de la phrase, le segment 7, l'ajout du *Type de phrase* comme effet fixe n'a pas amélioré le modèle ($\chi(1) = .12$, $p = .73$). En revanche, l'ajout du *Type de relation de cohérence* comme seul effet fixe a amélioré le modèle ($\chi(1) = 15.95$, $p = 6.52^{e-05}$). Le modèle ne s'est pas amélioré davantage lors de l'ajout de l'interaction avec le *Type de phrase* ($\chi(2) = .64$, $p = .73$). L'output du modèle final du segment 7, contenant donc (à part de la structure aléatoire) le *Type de relation de cohérence* commet seul effet fixe est reporté dans le Tableau 6.11.

Tableau 6.11. Expérience 2. L'output du modèle final pour segment 7 (personnes non-natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.52	.06	68.65	-9.31	8.68^{e-14}
Relation (*conséquence*)	-0.06	.02	1866.51	-4.00	6.53^{e-05}

Comme l'illustre la Figure 6.5, les personnes non-natives ont lu le segment 7 plus lentement lorsque les phrases étaient concessives, indépendamment de la présence ou de l'absence d'un connecteur.

Figure 6.5. Expérience 2. Les temps de lecture pour les segments 5, 6 et 7 (personnes non-natives, limites supérieures de l'IC 95% indiquées par des barres d'erreur)



Les temps de réponses aux questions de vérification (personnes non-natives)

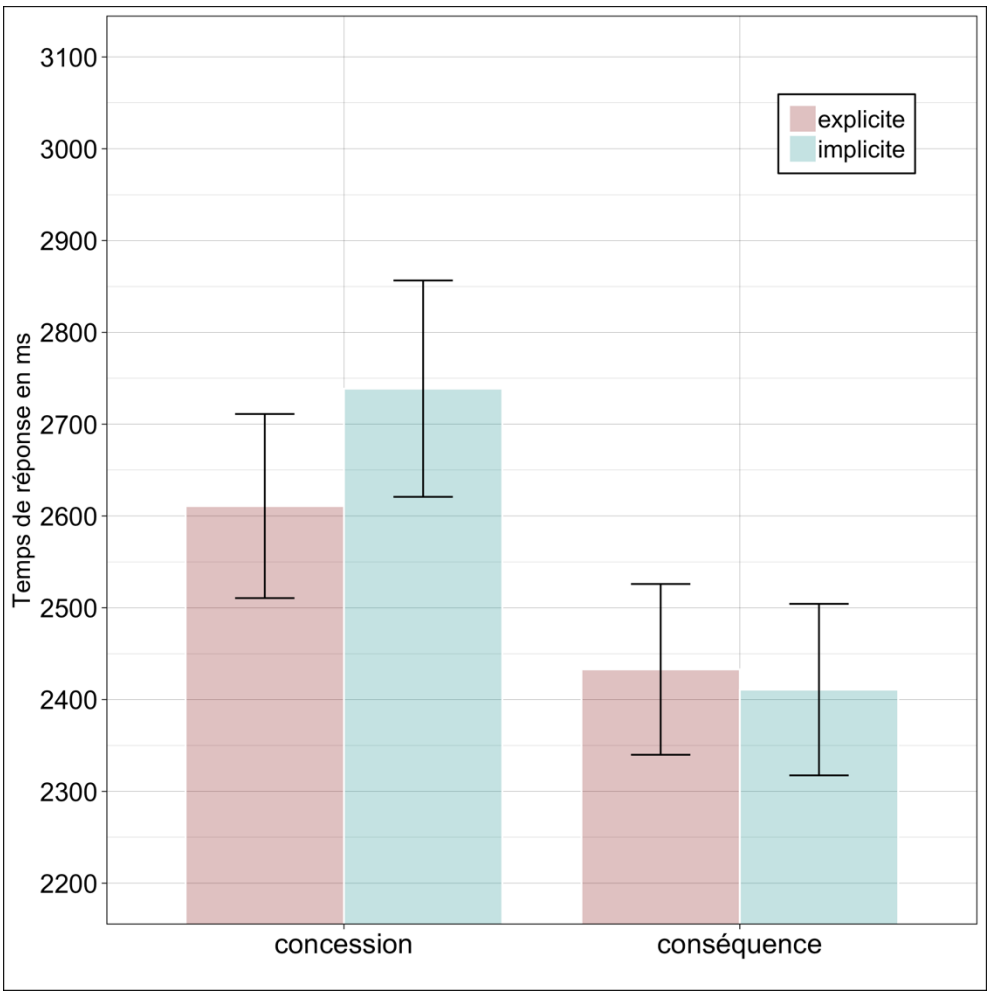
Pour les temps de réponses aux questions de vérification, l'ajout du *Type de phrase* comme effet fixe n'a pas amélioré le modèle ($\chi(1) = 1.21$, $p = .27$). Le modèle s'est amélioré lors de l'ajout du *Type de relation de cohérence* ($\chi(1) = 40{,}27$, $p = 2.21^{e-10}$). L'ajout de l'interaction avec le *Type de phrase* n'a pas amélioré le modèle davantage ($\chi(2) = 5.06$, $p = .08$). L'output du modèle qui contenait le *Type de phrase* comme seul effet fixe (ainsi que les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires) est reporté dans Tableau 6.12.

Tableau 6.12. Expérience 2. L'output du modèle final pour les temps de réponses aux questions de vérification (personnes non-natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	2.92	0.15	86.26	18.93	$< 2^{e-16}$
Relation (*conséquence*)	-0.34	.10	1882.13	-3.35	8.25^{e-04}

Nous observons que les personnes non-natives étaient toujours plus lentes à répondre aux questions relatives aux phrases introduites par des concessions que par des conséquences, indépendamment du fait que les phrases soient explicitement ou implicitement marquées. Les temps de réponse sont également illustrés dans la Figure 6.6.

Figure 6.6. Expérience 2. Les temps de réponse aux questions de vérification (personnes non-natives)



6.3.2 Discussion

Dans cette expérience, nous avons testé si les personnes non-natives sont également sensibles à l'incohérence lorsque celle-ci est créé par l'absence de connecteur. A cette fin, nous avons mesuré les temps de lecture des phrases de conséquences et de concession qui ont été présentées soit avec un connecteur, soit simplement juxtaposées. Dans le cas d'une concession, l'absence de connecteur crée de l'incohérence tandis qu'une conséquence peut être comprise même sans connecteur (Murray, 1997; Sanders, 2005).

Premièrement, les résultats démontrent que les personnes natives ont été affectées par l'incohérence dans les phrases concessives sans connecteur. Nous avons observé un ralentissement pour cette condition dans le dernier segment de la phrase ainsi que dans les réponses aux questions de vérification. Cela soutient donc l'hypothèse qu'une concession non-marquée crée une incohérence qui n'est pas facilement résolue (Murray, 1994). En revanche, nous n'avons pas observé une telle réaction à l'incohérence pour les personnes non-natives. A aucun moment de la lecture (ni lors de la réponse aux questions de vérification), les personnes non-natives montraient une perturbation du traitement mental pour les concessions implicites. Cela est curieux, tenant compte du fait que les personnes non-natives dans la première expérience réagissaient immédiatement à l'incohérence. Comme cette fois l'incohérence a été créé par l'absence de connecteur plutôt que par une mauvaise utilisation, il semble que les personnes non-natives manquaient de sensibilité pour détecter l'incohérence. Il convient de souligner à cet égard que les participant-e-s de cette expérience étaient même linguistiquement plus compétent-e-s que ceux de la première expérience. Cette différence ne peut donc pas s'expliquer par un manque de compétences linguistiques.

Deuxièmement, nous avons observé que la lecture des personnes non-natives a été fortement affectée par la relation de cohérence. Les relations de conséquence ont été lues plus rapidement que les concessions au segment 7 et ont suscité des temps de réponses plus rapides aux questions de vérification. Cette différence reflète les différences de complexité cognitive entre les deux relations testées, à savoir le fait qu'une conséquence est plus facile à traiter qu'une concession (Sanders et al., 1992; Murray, 1997; Sanders, 2005).

Finalement, tous les participant-e-s lisaient les phrases explicites plus rapidement que les phrases implicites au segment 5. Une explication pour cette observation serait que la présence d'un connecteur a temporairement accéléré la lecture, même pour les personnes non-natives. L'effet inverse dans le segment 6 (c'est-à-dire que les phrases explicites ont été lues plus lentement que les phrases implicites) s'expliquerait ainsi par le besoin d'intégrer la signification du connecteur. Une autre explication pour ces effets serait que les deux clauses dans la condition explicite faisaient partie de la même phrase, tandis que les deux clauses dans la condition implicite étaient séparées par un point et formaient donc des phrases à part entière. Il se pourrait donc que les temps de lecture élevés pour la condition explicite ne reflètent pas un effet d'accélération pour cette condition, mais plutôt le temps nécessaire à la création d'une représentation mentale de la première clause dans la condition implicite (achevée en tant que phrase autonome).

Quoi qu'il en soit, en résumé nous avons vu que, contrairement aux personnes natives, les personnes non-natives n'étaient pas affectées par les connecteurs manquants dans des phrases concessives. Elles ont seulement montré des temps de lecture plus longs pour les phrases concessives en comparaison avec les phrases de conséquence.

6.4 Conclusion

Au cours de deux expériences, nous avons examiné la sensibilité des personnes non-natives aux connecteurs discursifs. A l'aide d'une tâche de lecture segmentée, nous avons mesuré la perturbation de la lecture des phrases concessives et de conséquence qui contenaient des connecteurs incorrects (Expérience 1) et manquants (Expérience 2).

De manière générale, nous concluons que, malgré l'effort mental élevé lié au fait de lire dans une L2 (Clahsen & Felser, 2006; Recio Fernández, 2020) les personnes non-natives peuvent être très sensibles aux instructions fournies par les connecteurs. Plus précisément, nous avons observé dans la première expérience que les personnes non-natives réagissaient très tôt aux mauvaises utilisations des connecteurs. Nous avons également observé que la sensibilité des personnes non-natives est toutefois un peu moins prononcée lorsque l'incohérence est moins saillante. Lorsque l'incohérence provient de l'absence de connecteur, le traitement

mental des personnes non-natives n'était pas du tout affecté. En d'accord avec l'hypothèse de Papadopoulou & Clahsen (2003) selon laquelle les personnes non-natives dépendent plus des formes lexicales que fonctionnelles, cette observation pourrait indiquer que les personnes non-natives, bien qu'elles soient sensibles aux connecteurs, y font moins recours que les personnes natives.

Toutefois, le fait que la sensibilité était moins prononcée dans la deuxième expérience pourrait s'expliquer par une légère différence de design. Lors de l'Expérience 1, les participant-e-s lisaient deux conditions incohérentes (*conséquence incohérente* et *concession incohérente*), alors que dans l'Expérience 2, seulement une condition était incohérente (*concession implicite*). Il n'est toutefois pas certain que cela ait entrainé un autre type de lecture, par exemple de sorte que les participant-e-s de la deuxième expérience s'attendaient plus à lire des phrases cohérentes. Premièrement, dans les deux expériences, nous avons testé des phrases de distraction qui étaient également incohérentes. Deuxièmement, si le nombre réduit de phrases incohérentes dans l'Expérience 2 avait un effet, l'effet attendu serait que l'incohérence serait encore plus saillante. Cependant, nous n'avons observé aucune réaction à l'incohérence, contrairement à la première expérience. Troisièmement, le groupe de contrôle natif a réagi très fort à l'incohérence dans les relations concessives implicites, même s'il s'agissait de la seule condition incohérente. Et quatrièmement, les participant-e-s de la deuxième expérience étaient linguistiquement plus compétent-e-s, ce qui aurait pu niveler l'effet potentiel. Cela souligne d'ailleurs que le niveau de compétence linguistique d'un-e apprenant-e est n'est pas le seul facteur responsable de la maitrise des connecteurs, mais qu'il s'agit plutôt d'une interaction entre des facteurs linguistiques et cognitifs, telle que la complexité de la relation de cohérence.

Enfin, les résultats de nos expériences indiquent que la complexité de la relation de cohérence a une influence sur le traitement mental en L2. Nous avons observé des temps de réponse aux questions de vérification plus longs pour les concessions incohérentes que pour les conséquences incohérentes lors de la première expérience. Encore plus distinctement, lors de la deuxième expérience, les phrases concessives (cohérentes) provoquaient des temps de lecture plus longs au dernier segment de la phrase, ainsi que des temps de réponses aux questions de vérification plus longs que les phrases de conséquence (cohérentes).

En résumé, on peut conclure que les personnes non-natives sont capables d'activer les fonctions procédurales des connecteurs lors du traitement mental d'une phrase, pourvu que les connecteurs soient bien maitrisés (voir aussi Degand & Sanders, 2002). Cependant, nous avons également observé que cela ne veut pas dire que les personnes non-natives arrivent automatiquement à remarquer quand un connecteur est nécessaire pour que la relation soit cohérente. A cet égard, leur sensibilité est moins prononcée que celle des personnes natives.

Ces résultats indiquent donc que la catégorie des connecteurs n'est pas difficile en soi pour des apprenant-e-s et que la maitrise insuffisante des connecteurs en L2 (discutée au chapitre 3.3) dépend, en fait, de facteurs linguistiques spécifiques, tel que les caractéristiques du connecteur en question. Dans le chapitre suivant, nous examinerons plus en détail comment un connecteur plus complexe réduit la sensibilité à l'incohérence d'une personne native lors de son traitement mental par rapport à un connecteur plus simple.

7 L'impact du type de connecteur sur le traitement mental en L2

Ce chapitre est motivé par plusieurs observations faites dans les chapitres précédents. La première observation est que les différents connecteurs sont inégalement maitrisés par les apprenant-e-s d'une L2 lorsqu'ils accèdent à leurs connaissances explicites (chapitre 5). Tandis que des connecteurs comme *néanmoins* semblent maitrisés dans la tâche d'insertion à un niveau similaire à celui des personnes natives, d'autres comme *cependant* ont suscité des scores considérablement plus bas. La question se pose inévitablement de savoir si les différences dans la maitrise de différents connecteurs se manifestent également lorsque les personnes non-natives accèdent à leurs connaissances implicites. De manière plus générale, nous avons déjà démontré que les personnes non-natives sont en principe sensibles aux connecteurs, bien que cette sensibilité soit moins prononcée que celle des personnes natives (chapitre 6). Ainsi, il est possible que les personnes non-natives soient plus sensibles à certains connecteurs qu'à d'autres lorsqu'elles lisent un discours.

Pour les personnes natives, nous avons observé un tel effet: nous avons observé que les connecteurs moins fréquents rendent la détection de l'incohérence plus difficile (chapitre 4). Si même les personnes natives sont affectées par les connecteurs difficiles, on peut supposer qu'il en va de même pour les personnes non-natives.

Malheureusement, nous manquons, jusqu'à présent, des données pour vérifier cette hypothèse : au chapitre 4, nous n'avons testé que des personnes *natives*; au chapitre 5, nous n'avons testé que les connaissances *explicites* et au chapitre 6, nous n'avons testé que des connecteurs *fréquents* (et dont on peut faire l'hypothèse qu'ils sont généralement bien maitrisés par les personnes non-natives). Ainsi, dans ce chapitre, nous allons tester la connaissance *implicite* des personnes *non-natives* pour un connecteur *difficile*. En particulier, nous nous intéressons à la question de savoir si ce connecteur complique la détection de l'incohérence en comparaison avec un connecteur plus accessible – similaire à ce que nous avons

observé pour les personnes natives (chapitre 4). Pour répondre à cette question, nous allons mener deux expériences.

Dans la première, nous testerons la sensibilité à l'incohérence des personnes non-natives en utilisant une tâche de lecture segmentée. Nous mesurerons les temps de lecture pour des phrases cohérentes et incohérentes qui sont indiquées soit par un connecteur dont on peut supposer qu'il est bien maitrisé, soit par un connecteur censé être plus difficile pour les personnes non-natives. Comme nous allons comparer cette fois directement les temps de lecture pour des phrases marquées avec des connecteurs, nous utilisons une méthode légèrement différente de celle employée dans le chapitre 4. Tandis que la comparaison directe nous permet de déterminer sans équivoque l'impact du connecteur sur la lecture des personnes non-natives, cela n'a pas été nécessaire pour les personnes natives du chapitre 4, car nous pouvions supposer que les personnes natives connaissaient les fonctions de tous les connecteurs testés.[25]

Pour les personnes non-natives, la situation est aussi quelque peu différente parce que l'interprétation des effets potentiellement émergeants est plus difficile. Par exemple, un ralentissement de la lecture peut refléter différentes choses: d'une part, elle pourrait signifier que les participant-e-s mettent plus de temps à comprendre le connecteur. D'autre part, elle pourrait également indiquer qu'ils ne le comprennent pas du tout et sont perturbés par leur incompréhension. Pour faire face à ce problème d'interprétation, nous allons mener une deuxième expérience dans laquelle nous demanderons aux participant-e-s d'évaluer des phrases. Ainsi, nous contrôlerons si les effets potentiels observés dans la première expérience sont dus à un manque de connaissance des connecteurs testés.

Avant de faire ceci, nous allons discuter les raisons pourquoi nous pensons qu'un connecteur difficile pourrait affecter le traitement mental d'une personne non-native.

[25] Au chapitre 4, cela nous a permis d'inclure une autre variable (celle de la relation de cohérence) et de faire des comparaisons au travers des trois expériences.

7.1 Est-ce qu'un connecteur moins accessible aux personnes non-natives réduit leur sensibilité à l'incohérence ?

Lorsqu'on lit dans une L2, on a besoin d'un effort mental plus élevé en comparaison avec la lecture dans sa L1 (Clahsen & Feller, 2006, Recio Fernández, 2020). Nous l'avons nous-mêmes observé : à part tous les effets observés au chapitre 6, les personnes non-natives lisaient les phrases beaucoup plus lentement que les personnes natives. Cependant, la lecture en L2 peut être facilitée par des connecteurs. Non seulement il est connu que les connecteurs peuvent aider les personnes non-natives à mieux comprendre un discours (Degand & Sanders, 2002), mais nous avons également observé des effets de facilitation qui se produisent lors du traitement mental (chapitre 6).

Est-ce qu'on peut en conclure que les connecteurs sont toujours une aide pour la lecture dans une deuxième langue ? La réponse n'est pas si évidente : comme nous l'avons vu, tous les connecteurs ne sont pas bien maitrisés (chapitre 5). Ainsi, on peut supposer que plusieurs d'entre eux n'exercent pas un effet de facilitation pour la lecture en L2. Par ailleurs, il y a des indications dans la littérature qui soutiennent cette hypothèse. Zufferey et Gygax (2017), par exemple, ont démontré que des personnes non-natives n'étaient pas aussi perturbées par l'incohérence générée par l'absence du connecteur *en effet* que des personnes natives. De plus, dans une tâche d'évaluation des phrases, elles considéraient les phrases marquées avec ce connecteur comme moins cohérentes que celles qui ne le contenaient pas, bien que cette absence rendait ces phrases incohérentes. Zufferey et Gygax (2017) expliquent leurs résultats par le fait que les participant-e-s ne maitrisaient pas les fonctions du connecteur *en effet*. Selon les auteur-trice-s, sa difficulté réside dans sa polyfonctionnalité et dans le fait que ses fonctions ne correspondent que peu aux fonctions des connecteurs apparemment équivalents de la L1.[26]

Bien que ces observations indiquent que la simple présence d'un connecteur ne facilite pas automatiquement la lecture en L2, l'étude ne permet pas de déterminer dans quelle mesure un connecteur complexe influence le traitement mental en L2. En effet, dans cette étude, l'incohérence n'a été obtenue que par

[26] Une autre explication serait un transfert négatif d'une L3, d'anglais, (p.ex., angl. *in fact* ['en fait']), qui créerait encore plus de confusion sur les fonctions de *en effet* (voir chapitre 8 pour les effets crosslinguistiques qui rendent certains connecteurs plus difficiles pour les personnes non-natives).

l'absence d'un connecteur. Pour déterminer si la lecture en L2 dépend du type du connecteur, il convient plutôt de mesurer la réaction des personnes non-natives à des phrases incohérentes (voir aussi chapitres 4 et 7) qui sont marquées par un connecteur facile ou difficile. Compte tenu de la difficulté de lire dans une deuxième langue (Clahsen & Felser, 2006), un connecteur plus complexe, tel que *en effet*, pourrait créer une confusion de sorte qu'une incohérence serait plus lentement (voire, pas du tout) détectée. En outre, une fois l'incohérence identifiée, les connecteurs complexes peuvent, d'une certaine manière, déstabiliser la confiance des personnes non-natives : au lieu de rejeter une phrase incohérente, elles pourraient plutôt remettre en question leurs propres connaissances sur le connecteur.

Pour combler ce vide dans la littérature et déterminer l'impact du type du connecteur sur le traitement mental dans une L2, nous allons donc tester si un connecter complexe complique la détection de l'incohérence. Avant de faire ceci, il convient de s'interroger sur les caractéristiques d'un connecteur qui pourraient rendre la détection d'incohérences plus difficile. La polyfonctionnalité (telle que décrite par Zufferey & Gygax, 2017) ne peut pas être le seul facteur à cet égard, car les personnes non-natives au chapitre 6 étaient sensibles à l'incohérence pour des phrases qui contenaient le connecteur polyfonctionnel *mais*.

Pour classer un connecteur comme *difficile* pour les personnes non-natives, nous avons établi sept critères que nous allons maintenant présenter. Nous présentons ces critères en prenant l'exemple de deux connecteurs, *mais* et *cependant*, dont nous supposons qu'ils se distinguent par leur complexité pour les personnes non-natives.

7.2 Analyse de la complexité des connecteurs *mais* et *cependant*

Dans ce chapitre, nous nous intéresserons à la question de savoir si un connecteur plus complexe réduit la sensibilité des personnes non-natives à l'incohérence lors de la lecture. Cela soulève la question de savoir quels facteurs rendent un connecteur plus ou moins complexe. Nous allons discuter cela en prenant l'exemple de deux connecteurs : *mais* et *cependant.* Nous allons maintenant évaluer leur complexité en fonction de différents critères.

Le critère de la relation de cohérence

Le premier critère est celui de la relation de cohérence. Nous avons observé que pour les deux groupes linguistiques, natif et non-natif, la relation de cohérence était un facteur décisif qui influençait la vitesse de lecture (chapitre 6). Toutefois, pour *mais* et *cependant*, ce critère peut être négligé car les deux indiquent la même relation de cohérence : une concession (selon le dictionnaire des connecteurs *LexConn*, Roze et al., 2012).

Le critère de la transparence sémantique

Nous avons discuté au chapitre 5 que la transparence sémantique d'un connecteur, telle qu'observée pour le connecteur *c'est pourquoi* par exemple, peut faciliter sa maitrise en L2. Toutefois, comme pour le critère de la relation de cohérence, *cependant* et *mais* ne se distinguent pas selon ce critère, car les deux ne sont pas sémantiquement transparents.

Le critère de la fréquence

Bien que les connecteurs fréquents n'étaient pas systématiquement mieux maitrisés par les apprenante-e-s au chapitre 5, nous avons montré que les connecteurs fréquents influencent positivement la réaction à l'incohérence des personnes natives lors du traitement mental (chapitre 4). Nous supposons qu'un connecteur fréquent peut avoir le même effet pour les personnes non-natives.

Pour déterminer la fréquence de *mais* et de *cependant*, nous avons compté leurs occurrences à l'aide de la plateforme *SketchEngine* (Kilgarriff et al., 2014) et plus précisément du corpus *FrenchWeb 2017* (*frTenTen17*, plus de 5 milliards de mots, Jakubíček et al., 2014). Les fréquences des deux connecteurs diffèrent considérablement : le connecteur *mais* se trouve plus que 16 millions fois dans le corpus, tandis que *cependant* n'a que 1.18 millions d'occurrences. Suivant le critère de la fréquence, nous supposons donc que *mais* devrait être plus facile à maitriser que *cependant* pour une personne non-native.

Le critère de la modalité

La fréquence globale d'un connecteur peut être encore différenciée en regardant s'il est plutôt utilisé à l'écrit ou à l'oral. L'idée est que les connecteurs qui apparaissent plutôt dans la langue écrite sont plus difficiles à maitriser que les connecteurs qui sont plutôt utilisés à l'oral. Nous avons déterminé la fréquence de *mais* et *cependant* dans les corpus oraux et écrits dans la collection de corpus ORFEO (Benzitoun et al., 2016). Les nombres d'occurrences sont reportés dans le Tableau 7.1.

Tableau 7.1. Les nombres d'occurrences de *mais* et *cependant* selon leur modalité dans la collection des corpus ORFEO

	Nombre d'occurrences dans les…		
	…corpus oraux	*…corpus écrits*	*ratio*
mais	882	1143	1 : 1.23
cependant	13	212	1 : 15.14

Nous y voyons la tendance à utiliser *cependant* plus à l'écrit qu'à l'oral, tandis que *mais* semble être plus universel.

Le critère du registre

Nous avons observé au chapitre 5 qu'un registre élevé peut rendre un connecteur plus difficile (p.ex., *en revanche*), respectivement qu'un registre plus bas (p.ex., *par contre*), peut le rendre plus facile à maitriser. Pour déterminer dans quels registres *mais* et *cependant* sont le plus utilisés, nous avons classifié les différents types de corpus (dans lesquels *mais* et *cependant* se trouvent) dans la collection de corpus ORFEO (Benzitoun et al., 2016) indépendamment de la modalité selon leur registre. Selon notre classification, les corpus provenant d'un registre élevé ont les suivants:

- presse écrite
- écrit scientifique
- littérature
- discours
- entretien
- médias oraux
- présentation publique

Les corpus d'un registre moyen ou bas incluent :

- écrit non-planifié
- conversation
- explication

- transaction
- consultation
- activité
- repas

Les nombres d'occurrences de *mais* et *cependant* selon le registre du corpus sont reportés dans le Tableau 7.2.

Tableau 7.2. Nombres d'occurrences de *mais* et *cependant* selon le registre dans les corpus d'ORFEO

	Nombre d'occurrences dans les corpus de …		
	… registre élevés	*…registre moyen ou bas*	*ratio*
mais	1537	486	3.16 : 1
cependant	215	11	19.55 : 1

Nous y constatons que *cependant* est presque toujours utilisé dans un registre élevé tandis que *mais* est plus souvent utilisé dans un registre plus bas. On peut donc supposer que, selon le critère du registre, *cependant* est plus difficile que *mais* pour les personnes non-natives.

Le critère de la polyfonctionnalité

Zufferey et Gygax (2017) ont montré les difficultés des apprenante-e-s du français avec le connecteur polyfonctionnel *en effet*. Ainsi, les auteur-trice-s ont conclu que la polyfonctionnalité est un critère qui peut rendre un connecteur plus difficile pour une personne non-native (voir aussi 4.1.2).

Selon le dictionnaire *LexConn* (Roze et al., 2012), *mais* est polyfonctionnel car il peut indiquer soit des concessions, soit des contrastes (voir aussi Crible et al., 2021). En revanche, *cependant* indique seulement des concessions. Ainsi, selon le critère de la polyfonctionnalité, *mais* peut être considéré étant plus complexe que *cependant* pour les personnes non-natives. Toutefois, il convient de noter que la polyfonctionnalité de *mais* ne correspond pas tout à fait à celle de, par exemple, *en effet*. Les relations de cohérence que peut indiquer *mais* (i.e., contraste, concession) sont très proches l'une à l'autre et ne sont même pas séparées dans de nombreux modèles de discours.

Résumé

Après avoir analysé les connecteurs *mais* et *cependant* selon six critères, nous résumons nos classifications dans le Tableau 7.4.

Tableau 7.4. Résumé de l'analyse de la complexité des connecteurs *mais* et *cependant*

facteur	*mais*	*cependant*
transparence sémantique	opaque	opaque
relation de cohérence	complexe	complexe
fréquence	haute	basse
registre	universel	élevé
modalité	pas de préférence	plutôt à l'écrit
fonctionnalité	polyfonctionnel	monofonctionnel

Bien que *mais* soit polyfonctionnel, d'autres critères nous laissent croire que *mais* est plus facile à maitriser pour les personnes non-natives que le connecteur *cependant*. En effet, *mais* est utilisé extrêmement fréquent et émerge à la fois à l'écrit qu'à l'oral ainsi que dans des registres divers. En revanche, *cependant* semble être lié plutôt à un registre élevé, émerge en conséquence plutôt à l'écrit et est généralement beaucoup moins fréquent. Les critères de la relation de cohérence et de la transparence sémantique - deux critères pour lesquels une influence sur la maitrise en L2 a été observée - ne sont pas décisifs, car les deux connecteurs ne diffèrent pas selon ces critères. Prises ensemble, nos analyses indiquent donc que *mais* devrait être plus facile à maitriser que *cependant* par les personnes non-natives et que cela est principalement dû aux différences de fréquence, de modalité et de registre.

Cette classification est d'ailleurs également étayée par des observations dans les expériences précédentes. Par exemple, les participant-e-s des expériences dans les chapitres 6.2 et 6.3 ont montré très tôt des perturbations du traitement mental lorsque les phrases contenaient des usages incorrects de *mais*. Les fonctions procédurales de *mais* semblent être donc intériorisées par les personnes non-natives. En revanche, il y a des indications que cela n'est pas forcément le cas pour *cependant* : nous avons vu que le connecteur *cependant* a suscité des scores extrêmement bas dans le chapitre 5.

Basé sur ces observations et les analyses menées, nous formulons donc l'hypothèse selon laquelle l'utilisation de *mais* dans une phrase facilite la détection

de l'incohérence, tandis que l'utilisation de *cependant* devrait rendre cette détection un peu plus difficile. Dans l'expérience suivante, nous allons tester cette hypothèse.

7.3 Expérience 1 : la réaction à l'incohérence en fonction du type du connecteur lors du traitement mental

Dans cette expérience, nous allons mesurer les temps de lecture pour des phrases concessives qui sont cohérentes ou incohérentes. Tandis que certaines des phrases contiennent le connecteur *mais*, d'autres contiennent le connecteur *cependant*. Comme nous venons de le voir, on peut considérer que *mais* représente un exemple de connecteur facile tandis que *cependant* parait plus difficile à maitriser dans une L2.

De manière générale, nous supposons que l'incohérence devrait affecter le traitement mental des deux groupes (natif et non-natif), de sorte que nous allons observer des ralentissements pour des phrases incohérentes. Comme dans les expériences précédentes, nous attribuons ces ralentissements de la lecture aux perturbations du traitement mental.

Nous nous attendons toutefois à observer une différence entre les deux groupes. Pour les participant-e-s non-natifs, nous nous attendons également à un effet du connecteur : nous supposons que l'incohérence dans une phrase marquée avec *cependant* ne provoque pas autant de perturbation du traitement mental qu'une phrase qui est marquée avec *mais*. La raison en est que ces participant-e-s n'ont pas intériorisé la fonction procédurale de ce connecteur aussi bien que pour *mais*. Pour le groupe natif, il n'est pas clair si nous allons observer un tel effet : d'un côté, nous avons vu dans le chapitre 4 que les connecteurs rares peuvent ralentir la réaction à l'incohérence des personnes natives. D'un autre côté, il n'est pas clair si le connecteur testé dans cette expérience (*cependant*) provoque les mêmes effets que les connecteurs testés dans le chapitre 4 (*néanmoins* et *or*).

7.3.1 Les participant-e-s

Pour le groupe natif, nous avons recruté 71 personnes natives via la plateforme *Prolific* (Prolific, London, 2020). Ces participant-e-s ont indiqué avoir grandi avec une seule langue. Une personne a été enlevée à pour cause d'un taux de réponses

correctes trop bas aux contrôles d'attention (réponses inférieures au niveau de la chance, voir ci-dessous pour la tâche). Donc, pour le groupe contrôle, les données de 70 personnes ont été analysées (âge moyen = 31.65 ans, écart type = 11.24, 30 femmes). Pour le groupe non-natif, 68 personnes non-natives avec l'anglais comme L1 ont été recrutées dont 5 ont été enlevées à cause de taux de réponses correctes insatisfaisants lors des contrôles d'attention. Les données de 63 personnes ont donc été analysées (âge moyen= 44.88, écart type = 14.33, 40 femmes).

Tous les participant-e-s ont été rémunérés avec 3 GBP et ont montré dans d'autres études sur la plateforme *Prolific* des participations satisfaisantes dans au moins 95% des cas.

La compétence linguistique des participant-e-s, telle que mesurée par la tâche *Lextale* (version française, Brysbaert, 2013, voir ci-dessous) est reportée dans le Tableau 7.5.

Tableau 7.5. Expérience 1. La compétence linguistique des participant-e-s, telle que mesurée par Lextale

	moyenne	*SD*	*%*	*CI 95%*
Groupe natif	46.71	6.18	83.41	[82.83, 83.99]
Groupe non-natif	23.52	8.79	41.99	[23.03, 23.99]

7.3.2 Le design et la procédure

En utilisant une tâche de lecture segmentée, nous avons mesuré le temps de lecture des phrases segmentées. Ces phrases contenaient, selon un design factoriel 2^x2, deux *Types de connecteur* différents (soit *cependant, soit mais*) et étaient soit cohérentes, soit incohérentes.

Toutes les phrases expérimentales contenaient une première et une deuxième clause qui étaient reliées soit par le connecteur *mais*, soit par le connecteur *cependant*. Un exemple d'une phrase expérimentale dans ces quatre conditions est reporté dans le Tableau 7.6. Comme on peut le voir, la phrase utilisée pour les quatre conditions était quasiment la même (*mais*/cohérent, *cependant*/cohérent, etc.). Comme chaque phrase a été présentée dans les quatre conditions, le contenu des phases ne représentait pas un biais.

Tableau 7.6. Exemple d'une phrase expérimentale dans ses quatre conditions

connecteur	cohérence	phrase expérimentale
mais	cohérent	Jeanne connait très bien la ville mais elle se perd souvent en se promenant.
mais	incohérent	Jeanne ne connait pas très bien la ville mais elle se perd souvent en se promenant.
cependant	cohérent	Jeanne connait très bien la ville. Cependant, elle se perd souvent en se promenant.
cependant	incohérent	Jeanne ne connait pas très bien la ville. Cependant, elle se perd souvent en se promenant.

Nous avons inclus 5 phrases expérimentales par condition, ce qui fait un total de 20 phrases expérimentales à partir desquelles nous avons établi quatre listes suivant le design d'un carré latin. Chaque phrase expérimentale a été segmentée en sept segments, tel qu'indiqué dans l'exemple (1).

1) Jeanne connait (1) // très bien (2) // la ville (3) // mais (4) // elle se perd (5) // souvent (6) // en se promenant. (7)

Le segment 4 contenait toujours le connecteur. Seuls les temps de lecture des segment 5, 6 et 7 ont été analysés car ces segments ne différaient pas selon les conditions.

Comme au chapitre 4, après la lecture de chaque phrase, une question de vérification a été présentée. La moitié de toutes les questions se référait à la première clause, l'autre moitié se référait à la deuxième clause. Ainsi, nous avons pu contrôler que les participant-e-s lisaient les phrases attentivement : si une personne a donné moins de 60% de bonnes réponses pour ces questions, nous en avons conclu que cette personne n'avait pas lu attentivement les phrases et nous l'avons, en conséquence, exclue de l'analyse des données.[27]

La question était toujours soit correcte, soit incorrecte, indépendamment si la phrase était incohérente ou cohérente. Par exemple, dans (2), l'affirmation selon laquelle Jeanne trouve toujours son chemin est fausse, qu'elle connaisse la ville (cf. phrase cohérente) ou non (cf. phrase incohérente), car il est indiqué dans toutes les conditions qu'elle se perd souvent.

[27] Nous avons utilisé dans cette expérience un taux de bonnes réponses de 60% et non de 75% (comme dans le chapitre 4) car d'autres questions de vérification ont été utilisées. Dans l'expérience actuelle, toutes les questions portaient soit sur la première, soit sur la deuxième clause de la phrase (au chapitre 4, toutes les questions portaient sur la première clause pour les phrases expérimentales). Cela a eu comme conséquence que les questions de vérification devenaient parfois un peu plus difficiles que celles du chapitre 4.

Le seuil de 60% ne semblait toutefois pas être problématique : une analyse qualitative des données des participant-e-s qui donnaient plus de 60% et moins de 75% de bonnes réponses n'a pas révélé d'autres anomalies dans leurs réponses. Enfin, des analyses supplémentaires avec le taux plus conservateur (de 75%) montraient les mêmes effets que ceux avec le taux de 60%.

2) Jeanne trouve toujours son chemin. Vrai ou faux ?

Le matériel linguistique pour les phrases expérimentales et les questions se basait vaguement sur le matériel utilisé dans le chapitre 4. Plus précisément, nous avons dû apporter des modifications en fonction des questions posées. Dans cette expérience, nous voulions assurer que la moitié des questions de vérification se référait à la première clause de la phrase expérimentale, tandis que l'autre moitié se référait à la deuxième clause. Par exemple, au chapitre 4, nous avons utilisé la phrase (3) et la question (4).

3) Alice aime le sucre et les desserts mais elle mange beaucoup de gâteaux tous les soirs.
4) Alice mange beaucoup de gâteaux. Vrai ou faux ?

Contrairement au chapitre 4, la question de vérification de cette expérience devrait se référer à la première clause. Ainsi, nous avons converti la phrase expérimentale (3) en (5) ainsi que la question (4) en (6).

5) Alice est au régime depuis jeudi mais elle mange beaucoup de gâteaux tous les soirs.
6) Alice a récemment changé son mode d'alimentation. Vrai ou faux ?

En plus des phrases expérimentales, nous avons également présenté des phrases de distraction qui devaient détourner l'attention des participant-e-s du fait que toutes les phrases expérimentales contenaient une concession. Si les participant-e-s n'avaient lu que des phrases concessives, les personnes auraient attendu par anticipation une relation concessive pour toutes les phrases ce qui aurait biaisé (ou même annulé) l'effet potentiel du connecteur sur la vitesse de lecture. Pour cette raison, les phrases de distraction contenaient d'autres relations de cohérence (telles que des contrastes, conséquences, etc.) indiquées par les connecteurs *en effet*, *ainsi, alors que* et *c'est pourquoi*.

Les participant-e-s ont accédé à l'expérience via un lien internet et sont arrivés à une première interface, créée avec le site *Qualtrics* (Qualtrics LLC, Provo [UT], Etats-Unis). Le formulaire de consentement a d'abord été présenté et après avoir l'accepté, les participant-e-s ont été dirigé-e-s de manière aléatoire vers une version de l'expérience. L'expérience a été programmée à l'aide de *Psychopy* (version 2020.20) et hébergée sur les serveurs de *Pavlovia* (*www.pavlovia.org*).

Après la lecture de toutes les phrases, les participant-e-s ont réalisé la tâche *Lextale* (voir chapitre 5) pour mesurer leur compétence linguistique. Les mots de cette tâche ont été présentés de manière isolée au milieu de l'écran et les participant-e-s ont indiqué pour chaque mot s'ils le reconnaissaient (en cliquant sur la touche « v ») ou non (en cliquant sur la touche « f »). Au total, l'expérience a duré environ 15 minutes pour les participant-e-s natifs et 16 minutes pour les participant-e-s non-natifs.

7.3.3 L'analyse

Pour chaque segment, nous avons exclu tous les temps de lecture inférieurs à 50ms et supérieurs à 3 secondes de l'analyse. Nous avons estimé qu'il n'était pas possible de bien lire les segments en moins de 50 ms. De plus, un temps supérieur à 3 secondes pouvait aussi être lié à la distraction de la personne lectrice (voir aussi Crible et al., 2021 pour des seuils comparables). Enfin, ces seuils nous ont permis de réduire l'asymétrie de nos données.

Concernant les questions de vérification, nous avons autorisé un peu plus de temps car les participant-e-s devaient lire une phrase entière. En conséquence, nous n'avons exclu seulement les réponses inférieures à 50ms et supérieurs à 6 secondes. Afin de réduire l'asymétrie de nos données telle qu'indiquée par la fonction *skewness()* du package *stats* (R Core Team), nous avons fait une transformation logarithmique (comme dans Crible et al., 2021). De plus, une inspection visuelle a indiqué une distribution selon la loi normale.

Ensuite, pour analyser les temps de lecture, nous avons réalisé des modèles linéaires d'effets mixtes à l'aide de *R* (R Core Team, 2020) et, plus précisément, en utilisant la fonction *lmer()* du *lmerTest* package (Kuznetsova et al., 2017). Pour monter les modèles, nous avons ajouté successivement les effets fixes et avons comparé le modèle résultant avec un modèle qui ne contenait pas cet effet en faisant les tests du rapport de vraisemblance (angl. *likelihood-ratio tests,* voir Baayen, 2009) à l'aide de la fonction *anova()* du package *stats* (R Core Team, 2020). Une fois que le modèle ne s'est plus amélioré avec l'ajout d'un nouvel effet fixe, nous avons récupéré les tailles d'effet estimées et le niveau de signification en utilisant la fonction *summary()* du package *stats* (R Core Team, 2020). Des comparaisons *post-hoc* ont été effectuées en utilisant la fonction *glht()* du paquet *multcomp* (Hothorn et al., 2008).

Tous les modèles (y inclut les modèles qui ne contenaient pas d'effets fixes) contenaient toujours les *Phrases expérimentales* et les *Participant-e-s* comme effets aléatoires. Comme nous avons anticipé de grandes différences dans les temps de lecture entre le groupe natif et le groupe non-natif qui risquait de dissimuler des différences importantes (voir aussi le chapitre 6), nous avons fait des modèles séparés pour les deux groupes (comme dans Crible et al., 2021)

Temps de lecture des segments 5, 6 et 7 (personnes natives)

Pour le segment 5, suivant immédiatement le connecteur, l'ajout de la *Cohérence de la phrase* comme effet fixe n'a pas amélioré le modèle ($\chi(1) = 2.78$, $p = .10$). En revanche, en ajoutant le *Type de connecteur* comme effet fixe, le modèle s'est amélioré ($\chi(1) = 4.40$, $p < .05$). L'interaction avec l'effet fixe de la *Cohérence de la phrase* n'a pas amélioré le modèle davantage ($\chi 2) = 2.89$, $p = .24$). L'output du modèle final (incluant donc seulement le *Type de connecteur* comme effet fixe et les *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires) est reporté dans le Tableau 7.7. Comme nous le voyons, les phrases indiquées par *cependant* ont été lues dans le segment 5 légèrement plus lentement.

Tableau 7.7. Output du modèle final du segment 5 (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.73	.04	81.72	-17.22	$< 2^{e-16}$
Connecteur (*mais*)	-0.03	.01	1296.52	-2.10	.04

Pour la lecture du segment suivant, le segment 6 de la phrase, l'ajout de la *Cohérence de la phrase* comme effet fixe a amélioré le modèle ($\chi(1) = 8.31$, $p < .005$). Par contre, l'interaction avec l'effet fixe du *Type de connecteur* n'a pas amélioré le modèle davantage ($\chi(2) = .61$, $p = .74$). L'output du modèle final (comportant donc la *Cohérence de la phrase* comme seule effet fixe et les *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires) est reporté dans le Tableau 7.8.

Tableau 7.8. Output du modèle final du segment 6 (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.70	.05	86.97	-15.20	2.00^{e-16}
Cohérence (*incohérent*)	.04	.01	1291.03	2.89	< .005

Comme représenté dans la Figure 7.1 (ci-dessous), le segment 6 a été donc lu plus lentement dans des phrases incohérentes que dans des phrases cohérentes, et cela indépendamment du connecteur.

Pour le dernier segment de la phrase, le segment 7, le modèle s'est significativement amélioré en ajoutant successivement la *Cohérence de la phrase* ($\chi(1) = 32.28$, $p = 1.334^{e-08}$) ainsi que son interaction avec *le Type de connecteur* ($\chi(2) = 6.25$, $p < .05$) comme effets fixes. L'output du modèle final est reporté dans le Tableau 7.9.

Tableau 7.9. Output du modèle final du segment 7 (personnes natives)

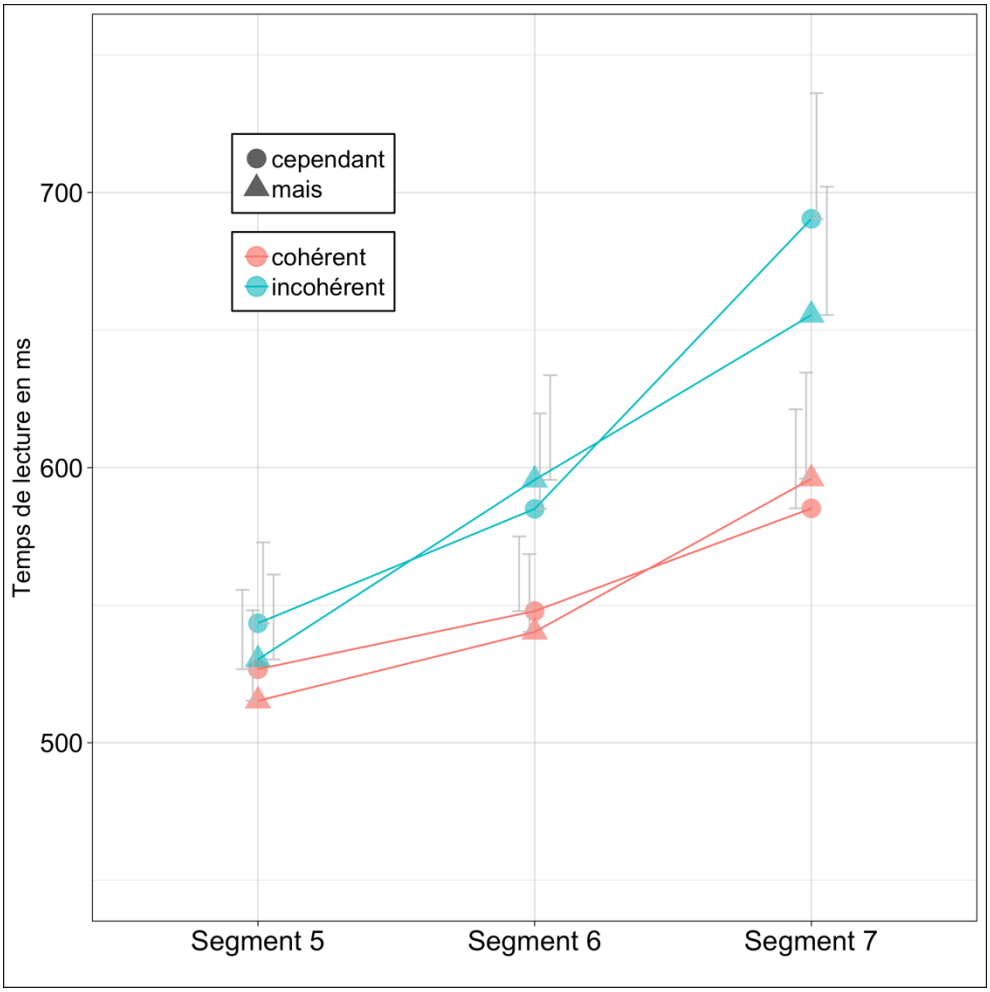
	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-6.4^{e-01}	5.2^{e-02}	9.7^{e+01}	-1.2^{e+01}	2.00^{e-16}
Cohérence (*incohérent*)	1.3^{e-01}	2.2^{e-02}	1.3^{e+03}	5.6^{e+00}	2.12^{e-08}
Connecteur (*mais*)	2.8^{e-03}	2.2^{e-02}	1.3^{e+03}	1.3^{e-01}	.90
Cohérence (*incohérent*) : connecteur (*mais*)	-6.7^{e-02}	3.2^{e-02}	1.3^{e+03}	-2.1^{e+00}	.03

Les comparaisons *post-hoc*, reportées dans Tableau 7.10, montrent non seulement que les phrases incohérentes ont été lues plus lentement, mais aussi la tendance des participant-e-s à lire les phrases incohérentes indiquées par *cependant* plus lentement que les phrases incohérentes indiquées par *mais*.

Tableau 7.10. Comparaisons *post-hoc* pour le segment 7 (personnes natives)

	β	SE	z	Pr(>|z|)
incohérent / *cependant* – cohérent / *cependant*	.13	.02	5.64	< .001
cohérent / *mais* - cohérent / *cependant*	.00	.02	.13	.99
incohérent / *mais* - cohérent / *cependant*	.06	.02	2.80	.03
cohérent / *mais* - incohérent / *cependant*	-0.12	.02	-5.51	< .001
incohérent / *mais* - incohérent / *cependant*	-0.06	.02	-2.85	.02
incohérent / *mais* - cohérent / *mais*	.06	.02	2.67	.04

Figure 7.1. Expérience 1. Les temps de lecture des segments 5, 6 et 7 (personnes natives, limites supérieures de l'IC 95% indiquées par des barres d'erreur)



On peut donc conclure que les personnes natives ont lu le segment 7 plus lentement lorsque la phrase était incohérente et cela pour les deux connecteurs. Au sein des phrases incohérentes, celles qui ont été indiquées par *cependant* ont été lues encore plus lentement.

Tandis que les temps de lectures des personnes natives étaient visiblement affectés par l'(in)cohérence de la phrase, le temps de réponse pour la question de compréhension ne semblait pas être affecté : ni l'ajout du facteur de la *Cohérence de la phrase* ($\chi(1) = .02$, $p = .89$), ni du *Type de connecteur* ($\chi(1) = .72$, $p = .40$) comme effets fixes ont pu prédire le temps nécessaire pour répondre à la question de compréhension.

La lecture des segment 5, 6 et 7 (personnes non-natives)

En analysant la lecture des personnes non-natives pour le 5ème segment, l'ajout de la *Cohérence de la phrase* comme effet fixe n'a pas amélioré le modèle ($\chi(1) = .98$, $p = .32$). En ajoutant le *Type de connecteur* comme seul effet fixe, nous observons une tendance ($\chi(1) = 3.36$, $p = .07$) à lire les phrases avec *cependant* plus lentement que les phrases contenant *mais*. Pour le segment 6, l'ajout de la *Cohérence de la phrase* comme effet fixe a amélioré le modèle ($\chi(1). = 11.62$, $p < .001$). L'interaction avec l'effet fixe du *Type de connecteur* n'a pas amélioré le modèle davantage ($\chi(2) = 4.35$, $p = .11$). L'output est reporté dans le Tableau 7.11.

Tableau 7.11. Output du modèle final du segment 6 (personnes non-natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.12	0.06	62.25	-2.03	< .05
Cohérence (*incohérent*)	.07	0.02	1168.73	3.42	< .001

Pour le dernier segment de la phrase, le segment 7, le modèle s'est amélioré lors de l'ajout de la *Cohérence de la phrase* ($\chi(1) = 13.05$, $p < .0005$) ainsi que son interaction avec le *Type de connecteur* ($\chi(2) = 7.34$, $p < .05$). L'output est reporté dans le Tableau 7.12.

Tableau 7.12. Output du modèle final du segment 7 (personnes non-natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	.00	.06	94.91	.06	.95
Cohérence (*incohérent*)	.03	.03	1156.00	1.11	.27
Connecteur (*mais*)	-0.01	.03	1152.00	-0.19	.85
Cohérence (*incohérent*) : Connecteur (*mais*)	.09	.04	1154.00	2.05	.04

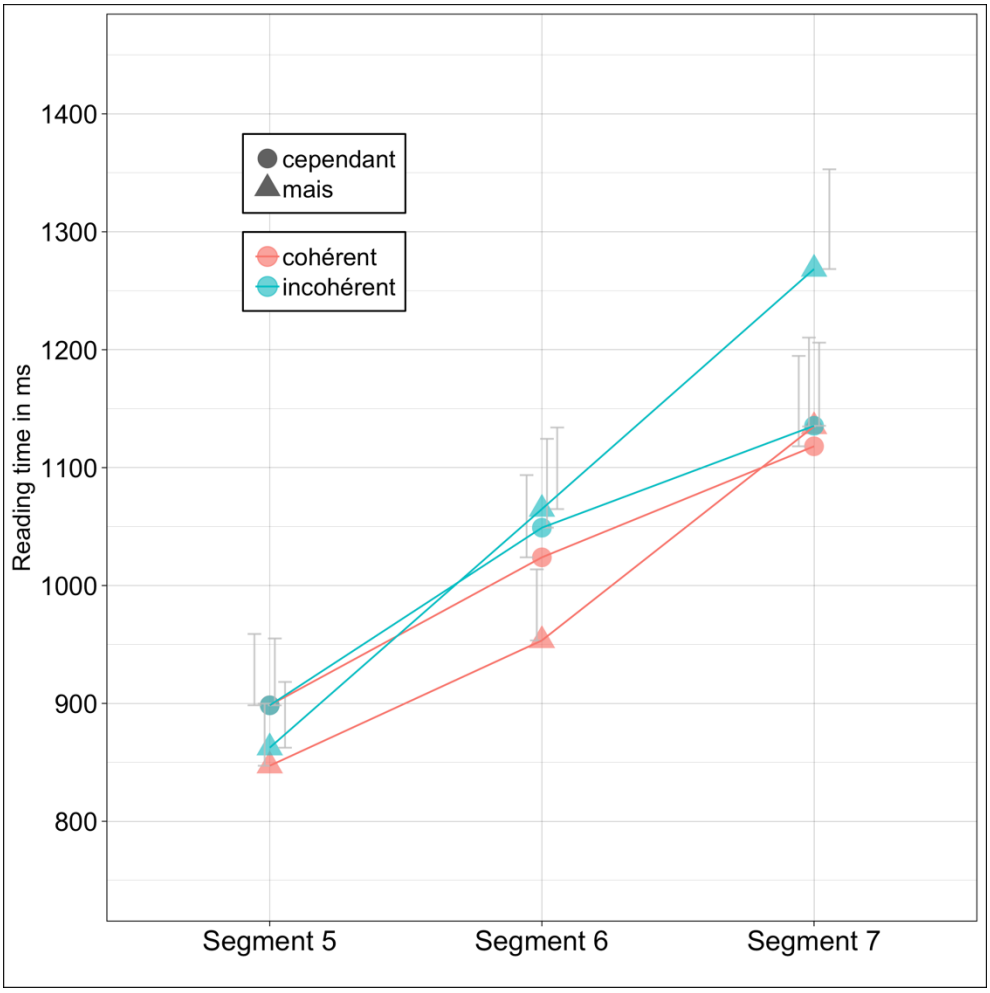
Les comparaisons *post-hoc*, reportées dans le Tableau 7.13, montrent que les phrases avec *cependant* ont été lues à la même vitesse indépendamment si elles étaient cohérentes ou incohérentes, tandis que les phrases indiquées par *mais* ont été lues plus lentement lorsque celles-ci étaient incohérentes.

Tableau 7.13. Comparaisons *post-hoc* pour le segment 7 (personnes non-natives)

	β	ET	z	Pr(>|z|)
incohérent / *cependant* – cohérent / *cependant*	.03	.03	1.11	.69
cohérent / *mais* - cohérent / *cependant*	-0.01	.03	-0.19	1.00
incohérent / *mais* - cohérent / *cependant*	.12	.03	3.81	< .001
cohérent / *mais* - incohérent / *cependant*	-0.04	.03	-1.30	.56
incohérent / *mais* - incohérent / *cependant*	.08	.03	2.71	.03
incohérent / *mais* - cohérent / *mais*	.12	.03	4.03	< .001

Cet effet est également visible dans la Figure 7.2.

Figure 7.2. Expérience 1. Temps de lecture pour des segments 5, 6 et 7 (personnes non-natives, limites supérieures de l'IC 95% indiquées par des barres d'erreur)



Enfin, comme chez les personnes natives, ni la *Cohérence de la phrase* ($\chi(1) < .01$, $p > .99$), ni le *Type de connecteur* ($\chi(1) = .60$, $p = .44$) ne provoquait un temps de réponse plus long à la question de vérification.

7.3.4 Discussion

Dans cette expérience, nous avons mesuré les temps de lecture des personnes natives et non-natives pour des phrases cohérentes et incohérentes qui contenaient des connecteurs différents (*mais* ou *cependant*). En ce faisant, nous avons testé si la réaction à l'incohérence lors du traitement mental d'une phrase dépend du type de connecteur utilisé.

De manière générale, les résultats montrent que les personnes non-natives étaient affectées par l'incohérence. Nous avons observé des ralentissements de lecture pour les phrases incohérentes dans le segment 6, ce qui démontre que le groupe non-natif comprenait bien les phrases et réalisait attentivement la tâche. Il est même remarquable que l'incohérence ait provoqué une réaction déjà dans le sixième segment (voir chapitre 4).

Encore plus important, nous avons également observé un effet du type de connecteur. Lors de la lecture du 7ème segment, les personnes non-natives lisaient toutes les phrases qui contenaient le connecteur *cependant* à la même vitesse, indépendamment du fait que la phrase soit incohérente ou cohérente. En d'autres termes, *cependant* dans une phrase incohérente ne provoquait pas de perturbation du traitement mental, tandis que cet effet a été trouvé pour les personnes natives. Certes, on a observé une réaction temporaire à l'incohérence lors de la lecture du segment 6 indépendamment du connecteur, mais nous avons observé que cet effet ne se prolongeait pas au dernier segment, dans lequel un effet d'incohérence devrait être plus fort, car il ne peut plus être résolu à cet endroit (voir chapitre 4). Ainsi, ces observations laissent supposer que *cependant* n'est pas seulement non-maitrisé (comme observé dans le chapitre 5), mais également que l'utilisation de ce connecteur empêche la détection d'incohérences lorsque les personnes accèdent à leurs connaissances implicites. La raison pourrait être qu'en comparaison avec *mais*, *cependant* est moins fréquent, est plutôt utilisé à l'écrit et évoque un registre élevé. Notre hypothèse selon laquelle un connecteur plus complexe complique la détection de l'incohérence d'une personne non-native est donc renforcée par ces données.

Pour le groupe natif, nous avons également observé, dès le segment 6, un ralentissement pour les phrases incohérentes. Cette observation démontre que le design et les phrases expérimentales étaient appropriés pour révéler les

perturbations du traitement mental lors de la lecture. En comparaison avec le chapitre 4, il est remarquable que nous ayons observé des perturbations du traitement mental pour des phrases incohérentes contenant *mais* qui se prolongeaient jusqu'au dernier segment de la phrase. Pour rappel, nous n'avons observé dans le chapitre 4 qu'un ralentissement temporaire pour les phrases incohérentes qui ont été marquées par *mais*. Nous avions attribué cela, entre autres, aux phrases expérimentales utilisées. Comme nous avons utilisé dans l'expérience actuelle des questions de vérification (et quelques phrases expérimentales) modifiées, l'interprétation pour l'absence d'effet au segment 7 trouvée au chapitre 4 reste tout-à-fait pertinente.

Par ailleurs, nous avons également fait des observations pour le groupe natif qui soutiennent clairement les conclusions tirées au chapitre 4. Premièrement, nous avons observé un effet du type de connecteur lors de la lecture du segment 7 : les phrases incohérentes marquées par *cependant* ont été lues plus lentement que celles marquées par *mais*. Nous attribuons cette différence au type de connecteur testé : comme les participant-e-s ont une très bonne connaissance implicite pour *mais*, il leur était plus facile de rejeter avec confiance une phrase incohérente marquée avec ce connecteur. Deuxièmement, les phrases indiquées par *cependant* ont été lues plus lentement dans le segment 5. Le cinquième segment est le segment qui suit directement le connecteur et est donc l'endroit de la phrase où la signification du connecteur est activée (Zufferey & Gygax, 2016, 2017). Le fait que les personnes natives lisaient ce segment légèrement plus lentement lorsque la phrase contenait le connecteur *cependant* démontre donc que l'activation du sens de *cependant* n'était pas aussi fluide que l'activation du sens de *mais*. Cela doit être attribué aux différences entre les connecteurs. Ensemble, ces deux observations soutiennent donc l'idée selon laquelle le traitement mental des personnes natives dépend (de manière subtile, certes) du type de connecteur.

En conclusion, nous avons vu que les personnes non-natives semblent avoir une meilleure connaissance implicite pour les connecteurs fréquents (tels que *mais*) que pour les connecteurs d'un registre élevé, moins fréquents et qui sont utilisés à l'écrit (tels que *cependant*). Dans le cas de *cependant*, cela a même conduit à ce que l'incohérence ne soit pas détectée (ou, du moins, à ce que les participant-e-s n'y réagissent pas). En outre, nous avons observé un effet du type de connecteur

pour le groupe natif qui soutient les conclusions tirées dans le chapitre 4, à savoir que les fonctions des connecteurs fréquents (tels que *mais*) sont plus « intériorisées » que celles d'un connecteur plus rare. Cela signifie que les fonctions du connecteur peuvent être activées rapidement et efficacement en raison d'une connaissance implicite élevée.

Alors que l'on peut supposer que dans le cas du groupe natif, il s'agit plutôt d'une plus activation plus facile de la fonction d'un connecteur, le cas est moins clair pour le groupe non-natif. Comme les participant-e-s n'étaient pas affecté-e-s par l'incohérence des phrases indiquées par *cependant*, on peut même se poser la question de savoir s'ils ne connaissent pas du tout les fonctions de *cependant*.

Pour répondre à cette question, nous allons mener une deuxième expérience qui nous permettra de mieux expliquer l'absence d'effet de cohérence pour le groupe non-natif pour *cependant* dans l'expérience actuelle.

7.4 Expérience 2 : la réaction à l'incohérence en fonction du type de connecteur lors d'une tâche d'évaluation

Dans cette expérience, nous allons demander aux personnes non-natives d'évaluer la cohérence des phrases testées dans l'Expérience 1 en leur permettant d'accéder à leur connaissances explicites. Nous avons observé dans l'Expérience 1 que l'incohérence ne provoquait pas de perturbations du traitement mental des personnes non-natives lorsque la phrase contenait le connecteur *cependant*. Deux explications sont possibles pour cela.

D'une part, cela peut tout simplement refléter le fait que les personnes non-natives ne connaissent pas la signification de *cependant*. Si tel est le cas, les personnes non-natives ne devraient pas être en mesure de détecter dans l'expérience actuelle l'incohérence des phrases marquées par *cependant*.

D'autre part, il est également possible que les personnes non-natives connaissent, en fait, les fonctions de *cependant*, mais qu'elles n'arrivent pas à reconnaitre des mauvais emplois de ce connecteur lorsqu'elles accèdent à leurs connaissances implicites. En d'autres termes, il est possible que le manque de sensibilité à *cependant* ne se manifeste que durant le traitement en ligne. Si cela est le cas, les personnes non-natives devraient reconnaitre l'incohérence des

phrases contenant *cependant* aussi bien que pour des phrases qui contiennent *mais* dans une tâche hors-ligne.

Outre le fait de déterminer si les personnes non-natives maitrisent *cependant*, cette expérience touche également une question plus générale. Au chapitre 4, nous avons déjà constaté que différentes mesures (telle que la compétence linguistique ou l'exposition à l'écrit) prédisaient la maitrise des connecteurs des personnes non-natives. La question se pose de savoir si de meilleurs résultats à ces tâches prédisent non seulement une meilleure maitrise dans une tâche d'insertion (comme au chapitre 4), mais permettent également de mieux détecter l'utilisation incorrecte d'un connecteur.

En mesurant également l'exposition à l'écrit et la compétence linguistique des participant-e-s, cette expérience nous permettra d'affiner l'image des différences individuelles pour les connaissances explicites sur les connecteurs.

7.4.1 Les participant-e-s

Nous avons recruté 70 personnes natives (âge moyen = 32.13 ans, écart type = 9.98, 29 femmes) et 70 personnes non-natives (âge moyen = 41.09, écart type = 13.81, 46 femmes). Les personnes natives du français ont indiqué d'avoir grandi avec une seule langue. Les personnes non-natives étaient de langue maternelle anglaise et ont également indiqué avoir grandi avec une seule langue. Aucune de ces personnes n'avait participé à l'Expérience 1. Tous les participant-e-s ont montré des participations satisfaisantes dans au moins 95% des études précédentes sur la plateforme *Prolific.* Tous les participant-e-s ont été rémunéré-e-s par 2.54 GBP. Leurs scores de compétences linguistiques, mesurés par la tâche *Lextale* (voir Tableau 7.14), ne différaient pas significativement de ceux des participant-e-s de la première expérience ($\chi(1) = 6.94^{e-30}$, $p > .99$, calculé à l'aide de la fonction *chisq.test()* du package *stats*, R Core Team, 2020).

Tableau 7.14. Expérience 2. Les scores des participant-e-s dans la tâche *Lextale*

	moyenne	*Écart type*	*%*	*IC de 95%*
Groupe natif	48.70	5.50	86.98	[86.46, 87.49]
Groupe non-natif	21.51	10.61	38.41	[20.96, 22.07]

7.4.2 Le design et la procédure

Dans cette expérience, nous avons repris les phrases de l'Expérience 1, y compris les phrases de distraction. Nous avons demandé cette fois aux personnes non-natives d'évaluer la cohérence de ces phrases. Afin d'assurer que les personnes interrogées pouvaient accéder à leurs connaissances explicites, nous n'avons pas imposé de contraintes de temps. Comme dans l'expérience précédente, les facteurs testés étaient la *Cohérence de la phrase* (incohérente ou cohérente) et le *Type du connecteur* (*cependant* ou *mais*).

Les participant-e-s ont accédé à l'expérience via le site *Qualtrics* (Qualtrics LLC, Provo [UT], Etats-Unis) sur Internet. Après avoir accepté le formulaire de consentement, les participant-e-s ont dû évaluer des phrases en déplaçant un curseur sur une échelle allant de « Je suis sûr-e que c'est incohérent » à « Je suis sûr-e que c'est cohérent ». Les réponses ont été converties en valeur numériques allant de 0.0 (correspondant à « Je suis sûr-e que c'est incohérent) à 10.0 (correspondant à « Je suis sûr-e que c'est cohérent »). Les phrases ont été présentées l'une après l'autre et dans un ordre aléatoire pour tous les participant-e-s.

Comme les phrases étaient toujours concessives, il fallait éviter que les participant-e-s jugent toutes les phrases expérimentales incohérentes. En effet, les phrases concessives testées contenaient souvent des actions qui paraissent au premier coup d'œil illogiques. Considérons par exemple (7).

7) Eva a eu une très mauvaise note à l'examen mais elle est très fière d'elle-même.

Dans cet exemple, il ne semble pas logique d'être fier de soi lorsqu'on a obtenu une mauvaise note à un examen. Toutefois, la réaction d'Eva peut s'expliquer autrement : Eva pourrait être fière d'avoir surmonté sa peur de l'examen ou de s'être améliorée par rapport à un examen précédent. Par conséquent, cette phrase doit être considérée comme étant cohérente. En revanche, on ne peut pas trouver une raison pertinente pour une phrase telle que (8), car ici le connecteur indique une concession tandis que la première et deuxième clause sont liées de manière causale.

8) Eva a eu une très bonne note à l'examen mais elle est très fière d'elle-même.

Ces actions ‘illogiques’ dans les phrases concessives auraient pu amener les participant-e-s à croire que ces phrases sont incohérentes. Pour éviter cela, avant la tâche principale, on a indiqué aux participant-e-s qu’une phrase pouvait être cohérente même si les actions décrites paraissaient illogiques. A l’aide de deux exemples (l’un d’une phrase cohérente, l’autre d’une phrase incohérente), nous avons précisé qu’une phrase était cohérente lorsqu’on pouvait trouver d’autres raisons qui pouvaient expliquer les actions illogiques.

Après avoir fait la première expérience, tous les participant-e-s faisaient deux versions de la tâche du *ART* (voir chapitre 5), une fois la version française, une fois la version anglaise. La version française provenait de Zufferey & Gygax (2020b) et la version anglaise était une adaptation de la version originale de Stanovich & West (1989) par Acheson, Wells et MacDonald (2008, ainsi que 65 noms inventés de Martin-Chang & Gould, 2008).

Après avoir fait ces deux tâches, tous les participant-e-s réalisaient la version française de la tâche de vocabulaire *Lextale* (Brysbaert, 2013, voir chapitre 5).

Pour examiner plus en détail les connaissances que nos participant-e-s non-natifs avaient sur les fonctions du connecteur *cependant*, nous leur avons demandé de réaliser deux petites tâches supplémentaires. Dans la première, on leur a demandé d’indiquer la meilleure traduction anglaise pour *cependant*. Dans la deuxième, ils devaient indiquer quelle relation de cohérence *cependant* pouvait indiquer. Ils ont déplacé un curseur sur une échelle allant de « extrêmement inadéquat » à « extrêmement adéquat » pour cinq relations de cohérence : cause, conséquence, concession, contraste et continuation thématique. Toutes les relations de cohérence ont été présentées avec un exemple correspondant. Selon le dictionnaire des connecteurs *LexConn* (Roze et al., 2012), seule la concession était la bonne réponse à cette question.

7.4.3 Les résultats et l’analyse

Nous avons mené des modèles linéaires d'effets mixtes à l’aide de *R* (R Core Team, 2020) qui ont été montés et effectués de la même manière que celle décrite au chapitre 7.3.3. La variable indépendante était la *Réponse à la question* (la valeur de 0.0 à 10.0), les variables dépendantes (et donc effets fixes) étaient la *L1* de *participant-e-s* (français ou anglais), la *Condition de la phrase* (cohérente ou incohérente) et le *Type de connecteur* (*mais* ou *cependant*).

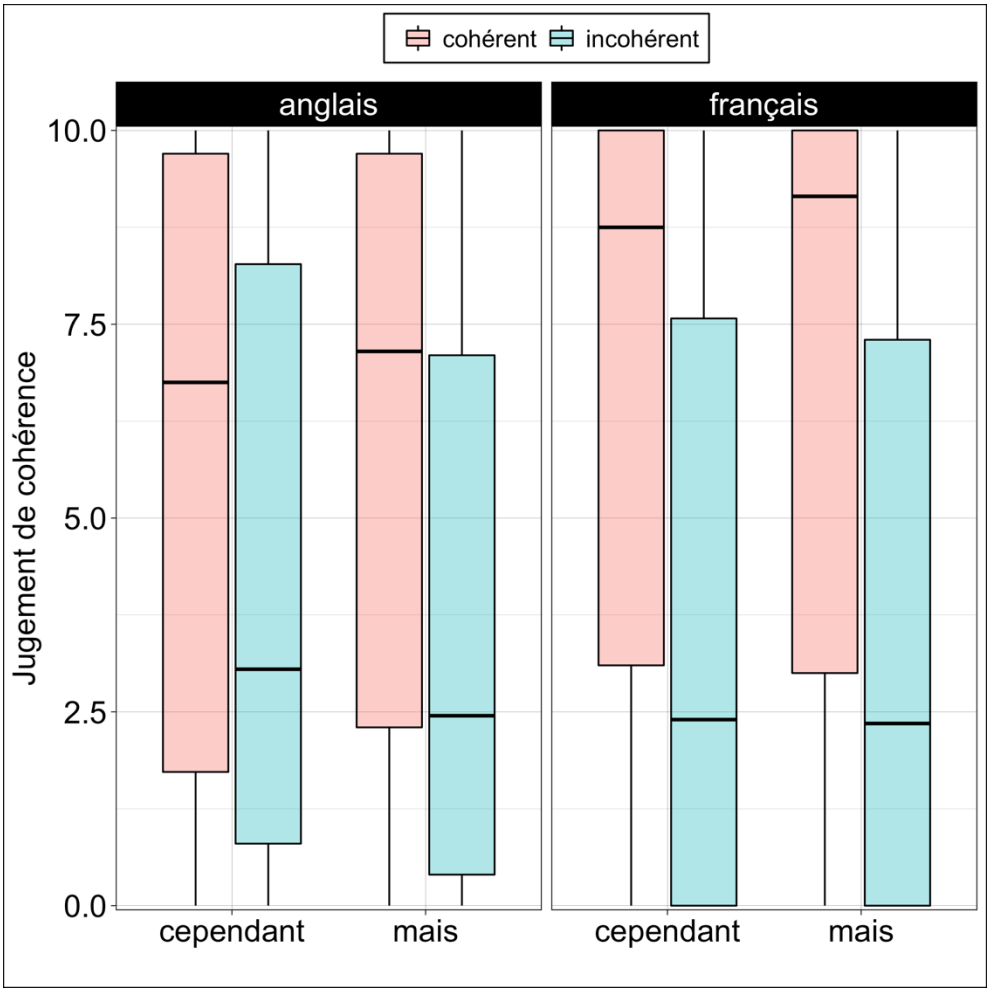
L'ajout de la *Langue des participant-e-s* (anglais ou français) comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 4.13$. $p = .04$). Lors de l'ajout de l'interaction avec la *Condition* (phrase cohérente ou incohérente) comme effet fixe, le modèle s'est amélioré de nouveau ($\chi(2) = 305.74$, $p < 2.2^{e-16}$). L'ajout de l'interaction avec le *Type de connecteur* (*mais* ou *cependant*) comme effet fixe n'a pas amélioré le modèle davantage ($\chi(4) = 6.38$, $p = .17$). L'output du modèle final - contenant donc l'interaction entre les effets fixes de la *Langue des participant-e-s* et la *Condition* ainsi que *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires - est reporté dans le Tableau 7.15.

Tableau 7.15. Expérience 2. Output du modèle final

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	5.84	.23	59.96	25.78	$< 2^{e-16}$
Langue (*français*)	.99	.23	324.30	4.26	2.72^{e-05}
Condition (*incohérent*)	-1.80	.20	2640.69	-9.23	$< 2^{e-16}$
Condition (*incohérent*) : Langue (*français*)	-1.22	.28	2640.87	-4.41	1.09^{e-05}

Nous observons donc que les participant-e-s jugeaient les phrases incohérentes significativement plus incohérentes que les phrases cohérentes et cela indépendamment du type de connecteur utilisé. De plus, nous observons un effet de la langue : les personnes natives donnaient des réponses avec plus de certitude que les personnes non-natives. Ces effets peuvent être également observés dans la Figure 7.3.

Figure 7.3. Expérience 2. Les jugements de cohérence des phrases (le jugement à 10.0 correspond à « Je suis sûr-e que c'est cohérent »).

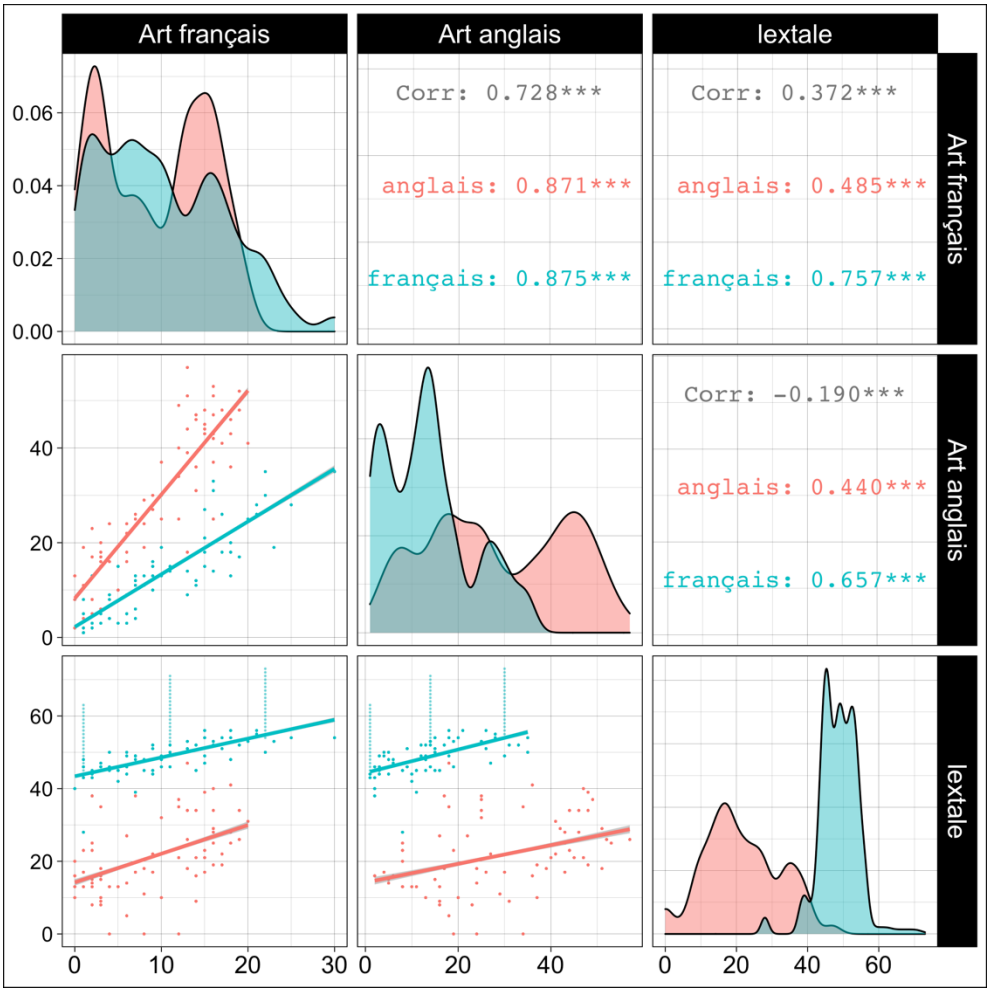


Mesures supplémentaires

Nous nous sommes également intéressés à la question de savoir quels facteurs individuels influençaient ces jugements. Pour examiner cela, nous avons fait passer trois mesures: l'exposition à l'écrit en français et en anglais en utilisant deux versions de la tâche *ART* (d'après l'anglais *Author Recognition Task*) ainsi que la taille du vocabulaire à l'aide de la tâche de *Lextale* (Brysbaert, 2013).

Pour ces trois mesures, nous avons observé des corrélations moyennes à fortes (voir Figure 7.4).

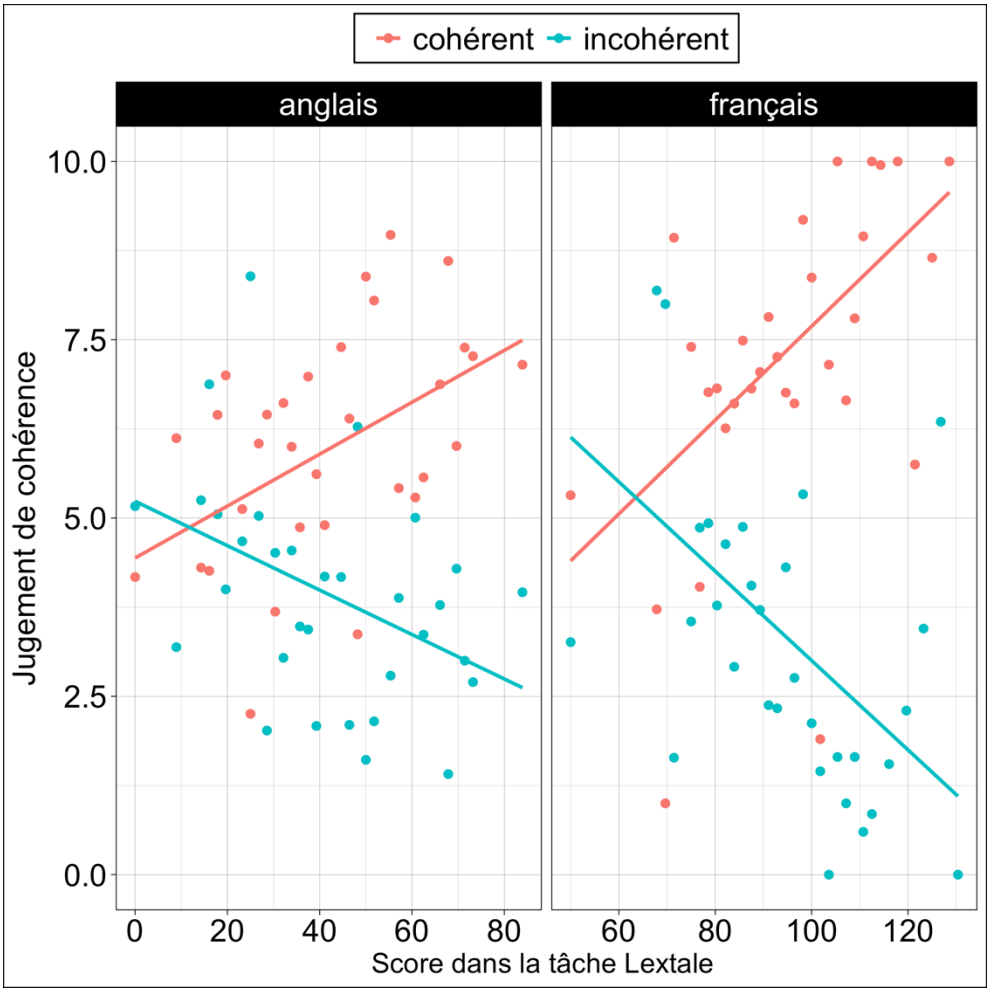
Figure 7.4. Les corrélations entre les scores des tâches ART (versions française et anglaise) et *Lextale*



Pour s'assurer que la contrainte d'absence d'autocorrélation entre les effets fixes sont respectée (Schreiber-Gregory, 2018), nous avons fait des modèles séparés pour chacune de ces mesures. Notre modèle de départ était toujours le modèle final de l'analyse précédente, c'est-à-dire le modèle qui contenait déjà l'interaction entre les effets fixes de la *Langue* (français, anglais) et la *Condition* (phrase cohérente, phrase incohérente) ainsi que les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires.

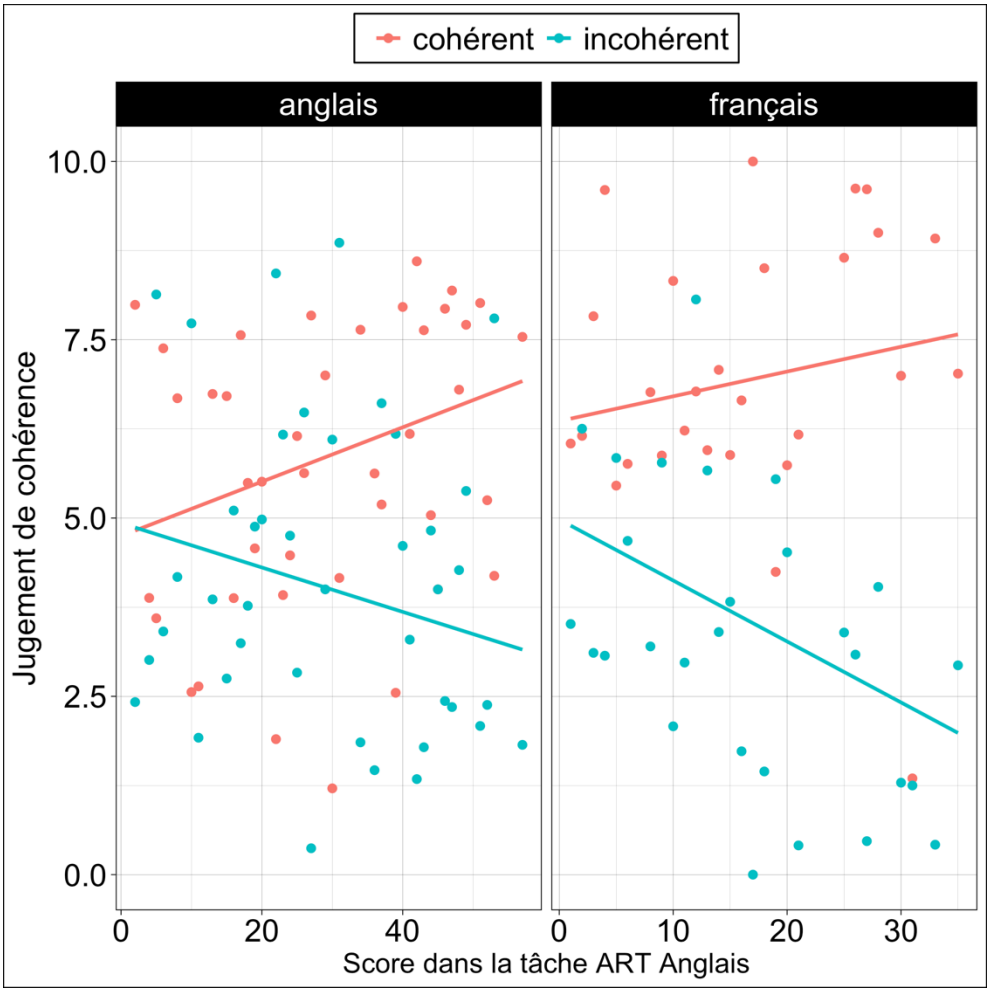
En ajoutant l'interaction avec le score obtenu dans la tâche de *Lextale* comme effet fixe, ce modèle s'est amélioré de manière significative ($\chi(4) = 89.49$, $p < 2.2^{e-16}$). Comme nous voyons dans la Figure 7.5, un score plus élevé dans la tâche *Lextale* prédisaient donc une meilleure détection de l'incohérence pour les deux groupes linguistiques.

Figure 7.5. Expérience 2. Les jugements de cohérence des phrases (le jugement à 10.0 correspond à « Je suis sûr-e que c'est cohérent ») corrélés avec les scores dans la tâche *Lextale*



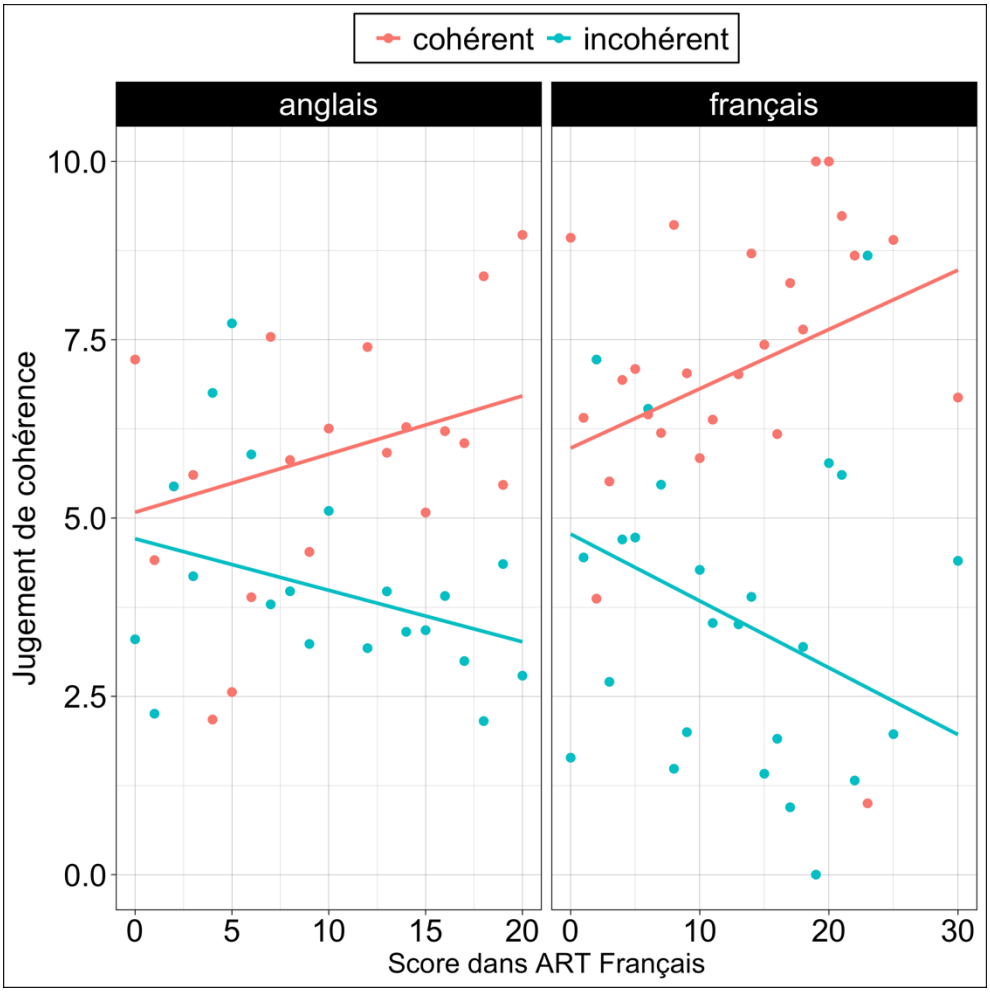
En ajoutant le score de *ART* dans la version anglaise au modèle de départ, le modèle s'est amélioré également de manière significative ($\chi(4)= 5.35$, $p = 2.18^{e-13}$). Comme visualisé dans la Figure 7.6, des scores dans la tâche *ART* (version anglaise) plus élevés prédisaient également une meilleure détection des phrases incohérentes pour les deux groupes linguistiques.

Figure 7.6. Expérience 2. Les jugements de cohérence des phrases (le jugement à 10.0 correspond à « Je suis sûr-e que c'est cohérent ») corrélés avec les scores dans la tâche *Art* (version anglaise)



Finalement, lors de l'ajout des scores de la tâche *ART* (version française) au modèle de départ, le modèle s'est également amélioré de manière significative ($\chi(4) = 65.14$, $p = 2.40^{e-13}$). De nouveau, un meilleur score dans cette tâche prédisait une meilleure détection de l'incohérence pour les deux groupes linguistiques. Cet effet est visualisé dans la Figure 7.7.

Figure 7.7. Expérience 2. Les jugements de cohérence des phrases (le jugement à 10.0 correspond à « Je suis sûr-e que c'est cohérent ») corrélés avec les scores dans la tâche Art (version française)



Les outputs de ces trois modèles sont reportés dans le Tableau 7.16.

Tableau 7.16. Expérience 2. Les outputs des modèles pour les trois mesures supplémentaires: Lextale, *ART* (version anglaise) et *ART* (version française)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
Modèle pour les scores de Lextale					
(Intercept)	4.36	.41	244.30	10.62	$< 2^{e-16}$
Langue (*français*)	-3.20	1.51	368.20	-2.12	3.5^{e-02}
Condition (incohérent)	.96	.44	2656.00	2.17	3.0^{e-02}
Lextale	.04	.01	316.10	4.38	1.7^{e-05}
Langue (*français*) : Condition (incohérent)	7.29	1.77	2638.00	4.13	3.8^{e-05}
Langue (*français*) : Lextale	.03	.02	358.70	1.42	1.6^{e-01}
Condition (incohérent) : Lextale	-0.07	.01	2654.00	-6.91	6.2^{e-12}
Langue (*français*) : Condition (incohérent) : Lextale	-0.06	.02	2643.00	-2.62	9.0^{e-03}
Modèle pour la tâche ART (version anglaise)					
(Intercept)	4.74	.39	230.90	12.25	2.00^{e-16}
Langue (*français*)	1.62	.46	315.90	3.53	4.80^{e-04}
Condition (*incohérent*)	.19	.42	2648.00	0.47	.64
ART A	0.04	.01	315.40	3.52	5.00^{e-04}
Langue (*français*): Condition (*incohérent*)	-1.58	.54	2652.00	-2.91	3.60^{e-03}
Langue (*français*): ART A	.00	.02	315.90	-0.18	.86
Condition (*incohérent*) : ART A	-0.07	.01	2650.00	-5.42	6.55^{e-08}
Langue (*français*): Condition (*incohérent*) : ART A	-0.05	.02	2652.00	-2.02	4.33^{e-02}
Modèle pour la tâche ART (version française)					
(Intercept)	5.07	.34	181.30	14.83	2.00^{e-16}
Langue (*français*)	.92	.42	310.20	2.19	.03
Condition (*incohérent*)	-0.34	.35	2651.00	-0.96	.34
ART F	0.08	.03	312.60	3.06	< .005
Langue (*français*): Condition (*incohérent*)	-0.87	.49	2650.00	-1.78	.08
Langue (*français*): ART F	0.00	.04	311.40	-0.01	.99
Condition (*incohérent*) : ART F	-0.16	.03	2655.00	-4.92	9.14^{e-07}
Langue (*français*): Condition (*incohérent*) : ART F	-0.02	.04	2653.00	-0.45	.65

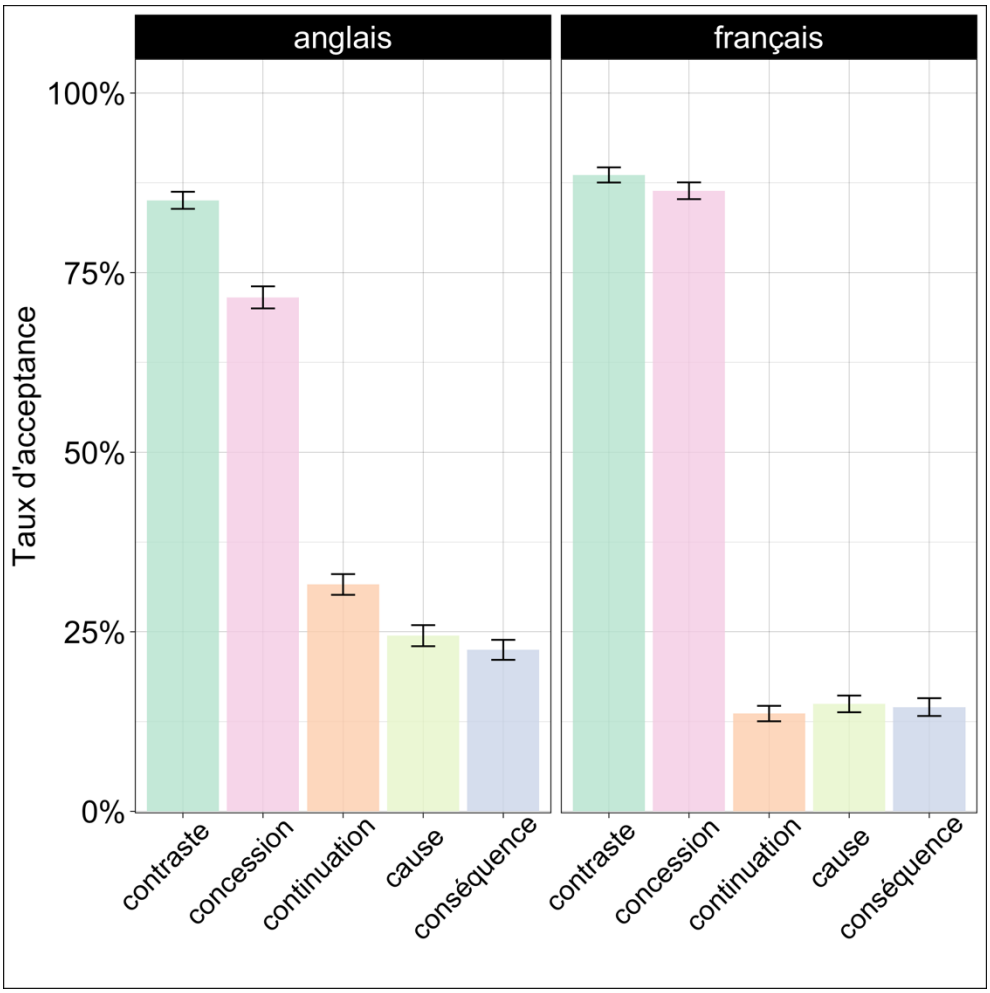
Résultats descriptifs pour les questions sur la fonction de cependant

Pour tester si les personnes non-natives connaissaient la fonction du connecteur *cependant,* nous leur avons posé deux questions à la fin de l'expérience. Dans la première, on leur a demandé de donner la meilleure traduction anglaise pour *cependant.* 80% des participant-e-s non-natifs ont indiqué le connecteur

anglais *however*. Au total, 87% de tous les participant-e-s ont indiqué un connecteur concessif (angl. *however*, *nonetheless*, *though*, fr. *mais*).

Dans la deuxième question, on leur a demandé d'indiquer quelles relations de cohérence le connecteur *cependant* pouvait indiquer. Comme visualisé dans la Figure 7.8, les deux groupes ont montré des taux d'acceptation très hauts pour la relation concessive ainsi que pour la relation contrastive.

Figure 7.8. Les taux d'acceptation des participant-e-s pour les cinq relations de cohérence proposées que *cependant* pourrait indiquer



7.5 Discussion

Dans cette expérience, nous avons demandé à des personnes natives et non-natives d'évaluer la cohérence des phrases que nous avions testées dans la première expérience. Ces phrases étaient soit cohérentes ou incohérentes et contenaient soit le connecteur *mais* soit le connecteur *cependant*. Ce faisant, nous avons pu déterminer si l'absence de perturbation du traitement mental pour une

phrase incohérente qui contenait *cependant* (Expérience 1) était liée au fait que les personnes non-natives ne comprennent pas *cependant* ou si cela était dû au fait que les participant-e-s de la première expérience n'avait pas accès à leurs connaissances implicites de ce connecteur durant la lecture.

Les résultats de cette expérience montrent que les personnes non-natives sont, tout comme leurs paires natives, capables de discriminer les phrases incohérentes des phrases cohérentes lorsqu'elles accèdent à leurs connaissances explicites, et ce quel que soit le type de connecteur utilisé.[28] De plus, dans les questions supplémentaires visant à tester les connaissances métalinguistiques sur les fonctions de *cependant*, les participant-e-s ont pu indiquer que ce connecteur est utilisé pour indiquer une concession. A cet égard, nous avons également observé que les deux groupes n'ont pas accepté seulement *cependant* pour indiquer une concession, mais également une relation *contrastive*. Une explication pour cette évaluation erronée pourrait être le fait que ces deux relations sont parfois difficiles à distinguer, notamment pour un public non-instruit (Zufferey & Degand, 2017).

Le fait que les fonctions de *cependant* étaient connues par les participant-e-s non-natifs se reflète également dans le fait qu'ils ont donné dans plus de 80 % des cas un connecteur concessif comme la meilleure traduction de *cependant*.

Nous en concluons donc que - lors de l'accès aux connaissances explicites - la difficulté de *cependant* non seulement n'a pas été un obstacle pour la détection de l'incohérence, mais que ses fonctions sont maitrisées par les personnes non-natives. Ces résultats ne sont d'ailleurs pas nécessairement en contradiction avec les résultats du chapitre 5, où nous avons observé que les élèves germanophones maitrisaient *cependant* très mal. La différence entre les expériences peut tenir au fait que nous avons utilisé des tâches différentes. Alors que dans le chapitre 5 il s'agissait d'une tâche d'insertion du connecteur, la tâche de l'expérience actuelle portait sur la détection de l'incohérence. Les recherches futures pourraient explorer plus en profondeur comment les différentes tâches captent les connaissances explicites sur les connecteurs.

[28] Le fait que les deux groupes linguistiques n'ont pas clairement rejetés les phrases incohérentes (respectivement ont accepté les phrases cohérentes) à 100% peut être dû au fait que les concessions contenaient souvent les actions illogiques ce qui aurait pu affecter le jugement de cohérence.

Finalement, nous avons vu que les scores de *Lextale*, ART Anglais et ART Français prédisaient pour les deux groupes une meilleure détection de l'incohérence, respectivement un meilleur jugement lorsque la phrase était cohérente. Ces observations soutiennent donc nos observations du chapitre 5, à savoir qu'une bonne connaissance explicite des connecteurs peut être prédite par une exposition à l'écrit élevée (à la fois en L1 et L2) ainsi qu'un vocabulaire plus grand.

En conclusion, les résultats de cette expérience indiquent que lors de l'accès aux connaissances explicites, les personnes non-natives sont capables de détecter de l'incohérence même si la phrase est marquée avec le connecteur *cependant*.

7.6 Conclusion

Dans deux expériences, nous avons examiné si la détection de l'incohérence chez les personnes non-natives dépendait du type de connecteur utilisé. Nous avons formulé l'hypothèse selon laquelle un connecteur plus compliqué (dû à sa fréquence, son registre et sa modalité) complique la détection de l'incohérence, notamment lors du traitement mental. Pour tester cette hypothèse, nous avons testé des phrases incohérentes et cohérentes qui contenaient deux connecteurs qui différaient dans leur difficulté pour les personnes non-natives : le connecteur *mais* et le connecteur *cependant*.

Les résultats de la première expérience confirment clairement que le type de connecteur a un impact sur la manière dont on lit en L2. Nous avons vu que l'incohérence ne provoquait pas de perturbations du traitement mental lorsque les phrases contenaient le connecteur *cependant*. Par contre, lorsque *mais* a été utilisé, l'incohérence de la phrase a provoqué une perturbation du traitement mental. Comme nous avons testé les mêmes phrases pour ces deux connecteurs, cette différence doit être due aux différences entre les connecteurs testés. Plus précisément, nous attribuons cet effet à des différences de fréquence, de modalité et de registre entre ces deux connecteurs, qui ont rendu *cependant* moins accessible aux personnes non-natives. Ainsi on peut conclure que la lecture dans une langue seconde dépend dans une large mesure du connecteur utilisé. De manière plus générale, cette expérience soutient donc l'hypothèse selon laquelle

la grande hétérogénéité des connecteurs affecte leur maitrise dans une langue seconde.

A cet égard, les résultats de la deuxième expérience permettent d'affiner encore ces observations. Dans la tâche d'évaluation, nous avons constaté que les personnes non-natives connaissaient la signification de *cependant* et étaient capables de discriminer les emplois erronés de ce connecteur de ses emplois corrects. Le fait que l'incohérence n'ait pas été détectée lors de la première expérience ne peut donc pas être lié au fait que les participant-e-s ne connaissaient pas la signification de *cependant*. Cette différence reflète plutôt une différence entre les deux types de connaissances testées. Comme il est peu probable que la différence entre les expériences soit due à des différences de compétences linguistiques entre les deux groupes testés (les scores de *Lextale* ne montrent pas de différence significative entre les deux groupes), nous concluons que la connaissance implicite (Expérience 1) des personnes non-natives pour *cependant* est moins développée que la connaissance explicite (Expérience 2). En revanche, nous n'avons pas observé une telle différence pour le connecteur fréquent *mais* : il semble que les personnes non-natives n'ont pas seulement acquis sa fonction d'indiquer une concession mais l'ont également intériorisée de telle sorte qu'ils y ont accès de manière automatique durant la lecture.

Les résultats de la deuxième expérience permettent avec les résultats du chapitre 5 d'obtenir une image plus large et cohérente des facteurs individuels qui permettent une meilleure connaissance explicite des connecteurs en L2. Comme dans le chapitre 5, une exposition à l'écrit plus élevée ainsi qu'une meilleure compétence linguistique prédisaient une meilleure maitrise des connecteurs. Cela démontre que ces corrélations n'émergent pas uniquement dans les tâches d'insertion de connecteur (chapitre 5), mais aussi dans des tâches qui testent la maitrise des connecteurs de manière plus subtile (à savoir les tâches de jugement comme celle utilisée dans ce chapitre).

Enfin, nos résultats indiquent également que la lecture est également affectée par le type de connecteur chez les personnes natives. Dans la première expérience, nous avons constaté que l'intégration de la signification du connecteur *cependant* nécessitait un temps plus long que pour celle de *mais*. Cela soutient donc nos observations faites au chapitre 4.

La série d'expériences réalisées dans ce chapitre montre donc clairement que l'hétérogénéité des connecteurs a une influence sur leur maitrise en L2 et que la maitrise de différents connecteurs dépend du type de connaissance (explicite ou implicite). Dans la série d'expériences de ce chapitre (ainsi que celles du chapitre 5), nous avons montré que les difficultés d'un connecteur résident dans sa basse fréquence, le fait qu'il fasse partie d'un registre de langue élevé et qu'il apparaisse en particulier dans la langue écrite. Nous avons également démontré que ces facteurs s'articulent en fonction du type de la connaissance auquel les personnes peuvent accéder.

Toutefois, comme nous l'avons déjà suggéré, ces facteurs ne sont pas les seuls à pouvoir expliquer ces observations et il doit y avoir plus des facteurs qui rendent un connecteur plus difficile qu'un autre. Nous allons donc examiner dans les chapitres suivants plus en détail encore d'autres facteurs qui pourraient rendre un connecteur plus difficile. Parmi les facteurs examinés se trouvent l'optionnalité du connecteur et un transfert négatif de la L1 (chapitre 8) ainsi que les caractéristiques syntaxiques et pragmatiques d'un connecteur (chapitre 9).

8 Les influences crosslinguistiques

Nous avons observé jusqu'à présent que le type de connecteur a une influence sur la manière dont l'incohérence est détectée dans un discours. Par exemple, nous avons vu que les personnes natives réagissaient plus tôt durant la lecture d'une phrase à l'incohérence causée par une mauvaise utilisation d'un connecteur lorsque celui-ci était fréquent (chapitre 4). Le type de connecteur semble également influencer la lecture en L2 : nous avons vu – bien que les personnes non-natives soient généralement sensibles à des mauvaises utilisations d'un connecteur (chapitres 6) – qu'elles ne l'étaient moins pour un connecteur rare qui est utilisé plutôt à l'écrit (chapitre 7). Ces observations soutiennent donc notre hypothèse selon laquelle le type de connecteur exerce une grande influence sur la maitrise des connecteurs en L2 et que certains connecteurs sont plus difficiles à utiliser et comprendre que d'autres (voir aussi chapitre 5). Nous avons également pu distinguer différents facteurs qui en sont responsables. Entre autres, nous avons discuté des facteurs tels que le registre de langue que le connecteur évoque, sa modalité, sa fréquence, sa transparence sémantique ou la relation de cohérence qu'il indique.

Dans ce chapitre, nous allons compléter ces discussions avec un autre facteur qui est supposé influencer considérablement la maitrise de la L2 en général, et la détection d'une mauvaise utilisation d'un connecteur en particulier : les influences crosslinguistiques de la L1 des apprenant-e-s (Smith & Kellerman 1986; Alonso, 2016).

Selon VanPatten et Benati (2010), le transfert crosslinguistique se caractérise par l'hypothèse inconsciente de l'apprenant-e que la L2 se comporte comme la L1. Cela conduit à des erreurs dans la langue cible. Par exemple, il a été démontré que les mots qui encodent de l'information conceptuelle et qui ressemblent orthographiquement, phonologiquement et même morphologiquement un mot de la L1 sans partager son contenu sémantique (p.ex., fr. *liquoriste* vs. angl. *liquorice* 'reglisse') provoquent plus des traductions erronées en L2 (Janke & Kolokonte, 2015). Concernant les connecteurs de discours, plusieurs études

expliquent également leurs utilisations erronées dans des productions de texte d'apprenant-e-s par une influence négative de la L1 (p.ex., Granger & Tyson, 1996).

Comme l'influence crosslinguistique semble avoir un impact important sur la maitrise des connecteurs en L2, on peut se poser la question de savoir si un transfert négatif empêche les personnes non-natives de reconnaitre les emplois incorrects de connecteurs. A cet égard, il se peut que le transfert négatif ne soit pas un obstacle à la détection de l'incohérence, vu que les personnes non-natives sont capables d'intégrer les significations des connecteurs lors de la lecture et réagissent généralement rapidement aux mauvaises utilisations (chapitre 6). Cependant, nous avons également observé que la détection d'une mauvaise utilisation lors de la lecture semblait plus difficile lorsque le connecteur *cependant* était utilisé, un connecteur que l'on pensait moins accessible (chapitre 7). Ainsi, il se peut aussi que le transfert négatif entrave la détection de l'incohérence lors de la lecture car celui-ci, tout comme l'inaccessibilité de *cependant*, représente une difficulté ajoutée à la lecture en L2.

A notre connaissance, seule l'étude de Zufferey et al. (2015) a examiné la détection des mauvaises utilisations des connecteurs dues au transfert négatif chez les personnes non-natives. Dans cette étude, les participant-e-s ne détectaient les erreurs de transfert négatif que lorsqu'ils accédaient à leurs connaissances implicites. Lorsqu'ils évaluaient consciemment les mêmes phrases, ils ne détectaient pas les erreurs.

Bien que les résultats de Zufferey et al. (2015) soient certainement valables pour les relations de cohérence étudiées, nous pensons qu'ils ne peuvent pas être généralisés à toutes les relations de cohérence. Plus précisément, nous supposons que la capacité à détecter des mauvaises utilisations d'un connecteur est réduite pour certaines relations de cohérence. Dans ce chapitre, nous examinons l'une d'entre elles plus en détail, à savoir les relations de cohérence de *spécification*, telle que présentées à l'exemple (1).

1) Le symbole de la nature, c'est-à-dire un vieil arbre, est très vieux.

Les spécifications se caractérisent par une cohérence intra-phrastique et, en même temps, par une forte optionnalité du connecteur. En effet, un-e lecteur-rice comprendrait (1) tout aussi bien même sans le connecteur *c'est-à-dire*. Ces caractéristiques des spécifications nous amènent à penser que les personnes non-

natives auraient plus de difficulté à détecter les effets de transfert. Un tel cas de transfert négatif pour une spécification en français serait l'utilisation incorrecte de *alors* par des personnes germanophones, due à une confusion avec le connecteur allemand *also*, voir exemple (2).

2) *Le symbole de la nature, alors un vieil arbre, est très vieux.

Au cours de deux expériences, nous allons examiner si les personnes non-natives germanophones arrivent à détecter ce faux emploi lors de la lecture en accédant à leurs connaissances implicites (Expérience 1) ainsi que lorsqu'elles jugent consciemment les phrases, en accédant à leurs connaissances explicites (Expérience 2). Avant de présenter ces expériences, nous allons discuter plus en détail les influences que la L1 peut avoir sur la maitrise des connecteurs en L2.

8.1 Les influences de la L1 pour la maitrise des connecteurs en L2

Il est bien connu que la L1 peut avoir une grande influence sur l'acquisition d'une L2 et constitue même un « composant majeur de l'acquisition de la langue seconde » (VanPatten & Benati, 2010 : 160, ma traduction). Nous allons maintenant discuter de quelle manière cette influence peut se manifester pour les connecteurs de discours. Nous allons distinguer trois cas à cet égard : le transfert positif, le transfert stylistique et le transfert négatif.

8.1.1 Les différents types d'influence de la L1 pour la maitrise des connecteurs en L2

Premièrement, l'influence de la L1 peut avoir un effet positif pour l'apprentissage des connecteurs de la langue cible. Ce phénomène est appelé *transfert positif* (Janet, 2009). De manière générale, le transfert positif est bien étudié, notamment pour les mots conceptuels. Il a été démontré, par exemple, que lorsque deux mots conceptuels ont, outre leur similarité phonologique et orthographique, une superposition sémantique – tel que le mot anglais *tourist* et français *touriste* – leur acquisition est facilitée (Carroll, 1992, Janke & Kolonke, 2015). Ces paires de mots accélèrent également les temps de réaction dans des tâches de décision lexicale (Sánchez-Casas, Davis, & García-Albea, 1992; voir aussi Costa, Santesteban & Caño, 2005 pour un résumé).

Il est également possible que cet effet de facilitation ne s'applique pas uniquement aux mots conceptuels mais aussi aux mots procéduraux comme les connecteurs. Pour donner un exemple, la similarité orthographique entre le connecteur du français *alors* et celui de l'allemand *also* pourrait, dans certains cas, faciliter leur maitrise dans une L2. Considérons les exemples (3) et sa traduction en (4) où les deux connecteurs indiquent des conséquences.

3) *Ich mag die Natur, also gehe ich im Wald spazieren*.
4) J'aime la nature, alors je me promène dans la forêt.

Comme les deux connecteurs partagent la fonction d'indiquer une conséquence, et sont orthographiquement très similaires, un-e apprenant-e germanophone pourrait bénéficier du transfert positif et utiliser correctement le connecteur français en suivant son intuition de la L1.

Évidemment, il est difficile de retracer les occurrences du transfert positif dans des productions de texte des apprenant-e-s. La seule indication que certains connecteurs sont plus faciles pour les apprenant-e-s grâce à un transfert positif serait de constater qu'ils sont utilisés de manière excessive. Cependant, le fait que l'utilisation d'un connecteur soit facilitée par un transfert positif ne signifie pas nécessairement que l'apprenant-e le surutilise. De plus, une surutilisation d'un connecteur peut indiquer également un autre type de transfert, à savoir le transfert *stylistique*.

Un énoncé, notamment à l'écrit, possède des caractéristiques différentes en fonction de la langue dans laquelle il est conçu, comme par exemple, sa progression thématique (Mauranen, 1996; Wei, Zhang & Zhang, 2020). Holtgraves (1997), par exemple, a observé que ses participant-e-s d'origine coréenne déclaraient communiquer généralement plus souvent de manière indirecte que les personnes originaires des États-Unis. Elles indiquaient également d'être plus susceptibles de relever des significations indirectes que des participant-e-s américain-e-s. Selon Holtraves (1997), cette observation reflète les différences culturelles dans la communication. Cuenca (2003) résume :

> « En conclusion, ce qui peut être considéré comme trop informatif, non pertinent, trop verbeux et obscur dans un contexte culturel, est interprété comme un signe d'autorité intellectuelle dans un autre contexte culturel. » (Cuenca, 2003 : 1086, ma traduction)

Les différences stylistiques ont, par conséquent, également une influence sur la manière d'utiliser les connecteurs dans une L2. Dans les études de corpus en

particulier, le transfert stylistique est une explication souvent avancée pour expliquer la sur- ou sous-utilisation des connecteurs (p.ex., Carrió-Paster, 2013; Field & Yip, 1992; Gao, 2016; Wei et al., 2020;). Par exemple, en comparant des essais argumentatifs écrits en anglais par des apprenant-e-s suédois-e-s, Tapper (2005) a observé une surutilisation des connecteurs de clarification et de corroboration. Selon la chercheuse, cette surutilisation s'explique par la stylistique de la L1 des apprenant-e-s car rhétoriquement et stylistiquement, l'exemplification serait un trait caractéristique d'un texte argumentatif écrit en suédois. De même, Kanno (1989) a analysé 41 essais rédigés par des étudiant-e-s d'origine japonaise et a constaté que les effets d'une influence stylistique de la L1 étaient particulièrement évidents dans les relations causales. Elle explique que la structure argumentative diffère entre le japonais et l'anglais et que cette différence amène les apprenant-e-s de l'anglais à surutiliser certains connecteurs causaux. Kanno (1989) précise que l'influence de la L1 se manifeste également dans le choix de ne pas indiquer les relations de cohérence. La chercheuse indique qu'en japonais, pour des raisons stylistiques, les liens de cohérence sont laissés beaucoup plus implicites afin que les lecteur-trice-s puissent découvrir les liens par eux-mêmes, tandis qu'en anglais la personne lisant le texte serait plus guidée. Dans le cadre de cette étude, il est également démontré que certaines utilisations de connecteurs causées par un transfert de la L1 sont totalement erronées. Cela représente le troisième type d'influence que la L1 peut exercer sur la L2, le transfert *négatif*. Il consiste dans le cas des connecteurs à attribuer faussement des fonctions d'un connecteur de la L1 à un connecteur de la L2.

Selon Fraser (1998), « il existe une correspondance générale entre les marqueurs [de deux langues], mais certainement pas une correspondance exacte » (Fraser 1998 : 950, ma traduction; voir aussi Borreguero Zuloaga 2011; Zufferey & Cartoni 2012). Il en résulte que la similitude entre un connecteur de la L1 et un connecteur de la L2 peut conduire à l'hypothèse erronée que les deux connecteurs ont exactement les mêmes fonctions - ce qui entraine ensuite une mauvaise utilisation du connecteur de la L2. Ces utilisations incorrectes sont bien documentées dans les études de corpus (p.ex., Petchprasert, 2013; Aliyar, 2020). Granger & Tyson (1996), par exemple, ont comparé l'utilisation de 108 connecteurs dans des productions de textes de participant-e-s de langue maternelle française et anglaise (étudiant-e-s universitaires en lettres écrivant sur un sujet similaire). Les

chercheur-se-s rapportent, entre autres, l'utilisation erronée du connecteur anglais *on the contrary* qui était souvent confondu avec certaines fonctions du connecteur français *au contraire*.

En conclusion, il semble très probable que la langue première influence l'acquisition et la maitrise des connecteurs en langue seconde. A part des potentielles déviances stylistiques qui ne sont pas, ou pas forcément, préjudiciables à la compréhension, l'influence de la langue maternelle peut également susciter de l'incohérence et donc entraver la compréhension, à savoir dans les cas où une personne confond les fonctions des connecteurs de sa L1 et de la L2.

Alors que la production d'erreurs dues au transfert négatif est bien documentée dans les études de corpus (p.ex., Granger & Tyson, 1996), la question se pose de savoir dans quelle mesure les personnes non-natives sont également capables de les *détecter* dans un texte. Nous avons déjà vu au chapitre précédent qu'en principe, les personnes non-natives sont capables à détecter de l'incohérence lorsqu'un connecteur est mal-utilisé, pourvu qu'il soit fréquent. Cependant, si un connecteur est utilisé de manière erronée et qu'en plus les personnes non-natives l'associent à un connecteur similaire dans leur L1, il se peut qu'elles ne remarquent pas l'erreur.

8.1.2 Les apprenant-e-s sont-ils capables de détecter un mauvais emploi d'un connecteur dû au transfert négatif ?

Pour savoir si les apprenant-e-s de langue arrivent à détecter un mauvais emploi d'un connecteur malgré un transfert négatif, Zufferey et collègues (2015) ont demandé à des personnes néerlandophones et francophones de simplement lire (Expérience 1) et ensuite d'évaluer (Expérience 2) des phrases en anglais contenant des mauvaises utilisations des connecteurs. Un exemple d'une erreur testée était l'utilisation erronée du connecteur anglais *if* dans des phrases contrastives, telle que la phrase (5).

5) * *Many products are exported from Switzerland at various scales. If Swiss chocolate is universally known, Swiss wines are known only by the specialists*
'De nombreux produits sont exportés de Suisse à des échelles différentes. CONNECTEUR le chocolat suisse est universellement connu, les vins suisses ne sont connus que des spécialistes.'

Les auteur-trice-s ont supposé que cette erreur ne serait pas reconnue par les lecteur-trice-s francophones dû à un effet de transfert négatif, à savoir la confusion avec le connecteur français *si*. Il est vrai que le connecteur *si* peut être considéré comme étant l'équivalent approprié pour *if* lorsqu'il s'agit d'une condition (Zufferey et al., 2015 : 396), mais il se distingue du connecteur anglais *if* dans un autre cas : *si* peut être utilisé pour indiquer un contraste (LexConn, Roze et al., 2012), tandis que anglais *if* ne sert pas cette fonction.

Comme supposé par Zufferey et al. (2015), les participant-e-s francophones ne détectaient pas les erreurs telles que (5) lorsqu'ils évaluaient les phrases. Cependant, lorsque les participant-e-s recouraient à leurs connaissances implicites (i.e., dans une tâche de lecture en ligne), ces types d'erreurs provoquaient visiblement des perturbations de la lecture. L'équipe de recherche en a conclu que la capacité de détecter des faux emplois malgré un transfert négatif dépend du type de connaissance auquel les personés non-natives accèdent. Cependant, la généralisation de ces résultats pour toutes les mauvaises utilisations des connecteurs par transfert négatif n'est pas assurée et nous supposons que leur détection dépend de la relation de cohérence en question. Pour une relation pour laquelle l'utilisation d'un connecteur est hautement optionnelle (encore plus que pour un contraste ou une condition, les relations testées dans Zufferey et al., 2015), il se peut qu'un faux emploi du connecteur soit plus difficilement détecté car l'indication erronée fournie par le connecteur peut être plus facilement négligée. Un exemple d'une telle relation de cohérence est la relation de *spécification*.

8.1.3 Les caractéristiques des spécifications et leur rôle dans le transfert négatif

La cohérence d'un discours ne lie pas nécessairement deux phrases, mais peut également se trouver au niveau intra-phrastique, par exemple dans le cas des spécifications. Comme on peut voir dans (6), les spécifications ne reformulent ou répètent pas seulement ce qui est dit avant en d'autres termes ('hôtels de luxe'), mais les précisent davantage (Cuenca, 2003).

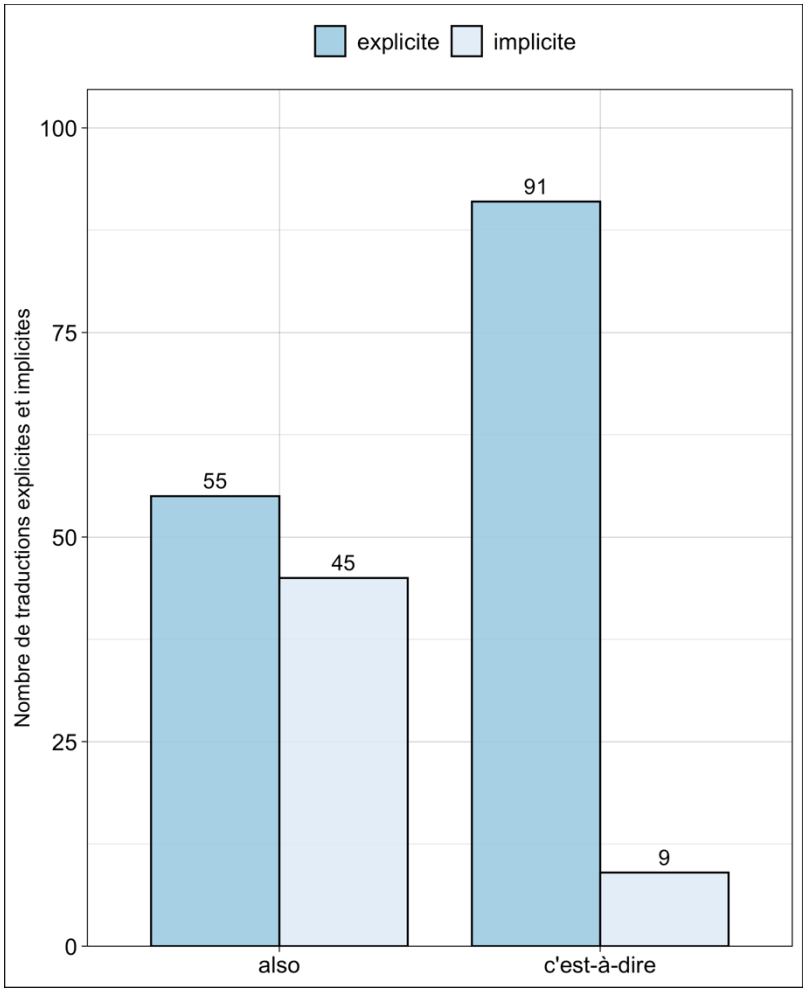
6) Les hôtels de luxe, les établissements qui ont 4 ou 5 étoiles, sont très chers.

Nous voyons également dans l'exemple (6) que l'utilisation d'un connecteur est hautement optionnelle dans une spécification. Pour l'anglais, Asr & Demberg

(2012a, 2012b) ont montré que les relations similaires d'une spécification (telle qu'un *restatement* 'réaffirmation') sont les relations les moins souvent marquées explicitement par un connecteur dans le corpus *Penn Discourse Treebank* (Prasad et al., 2008). Taboada (2006, voir aussi Das, 2014; Das & Taboada, 2018) a montré que ces types de relations sont en outre le plus souvent marquées par d'autres moyens que les connecteurs. Cela soulève la question de savoir à quel point la nécessité d'indiquer une spécification dépend de la langue en question.

Pour savoir si les spécifications sont indiquées plus explicitement ou implicitement en allemand qu'en français, nous avons extrait aléatoirement 100 spécifications indiquées par *also* en allemand ainsi que 100 spécifications en français indiquée par *c'est-à-dire* dans le corpus parallèle allemand-français des débats du Parlement européen Europarl (Koehn, 2005). Ensuite, nous avons examiné comment ces spécifications ont été traduites vers l'allemand ou vers français, c'est-à-dire si par exemple *c'est-à-dire* était traduit explicitement avec un connecteur équivalent de l'allemand ou si la spécification était implicite. L'analyse a révélé que les spécifications en allemand, indiquées par *also,* ont été traduites en français dans environ la moitié des cas par un connecteur équivalent (p.ex., en utilisant des connecteurs comme *c'est-à-dire* ou *à savoir*). En français, les spécifications indiquées par *c'est-à-dire* ont été traduites vers l'allemand dans plus de 90% de manière explicite (p.ex., en utilisant des connecteurs comme *also* ou *das heisst*). Les nombres de traductions sont illustrés dans la Figure 8.1.

Figure 8.1. Les nombres de traduction explicites et implicites pour *also* (‘alors’) et *c’est-à-dire* dans des corpus parallèles allemand-français des débats du Parlement européen



Il semble donc que les spécifications sont plus souvent laissées implicites en français par rapport à l’allemand.

Concernant notre hypothèse de départ – à savoir qu’une erreur due au transfert négatif n’est que très difficilement détectée dans une spécification – il convient de noter que le marquage d’une spécification en français semble être, par transfert négatif, parfois difficile pour des apprenant-e-s germanophones du français. Considérons l’exemple authentique (7), provenant d’une étudiante dans un travail de séminaire universitaire, dans lequel l’utilisation erronée du connecteur *alors* provoque de l’incohérence intra-phrastique.

7) *Même si certains morphèmes ne se constituent que d’une seule lettre, alors un graphème ou phonème, ils sont importants puisqu’ils portent un sens.

Cette erreur s'explique de la façon suivante. Le connecteur français utilisé, *alors*, est très fréquent (plus de 4 millions d'occurrences dans le corpus *frTenTen17*, Jakubíček et al., 2013) et indique, selon la taxonomie de Sanders et al. (1992), des relations de conséquence. Outre cette fonction, *alors* indique également de la temporalité (Bras, Draoulec & Asher, 2009) et peut être utilisé comme un marqueur discursif, indiquant par exemple les transitions, les prises de tours ou les changements de sujet (Essen, 1987; voir aussi Ferre 2011; Degand & Fagard, 2011; Degand & Evers-Vermeul, 2015). Fait important, il est orthographiquement très similaire au connecteur allemand *also* ('alors' ou 'c'est-à-dire') et peut, tout aussi bien que français *alors*, indiquer une conséquence (DimLEX, Stede & Umbach, 1998) ainsi que des changements de sujet et prises-de-parole dans la langue parlée. Hormis ces similitudes, il y a également une différence importante entre les connecteurs *alors* et *also* : contrairement au connecteur français, le connecteur allemand peut être utilisé pour indiquer une spécification, comme dans (8).

8) Luxushotels, also 4- oder 5-Sterne-Hotels, sind sehr teuer.
'Les hôtels de luxe, c'est-à-dire les établissements de 4 ou 5 étoiles, sont très chers.'

L'erreur dans (7) s'explique donc par une confusion entre les fonctions du connecteur français *alors* avec les fonctions du connecteur allemand *also*. Plus précisément, l'apprenante a fallacieusement assimilé par transfert crosslinguistique les deux connecteurs et utilisé *alors* pour indiquer une spécification alors que seul le connecteur allemand remplit cette fonction.

Comme les connecteurs sont hautement optionnels dans des spécifications et en raison du fait que les spécifications ne sont pas toujours marquées explicitement (en français en particulier), nous supposons que les apprenant-e-s germanophones du français tombent dans le piège du transfert négatif et n'identifient pas facilement l'emploi erroné de *alors*.

Nous allons tester cette hypothèse dans le cadre de deux expériences[29], d'abord en visant les connaissances explicites (Expérience 1) ensuite les connaissances implicites (Expérience 2) des personnes non-natives.

[29] Les résultats des deux expériences ont été publiés, sous une autre forme, sous le titre « Processing clause-internal discourse relations in a second language. A case study of specifications in German and French » dans le journal *Studies of Second Language Acquisition* (Wetzel, Crible & Zufferey, 2022)

8.2 Expérience 1 : la détection de l'emploi erroné de *alors* lors de l'accès aux connaissances explicites

Dans cette expérience, nous allons examiner si les personnes non-natives sont sensibles à une utilisation erronée d'un connecteur par transfert négatif lorsqu'elles ont accès à leurs connaissances explicites. Pour tester cela, nous allons demander aux participant-e-s germanophones d'évaluer les phrases contenant une spécification qui est 1) indiquée par le bon connecteur *c'est-à-dire*, 2) sans connecteur, ou 3) par le connecteur *alors*, qui pourrait – par similarités orthographiques et fonctionnelles avec le connecteur allemand *also* – déclencher des effets de transfert négatif chez les germanophones.

Malgré le transfert négatif, nous supposons que les participant-e-s non-natifs arrivent à discriminer les emplois corrects des emplois incorrects car ils bénéficient de l'accès à leurs connaissances explicites. Plus précisément, le fait que les participant-e-s n'aient pas de contrainte de temps pour évaluer les phrases devrait leur permettre de détecter les phrases incorrectes.

En outre, la capacité à détecter les faux emplois de *alors* devrait être prédite par la compétence linguistique des participant-e-s non-natifs. En effet, nous avons vu au chapitre 7 que la détection des faux emplois de *mais* et *cependant* (lors de l'accès aux connaissances explicites) a été prédite par les scores de compétence linguistique, à savoir les scores de *Lextale* (Brysbart, 2013). De même, nous avons observé au chapitre 5 que ces scores prédisaient la maitrise des connecteurs dans une tâche d'insertion. S'il existe effectivement un lien entre les connaissances explicites sur les connecteurs et la compétence linguistique, nous devrions également l'observer dans les cas de transfert négatif.

Enfin, il n'est pas clair si les participant-e-s évaluent les spécifications implicites différemment des spécifications marquées explicitement (avec *c'est-à-dire*). Trois hypothèses peuvent être formulées à ce sujet. Tout d'abord, il se peut que les spécifications implicites soient considérées comme plus correctes parce que la présence d'un connecteur représente pour les personnes non-natives une difficulté ajoutée à l'interprétation de la phrase (p.ex. Zufferey & Gygax, 2017; Crible et al., 2021). De plus, nous avons montré que les spécifications sont laissées plus souvent implicites en français. Ainsi, la présence d'un connecteur, bien qu'il soit correct, peut être considéré comme un élément divergeant de la norme apprise.

En revanche, il se peut également que les spécifications explicites soient considérées comme étant plus correctes, car la présence d'un connecteur facilite l'interprétation de la phrase. Comme mentionné auparavant, l'étude de Degand & Sanders (2002) a montré que les personnes non-natives comprennent mieux des textes lorsque des connecteurs y sont présents. Ainsi, une phrase explicite pourrait être comprise mieux qu'une phrase sans connecteur - ce qui pourrait se refléter dans les évaluations de cohérence. Par ailleurs, nous avons montré que les spécifications sont marquées plus souvent explicitement en allemand qu'en français. Ainsi, des germanophones pourraient supposer par transfert de leur L1 que le marquage explicite d'une spécification est plus acceptable en français que le fait de les laisser implicites.

Enfin, il se peut aussi que les participant-e-s non-natifs jugent les deux types de spécifications comme correctes. Les deux types sont effectivement corrects et apparaissent comme étant relativement fréquents en français ainsi qu'en allemand.

8.2.1 Les participant-e-s

Les participant-e-s ont été recruté-e-s sur la plateforme *Prolific* (Prolific, London, 2020) ainsi que parmi des étudiant-e-s des universités en Suisse romande et alémanique. Les participant-e-s de *Prolific* ont montré une bonne participation aux expériences précédentes sur la plateforme (un minimum de 95% de bonnes évaluations) et ont été rémunérés avec 2.10 GBP dans notre expérience. Les participant-e-s recruté-e-s dans les universités ont été rémunéré-e-s par un bon pour un commerce de vente d'une valeur de 5 CHF.

Au total, nous avons analysé les données de 72 participant-e-s germanophones qui ont tous indiqué parler le français comme langue seconde (âge moyen = 28.5, écart type = 11.9, 45 femmes) ainsi que 103 personnes natives francophones (âge moyen = 28, écart type = 11.4, 22 femmes). Tous les participant-e-s ont donné leur consentement à participer à cette expérience.

Concernant leur compétence linguistique, telle que mesurée par la tâche de *Lextale* (Brysbart, 2013; voir aussi chapitre 5), le groupe non-natif a obtenu un score moyen de 38% (IC 95% [37.28; 39.16]). En comparaison, le groupe natif a obtenu un score moyen de 84% (IC 95% [83.66; 84.27]).

A part les mesures de *Lextale*, nous avons également demandé à tous les participant-e-s d'auto-évaluer leurs compétences linguistiques en français sur trois

niveaux : la lecture, la compréhension à l'oral et la production à l'oral (suivant le questionnaire LEAP de Kaushanskaya, Blumenfeld & Marian, 2019). Les réponses ont été converties en valeurs numériques (100% = *parfait*, 80% *très bon*, 60% = *bon, etc.*). Le groupe non-natif a indiqué pour la *lecture* en moyen 68 (écart type = 17), pour la *compréhension à l'oral* en moyenne 63 (écart type = 20), et pour la *production orale* en moyenne 58 (écart type = 19). Les personnes non-natives estimaient donc avoir en moyenne un niveau en français *bon* à *très bon*.

8.2.2 Le design et la procédure

L'expérience a été programmée sur *Qualtrics* (Qualtrics LLC, Provo [UT], Etats-Unis) et les participant-e-s y ont accédé via un lien Internet. Dans la tâche principale, les participant-e-s ont dû juger des phrases en déplaçant un curseur sur une échelle allant de « Je suis sûr-e que c'est incorrect » à « Je suis sûr-e que c'est correct ». Pour s'assurer que les participant-e-s puissent accéder à leurs connaissances explicites, aucune contrainte de temps a été imposée.

Au total, 30 phrases expérimentales ont été créées, qui contenaient toutes une spécification. Chaque phrase était présentée dans trois conditions. Soit les spécifications ont été laissées implicites (sans connecteur, voir 9), soit marquée correctement avec le connecteur *c'est-à-dire* (10), soit marquée incorrectement avec le connecteur *alors* (11).

9) Mon restaurant de tous les jours, une pizzeria qui se trouve derrière la gare, vient de fermer.
10) Mon restaurant de tous les jours, c'est-à-dire une pizzeria qui se trouve derrière la gare, vient de fermer.
11) *Mon restaurant de tous les jours, alors une pizzeria qui se trouve derrière la gare, vient de fermer.

Avec ces phrases, trois listes ont été établies selon le design du carré latin (30 phrases expérimentales x 3 conditions) ce qui évitait que le contenu des phrases biaise les jugements.

Outre les phrases expérimentales, nous avons également présenté 40 phrases de remplissage. Parmi ces 40 phrases, 28 contenaient un emploi correct ou incorrect d'un pronom (12) et 12 contenaient le connecteur *alors* qui indiquait correctement une conséquence (13). Ceci assurait que les phrases contenant *alors* étaient aussi correctes de temps en temps.

12) *Jérôme adorait la voiture automatique de laquelle il avait appris à conduire tout seul.
13) Connor cuisine très épicé, alors Betty a souvent des aigreurs d'estomac quand elle mange chez lui.

Après avoir complété la tâche principale, les participant-e-s réalisaient la tâche de Lextale (Brysbaert, 2013; voir aussi chapitre 5) et répondaient aux questions d'autoévaluation.

8.2.3 L'analyse

Les positionnements du curseur de la tâche principale ont été convertis en valeurs numériques, de 0.0 à 10.0 (10 correspondait à « Je suis sûr-e que c'est correct »). Ces scores étaient notre variable dépendante. Nos variables indépendantes étaient le *Groupe linguistique* (*natif* ou *non-natif*) et la *Condition de la phrase* (spécification indiquée par *c'est-à-dire*, indiquée par *alors* ou laissée *implicite*).

Pour l'analyse, nous avons monté des modèles linéaires à effets mixtes à l'aide de *R* (R Core Team, 2020). Nous les avons construits en suivant l'approche de Baayen (2008), à l'aide de la fonction *lmerTest()* du package *lme4* (Bates et al, 2014). Nous avons évalué l'amélioration des modèles lorsqu'un effet fixe a été ajouté à l'aide de la fonction *anova()* du package *base R* (R Core Team, 2020). Pour obtenir les valeurs de *p*, nous avons utilisé la fonction *summary*() du package *base R* (R Core Team, 2020). Pour les comparaisons *post-hoc*, nous avons utilisé la fonction *emmeans()* du package *emmeans* (Lenth, 2020). Pour tous nos modèles, nous avons inclu les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires.

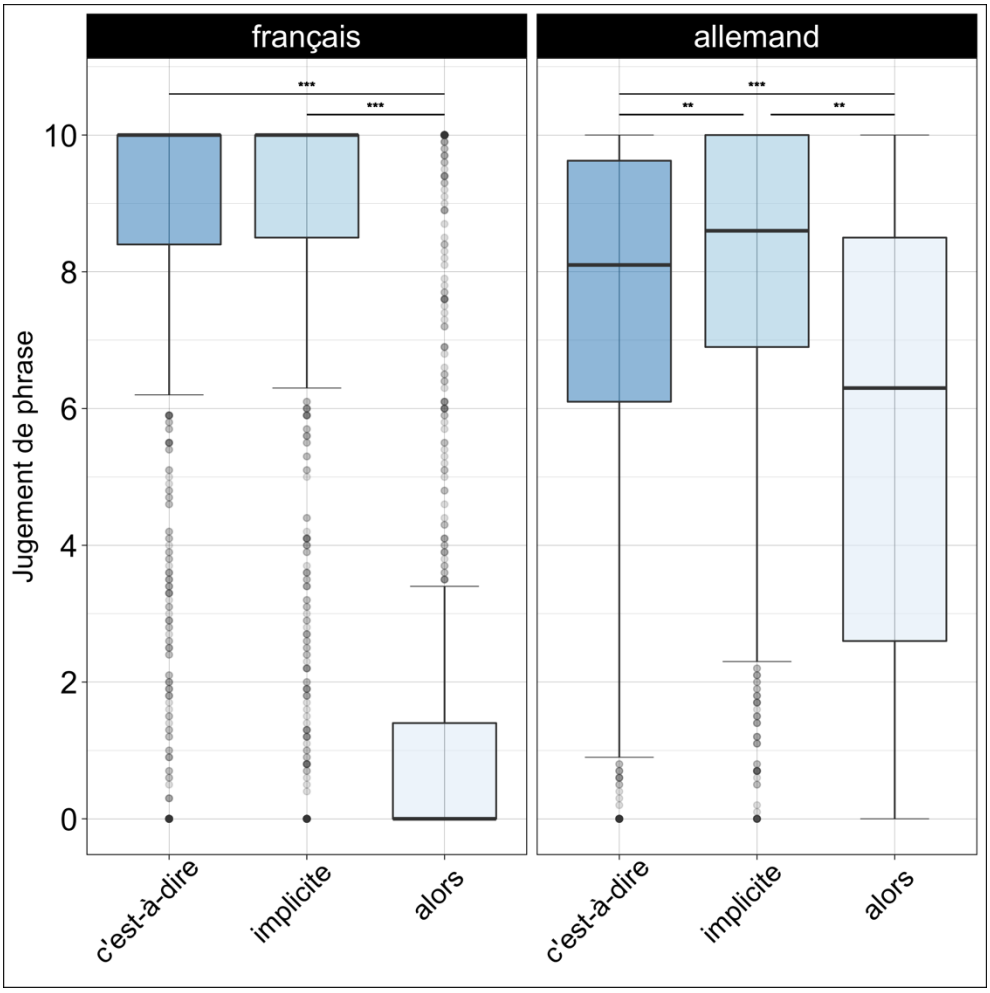
L'ajout du *Groupe linguistique* (*natif* ou *non-natif*) comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 15.12$, $p < .0005$). L'ajout de l'interaction avec le *Type de connecteur* (*alors*, *c'est-à-dire* ou *implicite*) a amélioré le modèle davantage ($\chi(4) = 3670.5$, $p < 2.2^{e-16}$). L'output du modèle final, contenant l'interaction entre le *Groupe linguistique* et le *Type de connecteur* ainsi que les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires, est reporté dans le Tableau 8.1.

Tableau 8.1. Expérience 1, output du modèle final.

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	5.53	.17	196.05	33.29	< .001
Groupe (*natif*)	-4.12	.18	367.60	-22.36	< .001
Type de connecteur (*c'est-à-dire*)	1.70	.14	5019.26	12.31	< .001
Type de connecteur (*implicite*)	2.23	.14	5019.26	16.17	< .001
Groupe (*natif*) : Type de connecteur (*c'est-à-dire)*	5.47	.18	5018.74	30.41	< .001
Groupe (*natif*): Type de connecteur (*implicite*)	4.96	.18	5018.74	27.54	< .001

Les comparaisons *post-hoc* indiquent que les germanophones jugeaient les phrases contenant *alors* comme étant moins correctes que les phrases qui ne contenaient pas de connecteur ($\beta(5214) = 2.20$, ET = .14, $p < .0001$) ainsi que les phrases marquées par *c'est-à-dire* ($\beta(5214) = 1.70$, ET = .15, $p < .0001$). En outre, ils jugeaient ces phrases comme plus correctes que les personnes natives ($\beta(5214) = 4.12$, ET = .14, $p < .0001$). Les personnes non-natives ont également évalué différemment les deux versions des spécifications qui étaient marquées correctement : ils ont jugé les phrases qui ne contenaient pas de connecteur comme plus correctes que les phrases qui contenaient *c'est-à-dire* ($\beta(5214)= -0{,}51$, ET = .14, $p < .01$). Les personnes natives, par contre, n'ont pas évalué l'une des deux conditions comme plus correcte ($\beta(5214)= -0{,}01$, ET = .12, $p > .99$). Les résultats sont illustrés à la Figure 8.2

Figure 8.2. Les jugements de phrases (10 correspondant à « Je suis sûr-e que c'est correct).



Pour mieux comparer ces résultats, il convient également de tenir compte des jugements pour les phrases de remplissage. Les phrases de remplissage qui contenaient des pronoms faux étaient en moyenne évaluée à 4.16 (écart type = 3.55), et celles qui contenaient des pronoms corrects étaient évaluées en moyenne à 6.54 (écart type = 3.14). Le jugement des phrases de remplissage qui contenaient des utilisations correctes de *alors* était en moyenne de 6.92 (écart type = 3.16).

Scores au test Lextale

Pour savoir si la capacité à détecter le faux emploi de *alors* était lié à la compétence linguistique des participant-e-s, nous avons analysé si les scores obtenus dans la tâche *Lextale* étaient prédictifs d'une meilleure détection de la mauvaise utilisation de *alors*.

Une corrélation entre la taille du vocabulaire en français et le fait d'être natif ou non semble très probable. Ainsi, pour respecter l'hypothèse de l'absence d'autocorrélation des effets fixes (Schreiber-Gregory, 2018), nous avons fait des modèles séparés pour les deux groupes linguistiques. La variable indépendante était cette fois-ci seulement les jugements pour les phrases qui contenaient *alors*, car nous étions seulement intéressés à déterminer si les scores de *Lextale* prédisaient une meilleure détection dans cette condition.

Pour les personnes non-natives, l'ajout des *Scores de Lextale* comme effet fixe a amélioré le modèle de manière significative ($\chi(1) = 5.21$, $p < .05$). L'output de ce modèle, contenant les Participant-e-s et les Phrases comme effet aléatoires ainsi que les *Scores de Lextale* comme effet fixe est reporté dans le Tableau 8.2.

Tableau 8.2. Output du modèle final (variable indépendante : les scores pour les phrases marquées avec *alors*), personnes non-natives

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	6.58	.53	71.28	12.48	$< 2^{e-16}$
Scores de Lextale	-0.05	.02	70.35	-2.29	.03

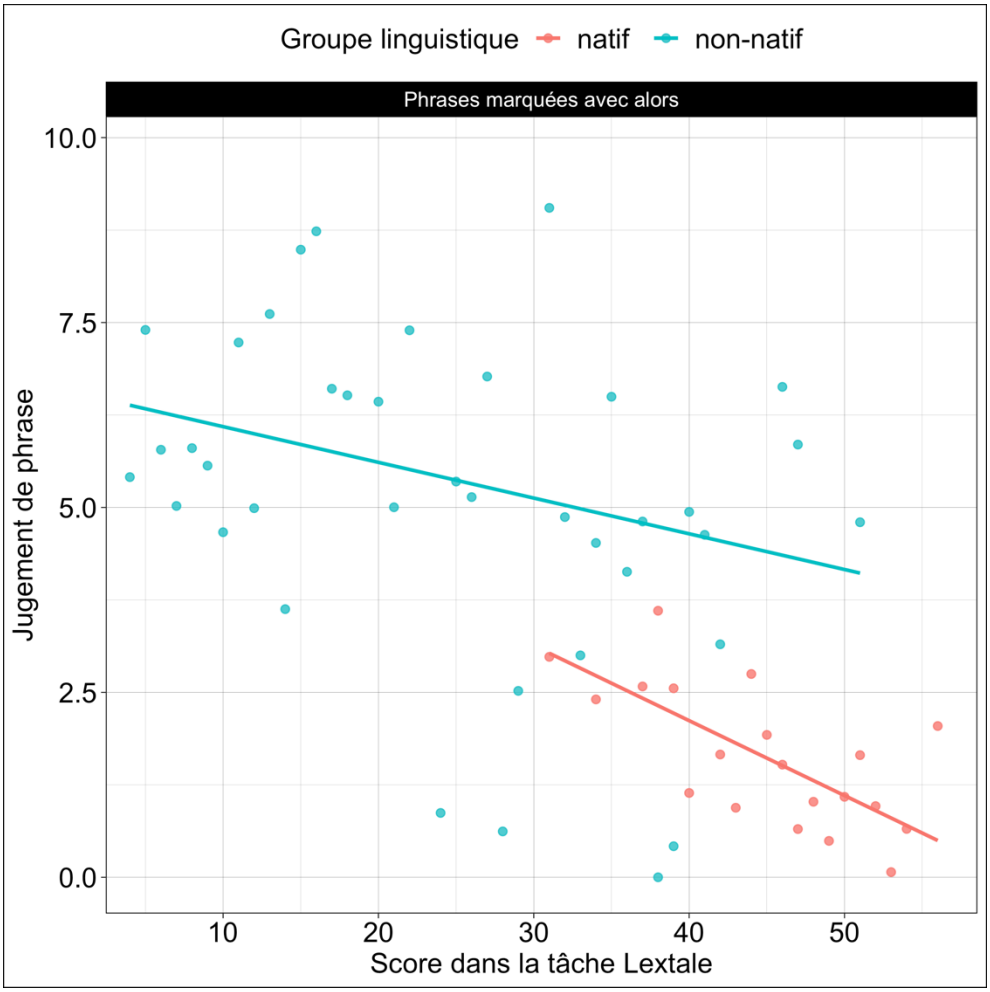
Ces résultats indiquent que plus une personne obtenait un score élevé dans la tâche de *Lextale*, plus elle jugeait les phrases marquées avec *alors* comme incorrectes. D'ailleurs, cela était également le cas pour les personnes natives : lors de l'ajout des *Scores de Lextale* comme effet fixe, le modèle s'est également amélioré de manière significative ($\chi(1) = 110.95$, $p < .001$). L'output de ce modèle, contenant également les *Scores de Lextale* comme effet fixe et les *Participant-e-s* et les *Phrases* comme effet aléatoires, est reporté dans le Tableau 8.3.

Tableau 8.3. Output du modèle final (variable indépendante : les scores pour les phrases marquées avec *alors*), personnes natives

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	6.23	1.45	102.75	4.31	$< 3.81^{e-05}$
Scores de Lextale	-0.10	.03	98.35	-3.38	< .005

Les résultats sont également illustrés par la Figure 8.3.

Figure 8.3. Expérience 1, les scores dans la tâche *Lextale* et les jugements pour les phrases incorrectement marquées avec *alors*.



8.2.4 Discussion

Dans cette expérience, des participant-e-s natifs et non-natifs (germanophones) ont évalué 30 phrases contenant des spécifications en français. Certaines spécifications étaient correctement indiquées par le connecteur *c'est-à-dire*, d'autres ne contenaient pas de connecteur et d'autres encore contenaient le connecteur *alors* pour lequel nous nous attendions à des effets de transfert négatif. Malgré cette difficulté pour les germanophones, nous nous attendions également à ce qu'ils soient en mesure de distinguer les phrases correctes des phrases incorrectes.

Nous avons observé que les personnes non-natives étaient effectivement capables de discriminer les phrases correctes (implicites ou marquées avec *c'est-à-dire*) des phrases incorrectes (marquées avec *alors*), malgré un transfert négatif potentiel. Cette observation semble être en contradiction avec les résultats de

Zufferey et collègues (2015) qui ont démontré que les participant-e-s ne détectaient pas les emplois erronés des connecteurs dans une tâche d'évaluation similaire. Une des raisons de cette différence pourrait être la relation de cohérence testée. Par rapport aux relations testées par Zufferey et collègues (conditions et contrastes), un connecteur indiquant une spécification est beaucoup plus optionnel et risque d'être ignoré plus facilement. Nous y reviendrons dans la conclusion.

Bien que les personnes non-natives discriminaient les marquages corrects des marquages incorrects d'une spécification, on observe également des indices que cela n'était pas si facile pour eux. En effet, les jugements des spécifications indiquées par *alors* réalisés par les personnes non-natives étaient en moyenne de 5.54 (écart type = 3.33, la médiane étant même à 6.3) ce qui correspond sur l'échelle utilisée plutôt à « Je suis sûr-e que c'est correct » qu'à « Je suis sûr-e que c'est incorrect ». Cela ne peut pas être dû à un manque de compétence linguistique, car les phrases de remplissage incorrectes ont été plus clairement rejetées.

Nous avons également pu prédire la capacité à détecter les faux emplois de *alors* selon le niveau de compétence linguistique de nos participant-e-s, tel que mesuré par *Lextale*. Il semble donc que le pouvoir prédictif de la compétence linguistique (mesuré par *Lextale*) pour les connaissances explicites des connecteurs soit un effet stable. Nous avions en effet également observé que les scores plus hauts dans cette tâche prédisaient des meilleurs scores dans une tâche d'insertion de connecteur (chapitre 5) ainsi qu'une meilleure détection des emplois erronés sans transfert négatif (chapitre 7). Il convient de noter à cet égard que l'effet prédictif des scores de *Lextale* s'applique également aux personnes natives. Cela souligne la grande variance de la maitrise des connecteurs chez les personnes natives (p.ex., Nippold et al., 1992; Lamiroy, 1994; Zufferey & Gygax, 2020a, 2020b; Tskhovrebova et al., 2022).

Finalement, il semble que les personnes non-natives hésitaient plus à accepter une spécification qui était marquée explicitement (et correctement) avec le connecteur *c'est-à-dire* qu'une spécification qui n'était pas indiquée par un connecteur. Cela indique que la présence d'un connecteur, même s'il est correct, n'est pas automatiquement bénéfique pour les personnes non-natives.

En conclusion, nous avons observé que les personnes non-natives sont généralement capables de détecter des emplois erronés des connecteurs malgré

un transfert négatif potentiel. Elles ont toutefois évalué les spécifications par *alors* comme étant plus correctes qu'incorrectes. Ainsi, la question se pose de savoir si les personnes non-natives réagissent aux emplois erronés de *alors* lorsqu'elles accèdent à leurs connaissances implicites. Pour rappel, nous avons vu au chapitre 7 que la mauvaise utilisation de *cependant* n'a été remarquée que lorsque les participant-e-s accédaient à leurs connaissances explicites.

8.3 Expérience 2 : la détection de l'emploi erroné de *alors* lors de l'accès aux connaissances implicites

Nous avons observé dans l'expérience précédente que les personnes non-natives étaient capables de discriminer les spécifications correctement marquées des spécifications qui étaient incorrectement marquées avec *alors*, lorsqu'elles accédaient à leurs connaissances explicites. Comme nous savons que la détection d'un emploi erroné d'un connecteur dépend du type de connaissances (chapitre 7), la question se pose de savoir si les personnes non-natives réagissent également à la mauvaise utilisation de *alors* lors de la lecture.

Comme nous nous intéressons à la relation de spécification dans laquelle la cohérence est établie de manière intra-phrastique et où le marquage par un connecteur est hautement optionnel, nous pensons que les effets de transfert négatif pour la connaissance implicite seront un peu différents de ceux de Zufferey et al. (2015). Rappelons que Zufferey et al. (2015) ont observé que les personnes non-natives ne réagissaient aux erreurs de connecteurs par transfert négatif que lorsqu'elles accédaient à leurs connaissances implicites.

Nos hypothèses diffèrent à cet égard : bien que nous sachons que les personnes non-natives peuvent être sensibles aux mauvaises utilisations des connecteurs lorsqu'elles accèdent à leurs connaissances implicites (chapitres 6 et 7), nous pensons que cela n'est pas le cas dans le cadre de l'étude actuelle car nous compliquons la détection de l'incohérence en utilisant une relation de cohérence dans laquelle l'utilisation d'un connecteur est hautement optionnelle. Nous supposons ainsi que les germanophones ne vont pas ralentir leur lecture pour une spécification qui est introduite avec *alors,* car ils vont négliger l'instruction erronée du connecteur.

Concernant la différence entre les spécifications implicites et explicites (indiquées par *c'est-à-dire*), nous supposons que les phrases avec le connecteur *c'est-à-dire* seront lues aussi rapidement que les phrases implicites. Même si les participant-e-s de la première expérience jugeaient les phrases contenant *c'est-à-dire* comme étant moins correctes que les phrases implicites, ils les jugeaient tout de même comme étant plutôt correctes. Ainsi, ils ne devraient pas ralentir leur lecture après avoir lu un connecteur qu'ils estiment être correct.

8.3.1 Les participant-e-s

Nous avons recruté 51 participant-e-s germanophones (âge moyen = 28.26 ans, écart type = 8.62, 21 femmes) sur la plateforme *Prolific* (Prolific, London, 2020) qui indiquaient parler le français. Pour le groupe de contrôle, nous avons recruté 63 participant-e-s francophones (âge moyen = 29.3, écart type = 10.3, 13 femmes).

Le groupe natif a obtenu un score moyen de 80.95% (95% IC [80.50; 81.40]) dans la tâche de *Lextale*, tandis que le groupe non-natif a obtenu un score moyen de 36.24% (95% IC [35.24; 37.23]).

Un test de χ^2 (calculé à l'aide de la fonction *chisq.test()* du package *stats*, R Core Team, 2020) ne montrait pas de différence significative ($\chi(1) = 5.68^{e-30}$, $p = 1$) entre les scores des participant-e-s de l'expérience actuelle et de la première expérience (pour rappel, dans la première expérience le score du groupe non-natif était de 38% et celui du groupe natif de 84%).

Comme les scores de *Lextale* indiquaient un niveau de langue comparable à celui des participant-e-s de la première expérience, nous n'avons pas recruté davantage parmi les étudiant-e-s universitaires. Comme dans la première expérience, le groupe non-natif de l'expérience actuelle estimait avoir un *bon* à *très bon* niveau en français (*lecture* : moyenne = 61, écart type = 22; *compréhension à l'oral* = 61, écart type = 22; *production à l'oral* = 59, écart type = 22).

Tous les participant-e-s ont montré une bonne participation aux expériences précédentes sur la plateforme *Prolific* (un minimum de 95% de participations approuvées) et ont été rémunérés avec 2.10 GBP dans notre expérience. Tous les participant-e-s ont donné leur consentement à participer à l'expérience.

8.3.2 La procédure

L'expérience a été programmée avec le logiciel *Psychopy* (version 2020.20) et hébergée sur les serveurs de *Pavlovia*. Les participant-e-s y ont accédé via un lien Internet. D'abord le formulaire de consentement a été présenté sur le site *Qualtrics* (Qualtrics LLC, Provo [UT], Etats-Unis). Ensuite, les participant-e-s ont été dirigés de manière aléatoire vers l'une des listes.

Dans celle-ci, les phrases testées étaient les mêmes que dans la première expérience. Nous avons divisé ces phrases en cinq segments, comme représenté dans (14).

14) Mon restaurant de tous les jours, (1) // [*c'est-à-dire*] [*alors*] [Ø] une pizzeria (2) // qui se trouve derrière la gare, (3) // vient (4) //de fermer. (5) //

La spécification figurait toujours dans les segments 2 et 3. La manipulation (utilisation de *c'est-à-dire*, *alors* ou omission du connecteur) a été faite dans le deuxième segment. Pour permettre une fluidité de lecture naturelle, nous avons ajouté dans le segment manipulé un premier élément de la spécification (dans (14) p.ex., *une pizzeria*). Ainsi, seuls les temps de lecture pour les segments 3, 4 et 5 ont pu être analysés, car ils étaient identiques dans toutes les conditions. Contrairement aux expériences précédentes (chapitres 4, 6 et 7), les effets de perturbation de la lecture étaient plutôt attendus au milieu de la phrase, à savoir dans les segments 3 et 4, car l'insertion de la spécification devait être intégrée dans la phrase principale.

Afin de nous assurer que les participant-e-s lisaient attentivement les phrases, nous avons ajouté après chaque phrase une question de vérification, comme en (15) (voir également les chapitres 4, 6 et 7).

15) Le restaurant a fermé. Vrai ou faux ?

Les questions ne se référaient pas au contenu de la spécification insérée (p.ex., *une pizzeria qui se trouve derrière la gare*) mais plutôt à la proposition principale (*le restaurant vient de fermer*).

8.3.3 L'analyse

Nous avons analysé les temps de lecture en utilisant la même approche que dans la première expérience. Seulement cette fois, nos variables dépendantes étaient

les temps de lecture pour chaque segment. Comme dans les chapitres 6 et 7, nous avons fait des modèles séparés pour les deux groupes linguistiques (Crible et al., 2021). Comme les segments étaient un peu plus longs en comparaison avec les phrases testées aux chapitres 4, 6 et 7, nous avons utilisé un seuil minimum de 50ms et un seuil maximum de 5s.[30] Nous avons également fait une transformation logarithmique ce qui a encore réduit l'asymétrie de nos données (contrôlée à l'aide de la fonction *skewness()* du package *stats*, R Core Team, 2020).

Pour le segment 3, suivant directement le connecteur, l'ajout du *Type de connecteur* pour le groupe non-natif n'a pas amélioré le modèle ($\chi(2)= .80$, $p = .67$). En revanche, son ajout a amélioré le modèle de manière significative pour les personnes natives ($\chi(2) = 7.89$, $p < .05$). L'output du modèle final pour le groupe natif, contenant le *Type de connecteur* comme seul effet fixe et les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires, est reporté dans le Tableau 8.4.

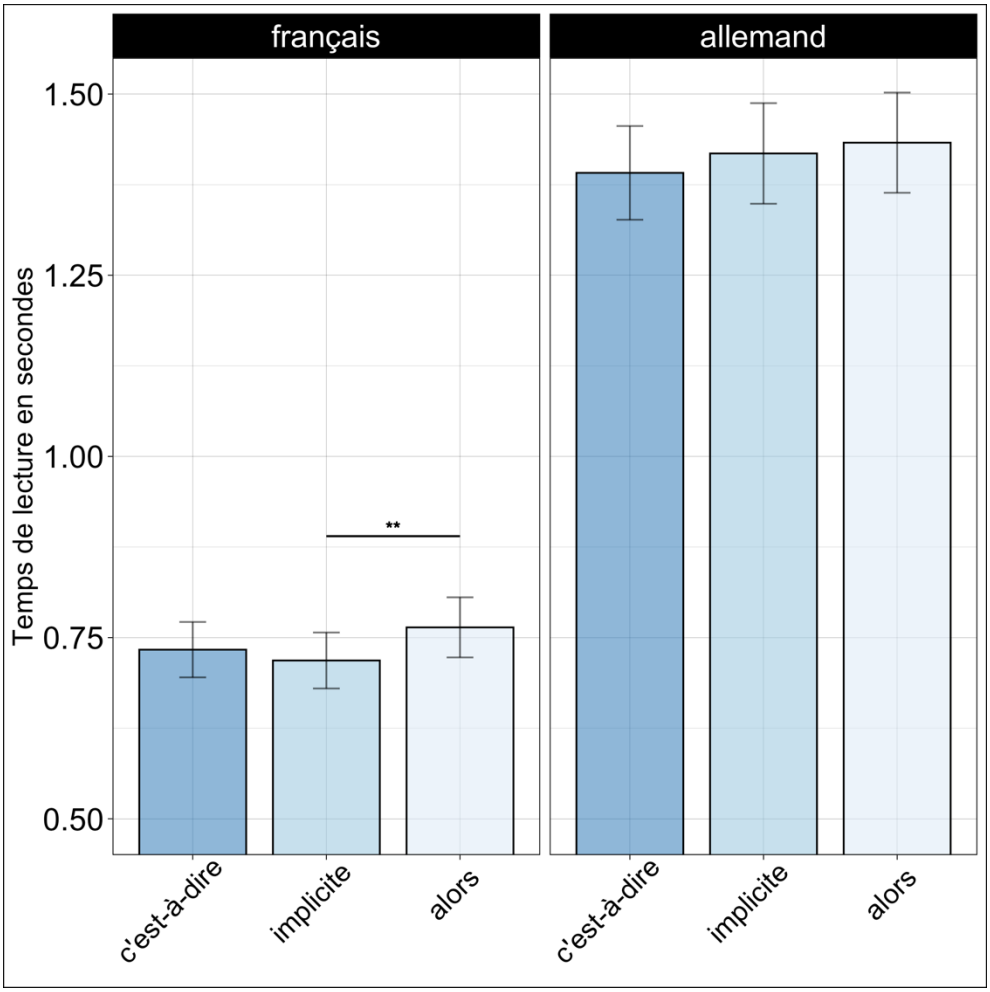
Tableau 8.4. L'output du modèle final pour le segment 3 (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.47	.06	82.50	-7.28	$< 1.78^{e-10}$
Type de connecteur (*c'est-à-dire*)	-0.02	.02	1787.36	-0.98	.33
Type de connecteur (*implicit*)	-0.06	.02	1787.33	-2.77	< .01

Nous observons que le groupe natif a lu le troisième segment des phrases marquées avec *alors* significativement plus lentement que les phrases implicites, un effet qui est également illustré, de même que les temps de lecture des personnes non-natives, à la Figure 8.4.

[30] Des analyses supplémentaires ont montré exactement les mêmes effets lorsque le seuil maximal des temps de lecture a été fixé à 3 s.

Figure 8.4. Les temps de lecture pour le segment 3 (groupe natif et non-natif). L'IC de 95% est indiquée par les barres d'erreurs



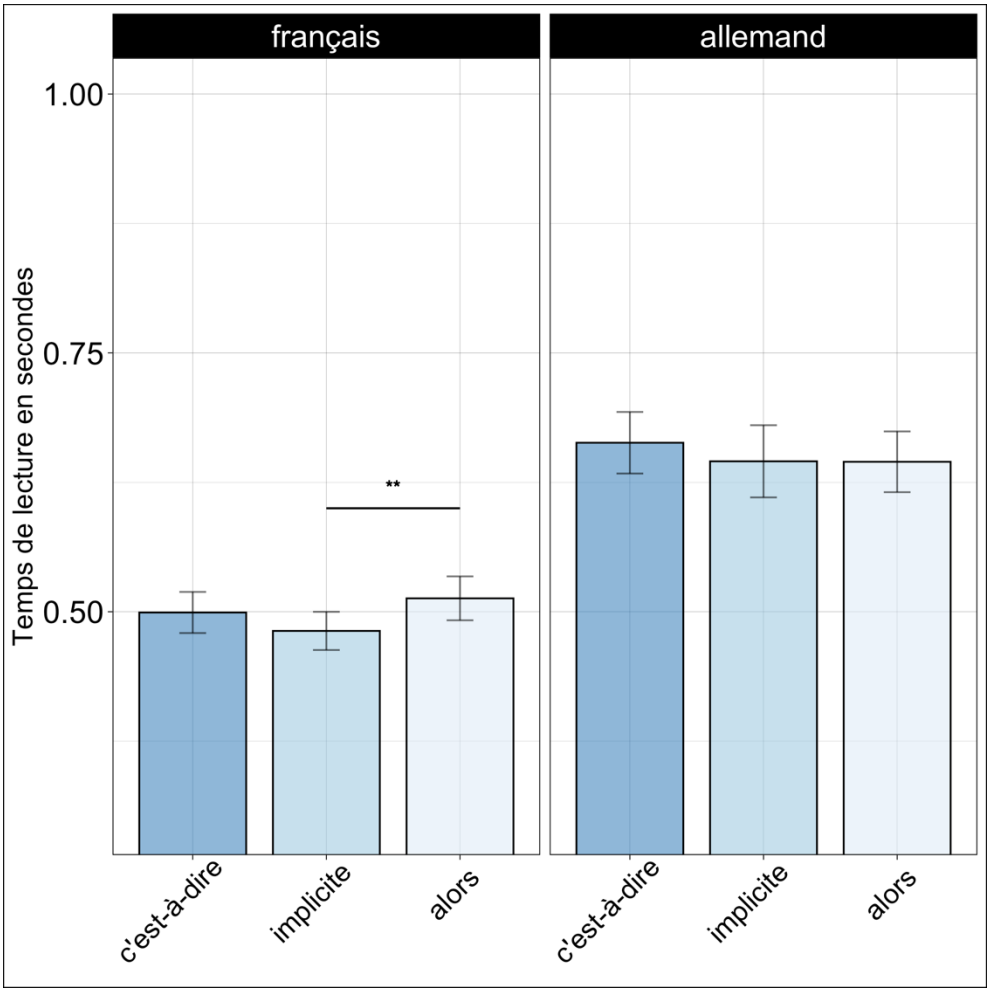
Cet effet se prolonge durant la lecture du segment suivant : lors de l'ajout du *Type de connecteur* comme effet fixe pour le modèle du groupe natif, le modèle s'est amélioré de manière significative ($\chi(2) = 7.76$, $p < .05$). L'output du modèle final pour les personnes natives, contenant le *Type de connecteur* comme seul effet fixe et les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires, est reporté dans le Tableau 8.5.

Tableau 8.5. L'output du modèle final pour le segment 4 (personnes natives)

	β	*ET*	*ddl*	*t*	*Pr(>|t|)*
(Intercept)	-0.76	.04	86.51	-21.05	$< 2^{e-16}$
Type de connecteur (*c'est-à-dire*)	-0.02	.02	1790.34	-1.41	.16
Type de connecteur (*implicit*)	-0.05	.02	1790.40	-2.79	< .01

Pour le groupe non-natif, l'ajout du *Type de connecteur* comme effet fixe n'a pas amélioré le modèle ($\chi(2) = 2.86$, $p = .24$). Les temps de lecture pour le segment 4 sont illustrés à la Figure 8.5.

Figure 8.5. Les temps de lecture pour le segment 4 (groupe natif et non-natif). L'IC de 95% est indiquée par les barres d'erreurs



Concernant le dernier segment de la phrase, le segment 5, l'ajout du *Type de connecteur* comme effet fixe n'a pas amélioré le modèle, ni pour le groupe natif ($\chi(2) = 2.01$, $p = .37$), ni pour le groupe non-natif ($\chi(2) = .13$, $p = .94$). Cela indique que tous les participant-e-s ont lu le dernier segment avec la même rapidité dans les trois conditions. Finalement, nous n'avons pas non plus trouvé d'effet en analysant les temps de réponse aux questions de vérification : l'ajout du *Type de connecteur* comme effet fixe n'a pas amélioré le modèle, ni pour le groupe natif ($\chi(2) = 1.16$, $p = .56$), ni pour le groupe non-natif ($\chi(2) = 1.97$, $p = .37$).

En résumé, nous observons un ralentissement de lecture temporaire (dans les segments 3 et 4) chez le groupe natif pour les phrases marquées par le

connecteur *alors*. Pour les personnes non-natives, nous n'observons pas d'effet en fonction du type de connecteur tout au long de la phrase.

Les scores *Lextale* n'ont été à aucun moment prédictifs des temps de lecture, ni pour le groupe natif, ni pour le groupe non-natif.

8.3.4 Discussion

Dans cette expérience, nous avons mesuré les temps de lecture des personnes natives et non-natives (germanophones) lorsqu'elles lisaient des phrases contenant des spécifications, présentées dans trois conditions. Soit les spécifications étaient marquées correctement par le connecteur *c'est-à-dire*, soit elles étaient marquées incorrectement par *alors*, soit elles n'étaient pas marquées par un connecteur, ce qui représente également une manière correcte de communiquer une spécification. L'utilisation de *alors* pour indiquer une spécification était supposé rendre – par similarité avec le connecteur allemand *also* – la détection de l'incohérence plus difficile pour les personnes non-natives.

Nous avons observé que les temps de lecture des personnes non-natives n'ont été à aucun moment affectés par ces trois conditions. Pour le groupe natif, en revanche, nous avons observé que les spécifications indiquées par *alors* provoquaient des ralentissements de la lecture aux segments 3 et 4. Cet effet est réminiscent des résultats obtenus par Murray (1997) qui montrent que certains connecteurs mal-utilisés perturbent davantage la lecture que d'autres en fonction de la relation de cohérence qu'ils indiquent.

Cependant, dans notre expérience, les effets de perturbation de la lecture chez le groupe natif étaient plutôt subtils et seulement temporaires, bien que les phrases aient été jugées clairement fausses par des personnes natives lors de la première expérience. Le fait que l'incohérence n'ait pas entrainé une perturbation persistante de la lecture s'explique par le fait que l'utilisation incorrecte de *alors* ne change pas le sens principal de la phrase et que la spécification est toujours comprise. A cela s'ajoute que nous n'avons observé une différence qu'entre les spécifications qui ont été indiquées par *alors* et les versions implicites. Le fait que les spécifications explicites aient été lues à la même vitesse, que les connecteurs utilisés soient corrects ou incorrects, montre à quel point l'incohérence intra-phrastique est subtile.

Nous concluons donc que l'incohérence intra-phrastique et l'optionalité du connecteur ne provoquent qu'un effet subtil de perturbation de la lecture lorsque le connecteur est utilisé incorrectement – et cela seulement pour le groupe natif.

8.4 Conclusion

Cette étude s'inscrit dans la lignée des expériences qui investiguent la détection d'incohérences en L2 causées par une mauvaise utilisation des connecteurs (voir aussi les chapitres 6 et 7). Dans cette expérience, nous avons examiné dans quelle mesure des effets crosslinguistiques négatifs empêchent les personnes non-natives de détecter un mauvais emploi d'un connecteur. Dans une première expérience, nous avons demandé aux participant-e-s d'évaluer consciemment des phrases contenant des spécifications qui ont été laissées implicites, marquées incorrectement par le connecteur *alors* ou marquées correctement par *c'est-à-dire*. Dans une deuxième expérience, nous avons mesuré les temps de lecture pour ces phrases.

Nous avons observé que les personnes non-natives détectaient l'erreur lorsqu'elles accédaient à leurs connaissances explicites (Expérience 1), alors que l'erreur ne provoquait pas de perturbation de lecture (Expérience 2). Ainsi, nos résultats semblent être en contradiction avec ceux de Zufferey et al. (2015). Pour rappel, dans cette étude, les chercheur-se-s ont trouvé le pattern inverse : leurs participant-e-s n'ont détecté les erreurs qu'en accédant à leurs connaissances implicites, alors qu'elles sont passées inaperçues lorsque les participant-e-s ont accédé à leurs connaissances explicites. Cette divergence entre nos résultats et ceux de Zufferey et al. (2015) peut s'expliquer par le fait que nous avons testé des *spécifications*.

Les spécifications peuvent être comprises même avec un connecteur qui donne une indication erronée. Ainsi, la représentation mentale de la phrase (Kintsch, & Van Dijk, 1978) ne dépend pas toujours du connecteur. En outre, il n'est même pas obligatoire de marquer une spécification explicitement avec un connecteur, en fait, il semble même que la norme du français soit plutôt de ne pas marquer une spécification explicitement. Comme la phrase peut être interprétée grâce à d'autres moyens (tel que le contenu sémantique ou la ponctuation), l'indication (erronée ou non) d'un connecteur est plus facilement négligée. Des

études futures pourraient explorer davantage dans quelle mesure les connecteurs optionnels sont plus facilement ignorés par rapport aux connecteurs qui sont plus nécessaires pour l'identification de la relation de cohérence (voir aussi Murray, 1997).

Les phrases à cohérence intra-phrastique semblent ainsi avoir des propriétés qui empêchent les personnes non-natives de détecter une utilisation incorrecte d'un connecteur qui est due au transfert négatif lors de la lecture. Cela n'a cependant pas empêché les personnes non-natives de notre expérience de distinguer les phrases correctes des phrases incorrectes lorsqu'elles accédaient à leurs connaissances explicites. A cet égard, nous avons également vu que la capacité à détecter l'erreur a été prédite par les scores de la tâche *Lextale*. Le lien entre les connaissances explicites des personnes non-natives sur les connecteurs et la taille de leur vocabulaire a déjà été observé pour les scores d'une tâche d'insertion de connecteur (chapitre 5) ainsi que la détection de l'incohérence sans effet de transfert (chapitre 7). Comme cette corrélation semble également exister pour la détection de l'incohérence malgré un transfert négatif, nous pouvons donc conclure qu'elle est assez persistante. Il est intéressant à cet égard que nous n'ayons pas trouvé un lien entre la taille du vocabulaire et les connaissances implicites, ni dans ce chapitre, ni dans les chapitres 4, 6 ou 7. Il semble donc que la taille du vocabulaire ne prédise pas la facilité avec laquelle les personnes non-natives intègrent les significations des connecteurs lors de la lecture. Nous reviendrons sur cette question dans la conclusion de la thèse au chapitre 10.

Évidemment, notre étude a certaines limites que nous aimerions brièvement discuter. Premièrement, nous n'avons pas testé, contrairement à Zufferey et al., (2015), les mêmes participant-e-s dans les deux expériences. Bien que cela puisse constituer un biais, le nombre de participant-e-s était suffisamment grand dans nos expériences (voir d'autres études sur le sujet, p.ex., Degand & Sanders, 2002) et nous avons veillé à ce que les participant-e-s des deux expériences aient le même niveau de langue (mesuré par *Lextale*). La question de savoir dans quelle mesure les scores de *Lextale* reflètent les compétences linguistiques est abordée au chapitre 10.

Deuxièmement, comme nous n'avons testé que trois conditions (pas de connecteur, *c'est-à-dire*, *alors*), ne savons pas comment un connecteur complétement faux, pour lequel aucun effet de transfert négatif n'existe, aurait été

évalué par les personnes non-natives.[31] De plus, il se peut théoriquement que les personnes non-natives n'auraient pas réagi non plus à une utilisation erronée d'un autre connecteur lors de la deuxième expérience. Nous n'avons pas inclu une quatrième condition parce que cela aurait mis en évidence la présence des spécifications dans les phrases expérimentales et aurait attiré l'attention sur *alors*. Par conséquent, beaucoup plus de phrases de remplissage auraient été nécessaires, ce qui aurait rendu l'expérience impraticable.

Se peut-il donc que les effets observés soient uniquement dus aux propriétés de la spécification et non au transfert négatif avec l'allemand *also* ? Nous pensons que cela est peu probable. En effet, nous avons observé dans l'Expérience 1 plusieurs indications que le connecteur du français *alors* était effectivement confondu avec l'allemand *also* : les phrases contenant *alors* étaient évaluées comme étant plus correctes qu'incorrectes. Cela indique que, contrairement au groupe natif, les germanophones hésitaient à rejeter systématiquement *alors* pour indiquer une spécification. Comme d'autres phrases incorrectes (à savoir les phrases de remplissage) ont été plus facilement identifiées comme incorrectes, nous pouvons conclure que cela n'est pas dû seulement à un manque de compétences linguistiques.

En conclusion, notre étude a montré que le transfert négatif est un facteur qui n'affecte pas seulement la production des connecteurs en L2 (p.ex., Granger & Tyson, 1996), mais aussi la détection d'erreurs lorsque les connecteurs sont mal utilisés (Zufferey et al., 2015). A cet égard, nous avons pu nuancer cette image en soulignant l'importance du type de relation de cohérence. De plus, les résultats obtenus dans ce chapitre, de même que ceux du chapitre précédent, soulignent l'importance de la distinction entre les connaissances explicites et implicites et montrent que la maitrise des connecteurs d'une L2 peut être développée davantage dans un type de connaissance que dans l'autre.

Avant de conclure la présente thèse au chapitre 10, nous allons maintenant examiner encore deux propriétés des connecteurs qui sont, tout comme le transfert

[31] Une autre idée serait de tester deux groupes non-natifs, l'un qui pourrait être affecté par transfert négatif (germanophone), l'autre qui ne serait pas affecté (p.ex., anglophone). Cependant, il est très difficile de recruter deux groupes non-natifs avec niveau de langue totalement équivalent.

de la L1, censées influencer la maitrise des connecteurs dans une deuxième langue. Il s'agit des contraintes pragmatiques et syntaxiques des connecteurs.

9 La sensibilité aux contraintes syntaxiques et pragmatiques lors de l'accès aux connaissances explicites

Nous avons observé à plusieurs reprises que les problèmes des personnes non-natives avec les connecteurs ne s'appliquent pas systématiquement à tous les connecteurs, mais se limitent plutôt à certains connecteurs spécifiques. Par exemple, tandis que les connecteurs *donc* et *mais* ont provoqué des ralentissements de la lecture lorsqu'ils étaient utilisés incorrectement (chapitre 6), cela n'était pas le cas pour *cependant* (chapitre 7). Il ne semble toutefois pas y avoir pas une seule raison qui permette d'expliquer pourquoi certains connecteurs tels que *cependant* sont plus difficiles que d'autres. Nous avons vu par exemple que certains connecteurs semblent être plus difficiles à maitriser parce qu'ils ne sont pas fréquents (chapitres 4, 5 et 7), alors que d'autres, bien qu'ils soient tout aussi rares, sont bien maitrisés, ce qui s'explique par leur transparence sémantique (chapitre 5). De plus, nous avons vu que la maitrise des connecteurs ne semble pas dépendre seulement des connecteurs eux-mêmes, mais aussi du type de connaissances que les apprenant-e-s peuvent activer. Nous avons observé à cet égard que les personnes non-natives étaient capables de reconnaitre une utilisation incorrecte d'un connecteur lorsqu'elles accédaient à leurs connaissances explicites, mais qu'elles n'y réagissaient pas lorsqu'elles accédaient à leurs connaissances implicites (chapitres 7 et 8).

Bien que nos observations montrent une image plutôt cohérente de la maitrise des connecteurs en L2, nos études expérimentales n'ont pas encore tenu compte d'une caractéristique importante des connecteurs : leurs fonctions *pragmatiques*. En effet, dans toutes les expériences menées, les phrases expérimentales utilisées n'évoquaient pas la dimension pragmatique.

En général, la dimension pragmatique d'une langue se réfère aux implications contextuelles d'une conversation, c'est-à-dire la prise en compte du contexte dans lequel un énoncé est produit. Ainsi, une variable importante sont les intentions des locuteurs-trice-s. Moeschler et Reboul (1996) précisent :

> « Une des tâches de la pragmatique est d'expliquer comment un auditeur peut arriver à comprendre une énonciation de manière non littérale et pourquoi le locuteur a choisi un mode d'expression non littérale plutôt qu'un mode d'expression littérale. » (1994 : 22)

La pragmatique inclut ainsi les *actes de langage* (Austin, 1962; Searle 1969), les *implicatures conversationnelles* (Grice, 1975), ainsi que d'autres propriétés sous-jacentes des énoncés, comme par exemple les utilisations des particules modales pour montrer des attitudes différentes envers le contenu propositionnel (Weydt, 1969; Walterweit, 2001; Cuenca, 2013). VanPatten et Benati (2010) résument :

> « La pragmatique s'intéresse (...) au rôle du contexte dans le langage et à la façon dont les locuteur-trice-s et les interlocuteur-trice-s s'appuient sur ce contexte pour communiquer avec succès » (2010 : 164, ma traduction).

Nous savons que certains connecteurs peuvent avoir un profil pragmatique très spécifique (Moeschler & Reboul, 1994; voir aussi le chapitre 2.6) et il est probable que ces fonctions pragmatiques puissent rendre ces connecteurs plus difficiles à maitriser que d'autres. Une première indication en ce sens est l'observation que les personnes interrogées choisissaient dans la tâche d'insertion au chapitre 5 pour les phrases ciblant *puisque* – un connecteur pragmatiquement complexe - seulement dans 56% des cas ce connecteur. La question se pose ainsi de savoir si les personnes non-natives connaissent les contraintes pragmatiques de *puisque* et en tiennent compte lorsqu'il est question de choisir entre des connecteurs causaux différents.

Outre les contraintes pragmatiques, il convient de noter que les connecteurs possèdent également des contraintes *syntaxiques*. Alors que certains se placent de préférence à certaines positions syntaxiques, d'autres doivent même obligatoirement se trouver à certaines positions dans une phrase. Ces contraintes sont particulièrement importantes à connaitre pour les apprenant-e-s : pour pouvoir rédiger correctement un texte en L2, il faut savoir où un connecteur peut être placé dans la phrase afin d'éviter les erreurs - d'autant plus que certains connecteurs changent leur fonction selon leur position syntaxique.

Pour examiner plus en détail le rôle des contraintes qui conduisent à utiliser certains connecteurs et à en exclure d'autres, nous allons explorer dans ce chapitre la question de savoir si – et à quel degré – les personnes non-natives sont sensibles aux contraintes pragmatiques et syntaxiques des connecteurs causaux *puisque* et

parce que. Pour ce faire, nous allons mesurer la connaissance explicite des participant-e-s pour ces contraintes en utilisant une tâche d'insertion.

9.1 Les contraintes qui conduisent à utiliser un connecteur spécifique

Lors de la rédaction d'un texte, il convient de choisir entre plusieurs options le meilleur connecteur pour indiquer une relation de cohérence donnée. Dans certains cas, plusieurs options semblent appropriées, tandis que dans d'autres cas, le choix est plus limité.

Par exemple, en (1), les connecteurs *mais*, *bien que* ou *en dépit du fait que* sont - ou au moins paraissent théoriquement - des choix adéquats.

1) Il ne fait pas froid, ____ il y a un fort vent.

Cependant, tous ces connecteurs ne sont pas possibles. Par exemple, le connecteur *bien que* n'est pas possible car il exige le subjonctif. De même pour *en dépit du fait que* qui pourrait être considéré comme inadéquat car il ne correspond pas au registre de la situation.[32] Parmi les connecteurs mentionnés, seul *mais* convient dans ce cas.

Cet exemple permet également de constater que les contraintes diffèrent l'une à l'autre : tandis que l'utilisation de *bien que* est inappropriée car agrammaticale, l'utilisation de *en dépit du fait que* est inappropriée sur le niveau pragmatique. Le défi d'un-e apprenant-e consiste à identifier ces contraintes afin d'utiliser non seulement un connecteur qui indique la bonne relation de cohérence, mais qu'il le fasse de manière grammaticalement et pragmatiquement appropriée.

Dans ce qui suit, nous allons discuter si les contraintes syntaxiques et pragmatiques des connecteurs compliquent leur maitrise dans une L2.

9.1.1 Les contraintes syntaxiques des connecteurs

Nous avons mentionné au chapitre 2 que les connecteurs sont des éléments à la jonction de plusieurs dimensions de la langue, comme la pragmatique ou la sémantique. Qu'en est-il de la syntaxe ?

[32] Pourvu que la situation soit une conversation décontractée entre amis

Premièrement, il convient de noter que de nombreux connecteurs ne peuvent pas être placés partout dans une phrase, comme nous voyons dans l'exemple (2).

2) *Il pleut je vais dehors mais.

Évidemment, la position syntaxique de *mais* est incorrecte dans cet exemple, mais il existe également des cas un peu moins clairs, par exemple lorsqu'une position syntaxique d'un connecteur est incorrecte alors qu'un autre connecteur - indiquant pourtant la même relation de cohérence – convient parfaitement. Considérons l'exemple (3), tiré de LexConn (Roze et al., 2012), dans lequel le connecteur contrastif *au contraire* est correctement positionné alors que d'autres connecteurs contrastifs, tel que *mais*, seraient incorrects dans cette position, comme illustré en (4).

3) Julie adore le cinéma. Marie, *au contraire*, déteste ça.
4) *Julie adore le cinéma. Marie *mais* déteste ça.

Par ailleurs, la syntaxe peut être également décisive pour la fonction d'un connecteur polyfonctionnel. Considérons le connecteur français *aussi*, qui, en position initiale de la phrase, indique une conséquence, (5), tandis qu'en position médiane, (6), il indique une relation parallèle (*Lexconn*, Roze et al., 2012).

5) Marc a faim. Aussi, il vient au restaurant.
6) Marc a faim. Il vient aussi au restaurant.

Ainsi, pour éviter des erreurs, une personne non-native doit apprendre les contraintes syntaxiques d'un connecteur. A cet égard, la question se pose de savoir si les contraintes syntaxiques des connecteurs posent des problèmes aux personnes non-natives ou si elles représentent plutôt une aide pour choisir le bon connecteur.

D'une part, il est connu que les personnes non-natives font des usages non-natifs de la syntaxe lorsqu'elles utilisent les connecteurs. Par exemple, il a été démontré à plusieurs reprises que les personnes non-natives ont tendance à mettre les connecteurs en position initiale de la phrase (p.ex., Leedham & Cai, 2013; Park, 2013; Merilaine, 2015; Don & Sriniwass, 2017; Shi, 2017).

D'autre part, il est vrai que la syntaxe n'est pas le domaine le plus difficile à apprendre en L2 (Hopp, 2010). Bien que l'acquisition de la syntaxe puisse être affecté par un transfert négatif de la L1 (p.ex., Juffs, 1996), un bon apprentissage de « l'interface syntaxe-discours n'est pas a priori contraint par les propriétés de la

L1 », (Hopp, 2010 : 478, ma traduction). Ainsi, l'acquisition de la syntaxe d'une L2 « ne représente pas nécessairement une difficulté insurmontable lors de l'acquisition d'une L2 par un adulte, ni dans la compréhension hors ligne, ni dans le traitement en ligne » (Hopp, 2010 : 478, ma traduction). Les chercheuses Dulay et Burt (1973) ne recommandent ainsi même pas d'enseigner spécifiquement la syntaxe pour pouvoir réorienter les efforts d'enseignement vers d'autres aspects de la langue plus importants. Comme l'acquisition de la syntaxe en L2 ne semble donc pas impliquer une difficulté élevée, on peut supposer que les contraintes syntaxiques des connecteurs sont plutôt bien maitrisées.

En résumé, nous avons vu que la position d'un connecteur dans une phrase n'est pas arbitraire et qu'une meilleure sensibilité à la syntaxe, c'est-à-dire à ces contraintes et préférences, pourrait faciliter l'utilisation des connecteurs. Tandis qu'il est supposé que les contraintes syntaxiques sont relativement faciles à maitriser en L2 (p.ex., Hopp, 2010), d'autres types de contraintes, telles que les contraintes pragmatiques, semblent être plus difficiles (p.ex., Bardovi-Harlig & Hartford, 1990, 1993).

9.1.2 Les contraintes pragmatiques des connecteurs

Nous avons vu au chapitre 2.5.4 que les connecteurs servent, outre leurs fonctions structurales, des fonctions pragmatiques et il existe, en conséquence, de nombreuses différences subtiles entre les connecteurs sur cette dimension (p. ex., Bell, 2010; Zufferey, 2012).

Pour en donner un exemple, reconsidérons la distinction entre une causalité *objective* et *subjective* (voir section 2.5.4). Comme, dans plusieurs langues, des connecteurs spécifiques sont utilisés pour indiquer soit l'un soit l'autre, le choix entre, par exemple, *da* et *weil* (en allemand) ou entre *omdat* et *want* (en néerlandais) n'est pas arbitraire et dépend du degré d'objectivité de la phrase (Waltereit, 2001; Sanders & Spooren 2015). Cette différence pourrait représenter une difficulté pour les apprenant-e-s et en particulier pour ceux ayant une L1 qui n'indique pas cette distinction à l'aide des connecteurs (telle que l'anglais, Traxler et al., 1997a).

Au chapitre 2.5.4, nous avons également vu un autre exemple d'un type d'une contrainte pragmatique pour un connecteur, l'exemple (7), fourni par Fraser (1998).

7) *A : I thought you would come.*
‘A : Je pensais que tu allais venir.’
B : But / However I did !*
‘B : Mais / *Cependant je l’ai fait !’

Dans ce cas, seul un connecteur concessif convient, à savoir un connecteur impliquant pragmatiquement une protestation (tel que *but* ‘mais’). De nouveau, les apprenant-e-s doivent connaitre non seulement la fonction des connecteurs dans ce cas, mais également leurs propriétés pragmatiques afin de pouvoir choisir le bon.

La question se pose alors de savoir si les personnes non-natives sont capables d’identifier ces contraintes pragmatiques et si certaines contraintes pragmatiques sont plus saillantes que d’autres.

De manière générale, de nombreux travaux indiquent que les personnes non-natives rencontrent des difficultés considérables quant aux éléments pragmatiques d’une langue (Bardovi-Harlig & Hartford, 1990, 1993; Bardovi-Harlig & Dörnyei, 1998). Ainsi, on peut faire l’hypothèse que la maitrise des dimensions pragmatiques subtiles des connecteurs est difficile en L2. Il ne s’agit toutefois que d’une hypothèse, car, jusqu’au présent, nous manquons de données qui permettraient de tirer des conclusions sur la sensibilité des apprenant-e-s à ces contraintes.

Dans ce chapitre, nous allons combler ce vide en analysant dans quelle mesure les personnes non-natives sont capables d’identifier les contraintes pragmatiques et syntaxiques des connecteurs *parce que* et *puisque*. Ces deux connecteurs ont été choisis car ils possèdent, dans certains cas, des contraintes qui les imposent comme seul connecteur possible dans une phrase, comme nous allons le voir maintenant.

9.1.3 Les contraintes pragmatiques et syntaxiques de *parce que* et *puisque*

Parce que et *puisque* sont deux connecteurs français indiquant une relation de *cause-conséquence* (selon Sanders et al., 1992). Tandis que *puisque* est classifié comme un connecteur indiquant une forte implication des locuteurs-trice-s – ou en d’autres termes un haut degré de subjectivité - (e.g., Degand, 2004, Zufferey, 2012, Schumann et al., 2020), *parce que* est généralement classifié comme un connecteur de causalité objective (Degand & Pander Maat, 2003, Degand &

Fagard, 2018; Fagard & Degand, 2008; Lambrecht et al., 2006). Toutefois, cela dépend aussi de la modalité : alors que *parce que* exprime souvent une causalité objective lorsqu'il est utilisé à l'écrit, il est également utilisé dans la langue parlée pour exprimer une causalité subjective (Simon & Degand, 2007, voir également Zufferey et al., 2018).

Bien que *puisque* et *parce que* indiquent la même relation de cohérence, les deux connecteurs ne sont pas toujours interchangeables et le travail de Zufferey (2012, voir aussi 2014) a montré que les personnes natives préfèrent dans certains cas l'un à l'autre.

Dans ce travail, Zufferey (2012) a mesuré les préférences des personnes natives pour les trois connecteurs causaux *car*, *puisque* et *parce que* dans des phrases différentes. Plus précisément, les participant-e-s de cette étude évaluaient l'acceptabilité de phrases qui contenaient ces trois connecteurs. Dans une autre tâche de l'étude, une tâche d'insertion, des participant-e-s choisissaient parmi les trois connecteurs le connecteur approprié pour une phrase donnée qui contenait un blanc. Ensemble, les résultats de ces deux tâches permettent d'identifier les préférences des personnes natives quant à la question de savoir dans quel cas quel connecteur convient le mieux.

Concernant *parce que*, les participant-e-s de Zufferey (2012) montraient une préférence claire pour ce connecteur lorsque la phrase contenait une réponse à une question, tel que (8).

8) Jean: Pourquoi est-il parti? Pierre: ____ il était fatigué.

Aucun-e des participant-e-s ne choisissaient *puisque* dans ce cas. Les personnes se sont également accordées sans univoque à utiliser ce connecteur lorsqu'il figurait dans une structure clivée avec *c'est,* comme dans (9) (tous les exemples tirés de Zufferey, 2012).

9) C'est parce qu'il a trop mangé qu'il est malade.

Non seulement cette phrase a été jugée comme correcte dans 100% des cas, les participant-e-s choisissaient également dans la tâche d'insertion uniquement ce connecteur.

Ces deux observations confirment les descriptions théoriques des connecteurs du Groupe λ-l (1975). En analysant des exemples d'utilisation de *car*,

parce que et *puisque*, ce groupe a identifié plusieurs caractéristiques individuelles de ces trois connecteurs.

Pour ce qui est de *parce que*, le groupe a conclu que ce connecteur « sert à constituer (…) une idée nouvelle » (Groupe λ-l, 1975 : 255) - par exemple dans une réponse à une question en « pourquoi », telle qu'en (8).[33] Dans cet exemple, l'information introduite dans la deuxième clause (i.e., *il était fatigué*) représente la raison pour la conséquence exprimée dans la première (i.e., *il est parti*), et cette raison est inconnue de Jean (car autrement il n'aurait pas demandé). Conformément à cette description théorique, les participant-e-s de Zufferey (2012) ont clairement préféré *parce que* dans ce cas.

Le Groupe λ-l a également constaté que *parce que* est le seul connecteur possible dans une structure clivée telle que (9). Ainsi, tout comme pour la phrase (8), les participant-e-s dans Zufferey (2012) ont montré pour la phrase (9) une préférence claire pour *parce que*.

Quant aux connecteurs *car* et *puisque*, le Groupe λ-l (1975) a constaté qu'ils relient non pas des contenus (tel que *il était fatigué* dans exemple 8) mais sont utilisés pour indiquer des *actes de langages* (voir aussi chapitre 2.5.4). Leur analyse a montré que ces connecteurs sont notamment utilisés pour accomplir des actes d'inférence ou de justification, comme par exemple dans l'exemple (10) (tiré de Groupe λ-l, 1975 : 266)

10) C'est un franc salaud, car il faut appeler les choses par leur nom.

Il s'ensuit que lors de l'utilisation de *puisque,* la personne destinataire « n'est plus un spectateur neutre devant lequel on se justifie, mais un acteur que l'on essaye d'influencer, de persuader, ou de dissuader, en tout cas de transformer » (Groupe λ-l, 1975 : 279). En d'autres termes, *puisque* « est (…) une conjonction qui prend en compte 'l'autre' » (Leeman; 1994 : 121).

Mais il y a également des différences entre *puisque* et *car* : contrairement à *car*, *puisque* a une *fonction échoïque*. Cela veut dire que *puisque* introduit une information qui est censée être connue de l'interlocuteur-trice (e.g., Schuman et al,

[33] Il convient de noter que l'information ne doit pas nécessairement être nouvelle, mais plutôt le fait que l'information introduite est causalement liée à la première. Considérons l'exemple suivant : « Je suis là parce que tu m'as appelé, et non pour mon plaisir » (tiré de Groupe λ-l, 1975 : 261). Ici, l'information introduite par le connecteur (i.e., *tu m'as appelé*) est connue mais pas en tant que raison pour laquelle le locuteur-trice est venu. Ainsi, *parce que* convient comme connecteur.

2020). Le Groupe λ-l (1975) exemplifie cette fonction en contrastant les utilisations de *car* et *puisque* pour la phrase (11) (tirée et adaptée de Groupe λ-l, 1975 : 278).

11) Raconte, ____ tu es venu pour ça.

Il est précisé qu'employant *car* dans cette phrase « le locuteur 'a l'air' de révéler la raison, jusque-là cachée, pour laquelle son interlocuteur est venu. En employant *puisque*, en revanche, il 'a l'air' de se référer à un aveu préalable » (1975 : 278).
La phrase utilisée pour la fonction *échoïque* de *puisque* dans Zufferey (2012) est l'exemple (12).

12) Il est nul en maths, ____ 2 et 2 font 4 et pas 5.

Nous appellerons ce cas de figure *évidence*, car il s'agit de l'introduction d'une information qui est mutuellement partagée[34] et donc évidente. Cependant, les résultats pour l'*évidence* ne sont pas concluants dans Zufferey (2012) : dans la tâche d'insertion les participant-e-s ne préféraient pas *puisque* aux autres connecteurs. Comme indique Zufferey (2012), « ce résultat pourrait être dû au choix de la phrase, pour laquelle une interprétation échoïque n'aurait pas été assez saillante. » (2012 : 152, ma traduction).

La fonction échoïque de *puisque* se réalise également dans d'autres cas, à savoir lorsqu'une information initialement présentée par l'interlocuteur-trice est répétée. Considérons l'exemple (13), pris de Zufferey (2012), ou l'information (que la personne adressée aime la montagne) est répétée par le locuteur-trice.

13) Viens te promener! ____ tu dis que tu aimes la montagne.

De nouveau, l'information n'est pas nouvelle pour la destinataire puisque c'est elle-même qui l'avait initialement introduite. Dans ce cas, que nous appellerons *citation*, les participant-e-s de Zufferey (2012) montraient une claire préférence pour *puisque*.

[34] Il convient de noter que, selon Sperber et Wilson (1986, voir aussi chapitre 2.5.1), nous ne saurons jamais avec certitude si une information est *mutuellement connue* ou *supposée connue* par deux locuteur-trice-s. Ainsi, Sperber et Wilson proposent le terme d'une information *mutuellement manifeste* (1986 : 40), c'est-à-dire une information qui est mutuellement perceptible (comme le bruit d'une voiture) ou mutuellement inférable du contexte (comme des anachronismes). Contrairement à d'autres types d'informations (*connues* ou *supposées*), une information mutuellement manifeste est seulement à la disposition des locuteur-trice-s, même s'ils n'en sont pas forcément conscients.

En résumé, bien que *puisque* et *parce que* indiquent les deux une relation de *conséquence-cause*, ils ont des contraintes qui excluent dans certains cas l'utilisation de l'autre. Nous récapitulons sur la base du travail de Zufferey (2012) dans le Tableau 9.1 les types de contraintes qui rendent l'utilisation des deux connecteurs appropriée ou inappropriée.

Tableau 9.1.Type des contraintes pour *parce que* et *puisque*

	contrainte	*type*
parce que		
	Utilisation en structure clivée (p.ex., avec *c'est*)	syntaxique
	Utilisation pour constituer une nouvelle information (p.ex., en réponse à une question « pourquoi »)	pragmatique
puisque		
	Utilisation pour donner une information déjà connue (*citation* de ce qui a été dit auparavant)	pragmatique
	Utilisation pour donner une information déjà connue (*évidence* de la situation)	pragmatique

Nous allons maintenant tester si les personnes non-natives sont capables d'identifier ces contraintes et si oui, quelles compétences individuelles les aident dans cette tâche.

Il convient toutefois de noter que Zufferey (2012) n'a utilisé qu'une seule phrase expérimentale pour chaque contrainte, ce qui induit le risque que la phrase utilisée ait biaisé le jugement des participant-e-s (voir aussi Zufferey, 2012 : 152). De plus, nous ne savons pas si les résultats s'appliquent à tous les locuteur-trice-s. Les personnes interrogées de cette étude peuvent être considérées comme étant linguistiquement très compétentes, car elles étaient toutes des étudiant-e-s de français comme L1. Tenant compte de la variation individuelle au sein d'un groupe de locuteur-trice d'une langue (voir chapitre 3.5), les résultats ne peuvent pas être généralisés pour tous les locuteurs-trice-s de français.

Ces deux limitations font que les contraintes élaborées ne sont pas nécessairement complètes et nous ne savons pas avec certitudes dans quelle mesure les contraintes mentionnées sont vraiment généralisables. En testant également un groupe contrôle de francophones natifs, notre étude va éclairer davantage la question de savoir comment ces contraintes sont identifiées par un ensemble plus hétérogène de locuteur-trices du français.

9.2 Expérience : La sensibilité aux contraintes syntaxiques et pragmatiques de *parce que* et *puisque*

Dans cette expérience[35], nous allons tester si les personnes natives et non-natives du français sont capables d'identifier les contraintes pour les deux connecteurs français, *parce que* et *puisque*, en utilisant une tâche d'insertion de connecteur, similaire à celle utilisée au chapitre 5. Bien que *parce que* et *puisque* indiquent la même relation de cohérence (selon la taxonomie de Sanders et al., 1992), un seul des deux connecteurs sera toujours approprié dans une phrase, car les phrases expérimentales contiendront des contraintes pragmatiques et syntaxiques qui exigent un seul connecteur et excluent l'autre.

Les phrases ciblant *parce que* seront présentées avec des contraintes syntaxiques et pragmatiques. La première consiste à présenter une phrase avec une structure clivée avec *c'est* ou un adverbe tel que *probablement*. Ces contraintes excluent l'utilisation de *puisque* et imposent *parce que* comme le connecteur correct. La deuxième contrainte est pragmatique : nous présenterons une phrase qui est une réponse à une question. Ainsi, dans ces phrases, seul *parce que* sera correct. En testant ces deux contraintes, nous allons observer si les personnes non-natives sont plus sensibles à une contrainte qu'à l'autre.

Nous supposons qu'elles seront plus sensibles aux contraintes syntaxiques car celles-ci sont comparablement plus faciles à apprendre que les contraintes pragmatiques (Hopp, 2010; Dulay & Burt, 1973). Nous supposons en outre que les locuteurs-trice-s linguistiquement compétent-e-s devraient, en général, identifier plus facilement les contraintes et donc obtenir de meilleurs scores dans la tâche que les personnes moins compétentes. Quant aux personnes natives, les participant-e-s de Zufferey (2012) montraient des préférences claires pour *parce que*, à la fois lorsque la contrainte était syntaxique ou pragmatique. Ainsi, nous supposons que les personnes natives vont être également sensibles à ces deux contraintes.

Concernant *puisque*, les phrases contiendront des contraintes pragmatiques. Plus précisément, nous testerons l'utilisation de *puisque* pour indiquer une information déjà connue, ce qui rend *parce que* inapproprié. Cela sera réalisé en

[35] Les données de cette étude ont été publiées, sous une autre forme, dans le *Journal of Pragmatics* sous le titre « Pragmatic and syntactic constraints on French causal connectives: An evaluation of native and non-native speakers' sensitivity » (Wetzel, Tskhovrebova, Gygax & Zufferey, 2023).

utilisant soit une contrainte de type *citation,* soit en présentant une information qui est nécessairement connue (*évidence*). En comparant l'identification de ces deux types de contraintes pragmatiques, nous pourrons observer si l'une d'entre elles est plus saillante que l'autre pour les personnes non-natives.

Nous supposons que la présentation d'une information répétée (i.e., *citation*) devrait faciliter l'identification de la contrainte en comparaison avec une information généralement connue (i.e., *évidence*). Nous le pensons parce qu'il est impossible de savoir avec certitude si une information est effectivement mutuellement connue et identifiée comme telle (voir aussi Sperber & Wilson, 1986). Pour une information répétée (i.e., *citation*) par contre, il est plus clair qu'il s'agit d'une information déjà connue. De plus, dans l'étude de Zufferey (2012), la phrase contenant une *évidence* ne semblait pas générer de préférences claires pour *puisque*.

Finalement, nous nous attendons également à un impact de la compétence linguistique : nous supposons que les personnes plus compétentes identifient plus facilement les contraintes pragmatiques de *puisque* que les personnes moins compétentes. Nous n'attendons pas à une différence pour ces deux contraintes chez les personnes natives.

9.2.1 Les participant-e-s

Pour le groupe natif, nous avons recruté 63 personnes francophones (âge moyen = 23, écart type = 9.2, 23 femmes) via la plateforme *Prolific* (Prolific, London, 2020). Pour le groupe non-natif, nous avons recruté 75 personnes anglophones qui indiquaient parler couramment le français (âge moyen = 40.8, écart type = 15.8, 60 femmes). Tous les participant-e-s ont été rémunérés par 3.15 GBP et leurs participations dans d'autres études sur la plateforme *Prolific* ont été approuvées dans plus de 95% des cas.

Nous avons mesuré la compétence linguistique en français des deux groupes testés à l'aide de la tâche *Lextale* (Brysbaert, 2013, voir chapitres 4, 5, 6, 7 et 8) et de la *Tâche de Grammaire* (Zufferey & Gygax, 2020, voir chapitres 5, 7). Dans la tâche *Lextale*, le groupe natif a obtenu un score moyen de 48.33 sur 56 (écart type = 4.41, CI 95 % [48.19; 48.47]) ce qui correspond à 86% de bonnes réponses. Dans la *Tâche de Grammaire*, ce groupe a obtenu un score moyen de 66% (écart type = 8.78, CI 95 % [65.72; 66.41]). Le groupe non-natif a obtenu dans la tâche *Lextale* un score moyen de 23.4 sur 56 (écart type = 11.58, CI 95% [23.06;

23.74]) ce qui correspond à 41.79% de bonnes réponses. Dans la tâche de Grammaire, les personnes non-natives ont eu un score moyen de 56.05% (écart type = 7.65, CI 95% [55.77; 56.32].

9.2.2 Le design et la procédure

Les participant-e-s ont pu accéder à l'expérience via un lien qui les dirigeaient vers le site de *Qualtrics* (Qualtrics LLC, Provo [UT], Etats-Unis) sur lequel l'expérience a été programmée. Après avoir accepté le formulaire de consentement, affiché soit en anglais soit en français, les participant-e-s ont réalisé la tâche principale. Dans celle-ci, les participant-e-s remplissaient le blanc dans 60 phrases données (40 phrases expérimentales + 20 phrases de distraction) en choisissant le connecteur approprié parmi une liste de quatre options. Les phrases s'affichaient seules et dans un ordre aléatoire. Les options proposées étaient toujours les quatre connecteurs *mais*, *alors*, *parce que* et *puisque*, également affichés dans un ordre aléatoire. Aucune contrainte par le temps a été imposée.

Parmi les phrases expérimentales, 20 exigeaient le connecteur *parce que* (et excluaient *puisque*), et 20 exigeaient *puisque* (et excluaient l'utilisation de *parce que*). Nous avons réalisé cela en utilisant des contraintes spécifiques de chaque connecteur, deux contraintes par connecteur.

Les contraintes dans les phrases qui exigeaient seulement *parce que* étaient soit syntaxiques, soit pragmatiques. Cela nous a permis de déterminer à quelles contraintes les participant-e-s étaient plus sensibles, c'est-à-dire si les contraintes syntaxiques étaient plus saillantes que les contraintes pragmatiques.

La contrainte syntaxique de *parce que* consistait à l'utiliser dans une structure clivée, soit avec *c'est*, voir (14), soit avec des adverbes tels que *probablement*, voir (15). Comme discuté auparavant, ces deux structures sont impossibles pour *puisque* (Groupe λ-l, 1975; Zufferey, 2012).

14) Julie a beaucoup d'amis. Je pense que c'est _____ elle est toujours sympa avec tout le monde.
15) Jean ne s'est jamais perdu dans la ville, probablement _____ il y vit depuis longtemps.

La contrainte pragmatique consistait dans le fait que la cause indiquée par *parce que* figurait dans une réponse à une question, voir (16).

16) Tu as reparlé avec elle ? - Je n'ai pas pu _____ elle est partie en vacances.

Ici, l'idée est que lorsqu'une personne pose une question, elle ne connaît généralement pas la réponse. Comme la personne vient d'être informée seulement dans la réponse, la cause dans (16) est un nouvel élément. Par conséquent, seul *parce que* convient comme connecteur dans ce cas.[36]

Les contraintes pour les phrases visant *puisque* étaient toutes les deux de type pragmatique. Cela nous a permis d'examiner si une contrainte pragmatique était plus saillante que l'autre. Alors que les deux se basaient sur le fait que *puisque* s'utilise, contrairement à *parce que*, pour indiquer une information déjà connue (Zufferey, 2012), les deux types de contraintes se distinguaient l'une de l'autre par la raison pour laquelle l'information est connue.

Dans la première contrainte, la *citation*, l'information est répétée par la personne l'interlocutrice, voir (17).

17) Je savais que tu allais réparer la table _____ tu me l'avais annoncé hier.

Dans la deuxième contrainte, l'*évidence*, l'information est connue parce qu'il s'agit d'une information généralement partagée, voir (18).

18) On peut commencer la réunion _____ tout le monde est là.

Pour résumer, nous avons testé 20 phrases ciblant *parce que* (dont 10 contenaient une contrainte syntaxique et 10 une contrainte pragmatique) et 20 phrases ciblant *puisque*.

Outre les phrases expérimentales, nous avons également présenté 20 phrases de remplissage. 10 de ces phrases visait le connecteur *alors*, voir (19), 10 le connecteur *mais*, voir (20).

19) J'ai lu un livre très intéressant _____ je voulais vous en parler.
20) J'aime bien Jean _____ parfois il m'énerve.

Non seulement ces phrases servaient à détourner l'attention des participant-e-s des connecteurs *parce que* et *puisque*, elles servaient également comme base de comparaison : contrairement aux phrases visant *puisque* ou *parce que*, les phrases

[36] Il est vrai pourtant que les participant-e-s auraient pu utiliser *puisque* dans (16), donnant le sens suivant à la réponse : « Je n'ai pas pu, *je te rappelle :* elle est partie en vacances ». Les résultats obtenus n'indiquent toutefois pas une telle interprétation.

de remplissage ne contenaient pas de contraintes spécifiques. Pour ces phrases, il s'agissait simplement d'identifier la bonne relation de cohérence et de lui associer le connecteur correspondant. Nous avons ainsi pu déterminer si les participant-e-s étaient capables d'identifier des relations de cohérence.

Les relations de cohérence dans les phrases de remplissage se distinguaient par ailleurs par leur complexité (Sanders et al., 2012, Sanders et al., 2018). La relation indiquée par *alors* (conséquence) a un ordre d'implication différent que celle indiquée par *puisque* / *parce que* (cause). La relation indiquée par *mais* (concession) a une polarité négative, contrairement à celle indiquée par *puisque* / *parce que* (conséquence). Comme la relation indiquée par *mais* est plus complexe, nous avons pu obtenir un aperçu plus nuancé de la maitrise des connecteurs des participant-e-s.

Après avoir réalisé la tâche principale, les participant-e-s passaient à la tâche de *Lextale* (Brysbaert, 2013; voir chapitre 5) et à la *Tâche de Grammaire* (Zufferey & Gygax, 2020b; voir chapitre 5). Ensuite, le groupe non-natif répondait à des questions visant - entre autres - leur niveau socio-économique, ainsi que des questions liées à leur apprentissage du français. Ces dernières concernaient leur motivation et anxiété de parler le français, leurs histoires d'apprentissage personnelles et la durée et fréquence des contacts avec des francophones.

Finalement, les participant-e-s répondaient à la question ouverte et méta-linguistique suivante : « Quelle est la différence entre *puisque* et *parce que* ? ». Les réponses nous permettaient d'analyser si les personnes non-natives étaient au courant des fonctions différentes des deux connecteurs. Les participant-e-s pouvaient soit indiquer « Je ne sais pas », soit écrire une réponse sans limite d'espace.

9.2.3 Les résultats et l'analyse

Premièrement, nous allons analyser si les personnes ont donné des réponses correctes plus fréquemment pour un des deux connecteurs, indépendamment des contraintes utilisées. Les résultats descriptifs sont reportés dans le Tableau 9.2.

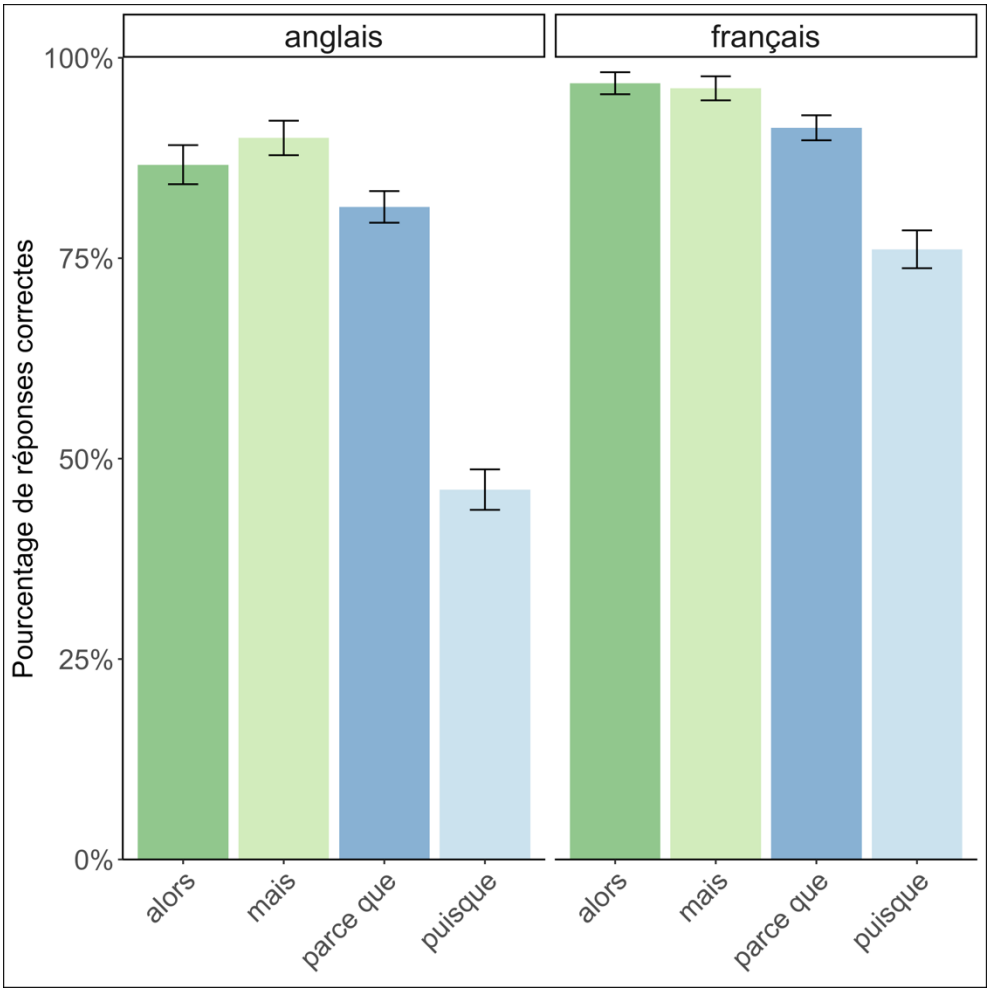
Tableau 9.2. Résultats descriptifs pour parce que et puisque

Groupe natif	*Pourcentage de réponses correctes*	*Écart type*	*SE*	*CI de 95%*
parce que	91%	.28	.01	[89.71; 92.83]
puisque	76%	.43	.01	[73.74; 78.47]
Groupe non-natif				
parce que	81 %	.39	.01	[79.44; 83.73]
puisque	46 %	.50	.01	[43.20; 48.75]

Nous observons que les personnes natives, de même que les personnes non-natives, ont donné des réponses plus correctes pour les phrases ciblant *parce que* que pour les phrases ciblant *puisque*.

La question se pose de savoir si les scores pour *puisque* sont particulièrement bas, ou les scores pour *parce que* particulièrement hauts. Pour répondre à cette question, il convient de comparer ces scores avec les scores pour les connecteurs testés dans les phrases de remplissage (*alors* et *mais*) : pour le groupe non-natif, les phrases pour *alors* ont suscité 87% de réponses correctes (écart type= .34), les phrases de distraction visant *mais* 90% de réponses correctes (écart type = .30). Ces comparaisons sont également représentées par la Figure 9.1.

Figure 9.1. Scores pour les toutes les phrases. Les intervalles des confiances (95%) sont indiqués par des barres d'erreur.



Nous voyons donc que *puisque* a suscité des scores bas comparés aux trois autres connecteurs, notamment pour le groupe non-natif.

Avant de se concentrer plus sur ce connecteur, nous allons maintenant analyser si les réponses pour *parce que* dépendaient de la contrainte exigée par ce connecteur.

Comparaisons des contraintes pour parce que

Nous allons analyser si les contraintes utilisées pour *parce que* (contrainte pragmatique vs. contrainte syntaxique) ont suscité des scores différents. Pour ce faire, nous analyserons les données pour ce connecteur uniquement. Nous avons fait cela en réalisant des modèles logistiques généralisés à l'aide de *R* (R Core Team 2020).

Comme dans les chapitres précédents, les modèles ont été montés selon la procédure de Baayen (2008). Pour calculer l'amélioration du modèle lorsqu'un effet fixe a été ajouté, nous avons utilisé la fonction *anova()* de *base-R* (R Core Team 2020). Pour exécuter les modèles, nous avons utilisé la fonction *glmer()* du package *lm4* (Bates et al., 2015). Nous avons obtenu les valeurs de *p* ainsi que les tailles d'effet estimées en utilisant la fonction *summary()* du package *car* (Fox & Weisberg 2019). Tous les modèles contenaient toujours les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires.

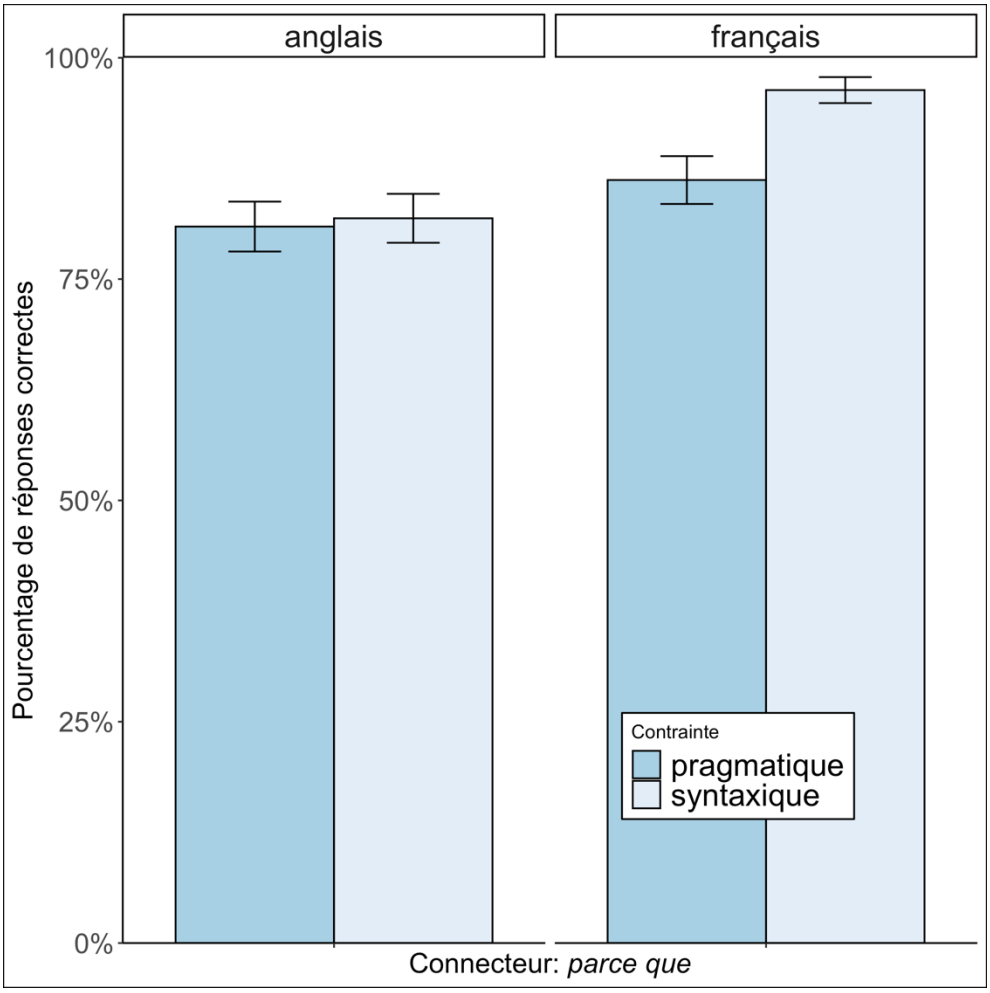
L'ajout de la *L1 des participant-e-s* (français ou anglais) comme effet fixe a amélioré le modèle de manière significative ($\chi^2(1) = 26.62$, $p = 2.48^{e-07}$). L'ajout de l'interaction avec le *Type de contrainte* (contrainte syntaxique vs. contrainte pragmatique) a amélioré le modèle davantage ($\chi^2(2) = 31.82$, $p = 1.23^{e-07}$). L'output du modèle final, contenant l'interaction entre la *L1 des Participant-e-s* et le *Type de Contrainte* ainsi que les *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires, est reporté dans le Tableau 9.3.

Tableau 9.3. Résultats du modèle final

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	1.72	.26	6.59	4.48^{e-11}
L1 (*français*)	.42	.20	2.11	.04
Contrainte (*syntaxique*)	.10	.35	.28	.78
L1 (*français*) : Contrainte (*syntaxique*)	1.48	.28	5.33	9.67^{e-08}

Nous pouvons faire deux constatations : d'abord les non-natifs ont généralement obtenu de moins bons scores que les personnes natives pour *parce que*. Ensuite, les personnes non-natives n'ont pas obtenu un score différent en fonction du type de contrainte, alors que le groupe natif a obtenu des scores moins hauts pour les contraints pragmatiques. Ces observations sont également représentées visuellement par la Figure 9.2.

Figure 9.2. Résultats pour les contraintes pragmatiques et syntaxiques de parce que (IC 95%)



Comparaison des contraintes pour puisque

Maintenant, nous allons analyser si les participant-e-s donnaient plus ou moins de réponses correctes selon la contrainte exigée par *puisque*. Pour ce faire, nous analysons les données pour ce connecteur uniquement, en procédant comme dans l'analyse précédente. Pour rappel, les deux types de contraintes testés étaient de nature pragmatique (*citation* vs. *évidence*).

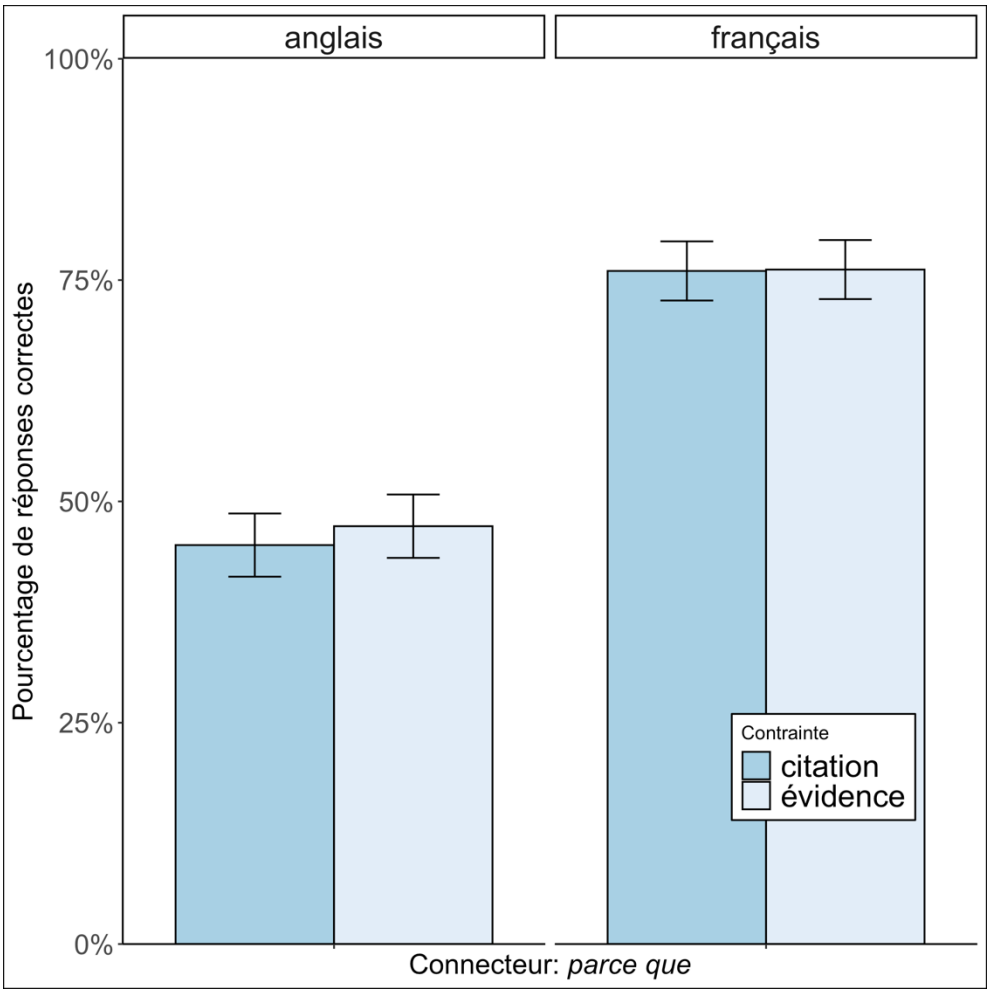
Lors de l'ajout de la *L1 des participant-e-s* comme effet fixe, le modèle s'est amélioré de manière significative ($\chi^2(1) = 63.63$, $p = 1.50^{e-15}$). Cependant, en ajoutant son interaction avec le *Type de Contrainte,* le modèle ne s'est pas amélioré davantage ($\chi^2(2) = .12$, $p = .94$). L'output du modèle final, contenant seulement la *L1 des participant-e-s* comme effet fixe (ainsi que les *Participant-e-s* et les *Phrases expérimentales* comme effets aléatoires) est reporté dans le Tableau 9.4.

Tableau 9.4. Résultats du modèle final pour *puisque*,

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	-0.20	.18	-1.10	.27
Langue (*français*)	1.61	.18	8.80	$< 2^{e-16}$

Comme nous voyons également dans la visualisation des résultats à la Figure 9.3, le groupe natif a obtenu de meilleurs scores que le groupe non-natif pour *puisque*, indépendamment du type de contrainte.

Figure 9.3. Résultats pour les contraintes pragmatiques (évidence ou citation) de puisque



Facteur prédictifs : la taille du vocabulaire et les compétences en grammaires

Nous allons maintenant examiner si les mesures supplémentaires[37] prédisent les compétences avec les connecteurs.

Nous observons pour les deux groupes linguistiques des corrélations modérées entre les scores de *Lextale* et la *Tâche de Grammaire* (groupe natif : r = .44, t = 30.38, ddl = 38, $p < 2.2^{e-16}$; groupe non-natif : r = .40, t = 28.79, ddl = 4498, $p < 2.2^{e-16}$, testée avec la fonction *cor.test()* du package stats de *R*, R Core Team, 2020). Afin de respecter l'hypothèse de l'absence de corrélation entre les effets fixes pour les modèles linéaires (Schreiber-Gregory, 2018), nous allons faire de modèles séparés pour ces deux mesures. De plus, nous allons faire de modèles séparés également en fonction de la langue (groupe natif et groupe non-natif) car on peut faire l'hypothèse d'une corrélation entre le fait d'avoir un score haut dans les tâches supplémentaires et le fait d'être natif.

Compte tenu de différentes moyennes des scores en fonction du connecteur (voir Tableau 9.2) nous allons ajouter toujours le *Type de connecteur* (i.e., *puisque* ou *parce que*) comme premier effet fixe avant d'ajouter l'interaction avec la mesure en question. Cela nous permet d'observer plus en détail si les scores de nos mesures supplémentaires prédisent un meilleur score en fonction du connecteur testé.

Les scores de Lextale

Pour le groupe natif, l'ajout du *Type de connecteur* comme effet fixe a amélioré le modèle ($\chi^2(1) = 16.41$, $p = 5.10^{e-05}$). Lors de l'ajout de l'interaction avec les *Scores de Lextale,* le modèle ne s'est pas amélioré davantage ($\chi^2(2) = 4.02$, $p = .13$).

Pour le groupe non-natif, l'aout du *Type de connecteur* comme effet fixe a amélioré le modèle ($\chi^2(1) = 47.28$, $p = 6.16^{e-12}$). Cette fois, l'ajout de l'interaction avec les *Scores de Lextale* a amélioré le modèle d'avantage ($\chi^2(2) = 9.25$, $p < .01$). L'output est reporté dans 9.5.

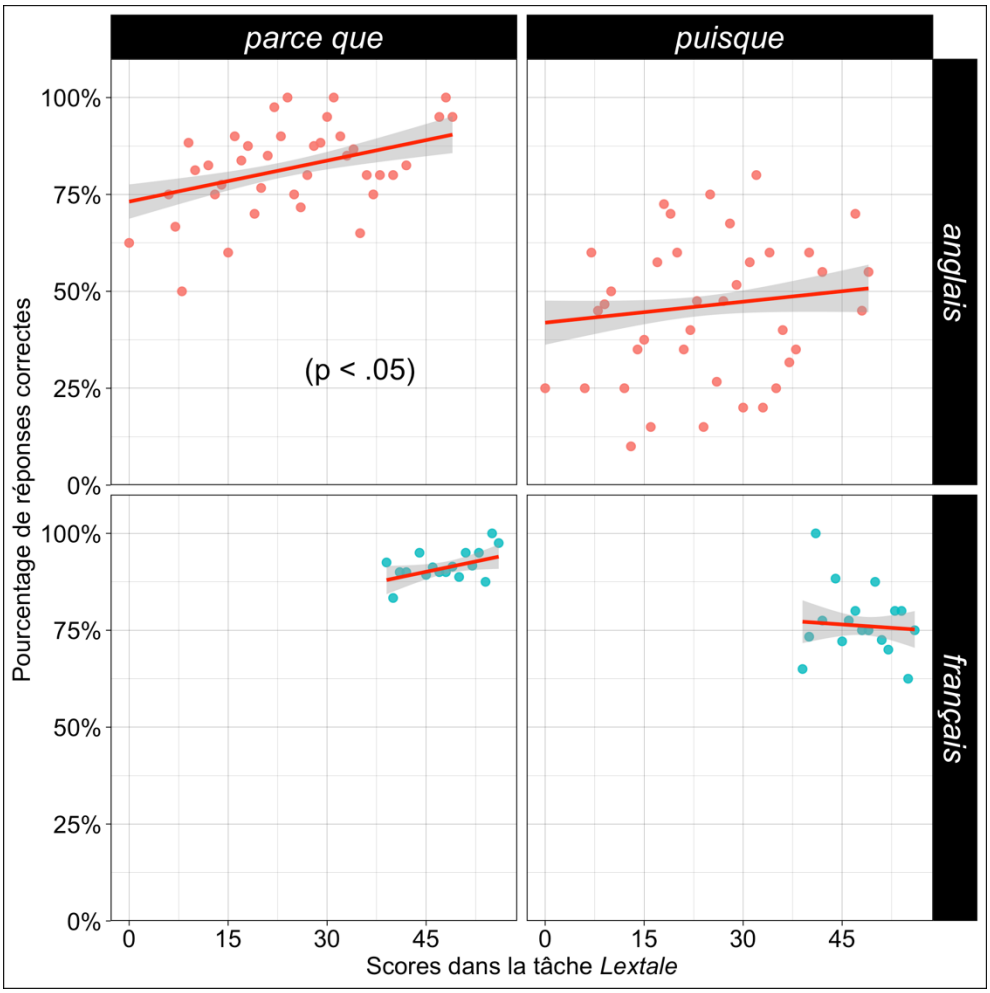
[37] Des analyses supplémentaires ont montré que les réponses aux questions démographiques/socio-économiques ainsi qu'aux questions sur l'apprentissage du français n'étaient pas prédictifs pour les scores dans la tâche principale.

Tableau 9.5. Mesures supplémentaires (Lextale): résultats du modèle final (personnes non-natives).

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	1.10	.25	4.41	1.03^{e-05}
Connecteur (*puisque*)	-1.47	.26	-5.56	2.64^{e-08}
Lextale	.03	.01	3.00	< .005
Connecteur (*puisque*) : Lextale	-0.02	.01	-2.22	< .05

Nous observons donc pour le groupe non-natif que les scores pour le connecteur *puisque* ne sont pas prédits par la taille du vocabulaire, telle que mesurée par *Lextale*. Ces effets sont également visualisés dans la Figure 9.4.

Figure 9.4. Scores obtenus dans Lextale et les pourcentages de réponses correctes pour les phrases ciblant parce que et puisque. Effets significativement prédictifs indiqués avec (p<.05).



Des analyses supplémentaires ont montré, en revanche, que les scores de *Lextale* prédisaient le pourcentage de réponses correctes des phrases de remplissage.

Tâche de grammaire

Pour le groupe natif, de manière peu surprenante, l'ajout du *Type de connecteur* comme effet fixe a amélioré le modèle de manière significative ($\chi^2(1) = 16.41$, $p = 5.10^{e-05}$). L'ajout de l'interaction avec les scores de la *Tâche de grammaire* n'a pas amélioré le modèle davantage ($\chi^2(2) = .93$, $p = .63$)

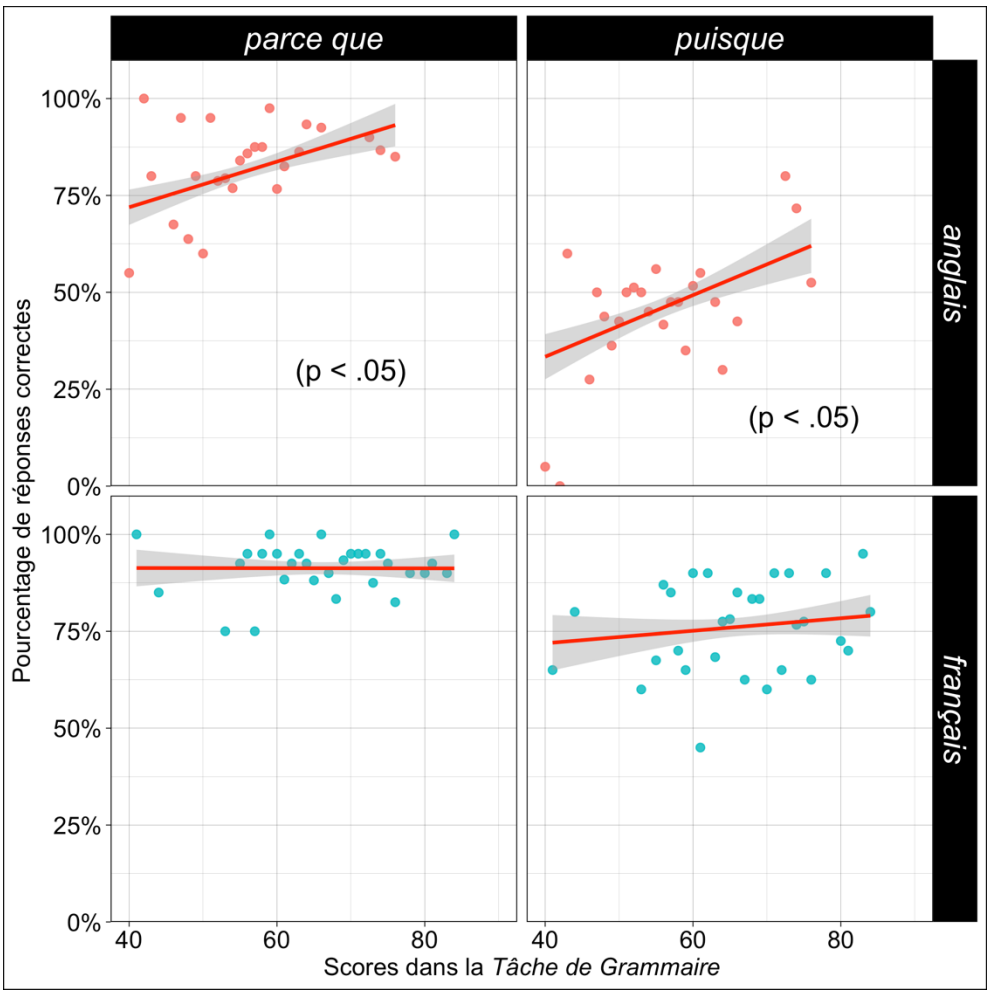
Pour le groupe non-natif, l'ajout du *Type de connecteur* a amélioré le modèle de manière significative ($\chi^2(1) = 47.28$, $p = 6.16^{e-12}$). De plus, l'interaction avec les scores de la *Tâche de grammaire* a également amélioré le modèle de manière significative ($\chi^2(2) = 14.60$, $p < .01$). L'output du modèle final - contenant l'interaction entre la *Tâche de grammaire* et le *Type de Connecteur* ainsi que les *Participant-e-s* et *Phrases expérimentales* comme effets aléatoires - est reporté dans le Tableau 9.6.

Tableau 9.6. Mesures supplémentaires (Tâche de grammaire) : résultats du modèle final (personnes non-natives).

	β	*ET*	*z*	*Pr(>|z|)*
(Intercept)	-0.79	.72	-1.10	.27
Connecteur (*puisque*)	-1.45	.71	-2.04	.04
Tâche de grammaire	.04	.01	3.47	< .01
Connecteur (*puisque*) : Tâche de grammaire	-0.01	.01	-0.60	.55

Nous observons donc que les scores de la *Tâche de grammaire* ont prédit un meilleur score pour les deux connecteurs, cependant seulement pour le groupe non-natif. Ces effets sont également visualisés dans la Figure 9.5.

Figure 9.5. Scores obtenus dans la *Tâche de Grammaire* et les pourcentages de réponses correctes pour les phrases ciblant *parce que* et *puisque*. Effets significativement prédictifs indiqués avec (p <. 05).



De nouveau, des analyses supplémentaires ont démontré que les scores de la *Tâche de Grammaire* étaient également prédictifs de meilleurs scores pour les personnes non-natives pour les phrases de remplissage.

9.2.4 Analyse exploratoire des connaissances métalinguistiques

Les résultats montrent clairement que *puisque* n'était pas bien maitrisé par les personnes non-natives, quel que soit la contrainte testée. Cela pose la question de savoir quels facteurs permettent une meilleure maitrise de ce connecteur. Pour répondre à cette question, nous allons faire une *analyse exploratoire* des données (selon Tukey, 1977; voir aussi Morgenthaler, 2009).

Plus précisément, nous allons évaluer si les connaissances métalinguistiques, mesurée par les réponses à la question « Quelle est la différence

entre *puisque* et *parce que* ? », permettent de prédire une meilleure compétence pour ce connecteur. Pour ce faire, nous (i.e., trois évaluateur-e-s) avons classifié les réponses à cette question métalinguistique ouverte dans quatre catégories. Voici le quatre groupes établis :

1. *Je ne sais pas* / *Pas de réponse*
   Ce premier groupe comprenait les personnes qui avaient cliqué sur *Je ne sais pas* et qui n'avaient donc pas donné de réponse.

2. *Réponse incorrecte*
   Ce groupe était constitué de personnes qui ont donné une explication incorrecte.
   Voici un exemple :
   > « They both mean roughly the same but puisque is more formal/literary »
   > 'Les deux signifient à peu près la même chose, mais puisque est plus formel/littéraire.'

3. *Réponse simplifiée*
   Ce groupe était constitué de réponses qui ont seulement donné une traduction des mots anglais sans aborder les fonctions entre les deux connecteurs en détail. Nous avons compté cela comme un groupe entier, car certaines caractéristiques du connecteur anglais *since* correspondent effectivement à celles de *puisque* (Zufferey & Cartoni, 2012).
   Voici un exemple :
   > *« Puisque* means *since* and *parce que* means *because*. »
   > '*Puisque* signifie ou *since* et *parce que* signifie *because'*

4. *Réponse correcte*
   Ce groupe était composé de personnes qui étaient capables de nommer et de décrire les différentes fonctions de *puisque* et *parce que*.
   Voici un exemple :

   > « Puisque signifie que la personne avec qui vous parle sait la raison avant que vous l'explique, et en revanche parce que signifie que la personne ne le sait pas avant que vous l'explique »

Comme nous voyons dans le Tableau 9.7, le calcul du Kappa de Fleiss a montré un accord inter-juges presque parfait (selon Landis & Koch, 1977) pour la classification de tous les groupes.

Tableau 9.7. Les Kappas de Fleiss pour les accords inter-juges

Classification du groupe	κ	*z*	*p*
Je ne sais pas	1	15	< .001
Réponse incorrecte donnée	.95	14.29	< .001
Réponse simplifiée	.94	14.13	< .001
Réponse correcte	.97	14.49	< .001

Pour ces quatre groupes établis, les scores moyens pour les phrases qui exigeaient *puisque* sont reportés dans le Tableau 9.8.

Tableau 9.8. Les scores moyens pour les phrases ciblant puisque selon les quatre groupes établis.

	Groupe	*n*	*Score moyen pour les phrases ciblant puisque*	*Écart type*	*SE*	*Ci 95%*
1	Je ne sais pas	18	37 %	.48	.03	[.32; .42]
2	Réponse incorrecte	18	39 %	.49	.03	[.34; .44]
3	Réponse simplifiée	27	49 %	.50	.02	[.44; .53]
4	Réponse correcte	12	66 %	.48	.03	[.60; .72]

Nous voyons que les personnes qui ont donné des réponses que nous avons classifié comme tout à fait correcte (*Groupe 4*) ont obtenu de meilleurs scores pour *puisque* que les personnes qui ont donné des réponses incorrectes à la question métalinguistique (*Groupes 1* et *Groupe 2*). Il convient de noter pourtant que les scores de ce groupe (66% de bonnes réponses) ne correspondent tout de même pas à ceux du groupe natif (76% de bonnes réponses).

9.2.5 Discussion

Dans cette expérience, nous avons évalué la connaissance explicite de participant-e-s natifs et non-natifs pour deux connecteurs causaux du français, *parce que* et *puisque*. L'objectif était de déterminer dans quelle mesure les personnes non-natives sont aidés dans le choix d'un connecteur approprié par des contraintes syntaxiques et pragmatiques. Dans la tâche principale, les participant-e-s devaient choisir le connecteur approprié pour des phrases à trous. Les phrases qui exigeaient *parce* que (et excluaient *puisque*) étaient présentées selon deux contraintes, soit pragmatique soit syntaxique. Les phrases qui devaient être marquées avec *puisque* contenaient deux contraintes pragmatiques : l'une était la répétition de ce qui avait été dit (*citation*), l'autre contenait un fait généralement connu (*évidence*).

Premièrement, il convient de noter que les personnes non-natives étaient capable d'identifier correctement la relation de cohérence (de type conséquence-cause) dans 92% des cas, c'est-à-dire qu'ils ont cliqué plus de 9 fois sur 10 sur soit *puisque*, soit *parce que* lorsqu'une phrase expérimentale a été affichée. Cela

indique que les participant-e-s comprenaient les phrases et étaient capables de faire correctement la tâche.

Les résultats indiquent que bien que *parce que* ait été bien maitrisé, le groupe non-natif ne profitait pas dans la même mesure que le groupe natif de la contrainte syntaxique pour ce connecteur. En effet, le groupe natif choisissait dans quasiment tous les cas (96% des bonnes réponses) *parce que* lorsque la contrainte était syntaxique. Lorsque la contrainte était pragmatique, ce groupe n'identifiait *parce que* comme connecteur correct que dans 86% des cas. Nous n'avons pas observé une telle différence pour le groupe non-natif : ce groupe choisissait dans 81% des cas *parce que* pour cette contrainte, donc aussi souvent que pour la contrainte pragmatique. Nous concluons donc que les personnes non-natives ne bénéficiaient pas dans la même mesure que les personnes natives des indices syntaxiques.

Quant à *puisque*, les résultats montrent que les personnes non-natives le choisissaient beaucoup plus rarement que *parce que*. Dans moins que la moitié des phrases qui exigeaient ce connecteur, *puisque* a été choisi. Si on considère que le choix se faisait notamment entre deux connecteurs qui indiquaient la même relation de cohérence (i.e., *puisque* et *parce que*) le score à moins de 50% pour *puisque* peut être considéré même en dessous du niveau de chance. De plus, comme les personnes non-natives étaient plutôt compétentes sur le plan linguistique (cf. les scores de *Lextale* et de la *Tâche de Grammaire*) et que les connecteurs ont été correctement identifiés dans les phrases de remplissage dans environ neuf cas sur dix, on peut supposer qu'elles étaient capables de reconnaitre la bonne relation de cohérence. Cela indique que ni le contenu linguistique, ni l'identification de la relation, ni la relation de cohérence en soi posait de problèmes aux personnes non-natives. Les scores extrêmement bas pour *puisque* indiquent plutôt que le groupe non-natif ne maitrisait pas ce connecteur. Ainsi, il n'est pas surprenant que nous n'ayons pas observé des scores différents en fonction des deux contraintes testées.

Bien que *puisque* n'ait pas été bien maitrisé, il semble y avoir une interaction avec la compétence linguistique de sorte qu'une meilleure compétence de grammaire (telle que mesurée par notre *Tâche de Grammaire)* prédisait des meilleurs scores pour ce connecteur. En fait, la compétence grammaticale du groupe non-natif prédisait des meilleurs scores pour tous les connecteurs testés, y compris les connecteurs des phrases de remplissage. Cela soutient donc nos

résultats au chapitre 5 où nous avons observé qu'une meilleure compétence grammaticale prédisait également une meilleure maitrise des connecteurs.

La taille du vocabulaire, telle que mesurée par *Lextale*, était également prédictive de la maitrise des connecteurs *mais*, *alors* et *parce que,* ce qui souligne également nos observations des chapitres 5, 7 et 8, à savoir que la taille du vocabulaire prédit les connaissances explicites des personnes non-natives sur les connecteurs. Cependant, les scores de *puisque* n'ont pas pu être prédits par cette mesure. Cela pourrait indiquer que les connaissances grammaticales sont mieux corrélées aux connaissances pragmatiques que les connaissances lexicales. Cependant, davantage de recherches sont nécessaires pour vérifier cette hypothèse.

Nos résultats permettent également d'avancer quelques explications permettant de comprendre pourquoi *puisque* semble être aussi difficile pour les personnes non-natives. Dans de nombreuses réponses des participant-e-s, nous avons observé une mauvaise mise-en-correspondance de *since* et *puisque* qui menait à une confusion des fonctions du connecteur français. Comme illustré dans les exemples (21) – (24), plusieurs personnes ont indiqué que *puisque* implique un élément de sens temporel.

21) « Puisque is "since", so there's a time element. For example, if I was saying that I hadn't spoken to someone since they got back from a holiday, I would use "puisque". Parce que means because so there's no time involved in that, it's just an explanation »
'Puisque est "since", il y a donc un élément temporel. Par exemple, si je disais que je n'ai pas parlé à quelqu'un depuis son retour de vacances, j'utiliserais "puisque". Parce que signifie "because", il n'y a donc pas d'élément temporel, c'est juste une explication'

22) « Puisque a une aspect de temps "simultaneous", mais parce que est avant et apres. »

23) « ' Puisque ' peut être utilisé de deux manières différentes dans une phrase, c'est-à-dire qu'il parle soit de « temps », soit il donne la « raison de quelque chose ». Au contraire, « Parce que » fait référence à 'par cause de' »

24) « 'Puisque' can be used in replacement of 'parce que' as it can mean not only 'because' but also can start a sentence as it can mean 'since'. Whereas 'parce que' only translates to 'because'. »

Cela peut être expliqué par le fait que le connecteur anglais *since* remplit deux fonctions différentes : comme marque de temporalité où de causalité (Zamel, 1983). *Puisque*, par contre, ne sert que pour indiquer de la causalité. En transférant toutes les fonctions de *since* au connecteur français, les personnes non-natives se sont fait piéger par leur L1 et ont été amenées à une confusion entre les deux connecteurs. Cette confusion entre les langues est également révélée dans d'autres commentaires, tels que (25).

25) *« Since is conditional. Like the first part of the sentence is the cause of the second part after conditional. Because is just an explanation word. »*
'Since est conditionnel. Comme la première partie de la phrase est la cause de la deuxième partie après le conditionnel. because est juste un mot d'explication.'

Il convient de noter ici que la question se référait à la différence entre *puisque* et *parce que* et non à la différence entre *since* et *because*.

9.3 Conclusion

Dans ce chapitre, nous avons vu que les connecteurs indiquant les mêmes relations de cohérence ne sont pas interchangeables, car il y existe des contraintes pragmatiques et syntaxiques qui imposent (ou excluent) certains connecteurs. Cependant, les différents types de contraintes, à savoir pragmatiques ou syntaxiques, n'ont pas semblé influencer le jugement des personnes non-natives dans notre expérience. Nous avons montré par exemple que les personnes non-natives, malgré une compétence linguistique plutôt élevée, n'étaient pas plus sensibles aux contraintes syntaxiques de *parce que* (contrairement au groupe natif). Bien que nous n'ayons pas constaté de différences quant au type de contrainte pour les personnes non-natives, l'étude nous a permis tout de même de constater plusieurs choses.

Premièrement, nous avons observé que *puisque* est extrêmement difficile à maitriser malgré le fait qu'il indique une relation de cohérence simple (Sanders et al., 1992). Nous attribuons cela au fait que ce connecteur est pragmatiquement complexe (Zufferey, 2012, 2014). Comme nous avons observé que le groupe non-natif évitait visiblement d'utiliser ce connecteur pour remplir les blancs des phrases, notre étude confirme une constatation faite dans de nombreuses études de corpus :

les personnes non-natives tendent à utiliser seulement les connecteurs avec lesquels ils se sentent à l'aise et évitent les autres (p.ex., Leedham & Cai, 2013; Chen, 2014; Casteele & Collewart, 2013). Ainsi, les résultats obtenus dans cette expérience soutiennent également les résultats obtenus au chapitre 5, où nous avons constaté que les apprenant-e-s évitaient de choisir *puisque*. Notre étude indique en outre que la préférence pour un seul connecteur se manifeste même dans des cas où celui-ci est incorrect.

Bien que *puisque* semble être extrêmement difficile à maitriser en L2, notre étude met également en évidence certains facteurs qui facilitent l'utilisation de *puisque* en langue seconde. Tout d'abord, nous avons constaté qu'une meilleure compétence grammaticale peut prédire une meilleure maitrise de ce connecteur. Cela confirme les résultats des chapitres 5, 7 et 8, à savoir que la compétence linguistique est corrélée avec la connaissance explicite des connecteurs.

Finalement, nos résultats indiquent que la présence d'un connecteur de la L1 partageant partiellement les mêmes fonctions que *puisque* (i.e., *since*) n'est pas automatiquement bénéfique pour sa maitrise. Ainsi, nos observations complètent celles faites au chapitre 8 sur le rôle du transfert. Plus précisément, dans notre étude, les personnes qui faisaient simplement une superposition des fonctions du connecteur de leur L1 *since* au connecteur L2 *puisque* (groupe 3), avaient de la peine à identifier les phrases qui exigeaient l'utilisation de *puisque*. Seuls les participant-e-s qui avaient des connaissances métalinguistiques sur les fonctions de *puisque* (groupe 4) obtenaient des scores plus hauts pour les phrases ciblant ce connecteur. Mais même ce groupe, malgré ses bonnes connaissances métalinguistiques, a montré une certaine incertitude par rapport à *puisque* : contrairement à ses réponses pour les phrases ciblant les connecteurs *parce que* (90% de bonnes réponses), *alors* (98% de bonnes réponses) et *mais* (99% de bonnes réponses), ce groupe n'a pu identifier le connecteur *puisque* que dans 66% des cas. Il convient toutefois de préciser que ces résultats de l'analyse exploratoire doivent être considérés à la lumière du nombre relativement faible de locuteurs-trice-s. Ainsi, des recherches supplémentaires sont nécessaires pour explorer dans quelle mesure les compétences pragmatiques peuvent être apprises par le biais de connaissances métalinguistiques.

En résumé, nous voyons que les personnes non-natives identifient plus facilement les contraintes pour *parce que* (qu'elles soient syntaxiques ou

pragmatiques) tandis que les contraintes pragmatiques pour *puisque* ne sont pas véritablement identifiées. Quant aux personnes natives, nos résultats affinent ceux de Zufferey (2012). Nous avons observé par exemple que la contrainte pragmatique d'une *évidence* pour *puisque* était tout aussi bien identifiée par le groupe natif qu'une contrainte qui se base sur une *citation*. Dans Zufferey (2012), cette contrainte n'a pas donné une préférence claire pour ce connecteur, ce qui est probablement dû au fait qu'une seule phrase contenant cette contrainte a été testée.

Comme toute expérience, notre évaluation de la sensibilité aux contraintes a ses limites. Par exemple, on ne peut pas éviter de discuter le choix des phrases expérimentales utilisées. Comme la dimension pragmatique implique forcément au moins deux locuteur-trice-s, la question se pose de savoir si les phrases utilisées correspondent vraiment à des exemples authentiques. Il se peut en outre que les contraintes dans certaines phrases étaient beaucoup plus saillantes et concises que dans d'autres. De plus, nous ne savons pas avec certitude si les phrases contenant une dimension pragmatique sont identifiées telles quelles, c'est-à-dire si les personnes ont réussi à identifier la dimension pragmatique vu que les phrases ont été présentées à l'écrit et sans contexte. Par exemple, en ce qui concerne notre condition *évidence*, il convient de noter que, selon Sperber et Wilson (1986), une information ne peut jamais être considérée comme *mutuellement connue* (Sperber & Wislon, 1986 : 19). La question se pose donc de savoir dans quelle mesure les phrases de cette condition était effectivement évidentes pour tous les locuteur-trice-s. Nous reviendrons encore plus en détail au chapitre 10.3 sur la question de savoir dans quelle mesure les phrases artificiellement construites dans une approche expérimentale représentent un biais pour la généralisation des résultats.

En conclusion, les observations faites dans ce chapitre rappellent et complémentent celles faites dans d'autres chapitres, par exemple concernant le rôle du transfert négatif (chapitre 8) ou celui de la compétence linguistique (chapitres 5, 7 et 8). Nous allons maintenant conclure cette thèse de doctorat en synthétisant plus en détail ces observations ainsi que les tendances concordantes que nous avons observées au travers des chapitres.

10 Conclusion

Dans la présente thèse de doctorat, nous avons examiné la maitrise des connecteurs du discours en français langue seconde. Plus précisément, nous nous sommes intéressés à la question de savoir si certains connecteurs sont plus difficiles à maitriser en L2 que d'autres et en avons évalué les causes. Nous avons également examiné les compétences individuelles prédisant une meilleure maitrise des connecteurs en L2 ainsi investigué la question de savoir à quel point la maitrise d'un connecteur dépend du type de connaissance auxquelles une personne peut accéder (i.e., connaissances explicites vs. implicites).

Dans ce dernier chapitre, nous synthétiserons et discuterons les observations que nous avons faites au travers de nos études empiriques et développerons quelques nouveaux axes de recherches pour des travaux futurs.

10.1 Les principaux résultats

La présente thèse de doctorat avait pour but d'examiner si la maitrise des connecteurs en L2 dépend du type de connecteur en question, du type des connaissances auxquelles l'apprenant-e peut accéder ainsi que des compétences individuelles de l'apprenant-e. Pour faire cela, nous avons mené dix expériences que nous avons présentés dans les chapitres 4 à 9. Pour faire une première synthèse des résultats obtenus dans ces chapitres, nous pouvons faire les quatre constatations suivantes :

- Les connecteurs possèdent des degrés de complexité différents pour les apprenant-e-s.
- Les raisons pour lesquelles certains connecteurs sont difficiles en L2 sont multifactorielles et émergent des interactions entre ces facteurs.
- La maitrise des connecteurs en L2 peut être prédite par l'exposition à l'écrit et par la compétence linguistique de l'apprenant-e.
- La maitrise dépend également du type de connaissances auquel l'apprenant-e peut accéder.

Nous allons maintenant développer chacune de ces constatations.

10.1.1 « Les connecteurs ont des degrés de complexité différents pour les apprenant-e-s »

La première question de recherche que nous avons formulée était de savoir si les connecteurs diffèrent en termes de complexité pour les apprenant-e-s d'une L2. Nous avons observé à plusieurs reprises que cela semble effectivement être le cas.

D'abord, nous avons constaté au chapitre 5 que les gymnasien-ne-s maitrisent certains connecteurs (comme *c'est pourquoi*) considérablement mieux que d'autres (comme *cependant*). Cette observation ne peut être due qu'aux différents degrés de difficulté des connecteurs, car nous avons testé les mêmes participant-e-s pour toutes les phrases et la validité de celles-ci avait été contrôlée dans un pré-test. Il est également peu probable que les différences observées s'expliquent par le fait que les connecteurs testés indiquent des relations différentes. Non seulement nous avons observé des différences entre les connecteurs indiquant la même relation de cohérence (comme *ainsi* et *c'est pourquoi*), mais les participant-e-s du chapitre 9 ont également montré des différences dans leur maitrise d'un connecteur (i.e., *puisque*) alors qu'un autre qui indiquait la même relation de cohérence (i.e., *parce que*) était nettement mieux maitrisé.

Nous pouvons conclure aussi que les différences de complexité des connecteurs sont également visibles lorsque les personnes non-natives accèdent à leurs connaissances implicites. Au chapitre 7, les personnes non-natives ne ralentissaient pas leur lecture lorsque le connecteur *cependant* était incorrectement utilisé – contrairement à l'utilisation incorrecte de *mais* qui provoquait des ralentissements de lecture. En d'autres termes : certains connecteurs étaient nettement plus difficiles à maitriser que d'autres.

Le fait que certains connecteurs sont plus difficiles à maitriser implique nécessairement qu'il existe des connecteurs qui sont très *faciles* à maitriser. Nous avons ainsi observé au chapitre 5 que les connecteurs *néanmoins* et *par contre* étaient maitrisés dans la tâche d'insertion à un niveau similaire à celui des personnes natives[38]. De plus, nous avons observé au chapitre 6 que les personnes non-natives ralentissaient leur lecture lorsque les connecteurs *donc* et *mais* étaient

[38] en sachant que le niveau natif est très hétérogène aussi, voir section 10.2

incorrectement utilisés (voir aussi chapitre 7 où *mais* provoquait également des ralentissements de lecture lorsqu'il était mal utilisé).

Dans le même chapitre, nous avons en outre observé que les personnes non-natives montraient une réaction différente en fonction de la relation de cohérence. Il leur était particulièrement difficile de répondre aux questions de vérification lorsque les phrases précédentes étaient incohérentes et concessives. Le fait que nous avons pu observer des effets subtiles (et similaires à ceux de personnes natives) en ce qui concerne la complexité de la relation indique également que l'activation de la signification du connecteur *mais* se réalisait rapidement et ne posait pas de problèmes particuliers aux participant-e-s.

Prises ensemble, nos observations suggèrent qu'un connecteur n'est pas toujours difficile à maitriser et que certains connecteurs pourraient même être maitrisés à un niveau natif. Ainsi, nos observations nuancent le constat de Crewe (1990) selon lequel les mauvais usages des connecteurs sont une caractéristique presque universelle à tous les apprenant-e-s d'une L2. En effet, nous pouvons ajouter que cela dépend hautement du connecteur en question.

Cette conclusion a également des implications pour la recherche. Si on veut examiner par exemple les connaissances implicites, il faut faire attention à ce que l'utilisation d'un connecteur complexe ne biaise pas les résultats. Cela signifie également que l'on ne peut pas simplement déterminer la difficulté d'une relation de cohérence sans tenir compte du type de connecteur testé (p.ex., Goldman & Murray, 1992; Ozono & Ito, 2003).

Notre observation selon laquelle les connecteurs diffèrent en complexité pour les apprenant-e-s soulève la question de savoir pourquoi certains connecteurs sont plus difficiles que d'autres.

10.1.2 « Les causes à la difficulté des connecteurs en L2 sont multifactorielles »

Notre deuxième question de recherche était de savoir quels facteurs rendent certains connecteurs plus difficiles que d'autres. Nous avons mis l'accent sur les facteurs linguistiques et cognitifs, donc des facteurs qui sont directement liés aux connecteurs. A cet égard, nous n'avons pas pu identifier un seul facteur qui expliquerait systématiquement pourquoi un connecteur est difficile à maitriser. Nos

observations indiquent plutôt que la difficulté d'un connecteur résulte de plusieurs facteurs en interaction.

Considérons, par exemple, le facteur de la *fréquence* d'un connecteur. Nous avons vu aux chapitres 4 et 7 que des connecteurs rares, tels que *néanmoins,* ne provoquaient pas les mêmes ralentissements de la lecture lorsqu'ils étaient incorrectement utilisés que des connecteurs plus fréquents, comme *mais*. Cette observation en soi pourrait indiquer qu'un connecteur rare est plus difficile à maitriser en L2 qu'un connecteur fréquent. Cependant, nous avons vu dans le chapitre 5 que la fréquence d'un connecteur n'était pas un facteur qui rendait systématiquement tous les connecteurs rares plus difficiles (ou tous les connecteurs fréquents plus faciles). Dans cette expérience, le connecteur *c'est pourquoi,* par exemple, était, malgré le fait qu'il était moins fréquent, mieux maitrisé que *ainsi*. Nous attribuons plutôt cela à un autre facteur qui entrait en jeu, la *transparence sémantique* de *c'est pourquoi*.

Pour expliquer le fait que la fréquence n'a pas influencé les résultats au chapitre 5, alors qu'il semble y avoir un effet au chapitre 7, on pourrait avancer que la fréquence n'est qu'un facteur systématique lorsque les personnes non-natives activent leurs connaissances implicites (voir aussi la section 10.1.4). Cela est toutefois peu probable. Au chapitre 8, une utilisation incorrecte du connecteur *alors*, connecteur extrêmement fréquent, n'a pas été détectée lorsque les participant-e-s accédaient à leurs connaissances implicites, car un autre facteur entrait en jeu : le *transfert négatif*. Nous constatons donc que la raison pour laquelle certains connecteurs sont plus difficiles à maitriser que d'autres doit être clairement multifactorielle et résulte de l'interaction de plusieurs facteurs.

Pour démêler les facteurs qui pourraient jouer un rôle à cet égard, il convient d'aborder maintenant tous les facteurs que nous avons rencontrés au cours de cette thèse. Cela nous permettra d'établir des zones de difficulté qui aideront par la suite de mieux déterminer la difficulté d'un connecteur pour un-e apprenant-e d'une L2. Les facteurs que nous allons discuter maintenant sont les suivants[39] :

- L'opacité sémantique
- La complexité de la relation de cohérence
- L'optionnalité

[39] Il faut noter que notre liste ne représente qu'un premier pas envers l'identification des facteurs en question. Ainsi, notre liste n'est pas exhaustive et il est évident, comme nous allons le voir, que certains facteurs se chevauchent fortement.

- La fréquence
- La modalité / registre
- Les usages pragmatiques
- L'influence crosslinguistique
- La polyfonctionnalité

L'opacité sémantique

Premièrement, nous avons trouvé des indices qu'un connecteur sémantiquement opaque est plus difficile à maitriser qu'un connecteur sémantiquement transparent. Au chapitre 5, nous avons observé que le connecteur *c'est pourquoi* était extrêmement bien maitrisé. Inversement, le connecteur indiquant la même relation de cohérence, *ainsi*, n'était pas aussi bien maitrisé. Cela nous a fait conclure que la grammaticalisation avancée de *ainsi* rendait ce connecteur plus difficile à maitriser en L2 que *c'est pourquoi* puisque sa signification ne pouvait pas être déduite de sa forme lexicale.

Le facteur de la transparence sémantique permet par ailleurs de formuler des hypothèses nouvelles concernant la difficulté d'autres connecteurs. Par exemple, les connecteurs sémantiquement opaques comme *cependant* ou *or* peuvent être considérés comme plus difficiles à maitriser que les connecteurs *par conséquent* ou *dans l'intention de* qui sont sémantiquement plus transparents.

Bien que la transparence sémantique semble être un facteur prometteur, il faut néanmoins noter que les connecteurs *mais* et *donc* sont sémantiquement opaques, mais étaient pourtant bien maitrisés aux chapitres 6 et 7. Cela pourrait s'expliquer par le fait déjà discuté que la difficulté d'un connecteur ne s'explique pas par un facteur seul, mais plutôt par l'interaction avec d'autres facteurs. En effet, dans le cas de *donc*, il se peut que la difficulté de son *opacité* soit atténuée par sa *fréquence* élevée.

La fréquence

Nous avons déjà mentionné que la *fréquence* d'un connecteur est un facteur censé influencer sa difficulté pour un-e apprenant-e. Nos résultats pour ce facteur ne sont cependant pas univoques.

D'une part, nous avons observé que lorsque les personnes non-natives accèdent à leurs connaissances explicites, les connecteurs rares sont aussi bien maitrisés que les connecteurs fréquents (chapitre 5). De plus, elles connaissaient

les fonctions du connecteur rare *cependant* et identifiaient plutôt bien les cas où il était incorrectement utilisé (chapitre 7).

D'autre part, les résultats diffèrent lorsque les personnes non-natives accèdent à leurs connaissances implicites. Ainsi, nous avons observé que les personnes interrogées au chapitre 7 ne ralentissaient pas leur lecture lorsque la phrase était incorrectement indiquée par *cependant*, tandis que le connecter fréquent *mais* provoquait des ralentissements. Cela indique que l'activation de la signification d'un connecteur est plus facile lorsqu'il est fréquent.

Il convient de noter que nous avons déterminé la fréquence des connecteurs dans les chapitres 5 et 7 à l'aide d'un corpus *écrit*. Il se peut cependant que la fréquence d'un connecteur ne soit pertinente pour une personne non-native lorsqu'on tient compte de sa fréquence *orale*. La question qui se pose également est de savoir si les effets observés sont dus à la fréquence d'un connecteur ou si cela est également lié à la modalité et au registre. En effet, comme nous avons vu au chapitre 7, la fréquence, le registre et la modalité se chevauchent fortement. Ainsi, un connecteur rare dans la langue parlée, comme par exemple *hormis le fait que*, appartient à un registre plus élevé et s'utilise plutôt à l'écrit (voir aussi Cuenca, 2013; Crible & Cuenca, 2017). A cet égard, nous avons attribué le fait que le connecteur *par contre* a été bien maitrisé au chapitre 5 au fait qu'il est fréquemment utilisé au quotidien. Inversement, le connecteur *en revanche* a été moins bien maitrisé dans cette tâche, ce qui s'explique par le fait qu'il appartient plutôt à un registre élevé.

En résumé, la fréquence d'un connecteur n'a pas nécessairement d'influence sur la manière dont on le maitrise en L2 lorsqu'on accède à des connaissances explicites (chapitre 5). Cependant, lorsqu'une personne non-native accède à ses connaissances implicites, il semble qu'il lui soit plus facile d'activer la signification d'un connecteur fréquent par rapport aux connecteurs plus rares (chapitre 7). Enfin, nous avons vu que la fréquence doit aussi être considérée en fonction du registre et de la modalité.

La complexité de la relation de cohérence

Nous avons supposé que les connecteurs qui indiquent des relations plus complexes, telles que les concessions, sont plus difficiles à maitriser dans une

deuxième langue (voir la section 5.1.1). Cela s'explique notamment par le fait que les personnes natives mettent plus de temps à lire et à traiter des relations complexes (p.ex., Köhne & Demberg, 2013).

En ce qui concerne les personnes non-natives, nos résultats indiquent d'abord que la relation n'est pas un facteur qui rend certains connecteurs plus difficiles que d'autres. Nous avons vu au chapitre 5 que les connecteurs indiquant une relation de cohérence plus complexe, telle qu'une concession, n'ont pas été systématiquement moins bien maitrisés par les apprenant-e-s que d'autres. Au contraire, le connecteur concessif *néanmoins* était l'un des rares connecteurs testés dans ce chapitre à être maitrisé à un niveau natif. Cette observation est également corroborée par les résultats du chapitre 9. Dans ce chapitre, nous avons observé que les phrases qui devraient être indiquées par *mais* n'étaient pas plus difficiles à identifier que celles indiquées par *puisque*, *parce que* et *alors*, c'est-à-dire des phrases qui contenaient des relations plus simples.

Ces observations s'expliquent aussi par les tâches utilisées. Aux chapitres 5 et 9, nous avons utilisé des tâches d'insertion, c'est-à-dire des tâches où il fallait identifier la relation de cohérence et choisir le bon connecteur. Comme les concessions doivent être marquées explicitement (Murray, 1994, 1997), ces connecteurs étaient potentiellement plus saillants et donc plus faciles à identifier comme connecteurs appropriés. Ainsi, nous concluons qu'un connecteur concessif n'est pas automatiquement plus difficile à maitriser dans une L2 qu'un connecteur causal, notamment lorsque les personnes non-natives accèdent à leurs connaissances explicites.

Mais évidemment, ces observations ne veulent pas dire que la complexité de la relation n'ait aucun effet sur la lecture en L2. En effet, nous avons constaté au chapitre 6 une influence de la relation de cohérence lorsque les personnes non-natives accédaient à leurs connaissances implicites.[40] Dans chapitre, nous avons observé qu'une relation concessive provoquait des ralentissements plus prononcés qu'une relation causale, notamment lorsque les connecteurs étaient incorrectement

[40] Il convient pourtant de noter que le chapitre 6 ne teste qu'un seul connecteur par relation de cohérence, ce qui complique la généralisation des résultats pour les relations testées. Comme discuté au chapitre 6, nous estimons cependant que les connecteurs testés, *mais* et *donc*, sont bien maitrisés par les personnes interrogées.

utilisés. De plus, nous avons vu que les *concessions* provoquaient des temps de réponse plus longs que les *causes*, que la phrase contienne ou non un connecteur.

Cependant, nous pensons que cela soit dû au fait que la complexité de la relation rend un connecteur plus difficile à utiliser lorsque les personnes accèdent à leurs connaissances implicites. Au contraire, les résultats obtenus au chapitre 6 reflètent des processus cognitifs généraux de la lecture, à savoir le fait que les relations concessives sont plus difficiles à traiter que les relations causales (Morera et al. 2017; Köhne & Demberg, 2013; Xu et al., 2018, voir aussi chapitre 2.5.3). Le fait que nous ayons pu observer un effet de la relation de cohérence n'a été donc possible que par la maitrise des connecteurs *mais* et *donc*.

Nous constatons donc que la complexité d'une relation de cohérence peut influencer - à condition que les connecteurs soient maitrisés - la lecture tant en L1 qu'en L2 (chapitre 6), tandis qu'elle n'est pas un facteur qui rend un connecteur nécessairement plus difficile à maitriser lorsqu'on accède aux connaissances explicites (chapitres 5 et 9).

Il convient de noter que l'influence de la complexité de la relation doit être considérée, comme tous les facteurs, en interaction avec d'autres. Ainsi, nous avons constaté qu'une relation, telle qu'une spécification, ne favorise pas toujours la détection immédiate de l'incohérence, même si la relation peut être considérée comme étant simple (chapitre 8). La question de savoir quand il faut employer un connecteur dans une phrase représente donc une autre difficulté de la maitrise des connecteurs en L2. Comme les connecteurs diffèrent à cet égard, *l'optionnalité* d'un connecteur représente un autre facteur qui peut compliquer sa maitrise.

L'optionnalité du connecteur

De manière générale, le degré d'un marquage explicite d'une relation de cohérence ne dépend non seulement du type de la relation (Murray, 1994; Sanders, 2005), mais aussi de la langue et sa stylistique (Kanno, 1989). Nous avons observé à cet égard que lorsque le connecteur est hautement optionnel (par exemple dans une spécification) une phrase sans connecteur a été jugée plus correcte qu'une phrase qui contenait un connecteur correct (chapitre 8). Cela n'indique pourtant pas qu'un connecteur optionnel est automatiquement plus difficile qu'un connecteur qui est nécessaire pour assurer la cohérence d'une phrase. En effet, nous avons vu au

chapitre 6 que le connecteur *donc*, un connecteur causal qui peut être considéré comme étant plutôt optionnel (Sanders, 2005), provoquait des ralentissements de lecture lorsqu'il était incorrectement utilisé. En revanche, les personnes non natives ne réagissent pas à l'absence d'un connecteur concessif, qui est pourtant nécessaire pour assurer la cohérence.

Nous concluons donc qu'un connecteur optionnel n'est pas automatiquement plus difficile qu'un connecteur non optionnel, mais plus généralement que la question de savoir quand un connecteur doit être utilisé représente une difficulté considérable en L2. En effet, au chapitre 8, les personnes non-natives ont certes jugé les phrases marquées avec le connecteur optionnel *c'est-à-dire* comme correctes, mais l'absence d'effet de plafond pour ces phrases (comme observé pour la condition implicite) indique qu'elles avaient une certaine incertitude. Comme nous avons vu, une raison à cette incertitude pourrait être que les spécifications sont marquées explicitement de manière différente dans leur L1 (chapitre 8).

Les fonctions pragmatiques

De manière générale, les compétences pragmatiques sont très difficilement atteintes dans une langue seconde (p.ex., Bardovi-Harlig & Dörnyei, 1998). Nos résultats corroborent cette observation en ce qui concerne la maitrise des connecteurs.

Nous avons observé aux chapitres 5 et 9 que *puisque*, connecteur pragmatiquement très complexe, n'était pas bien maitrisé. Cela ne s'explique pas forcément par le fait que les personnes interrogées ne connaissaient pas la signification de *puisque* : plus de la moitié des personnes non-natives au chapitre 9 faisaient un lien entre ce connecteur et un connecteur correspondant dans leur L1. Il semble plutôt que les personnes non-natives l'ont évité dans les deux tâches où nous l'avons testé parce qu'ils ne connaissaient pas bien ses fonctions pragmatiques. D'autres recherches ont montré que les personnes non-natives s'appuient sur un petit ensemble de connecteurs bien maitrisés et n'utilisent ainsi qu'un ensemble restreint de connecteurs (p.ex., Leedham & Cai, 2013; AlAttar & Abu-Ayyash, 2020). Ainsi, nous pensons que les personnes non-natives ont évité *puisque* aux chapitres 5 et 9 parce qu'ils ne se sentaient pas à l'aise avec ce connecteur pragmatiquement complexe.

Il faut cependant relativiser la constatation selon laquelle tout usage pragmatique rend automatiquement un connecteur plus difficile en L2. Considérons la citation suivante :

> « Certains-connecteurs pragmatiques (mais, puisque, parce que) ont des emplois dans lesquels les connexions qu'ils opèrent ne portent pas sur des contenus, mais sur des actes d'énonciation. Ainsi, dans *Pierre s'est marié, mais c'est un secre*t, la connexion avec *mais* ne porte pas sur le fait que Pierre s'est marié, mais sur le fait de dire que Jean s'est marié. » (Moeschler & Reboul, 1994 : 21)

Comme il s'agit dans l'exemple d'un acte relativement simple et facile à comprendre, on peut supposer que l'utilisation de *mais* dans cet exemple serait moins problématique pour les personnes non-natives que l'utilisation de *puisque* testée au chapitre 9. Ceci est d'autant plus vrai que *mais* semble être un connecteur que les personnes non-natives maitrisent plutôt bien (chapitres 6 et 7). L'exemple montre ainsi qu'il existe aussi des différences dans l'utilisation pragmatique des connecteurs et que certains pourraient être plus ou moins problématiques. Nous concluons que plus de recherches sont nécessaires pour comprendre l'impact des fonctions pragmatiques des connecteurs sur leur maitrise dans une langue seconde.

Dans l'ensemble, nous avons observé à plusieurs reprises des indices que les fonctions pragmatiques d'un connecteur peuvent entrainer certaines difficultés. Mais là encore, les facteurs qui rendent un connecteur plus difficile doivent toujours être évalués en lien avec d'autres facteurs : nous supposons qu'une utilisation pragmatique d'un connecteur peut être maitrisée plus facilement en L2 lorsqu'elle existe également en L1. Ceci révèle le facteur suivant, l'influence *crosslinguistique*.

L'influence crosslinguistique

Nous avons vu au chapitre 8 que la présence d'un connecteur dans la L1 qui semble partager les mêmes fonctions qu'un connecteur de la L2 peut compliquer la maitrise de ce dernier. Plus précisément, nous avons observé que les participant-e-s germanophones au chapitre 8 ne ralentissaient pas leur lecture, contrairement aux personnes natives, lorsqu'ils lisaient des phrases qui contenaient une utilisation incorrecte de *alors*. Cela s'explique par une confusion avec le connecteur allemand *also*, lequel représenterait un choix adéquat si les phrases étaient en allemand. Comme nous savons que les connecteurs incorrectement utilisés peuvent provoquer des ralentissements de la lecture chez les personnes non-natives

(chapitre 6), le manque de la réaction à l'incohérence au chapitre 8 doit être attribué au transfert négatif.

Nous constatons donc que les connecteurs de la L1 exercent une influence sur la maitrise des connecteurs de la L2 et que le transfert négatif est effectivement un facteur à prendre en compte dans une analyse qui détermine la difficulté d'un connecteur.

Outre le transfert négatif, la question se pose de savoir si la présence d'un connecteur équivalent dans la L1 peut représenter un avantage, en d'autres termes, si un transfert *positif* est possible. Nous avons observé à cet égard au chapitre 9 que même si la L1 possède un connecteur correspondant à une fonction pragmatique d'un connecteur de la L2, la maitrise de ce dernier n'est pas assurée.

Il serait intéressant de déterminer si un transfert est également possible d'une L3. Considérons par exemple les connecteurs français *or* ou *car* qui, dû à des similarités orthographiques avec l'anglais *or* ('ou') ou *car* ('voiture'), pourraient éventuellement amener à des erreurs (ou, au moins, à une confusion) chez les personnes germanophones ayant une bonne compétence en anglais. En revanche, une bonne compétence d'une L3 pourrait également faciliter la maitrise de certains connecteurs. Par exemple, pour une personne germanophone qui connaitrait la signification de l'anglais *except* ('sauf'), le connecteur français *excepté que* deviendrait sémantiquement plus transparent. Plus de recherches sont cependant nécessaires pour investiguer à quels égards des connaissances d'une L3 influencent la maitrise des connecteurs d'une L2 (ou, respectivement l'influence d'une L2 sur la maitrise des connecteurs en L3).

La polyfonctionnalité

Selon Steffani et Nippold (1997 : 1054) ainsi que Crewe (1990 : 324), les connecteurs polyfonctionnels représentent une difficulté potentielle pour les apprenant-e-s de langue seconde. En revanche, pour les personnes natives, nous avons observé au chapitre 4 que la polyfonctionnalité ne semble pas être un obstacle majeur pour la lecture. En effet, nous avons observé que les personnes natives ont identifié et activé la bonne signification des connecteurs *aussi* et *or* assez rapidement. Qu'en est-il pour les personnes non-natives ?

Nous avons observé que les personnes non-natives n'avaient pas eu de problème à comprendre le connecteur *mais*. Aux chapitres 6 et 7, elles ont réagi très tôt à son utilisation incorrecte malgré sa polyfonctionnalité. Nous constatons donc que le fait qu'un connecteur soit polyfonctionnel ne le rend pas *forcément* plus complexe. Cependant, les données manquent pour déterminer si la polyfonctionnalité d'un connecteur ne peut pas s'ajouter sa difficulté dans d'autres contextes. Enfin, la difficulté de la mise en correspondance d'un connecteur L1 et d'un connecteur L2 observée au chapitre 8 pourrait être due non seulement au transfert négatif, mais aussi lié au fait que les deux connecteurs *alors* et *also* sont polyfonctionnels.

Nous concluons ainsi que plus des recherches sont nécessaires pour comprendre dans quelle mesure les connecteurs polyfonctionnels dont les relations ne correspondent pas à celles d'un connecteur de L1 sont difficiles à maitriser en L2.

Résumé des facteurs

Au cours de cette thèse, nous avons identifié plusieurs facteurs qui rendent certains connecteurs en langue seconde plus difficiles que d'autres. Nous avons constaté, par exemple, que la mise en correspondance d'un connecteur de la L2 avec les fonctions d'un connecteur de la L1 représente une difficulté de la maitrise des connecteurs en langue seconde. Cela pose notamment un problème lorsque les fonctions ne correspondent pas entre les deux connecteurs (chapitre 8). Mais même lorsqu'il y a des similitudes (comme par exemple quant aux fonctions pragmatiques, chapitre 9), la maitrise du connecteur de la L2 n'est pas assurée.

De plus, nous avons vu que les connecteurs moins fréquents et appartenant à un registre élevé de l'écrit rendent l'activation de leur signification pendant la lecture plus difficile (chapitre 7). La connaissance explicite, par contre, n'est pas forcément affectée par ces facteurs (chapitres 5 et 7).[41]

Nous avons déjà évoqué au chapitre 7 que la difficulté d'un connecteur ne s'explique pas par un seul facteur, mais de manière multiple. Pour déterminer la difficulté d'un connecteur pour un apprenant-e d'une L2, il faut donc analyser les

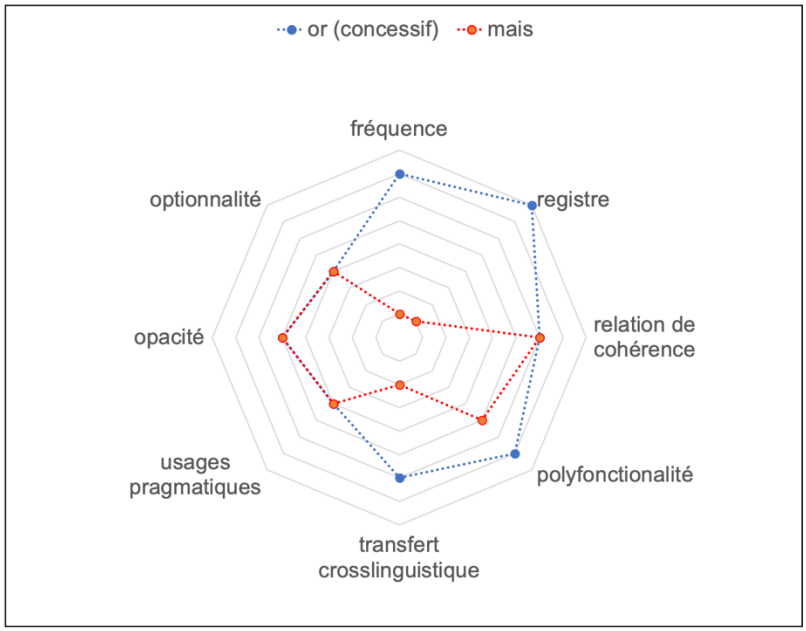
[41] Nous allons revenir sur la différence entre les connaissances implicites et explicites dans la section 10.1.4

facteurs mentionnés pour un connecteur donné, les mettre en balance et les comparer avec d'autres connecteurs. Ce type d'analyse peut contribuer à rendre l'enseignement plus adapté aux élèves, car il permet d'identifier les connecteurs problématiques. Pour donner un exemple d'une telle analyse, nous allons brièvement estimer la difficulté du connecteur français *or* et la comparer avec la difficulté du connecteur *mais.*

Or est un connecteur polyfonctionnel qui indique deux relations de cohérence différentes, une continuation et une concession (voir *LexConn*, Roze et al., 2012). Ces deux relations de cohérence sont opposées en ce qui concerne leur continuité. De plus, le connecteur est hautement optionnel dans l'une d'entre elles. Sa mise en correspondance avec un connecteur de la L1 (par exemple en anglais) n'est pas évidente - elle peut même mener à des confusions crosslinguistiques (par exemple avec l'anglais *or* signifiant 'ou'). En outre, il est sémantiquement opaque et appartient à un registre élevé lié au discours écrit. *Mais* en revanche, est un connecteur utilisé aussi bien à l'écrit qu'à l'oral. Même s'il est également opaque et polyfonctionnel, il est beaucoup plus fréquent que *or* (chapitre 4) et n'est pas optionnel dans sa fonction primaire (à savoir, indiquer une concession).

Nous voyons donc que le connecteur *or* doit être classifié comme un connecteur plus problématique pour un apprenant-e L2, alors que *mais* semble avoir des caractéristiques qui le rendent plus accessible. Notre analyse de *mais* et *or* à la base des facteurs identifiés au cours de la thèse peut également être visualisée dans la Figure 10.1

Figure 10.1. Visualisation des zones de difficultés de *mais* et *or*.



L'idée dans cette visualisation est que plus la surface d'un connecteur est petite, plus on peut supposer qu'il est facile à maitriser en L2. Bien que la visualisation soit quelque peu subjective et approximative, elle permet tout de même de comparer la difficulté estimée des différents connecteurs, ou même des fonctions différentes d'un même connecteur (p.ex., *or concessif* vs *or continuation*).

Cependant, une telle visualisation ne prend pas en compte le fait que la difficulté d'un connecteur dépend toujours de l'apprenant-e. Comme nous allons le voir maintenant, il existe en effet une grande variation au sein du groupe des personnes non-natives quant à la maitrise des connecteurs en L2.

10.1.3 « La maitrise des connecteurs en L2 peut être prédite par l'exposition à l'écrit et par la compétence linguistique de l'apprenant-e»

Nous avons mentionné au chapitre 3.5 qu'il semble exister une grande variation interpersonnelle dans la maitrise des connecteurs en L2. Nos résultats confirment cela et nous indiquent plusieurs compétences qui corrèlent avec une meilleure maitrise des connecteurs.

Dans nos expériences, une mesure qui a permis de prédire la maitrise des connecteurs est la taille du vocabulaire, telle que mesurée par le test *Lextale*. Cela était le cas notamment lorsque les personnes accédaient à leurs connaissances

explicites : une taille de vocabulaire plus grande prédisait ainsi au chapitre 5 des meilleurs scores dans la tâche d'insertion.
De plus, les scores de *Lextale* prédisaient une meilleure détection des utilisations incorrectes des connecteurs. Au chapitre 7, par exemple, les gens qui avaient des scores élevés dans cette tâche identifiaient mieux que les autres les emplois incorrects de *cependant* et de *mais*. Cette observation est également soutenue par les résultats du chapitre 8. Dans ce chapitre, les participant-e-s ayant un score élevé dans la tâche de *Lextale* identifiaient plus facilement l'usage incorrect de *alors*. Et finalement, au chapitre 9, la capacité à identifier les contraintes pragmatiques et syntaxiques et à choisir le bon connecteur a été également prédit par ces scores.

Nous pensons que le fait que la taille du vocabulaire prédisait la compétence des apprenant-e-s lors de tâches explicites reflète le fait que les connecteurs sont souvent enseignés en tant que simples éléments lexicaux, en les présentant par exemple sous forme de liste (Zamel, 1983). Cette observation pourrait aussi expliquer les résultats obtenus avec le connecteur *puisque* aux chapitres 5 et 9 : une personne qui apprend uniquement la signification d'un connecteur en tant qu'élément du lexique négligera davantage ses usages pragmatiques.

De plus, en tenant compte des corrélations observées au chapitre 5, il faut considérer que la taille du vocabulaire pourrait être aussi représentative de la compétence linguistique en général (voir aussi la section 10.3). Nos résultats indiquent ainsi que la maitrise des connecteurs peut s'améliorer avec une meilleure compétence linguistique (voir aussi, p.ex., Geva, 1986; Cho & Sin, 2014; Tazegül, 2015).

Il est intéressant de noter à cet égard que la taille du vocabulaire ne prédisait dans nos expériences qu'une meilleure connaissance *explicite* (chapitres 5, 7, 8, 9) et pas une meilleure connaissance *implicite*. En effet, dans les chapitres 6, 7 et 8, les scores de *Lextale* ne prédisaient pas les temps de lecture plus rapides (ou plus lents). Cette absence d'effet est intriguante, notamment pour ce qui est de la détection de l'incohérence, car on aurait pu imaginer un lien entre le fait d'être compétent-e dans une langue et celui de détecter plus rapidement de l'incohérence. Nos résultats indiquent néanmoins qu'il n'y a pas de lien entre la taille du vocabulaire et la capacité à activer la signification d'un connecteur et à l'intégrer lors de la lecture. Il se peut que les connaissances implicites, en tant que

connaissances inconscientes, soient en quelque sorte trop vagues et subtiles pour être affectées par la taille du vocabulaire.
Le fait que nous pas pu documenter un lien entre les deux mesures ne peut d'ailleurs pas être attribué aux problèmes méthodologiques de la tâche de lecture segmentée, puisque nous avons observé des réactions à l'incohérence (voir aussi la section 10.3 sur les limites de la méthodologie choisie).

La deuxième compétence qui a prédit dans plusieurs expériences la maitrise des connecteurs en L2 était les compétences de grammaire, telles que mesurées par la *Tâche de grammaire* (Zufferey & Gygax, 2020a). Les compétences grammaticales ont ainsi prédit un meilleur score dans la tâche d'insertion au chapitre 5. De plus, de meilleures compétences grammaticales ont également prédit une meilleure identification des contraintes pragmatiques et syntaxiques au chapitre 9. Pris ensemble avec les résultats de la tâche *Lextale*, nos résultats pourraient indiquer que la maitrise des connecteurs n'implique pas seulement un vocabulaire étendu, mais aussi des compétences grammaticales.

La troisième mesure individuelle qui a prédit à plusieurs reprises la maitrise des connecteurs était leur exposition à l'écrit. Par exemple, au chapitre 7, nous avons vu qu'une meilleure exposition à l'écrit prédisait une meilleure connaissance explicite des connecteurs. Cela peut s'expliquer par le fait qu'une personne qui a beaucoup lu dans sa vie a pu développer un bon sens de la cohérence d'un texte. Cette sensibilisation à la structuration d'un texte lui fait prendre davantage conscience de l'importance des connecteurs, ce qui facilite leur acquisition et augmente leur maitrise.

Nous avons vu à cet égard que cette *compétence de discours* se manifeste indépendamment de la langue et doit être considérée comme étant universelle. En effet, une plus grande exposition à la L1 a prédit à plusieurs reprises une meilleure compétence avec les connecteurs de la L2 (chapitres 5 et 7). En suivant l'*hypothèse de l'interdépendance* de Cummins (1979), nous pensons que le développement de d'une compétence discursive peut faciliter la maitrise des connecteurs en L2.

Il est vrai à cet égard que la tâche utilisée (ART) mesure simultanément d'autres facteurs individuels, comme le statut socio-économique et le milieu éducatif de la personne. Ainsi, il est possible qu'une personne qui a obtenu un score élevé dans cette tâche n'ait pas lu les livres d'un auteur-trice correctement identifié-e. Cela n'enlève pourtant rien à notre interprétation : on peut supposer qu'une

personne ayant un accès facile à l'éducation et à la langue écrite est plus susceptible de développer une compétence discursive qu'une personne issue d'un milieu défavorisé. De plus, comme les noms d'auteurs-trice-s sont parfois peu connus, le seul fait de les connaitre indique que la personne est, d'une manière ou d'une autre, en contact avec la littérature.

Finalement, nous avons observé que les personnes ayant des connaissances métalinguistiques sur le connecteur *puisque* semblaient mieux le maitriser (chapitre 9). Il faut bien sûr tenir compte du fait que cette observation a été faite dans le cadre d'une étude exploratoire avec un nombre relativement faible de participant-e-s. Néanmoins, elle est très concluante : les personnes qui comprennent consciemment les fonctions d'un connecteur et qui arrivent les verbaliser semblent avoir de meilleures chances de maitriser ce connecteur que les personnes qui le mettent seulement en correspondance avec un connecteur de la L1. Dans ce contexte, il serait intéressant d'examiner si une meilleure connaissance méta d'une fonction pragmatique d'un connecteur facilite également la maitrise lorsqu'on accède à la connaissance implicite. Cela n'est pas nécessairement le cas, car la maitrise des connecteurs varie considérablement en fonction du type de connaissance auxquelles on a accès, comme nous allons le voir maintenant.

10.1.4 « La maitrise des connecteurs dépend du type de connaissances auxquelles l'apprenant-e peut accéder »

Notre dernière question de recherche était de savoir si la maitrise des connecteurs dépendait du type de connaissances, implicite ou explicite (p.ex., VanPatten & Benati, 2010; Ellis, 2009), auquel l'apprenant-e peut accéder. Afin d'examiner cette question, nous avons utilisé d'une part une tâche de lecture segmentée qui ne permettait que l'accès aux connaissances implicites (chapitres 4, 6, 7 et 8). D'autre part, nous avons utilisé des tâches qui permettaient aux participant-e-s d'accéder à leurs connaissances explicites (chapitres 5, 7, 8 et 9). De plus, dans les chapitres 7 et 8, nous avons utilisé les deux tâches, ce qui nous a permis de comparer directement les deux types de connaissances.

De manière générale, nous avons constaté que la maitrise des connecteurs en L2 dépend dans une large mesure du type de connaissances auxquelles l'apprenant-e peut accéder. Comme nous avons déjà mentionné dans la section 10.1.2, certains facteurs tels que la fréquence d'un connecteur ainsi que le type de

relation de cohérence qu'il indique semblent jouer un rôle important lorsque les personnes accèdent à leurs connaissances implicites (chapitre 6), tandis qu'ils ne rendent pas les connecteurs nécessairement plus difficiles lorsqu'elles accèdent à leurs connaissances explicites (chapitre 5).

Nous avons observé notamment aux chapitre 7 et 8 à quel degré la connaissance implicite peut différer de la connaissance explicite. Au chapitre 7, les personnes ne réagissaient pas à l'incohérence provoquée par l'utilisation incorrecte de *cependant* lorsqu'elles accédaient à leurs connaissances implicites. Par contre, lorsque les connaissances explicites ont été testées, les participant-e-s étaient capables de détecter ces emplois erronés.

Au chapitre 8, nous avons observé la même dissociation : lorsque les participant-e-s accédaient à leurs connaissances implicites, ils ne ralentissaient pas lorsqu'ils lisaient une utilisation erronée de *alors*. Lorsqu'ils devaient évaluer les phrases en accédant à leurs connaissances explicites, ils arrivaient en revanche à discriminer les emplois incorrects des emplois corrects des connecteurs.

Les résultats montrent cependant qu'on ne peut pas en conclure que l'incohérence n'est jamais détectée lorsque les gens accèdent à leurs connaissances implicites. En effet, au chapitre 6 ainsi qu'au chapitre 7, nous avons observé que lorsque les connecteurs plus faciles ont été utilisés (*donc*, *mais*), l'incohérence a provoqué également des ralentissements de la lecture.

En résumé, notre travail illustre la différence considérable entre les deux types de connaissances. Cela souligne l'importance de les distinguer, à la fois pour la recherche et pour l'enseignement. En effet, si on veut évaluer si une personne maitrise des connecteurs (pour des raisons scientifiques ou éducatives), il faut bien tenir compte du fait que cette maitrise va dépendre de la définition de la maitrise et de la manière dont on l'évalue.

Quant à la constatation de Crewe (1990) selon laquelle la mauvaise utilisation des connecteurs d'une L2 est une caractéristique presque universelle des apprenant-e-s, nos études permettent de nuancer fortement son propos. Nous pouvons y ajouter que cette difficulté dépend du connecteur en question, de la personne en question et, finalement, de ce qu'on entend par la maitrise, c'est-à-dire si la personne accède à ses connaissances implicites ou explicites.

10.2 Les observations sur la maitrise des connecteurs en L1

Bien que l'objectif de la présente thèse était d'étudier la maitrise des connecteurs dans une langue seconde, nos résultats nous permettent également des tirer des conclusions nouvelles sur la maitrise des connecteurs en L1, car nous avons testé dans toutes les expériences des personnes natives comme groupe de contrôle.

Premièrement, nous avons observé une grande variation dans la maitrise que les personnes natives ont de leur L1. Les scores des tâches qui mesuraient la taille du vocabulaire (*Lextale*), les compétences grammaticales (*Tâche de grammaire*) et l'exposition à l'écrit (*ART*) ont toujours montré une grande variance (chapitres 5, 7, 8 et 9). Nos résultats corroborent ainsi la littérature grandissante qui indique une grande variation dans les compétences linguistiques des personnes natives (Kamalski, Lentz, Sanders & Zwaan, 2008; Van Silfhout, Evers-Vermeul, Mak & Sanders, 2014; Scholman, Demberg & Sanders, 2020). De plus, nous avons vu que les personnes natives ne maitrisent pas tous les connecteurs de la même manière. Bien que les personnes natives étaient généralement sûrs d'eux au moment d'indiquer dans des tâches d'évaluation si un connecteur était incorrectement utilisé ou non (chapitres 7 et 8), nous avons observé qu'ils leur étaient plus difficile d'identifier certains connecteurs adéquats dans des tâches d'insertion (chapitres 5 et 9). A cet égard, il leur a semblé plus difficile de choisir le bon connecteur lorsque celui-ci indiquait une relation complexe (chapitre 5) ou s'imposait par des contraintes pragmatiques (chapitre 9).

Nous avons observé en outre que la lecture des personnes natives est également affectée par le type de connecteur utilisé. Plus précisément, nous avons observé au chapitre 4 que lorsque les connecteurs *ainsi*, *néanmoins*, *or* et *aussi* ont été incorrectement utilisés, les personnes natives ne ralentissaient leur lecture qu'au segment final de la phrase, alors que l'utilisation incorrecte des connecteurs plus fréquents *donc* et *mais* provoquaient des ralentissements déjà à l'avant-dernier segment de la phrase. Nous attribuons cela au fait que l'activation et l'intégration de la signification de *mais* et *donc* est, grâce à leur fréquence, facilitée par rapport aux connecteurs plus rares. Ces observations sont également étayées par les observations au chapitre 7. Dans ce chapitre, nous avons constaté que le dernier segment d'une phrase incohérente marquée avec *cependant* a été lue plus lentement par rapport à une phrase incohérente marquée avec *mais*. De nouveau,

nous attribuons cela au fait que la signification de *cependant* n'a pas été activée et intégrée au contenu de la phrase aussi rapidement que pour *mais*.

Deuxièmement, nous avons vu que la capacité à maitriser les connecteurs a été prédite par différentes mesures. Par exemple, nous avons constaté au chapitre 5 que la taille du vocabulaire prédisait un meilleur score dans la tâche d'insertion. De même, un meilleur score dans la tâche *Lextale* prédisait dans les tâches de jugement aux chapitres 7 et 8 une meilleure détection des utilisations incorrectes des connecteurs. En fait, dans toutes les tâches d'évaluation ou d'insertion (chapitres 5, 7 et 8), *Lextale* était une mesure prédictive. Il n'y a qu'au chapitre 9 que nous avons constaté qu'un vocabulaire plus étendu ne prédisait pas une meilleure maitrise de *parce que* et *puisque*.[42]

Les compétences grammaticales semblaient également corréler avec une meilleure maitrise des connecteurs chez les natifs également. Ainsi, les scores dans la *Tâche de grammaire* (Zufferey & Gygax, 2020a) prédisaient au chapitre 5 des meilleurs scores dans la tâche d'insertion.

Comme chez les personnes non-natives, la maitrise des connecteurs par les personnes natives a été prédite par une meilleure exposition à l'écrit. En effet, le score ART prédisait à la fois au chapitre 5 une meilleure utilisation des connecteurs ainsi qu'une meilleure détection des utilisations incorrectes au chapitre 7. Cela indique qu'une meilleure compétence de discours influence également la lecture en L1. Il convient de noter à cet égard que nous avons trouvé au chapitre 9 qu'une meilleure exposition à l'écrit de la L2 prédisait aussi une meilleure maitrise des connecteurs en L1. Cela corrobore donc notre conclusion selon laquelle la compétence discursive joue un rôle indépendamment de la langue testée.

Une autre observation intéressante concerne la détection de l'incohérence. Nous avons vu que des connecteurs incorrects entrainaient toujours (parfois plus tôt, parfois plus tard) un ralentissement de la fluidité de lecture, ce que nous attribuons à une perturbation du traitement mental de la phrase (chapitres 4, 6 et 7). Cependant, lorsque le connecteur était hautement optionnel et indiquait une spécification, donc un ajout à la clause principale qui n'est pas forcément nécessaire, le ralentissement était beaucoup moins prononcé et ne persistait pas (chapitre 8). Cela illustre le fait que, lorsque les personnes natives font attention au

[42] Pour *parce que*, cela s'explique par le fait que ce connecteur a généré des scores plafonds. En revanche, nous n'avons pas d'explication pour l'absence d'effet pour *puisque*.

message principal d'une phrase, elles en arrivent à négliger des instructions erronées des connecteurs et ne sont que temporairement affectées dans leur lecture.

En résumé, notre thèse a montré que la maitrise des connecteurs par les personnes natives varie d'une personne à l'autre (chapitres 5, 7, 8 et 9) et d'un connecteur à l'autre (chapitres 4, 5, 7 et 9). Il convient de noter en outre que les personnes natives ont toujours montré une réaction lorsqu'un connecteur était utilisé de manière incorrecte, indépendamment de la tâche mais à des degrés différents : lorsqu'un connecteur optionnel était utilisé (chapitre 8), il n'y a eu que des ralentissements légers et temporels de la lecture, alors que lorsqu'un connecteur fréquent et nécessaire pour la cohérence de la phrase était utilisé, les effets de ralentissement étaient plus prononcés (chapitres 4 et 6).

10.3 Limites

Cette thèse avait pour but d'examiner la maitrise des connecteurs des personnes non-natives et investiguer si, et pourquoi, certains connecteurs semblent être plus difficiles que d'autres. Cela a été atteint en adoptant une approche expérimentale. Évidemment, comme toute recherche quantitative et réductionniste, le travail présent connait des limitations. Quelques-unes de ces limites ont déjà été mentionnées dans certains chapitres, nous nous contenterons ici de résumer et de discuter les limites générales.

Limites concernant l'artificialité de l'approche expérimentale

De manière générale, le dilemme d'une approche expérimentale, notamment en sciences sociales, consiste à vouloir contrôler les variables autant que possible, tout en restant authentique (Dipboye & Flanagan, 1979). Zufferey (2020) souligne que cela peut être un inconvénient par rapport aux études de corpus qui travaillent avec des textes plus authentiques. Dans la présente thèse, le dilemme entre l'artificialité/authenticité et la contrôlabilité s'est manifesté notamment de deux manières.

D’abord les phrases utilisées étaient artificiellement construites. Certes, cela nous a permis de contrôler les variables testées, mais nous ne savons pas si les participant-e-s lisaient ces phrases de la même manière qu’ils lisent un discours au quotidien. De plus, contrairement aux discours du quotidien, nos phrases ont été présentées sans contexte et sans implication émotionnelle ou personnelle des participant-e-s. Comme déjà discuté au chapitre 9, cela est spécialement problématique lorsque la phrase contient un niveau pragmatique et nécessite un certain contexte pour être comprise. Ainsi, nous ne savons pas si les participant-e-s ont été capables d’identifier ce niveau pragmatique dans tous les cas.

Il convient cependant de noter que le fait que nos phrases aient été artificiellement construites et n’impliquaient pas émotionnellement les participant-e-s ne nous a pas empêché de trouver plusieurs effets, comme celui de l'incohérence (chapitre 4, 6, 7 et 8). Cela indique que les participant-e-s ont dû lire attentivement les phrases. Il se pourrait même, qu’au contraire, l’utilisation de phrases ayant une implication émotionnelle (comme l'utilisation de déclarations politiques) aurait biaisé les résultats. De plus, l’ajout de plus de contexte n'aurait pas seulement rendu la tâche plus difficile (Geva, 1986; Tskhovrebova et al., 2022), mais aurait introduit des éléments supplémentaires et ce faisant, d’autres biais potentiels.

A cela s’ajoute le fait que nous avons parfois utilisé des connecteurs différents pour les mêmes phrases de manière interchangeable (chapitres 5 et 7). Théoriquement, cela aurait pu amener à des problèmes de registre, par exemple au chapitre 7, où *cependant* aurait été inapproprié dans une phrase alors que *mais* convenait. Cependant, il convient de noter que toutes les phrases ont été relues par des locuteur-trice-s francophones et aucun problème de registre n'a été identifié. De plus, il était crucial d’utiliser les mêmes phrases pour pouvoir comparer les temps de lecture et en tirer des conclusions sur la maitrise des connecteurs utilisés.

Le dilemme de l’artificialité de l’approche expérimentale ne concerne pas seulement les phrases utilisées mais aussi les types de tâches utilisées. Dans nos expériences, nous avons utilisé des tâches diverses qui ne correspondent pas à une lecture normale. A cet égard, il convient de noter que nous avons testé uniquement la lecture au niveau inter-phrastique, c’est-à-dire entre deux clauses ou deux phrases. Cependant, les connecteurs peuvent également indiquer des

relations de cohérence plus éloignées dans le discours, par exemple entre deux paragraphes.

Nos études sont certes limitées à cet égard, mais il est vrai aussi que les apprenant-e-s, notamment ceux qui ont un faible niveau linguistique, comprennent mieux les relations de cohérence lorsque le contexte est limité (Geva, 1986). Il est également connu que davantage de contexte rend plus difficile l'identification des relations de cohérence (Tskohovrebova et al., 2022). Tester au niveau inter-phrastique nous a donc semblé être la meilleure façon de tester la maitrise des connecteurs en L2.

La question se pose de savoir si nos résultats permettraient des conclusions sur la compréhension d'un texte entier. En effet, il est très difficile de généraliser une observation faite pour une simple paire de phrases à la compréhension d'un texte entier. Nous aimerions préciser à cet égard que nos études ne portaient pas sur la *compréhension de texte*, mais plutôt sur la *maitrise des connecteurs*. C'est une distinction importante, car il se peut que la compréhension de texte soit mauvaise malgré une bonne utilisation des connecteurs (ou *vice versa* si on tient compte de l'utilité des signaux alternatifs, voir chapitre 3.2). Nous concluons qu'il est nécessaire de mener davantage de recherches pour étudier la maitrise des connecteurs dans une langue seconde au niveau du texte et en tirer des conclusions sur la manière dont une personne non-native le comprend.

Limites concernant la généralisation des résultats

Une autre limite tient au fait que les personnes incluses dans nos études ne sont pas nécessairement représentatives de toutes les personnes non-natives apprenant-e-s du français. Pour permettre une meilleure généralisation des résultats, il convient de tester un nombre assez grand de participant-e-s. Nous avons fixé le nombre de nos participant-s-s sur la base de la littérature de référence (p.ex., Degand & Sanders, 2002; Zufferey & Gygax, 2017). Néanmoins, il est possible qu'un nombre de participant-e-s encore plus élevé aurait nuancé davantage les résultats.

A cela s'ajoute une limite quant à la plateforme *Prolific* qui a été souvent utilisée dans nos études pour recruter des participant-e-s (chapitres 4, 6, 7, 8 et 9) : le recrutement via *Prolific* a impliqué que nous ne connaissions pas nos participant-

e-s, c'est-à-dire que nous ne savions pas, par exemple, si les participant-e-s du groupe contrôle étaient réellement des personnes natives. De plus, nous ignorons les conditions dans lesquelles les expériences ont été faites.

Cependant, notre approche expérimentale nous permet, contrairement aux études de corpus[43], de contrôler ces biais potentiels. En analysant les réponses quantitativement (c'est-à-dire avec des seuils) et qualitativement, nous avons pu contrôler que les réponses ont été effectivement données de manière attentive. Par exemple, aux chapitres 4, 6 et 7, nous avons utilisé des questions de vérification qui nous ont permis de déterminer si un-e participant-e avait lu les phrases (nous reviendrons encore à ce contrôle dans un instant) et les compétences de langue ont été contrôlées par le score obtenu au test *Lextale*.

L'approche consistant à recruter sur Internet a même des avantages, car elle permet de cibler un public plus hétérogène et contribue ainsi à une meilleure généralisation des données (voir aussi Santana et al., 2021 pour une discussion). Dans le cadre de nos expériences, il y a aussi des indices que le recrutement via *Prolific* n'a pas été défavorable, car nous avons observé les mêmes tendances indépendamment du moyen de recrutement. Nous avons observé au chapitre 5 que les connecteurs *puisque* et *cependant* n'ont pas été bien maitrisés par les gymnasien-ne-s et les résultats du chapitres 7 et 9 – dans lesquels nous avons recruté via *Prolific* – ont donné des résultats concordants. Ceci indique que le recrutement via Internet peut être considéré comme un moyen valide pour la recherche quantitative.

De manière générale, il est également connu que la qualité des données obtenues via *Prolific* est élevée, notamment en comparaison avec d'autres plateformes de recrutement (Palan & Schitter, 2018; Eyal, Rothschild, Gordon, Evernden, & Damer, 2021).

Limites concernant le niveau de compétences linguistiques

Comme discuté au chapitre 3.5, il est connu que les compétences linguistiques des personnes non-natives influence leur maitrise des connecteurs (p.ex., Geva

[43] en effet, dans une analyse d'un corpus qui n'est pas établi par nous-mêmes, nous ne pouvons encore moins contrôler si des textes analysés ne sont pas écrits par des personnes fatiguées, distraites ou dans d'autres conditions.

1986). Alors que notre approche était plutôt axée sur les connecteurs, nous avons ainsi dû évaluer les compétences linguistiques et en tenir compte. Pour faire cela, nous avons reporté les scores des participant-e-s à la tâche *Lextale* tout au long de la thèse. Les questions se posent donc : *Lextale* reflète-t-il vraiment le niveau général de compétences linguistiques ?

D'abord, il convient de noter que bien que Lextale ne capture qu'une partie de la compétence linguistique, nous avons observé des corrélations de ces scores avec d'autres compétences, telles que la maitrise des connecteurs ou les connaissances en grammaire (chapitres 5 et 9). Cela indique que les scores de *Lextale* peuvent effectivement être généralisés dans une certaine mesure aux compétences linguistiques en général. Évidemment, cette généralisation a des limites (voir aussi Puig Mayenco, Chaouch-Orozco, Liu & Martín-Villena, 2023, pour une discussion sur la validité de la version anglaise de *Lextale*), car la notion de compétence linguistique comprend un large éventail de compétences différentes. Cependant, il faut noter que les corrélations que nous avons observées (p.ex., chapitres 7, 8 ou 9) étaient compatibles avec nos hypothèses et justifiaient ainsi de manière *post-hoc* l'implémentation de cette tâche.

Finalement, la tâche de *Lextale* est d'un point de vue méthodologique très pratique, car elle est assez rapide à implémenter, et aussi à réaliser. Cela a deux avantages : les effets de fatigue sont limités et la facilité d'implémentation facilite la reproductibilité de nos expériences dans des recherches futures. En effet, la tâche de *Lextale* est une mesure fréquemment utilisée dans des recherches en langue seconde (Dewaele & Dewaele, 2020).

Limites concernant la lecture segmentée

Aux chapitres 4, 6, 7 et 8, nous avons utilisé la tâche de lecture segmentée pour mesurer la connaissance implicite des connecteurs. Bien que cette tâche soit tout à fait valide pour mesurer la connaissance implicite (Cho, 2020), elle connait des limitations importantes.

Premièrement, la lecture segmentée permet en comparaison avec les appareils d'oculométrie des mesures beaucoup moins fines. Outre le fait qu'elle ne correspond pas tout à fait une lecture naturelle (comme discuté auparavant), elle ne permet pas non plus de mesurer les régressions, ni les points de fixation sur des

éléments précis au sein d'un segment affiché (voir aussi la section 3.1.1). Nous concluons ainsi que plus de recherche est nécessaire pour valider nos observations avec d'autres méthodologies (voir aussi la section suivante).

Une autre limite est que nous avons utilisé dans cette tâche à certains endroits des connecteurs qui se trouvaient en début de phrase, séparés par un point de la phrase précédente (les connecteurs *cependant* et *aussi* dans les chapitres 4 et 7), alors que d'autres introduisaient une nouvelle clause après une virgule. Ceci est évidemment désavantageux, car il se peut que le fait que le connecteur introduise une phrase modifie la lecture. Cependant, cela était nécessaire pour garantir une utilisation correcte du connecteur, car ces positions correspondent à leurs positions naturelles dans un discours authentique.

De plus, de manière plus importante, le fait que certains connecteurs figuraient en position initiale de la phrase n'a pas changé les résultats de nos expériences, puisque nous n'avons constaté les effets de cohérence qu'en fin de phrase (chapitres 4, 6 et 7), donc des segments qui ne devraient pas être touchés par le fait que le connecteur introduisait soit une nouvelle phrase ou soit une deuxième clause.

Une autre limitation concerne les questions de vérification utilisées aux chapitres 4, 6, 7. Une première question qui se pose dans ce contexte est de savoir si le temps de lecture pour une phrase ne devrait pas être analysée si la réponse donnée à la question était fausse. En effet, lorsqu'une personne ne répond pas correctement à la question, il est possible qu'elle n'ait pas bien lu la phrase, ce qui représenterait un biais important.

Il convient cependant de noter que les notions de *vrai* et de *faux* ne sont pas toujours aussi claires et peuvent avoir conduit les participant-e-s à choisir une réponse que nous avons qualifiée comme incorrecte. Par exemple, en (1), il n'est pas précisé que Nadia n'a pas de chien.

1) Nadia a peur de tous les animaux à fourrure mais elle a toujours eu un chat. Nadia a toujours eu un chien. Vrai ou faux?

Ainsi, donner une réponse négative est sûrement la chose la plus rationnelle à faire, mais les opinions des gens peuvent différer. Nous avons voulu être tolérants vis-à-vis de ces opinions divergentes. Pour ne pas perdre des données légitimes, nous

avons donc décidé de garder les temps de lecture d'une phrase même lorsqu'une question de vérification a engendré une réponse incorrecte.

De manière générale, il convient aussi de noter que les participant-e-s ont donné des réponses correctes aux questions de vérification (les réponses trop basses ont été éliminées). Par exemple, au chapitre 6 les personnes non-natives avaient dans la première expérience un score moyen de 88% de réponses correctes et dans la deuxième expérience un score moyen de 83%. Cela nous montre que les participant-e-s ont bien lu les phrases.

Il est vrai que même si les participant-e-s lisent attentivement les phrases, nous ne savons pas comment ils les ont interprétées, c'est-à-dire s'ils ont été capables d'identifier correctement la relation de cohérence. En effet, les questions de vérification visaient à contrôler que les participant-e-s aient bien *lu* les phrases mais les réponses données ne reflètent pas forcément le fait que les participant-e-s ont *compris* ce qu'ils ont lu.

Évidemment, nous ne pouvons pas affirmer avec certitude que les personnes ont vraiment compris les phrases en nous basant seulement sur leurs réponses. Nous pensons toutefois que cela a été le cas, du fait que nous avons observé des effets de cohérence : si les personnes non-natives avait interprété les phrases différemment, nous n'aurions pas observé de réactions à l'incohérence aux chapitres 5, 6 et 7. De plus, il faut rappeler que les participant-e-s étaient généralement compétent-e-s sur le plan linguistique, comme en témoignent leurs scores élevés dans les mesures de *Lextale* ou de la *Tâche de grammaire*.

En dépit de toutes ces critiques et malgré le fait que la mesure de la connaissance implicite reste un défi général (Ellis, 2006), la lecture segmentée est une approche largement utilisée dans la littérature pour tester les connaissances implicites (p.ex., Jiang, 2011; Zufferey & Gygax, 2017) et nous a permis d'observer l'effet que l'utilisation incorrecte d'un connecteur peut (ou ne peut pas) avoir sur la lecture des personnes non-natives.

Négligence de l'enseignement ?

Nous avons mentionné que la présente thèse visait principalement à tester les facteurs liés aux connecteurs, mais prenait également en compte d'autres variables qui expliquent la maitrise des connecteurs en L2, telle que la variation

interpersonnelle ainsi que le type de connaissance auxquelles les personnes ont accès. A cet égard, nous n'avons pas mis l'accent sur la question de savoir comment les connecteurs ont été enseignés et appris, un facteur qui semble pourtant crucial pour leur maitrise en L2 (Crewe, 1990). On pourrait ainsi argumenter que le fait que certaines personnes maitrisaient mieux *puisque* n'était pas forcément dû à leurs connaissances métalinguistiques, mais plutôt au fait qu'ils ont appris à utiliser ce connecteur d'une certaine manière (qui a résulté dans des meilleures connaissance métalinguistiques).

Évidemment, on ne peut pas négliger l'influence de l'enseignement sur la maitrise des connecteurs. Cependant, nous pensons également que cette influence ne devrait être surestimée :

> « Selon Krashen (1982) l'enseignement formel a un rôle très limité dans l'ALS [acquisition langue seconde, note de l'auteur] car les connaissances apprises qui sont le résultat de l'enseignement n'aident pas les apprenant-e-s dans l'acquisition de la langue cible. Krashen affirme que l'instruction formelle contribue à l'apprentissage de connaissances explicites et n'est pas liée à l'acquisition. » (VanPatten & Benati, 2010 : 47)

Ainsi, plusieurs études ont dû constater que « l'enseignement est incapable de modifier les séquences d'acquisition » (VanPatten & Benati, 2010 : 47), qui sont dues aux facteurs linguistiques et cognitifs. Il semble plutôt que l'enseignement ne peut faciliter l'acquisition d'une langue seulement lorsqu'il coïncide avec le moment où l'apprenant-e est prêt à progresser dans son acquisition (voir aussi Pienemann, 1998).

Cela nous a fait conclure que l'impact de l'enseignement, notamment dans le cadre des études sur la connaissance implicite (chapitres 6, 7 et 8), est négligeable. Pour les contrôler dans les chapitres qui visaient la connaissance explicite, (p.ex., chapitres 8 et 9) nous avons posé aux participant-e-s des questions sur leur histoire personnelle d'apprentissage du français. Les réponses à ces questions n'étaient cependant pas prédictives des résultats obtenus.

Un autre indice que l'enseignement ne peut pas expliquer les difficultés provient du chapitre 5. On aurait pu croire que les participant-e-s testés dans ce chapitre constituent un groupe homogène : tout le monde était issu du même type de formation et environ au même stade de formation (rappelons que l'âge moyen des participant-e-s était 18 ans avec un écart type de 1.13). Toutefois, c'est

notamment ce chapitre qui a révélé des différences considérables entre les personnes non-natives.

Cependant, plus de recherche est nécessaire pour déterminer comment et à quel moment l'enseignement peut faciliter la maitrise des connecteurs en L2. Nous reviendrons à cette question à la section 1.4.

Limites concernant l'utilisation des statistiques

Chaque analyse statistique est réalisée après avoir pris de nombreuses décisions. Ces décisions ne sont pas toujours évidentes et doivent être considérées avec prudence.

Un exemple d'une telle décision est le fait d'exécuter des modèles séparés pour les analyses des temps de lecture des personnes natives et non-natives (p.ex., chapitre 6). D'une part, l'utilisation des modèles séparés risque de gonfler exagérément les différences potentielles entre les groupes. D'autre part, l'exécution d'un modèle qui inclut les deux groupes linguistiques risque de diluer les différences entre les groupes linguistiques. En effet, d'autres études ont montré qu'il existe une grande différence de temps de lecture entre une personne native et une personne non-native. Comme nous avons investigué des effets plutôt subtiles (p.ex., l'activation et intégration de la signification des connecteurs; chapitres 6, 7 et 8), nous avons opté pour des modèles différents, afin que les grandes différences entre groupes n'obscurcissent pas le rôle des autres facteurs testés.

Quant aux traitement des données, nous avons réalisé (en suivant la littérature de référence, par exemple Crible et al., 2021, Kapteijns, & Hintz, 2021) des transformations logarithmiques, ce qui représente la transformation la plus courante pour ce types des données (Nicklin & Plonsky, 2020). Toutefois, cette approche peut également être critiquée, car en analysant le temps de lecture de manière logarithmique, nous supposons que les effets du temps brut sont multiplicateurs, c'est-à-dire proportionnels à la nouvelle échelle utilisée.

Un autre exemple d'une décision statistique concerne les seuils utilisés pour exclure les personnes non-natives du groupe de contrôle ainsi que les temps de lecture trop lents ou les réponses incorrectes aux questions de vérification. En sachant que tout seuil imposé reste toujours quelque peu arbitraire, nous aimerions souligner que les décisions ont été prises avec la meilleure conviction possible, en

respectant la littérature de référence, tout en étant documentées de manière transparente.

Pour ce qui est des seuils utilisés pour exclure les temps de lecture trop lents, Nicklin et Plonsky (2020) argumentent :

> « Si les données sont distribuées de façon anormale, il faut utiliser la transformation logarithmique, car elle rend généralement la distribution acceptable pour les analyses statiques **sans avoir recours à l'élimination ou à l'altération de points de données potentiellement légitimes** » (2020:49, ma traduction, mise en évidence de l'auteur)

En d'autres termes, l'utilisation des seuils pour éliminer aveuglément des données aberrantes (c'est-à-dire les temps de lecture trop lents) n'est pas nécessaire. Plutôt, après avoir réanalysé des données d'études utilisant une tâche de lecture segmentée en L2, Nicklin et Plonsky (2021) proposent de supprimer les données aberrantes séparément après une analyse qualitative ou de transformer les données au lieu de les éliminer.

Cependant, une analyse qualitative n'a pas été possible chez nous en raison de la quantité de données. De plus, il faut noter que nos expériences ont été réalisées sur Internet, ce qui nous a obligés à être particulièrement attentifs aux éventuelles distractions des participant-e-s. Dans ce contexte, les seuils que nous avons fixés étaient généralement plutôt « généreux » (ils permettaient la lecture d'un court segment jusqu'à 3 secondes). Ainsi, nous ne pensons pas d'avoir supprimé des données potentiellement légitimes en utilisant des seuils.

Limites concernant le contrôle des facteurs

Nous avons conclu que la difficulté d'un connecteur émerge de plusieurs facteurs en interaction. Comme un connecteur a plusieurs propriétés (par exemple sa fréquence, son registre, sa transparence), la question se pose donc de savoir si nous avons pu contrôler que nous avons toujours testé le facteur qui nous intéressait. En d'autres termes, si on s'intéressait au transfert négatif (chapitre 8) ou aux usages pragmatiques (chapitre 9) n'était-il pas possible que d'autres facteurs, tels que la fréquence des connecteurs, aient influencé les résultats ?

Évidemment, un contrôle absolu de toutes les variables n'existe pas dans une approche expérimentale. Nous avons essayé cependant d'éliminer tant que possible les variables confondantes. Par exemple, nous avons veillé à ce que les connecteurs au chapitre 5 soient tous utilisés à l'écrit et monofonctionnel. Dans

d’autres cas, nous n’avons pas pu faire cela, ou alors nous avons estimé que les facteurs n’intervenaient pas de manière déterminante. Un exemple déjà discuté est le fait de tester *cependant* et *aussi* aux chapitres 4 et 7 en position initiale de la phrase.

Enfin, toute recherche doit toujours être considérée de manière critique. Il existe dans chaque travail expérimental la possibilité d'erreurs de type 1 et 2 ainsi que des variables confondantes. Il convient de rappeler que les données d’une approche expérimentale ne constituent pas automatiquement une vérité universelle et généralisable, et que le véritable apport d'un travail de recherche réside dans sa contribution au discours du domaine qui change et évolue en permanence. Ainsi, la présente thèse ne peut constituer qu’un premier pas vers l’investigation de la maitrise des connecteurs en L2. Pour terminer, nous abordons maintenant plus en détail les pistes les plus prometteuses pour de futures recherches.

10.4 Futures directions de recherches

Au cours de la présente thèse, nous avons mentionné par endroits quelques pistes pour de futures recherches. Nous allons compléter ces idées avec un aperçu plus global des directions qui sont selon nous les plus prometteuses.

Premièrement, des recherches futures pourraient analyser plus en détail les facteurs de notre liste établie dans la section 10.1.2. Par exemple, nous avons conclu que les usages pragmatiques peuvent rendre un connecteur plus difficile, mais nous avons également mentionné qu’il peut y avoir des différences quant à l’acte de langage réalisé. Il serait intéressant de déterminer dans ce contexte si certains usages pragmatiques sont plus difficiles que d’autres à comprendre et produire pour des personnes non-natives. Il se peut en outre que d’autres facteurs, tel qu’un transfert négatif, interviennent également. De même, une analyse approfondie pourrait investiguer encore d’autres facteurs, comme par exemple l’objectivité ou la subjectivité de la causalité (voir chapitre 2.5.4) où, de manière encore plus détaillée la polyfonctionnalité, pour laquelle nous n’avons pu donner qu’une réponse limitée.

Une autre direction de recherche serait d’examiner encore plus en détail les interactions entre facteurs car, de manière générale, nous n’avons que partiellement examiné cette question. Ainsi, les interactions entre d’autres facteurs

tel que la fréquence, la relation de cohérence ou le transfert restent encore à découvrir. Ce faisant, les facteurs pourraient également être classés selon leur puissance d'influence.

D'autres études pourraient chercher à déterminer comment la maitrise des connecteurs interagit avec celle des signaux alternatifs. On pourrait argumenter que les signaux alternatifs représentent des éléments relativement faciles à identifier en L2 qui, par conséquent, aideraient les lecteur-trice-s à mieux identifier les relations de cohérence. Selon nos connaissances, seule l'étude de Crible et al., (2021) a investigué la question de savoir si le signal alternatif d'un parallélisme ('Pierre *part en vacances*, alors que Marie ne *part* pas *en vacances'*) facilite la détection des relations contrastive dans une langue seconde. Comme nous l'avons mentionné au chapitre 2.4.1, il existe cependant de multiples autres signaux alternatifs, comme les titres (Christofalos et al., 2020), les mots de listes (Scholman et al., 2020) ou les verbes de causalité (Rhode & Horton, 2014), pour lesquels nous ne savons pas encore s'ils influencent la manière dont on lit un texte en L2.

Une autre direction de recherche future serait d'investiguer encore plus en détail la différence entre connaissances implicites et explicites. Ainsi, chaque facteur mentionné peut être évalué sur les deux types de connaissances, comme nous l'avons fait aux chapitres 7 et 8. Mais même au sein d'un type de connaissance, il se peut que des mesures différentes conduisent à des résultats distincts. Müller et Mari (2021) ont ainsi constaté que les résultats obtenus par des mesures de lecture segmentée ne correspondaient pas forcément aux résultats obtenus par oculométrie. Quant à la connaissance explicite, il serait également intéressant de complémenter nos observations avec d'autres mesures, comme par exemple des tâches visant le niveau de discours.

D'autres travaux encore pourraient investiguer les implications pédagogiques de nos résultats. Par exemple, nous avons observé au chapitre 5 une corrélation de l'exposition à l'écrit en L1 avec la maitrise des connecteurs en L2. Cela soulève la question de savoir s'il est possible de profiter de l'aspect interlingual du discours de sorte que l'enseignement se concentre sur la structuration d'un texte indépendamment de la langue enseignée. A cet égard, on pourrait sensibiliser les élèves aux différentes manières de marquer les relations de cohérence. Cela pourrait être effectué en plusieurs étapes, par exemple :

- Quelle est la relation de cohérence qui est communiquée ?
- Est-ce qu'elle peut être laissée implicite ?
- Si un marquage explicite est nécessaire ou avantageux, quels seraient les moyens de le réaliser ? (p.ex. en répétant les mots, avec un connecteur)
- Si un connecteur s'impose : quel connecteur est le plus adapté ?

Si on arrivait à améliorer la compétence de discours d'élèves de cette façon, cela aurait, selon nos résultats, également une bonne répercussion sur la maitrise des connecteurs en L2.

En plus de s'intéresser à la manière dont on *apprend* les connecteurs en L2, il serait intéressant de se concentrer sur leur *acquisition*. A cet égard, un paradoxe de l'acquisition d'une langue seconde est le fait que l'acquisition d'une L2 se réalise notamment à travers des connaissances explicites, alors que l'acquisition de la L1 se réalise quasi uniquement par des connaissances implicites (Bialystok & Ryan, 1985; Ellis, N., 2005; Ellis, R., 2006, aussi, Cho 2020).

Quant aux connecteurs, il serait ainsi intéressant d'investiguer à quel degré une acquisition par la connaissance implicite facilite la maitrise des connecteurs, et si leur acquisition diffère à cet égard d'autres éléments linguistiques, qu'ils soient de nature grammaticale, syntaxique, pragmatique ou lexicale.

De la même manière, il serait intéressant de comparer à quel égard l'acquisition des connecteurs en L2 ressemble à celle en L1. Comme les personnes non-natives semblent être affectées en partie par les mêmes facteurs que les personnes natives (p.ex., la complexité de la relation de cohérence, chapitre 6), on peut se demander dans quelle mesure il existe une similarité de l'acquisition des connecteurs en L1 et en L2. De manière générale, la littérature a distingué trois raisons pour lesquelles l'acquisition de la L1 diffère de l'apprentissage d'une L2 (Boers et al., 2010). Premièrement, la catégorisation des concepts du monde est déjà ancrée et liée à la L1 des apprenant-e-s. Deuxièmement, l'exposition à sa L1 est extrêmement haute et grâce à cette exposition, les jeunes apprenant-e-s de la L1 ont très fréquemment des possibilités de tester et d'imiter les modèles linguistiques qui les entourent. Et troisièmement, la transmission de la langue entre parents et enfants est forcément différente de l'activité dans des cours de langues (Boers et al., 2010). Bien que ces trois points limitent inévitablement la comparaison entre l'acquisition de la L1 et de la L2, la recherche a tout de même identifié quelques ressemblances dans l'acquisition du vocabulaire (p.ex., Xia & Wolf, 2010).

Ainsi, il est possible qu'il existe également des similitudes entre l'acquisition de la L1 et de la L2 pour les connecteurs de discours. Par exemple, nous avons nous-

mêmes observé que les personnes non-natives, tout comme les personnes natives, ont été affectées par la complexité de la relation de cohérence. Cependant, plus de recherche est nécessaire pour examiner si l'acquisition des connecteurs se poursuit effectivement selon les séquences développementales d'un enfant acquérant sa L1.

Finalement, des recherches futures pourraient également investiguer plus en détail dans quelle mesure la maitrise des connecteurs influence la compréhension d'un texte. Il est vrai qu'il a été montré que la présence des connecteurs peut faciliter la compréhension d'un texte en L2 (Degand & Sanders, 2002), mais il reste à savoir à quel égard des connecteurs compliqués (tel que *puisque*, *or* ou *cependant*) facilitent également la compréhension (voir aussi Degand et al., 1999 pour des résultats et discussion à cette question en L1).

En résumé, notre travail ouvre de nombreuses pistes différentes pour des recherches futures, qui consistent non seulement à réviser ou compléter la liste de facteurs proposés, mais aussi à tester l'application des résultats dans le cadre de l'enseignement des langues.

10.5 Le mot de la fin

La présente thèse de doctorat a montré que la maitrise souvent insuffisante des connecteurs en langue seconde varie en fonction du connecteur spécifique, de l'aptitude générale de l'apprenant-e ainsi que du type de connaissances. Son apport ne consiste pas uniquement dans la nouveauté de ces constatations, mais aussi dans le fait que ces conclusions ont été tirées par une évaluation expérimentale systématique. Notre approche expérimentale nous a permis de compléter les observations faites dans des études de corpus et mettre en évidence la signification de la variété des connecteurs pour leur maitrise en L2.

L'importance d'une telle démarche est évidente, car plus nous en savons sur la manière dont les connecteurs posent des difficultés aux apprenant-e-s, mieux nous pouvons adapter la manière dont elles sont appris et acquis. Ceci est essentiel car une bonne maitrise des connecteurs en L2 ne permet non seulement de structurer et relier les idées dans un discours, elle permet également de les nuancer et évaluer. Dans un mot : une bonne maitrise des connecteurs permet de *se faire comprendre* de manière nuancée lorsqu'il s'agit de communiquer des contenus différenciés. Il reste donc à espérer que la présente thèse puisse représenter une contribution pour surmonter le défi de rendre les apprenant-e-s de langue plus compétent-e-s dans l'utilisation des connecteurs du discours.

11 Références

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior research methods, 40*(1), 278-289.

Adeyemi, R. I. (2017). *Corpus-based analysis of cohesion in written English essays of Nigerian tertiary learners* [Doctoral dissertation, North-West University, Vaal Triangle Campus, South Africa]. https://repository.nwu.ac.za/handle/10394/24878

Afantenos, S., Asher, N., Benamara, F., Bras, M., Fabre, C., Ho-Dac, L. M., Le Draoulec, A., Muller, P., Péry-Woodley, M.-P., Prevot, L., Rebeyrolle, J., Tanguy, L., Vergez-Couret, M. & Vieu, L. (2012). An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12),* 2727-2734. European Language Resources Association (ELRA).

Agudo, J. de D. M. (Ed.) (2017). *Native and Non-Native Teachers in English Language Classrooms: Professional Challenges and Teacher Education*. Berlin : Walter de Gruyter.

AlAttar, F. H., & Abu-Ayyash, E. A. S. (2020). An investigation of the use of conjunctive cohesive devices in emirati students' argumentative essays. *British Journal of English Linguistics*, *8*, 9–25.

Aliyar, M. (2020). A corpus-based analysis of linguistic interference of Italian native speakers in English and Spanish written production as a foreign language. *Journal of Second Language Teaching & Research*, *8*(1).

Alonso, R. A. (Ed.). (2016). *Crosslinguistic influence in second language acquisition*. Bristol : Multilingual Matters.

Altenberg, B. (2006). The function of adverbial connectors in second initial position in English and Swedish. In Aijmer, K., & A. M. Simon-Vandenbergen (Eds.), *Pragmatic markers in contrast* (pp. 11-37). Brill.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, *39*(2), 187–205. https://doi.org/10.1016/0093-934X(90)90011-5

Anderson, G. (2014). So, Transitions: Linking Adverbial Use of University ESL Students. *CATESOL Journal*, *26*(1), 1–13.

Asher, N. (1993). *Reference to abstract objects in discourse* (Vol. 50). Berlin : Springer Science & Business Media.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge : University Press.

Asr, F. T., & Demberg, V. (2012a). Implicitness of discourse relations. In *Proceedings of COLING 2012* (pp. 2669-2684).

Asr, F. T., & Demberg, V. (2012b). Measuring the Strength of Linguistic Cues for Discourse Relations. In *Proceedings of the Workshop on Advances in Discourse Analysis and Its Computational Aspects* (pp.33–42). https://aclanthology.org/W12-4703

Asr, F. T., & Demberg, V. (2015). Uniform Information Density at the Level of Discourse Relations: Negation Markers and Discourse Connective Omission. In *Proceedings of the 11th International Conference on Computational Semantics* (pp.118–128). https://www.coli.uni-saarland.de/~fatemeh/iwcs2015.pdf

Austin, L. (1962). *How to Do Things with Words.* Oxford : Clarendon.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(1), 31-56

Baayen, H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge : University Press.

Bardovi-Harlig, K., & Dörnyei, Z. (1998). Do Language Learners Recognize Pragmatic Violations? Pragmatic Versus Grammatical Awareness in Instructed L2 Learning*. *TESOL Quarterly*, *32*(2), 233–259. https://doi.org/10.2307/3587583

Bardovi-Harlig, K., & Hartford, B. S. (1990). Congruence in Native and Nonnative Conversations: Status Balance in the Academic Advising Session *. *Language Learning*, *40*(4), 467–501. https://doi.org/10.1111/j.1467-1770.1990.tb00603.x

Bardovi-Harlig, K., & Hartford, B. S. (1993). Learning the Rules of Academic Talk: A Longitudinal Study of Pragmatic Change. *Studies in Second Language Acquisition*, *15*(3), 279–304. https://doi.org/10.1017/S0272263100012122

Barnett, M. A. (1988). Reading through context: How real and perceived strategy use affects L2 comprehension. *The modern language journal*, *72*(2), 150-162.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Bell, D. M. (2010). Nevertheless, still and yet: Concessive cancellative discourse markers. *Journal of Pragmatics*, *42*(7), 1912–1927. https://doi.org/10.1016/j.pragma.2009.12.010

Benzitoun, C., Debaisieux, J.-M., & Deulofeu, H.-J. (2016). Le projet ORFÉO: Un corpus d'étude pour le français contemporain. *Corpus*, *15*, Article 15. https://doi.org/10.4000/corpus.2936

Berman, M. (1979). *Advanced language practice for EFL.* London: Hodder & Stoughton.

Bialystok, E., & Ryan, E. B. (1985). Toward a definition of metalinguistic skill. *Merrill-Palmer Quarterly,* 229-251.

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97-116.

Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, *20*(3), 267-275.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford : Basil Blackwell.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Vol. 99). Cambridge : University press.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of child language*, *7*(2), 235-261.

Boers, F., De Rycker, A., & De Knop, S. (2010). Fostering language teaching efficiency through cognitive linguistics: Introduction. In S. De Knop, F. Boers, & A. De Rycker (Eds*.), Fostering Language Teaching Efficiency through Cognitive Linguistics* (pp. 1–26). De Gruyter Mouton. https://doi.org/10.1515/9783110245837.1

Bolton, K., Nelson, G., & Hung, J. (2002). A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, *7*(2), 165–182. https://doi.org/10.1075/ijcl.7.2.02bol

Borreguero Zuloaga, M. (2011). La traducción de los marcadores del discurso: Valores, funciones, posiciones y otros problemas. In *Últimas tendencias en traducción e interpretación* (pp. 123–139). https://doi.org/10.31819/9783865278777-009

Brantmeier, C. (2005). Effects of Reader's Knowledge, Text Type, and Test Type on L1 and L2 Reading Comprehension in Spanish. *The Modern Language Journal*, *89*(1), 37–53. https://doi.org/10.1111/j.0026-7902.2005.00264.x

Bras, M., Le Draoulec, A., & Asher, N. (2009). A formal analysis of the french temporal connective alors. *Oslo Studies in Language*, *1*(1).

Britton, B.K., Glynn, S.M., Meyer, B.J.F. & Penland, M.J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology 74*. 51–61.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Brysbaert, M. (2013). Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, *53*(1), Article 1.

Britton, B. K., Glynn, S. M., Meyer, B. J., & Penland, M. J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology, 74*(1), 51.

Busquets, J., Vieu, L., & Asher, N. (2001). La SDRT: Une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum (Presses Universitaires de Nancy)*, *13*(1), 73–101.

Callow, K., & Callow, J. (1992). Text as purposive Communcation: A meaning-based Analyses. In W. Mann & S. Thompson, *Discourse Description. Diverse Linguistic analyses of a fund-raising text* (pp. 5–39). Amsterdam : John Benjamins Publishing.

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. M. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*(9), 1394–1413. https://doi.org/10.1080/01690965.2012.685885

Canestrelli, A. R., Mak, P., & Sanders, T. (2016). The influence of genre on the processing of objective and subjective causal relations: Evidence from eyetracking. In *Genre in Language, Discourse and Cognition*. Berlin : Walter de Gruyter.

Caron, J., Micko, H. C., & Thüring, M. (1988). Conjunctions and the recall of composite sentences. *Journal of memory and language*, *27*(3), 309-323.

Casteele, A. V., & Collewaert, K. (2013). The Use of Discourse Markers in Spanish Language Learners' Written Compositions. *Procedia - Social and Behavioral Sciences*, *95*, 550–556. https://doi.org/10.1016/j.sbspro.2013.10.681

Cain, K., Patson, N., & Andrews, L. (2005). Age- and ability-related differences in young readers' use of conjunctions. *Journal of Child Language*, *32*, 877–892. https://doi.org/10.1017/S0305000905007014

Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429–441. https://doi.org/10.1037/a0022824

Carrió-Pastor, M. L. (2013). A contrastive study of the variation of sentence connectors in academic English. *Journal of English for Academic Purposes*, *12*(3), 192–202. https://doi.org/10.1016/j.jeap.2013.04.002

Carroll, S. E. (1992). On cognates. *Interlanguage Studies Bulletin (Utrecht)*, *8*(2), 93–119. https://doi.org/10.1177/026765839200800201

Chen, P. J. (2014). The comparison of intermediate and advanced Chinese learners' use of English adverbial connectors in academic writing. *International Journal on Studies in English Language and Literature*, *2*(8), 85-92.

Cho, H. Y., & Shin, J.-A. (2014). Cohesive Devices in English Writing Textbooks and Korean Learners' English Writings. *English Teaching*, *69*(1), 41–59. https://doi.org/10.15858/engtea.69.1.201403.41

Cho, J. (2020). Online processing and offline judgments of L2-English articles. *Linguistic Approaches to Bilingualism*, *12*(3), 280-309. https://doi.org/10.1075/lab.18053.cho

Christofalos, A., Raney, G., Daniel, F., & Demos, A. (2020). Titles support the development of coherent situation models. *Journal of Research in Reading*, *43*. https://doi.org/10.1111/1467-9817.12315

Cohen, A., Glasman, H., Rosenbaum-Cohen, P. R., Ferrara, J., & Fine, J. (1979). Reading English for Specialized Purposes: Discourse Analysis and the Use of Student Informants. *TESOL Quarterly*, *13*(4), 551–564. https://doi.org/10.2307/3586449

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL quarterly*, *33*(2), 185-209.

Cooper, M. (1984). Linguistic competence of practiced and unpracticed non-native readers of English. In J.C. Charles & A.H. Urquhart (Eds.), *Reading in a foreign language* (pp. 122138). London: Longman.

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(1), 107–126. https://doi.org/10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570. https://doi.org/10.1016/j.tics.2006.10.002

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, *94*(1), 94–103. https://doi.org/10.1016/j.bandl.2004.12.002

Crewe, W., Wright, C., & Leung, M. W. K. (1985). Connectives: On the other hand, who needs them. *Working papers in Linguistics and Language Teaching*, *8*, 61-75.

Crewe, W. J. (1990). The illogic of logical connectives. *ELT Journal*, *44*(4), 316–325. https://doi.org/10.1093/elt/44.4.316

Crible, L., & Cuenca, M.-J. (2017). Discourse Markers in Speech: Distinctive Features and Corpus Annotation. *Dialogue & Discourse*, *8*(2), Article 2.

Crible, L., & Pickering, M. J. (2020). Compensating for processing difficulty in discourse: Effect of parallelism in contrastive relations. *Discourse Processes*, *57*(10), 862-879.

Crible, L., & Demberg, V. (2020). When Do We Leave Discourse Relations Underspecified? The Effect of Formality and Relation Type. *Discours*, *26*. https://doi.org/10.4000/discours.10848

Crible, L., Wetzel, M., & Zufferey, S. (2021). Lexical and Structural Cues to Discourse Processing in First and Second Language. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.685491

Crible, L. (2021). Negation Cancels Discourse-Level Processing Differences: Evidence from Reading Times in Concession and Result Relations. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-021-09802-2

Cronbach, L. J. (1946). Response sets and test validity. *Educational and psychological measurement*, *6*(4), 475-494.

Crosson, A. C., Lesaux, N. K., & Martiniello, M. (2008). Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics*, *29*(4), 603–625. https://doi.org/10.1017/S0142716408080260

Crosson, A. C., & Lesaux, N. K. (2013). Does knowledge of connectives play a unique role in the reading comprehension of English learners and English-only students? *Journal of Research in Reading*, *36*(3), 241–260. https://doi.org/10.1111/j.1467-9817.2011.01501.x

Crothers, Edward, 1979. *Paragraph Structure Inference*. Norwood, NJ : Ablex Publishing Corporation.

Cuenca, M.-J. (2003). Two ways to reformulate: A contrastive analysis of reformulation markers. *Journal of Pragmatics*, *35*(7), 1069–1093. https://doi.org/10.1016/S0378-2166(03)00004-3

Cuenca, M.-J. (2013). The fuzzy boundaries between modal and discourse marking. In L. Degand, B. Cornillie, & P. Pietrandrea, *Discourse Markers and Modal Particles: Categorization and description*. Amsterdam : John Benjamins Publishing.

Cuenca, M. J., & Degand, L. (Eds.). (2022). *Discourse Markers in Interaction: From Production to Comprehension* (Vol. 376). Berlin : Walter de Gruyter.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of educational research*, *49*(2), 222-251.

Dahlgren, K. (1988). *Naive semantics for natural language understanding*. Boston: Kluwer Academic Publishers.

Das, D. (2014). *Signalling of Coherence Relations in Discourse* [Dissertation, Arts & Social Sciences: Department of Linguistics]. http://summit.sfu.ca/item/14446

Das, D., & Taboada, M. (2018). RST Signalling Corpus: A corpus of signals of coherence relations. *Language Resources and Evaluation*, *52*(1), 149–184. https://doi.org/10.1007/s10579-017-9383-x

Degand, L., Lefèvre, N., & Bestgen, Y. (1999). The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design*, *1*(1), 39–51. https://doi.org/10.1075/dd.1.1.06deg

Degand, L., & Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and writing*, *15*(7), 739-757.

Degand, L., & Pander Maat, H. (2003). A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. *LOT Occasional series*, *1*, 175-199.

Degand, L. (2004). Contrastive analyses, translation, and speaker involvement: The case of puisque and aangezien. *Language, culture and mind*, 1-20.

Degand, L., & Fagard, B. (2011). *Alors* between discourse and grammar: the role of syntactic position. *Functions of language, 18*(1), 29-56.

Degand, L., & Evers-Vermeul, J. (2015). Grammaticalization or pragmaticalization of discourse markers? : More than a terminological issue. *Journal of historical pragmatics, 16*(1), 59-85.

Degand, L., & Fagard, B. (2018). (Inter)subjectification des connecteurs: Le cas de car et parce que. *Linguística: Revista de Estudos Linguísticos da Universidade do Porto*, *3*(0), Article 0.

Dewaele, L., & Dewaele, J.-M. (2020). Actual and Self-Perceived Linguistic Proficiency Gains in French during Study Abroad. *Languages, 6*, 1–17. https://doi.org/10.3390/languages6010006

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important ? *Bilingualism: Language and Cognition*, *23*(1), 171-185.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2022). Formal versus informal L2 learning: How do individual differences and word-related variables influence French and English L2 vocabulary learning in Dutch-speaking children? *Studies in Second Language Acquisition*, *44*(1), 87-111.

Dipboye, R. L., & Flanagan, M. F. (1979). Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? *American Psychologist, 34*(2), 141.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*(5), 735–808. https://doi.org/10.1017/S0140525X99002186

Don, Z. M., & Sriniwass, S. (2017). Conjunctive Adjuncts in Malaysian Undergraduate ESL Essays: Frequency and Manner of Use. *Moderna Språk*, 19.

Doyle, A. E. (1982). The limitations of cohesion. *Research in the Teaching of English*, *16*(4), 390-393.

Drenhaus, H., Demberg, V., Köhne, J., & Delogu, F. (2014). Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).

Dressler, W. U. (Ed.). (1978). *Current trends in textlinguistics*. Berlin : de Gruyter.

Dufva, M., & Voeten, M. J. M. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied Psycholinguistics*, *20*(3), 329–348. https://doi.org/10.1017/S014271649900301X

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*(2), 245–258. https://doi.org/10.1111/j.1467-1770.1973.tb00659.x

Ellis, N. C. (2008). Implicit and explicit knowledge about language. *Encyclopedia of language and education, 6*, 1-13.

Ellis, R. (1993). Second language acquisition and the structural syllabus. *TESOL Quarterly, 27*, 91–113.

Ellis, R. (2004). The Definition and Measurement of L2 Explicit Knowledge. *Language Learning*, *54*(2), 227–275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A Psychometric Study. *Studies in Second Language Acquisition*, *27*(2), 141–172.

Ellis, R. (2006). Modelling Learning Difficulty and Second Language Proficiency: The Differential Contributions of Implicit and Explicit Knowledge. *Applied Linguistics*, *27*(3), 431–463. https://doi.org/10.1093/applin/aml022

Ellis, R. (2009a). 1. Implicit and Explicit Learning, Knowledge and Instruction. In *Implicit and Explicit Learning, Knowledge and Instruction* (pp. 3–26). Multilingual Matters. https://doi.org/10.21832/9781847691767-003

Ellis, R. (2009b). 2. Measuring Implicit and Explicit Knowledge of a Second Language. In *Measuring Implicit and Explicit Knowledge of a Second Language* (pp. 31–64). Multilingual Matters. https://doi.org/10.21832/9781847691767-004

Essen, U. (1987). *Analyzing Intercultural Communication*. Berlin : Walter de Gruyter.

Evers-Vermeul, J., & Sanders, T. (2011). Discovering domains – On the acquisition of causal connectives. *Journal of Pragmatics*, *43*(6), 1645–1662. https://doi.org/10.1016/j.pragma.2010.11.015

Evers-Vermeul, J., & Sanders, T. J. M. (2009). The emergence of Dutch connectives: How cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language*, *36*(4), 829–854. https://doi.org/10.1017/S0305000908009227

Evers-Vermeul, J., Hoek, J., & Scholman, M. C. (2017). On temporality in discourse annotation: Theoretical and practical considerations. *Dialogue & Discourse*, *8*(2), 1-20.

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1-20.

Fagard, B., & Degand, L. (2008). La fortune des mots: Grandeur et décadence de "car." *Congrès Mondial de Linguistique Française 2008*, 036. https://doi.org/10.1051/cmlf08213

Feller, D. P., Magliano, J., Sabatini, J., O'Reilly, T., & Kopatich, R. D. (2020). Relations between component reading skills, inferences, and comprehension performance in community college readers. *Discourse Processes*, *57*(5-6), 473-490.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453–489. https://doi.org/10.1017/S0142716403000237

Field, Y., & Yip, O. Y. L. M. (1992). A Comparison of Internal Conjunctive Cohesion in the English Essay Writing of Cantonese Speakers and Native Speakers of English. *RELC Journal*, *23*(1), 15–28. https://doi.org/10.1177/003368829202300102

Fillmore, C. J. (1975, September). An alternative to checklist theories of meaning. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 1, pp. 123-131).

Ferre, G. (2011). Multimodal Analysis of Discourse Markers "donc", "alors" and "en fait" in Conversational French. *ICPhS VXII*, 671–674. https://hal.archives-ouvertes.fr/hal-00617521

Fox, J., & Weisberg, S. (2018). *An R Companion to Applied Regression*. Newbury Park : SAGE Publications.

Fraser, B. (1998). Contrastive Discourse Markers in English. In A. H. Jucker & Y. Ziv (Eds.), *Discourse Markers. Description and Theory* (pp. 301–326). John Benjamins Publishing.

Gamut, L. T. F. (1991). *Logic, Language, and Meaning, Volume 1: Introduction to Logic.* Chicago : University of Chicago Press.

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second-language acquisition. *Canadian Journal of Psychology / Revue canadienne de psychologie*, *13*(4), 266.

Gao, X. (2016). A cross-disciplinary corpus-based study on English and Chinese native speakers' use of linking adverbials in academic writing. *Journal of English for Academic Purposes*, *24*, 14–28. https://doi.org/10.1016/j.jeap.2016.08.002

Geva, E. (1986). Reading Comprehension in a Second Language: The Role of Conjunctions. *TESL Canada Journal*, 85–96. https://doi.org/10.18806/tesl.v3i0.996

Geva, E. (1992). The Role of Conjunctions in L2 Text Comprehension. *TESOL Quarterly*, *26*(4), 731–747. https://doi.org/10.2307/3586871

Geva, E., & Ryan, E. B. (1993). Linguistic and Cognitive Correlates of Academic Skills in First and Second Languages. *Language Learning*, *43*(1), 5–42. https://doi.org/10.1111/j.1467-1770.1993.tb00171.x

Goldsmith E. (1982). *An investigation of the role of adversative connectives in helping good and poor readers to integrate information in text*. Research Report. (ERIC Document Reproduction Service No. ED 247 525).

Goldman, S., & Murray, J. (1992). Knowledge of Connectors as Cohesion Devices in Text: A Comparative Study of Native-English and English-as-a-Second-Language Speakers. *Journal of Educational Psychology*, *84*(4), 504–519.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and cognitive processes*, *25*(2), 149-188.

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text. *Rethinking reading comprehension*, 82, 98.

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, *15*(1), 17–27. https://doi.org/10.1111/j.1467-971X.1996.tb00089.x

Groenendijk, J., & Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and philosophy*, 39-100.

Grolig, L., Tiffin-Richards, S. P., & Schroeder, S. (2020). Print exposure across the reading life span. *Reading and Writing*. https://doi.org/10.1007/s11145-019-10014-3

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.

Groupe λ-l (1975). Car, parce que, puisque. *Revue Romane*, *2*. https://tidsskrift.dk/revue_romane/article/view/29139

Grice, H. P. (1975). Logic and Conversation. *Speech Acts*, 41–58. https://doi.org/10.1163/9789004368811_003

Grimes, J. E. (1975). *The thread of discourse*. Paris : Mouton.

Ha, M. (2016). Linking Adverbials in First-Year Korean University EFL Learners' Writing: A Corpus-Informed Analysis. *Computer Assisted Language Learning: An International Journal*, *29*(6), 1090–1101. https://doi.org/10.1080/09588221.2015.1068814

Haberlandt, K. (1982). Reader expectations in text comprehension. In *Advances in Psychology* (Vol. 9, pp. 239-249). North-Holland.

Haberlandt, K., & Bingham, G. (1978). Verbs contribute to the coherence of brief narratives: Reading related and unrelated sentence triples. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 419–425. https://doi.org/10.1016/S0022-5371(78)90247-5

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(2), 123–141. https://doi.org/10.1017/S1366728901000232

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London : Longman.

Hamed, M. (2014). Conjunctions in Argumentative Writing of Libyan Tertiary Students. *English Language Teaching*, *7*(3), p108. https://doi.org/10.5539/elt.v7n3p108

Hartnett. (1986). Static and dynamic cohesion: Signals of thinking in writing. Functional Approaches to Writing. *Research Perspectives*. London: Frances Pinter.

Hasselgård, H. (2019). Phraseological teddy bears: Frequent lexical bundles in academic writing by Norwegian learners and native speakers of English. In V. Wiegand & M. Mahlberg (Eds.), *Corpus Linguistics, Context and Culture* (pp. 339–362). Berlin : De Gruyter. https://doi.org/10.1515/9783110489071-013

Heim, I.: 1982, *The Semantics of Definite and Indefinite Noun Phrases* [Dissertation, University of Massachusetts, Amherst].

Heim, I.: 1983, 'File Change Semantics and the Familiarity Theory of Definiteness', in R. Baüerle, Ch. Schwarze, and A. von Stechow (Eds.), *Meaning, Use and Interpretation of Language*, De Gruyter, Berlin.

Hirst, D. (1989). Review of Sperber & Wilson 1986. Relevance: Communication and Cognition. *Mind and Language*, *4*(1/2), 138-146.

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive science*, *3*(1), 67-90. https://doi.org/10.1207/s15516709cog0301_4

Hobbs J.R. (1985). *On the Coherence and Structure of Discourse 5*(37). Stanford *:* Center for Study of Language and Information (CSLI).

Hobbs, J. R. (1990). *Literature and cognition* (No. 21). Stanford : Center for the Study of Language (CSLI).

Hoek, J., Evers-Vermeul, J., & Sanders, T. J. M. (2019). Using the Cognitive Approach to Coherence Relations for Discourse Annotation. *Dialogue & Discourse*, *10*(2), Article 2. https://doi.org/10.5087/dad.2019.201

Holtgraves, T. (1997). Styles of Language Use: Individual and Cultural Variability in Conversational Indirectness. *Journal of Personality and Social Psychology*, *73*, 624–637. https://doi.org/10.1037/0022-3514.73.3.624

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press. http://lup.lub.lu.se/record/1852359

Holtgraves, T. (1997). Styles of language use: Individual and cultural variability in conversational indirectness. *Journal of personality and social psychology*, *73*(3), 624.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. https://doi.org/10.1016/j.lingua.2009.06.004

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *50*(3), 346-363.

Huddleston, R. (1978). On classifying anaphoric relations: A review of MAK Halliday and R. Hasan, Cohesion in English. *Lingua*, *45*(3-4), 333-354.

Jackendoff, R. S. (1975). Semantic Interpretation in Generative Grammar. *Foundations of Language, 12*(4).

Jacques, M.-P., & Rinck, F. (2017). Un corpus de “littéracie avancée” : Résultat et point de départ. *Corpus*, *16*, 217-237. https://journals.openedition.org/corpus/2806

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013, Juillet). The TenTen Corpus Family. *In 7th international corpus linguistics conference CL* (pp. 125-127). Lancaster University. https://is.muni.cz/publication/1120431/en/The-TenTen-Corpus-Family/Jakubicek-Kilgarriff-Kovar-Rychly

Janke, V., & Kolokonte, M. (2015). False cognates: The effect of mismatch in morphological complexity on a backward lexical translation task. *Second Language Research*, *31*(2), 137–156.

Jiang, N. (2011). *Conducting Reaction Time Research in Second Language Studies.* London : Routledge. https://doi.org/10.4324/9780203146255

Johnson-Laird, P. N. (1988). *The computer and the mind: An introduction to cognitive science*. Harvard : University Press.

Johnson, E. D., & Arnold, J. E. (2022). The frequency of referential patterns guides pronoun comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0001137

Johnston, P., & Pearson, D. P. (1982). *Prior knowledge, connectivity, and the assessment of reading comprehension* (Tech. Rep. No. 245.). Urbana: University of Illinois, Center for the Study of Reading. (ERIC Document Reproduction Service No. ED 247 525)

Juffs, A. (1996). Semantics-syntax correspondences in second language acquisition. *Second Language Research*, *12*(2), 177–221.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. https://doi.org/10.1037/0096-3445.111.2.228

Kamalski, J., Lentz, L., Sanders, T., & Zwaan, R. A. (2008). The forewarning effect of coherence markers in persuasive discourse: Evidence from persuasion and processing. *Discourse Processes, 45*(6), 545-579.

Kamp, H. (1981). Événements, représentations discursives et référence temporelle. *Langages (64)*, 39-64.

Kamp, H., Van Genabith, J., & Reyle, U. (2011). Discourse Representation Theory. In D. M. Gabbay & F. Guenthner (Eds.), *Handbook of Philosophical Logic: Volume 15* (pp. 125–394). Amsterdam : Springer Netherlands. https://doi.org/10.1007/978-94-007-0485-5_3

Kanno, Y. (1989). The Use of Connectives in English Academic Papers Written by Japanese Students. *MITA Working Papers in Psycholinguistics, Volume 2*. https://eric.ed.gov/?id=ED358721

Kapteijns, B., & Hintz, F. (2021). Comparing predictors of sentence self-paced reading times: Syntactic complexity versus transitional probability metrics. *Plos one, 16*(7), e0254546.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2019). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, 1–6. https://doi.org/10.1017/S1366728919000038

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 115–126. https://doi.org/10.1016/S0022-5371(84)90082-3

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of semantics*, *25*(1), 1-44.

Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A Cognitive View of Reading Comprehension: Implications for Reading Difficulties. *Learning Disabilities Research & Practice*, *29*. https://doi.org/10.1111/ldrp.12025

Kerbrat-Orecchioni, C. (1991). La politesse dans les interactions verbales. *Dialoganalyse III*, *1*, 39-59.

Khemlani, S. S., Byrne, R. M., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive science*, *42*(6), 1887-1924.

Kiany, G. R., & Nejad, M. K. (2001). On the Relationship between English Proficiency, Writing Ability, and the Use of Conjunctions in Iranian EFL Learners' Compositions. *ITL - International Journal of Applied Linguistics*, *133*(1), 227–239. https://doi.org/10.1075/itl.133-134.03kia

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*(1), 7-36.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, *85*(5), 363.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163.

Kleijn, S., Mak, W. M., & Sanders, T. J. M. (2020). Causality, subjectivity and mental spaces: Insights from on-line discourse processing. *Cognitive Linguistics*, *32*(1), 35–65. https://doi.org/10.1515/cog-2018-0020

Knott, A., & Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, *30*(2), 135–175. https://doi.org/10.1016/S0378-2166(98)00023-X

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of machine translation summit x: papers* (pp. 79-86).

Köhne, J., & Demberg, V. (2013). The time-course of processing discourse connectives. In *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 35, No. 35).*

Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes, 28*(8), 1169–1206. https://doi.org/10.1080/01690965.2012.699076

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*, *62*, 621.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1-26.

Laali, M., & Kosseim, L. (2017, August). Automatic Mapping of French Discourse Connectives to PDTB Discourse Relations. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue* (pp. 1-6).

Lambrecht, Knud, Bordeaux, Julia, Reichle, Robert, (2006). Cognitive constraints on assertion scope: the case of spoken French parce que. In: Nishida, Chiyo, Montreuil, Jean-Pierre (Eds.), *New Perspectives on Romance Linguistics: Volume I: Morphology, Syntax, Semantics, and Pragmatics. Selected Papers from the 35th Linguistic Symposium on Romance Languages*, 143-154. Amsterdam: John Benjamin.

Lamillar, F. J. S., & Godart-Wendling, B. (2002). Grammaire de Montague, catégories et types: Une présentation des théories actuelles en sémantique et en interprétation du discours. *Langages*, 76-92.

Lamiroy, B. (1994). Pragmatic connectives and L2 acquisition: The case of French and Dutch. *Journal of Pragmatics*, *4*(2), 183–201.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Laparle, S. (2021). Tracking Discourse Topics in Co-speech Gesture. In *International Conference on Human-Computer Interaction* (pp. 233-249). Cham : Springer.

Leal, T. (2018). Data analysis and sampling: Methodological issues concerning proficiency in SLA research. In *Critical Reflections on Data in Second Language Acquisition* (pp. 63-88). Amsterdam : John Benjamins.

Lee, K. (2013). Korean ESL Learners' Use of Connectors in English Academic Writing. *English Language Teaching*, *25*, 81–103. https://doi.org/10.17936/pkelt.2013.25.2.005

Leech, G., & Svartvik, J. (2013). *A communicative grammar of English*. Oxfordshire : Routledge.

Leedham, M., & Cai, G. (2013). Besides … on the other hand: Using a corpus approach to explore the influence of teaching materials on Chinese students' use of linking adverbials. *Journal of Second Language Writing*, *22*(4), 374–389. https://doi.org/10.1016/j.jslw.2013.07.002

Leeman, D. (1994). Puisque et sur car. *Théories, données et pratiques en français langue étrangère, 3*(113).

Lehman, S., & Schraw, G. (2002). Effects of coherence and relevance on shallow and deep text processing. *Journal of Educational Psychology*, *94*(4), 738–750. https://doi.org/10.1037/0022-0663.94.4.738

Lenth. (2020). Package 'lsmeans'. *The American Statistician, 34*(4)*,* 216-22.

Lei, L. (2012). Linking adverbials in academic writing on applied linguistics by Chinese doctoral students. *Journal of English for Academic Purposes*, *11*(3), 267–275. https://doi.org/10.1016/j.jeap.2012.05.003

Levy, R., & Jaeger, T. F. (2006). Speakers optimize information density through syntactic reduction. In *Advances in Neural Information Processing Systems* (Vol. 19, p. 856).

Li, C., Dewaele, J.-M., & Hu, Y. (2021). Foreign language learning boredom: Conceptualization and measurement. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2020-0124

Li, M., & Kirby, J. R. (2015). The Effects of Vocabulary Breadth and Depth on English Reading. *Applied Linguistics*, *36*(5), 611–634. https://doi.org/10.1093/applin/amu007

Longacre, R. E. (1983). Switch-reference systems in two distinct linguistic areas: Wojokeso (Papua New Guinea) and Guanano (Northern South America). *Switch-reference and universal grammar*, 185-208.

Longacre, R. E. (1979). The paragraph as a grammatical unit. In *Discourse and syntax* (pp. 113-134). Brill.

Lorch Jr, R. F., & Lorch, E. P. (1986). On-line processing of summary and importance signals in reading. *Discourse processes*, *9*(4), 489-496.

Lyu, S., Tu, J.-Y., & Lin, C.-J. C. (2020). Processing Plausibility in Concessive and Causal Relations: Evidence from Self-Paced Reading and Eye-Tracking. *Discourse Processes*, *57*(4), 320–342. https://doi.org/10.1080/0163853X.2019.1680089

Maat, H. P. (1998). Classifying negative coherence relations on the basis of linguistic evidence. *Journal of Pragmatics*, *30*(2), 177-204.

Mahmoodi, M. H., & Yousefi, M. (2022). Second language motivation research 2010–2019: A synthetic exploration. *The Language Learning Journal*, *50*(3), 273-296.

Mann, W. C., & Thompson, S. A. (1986). Relational propositions in discourse. Discourse processes, 9(1), 57-90.

Mann, W. C., & Thompson, S. A. (1987). Rhetorical structure theory: Description and construction of text structures. In *Natural language generation* (pp. 85-95). Dordrecht : Springer.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, *8*(3), 243-281.

Mann, W. C., & Thompson, S. A. (Eds.). (1992). *Discourse description: Diverse linguistic analyses of a fund-raising text* (Vol. 16). Amsterdam : John Benjamins Publishing.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language proceissng. *Studies in Second Language Acquisition*, *27*(1), 53–78.

Marshall, N., & Glock, M. D. (1978). Comprehension of connected discourse: A study into the relationships between the structure of text and information recalled. *Reading Research Quarterly*, 10-56.

Martin-Chang, S. L., & Gould, O. N. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, *31*(3), 273-284.

Mauranen, A. (1996). Discourse competence-Evidence from thematic development in native and non-native texts. *Pragmatics and beyond new series*, 195-230.

McClure, E., & Geva, E. (1983). The development of the cohesive use of adversative conjunctions in discourse. *Discourse Processes*, *6*(4), 411–432. https://doi.org/10.1080/01638538309544575

McKeown, K. (1982, June). The TEXT system for natural language generation: An overview. In *20th Annual Meeting of the Association for Computational Linguistics* (pp. 113-120).

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, *14*(1), 1-43.

Merilaine, E. (2015) "The frequency and variability of conjunctive adjuncts in the Estonian–English Interlanguage Corpus." [MA dissertation, University of Tartu].

Meyer, B. J. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Reading Behavior*, *7*(1), 7-47.

Meyer, B.J., Brandt, D.M. & Bluth, G.J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly 16*, 72–103.

Meyer, B.J. (1992). An Analyses of a Plea for Moneay. In W. Mann & S. Thompson, *Discourse Description. Diverse Linguistic analyses of a fund-raising text* (pp. 97–108). Amsterdam : John Benjamins Publishing.

Milton, J. C., & Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In *Studies in lexis. Proceedings of a seminar on lexis organized by the Language Centre of the HKUST, Hong Kong (Language Centre, HKUST, Hong Kong, 1993)*.

Millis, K. K., Graesser, A. C., & Haberlandt, K. (1993). The impact of connectives on the memory for expository texts. *Applied Cognitive Psychology*, *7*(4), 317-339.

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of memory and language*, *33*(1), 128-147.

Mohamed, N. (2016). Use of conjunctions in argumentative essay by ESL undergraduates. *E-Academia Journal*, *5*(1).

Moeschler, J., & Reboul, A. (1994). *Dictionnaire encyclopédique de pragmatique*. Paris : Seuil. https://archive-ouverte.unige.ch/unige:110321

Moeschler, J. (1998). *Le temps des événements: Pragmatique de la référence temporelle*. Paris : Éditions Kimé.

Moeschler, J. (2005). Connecteurs pragmatiques, inférences directionnelles et représentations mentales. *Cahiers Chronos, 12*, 35-50.

Montague, R. (1970). Pragmatics and intensional logic. *Synthese*, *22*(1), 68-94.

Montague, R. (1974). *Formal philosophy: Selected papers*. New Haven: Yale University Press.

Moore, J. D., & Pollack, M. E. (1992). A problem for RST: The need for multi-level discourse analysis. *Computational linguistics*, *18*(4), 537-544.

Morera, Y., León, J. A., Escudero, I., & Vega, M. de. (2017). Do Causal and Concessive Connectives Guide Emotional Expectancies in Comprehension? A Double-Task Paradigm Using Emotional Icons. *Discourse Processes*, *54*(8), 583–598. https://doi.org/10.1080/0163853X.2015.1137445

Morgenthaler, S. (2009). Exploratory data analysis. *WIREs Computational Statistics*, *1*(1), 33–44. https://doi.org/10.1002/wics.2

Mosenthal, J. H., & Tierney, R. J. (1984). Commentary: Cohesion: Problems with Talking about Text. *Reading research quarterly*, *19*(2), 240-244.

Müller, M.-L., & Mari, M. A. (2021). Definite Descriptions in the Light of the Comprehension vs. Acceptance Distinction: Comparing Self-Paced Reading with Eye-Tracking Measures. *Frontiers in Communication, 6,* 77. https://doi.org/10.3389/fcomm.2021.634362

Murray, J. D. (1994). Logical connectives and local coherence. In R. F. Lorch & E. 1. O'Brien (Eds.), *Sources of cohesion in text comprehension* (pp. 107-125). Hillsdale, NJ: Lawrence Erlbaum Associates.

Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, *25*(2), 227–236. https://doi.org/10.3758/BF03201114

Nicklin, C., & Plonsky, L. (2020). Outliers in L2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics, 40*, 26-55.

Nippold, M., Schwarz, I., & Undlin, R. (1992). Use and Understanding of Adverbial Conjuncts. *Journal of Speech, Language, and Hearing Research*, *35*(1), 108–118. https://doi.org/10.1044/jshr.3501.108

Nippold, M. A., Taylor, C. L., & Baker, J. M. (1996). Idiom Understanding in Australian Youth: A Cross-Cultural Comparison. *Journal of Speech, Language, and Hearing Research*, *39*(2), 442–447. https://doi.org/10.1044/jshr.3902.442

Nishimoto, Y. (1997). Teaching the function of English connectives for expository text comprehension. *全国英語教育学会紀要* *8*, 103–115. https://doi.org/10.20581/arele.8.0_103

Oberg, A., & Daniels, P. (2013). Analysis of the effect a student-centred mobile learning instructional method has on language acquisition. *Computer assisted language learning*, *26*(2), 177-196.

Orfitelli, R., & Polinsky, M. (2017). When performance masquerades as comprehension: Grammaticality judgments in experiments with non-native speakers. In *Quantitative Approaches to the Russian Language* (pp. 197–214). https://doi.org/10.4324/9781315105048

Ozono, S., & Ito, H. (2003). Logical connectives as catalysts for interactive L2 reading. *System*, *31*(2), 283–297. https://doi.org/10.1016/S0346-251X(03)00025-3

Palan, S., & Schitter, C. (2018). Prolific. ac — A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22-27.

Pander Maat, H. (1998). Classifying negative coherence relations on the basis of linguistic evidence. *Journal of Pragmatics*, *30*(2), 177–204. https://doi.org/10.1016/S0378-2166(98)00024-1

Pander Maat, H., & Degand, L. (2002). *Scaling causal relations and connectives in terms of speaker involvement*. *12*(3), 211–245. https://doi.org/10.1515/cogl.2002.002

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing : A Study of Relative Clause Attachment in Greek. *Studies in Second Language Acquisition*, *25*(4), 501–528.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implication for bilingualism and second language acquisition. In N., Ellis (Ed), *Implicit and Explicit Learning of Languages*. London : Academic Press https://ci.nii.ac.jp/naid/10016108580/

Partee, B. H., & Hendriks, H. L. (1997). Montague grammar. In van Benthem, J., & A., Ter Meulen (Eds.), *Handbook of logic and language* (pp. 5-91). North-Holland.

Park, Y.-Y. (2013). How Korean EFL Students Use Conjunctive Adverbials in Argumentative Writing. *English Teaching*, *68*(4), 263–284. https://doi.org/10.15858/engtea.68.4.201312.263

Petchprasert, A. (2013). A study of cohesive markers used in L1 and L2 essay writing: Translation versus direct composition. *3L, Language, Linguistics, Literature*, *19*(1).

Pienemann, M. (1998). *Language processing and L2 development*. Amsterdam: John Benjamins Publishing.

Pienemann, M. (2015). An outline of processability theory and its relationship to other approaches to SLA. *Language Learning, 65*(1), 123-151.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*(5-6), 601-638.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008, Mai). The Penn Discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pp. 2961-2968, European Language Resources Association (ELRA).

Pretorius, E. J. (2006). The comprehension of logical relations in expository texts by students who study through the medium of ESL. *System*, *34*(3), 432–450. https://doi.org/10.1016/j.system.2006.02.003

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martín-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). *Linguistic Approaches to Bilingualism*, 2023.

Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology*, *48*(2), 521.

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL : https://www.R-project.org/

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Recio Fernández, I. M. (2020). *The Impact of Procedural Meaning on Second Language Processing: A Study on Connectives* [Dissertation, Heidelberg University]. https://doi.org/10.11588/heidok.00028641

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 Discourse: L1 Influence and General Learner Effects. *Studies in Second Language Acquisition*, *30*(3), 333–357.

Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition*, *133*(3), 667-691.

Roze, C., Danlos, L., & Muller, P. (2012). LEXCONN: A French Lexicon of Discourse Connectives. *Discours. Revue de Linguistique, Psycholinguistique et Informatique. A Journal of Linguistics, Psycholinguistics and Computational Linguistics*, *10*(10). https://doi.org/10.4000/discours.8645

Sáchez-Casas, R. M., García-Albea, J. E., & Davis, C. W. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology*, *4*(4), 293-310.

Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35. https://doi.org/10.1080/01638539209544800

Sanders, T. J., Spooren, W. P., & Noordman, L. G. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics, 4*(2).

Sanders, T. J. M., & Noordman, L. G. M. (2000). The Role of Coherence Relations and Their Linguistic Markers in Text Processing. *Discourse Processes*, *29*(1), 37–60. https://doi.org/10.1207/S15326950dp2901_3

Sanders, T., & Spooren, W. (2001). Text representation as an interface between language and its users. In Sanders, T., Schilperoord, J., and W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 1–25). Amsterdam : John Benjamins Publishing. https://doi.org/10.1075/hcp.8.02san

Sanders, T. (2005). Coherence, causality and cognitive complexity in discourse. In *Proceedings/Actes SEM-05, First International Symposium on the exploration and modelling of meaning* (pp. 105-114). Toulouse: University of Toulouse-le-Mirail.

Sanders, T. J., & Spooren, W. P. (2015). Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interactions and written text. *Linguistics*, *53*(1), 53-92.

Sanders, T. J. M., Demberg, V., Hoek, J., Scholman, M. C. J., Asr, F. T., Zufferey, S., & Evers-Vermeul, J. (2018). Unifying dimensions in coherence relations: How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory*, *1.* https://doi.org/10.1515/cllt-2016-0078

Santana, A., Spooren, W., Nieuwenhuijsen, D., & Sanders, T. J. M. (2021). Do Spanish causal connectives vary in subjectivity? What crowdsourcing data reveal about native speakers' preferences. *Text & Talk 41*(2), 211-237.https://doi.org/10.1515/text-2019-0102

Scholman, M. C. J., Rohde, H., & Demberg, V. (2017). "On the one hand" as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, *97*, 47–60. https://doi.org/10.1016/j.jml.2017.07.010

Scholman, M. C., & Demberg, V. (2017a). Examples and specifications that prove a point: Identifying elaborative and argumentative discourse relations. *Dialogue & Discourse*, *8*(2), 56-83.

Scholman, M., & Demberg, V. (2017b). Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. *Proceedings of the 11th Linguistic Annotation Workshop*, 24–33. https://doi.org/10.18653/v1/W17-0803

Scholman, M. C. J., Demberg, V., & Sanders, T. J. M. (2020). Individual differences in expecting coherence relations: Exploring the variability in sensitivity to contextual signals in discourse. *Discourse Processes*, 1–18. https://doi.org/10.1080/0163853X.2020.1813492

Schreiber-Gregory, D. N. (2018). Ridge Regression and multicollinearity: An in-depth review. *Model Assisted Statistics and Applications*, *13*(4), 359–365. https://doi.org/10.3233/MAS-180446

Schumann, J., Zufferey, S., & Oswald, S. (2020). The Linguistic Formulation of Fallacies Matters: The Case of Causal Connectives. *Argumentation*, *35*(3), 361-388. https://doi.org/10.1007/s10503-020-09540-0

Schiffrin, D. (1987). *Discourse markers* (No. 5). Cambridge : University Press.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Segal, E. M., Duchan, J. F., & Scott, P. J. (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. *Discourse Processes*, *14*(1), 27–54. https://doi.org/10.1080/01638539109544773

Segalowitz, N. (2000). Automaticity and attentional skill in fluent In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor: University of Michigan Press. https://dialnet.unirioja.es/servlet/articulo?codigo=5732214

Segalowitz, N. (2001). On the evolving connections between psychology and linguistics. *Annual Review of Applied Linguistics*, *21*, 3–22. https://doi.org/10.1017/S0267190501000010

Segalowitz, N. S., & Segalowitz, S. J. (2008). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385. https://doi.org/10.1017/S0142716400010845

Seuren, P. (1986). *Discourse Semantics*. Oxford : Blackwell.

Seuren, P. A., & Wekker, H. (1986). Semantic Transparency as a Factor in Creole Genesis. In *Substrata versus Universals in Creole Genesis: Papers from the Amsterdam Creole Workshop, April 1985* (Vol. 1, p. 57). Amsterdam : John Benjamins Publishing.

Shi, J. (2017). A Corpus-Based Study of Contrastive/Concessive Linking Adverbials in Spoken English of Chinese EFL Learners. *Studies in Literature and Language*, *14*(2), Article 2. https://doi.org/10.3968/9273

Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. *Complexity, accuracy, and fluency in second language use, learning, and teaching*, *207*, 226.

Smith, M. S., & Kellerman, E. (1986). Crosslinguistic influence in second language acquisition: An introduction. In M. S., Smith & E., Kellerman (Eds), *Crosslinguistic Influence in Second Language Acquisition.* New York: Pergamon.

Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic approaches to bilingualism*, *1*(1), 1-33.

Song, L. (2010). The Role of Context in Discourse Analysis. *Journal of Language Teaching and Research*, *1*(6), 876–879. https://doi.org/10.4304/jltr.1.6.876-879

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2012). Do L1 Reading Achievement and L1 Print Exposure Contribute to the Prediction of L2 Proficiency? *Language Learning*, *62*(2), 473–505. https://doi.org/10.1111/j.1467-9922.2012.00694.x

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (2nd ed). Hoboken, New Jersey : Blackwell Publishers.

Sperber, D., & Wilson, D. (1987). Précis of Relevance: Communication and Cognition. *Behavioral and Brain Sciences*, *10*(04), 697. https://doi.org/10.1017/S0140525X00055345

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Spooren, W., & Sanders, T. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of pragmatics*, *40*(12), 2003-2026.

Spyridakis, J. H., & Standal, T. C. (1987). Signals in expository prose: Effects on reading comprehension. *Reading Research Quarterly*, 285-298.

Staub, A., & Clifton Jr, C. (2006). Syntactic prediction in language comprehension: evidence from either... or. *Journal of experimental psychology: Learning, memory, and cognition*, *32*(2), 425.

Stanovich, K. E., & West, R. F. (1989). Exposure to Print and Orthographic Processing. *Reading Research Quarterly*, *24*(4), 402. https://doi.org/10.2307/747605

Stede, M., & Umbach, C. (1998). *DiMLex: A lexicon of discourse markers for text generation and understanding*. 1238–1242. https://doi.org/10.3115/980691.980771

Steffani, S. A., & Nippold, M. A. (1997). Japanese speakers of American English: Competence with connectives in written language. *Journal of Speech, Language, and Hearing Research : JSLHR*, *40*(5), 1048–1055. https://doi.org/10.1044/jslhr.4005.1048

Sterpin, L. F., Ortiz, S. S., Formoso, J., & Barreyro, J. P. (2021). The role of vocabulary knowledge on inference generation: A meta-analysis. *Psychology of Language and Communication*, *25*(1), 168–193. https://doi.org/10.2478/plc-2021-0008

Stukker, N., & Sanders, T. (2012). Subjectivity and prototype structure in causal connectives: A cross-linguistic perspective. *Journal of pragmatics*, *44*(2), 169-190.

Stukker, N., Sanders, T., & Verhagen, A. (2008). Causality in verbs and in discourse connectives: Converging evidence of cross-level parallels in Dutch linguistic categorization. *Journal of Pragmatics*, *40*(7), 1296-1322.

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure* (Vol. 40). Cambridge : University Press.

Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, *38*(4), 567–592. https://doi.org/10.1016/j.pragma.2005.09.010

Tapper, M. (2005). Connectives in advanced Swedish EFL learners' written English–preliminary results. *The Department of English: Working Papers in English Linguistics*, *5*, 116-144.

Taguchi, N. (2008). The effect of working memory, semantic access, and listening abilities on the comprehension of conversational implicatures in L2 English. *Pragmatics & Cognition*, *16*(3), 517–539. https://doi.org/10.1075/pc.16.3.05tag

Tavakoli, P. (2019). Chapter 2. Automaticity, fluency and second language task performance. In Z. (Edward) Wen & M. J. Ahmadian (Eds.), *Task-Based Language Teaching* (Vol. 13, pp. 39–52). Amsterdam : John Benjamins Publishing. https://doi.org/10.1075/tblt.13.03tav

Tazegül, A. B. (2015). Use, misuse and overuse of 'on the other hand': A corpus study comparing English of native speakers and learners. *International Online Journal of Education and Teaching (IOJET)*, *2*(2), 53-66.

Toulmin S. (1985). *The uses of Argument*. Cambridge : University Press.

Trabbasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In Mandl, H & Stein, & T. Trabasso (Eds), *Learning and Comprehension of Text* (pp83-109). Hillsdale, NJ : Lawrence Erlbaum Associates.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, *24*(5), 595–611. https://doi.org/10.1016/0749-596X(85)90048-8

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of Connectives on Language Comprehension: Eye tracking Evidence for Incremental Interpretation. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 481–497. https://doi.org/10.1080/027249897391982

Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 88.

Tsehaye, W., Pashkova, T., Tracy, R., & Allen, S. (2021). Deconstructing the Native Speaker: Further Evidence From Heritage Speakers for Why This Horse Should Be Dead! *Frontiers in Psychology*, *12*, 717352. https://doi.org/10.3389/fpsyg.2021.717352

Tskhovrebova, E., Zufferey, S., & Gygax, P. (2022). Individual Variations in the Mastery of Discourse Connectives from Teenage Years to Adulthood. *Language Learning*, *72*(2), 412–455. https://doi.org/10.1111/lang.12481

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160). London : Pearson.

Tyler, J. (2014). Prosody and the interpretation of hierarchically ambiguous discourse. *Discourse Processes*, *51*(8), 656-687.

Ucar, S., & Yukselir, C. (2017). A Corpus-Based Study on the Use of the Logical Connector 'Thus' in the Academic Writing of Turkish EFL Learners. *English Language Teaching*, *10*(2), 64. https://doi.org/10.5539/elt.v10n2p64

van den Bosch, L. J., Segers, E., & Verhoeven, L. (2018). Online processing of causal relations in beginning first and second language readers. *Learning and Individual Differences*, *61*, 59–67. https://doi.org/10.1016/j.lindif.2017.11.007

Van Dijk, T. A. (1979). Pragmatic connectives. *Journal of Pragmatics*, *3*(5), 447–456. https://doi.org/10.1016/0378-2166(79)90019-5

VanPatten, B., & Benati, A. (2010). *Key Terms in Second Language Acquisition*. New York / London : Continuum.

van Silfhout, G. van, Evers-Vermeul, J., & Sanders, T. J. M. (2014). Establishing coherence in schoolbook texts: How connectives and layout affect students' text comprehension. *Dutch Journal of Applied Linguistics*, *3*(1), 1–29. https://doi.org/10.1075/dujal.3.1.01sil

Waltereit, R. (2001). Modal particles and their functional equivalents: A speech-act-theoretic approach. *Journal of Pragmatics*, *33*(9), 1391-1417.

Webber, B. L. (1988). Tense as discourse anaphor. *Technical Reports (CIS)*, 441.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–256. https://doi.org/10.1162/jocn.1996.8.3.231

Wei, X., Zhang, L. J., & Zhang, W. (2020). Associations of L1-to-L2 rhetorical transfer with L2 writers' perception of L2 writing difficulty and L2 writing proficiency. *Journal of English for Academic Purposes*, *47*, 100907. https://doi.org/10.1016/j.jeap.2020.100907

Weydt, H. (1969). *Abtönungspartikel. Die deutschen Modalwörter und ihre französische Entsprechungen*. Bad Homburg: Gehl

Wetzel, M., Zufferey, S. & Gygax, P. (2020). Second Language Acquisition and the Mastery of Discourse Connectives : Assessing the Factors That Hinder L2 - Learners from Mastering French Connectives. *Languages, 5*(3), 35. https://doi.org/10.3390/languages5030035

Wetzel, M., Zufferey, S., & Gygax, P. (2022a). How Robust Is Discourse Processing for Native Readers? The Role of Connectives and the Coherence Relations They Convey. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.822151

Wetzel, M., Zufferey, S., & Gygax, P. (2022b). Do non-native readers rely on connectives? The processing of coherence relations in L2. In M.-J. Cuenca & L. Degand (Eds.), *Discourse Markers in Interaction* (pp. 89–119). De Gruyter. https://doi10.1515/9783110790351-005

Wetzel, M., Crible, L., & Zufferey, S. (2022). Processing clause-internal discourse relations in a second language. *Journal of Second Language Studies, 5(2),* 206-234. https://doi.org/10.1075/jsls.21032.wet

Wetzel, M., Tskhovrebova, E., Gygax, P. M., & Zufferey, S. (2023). Pragmatic and syntactic constraints on French causal connectives: An evaluation of native and non-native speakers' sensitivity. *Journal of Pragmatics*, *209*, 89–107. https://doi.org/10.1016/j.pragma.2023.03.001

Xia, X., & Wolf, H.-G. (2010). Basic-level salience in second language vocabulary acquisition. In S. De Knop, F. Boers, & A. De Rycker (Eds*.), Fostering Language Teaching Efficiency through Cognitive Linguistics* (pp. 79–98). De Gruyter Mouton. https://doi.org/10.1515/9783110245837.79

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, cognition and neuroscience*, *30*(6), 648-672.

Xu, X., Chen, Q., Panther, K.-U., & Wu, Y. (2018). Influence of Concessive and Causal Conjunctions on Pragmatic Processing: Online Measures from Eye Movements and Self-Paced Reading. *Discourse Processes*, *55*(4), 387–409. https://doi.org/10.1080/0163853X.2016.1272088

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied linguistics*, *24*(1), 1-27.

Yung, F., Scholman, M., & Demberg, V. (2021). A practical perspective on connective generation. *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, 72–83. https://doi.org/10.18653/v1/2021.codi-main.7

Zamel, V. (1983). Teaching those missing links in writing1. *ELT Journal*, *37*(1), 22–29. https://doi.org/10.1093/elt/37.1.22

Zhang, A. (2010). Use of cohesive ties in relation to the quality of compositions by Chinese college students. *Journal of Cambridge Studies, 5*(23), 78-86.

Zhang, L., & Seepho, S. (2013). Metacognitive strategy use and academic reading achievement: insights from a Chinese context. *Electronic Journal of Foreign Language Teaching*, *10*(1).

Zinar, S. (1990). Fifth-Graders' Recall of Propositional Content and Causal Relationships from Expository Prose. *Journal of Reading Behavior*, *22*(2), 181–199. https://doi.org/10.1080/10862969009547703

Zufferey, S. (2012). "Car, parce que, puisque" revisited: Three empirical studies on French causal connectives. *Journal of Pragmatics*, *44*(2), 138–153. https://doi.org/10.1016/j.pragma.2011.09.018

Zufferey, S., & Cartoni, B. (2012). English and French Causal Connectives in Contrast. *Languages in Contrast: International Journal for Contrastive Linguistics*, *12*(2), 232–250. https://doi.org/10.1075/lic.12.2.06zuf

Zufferey, S. (2014). Givenness, procedural meaning and connectives. The case of French puisque. *Journal of Pragmatics*, *62*, 121–135. https://doi.org/10.1016/j.pragma.2013.09.022

Zufferey, S., Mak, W., Degand, L., & Sanders, T. (2015). Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research*, *31*(3), 389–411. https://doi.org/10.1177/0267658315573349

Zufferey, S., & Gygax, P. M. (2016). The role of perspective shifts for processing and translating discourse relations. *Discourse Processes*, *53*(7), 532-555.

Zufferey, S., & Degand, L. (2017). Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguistics and Linguistic Theory*, *13*(2), 399-422.

Zufferey, S., & Gygax, P. (2017). Processing Connectives with a Complex Form-Function Mapping in L2: The Case of French "En Effet." *Frontiers in Psychology*, *8*, 1198. https://doi.org/10.3389/fpsyg.2017.01198

Zufferey, S., Mak, W., Verbrugge, S., & Sanders, T. (2018). Usage and processing of the French causal connectives 'car' and 'parce que.' *Journal of French Language Studies*, *28*(1), 85–112. https://doi.org/10.1017/S0959269517000084

Zufferey, S. (2020). *Introduction à la linguistique de corpus*. London : ISTE Group.

Zufferey, S., & Gygax, P. (2020a). Do teenagers know how to use connectives from the written mode? *Lingua*, *234*, 102779. https://doi.org/10.1016/j.lingua.2019.102779

Zufferey, S., & Gygax, P. (2020b). “Roger Broke His Tooth. <However> , He Went to the Dentist”: Why Some Readers Struggle to Evaluate Wrong (and Right) Uses of Connectives. *Discourse Processes*, *57*(2), 184–200. https://doi.org/10.1080/0163853X.2019.1607446

Zufferey, S., Tskhovrebova, E., Wetzel, M., & Gygax, P. (2022). Individual differences in the ability to master connectives: The importance of exposure to print. In M.-J. Cuenca & L. Degand (Eds.), *Discourse Markers in Interaction: From Production to Comprehension*, *376*, De Gruyter.