



HAL
open science

Inférence lexicale dans la compréhension écrite et acquisition subséquente du vocabulaire en chinois langue étrangère

Enhao Zheng

► **To cite this version:**

Enhao Zheng. Inférence lexicale dans la compréhension écrite et acquisition subséquente du vocabulaire en chinois langue étrangère. Linguistique. Université Toulouse le Mirail - Toulouse II, 2022. Français. NNT : 2022TOU20012 . tel-03922667

HAL Id: tel-03922667

<https://theses.hal.science/tel-03922667v1>

Submitted on 4 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par [Université de Toulouse 2-Jean Jaurès](#)

Présentée et soutenue par

Enhao ZHENG

Le 11 mars 2022

Inférence lexicale dans la compréhension
écrite et acquisition subséquente du
vocabulaire en chinois langue étrangère

Volume I

École doctorale et discipline ou spécialité
CLESCO, Science du langage

Unité de recherche
Laboratoire de NeuroPsychoLinguistique (LNPL)

Directrice de Thèse
Olga THEOPHANOUS

Jury

Mehmet-Ali AKINCI, Professeur des Universités, Université de Rouen (rapporteur)
Arnaud ARSLANGUL, Maître de conférences, INALCO Paris (examinateur)
Heather HILTON, Professeur des Universités, Université Lumière Lyon 2 (examinatrice)
Frédéric ISEL, Professeur des Universités, Université Paris Nanterre (rapporteur)
Vanessa TEILHET, Maître de conférences, Université de Toulouse 2 Jean Jaurès (examinatrice)
Olga THEOPHANOUS, Maître de conférences HDR, Université de Toulouse 2 Jean Jaurès (directrice de thèse)

REMERCIEMENTS

致远在中国的父母以及朝夕相伴的丈夫和女儿们
A mes parents en Chine et à Eric, Angèle et Sonia en France

Enseignante de chinois langue étrangère pendant neuf ans en France, j'ai constaté des difficultés chez les étudiants francophones dans l'apprentissage du vocabulaire chinois. Cela m'a amenée à me lancer dans la recherche afin d'explorer ce sujet en profondeur.

Je tiens à remercier avant tout Mme Olga Théophanous, Maître de conférences HDR en science du langage à l'Université de Toulouse Jean Jaurès (UT2J), ma directrice de thèse. Mme Théophanous m'a guidée dans la présente voie de recherche. Grâce à son accompagnement constant et son soutien avisé, j'ai pu concevoir mon projet de recherche, concrétiser le protocole expérimental et enfin achever cette thèse.

Mes remerciements vont également à M. Nicolas Cid, Mme Liwen Tardieu et Mme Qiyuan Ji, enseignants de chinois à l'UT2J pour leur aide pendant la conception des outils expérimentaux sur le plan de la sélection des mots cibles et la rédaction des textes. Je suis également reconnaissante envers mes consœurs, Mme Qiyuan Ji, Mme Solène Jabaud et Mme Xiaofan Molinier, pour leur aide précieuse lors du recrutement de participants pour notre expérience. Je remercie sincèrement Mme Qiyuan Ji, Mme Lin Zhang et Mme Wen Yu pour leur aide sur les transcriptions des entretiens d'inférence lexicale, ainsi que Mme Qiyuan Ji et Mme Xiaofan Molinier pour la double correction de la notation des réponses des participants.

J'adresse mes vifs remerciements à Mme Emilie Massa, chercheuse et enseignante en science du langage à l'UT2J, et M. Paul-Vincent Paubel, ingénieur d'étude CNRS, pour leur conseil précieux pendant la conception et la mise en place des manipulations en oculométrie, ainsi que lors du traitement de données oculométriques.

Un grand remerciement aussi va spécifiquement à M. Saïd Jmel, appui en statistiques aux projets scientifiques à l'UT2J, également à mes collègues du

laboratoire LNPL, Mme Lyanne Ahumada-Ebratt et Mme Lucille Soulier, pour leur aide généreuse et précieuse pendant les analyses statistiques de données.

Je tiens également à remercier tous mes participants, sans qui cette étude n'aurait pu être conduite.

De façon générale, j'exprime ma gratitude à toutes les personnes qui m'ont apporté leur soutien et m'ont rendu service.

Table des matières

Résumé	IX
Abstract	XI
Liste des abréviations	XIII
Liste des figures.....	XV
Liste des tableaux	XIX
INTRODUCTION GENERALE.....	1
CHAPITRE I – INFERENCE LEXICALE DANS LA COMPREHENSION ECRITE.....	9
Introduction	11
1.1 Compréhension écrite	11
1.1.1 Processus de compréhension écrite.....	11
1.1.1.1 Approche « Bas-haut » (« Bottom-up »).....	12
1.1.1.2 Approche « Haut-bas » (« Top-down »).....	12
1.1.1.3 Approche « Interactive ».....	13
1.1.2 Compréhension écrite en L2	14
1.2 Stratégies de lecture	17
1.3 Inférence lexicale, une stratégie importante dans la compréhension écrite	20
1.3.1 Processus d'inférence lexicale	22
1.3.2 Inférence lexicale en L1	23
1.3.3 Inférence lexicale en L2	26
1.3.3.1 Rôle de la distance typologique L1-L2 dans l'inférence lexicale en L2.....	32
1.3.3.2 L1 de l'apprenant mis en relief dans l'inférence lexicale en chinois L2	35
1.3.3.3 Lexicalisation dans la L1	37
1.3.3.4 Effet de la lexicalisation sur l'inférence lexicale	39
1.4 Récapitulation et questions de recherche sur l'inférence lexicale	46
CHAPITRE II – INFERENCE LEXICALE DANS L'APPRENTISSAGE DU VOCABULAIRE	49
Introduction	51
2.1 Inférence lexicale, une stratégie importante d'apprentissage du vocabulaire.....	51
2.1.1 Apprentissage du vocabulaire via la lecture en L1	52
2.1.2 Apprentissage du vocabulaire via la lecture en L2	56
2.1.3 Effet de la lexicalisation sur l'apprentissage du vocabulaire grâce au contexte.	58
2.2 Connaître un mot.....	61

2.3	Apprentissage incident vs intentionnel du vocabulaire.....	64
2.4	Interrogations sur l'apprentissage du vocabulaire via la lecture.....	67
2.4.1	Négligence et ignorance des mots inconnus	67
2.4.2	Disponibilité des informations lexicales dans le contexte	68
2.4.3	Fréquence d'exposition	69
2.4.4	Volume de lecture nécessaire.....	71
2.5	Éléments intervenant dans l'apprentissage incident du vocabulaire.....	73
2.5.1	Attention portée à un mot	74
2.5.2	Engagement de l'apprenant.....	76
2.6	Intégration des instructions dans l'apprentissage incident du vocabulaire	80
2.6.1	Approche « Instruction centrée sur la forme ».....	81
2.6.2	Apprentissage incident du vocabulaire avec l'approche « instruction centrée sur la forme »	82
2.7	Récapitulation et questions de recherche sur l'apprentissage subséquent du vocabulaire inféré.....	85
CHAPITRE III – APPRENTISSAGE DU VOCABULAIRE EN CHINOIS L2		89
Introduction		91
3.1	Enseignement du chinois L2 dans le monde et en France	91
3.2	Caractéristiques de la langue chinoise	93
3.2.1	Transcription phonétique	93
3.2.2	Caractère	94
3.2.2.1	Bloc composant sémantique.....	95
3.2.2.2	Bloc composant phonétique	96
3.2.2.3	Caractères en cas d'homophonie, de polysémie et de polyphonie.....	97
3.2.3	Mot.....	97
3.2.3.1	Caractère, morphème et mot	97
3.2.3.2	Mot composé.....	98
3.2.3.2.1	Sens d'un mot composé	99
3.2.3.2.2	Mode de constitution d'un mot composé.....	99
3.2.3.3	Etendue lexicale exigée pour un Chinois natif	100
3.2.4	Phrase.....	101
3.3	Distance L1-L2, élément mis en relief dans la recherche sur le chinois L2	102
3.4	Récapitulation et objectif de cette étude	105
CHAPITRE IV – PROTOCOLE EXPERIMENTAL		111
Introduction		113

4.1	Public visé.....	114
4.2	Matériels conçus pour l'expérience	117
4.2.1	Mots cibles	117
4.2.2	Textes expérimentaux	119
4.2.3	Test « Vocabulary Knowledge Scale »	124
4.2.4	Tâches post-inférence à réaliser.....	128
4.2.4.1	Exercice lacunaire	129
4.2.4.2	Consultation d'un dictionnaire.....	131
4.3	Méthodes d'expérience	135
4.3.1	La méthode de la « pensée à voix haute ».....	135
4.3.1.1	Interrogations sur la validité de la pensée à voix haute	138
4.3.1.2	Variables importantes de la pensée à voix haute	140
4.3.2	Oculométrie.....	142
4.3.2.1	Mesures d'oculométrie utilisées	143
4.3.2.2	Dispositifs informatiques de l'oculométrie.....	144
4.3.2.3	Format d'affichage des textes testés	146
4.4	Déroulement de l'expérience	149
4.4.1	Pré-test	149
4.4.2	Procédure	149
4.4.2.1	Etape I – Evaluation de connaissances lexicales avant l'inférence lexicale	150
4.4.2.2	Etape II - Jour de l'inférence lexicale	151
4.4.2.3	Etape III – Deuxième évaluation de connaissances lexicales (VKS2)	153
4.4.2.4	Etape IV – Troisième évaluation de connaissances lexicales (VKS3)	154
4.5	Récapitulation	155
CHAPITRE V – SOURCES DE CONNAISSANCE UTILISEES DANS L'INFERENCE LEXICALE		157
Introduction		159
5.1	Analyse qualitative.....	159
5.1.1	Connaissances générales (SC01)	161
5.1.2	Sources propres à Ln (SC02)	161
5.1.3	Contexte lointain (SC03)	163
5.1.4	Ponctuation (SC04).....	163
5.1.5	Contexte proche (SC05).....	163
5.1.6	Grammaire (SC06).....	164
5.1.7	Association/Collocation (SC07)	164

5.1.8	Sens d'un caractère constitutif du mot (SC08)	165
5.1.9	Forme d'un caractère constitutif du mot (SC09)	165
5.1.10	Son d'un caractère constitutif du mot (SC10).....	166
5.1.11	Mode de constitution du mot (SC11).....	166
5.2	Analyse quantitative.....	167
5.2.1	Diversité des sources de connaissance utilisées.....	168
5.2.2	Fréquence d'utilisation des sources de connaissance	172
5.2.2.1	Fréquence d'utilisation des sources de connaissance pour inférer les mots lexicalisés et non lexicalisés	175
5.2.2.2	Fréquence d'utilisation des sources de connaissance pour inférer les mots de différentes classes grammaticales	180
5.3	Discussion	181
5.3.1	Diversité des sources de connaissance utilisées.....	182
5.3.2	Fréquence d'utilisation des sources de connaissance	182
5.3.2.1	Mots lexicalisés vs non lexicalisés	182
5.3.2.2	Adjectif vs Nom vs Verbe.....	186
5.3.3	Interprétation de quelques cas intéressants	188
5.3.3.1	Contexte proche	188
5.3.3.2	Sens d'un caractère constitutif d'un mot inconnu.....	189
5.3.3.3	Connaissances grammaticales.....	192
5.3.3.4	Forme d'un caractère constitutif d'un mot inconnu.....	195
5.3.3.5	Information phonétique.....	197
5.3.3.6	Mode de constitution d'un mot	198
5.4	Conclusions.....	200
CHAPITRE VI – DEGRE DE REUSSITE DE L'INFERENCE DES MOTS CIBLES ET ATTENTION PORTEE AUX MOTS CIBLES PENDANT LA LECTURE ..		203
Introduction		205
6.1	Résultats d'inférence des mots cibles	206
6.1.1	Notation des réponses sur le sens.....	206
6.1.2	Degré de réussite d'inférence pour les mots cibles.....	207
6.1.3	Rapport entre le degré de réussite et les sources de connaissance utilisées.....	209
6.1.4	Comparaison du score entre les mots lexicalisés et non lexicalisés	212
6.1.5	Discussion	216
6.1.5.1	Degré de réussite d'inférence.....	216
6.1.5.1.1	Inférences les mieux réussies.....	216

6.1.5.1.2	Inférences les moins réussies	218
6.1.5.2	Facteurs intervenant dans la réussite d'inférence	220
6.1.5.2.1	Choix de sources de connaissance	220
6.1.5.2.2	Effet de la lexicalisation.....	222
6.1.5.2.3	Connaissances socio-culturelles.....	226
6.2	Attention portée aux mots cibles.....	228
6.2.1	Vitesse de lecture	229
6.2.2	Caractères sautés pendant la lecture.....	230
6.2.3	AOI observées.....	233
6.2.4	Comparaisons entre les AOI concernant les mots cibles et non cibles.....	234
6.2.5	Résultats pour les mots cibles	237
6.2.5.1	Comparaisons entre les AOI des mots lexicalisés et non lexicalisés.....	238
6.2.5.2	Rapport entre l'attention portée aux mots et le degré de réussite d'inférence ..	241
6.2.6	Discussion	242
6.2.6.1	Attention portée aux mots inconnus rencontrés en lecture	243
6.2.6.2	Comparaison entre les mots inconnus lexicalisés et non lexicalisés	245
6.2.6.3	Rapport entre l'attention portée aux mots cibles et le degré de réussite d'inférence	246
6.3	Conclusions.....	247
CHAPITRE VII – EVOLUTION DES CONNAISSANCES LEXICALES		249
Introduction		251
7.1	Evolution des connaissances lexicales multidimensionnelles pour les mots cibles	251
7.1.1	Evolution globale du niveau de connaissance pour les mots cibles.....	253
7.1.2	Comparaison du niveau de connaissance entre le Groupe I et le Groupe II.....	257
7.1.3	Comparaison du niveau de connaissance entre les mots lexicalisés et les mots non lexicalisés	259
7.1.4	Comparaison du niveau de connaissance entre les mots lexicalisés et les mots non lexicalisés pour les deux groupes.....	262
7.2	Evolution de la connaissance sémantique pour les mots cibles	264
7.2.1	Evolution globale du score sur le sens des mots cibles.....	264
7.2.2	Comparaison du score sur le sens des mots cibles entre le Groupe I et le Groupe II.....	269
7.2.3	Comparaison du score sur le sens entre les mots lexicalisés et les mots non lexicalisés	272

7.2.4	Comparaison du score sur le sens des mots lexicalisés et non lexicalisés pour les deux groupes	277
7.3	Discussion	278
7.3.1	Evolution des connaissances multidimensionnelles des mots cibles	278
7.3.1.1	Différence d'évolution du niveau de connaissance entre le Groupe I et le Groupe II	278
7.3.1.2	Différence d'évolution du niveau de connaissance entre les mots lexicalisés et les mots non lexicalisés	279
7.3.2	Evolution de la connaissance sémantique	282
7.3.2.1	Evolution du score du sens des mots cibles	284
7.3.2.2	Différence d'évolution du score sur le sens des mots cibles entre le Groupe I et le Groupe II	286
7.3.2.3	Différence d'évolution du score sur le sens entre les mots lexicalisés et les mots non lexicalisés	287
7.4	Conclusions	293
CHAPITRE VIII – CONCLUSIONS GENERALES		295
8.1	Réflexions pédagogiques	297
8.1.1	Apprentissage du lexique via la lecture en chinois L2	297
8.1.2	Accent sur les caractères dans l'enseignement du lexique chinois L2	300
8.1.3	Apprentissage plus soutenu des mots L2 non lexicalisés	304
8.2	Originalité et limites de cette étude	305
8.2.1	Sélection de mots cibles	306
8.2.2	Groupe de contrôle	306
8.2.3	Observation de l'attention portée aux mots inconnus	307
8.3	Futures pistes de recherche	308
Bibliographie		311
Annexe 1 – Exemple d'un test VKS		331
Annexe 2 – Stimuli oculométriques (dans l'ordre d'affichage sur l'écran)		333
Annexe 3 – Tâches post-inférence à réaliser		361
Annexe 4 - Réponses de différents niveaux de connaissance pour les mots cibles par participants au VKS1		363
Annexe 5 - Réponses de différents niveaux de connaissance pour les mots cibles par participants au VKS2		364
Annexe 6 - Réponses de différents niveaux de connaissance pour les mots cibles par participants au VKS3		365
Annexe 7 - Scores du sens pour les mots cibles par participant au VKS1		366
Annexe 8 - Scores du sens pour les mots cibles par participant à l'inférence lexicale		367

Annexe 9 - Scores du sens pour les mots cibles par participant au VKS2	368
Annexe 10 - Scores du sens pour les mots cibles par participant au VKS3	369

Résumé

Cette thèse traite de l'inférence lexicale dans la compréhension écrite en langue seconde/étrangère (L2), et de l'apprentissage subséquent du vocabulaire inféré.

L'inférence lexicale consiste en la déduction du sens d'un mot inconnu à la lumière des indices contextuels et des connaissances du lecteur. Ce processus de traitement d'informations concernant le mot inconnu peut aboutir à des gains lexicaux pour ce mot. L'apprentissage d'un mot inféré pendant la lecture étant le sous-produit de la compréhension du texte, l'attention et l'investissement cognitif de l'apprenant sur le mot peuvent influencer le résultat de l'apprentissage (Schmidt, 1990 ; Laufer & Hulstijn, 2001).

Nous nous sommes spécifiquement penchée sur le rôle de la lexicalisation, ou de l'existence d'un équivalent lexical attesté dans la langue maternelle (L1) de l'apprenant pour un mot L2. Les connaissances en L1 restent actives pendant les activités en L2. L'équivalent en L1 à un mot L2 lexicalisé permet à l'apprenant de comprendre plus facilement ce dernier en fournissant l'accès au concept correspondant dans le lexique mental de l'apprenant. En revanche, l'inférence du sens d'un mot L2 non lexicalisé dans la L1 est cognitivement plus coûteuse et plus difficile à cause du manque d'un concept correspondant (Paribakht, 2005), tout comme la rétention subséquente du mot (Heidari-Shahreza & Tavakoli, 2012).

La recherche sur la lexicalisation est rare et, à notre connaissance, la langue cible se limite à l'anglais. Nous avons mené cette étude auprès d'apprenants francophones en chinois L2, afin d'observer l'effet de la lexicalisation sur la lecture et sur l'inférence et l'apprentissage subséquent des mots inconnus présents dans la lecture. Nous cherchions aussi à savoir si une charge cognitive plus importante dans l'input aboutirait à de meilleurs gains lexicaux pour les mots inconnus. 33 participants francophones avec un niveau intermédiaire en chinois ont lu pour la compréhension deux textes comportant 20 mots chinois inconnus, 10 mots lexicalisés en français et 10 autres non lexicalisés. Ils ont ensuite inféré le sens des mots cibles surlignés dans les textes, avant d'accomplir un exercice lacunaire portant sur les mots cibles, soit avec à l'aide d'une liste des mots, soit à l'aide d'un dictionnaire chinois-français à leur disposition. Nous avons utilisé d'une part la technique d'oculométrie pour observer l'attention que les participants portaient aux mots cibles pendant la lecture et, d'une autre part, la méthode de la pensée à voix haute afin de déceler leurs raisonnements pour déduire le sens des mots cibles. Afin d'évaluer l'évolution des connaissances lexicales des mots cibles chez nos participants, nous avons administré

le test Vocabulary Scale Knowledge (Paribakht & Wesche, 1997) à trois reprises, un avant l'inférence et deux autres respectivement deux et quatre semaines après l'inférence.

Selon nos résultats, sans effet sur l'attention que les apprenants prêtent aux mots inconnus pendant la lecture, la lexicalisation influence le choix de sources de connaissance des apprenants pendant l'inférence lexicale ; la lexicalisation n'est pas l'élément concluant dans le résultat de l'inférence lexicale, mais les apprenants ont plus de difficultés à retenir un mot L2 non lexicalisé, par rapport à un lexicalisé, à court et à long terme ; enfin, la combinaison de deux tâches post-inférence aboutit à de meilleurs gains pour les mots inconnus notamment sur le plan formel, mais non sur le plan sémantique.

En nous basant sur nos résultats, nous avons conclu cette thèse en suggérant qu'il importe d'encourager les apprenants à apprendre de nouveaux mots L2 via la lecture et selon le guidage de l'enseignant ; davantage de travail explicite est nécessaire pour l'apprentissage des mots L2 non lexicalisés. Il est important d'aider les apprenants du chinois L2 à élaborer dans leur lexique mental des réseaux lexicaux à partir des caractères chinois déjà connus par eux.

Mots-clés : Inférence lexicale, acquisition incidente, lexicalisation, développement du vocabulaire en L2, chinois langue étrangère, oculométrie

Abstract

This thesis focuses on lexical inferencing in reading comprehension in second/foreign language (L2), as well as the subsequent learning of inferred vocabulary.

The lexical inferencing refers to deducing the meaning of an unknown word encountered in reading, in the light of contextual cues and reader's knowledge. The processing of lexical information concerning the unknown word can lead to lexical gains for this one. However, learning an inferred word while reading is a by-product of the comprehension of the text, the learner's attention to the word and his cognitive investment can impact learning outcomes (Schmidt, 1990; Laufer & Hulstijn, 2001).

We are specifically interested in the role of the existence of an attested lexical equivalent in learner's native language (L1) for an L2 word, or lexicalization. L1 knowledge remains active during L2 activities. The L1 equivalent of a lexicalized L2 word makes it easier for learners to understand the L2 word by providing access to the corresponding concept in their mental lexicon. However, inferencing the meaning of a non-lexicalized L2 word is cognitively more costly and more difficult due to the lack of a correspondent concept (Paribakht, 2005), as is the subsequent retention of that word (Heidari-Shahreza & Tavakoli, 2012).

Research on lexicalization is scarce, and so far, the target language has been limited to English according to our knowledge. We conducted this study with French-speaking learners of Chinese, aiming to observe the effect of lexicalization on lexical inferencing and subsequent learning of unknown words encountered in reading. We also investigated whether a more important cognitive load in the input would lead to better lexical gains for unknown words. 33 French-speaking participants with an intermediate level in Chinese read for understanding of two texts comprising 20 unknown Chinese words, 10 lexicalized in French and 10 others not lexicalized. They were then asked to infer the meaning of the highlighted words in the texts. Immediately after, they completed a gap-fill exercise on the target words, either provided with a list of target words and their meaning or having access to a Chinese-French dictionary. We used the eye tracking technique to observe the participants' attention to the target words during reading and, the think-aloud method to understand the participants' reasoning to deduce the target words' meaning. In purpose to assess the evolution of the participants' lexical knowledge for the target words, we conducted the Vocabulary Scale Knowledge test (Paribakht & Wesche, 1997) three

times, one pre-inference and two more, respectively two and four weeks after the inference session.

According to our results, with no effect on learner's attention to unfamiliar words during reading, lexicalization influences learners' choice of knowledge sources during lexical inferencing. Lexicalization is not the conclusive factor in the result of lexical inferencing, but learner has more difficulty retaining the meaning of an unlexicalized L2 word, compared to a lexicalized one, in both the short and long terms. Finally, the combination of the two post-inference tasks leads to better gains for unknown words, especially on the formal level, rather than on the semantic level.

Based on our results, we proposed in the end that it is important to encourage learners to learn new L2 words via reading, according to the teacher's guidance. More activities with explicit instructions are necessary for learning non-lexicalized L2 words. Especially in Chinese L2, it would be important to help learners to develop in their mental lexicon lexical networks based on Chinese characters they already know.

Keywords: Lexical inferencing, incidental acquisition, lexicalization, vocabulary development in L2, Chinese as a second/foreign language, eye-tracking

Liste des abréviations

AOI – Area of Interest (Aire d'intérêt)

L – Lexicalisé

L1 – Langue maternelle

L2 – Langue seconde/étrangère

MC – Mot cible

MS – Milliseconde

ND – Sources linguistiques propres à la L2 au niveau du discours

NM – Sources linguistiques propres à la L2 au niveau du mot

NP – Sources linguistiques propres à la L2 au niveau de la phrase

NL – Non lexicalisé

PVH – Pensée à voix haute (Think-aloud)

QCM – Questionnaire à choix multiple

SC – Source de connaissance

SLn – Sources propres à la langue maternelle ou à une autre langue que l'apprenant connaît

SNL – Sources non linguistiques

VKS – Vocabulary Knowledge Scale

VLТ – Vocabulary Level Test

UT2J – Université de Toulouse II Jean Jaurès

Liste des figures

Figure 1.1 – Modèle de la compréhension/production lexicale pour les modalités orales et écrites (De Bot et al., 1997, p. 315)	23
Figure 2.1 – Schéma théorique du processus d'apprentissage d'un mot via la lecture	81
Figure 4.1 – Procédure pour consulter un mot chinois par l'accès phonétique dans un dictionnaire chinois monolingue	133
Figure 4.2 – Procédure pour consulter un mot chinois par l'accès composant-clé dans un dictionnaire chinois monolingue	135
Figure 4.3 – Déroulement de la pensée à voix haute durant une tâche (Ericsson, 2006, p. 227)	136
Figure 4.4 – Schéma synthétique de l'utilisation de la méthode de la pensée à voix haute dans l'inférence lexicale (basé sur le schéma d'Ericsson, 2006, p. 227)	137
Figure 4.5 – Déroulement non séquentiel de la pensée à voix haute pendant une tâche (basée sur le schéma d'Ericsson, 2006, p. 227)	138
Figure 4.6 – Mise à l'écran des textes A et B et emplacement des mots cibles dans les textes	148
Figure 4.7 – Exemple des AOI mises en place pour les caractères	148
Figure 4.8 – Procédure de l'expérience	150
Figure 5.1 – Diversité en moyenne des sources de connaissance utilisées par mot	170
Figure 5.2 – Nombre d'inférences pour les mots lexicalisés (L) et non lexicalisés (NL), classées selon la diversité des sources de connaissance utilisées	171
Figure 5.3 – Poids relatif des catégories de source de connaissance	173
Figure 5.4 – Poids relatif des sources de connaissance	173
Figure 5.5 – Fréquence d'utilisation de la SC05 pendant l'inférence des mots cibles	174
Figure 5.6 – Fréquence d'utilisation de la SC06 pendant l'inférence des mots cibles	174
Figure 5.7 – Fréquence d'utilisation de la SC08 pendant l'inférence des mots cibles	174
Figure 5.8 – Fréquence d'utilisation de la SC09 pendant l'inférence des mots cibles	174
Figure 5.9 – Poids relatif des catégories de sources de connaissance	176

Figure 5.10 – Poids relatifs des sources de connaissance	176
Figure 5.11 – Fréquence d'utilisation de la grammaire (SC06) et du sens d'un caractère constitutif du mot (SC08)	178
Figure 6.1 – Score d'inférence en moyenne pour les mots cibles	207
Figure 6.2 – Utilisation du sens d'un caractère constitutif (SC08)	212
Figure 6.3 – Fréquence d'utilisation du sens d'un caractère constitutif du mot (SC08) et score d'inférence par mot	212
Figure 6.4 – Score d'inférence pour les mots lexicalisés (L) et non lexicalisés (NL)	215
Figure 6.5 – Temps total (ms) en moyenne pour les AOI contenant les mots cibles et non cibles	234
Figure 6.6 – Nombre de visite du regard en moyenne pour les AOI contenant les mots cibles et non cibles	234
Figure 6.7 – Première fixation (ms) en moyenne pour les AOI contenant les mots cibles et non cibles	235
Figure 6.8 – Temps total (ms) en moyenne par mot	237
Figure 6.9 – Nombre de visite du regard en moyenne par mot	237
Figure 6.10 – Première fixation (ms) en moyenne par mot	237
Figure 6.11 – Temps total (ms) en moyenne pour les AOI des mots cibles lexicalisés (L) et non lexicalisés (NL)	238
Figure 6.12 – Nombre de visite en moyenne pour les AOI des mots cibles lexicalisés (L) et non lexicalisés (NL)	239
Figure 6.13 – Première fixation (ms) en moyenne pour les AOI des mots cibles lexicalisés (L) et non lexicalisés (NL)	239
Figure 7.1 – Evolution de la répartition des réponses de différents niveaux de connaissance	255
Figure 7.2 – Evolution de la répartition des réponses de différents niveaux de connaissance pour le Groupe I et le Groupe II	259
Figure 7.3 – Evolution de la répartition des réponses de différents niveaux de connaissance pour les mots lexicalisés (L) et les mots non lexicalisés (NL)	262
Figure 7.4 – Evolution de la répartition des réponses de différents niveaux de connaissance du Groupe I et du Groupe II pour les mots lexicalisés (L) et non lexicalisés (NL)	263
Figure 7.5 – Evolutions du score moyen sur le sens des mots cibles pour le Groupe I et le Groupe II	271
Figure 7.6 – Evolutions du score moyen sur le sens pour les mots lexicalisés (L) et non lexicalisés (NL)	274

Figure 7.7 – Evolution du score moyen sur le sens des mots lexicalisés (L) et des mots non lexicalisés (NL)	277
Figure 7.8 – Transfert d’informations lexicales d’un lemme L1 vers un lemme L2 (adapté de N. Jiang, 2004, p. 418)	290
Figure 7.9 – Lien schématique « forme-concept » pour un mot L2 lexicalisé et « forme-conceptS » pour un mot L2 non lexicalisé	291
Figure 8.1 – Schéma des réseaux de mots établis à partir des caractères (basé sur les modèles de Y. H. Xu, 2014, p. 36)	304

Liste des tableaux

Tableau 1.1 – Théorie de la déduction du sens d’un mot inconnu via le contexte (extraite de Sternberg et al., 1982, pp. 163-169)	27
Tableau 1.2 – Typologie des sources de connaissance utilisées dans l’inférence lexicale en L2	42
Tableau 2.1 – Définitions de la profondeur des connaissances lexicales (Read 2004, adapté par Théophanous, 2019, p. 12)	62
Tableau 2.2 – Connaissance d’un mot (Nation, 2001, adapté par Théophanous, 2019, p. 14)	63
Tableau 2.3 – Volume de lecture en mot et en durée de temps hebdomadaire pour s’exposer autour de 12 fois aux familles de 1000-mots en anglais (Nation, 2014, p. 7)	72
Tableau 4.1 – Informations concernant les 33 participants	116
Tableau 4.2 – Mots cibles	118
Tableau 4.3 – Mots cibles présents dans le Texte A	121
Tableau 4.4 – Mots cibles présents dans le Texte B	123
Tableau 4.5 – Signification des scores du VKS (VKS scoring categories: Meaning of scores, d’après Paribakht & Wesche, 1997, p. 181)	125
Tableau 4.6 – Echelle d’éllicitation du VKS : la catégorie d’auto-évaluation (VKS elicitation scale: Self-report categories, d’après Paribakht & Wesche, 1997, p. 180)	125
Tableau 5.1 – Typologie des sources de connaissance utilisées pour inférer un mot chinois	162
Tableau 5.2 – Nombre d’inférences classé selon la diversité des sources de connaissance utilisées	169
Tableau 5.3 – Diversité des sources de connaissance utilisées pour inférer les mots lexicalisés (L) et non lexicalisés (NL)	171
Tableau 5.4 – Fréquence d’utilisation des sources de connaissance pendant l’inférence lexicale	172
Tableau 5.5 – Fréquence d’utilisation des sources de connaissance pendant l’inférence des mots lexicalisés (L) et non lexicalisés (NL)	175
Tableau 5.6 – Fréquence d’utilisation des sources de connaissance pour inférer les mots cibles de différentes classes grammaticale	180
Tableau 6.1 – Score d’inférence lexicale	207
Tableau 6.2 – Nombre des inférences observées et score d’inférence, classés selon la diversité des sources de connaissance utilisées	210

Tableau 6.3 – Nombre et répartition des inférences observées pour les mots lexicalisés (L) et non lexicalisés (NL), classés selon le score	212
Tableau 6.4 – Score d’inférence pour les mots lexicalisés (L) et non lexicalisés (NL)	213
Tableau 6.5 – Score d’inférence classé par lexicalisation pour les participants	214
Tableau 6.6 – Temps de lecture (minute) en moyenne pour les deux textes par participant	230
Tableau 6.7 – Nombre des AOI observées	234
Tableau 6.8 – Temps total, nombre de visite et première fixation en moyenne pour les AOI contenant les mots cibles (AOI MC) et non cibles (AOI non MC) par participants	236
Tableau 6.9 – Temps total, nombre de visite et première fixation en moyenne pour les AOI concernant les mots lexicalisés (L) et non lexicalisés (NL) par participants	240
Tableau 6.10 – Temps total, nombre de visite, première fixation et score d’inférence en moyenne par participant	241
Tableau 7.1 – Nombre et répartition des réponses de différents niveaux aux trois tests VKS	254
Tableau 7.2 – Nombre et répartition des réponses de différents niveaux de connaissance aux trois tests VKS pour le Groupe I et le Groupe II	257
Tableau 7.3 – Nombre et répartition des réponses de différents niveaux aux trois tests VKS pour les mots lexicalisés (L) et non lexicalisés (NL)	260
Tableau 7.4 – Nombre et répartition des réponses de différents niveaux de connaissance aux trois tests VKS pour les mots lexicalisés (L) et non lexicalisés (NL) des deux groupes d’apprentissage	263
Tableau 7.5 – Score sur le sens de mots cibles aux quatre phases	265
Tableau 7.6 – Nombre et répartition des réponses par échelon de score aux quatre phases d’apprentissage	266
Tableau 7.7 – Score sur le sens des mots cibles en moyen aux quatre phases d’apprentissage	267
Tableau 7.8 – Score sur le sens des mots cibles du Groupe I et du Groupe II aux quatre phases d’apprentissage	270
Tableau 7.9 – Nombre et répartition des réponses de différents scores du Groupe I et du Groupe II aux quatre phases d’apprentissage	271
Tableau 7.10 – Score sur le sens des mots lexicalisés (L) et non lexicalisés (NL) en moyenne aux quatre phases d’apprentissage, classé par participant	272
Tableau 7.11 – Score sur le sens pour les mots lexicalisés (L) et non lexicalisés (NL) aux quatre phases d’apprentissage	273

Tableau 7.12 – Nombre et répartition des réponses de différents scores pour les mots lexicalisés (L) et non lexicalisés (NL) aux quatre phases d'apprentissage	275
Tableau 7.13 – Nombre et répartition des réponses de différents scores pour les mots lexicalisés (L)	275
Tableau 7.14 – Nombre et répartition des réponses de différents scores pour les mots non lexicalisés (NL)	276
Tableau 7.15 – Nombre et répartition des réponses de différents niveaux de connaissance au VKS1 pour les mots lexicalisés (L) et non lexicalisés (NL)	281

INTRODUCTION GENERALE

La lecture est un acte important dans notre vie quotidienne. Selon la théorie interactive de la lecture de Kintsch & van Dijk (1978), pendant la lecture, le lecteur traite à la fois les informations envoyées par les mots et celles provenant de ses connaissances préalablement acquises. Ainsi, afin de mieux réussir la compréhension d'un texte, pendant la lecture, le lecteur peut utiliser des stratégies telles que l'inférence qui peut l'aider à déchiffrer des messages implicitement délivrés par l'auteur, en dehors de ceux explicitement montrés par le texte, afin d'avoir une compréhension intégrale d'un texte (Anderson & Pearson, 1984).

L'inférence lexicale est une stratégie qu'un lecteur peut utiliser en cas de lacune lexicale. Puisque personne ne maîtrise le lexique d'une langue de façon exhaustive, un lecteur peut rencontrer des mots inconnus pendant la lecture même dans sa langue maternelle (L1). A la rencontre d'un mot inconnu dans un texte, le lecteur peut utiliser l'inférence lexicale qui consiste à déduire le sens de ce mot inconnu à la lumière des indices contextuels, mais aussi des connaissances linguistiques et non linguistiques du lecteur (Haastrup, 1991; Sternberg et al., 1982)

L'inférence lexicale peut être utilisée en langue seconde/étrangère (L2) comme en L1. Un apprenant d'une L2 a normalement davantage de lacunes lexicales dans la L2 que dans sa langue maternelle. L'inférence lexicale représente donc un grand intérêt pour les apprenants d'une L2 qui tentent d'avoir une lecture fluide et d'obtenir une compréhension globale du texte.

L'inférence lexicale est un processus cognitif complexe. Du fait que les connaissances acquises en L1 restent actives pendant des activités en L2 (Kroll et al., 2010), la distance typologique entre la L1 de l'apprenant et la L2 peut influencer le processus et le résultat de l'inférence lexicale en L2 (Hamada & Koda, 2010). Quand la L2 est proche de la L1, afin de comprendre un mot L2 inconnu, l'apprenant peut s'appuyer sur ce qui est commun entre les deux langues, par exemple les mots

similaires à la forme et au sens ; cette ressource est moins disponible entre deux langues typologiquement lointaines (Paribakht & Tréville, 2007).

En outre, un apprenant L2 est généralement instruit dans sa L1 et possède un riche système de concepts qui sont acquis pendant le développement du vocabulaire en L1 (N. Jiang, 2004). Selon De Bot et al., (1997), l'inférence du sens d'un mot inconnu consiste à établir un lien entre la forme de ce mot et un concept par le biais d'un lemme dans le lexique mental de l'apprenant, à partir des indices repérés. Ce processus en L2 peut impliquer l'activation d'un lemme en L1 lié au même concept (ou largement similaire) (Paribakht, 2005). Autrement dit, quand un mot L2 a un mot ou une unité lexicale équivalente dans la L1 de l'apprenant, l'activation de cet équivalent en L1 peut aider l'apprenant à comprendre plus facilement le mot L2 concerné, en fournissant l'accès au concept correspondant. Quand la L2 est très différente de la L1, l'apprenant risque davantage de rencontrer des mots L2 n'ayant pas d'homologue dans sa L1, c'est-à-dire des mots L2 non lexicalisés. La non lexicalisation implique donc le manque du concept correspondant à un mot L2 dans le lexique mental de l'apprenant. La compréhension d'un mot L2 non lexicalisé est cognitivement plus coûteuse et, par conséquent, plus difficile pour l'apprenant, parce que ce dernier ne parvient pas à établir aisément le lien forme-concept pour ce mot (Wesche & Paribakht, 2009 ; L. Fan & Zhang, 2014).

Outre son rôle de la stratégie de lecture, l'inférence lexicale est également une stratégie importante d'apprentissage du vocabulaire. A force de repérer des informations lexicales concernant un mot inconnu, puis les traiter afin d'avoir un sens logique de ce mot par rapport à son contexte, l'inférence d'un mot inconnu peut aboutir à l'acquisition de ce dernier (Nguyen-Xuan, 1995). Des études ont confirmé que les enfants et les jeunes apprenants natifs sont capables de déduire le sens des mots nouveaux présents dans la lecture grâce au contexte et, par la suite, de les acquérir. La lecture extensive sans intervention pédagogique est la voie principale par laquelle les apprenants natifs développent leur vocabulaire pendant les années scolaires et au-delà (Nagy et al., 1985, 1987).

En L2, un apprenant peut parvenir à inférer correctement le sens d'un mot

inconnu présent dans la lecture et puis à obtenir des gains lexicaux sur ce mot inféré (Fraser, 1999). Cependant, beaucoup de facteurs peuvent intervenir dans le résultat de l'apprentissage d'un mot L2 subséquent à l'inférence lexicale, tels que la lexicalisation. Des études ont montré qu'un apprenant a davantage de difficultés à apprendre des mots L2 non lexicalisés grâce au contexte, par rapport à ceux lexicalisés (C. Chen & Truscott, 2010; Heidari-Shahreza & Tavakoli, 2012). L'existence d'un concept correspondant à un mot L2 dans le système conceptuel de l'apprenant semble favoriser l'apprentissage de ce mot via la lecture.

L'apprentissage des mots inférés pendant la lecture a lieu de façon incidente, puisque l'objectif principal de l'apprenant est la compréhension globale du texte (Laufer, 2017). En outre, la connaissance d'un mot comprend les aspects de la forme, du sens et de l'emploi du mot (Nation, 2001), et connaître un mot est un processus graduel et long. Les informations lexicales sur les plans divers d'un mot exigent des traitements cognitifs de différentes profondeurs. Une exposition massive et constante est nécessaire afin de combler le lemme dans le lexique mental de l'apprenant des informations de tous aspects concernant le mot et puis de les consolider (Nation, 2014; Schmitt, 2008; Uchihara et al., 2019). Développer un vocabulaire satisfaisant via la lecture dans le cadre naturel semble ardu à réaliser en L2, notamment dans un contexte d'apprentissage hétéroglotte à cause d'un input linguistique généralement réduit.

Selon l'hypothèse de noticing («Noticing Hypothesis », Schmidt, 1990, 2010) et l'hypothèse de la charge d'engagement (« Involvement Load Hypothesis », Laufer & Hulstijn, 2001), l'attention qu'un apprenant porte à un mot inconnu et l'engagement cognitif de sa part pour ce mot sont deux facteurs importants dans le résultat d'apprentissage du mot. Plus précisément, l'attention que l'apprenant porte à un mot inconnu est la condition sine qua non afin que l'input se transforme en saisie ; plus l'apprenant est engagé cognitivement pour connaître le mot inconnu, plus il peut obtenir de gains lexicaux pour ce mot. L'engagement cognitif de l'apprenant peut être imposé par un agent externe, par exemple l'objectif d'une tâche concernant ce mot à réaliser. Pour cela, certains chercheurs ont proposé d'intégrer l'approche « instruction centrée sur la forme » (Long, 1998) dans l'apprentissage incident du vocabulaire en

L2, afin d'orienter l'attention de l'apprenant vers les mots nouveaux présents dans la lecture et de créer la nécessité pour l'apprenant de connaître ces mots à l'aide d'interventions pédagogiques. Des études réalisées à ce sujet ont montré que le type de la/des tâche(s) à réaliser a un effet plus important que la simple fréquence d'exposition sur les gains lexicaux d'un mot inconnu (Hong & Wang, 2016; Laufer & Rozovski-Roitblat, 2015).

Pour notre part, nous nous sommes intéressée dans cette thèse à l'inférence et l'apprentissage subséquent du vocabulaire dans la compréhension écrite en L2, en nous penchant spécifiquement sur le rôle de la lexicalisation. La recherche sur la lexicalisation est très rare jusqu'à présent et, autant que nous le sachions, la langue cible étudiée se limite à la langue anglaise.

Nous visons les apprenants francophones du chinois L2. Etant non alphabétique, la langue chinoise est linguistiquement et culturellement lointaine de la langue française. Les apprenants francophones, en comparaison à une L2 similaire à sa langue maternelle, telle que l'anglais, pendant l'apprentissage de la langue chinoise, risquent davantage de rencontrer des mots non lexicalisés en français, en dehors des difficultés sur le plan de la mémorisation et de la reconnaissance des caractères chinois.

Nous cherchions à comprendre si la lexicalisation d'un mot chinois L2 inconnu a un effet a) sur l'attention qu'un apprenant francophone porte à ce mot pendant la lecture, b) sur les sources de connaissance que l'apprenant utilise pour inférer le sens du mot et le degré de réussite de l'inférence lexicale et c) sur la rétention subséquente de ce mot à court et à long terme. En outre, dans le but d'apprendre un mot chinois à la suite de son inférence, nous voulions savoir si la combinaison de deux tâches de vocabulaire à réaliser à l'issue de l'inférence, par rapport à une seule tâche, peut aboutir à de meilleurs gains lexicaux pour le mot.

Nous faisons les hypothèses que 1) un apprenant francophone aurait plus de difficultés à déduire le sens des mots chinois non lexicalisés et puis à les retenir, par rapport à des mots chinois lexicalisés, et 2) la combinaison de deux tâches de vocabulaire à réaliser à l'issue de l'inférence lexicale aboutirait à de meilleurs gains

lexicaux sur les mots chinois inconnus, par rapport à une seule tâche. Notre revue de littérature ne nous permettait pas d'émettre des hypothèses pour les questions concernant l'effet de la lexicalisation sur l'attention accordée aux mots chinois inconnus par un apprenant francophone pendant la lecture, ni sur le choix des sources de connaissance d'un apprenant francophone pendant l'inférence du sens des mots chinois.

Afin de vérifier nos hypothèses et chercher les réponses aux questions posées, nous avons mené une étude auprès d'apprenants francophones en chinois L2. L'objectif de notre étude est de contribuer à une meilleure compréhension du rôle de la lexicalisation dans la lecture et dans l'apprentissage du vocabulaire en L2, ainsi que de proposer des pistes didactiques pour l'enseignement du lexique en chinois L2.

Nous présenterons cette étude selon l'ordre suivant :

- Dans le **Chapitre I**, nous présenterons l'utilisation de la stratégie d'inférence lexicale dans la compréhension écrite en L2, comme en L1. Nous nous focaliserons sur l'effet de la lexicalisation d'un mot L2 inconnu dans la L1 de l'apprenant sur le processus et le résultat de l'inférence lexicale. Nous poserons à la fin du chapitre nos questions de recherche concernant l'effet de la lexicalisation sur l'inférence lexicale en chinois L2.
- Le **Chapitre II** est dédié au rôle de l'inférence lexicale dans l'apprentissage incident du vocabulaire. Conjointement à notre intérêt sur le rôle de la lexicalisation, nous nous focaliserons sur deux éléments clés dans l'apprentissage incident du vocabulaire, l'attention de l'apprenant et la charge cognitive dans l'input. Nous poserons à la fin du chapitre nos questions concernant l'effet de la lexicalisation et l'engagement cognitif de l'apprenant sur l'apprentissage du vocabulaire chinois subséquent à l'inférence lexicale à court et à long terme.
- Nous présenterons dans le **Chapitre III** certaines particularités de la langue chinoise et l'influence de ces particularités sur l'apprentissage du chinois L2 dans la pratique et dans la recherche.
- Dans le **Chapitre IV**, nous présenterons le protocole de notre expérience afin

de répondre aux questions de recherche et de vérifier nos hypothèses. Plus précisément, il sera question de nos participants, des matériels conçus, des méthodes expérimentales utilisées et de la procédure de l'expérience.

- Le **Chapitre V** est consacré à la présentation des résultats concernant le processus de l'inférence lexicale. Nous présenterons la typologie des sources de connaissance utilisées par les apprenants francophones pour inférer les mots chinois cibles. Nous comparerons les sources de connaissance utilisées pour inférer les mots chinois lexicalisés et les mots non lexicalisés. A la fin du chapitre, nous présenterons nos conclusions sur l'effet de la lexicalisation sur le choix de sources de connaissance d'un apprenant francophone pendant l'inférence du sens d'un mot chinois inconnu.
- Dans la première partie du **Chapitre VI**, nous montrerons les résultats sur le degré de réussite de l'inférence des mots cibles lexicalisés et ceux non lexicalisés. Nous examinerons par la suite le lien entre le degré de réussite et les sources de connaissance utilisées pendant l'inférence lexicale.
Dans la deuxième partie, nous présenterons les résultats concernant l'attention que les apprenants francophones portaient aux mots cibles pendant la lecture pour la compréhension. Nous examinerons si les participants ont prêté différemment leur attention aux mots cibles selon que les mots sont lexicalisés ou non en français, et également si les participants ont mieux réussi l'inférence des mots cibles auxquels ils ont porté davantage d'attention pendant la lecture. Nous terminerons ce chapitre avec nos conclusions sur l'effet de la lexicalisation sur l'attention qu'un apprenant francophone porte à un mot chinois inconnu présent dans la lecture et sur le résultat de l'inférence lexicale.
- Le **Chapitre VII** concerne la présentation des résultats sur la rétention des mots cibles à la suite d'une seule tâche de vocabulaire ou de deux tâches post-inférence. Nous observerons l'évolution de la connaissance lexicale des mots cibles chez nos participants selon différents aspects de la connaissance lexicale (forme, sens). Nous tirerons de ces résultats les conclusions sur l'effet de la charge cognitive dans l'input et l'effet de la lexicalisation sur l'apprentissage d'un mot chinois inconnu à court et à long terme.

- En nous basant sur les résultats et les conclusions présentées dans les trois chapitres précédents, nous proposerons dans le **Chapitre VIII** nos réflexions sur la didactique du lexique en chinois L2, ainsi que les perspectives qu'ouvrent les résultats de notre étude dans la recherche sur l'apprentissage du vocabulaire en L2.

La bibliographie figure à la fin du volume I de cette thèse, accompagnée des annexes qui montrent les résultats en détail.

Les données oculométriques concernant l'attention que les participants portaient aux mots cibles pendant la lecture, ainsi que les transcriptions de tous les entretiens pendant l'inférence lexicale, sont montrées dans le volume II de la thèse.

**CHAPITRE I – INFERENCE LEXICALE DANS LA
COMPREHENSION ECRITE**

Introduction

Nous commencerons ce chapitre par une discussion sur la compréhension écrite et le rôle du vocabulaire dans cet acte intellectuel, en langue maternelle (L1) et en langue seconde/étrangère (L2)

Par la suite, nous parlerons des stratégies de lecture qu'un lecteur peut utiliser pendant qu'il lit un texte, entre autres la stratégie d'inférence lexicale en cas de lacune lexicale. Nous présenterons les études qui investiguent ce processus cognitif de l'inférence lexicale et les éléments qui peuvent intervenir dans le processus et le résultat de l'inférence lexicale en L1 et en L2. Parmi tous les éléments impliqués dans l'inférence lexicale en L2, nous nous pencherons sur la distance typologique entre la L2 et la L1 de l'apprenant. Nous nous focaliserons sur l'inférence des mots L2 n'ayant pas d'équivalent dans la L1 de l'apprenant (ou non lexicalisés dans la L1) et présenterons les études réalisées à ce sujet.

Nous poserons, à la fin du chapitre, nos premières questions de recherche concernant l'effet de la lexicalisation d'un mot L2 inconnu, rencontré pendant la lecture, sur le processus et sur le résultat de l'inférence lexicale.

1.1 Compréhension écrite

La lecture, existant depuis des milliers d'années dans l'histoire des humains, est une activité quotidienne importante de nos jours. Il s'agit d'un processus pendant lequel un lecteur extrait des informations à partir des signes écrits ou imprimés, donc essentiellement les mots, pour comprendre le message délivré par l'auteur. Les mots sont donc porteurs d'informations que le lecteur cherche à comprendre et le vocabulaire est l'élément fondamental dans la compréhension d'un texte.

1.1.1 Processus de compréhension écrite

Dans les années 1960 et 1970, on a beaucoup discuté le processus de la compréhension d'un texte. Nous pouvons résumer trois approches principales de lecture représentatives de cette époque, l'approche « Bas-haut » (« Bottom-up »),

l'approche « Haut-bas » (« Top-down ») et l'approche « Interactive ».

1.1.1.1 Approche « Bas-haut » (« Bottom-up »)

L'approche « Bas-haut », avec comme représentant principal Gough (1972), schématise la lecture comme un cheminement à partir du décodage des mots individuels, de la compréhension des phrases et des paragraphes jusqu'à la compréhension globale d'un texte entier, accompagné de l'intégration de la grammaire. L'accent de cette approche « Bas-haut » est mis sur la capacité du lecteur de décoder ou reconnaître des mots, indépendamment du contexte dans lequel se trouvent les mots (Hudson, 2007). Selon cette approche, la capacité à décoder est critique dans la compréhension écrite, autant pour les enfants (Badian, 2005) que pour les lecteurs adultes (Holmes, 2009).

La capacité de décoder des mots reste l'un des éléments fondamentaux dans le modèle simple de la lecture (« Simple View of Reading »), proposé plus tard par Gough & Tunmer (1986). Le modèle, qui se présente sous la forme « Compréhension écrite = Décodage des mots + Compréhension du langage », implique que la performance de la compréhension écrite est le produit de la capacité à décoder des mots et la capacité de traiter des informations lexicales à divers niveaux (c'est-à-dire au mot, à la phrase et au discours) (Goigoux et al., 2016).

1.1.1.2 Approche « Haut-bas » (« Top-down »)

Le fait que le modèle simple de la lecture inclut la compétence linguistique du lecteur dans sa formule représente, pour des chercheurs, un compromis entre l'approche « Bas-haut » et l'approche « posé un jalon dans l'évolution des théories sur la compréhension écrite Dombey, 2009).

L'approche « Haut-bas » met en relief le rôle du lecteur dans le processus de la compréhension écrite. Examinant la lecture depuis l'angle psycholinguistique et sociolinguistique, Goodman (1967), le chercheur emblématique de cette approche, met en avant que « *La lecture est un processus sélectif qui comprend l'utilisation partielle des indices linguistiques que le lecteur sélectionne parmi les informations venant de sa perception visuelle selon ses attentes. Comme les informations*

sélectionnées sont partielles, le lecteur doit répéter à tout moment le cycle de confirmation, rejet et sélection des informations afin de faire progresser sa lecture. (pp. 126-127)¹ »

D'après cette approche, la compréhension d'un texte pourrait être élaborée avant la fin de la lecture, à l'aide des connaissances pré-acquises du lecteur qui, par la suite, filtre et réorganise les informations au fur et à mesure selon ses besoins et ses attentes. Goodman interprète la lecture comme « un jeu de devinette psycholinguistique » (« a psycholinguistic guessing game »). Cela souligne que le lecteur reste actif et proactif pendant sa lecture, et que les traitements qu'il fait des informations identifiées dans le texte ne sont pas mécaniquement procéduraux.

1.1.1.3 Approche « Interactive »

L'approche « Haut-bas » met la lumière sur les activités cognitives du lecteur et ouvre la voie à la troisième approche de lecture, l'approche « Interactive », combinaison de « Bas-haut » et de « Haut-bas » (Kintsch & van Dijk, 1978; Rumelhart, 1977).

Cette approche dévoile la relation interactive et dynamique entre deux sources d'informations principales, le texte et les connaissances du lecteur, ce qui est aujourd'hui généralement accepté par les chercheurs et les enseignants. Comme l'indique Kintsch (2005), « ...les processus « Haut-bas » et « Bas-haut » font partie intégrante de la perception, de la résolution de problèmes et de la compréhension. Sans apport sensoriel (« Bas-haut »), nous ne pourrions ni percevoir, ni comprendre, ni penser. Cependant, la perception, la compréhension et la pensée seraient également impossibles sans une composante mémoire ou connaissance (« Haut-bas »). (p. 126)² ». Autrement dit, selon cette approche, un lecteur traite simultanément les informations extraites des mots imprimés et celles provenant de

¹ « Reading is a selective process. It involves partial use of available minimal language cues selected from perceptual input on the basis of the reader's expectation. As this partial information is processed, tentative decisions are made to be confirmed, rejected, or refined as reading progresses. »

² « ...both top-down and bottom-up processes are integral parts of perception, problem solving, and comprehension. Without sensory input (bottom-up) we could neither perceive, nor comprehend, nor think. However, perception, comprehension, and thought would be equally impossible without a memory or knowledge component (top-down). »

ses connaissances préalablement acquises à l'aide des deux approches qui s'alternent et se compensent.

En dehors des éléments discutés dans les deux approches précédentes, les chercheurs se sont aussi penchés sur les facteurs tels que l'interaction entre l'intention de l'auteur et l'interprétation du lecteur (Smith, 1994) et l'influence des connaissances pré-acquises du lecteur sur sa façon d'interpréter le contenu du texte (Anderson & Pearson, 1984 ; Hudson, 1991). Selon Hudson (2007), le rôle du lecteur est de former arbitrairement une idée de ce qui est représenté par un texte dans un contexte approprié et de le transformer en sens. Le contexte dans lequel le lecteur interprète le sens du texte est un autre facteur mis en avant dans le processus de compréhension écrite.

1.1.2 Compréhension écrite en L2

La compréhension écrite en L2 est un processus plus complexe que celle en L1. La plus grande différence entre la compréhension écrite en L1 et celle en L2 est liée au fait que les apprenants L2 sont en général instruits et ont déjà acquis des connaissances encyclopédiques et des compétences de lecture dans leur langue maternelle grâce à leurs activités de compréhension écrite en L1. Ces connaissances et compétences acquises en L1 peuvent être transférées vers la L2 pendant les activités de compréhension (Koda, 2005).

Alderson (1984) a posé la fameuse question : la lecture en langue étrangère est-elle un problème de langue ou de compréhension ?³ Le problème de langue est lié au manque de connaissances et de compétences nécessaires pour traiter des informations linguistiques propres à la L2 (par exemple les connaissances orthographiques, syntaxiques, etc.) ; le problème de compétence renvoie à la faiblesse sur le plan des opérations mentales du niveau supérieur telles que la prédiction, l'analyse, le résumé et l'inférence, qui sont supposées fonctionner de manière universelle dans toutes les langues (Yamashita, 2001a).

La question d'Alderson (1984) renvoie à deux avis opposés à l'époque :

³ « Reading in a foreign language: a reading problem or a language problem? »

certains chercheurs pensaient que les lacunes lexicales importantes en L2 étaient la cause principale d'une mauvaise compréhension en lecture (Yorio, 1971), tandis que d'autres prétendaient que les apprenants doivent d'abord améliorer leurs habitudes de lecture en L1 pour avoir de meilleures compétences en lecture et mieux comprendre les textes en L2 (Coady, 1979). Après avoir fait une revue des études menées sur ce sujet, Alderson (1984) a conclu qu'une faible performance de compréhension en lecture en L2 résulte essentiellement de deux problèmes : le niveau linguistique en L2 est la cause majeure pour les apprenants de faible niveau en L2, tandis que la compétence en lecture contribue davantage à la performance de compréhension quand il s'agit des apprenants de niveau avancé. Allant plus loin, Alderson a émis deux hypothèses pour expliquer une faible performance de compréhension en lecture en L2 : 1) les connaissances et les compétences exigées pour la compréhension dans la langue cible ne sont pas les mêmes que celles pour la compréhension écrite en L1 ; 2) un niveau linguistique limité dans la langue cible empêche le lecteur de se servir correctement de ses connaissances et de ses compétences en lecture acquises en L1 dans la compréhension en L2 ; autrement dit, il y a un seuil linguistique à atteindre en L2 pour que les transferts, c'est-à-dire la capacité d'apprendre de nouvelles compétences en faisant appel aux ressources précédemment acquises (Genesee & Nicoladis, 2006, cité par Koda, 2007), depuis la L1 vers la L2 soient réalisés. (Cf. Hudson, 2007, Chapitre 3)

Selon Hudson (2007), parmi les compétences en lecture acquises en L1 et transférables vers la L2 figure la compétence de décodage (« decoding skill »). Il n'est pas difficile de comprendre que, pour décoder un mot français et un autre mot d'une langue entièrement non alphabétique comme le chinois, la compétence de décodage ne soit pas la même. En effet, les résultats des études ont montré que, par rapport aux apprenants dont la L1 est alphabétique, les apprenants dont la L1 est entièrement non alphabétique (par exemple la langue chinoise) ou partiellement alphabétique (par exemple la langue japonaise) sont moins performants pour décoder des mots L2 constitués d'une série de lettres alphabétiques (Akamatsu, 1999; Hamada & Koda, 2010). Ces résultats sont en accord avec la première hypothèse d'Alderson (1984) selon laquelle certaines compétences ne sont pas transférables à cause de la

distance typologique importante entre la L2 et la L1 (Nous reviendrons sur ce point dans la Section 1.3.3.1).

Au sujet du seuil linguistique en L2, les résultats de l'étude de Lee & Schallert (1997) ont fourni une preuve pour la deuxième hypothèse d'Alderson (1984). L'étude de Lee & Schallert (1997) a été menée auprès de lycéens coréens de différents niveaux en anglais L2. Les résultats montrent que seulement pour les lycéens ayant un meilleur niveau en anglais, les scores de la compréhension écrite en L2 corrélaient positivement avec ceux de la compréhension en L1, mais pas pour les autres lycéens ayant un niveau inférieur en anglais. Cela semble valider l'idée que c'est seulement au-delà d'un certain seuil linguistique en L2 qu'un apprenant parvient à employer de façon appropriée ses compétences en lecture acquises en L1 transférables vers la L2. En effet, de nombreuses études réalisées plus tard arrivent à une conclusion commune sur l'existence d'un seuil linguistique en L2 en dessous duquel les connaissances acquises en L1 ne peuvent pas être correctement utilisées pendant la lecture en L2 (ex. Haastруп, 1991; D. D. Qian, 2005; Théophanous & Dreyfus, 2004).

Concrètement dans une lecture en L2, quel est le seuil de connaissance des mots présents pour avoir une compréhension globale ? Pour Laufer (1992), en anglais L2, une connaissance de 95% des mots courants d'un texte est nécessaire pour qu'un apprenant ait une compréhension correcte. Proposant ce seuil à 98%, Hu & Nation (2000) mettent en avant que le pourcentage des mots connus dans un texte par un apprenant peut prédire la compréhensibilité du texte pour l'apprenant : si un apprenant peut comprendre 100% d'un texte qui ne contient aucun mot inconnu, plus le pourcentage des mots inconnus présents dans le texte augmente, moins l'apprenant arrive à avoir une compréhension appropriée du texte ; quand le manque de connaissance du vocabulaire atteint le niveau de 20% dans un texte, aucun apprenant ne parvient à comprendre le texte. Les chercheurs expliquent que cela est lié au fait que davantage de mots connus dans un texte représente un plus grand soutien informatif du contexte autour d'un mot inconnu et, par conséquent, plus de possibilités pour l'apprenant d'élaborer une compréhension globale du texte malgré les mots inconnus. Schmitt et al. (2011), d'après leur étude auprès d'apprenants de divers niveaux en anglais L2, ont également proposé le seuil de 98% et ont d'ailleurs

souligné que le thème du texte doit être pris en compte dans la discussion au sujet du seuil du vocabulaire et la compréhensibilité du texte. De nos jours, sans avoir un seuil lexical standard pour toutes les langues, il est généralement admis qu'une connaissance de 90%-95% des mots dans un texte est nécessaire pour le comprendre (Hirsch, 2003). Du point de vue de la charge cognitive, une bonne connaissance des mots présents dans un texte peut soulager cognitivement l'apprenant de ses traitements lexicaux (niveau bas) pour qu'il puisse alors s'investir davantage dans les autres opérations mentales (niveau haut) (Pulido, 2004, 2007).

Certains chercheurs s'intéressent à l'effet du type de la tâche à réaliser dans la compréhension en lecture en L2. Ils sont d'avis que différentes tâches ont des objectifs divers à atteindre et ces objectifs peuvent décider ce qui a un rôle plus important dans la compréhension écrite, la compétence en lecture en L1 ou le niveau linguistique en L2 (Alptekin & Erçetin, 2010 ; Taillefer, 1996).

Malgré de nombreuses études réalisées sur la relation entre la compétence en lecture en L1, celle en L2 et les connaissances linguistiques en L2, il n'existe toujours pas à nos jours une réponse nette et définitive à la question d'Alderson (1984). Cela étant dit, en L2 comme en L1, dans la compréhension écrite, le lecteur doit traiter à la fois les informations extraites des signes imprimés de divers niveaux (du mot, de la phrase jusqu'au texte entier) et celles de ses connaissances disponibles. Ces deux sources demeurent essentielles dans les activités de compréhension en lecture.

1.2 Stratégies de lecture

Comme rappelé dans les sections précédentes, dans le processus de la compréhension écrite, le lecteur doit rester actif et proactif pour traiter les informations provenant de diverses sources. Selon Koda (2005), pour mieux réussir sa compréhension, un lecteur compétent s'adapte continuellement pendant sa lecture aux facteurs contextuels variables, par exemple le niveau de difficulté du texte, l'objectif ciblé de cette lecture, etc. Cela dit, le lecteur peut effectuer une lecture stratégique.

Garner (1987, cité par Mokhtari & Kouider, 2002, p. 250) définit les stratégies de lecture comme « *les activités généralement délibérées et planifiées entreprises par des apprenants actifs, souvent pour remédier aux échecs cognitifs* »⁴ (p. 50) ». Carrell (1998), pour sa part, insiste également sur l'aspect métacognitif du terme « stratégie » qui implique une action activement sélectionnée et contrôlée par le lecteur pour atteindre les objectifs désirés.

Paris et al. (1991) regroupent les stratégies de lecture en fonction du moment d'utilisation, c'est-à-dire avant, pendant et après la lecture :

- Avant de lire un texte, le lecteur peut choisir un environnement physique approprié pour la lecture pour s'installer, définir l'objectif de la lecture, essayer de deviner le sujet traité dans le texte par le biais du titre, identifier la structure et le style du texte, etc. ;
- Pendant la lecture, les stratégies utilisées peuvent inclure l'identification de l'idée principale du texte, l'inspection des connaissances du vocabulaire et la mise en place de l'inférence, la prédiction du sens de chaque paragraphe et la relecture ou le balayage du texte pour trouver une idée générale ou spécifique ;
- À l'issue de la lecture, le lecteur peut évaluer sa compréhension du texte et décider s'il est nécessaire de le relire. Le lecteur peut aussi synthétiser et émettre des appréciations sur le texte et/ou sur le style de l'auteur.

Parmi toutes les stratégies de lecture, l'inférence est une stratégie de lecture majeure (Anderson & Pearson, 1984).

En termes psychologiques, Hamel & Joannette (2007) indiquent que la définition la plus traditionnelle de l'inférence est « *une opération logique portant sur des propositions tenues pour vraies (les prémisses) et concluant à la vérité d'une nouvelle proposition en vertu de sa liaison avec les premières.* (p. 12) » D'un point de vue plus pragmatique, Chernyshova & Traverso (2017) considèrent l'inférence comme l'acte d'« *...interpréter les éléments non littéraux, non-dits, ou encore implicites* (p. 2) ». Harris & Monaco (1978, cité par Hamel & Joannette, 2007)

⁴ « ...generally deliberate, playful activities undertaken by active learners, many times to remedy perceived cognitive failure. »

proposent pour leur part deux types d'inférence – l'inférence logique et l'inférence pragmatique : l'inférence logique renvoie à des déductions à partir des prémisses déterminées (par exemple, A est supérieur à B et C est inférieur à B. On en déduit alors que A est supérieur à C.) ; l'inférence pragmatique est le résultat de l'association du contexte et des connaissances d'une personne (par exemple, « La jeune fille monte dans la Rolls Royce ; la jeune fille est orpheline ». On en déduit donc que « La jeune fille a hérité. ») (cf. Hamel & Joannette, 2007, p. 13).

Selon Noordman & Vonk (2015), il y a deux façons de procéder l'inférence, l'activation des connaissances déjà disponibles ou la dérivation de nouvelles informations. L'inférence a donc à la fois la fonction explicative et la fonction interprétative dans une communication. L'exemple de Moescheler (2001, p. 89) ci-dessous montre concrètement la fonction interprétative de l'inférence dans la communication :

(Contexte : Il est huit heures du soir, heure à laquelle les enfants de la famille M se couchent.)

Le père à son fils Axel : Va te laver les dents.

Axel : Je n'ai pas sommeil.

Sous la condition que le fils soit conscient que 1) il fait nuit et 2) avant se coucher il doit se laver les dents, le fils décrypte le message de son père sans que ce dernier prononce un seul mot littéralement lié au sens de « se coucher ».

La compréhension écrite peut être considérée comme la communication à sens unique entre l'auteur et le lecteur, ce dernier essayant de déchiffrer les messages envoyés par ce premier. Ce déchiffrement du message implique non seulement les informations objectives fournies par le texte, mais également les connaissances du lecteur. Il en résulte que le lecteur parvient à comprendre des détails littéralement présentés dans le texte (le message explicite) et d'autres informations délivrées implicitement (Kintsch, 1998). Anderson & Pearson (1984), dans le but d'expliquer la théorie des schémas (« Schema theory ») sur le processus de la compréhension écrite, indiquent que différentes sortes d'inférence peuvent intervenir à diverses étapes du processus : l'inférence peut aider le lecteur à élaborer une prédiction globale d'un texte, à saisir des informations clés même si celles-ci ne sont pas explicitement

dites par l'auteur et à organiser les informations explicites et implicites selon une certaine logique afin d'avoir une compréhension intégrale d'un texte.

1.3 Inférence lexicale, une stratégie importante dans la compréhension écrite

Pendant la compréhension d'un texte, l'inférence peut assister le lecteur non seulement au niveau de la cohérence globale (« connexion du texte » ou « text-connecting »), mais également au niveau de la cohésion locale (« comblement de lacune » ou « gap-fill ») (Cain & Oakhill, 2014 ; Graesser et al., 1994).

Comme l'indiquent Anderson & Pearson (1984), l'inférence comprend l'acte de « ... *tirer une conclusion basée sur le manque de connaissance* (p. 269)⁵ ». Un manque de connaissance locale peut être un mot présent dans le texte que le lecteur ne connaît pas. Cela veut dire que l'inférence peut intervenir non seulement au niveau du texte, mais également au niveau du mot (Kintsch, 1998). Comme le dit Carton (1971, cité par Hastrup, 1991, p. 22), « *Dans l'inférence, les attributs et les contextes familiers sont utilisés pour reconnaître ce qui n'est pas familier.* (p. 45)⁶ » Quand le lecteur utilise l'inférence pour remédier à un manque de connaissance lexicale, l'inférence se réfère à l'acte de repérer des informations caractéristiques de ce mot inconnu⁷ afin de deviner un sens logique pour assurer la continuité d'un circuit de compréhension. Il s'agit alors de la stratégie d'inférence lexicale.

Hastrup (1991) définit l'inférence lexicale dans la compréhension écrite comme « *l'acte de deviner le sens d'un mot à la lumière des indices linguistiques disponibles, en lien avec les connaissances du monde de l'apprenant, de sa perception du co-texte et de ses connaissances linguistiques* (p. 13)⁸ ». Cette stratégie compense, plus ou moins, la lacune lexicale afin d'assurer une compréhension plus fluide, logique et complète du texte. Pour cette raison-là,

⁵ « ... draw a conclusion based upon lack of knowledge ».

⁶ « In inferencing, attributes and contexts that are familiar are utilized in recognizing what is not familiar. »

⁷ Un autre terme « mot non familier » est également utilisé dans les littératures. Dans notre thèse, nous employons le terme « mot inconnu » qui renvoie aux mots partiellement ou entièrement inconnus à un apprenant.

⁸ «...making informed guesses as to the meaning of a word in the light of all available linguistic cues in combination with the learner's general knowledge of the world, her awareness of the co-text and her relevant linguistic knowledge. »

Haastrup (2008, p. 70) considère également l'inférence lexicale comme une sous-catégorie de l'inférence du texte.

Nous avons discuté dans les sections précédentes que le vocabulaire est l'un des facteurs fondamentaux dans la compréhension écrite. Toutefois, pratiquement personne ne parvient à maîtriser de façon exhaustive le vocabulaire d'une langue, même de sa langue maternelle. Il n'est donc pas rare qu'un lecteur tombe sur un ou plusieurs mots inconnus dans sa lecture. A la rencontre d'un mot inconnu, le lecteur peut consulter un dictionnaire ou demander de l'aide auprès d'une personne à proximité. Mais ces moyens ne sont pas toujours disponibles, ni pratiques. Dans ce cas-là, l'inférence lexicale peut aider le lecteur à construire une « passerelle » au-dessus d'une lacune lexicale afin d'obtenir une cohésion locale pour assurer une compréhension fluide et une lecture en continuité.

Il est à noter que l'inférence lexicale s'applique également dans la compréhension orale. Cependant, par rapport à la lecture, les mots inconnus sont moins faciles à repérer à l'oral, parce que dans les communications orales les messages se délivrent vite sans donner aux auditeurs la possibilité de retourner aux parties précédentes. L'accent du locuteur peut aussi rendre compliqué le décodage auditif de l'auditeur. Pour cela, l'auditeur accepte de n'avoir qu'une compréhension globale du message oralement délivré. D'autant plus qu'à l'oral, les frontières entre les mots sont bien moins explicites qu'à la lecture. (Guo, 2015 ; Van Zeeland, 2014) Même si un mot inconnu est identifié, dans la compréhension orale, l'auditeur ne peut compter que sur les informations retenues dans sa mémoire à court terme pour essayer de comprendre le sens du mot. Des chercheurs sont même d'avis que la compréhension orale et l'inférence lexicale ne peuvent pas s'opérer en même temps à cause de la concurrence des ressources cognitives nécessaires pour les deux opérations (Ridgway, 2000, cité par Guo, 2015). D'ailleurs, Guo (2015) remarque dans son étude auprès des apprenants francophones de chinois en France que, quand la compréhension se déroule dans un contexte visio-audio (par exemple la compréhension d'un morceau de série télévisée), le résultat de l'inférence lexicale est influencé par les indices visuels non lexicaux tels que les expressions faciales et les gestes des personnages.

De ce fait, l'inférence lexicale est particulièrement intéressante dans la compréhension écrite. Selon Wesche & Paribakht (2009), la compréhension écrite offre aux apprenants le plus d'occasions d'exposition à des mots nouveaux.

1.3.1 Processus d'inférence lexicale

En se basant sur le modèle de la production et de la compréhension orale en L1 de Levelt (1989, 1993), le psycholinguiste De Bot et ses collègues linguistes Paribakht et Wesche (1997) ont essayé d'expliquer le processus de l'inférence d'un mot écrit sous l'angle cognitif au niveau du lexique mental (c'est-à-dire le réseau dans la mémoire d'un apprenant dans lequel tous les mots acquis sont liés les uns avec les autres avec des connexions plus ou moins fortes (Théophanous, 2019)).

De Bot et al. (1997, pp. 311-312) expliquent que le modèle de Levelt (1989, 1993) comprend trois niveaux distincts, le niveau conceptuel, le niveau du lemme et le niveau du lexème. Selon l'hypothèse lexicale (« Lexical hypothesis ») de Levelt (1989), le lexique mental a un rôle de médiateur entre la conceptualisation et le codage ou le décodage (respectivement pour la production et la compréhension) grammatical et phonologique. Dans un processus de production, une fois que les informations à envoyer sont sélectionnées, le message conçu passe du composant de conception (« conceptualization ») au composant de formulation (« formulator »). Dans ce dernier, les mots et les règles grammaticales appropriées sont sélectionnées depuis le lexique mental pour formuler le message. Pour cette étape, au moins quatre éléments lexicaux sont concernés – sémantique, syntaxique, morphologique et phonologique. Ces éléments sont respectivement représentés à deux niveaux dans une entrée lexicale, le lemme et le lexème. Le premier inclut les informations sémantiques et syntaxiques et le dernier contient la forme morphophonologique correspondante.

A partir de cela, De Bot et al. (1997) proposent pour leur part le modèle de la compréhension et de la production lexicale pour les modalités écrite et orale, synthétisé dans la Figure 1.1 plus bas. Selon De Bot et ses collègues, dans la compréhension écrite, la compréhension d'un mot écrit demande au lecteur de créer un lien entre la forme du mot imprimé et un concept enregistré dans sa mémoire : une

série de lettres alphabétiques perçues doit être associée à un lexème ; quand l'appariement est fait, la forme active un lemme correspondant qui est lié à un concept. La compréhension du mot s'accomplit seulement si cette création du lien aboutit avec succès. En d'autres termes, une reconnaissance de mot qui échoue peut être causée par une ou plusieurs défaillances à n'importe quel niveau dans ce lien. L'inférence lexicale peut alors être analysée depuis l'angle de la défaillance d'un maillon connectant les connaissances de deux niveaux adjacents. L'inférence du sens d'un mot inconnu est en effet un processus qui consiste à collectionner diverses sortes d'indices afin d'avoir suffisamment d'informations pour passer d'une étape à la suivante dans le lien forme-concept.

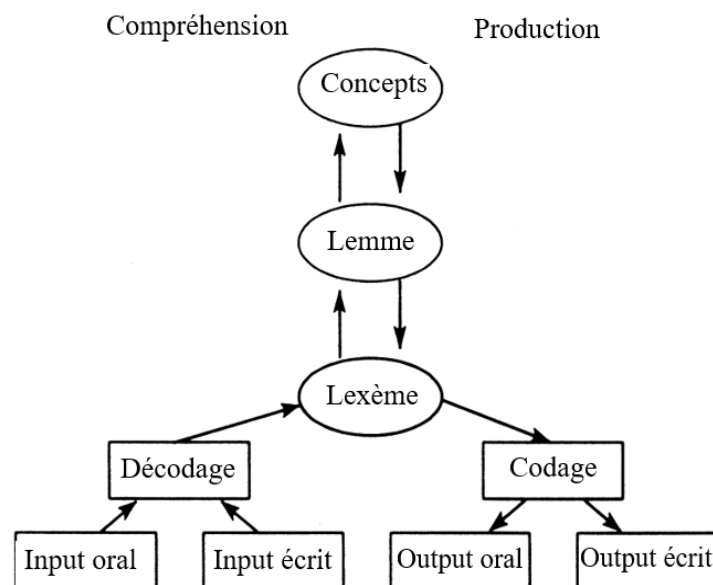


Figure 1.1 – Modèle de la compréhension/production lexicale pour les modalités orales et écrites (De Bot et al., 1997, p. 315)

1.3.2 Inférence lexicale en L1

L'inférence lexicale a d'abord attiré l'attention des chercheurs en L1. On a remarqué que, sans instructions ni aide externe, les enfants natifs parviennent à comprendre des textes dont tous les mots ne sont pas dans le périmètre de leur vocabulaire. Par exemple, dans une ancienne étude, Werner & Kaplan (1952) ont montré aux enfants âgés de 9 à 13 ans une série de 6 phrases qui fournissent différents

contextes informatifs pour un même pseudo mot qui est intégré dans ces phrases. Au total 12 pseudo mots ont été testés. Avec des degrés de réussite différents, les enfants de tous âges sont parvenus à comprendre des mots.

Pour Carnine et al. (1984), à la rencontre d'un mot inconnu dans un texte, deux stratégies majeures sont à la disposition de l'apprenant : l'une est la consultation du dictionnaire et l'autre « l'analyse du contexte » (« contextual analysis ») qui consiste à « *déterminer le sens d'un mot inconnu sur la base d'informations contextuelles ou d'indices contextuels déjà contenus dans la partie de la lecture sélectionnée* (p. 189)⁹ ». Selon Carnine et al. (1984), l'avantage de cette dernière stratégie est que son utilisation est indépendante 1) de la disponibilité d'un dictionnaire ou de la capacité de l'apprenant de se servir d'un dictionnaire et 2) des instructions spécifiques d'un enseignant sur le texte concerné. D'ailleurs, l'analyse du contexte permet à l'apprenant de maintenir son attention sur le texte sans être interrompu. Mais ces avantages sont valables sous condition que le contexte d'un mot inconnu fournisse suffisamment d'informations sur le sens du mot. Afin de comprendre l'effet des indices contextuels sur le résultat de la déduction du sens d'un mot inconnu, les chercheurs ont mené une étude auprès des enfants natifs anglais de 10 à 12 ans. Les trois éléments concernés sont le type de l'indice fourni par le contexte (par exemple un synonyme, un antonyme ou un ensemble de mots qui permet à l'apprenant de déduire le sens du mot), la proximité de l'indice contextuel par rapport au mot inconnu et le niveau linguistique de l'apprenant. Pendant l'expérience, on a demandé à ces enfants de donner le sens des mots anglais non fréquents, présentés isolément ou intégrés dans un passage. Les résultats montrent que 1) les enfants ont mieux réussi la déduction du sens des mots inconnus intégrés dans un passage (donc dans un contexte plus riche). Cela signifie que le contexte a un effet bénéfique à la compréhension d'un mot inconnu ; 2) les enfants ont davantage repéré les indices plus proches du mot inconnu, par rapport aux indices plus éloignés ; 3) parmi les indices contextuels de différents types, le synonyme du mot inconnu est le plus saillant pour les apprenants ; et 4) dans tous les cas, les apprenants plus âgés

⁹ « ...ascertain the meaning of an unfamiliar word based on contextual information or context clues already contained the reading selection. »

réussissent mieux la déduction du sens d'un mot inconnu, par rapport aux plus jeunes. En résumé, un contexte informatif avec des indices saillants et explicites a un effet positif sur l'inférence du sens d'un mot inconnu ; par ailleurs, le taux de réussite d'inférence augmente positivement avec le niveau linguistique de l'apprenant.

L'étude de Kaye et al. (1987) a été réalisée auprès d'apprenants un peu plus âgés, 58 collégiens et lycéens et 50 étudiants universitaires, dans le but d'observer chez les apprenants adolescents et adultes l'utilisation de la stratégie de décomposition lexicale (« lexical decomposition strategy », c'est-à-dire la capacité de déduire le sens d'un mot inconnu à partir des composants du mot, c'est-à-dire à partir du préfixe et du radical). On a présenté aux participants 58 mots, constitués d'un préfixe et d'un radical, et demandé aux participants de choisir pour chaque mot cible la définition correcte parmi quatre propositions. Les participants ont aussi répondu à des questions qui mesurent leurs connaissances au niveau métacognitif pour les mots cibles, les préfixes et les radicaux concernés (par exemple, noter sur une échelle de sept niveaux le niveau de ressemblance à la forme d'un mot cible à un autre mot que les participants connaissent). Les résultats montrent que les apprenants s'appuient davantage sur le sens des préfixes, par rapport aux radicaux, pour comprendre un mot inconnu ; les apprenants avec une compétence linguistique supérieure (en l'occurrence, les étudiants universitaires) sont plus capables d'extraire des informations au sein d'un mot inconnu pour en déduire le sens. En plus, tous les apprenants possèdent des connaissances métacognitives correctes, mais l'analyse des composants d'un mot inconnu pour deviner son sens n'est pas un acte systématique.

Le psychologue Sternberg et ses collègues (1982) ont étudié la capacité de compréhension verbale (« verbal comprehension », c'est-à-dire la capacité à comprendre un matériel linguistique tel qu'un magazine, un journal, un manuel pédagogique, etc.) en tant que partie intégrale de l'intelligence. Ils ont essayé de mettre en lumière les éléments qui peuvent intervenir lorsqu'un apprenant déduit le sens d'un mot inconnu via le contexte¹⁰. Selon ces chercheurs, on peut classer les

¹⁰ Le terme utilisé par Sternberg et al. (1982) est « learning from context », « apprendre grâce au contexte » en français, qui renvoie à l'acte de raisonner le sens d'un mot inconnu à l'aide du contexte constitué des mots connus dans lequel se trouve le mot inconnu (« ...learning meanings of single words embedded in large numbers of words whose meanings are known. (p. 181) ») Afin de ne pas confondre ce terme avec celui

indices contextuels en deux catégories : le contexte interne (les indices se situant au sein du mot, à savoir les préfixes, les affixes, etc.) et le contexte externe (les indices se situant au tour du mot dans le texte). Afin d'essayer d'expliquer pourquoi 1) certains mots ou concepts sont plus faciles à comprendre par rapport aux autres et 2) certains apprenants réussissent mieux la déduction de sens par rapport aux autres, les chercheurs ont proposé deux modèles : le modèle de compétence (« competence model ») qui spécifie les types d'indices disponibles dans le texte que l'apprenant peut utiliser pour inférer le mot inconnu à partir du contexte, et le modèle de performance (« performance model ») catégorisant les variables médiatrices qui ont un impact, positif ou négatif, sur l'utilisation des indices dans une situation donnée. On peut représenter de façon synthétique la théorie de Sternberg et al. (1982) par une matrice de 2 x 2, telle que dans le Tableau 1.1 plus bas.

A partir des résultats des études menées à ce sujet, les chercheurs sont aujourd'hui unanimement d'accord pour dire que les apprenants natifs sont capables de déduire le sens d'un mot inconnu rencontré dans leur lecture grâce au contexte. Les indices sur lesquels ils peuvent s'appuyer sont de divers types présents à différents niveaux du texte. Les éléments tels que la compétence linguistique, la représentation et l'emplacement de l'indice, ainsi que les connaissances métacognitives peuvent intervenir sur le résultat de l'inférence lexicale.

1.3.3 Inférence lexicale en L2

Un apprenant a normalement des lacunes lexicales plus importantes dans la L2 que dans la L1. L'inférence lexicale a donc un grand intérêt pour un apprenant L2. En effet, de nombreuses études ont montré que l'utilisation de l'inférence lexicale n'est pas rare chez les apprenants L2.

d'« apprentissage » sur lequel nous discuterons dans le chapitre suivant, nous traduisons ici le terme de Sternberg et al. (1982) comme « déduction du sens via le contexte ».

	Contexte externe	Contexte interne
<i>Modèle de compétence</i>	<ol style="list-style-type: none"> 1) Indices temporels (« temporal clues », la durée ou la fréquence d'occurrence du mot) 2) Indices spatiaux (« spatial clues », l'emplacement du mot) 3) Indices de valeur (« value clues », l'importance du sens du mot dans la compréhension du texte) 4) Indices descriptifs de la propriété (« stative descriptive clues », la description des propriétés telles que la couleur et la forme) 5) Indices descriptifs de la fonction (« functional descriptive clues », les informations concernant l'objectif, l'action ou l'utilisation possible) 6) Indices de cause ou de conditions favorables (« causal/enablement clues ») 7) Indices d'appartenance à une classe (« class membership clues ») 8) Indice d'équivalence (« equivalence clues », les informations sur le sens du mot) 	<ol style="list-style-type: none"> 1) Le préfixe 2) Le radical (« stem clue ») 3) Le suffixe 4) L'indice interactif (une combinaison des informations envoyées par plusieurs sortes d'indices mentionnées ci-dessus)
<i>Modèle de performance</i>	<ol style="list-style-type: none"> 1) Le nombre d'occurrence du mot inconnu 2) La variété des contextes dans lesquels se présente le mot inconnu 3) L'importance du sens du mot inconnu pour comprendre le contexte dans lequel il se trouve 4) Les informations utiles fournies par le contexte pour comprendre le sens du mot inconnu 5) La densité de mots inconnus présents dans le contexte 6) Le niveau de concrétude du sens du mot inconnu et du contexte 7) L'utilité des connaissances pré-acquises dans l'utilisation des indices 	<ol style="list-style-type: none"> 1) Le nombre d'occurrence du mot inconnu 2) L'importance du sens du mot inconnu pour comprendre le contexte dans lequel il se trouve 3) La densité de mots inconnus présents dans le contexte¹¹ 4) La densité de mots inconnus décomposables 5) L'utilité des connaissances pré-acquises dans l'utilisation des indices

Tableau 1.1 – Théorie de la déduction du sens d'un mot inconnu via le contexte (extraite de Sternberg et al., 1982, pp. 163-169)

¹¹ Une densité importante de mots inconnus dans le contexte rend plus difficile la compréhension et peut aboutir à un processus de compréhension mot par mot. Dans ce cas-là, davantage d'attention de l'apprenant est donc orientée vers les indices internes disponibles pour le mot inconnu.

D. D. Qian (2004), par exemple, a observé l'emploi des stratégies de lecture auprès des apprenants universitaires de l'anglais L2. Ses résultats montrent qu'à côté de la consultation d'un dictionnaire et des renseignements pris auprès de quelqu'un, l'inférence lexicale est la stratégie de lecture la plus utilisée par les apprenants L2 à la rencontre d'un mot inconnu. Les résultats de D. D. Qian (2004) confirment les conclusions de De Bot et al. (1997), de Fraser (1999) et de Paribakht & Wesche (1999) qui indiquent que l'inférence lexicale est la stratégie principale que les apprenants L2 utilisent pour initier leur compréhension des mots inconnus rencontrés dans une lecture.

Depuis la fin des années 1990, des études ont examiné l'inférence lexicale en L2 sous trois aspects principaux, le processus de l'inférence lexicale, les sources de connaissances impliquées pendant l'inférence lexicale et le taux de réussite (Haastrup, 2008).

En anglais L2, Nassaji (2004) a observé des apprenants de différentes L1 dans une tâche d'inférence lexicale et il a remarqué que les apprenants utilisent trois stratégies principales pour inférer un mot inconnu : l'identification (trouver des informations au sujet d'un mot inconnu), l'évaluation (vérifier les informations repérées et définir le sens) et la supervision (surveiller le sens inféré au fur et à mesure de la lecture). Fraser (1999) avance pour sa part que lors de l'inférence lexicale, les apprenants utilisent deux méthodes principales – l'identification du mot (« word identification ») et la création d'un sens (« sense creation ») : à la rencontre d'un mot inconnu, l'apprenant cherche d'abord à associer le mot avec un mot connu en L1 ou en L2 ; à défaut de celle-ci, l'apprenant élabore un nouveau sens à partir de divers indices repérés.¹²

Haastrup (1991) a proposé dans son étude une typologie de sources de connaissance en trois catégories, 1) les connaissances intralinguistiques (ou les connaissances linguistiques acquises dans la langue cible), 2) les connaissances interlinguistiques (c'est-à-dire les connaissances linguistiques acquises en L1 ou dans

¹² Cf. Noordman & Vonk (2015) pour les explications sous l'angle de vue psychologique des deux types d'inférence – l'inférence dérivant de nouvelles informations et celle activant des connaissances disponibles.

d'autres langues non L2), et 3) le contexte et les connaissances encyclopédiques. Selon Haastrup, pour inférer le sens d'un mot inconnu, les indices de ces trois catégories fonctionnent en interaction. L'étude de Haastrup (1991) a sollicité des collégiens et lycéens danois qui apprennent l'anglais, les seconds étant supposés avoir acquis un niveau supérieur en anglais à celui des premiers. L'objectif principal de cette étude est d'examiner le rôle du niveau en L2 de l'apprenant dans le processus d'inférence lexicale. La chercheuse a remarqué que tous les apprenants s'appuient sur une grande variété de sources de connaissance pour inférer un mot inconnu ; ayant un meilleur score d'inférence lexicale, les apprenants d'un niveau supérieur en anglais se servent davantage de connaissances intralinguistiques à tous niveaux (à savoir mot, phrase et discours) et inter-linguistiques, tandis que ceux d'un niveau moins avancé se contentent plus d'indices au niveau du mot. A partir de ses résultats, Haastrup propose un cadre de processus en continuum à sept niveaux, avec le niveau le plus bas concernant exclusivement des informations dérivant du mot et le niveau le plus haut qui est une intégration des indices de toutes sortes. Un apprenant de niveau inférieur n'arrive à traiter que les informations au niveau d'un mot (« Bottom-level processing »). Plus son niveau est élevé, plus il est capable d'intégrer les indices de toutes sortes en interaction (« Top-level processing »).

Morrison (1996) a également remarqué une différence au niveau du choix de sources de connaissance entre les apprenants de divers niveaux en français L2. La phrase dans laquelle se trouve le mot inconnu étant la source principale pour les apprenants de tous niveaux, les apprenants d'un niveau supérieur s'appuient sur un ensemble de sources de connaissance plus variées et réussissent mieux l'inférence des mots inconnus. Il est à noter que Morrison conclut de ses résultats que l'utilisation de l'inférence lexicale est spontanée et naturelle chez les apprenants L2. La conclusion de l'étude de Théophanous & Dreyfus (2004) réalisée auprès des jeunes arabophones dans un contexte de scolarisation en France semble montrer en revanche que ces jeunes apprenants n'utilisent pas l'inférence lexicale de façon spontanée.

L'étude de Mori (2003) en japonais L2 a examiné les rôles des indices au sein d'un mot inconnu et dans la phrase comportant le mot en question. Mori a sélectionné 45 mots cibles en Kanji, caractères chinois utilisés dans la langue japonaise à l'écrit,

qui sont jugés transparents sémantiquement (c'est-à-dire qu'on parvient à deviner le sens des mots à partir des significations des Kanjis constitutifs, par exemple, 外 (extérieur) + 見 (voir) → 外見 (apparence)). 74 apprenants anglophones en japonais L2 ont été sollicités pour inférer les sens des 45 mots cibles au total présentés sous trois conditions : 1) condition « Mot seul » dans laquelle les mots cibles sont présentés isolément, 2) condition « Contexte seul » (une phrase qui fournit des informations contextuelles dans laquelle le mot cible est absent) et 3) condition « Mot contextualisé » dans laquelle un mot cible est présenté dans une phrase qui fournit des informations contextuelles. Pour les mots cibles présentés dans les conditions « Mot seul » et « Mot contextualisé », les participants doivent deviner le sens des mots ; dans la condition « Contexte seul », les participants donnent un mot ou une petite phrase en anglais qui complètent le mieux la phrase lacunaire. Les résultats montrent que, sans différence significative entre les conditions à un seul indice (les conditions « Mot seul » et « Contexte seul »), les apprenants ont de meilleurs résultats d'inférence à l'aide de la combinaison des deux indices (la condition « Mot contextualisé »).

Atef-Vahid et al. (2013) ont examiné pour leur part le rôle des connaissances sur le thème du texte et l'effet de la densité du vocabulaire visuel (ou « sight vocabulary » qui renvoie aux mots dont la forme et le sens général sont reconnus instantanément, quel que soit le contexte) présent dans le texte sur le résultat de l'inférence lexicale. Leur étude est réalisée auprès de 114 étudiants universitaires persanophones qui apprennent l'anglais. À partir de leurs résultats, les chercheurs sont arrivés à la conclusion que les connaissances sur le thème du texte ont un effet considérable sur le résultat de l'inférence lexicale ; par ailleurs, les apprenants réussissent mieux l'inférence pour les mots inconnus intégrés dans un texte comportant davantage de vocabulaire visuel.

Les résultats des études réalisées en L2 s'accordent globalement avec celles obtenues en L1. En résumé, pour inférer un mot inconnu rencontré dans une lecture en L2,

– à l'exception des résultats de l'étude de Bensoussan & Laufer (1984) qui ne

montrent pas d'effet significatif du niveau linguistique en L2 de l'apprenant sur le résultat de l'inférence lexicale, la grande majorité des études ont validé l'idée qu'il existe un seuil lexical en L2 à atteindre afin que l'utilisation de l'inférence lexicale soit efficace pour l'inférence des mots inconnus (Dolgunsöz, 2016; Haynes, 1993; Théophanous & Dreyfus, 2004; M. J. Wu, 2005) ; plus le niveau est avancé dans la langue cible, mieux l'apprenant parvient à repérer et à traiter des indices utiles pour réussir son inférence lexicale (Dong, 2007; Gan, 2008; Y. Huang, 2008; S. H. Liu, 2001; Morrison, 1996; Nagy et al., 1997; Y. Zhu & Cui, 2005) ; un apprenant d'un niveau linguistique supérieur a tendance à intégrer des indices de diverses sortes pour inférer un mot, tandis qu'un apprenant d'un niveau moins avancé se contenterait plus d'extraire des informations lexicales au niveau du mot (Bengeleil & Paribakht, 2004; Dong, 2007; Haastrup, 1991; Nassaji, 2004; D. D. Qian, 2005) ;

- les apprenants s'appuient sur diverses sources de connaissance pour repérer des indices lorsqu'ils infèrent le sens d'un mot inconnu (Jelić, 2007; Nassaji, 2003; Paribakht & Wesche, 1999; X. J. Qian, 2005; Rousoulioti & Mouti, 2016; M. J. Wu, 2005) ;
- un contexte, interne ou externe (cf. le Tableau 1.1, supra p. 27), riche en information favorise la déduction du sens du mot (Dong, 2007; H. Fan, 2012; Y. X. Fang & Jiang, 2012; Gan, 2008; Mori, 2003; Y. H. Xu, 2014b; H. M. Zhang et al., 2019; J. L. Zhang, 2010; X. Y. Zhu & Zhou, 2007; Y. Zhu & Cui, 2005) ;
- le type de texte (par exemple narratif ou argumentatif) peut aussi influencer la performance d'inférence du sens des mots inconnus présents dans le texte (Shen, 2017) ;
- les connaissances non linguistiques de l'apprenant ont une influence importante sur le résultat d'inférence. Une meilleure connaissance du thème du texte faciliterait les inférence des mots inconnus (Atef-Vahid et al., 2013; Bengeleil & Paribakht, 2004; Haastrup, 1991; Pulido, 2004, 2007; Wesche & Paribakht, 2009).

1.3.3.1 Rôle de la distance typologique L1-L2 dans l'inférence lexicale en L2

Nous avons mentionné dans les sections précédentes que les connaissances acquises en L1 restent actives pendant les activités en L2 (Koda, 2005). Par exemple, dans la production orale, Théophanous & Perez-Bettan (2021) ont constaté des erreurs de transfert commises par les apprenants anglophones de français L2 sous l'influence de l'usage collocationnel en anglais, par exemple « *Il a pris une sieste.* » (au lieu de « *Il a fait une sieste.* » parce qu'en anglais on dit « *take a nap* ».) et « *Il est sommeil.* » (au lieu de « *Il a sommeil.* » parce qu'en anglais on dit « *He is sleepy.* »)

Cela signifie que, par rapport à l'inférence lexicale en L1, les connaissances linguistiques impliquées dans l'inférence lexicale en L2 concernent au moins deux langues, la langue cible et la langue maternelle de l'apprenant. Haastrup (1991), par exemple, classe les connaissances linguistiques impliquées dans l'inférence lexicale dans deux grandes catégories, intralinguistique et inter-linguistique : les connaissances intralinguistiques concernent les connaissances linguistiques acquises dans la langue cible (par exemple, un apprenant de l'anglais L2 comprend que « -able » est un suffixe adjectival qui représente en général une capacité ou une potentialité) ; les connaissances inter-linguistiques portent sur les connaissances acquises essentiellement dans la L1 de l'apprenant que ce dernier utilise pour déduire le sens d'un mot L2. Par exemple, un apprenant francophone peut facilement deviner que « *member* » en anglais veut dire « *membre* » en français, les deux mots étant congénères interlinguaux (ou « *cognates* », qui renvoient aux mots similaires à la forme et au sens dans différentes langues (Théophanous, 2001)).

Pour De Bot et al. (1997), « *les apprenants adultes sont susceptibles d'utiliser toutes les informations dont ils disposent pour interpréter des mots inconnus* (p. 317)¹³ ». En effet, des études montrent que les connaissances linguistiques acquises en L1 restent actives dans les traitements des informations lexicales en L2 (N. Jiang, 2002, 2004; Kroll et al., 2010). Par exemple, dans l'étude de Fraser (1999), l'une des erreurs les plus courantes commises par ses participants francophones dans

¹³ « *Adult language learners are likely to use whatever information is available to them to interpret unknown words.* »

l'inférence des mots anglais inconnus est d'associer ces mots anglais avec des congénères français inappropriés. Bensoussan & Laufer (1984) ont remarqué pour leur part que les « faux amis » sont un grand obstacle pour les apprenants hébreophones pour déchiffrer sémantiquement des mots cibles anglais. Nagy et al. (1997) ont pour leur part mené une étude qui consiste à examiner l'effet des connaissances syntaxiques en L1 (espagnol) sur l'inférence du sens des mots L2 (anglais) inconnus. Leurs résultats ont relevé une corrélation positive entre le niveau de la compétence en lecture en espagnol et le taux d'erreur d'inférence lié aux transferts syntaxiques inappropriés de l'espagnol vers l'anglais.

Or, dans le domaine de l'inférence lexicale en L2, pendant une longue période, on n'avait pas accordé autant d'importance à la L1 qu'elle aurait méritée. Comme l'indique Haastrup (2008), les études réalisées pendant le siècle dernier dans le domaine de l'acquisition des langues secondes/étrangères se focalisaient majoritairement sur l'anglais L2 du fait de sa popularité dans le monde. Les L1 cibles étaient typologiquement proches de l'anglais, par exemple le français et l'allemand. Suite à l'évolution de la mondialisation, d'autres L2 cibles ont fait l'objet de recherche dans le domaine, par exemple l'espagnol (Pulido, 2003, 2007), le japonais (Mori, 2003 ; Hamada & Koda, 2010), le chinois (M. J. Wu, 2005; Xiang et al., 2018), le russe (Comer, 2012), l'hébreu (Prior et al., 2014) et le turc (Fatih, 2015). L'effet de la distance typologique entre la L2 et la L1 de l'apprenant sur l'inférence lexicale en L2 a alors progressivement émergé.

Koda (2007, p. 27) définit la distance typologique linguistique comme le niveau de similarité structurale entre deux langues. Le partage des propriétés structurales signifie la similarité des processus linguistiques entre les deux langues, ce qui faciliterait les transferts des compétences en L1 « prêtes à utiliser » en L2 avec des ajustements minimaux. Pour Koda (2007), « ...l'input L2 est traité par le biais des compétences en L1 transférées (p. 18)¹⁴ » et « ...la distance linguistique devrait être directement en rapport avec les différences d'efficacité du traitement (des informations linguistiques) à un moment donné parmi les lecteurs en L2 de diverses

¹⁴ « ...L2 input is processed through the transferred L1 competencies. »

L1 (p. 27)¹⁵».

Les résultats de l'étude de Hamada & Koda (2010) ont fourni une bonne preuve sur l'effet de la distance L1-L2 quant au traitement d'informations lexicales pendant l'inférence lexicale. Hamada & Koda ont étudié deux groupes, l'un composé de 13 Chinois et 4 Japonais (groupe logographique) et l'autre de 15 Coréens et 1 Turc (groupe alphabétique), sur leurs choix de sources de connaissance pour inférer les pseudo-mots anglais dans trois passages. Elles ont constaté que, bien que les taux de réussite d'inférence lexicale étaient peu différents entre les deux groupes, le groupe logographique était moins compétent que le groupe alphabétique au niveau du décodage phonétique en anglais L2 ; par ailleurs, le score d'inférence lexicale du groupe alphabétique était en corrélation positive avec l'efficacité du décodage phonétique, ce qui n'était pas le cas pour le groupe logographique.

Les connaissances acquises en L1 « prêtes à utiliser » en L2 comprennent aussi le lexique. Comme le dit N. Ellis (1997, p. 134), contrairement aux jeunes enfants apprenant leur langue maternelle, les apprenants en L2 ont en général déjà développé de riches systèmes conceptuels qui sont liés à la L1. Ces connaissances acquises en L1 sont des ressources riches. La distance typologique L1-L2 décide en quelque sorte combien de connaissances lexicales acquises en L1 sont « transférables » de façon plus ou moins ajustée dans les activités en L2. Par exemple, un apprenant francophone peut facilement comprendre le mot « casa » en espagnol (« maison » en français) parce que dans leur lexique mental existe déjà les mots français « case » et « casanier ». Cette facilité peut provenir du partage de congénères entre les deux langues qui sont liés aux mêmes concepts dans le lexique mental de l'apprenant. Par exemple, dans l'étude de Paribakht & Tréville (2007) comparant l'utilisation de la stratégie d'inférence lexicale par des apprenants francophones et des apprenants persanophones, seuls les premiers ont utilisé la source de connaissance « Homonymie en L1 (le sujet confond le mot cible avec un mot de L1 de forme (sonore ou graphique) ressemblante) » pour inférer les mots anglais inconnus.

¹⁵ « ...linguistic distance should be directly related to differences in processing efficiency at a given point in time among L2 readers with diverse L1 backgrounds. »

Les connaissances non linguistiques sont relativement moins étudiées et discutées dans l'inférence lexicale. Toutefois, la distance typologique entre deux langues implique souvent l'écart entre les arrière-plans socio-culturels des milieux dans lesquels les langues sont utilisées. Dans l'étude de Wesche & Paribakht (2009) sur l'inférence lexicale, par rapport aux apprenants persanophones, les apprenants francophones ont utilisé plus fréquemment les connaissances non linguistiques pour inférer les mots anglais inconnus. Selon ces chercheuses, cela est probablement lié au rapprochement socio-culturel entre la communauté francophone et la communauté anglophone au Canada où leurs apprenants francophones recrutés étudiaient l'anglais L2. En outre, le groupe français a globalement mieux réussi les inférences des mots anglais inconnus que le groupe persan. Pour Wesche & Paribakht, par rapport aux apprenants persanophones, les apprenants francophones ont pu tirer profit à la fois de la similarité entre le français et l'anglais et des partages socioculturels entre les deux communautés. Dans leurs conclusions, Wesche & Paribakht (2009) ont mis l'accent sur l'importance du contexte social qui peut offrir plus d'occasions de contact et d'exposition à la langue et à la culture L2.

1.3.3.2 L1 de l'apprenant mis en relief dans l'inférence lexicale en chinois L2

L'arrière-plan linguistique et culturel de l'apprenant est notamment mis en relief dans la recherche en chinois L2 pour les particularités de cette langue (détaillées dans le Chapitre III). Le fait que la langue chinoise n'est pas alphabétique à l'écrit et que le lien forme-sens est faible pour les caractères, l'unité d'écriture chinoise, rend l'apprentissage du vocabulaire chinois complexe et difficile, notamment pour les apprenants dont la L1 est alphabétique comme la langue française. A cause de cela, dans les études empiriques réalisées sur le chinois L2 visant les apprenants de différentes L1, les apprenants japonais et coréens sont souvent classés dans un groupe différent des apprenants d'autres L1 parce que les langues japonaise et coréenne sont largement influencées par la langue chinoise¹⁶.

¹⁶ Les caractères chinois étaient les unités d'écriture pratiquées au Japon jusqu'au 8^{ème} siècle et en Corée jusqu'au 15^{ème} siècle, avant l'apparition de leurs propres alphabets. A présent, plus de 2000 Kanji (caractères chinois utilisés dans la langue japonaise) sont pratiqués au quotidien au Japon, parmi lesquels autour de 76% sont identiques au niveau de la forme à leurs homologues chinois (55% des caractères simplifiés et 21% des

Dans l'étude de M. J. Wu (2005), par exemple, les participants sont classés en trois groupes – le groupe japonais, le groupe coréen et le groupe autres L1 alphabétiques. Chaque groupe contient trois sous-groupes variés en niveau du chinois L2 – débutant, intermédiaire 1 (ou débutant-intermédiaire) et intermédiaire 2 (ou intermédiaire-avancé). Wu remarque que les participants du groupe japonais de tous niveaux s'appuient beaucoup sur les connaissances linguistiques L1 pour inférer les mots cibles chinois. Dans le groupe coréen, seuls les participants des niveaux intermédiaires cherchent des indices dans leur L1. La L1 est absente dans l'ensemble des sources de connaissance utilisées par les participants du groupe alphabétique.

L'effet positif de la similarité L1-L2 est également soutenu par les résultats des études de J. L. Zhang (2010) et de Y. X. Fang & Jiang (2012). J. L. Zhang (2010) a demandé à des apprenants en chinois L2 d'inférer les sens de 20 mots chinois inconnus présentés isolément, classés en cinq niveaux de transparence sémantique (Zwitserslood (1994, p. 344) explique qu'un mot est sémantiquement transparent si son sens est synchroniquement lié aux sens de ses composants, par exemple, « milkman »). Les participants sont regroupés en fonction de leur L1, le groupe japonais-coréen et le groupe alphabétique. Les résultats révèlent que le groupe japonais-coréen a un score d'inférence quasiment double de celui du groupe alphabétique. Y. X. Fang & Jiang (2012) ont adopté le protocole de Mori (2003, cf. supra p. 29) pour observer le rôle du contexte dans l'inférence lexicale. Dans cette étude auprès des apprenants du niveau intermédiaire en chinois, les participants sont conventionnellement classés en groupe japonais et groupe alphabétique. 36 mots chinois sont présentés chacun sous l'une des trois conditions – « Mot seul », « Contexte seul » et « Mot contextualisé ». Les participants doivent deviner le sens du mot dans les conditions « Mot seul » et « Mot contextualisé » et proposer un mot pour compléter la phrase lacunaire dans la condition « Contexte seul ». Les résultats révèlent une meilleure performance du groupe japonais par rapport au groupe

caractères traditionnels) et de 74% sémantiquement pareils (L. Y. Wang, 2011). En Corée (du sud), les caractères chinois furent entièrement remplacés par les alphabets coréens depuis les années 1970, mais les cours d'écriture chinoise sont intégrés dans le programme scolaire à l'école primaire. Linguistiquement, beaucoup de mots coréens peuvent retrouver leur homologue en chinois, phonétiquement proches. Les mots chinois représentent presque 60% du vocabulaire coréen utilisé aujourd'hui (Xi, 2018).

alphabétique, notamment dans la condition « Mot seul ». Les participants du groupe alphabétique, parmi les trois conditions, ont la meilleure performance sous la condition « Mot contextualisé ». Autrement dit, les apprenants japonais sont assistés par leurs connaissances pré-acquises sur les caractères grâce à la similarité entre leur L1 et le chinois, tandis que les apprenants d'autres L1, n'ayant pas cette ressource, doivent recourir davantage au contexte du mot pour chercher des indices.

1.3.3.3 Lexicalisation dans la L1

Si l'existence d'un concept dans la mémoire de l'apprenant peut faciliter l'inférence d'un mot inconnu, qu'en est-il de l'absence du concept correspondant au mot inconnu ? Plus précisément, l'apprenant parvient-il à réussir l'inférence d'un mot L2 qui n'a pas d'équivalent dans son lexique mental en L1, c'est-à-dire un mot non lexicalisé dans sa langue maternelle ?

Les termes « lexicaliser » et « lexicalisation » sont définis dans le dictionnaire « Le Grand Robert de la langue française » (Tome 4, 2001, p. 776) comme suit:

(se) Lexicaliser : (Ling.) se mettre à fonctionner comme une unité lexicale

Lexicalisation : (Ling.) Le fait d'être lexicalisé, de devenir une unité lexicale.

A l'opposé, la non lexicalisation se réfère à l'absence d'unité lexicale équivalente d'un mot L2 dans la L1 de l'apprenant. Le terme « unité lexicale » signifie qu'une signification spécifique quelconque est représentée sous forme d'un mot ou d'un mot composé, par exemple « pomme de terre », ou par expression figée ou idiomatique, par exemple « tomber dans les pommes ». Même si ces derniers sont composés de plusieurs mots, ensemble ils délivrent une seule signification, donc un seul concept dans le système conceptuel, ils sont donc considérés comme une seule unité.

Paribakht (2005), l'une des précurseurs dans la recherche sur le thème de la lexicalisation d'un mot L2 dans l'inférence lexicale, définit la lexicalisation en L1 comme « *l'existence en L1 d'unités lexicales correspondant aux mots cibles de L2* (p. 705)¹⁷ ». Par exemple, le mot anglais « testimony » est considéré lexicalisé dans la

¹⁷ Cela est la traduction en français de la phrase de Paribakht & Tréville (2007, p. 401). Une autre version en anglais plus proprement résumée de la définition sur la lexicalisation de Paribakht est « ...whether or not, there

langue française car il existe son homologue « témoignage » ; en revanche, le mot anglais « plight » (« situation critique » en français) est considéré non lexicalisé dans la langue française parce qu'aucune unité lexicale en français ne lui correspond.¹⁸

Selon le modèle de De Bot et al. (1997, cf. la Section 1.3.1), l'inférence lexicale est le processus qui consiste à collecter des informations nécessaires pour établir un lien entre la forme du mot et un concept dans le lexique mental de l'apprenant. Selon ce principe, il serait plus difficile pour un apprenant francophone d'inférer le mot anglais « plight » (« situation critique ») que le mot « testimony » (« témoignage » en français) à cause de l'absence du concept correspondant à « plight » dans son lexique mental en L1 ; cela signifie aussi que la non disponibilité d'un concept correspondant au mot « plight » obligerait l'apprenant à activer plusieurs concepts connexes à partir des indices repérés, par exemple « situation », « état », « difficile », « critique », etc., puis à les organiser afin d'obtenir une idée cohérente avec les contextes où se trouve le mot « plight ».

Paribakht (2005) est la première à formuler clairement l'hypothèse de lexicalisation. Elle indique que, « *L'hypothèse de lexicalisation... stipule qu'une fois que suffisamment de composants sémantiques et syntaxiques d'un mot inconnu sont extraits par inférence et qu'un lemme approprié, ou au moins un partiel, est activé dans le lexique mental du lecteur, une telle activation peut également impliquer la récupération par l'apprenant de son équivalent en L1... Tandis que l'inférence d'un mot lexicalisé en L1 peut être caractérisée comme un processus impliquant la reconnaissance d'un lemme, celui d'un mot non lexicalisé dans la langue maternelle implique un processus de construction d'un lemme.* (p. 729)¹⁹ » L'inférence du sens d'un mot L2 non lexicalisé dans la L1 de l'apprenant représente donc un processus

are lexical equivalents - single or compound lexical items, including lexical phrases – in the learner's L1 for the L2 target words. (Paribakht, 2010, p. 62) »

¹⁸ Le mot « testimony » est inclus dans la liste des mots anglais lexicalisés en français et le mot « plight » est inclus dans la liste des mots non lexicalisés dans l'étude de Paribakht & Tréville (2007).

¹⁹ « The lexicalization hypothesis... stipulates that once sufficient semantic and syntactic components of an unfamiliar word are extracted through inferencing and an appropriate, or at least a partial, lemma is activated in the reader's mental lexicon, such activation may also involve the learner's retrieval of the item's L1 equivalent... Whereas inferring the meaning of a word lexicalized in L1 can be characterized as a process involving lemma recognition, that of a word not lexicalized in the native language involves a process of lemma construction.»

cognitif plus coûteux et par conséquent, plus difficile à réussir, par rapport à un mot L2 lexicalisé.

1.3.3.4 Effet de la lexicalisation sur l'inférence lexicale

Les chercheurs ont déjà remarqué l'effet de la lexicalisation d'un mot L2 dans le cadre pédagogique. Par exemple, Laufer (1994) a abordé ce sujet dans les années 1990, le nommant alors « vide lexical ». Elle a remarqué que les apprenants israéliens en l'anglais L2 ont des difficultés à comprendre les mots anglais qui n'ont pas d'homologue en hébreu, par exemple « cosy » (*confortable* ou *douillet*), et qu'ils ont donc tendance à éviter ce genre de mots anglais.

Paribakht (2005) a examiné, auprès de 20 étudiants universitaires persanophones, qui apprennent l'anglais L2 en Iran, l'effet de la lexicalisation sur le choix des sources de connaissance pour inférer les mots inconnus et sur le résultat d'inférence. On a demandé aux participants d'inférer au total 50 mots cibles imprimés en gras dans des passages, dont 25 mots sont lexicalisés et les 25 autres non lexicalisés. Paribakht catégorise les sources de connaissance en « sources linguistiques propres à la L2 », « sources linguistiques propres à la L1 » et « source non linguistique ». Les sources linguistiques propres à la L2 sont regroupées en fonction de leur emplacement – mot, phrase et texte (cf. le Tableau 1.2, infra, p. 42, pour la typologie des sources de connaissance utilisée dans l'étude de Paribakht & Tréville (2007) qui est adaptée de celle de Paribakht (2005)). Les résultats montrent que la phrase comportant le mot inconnu est la piste principale pour inférer le sens d'un mot inconnu, quelle que soit la nature de lexicalisation du mot ; au niveau du mot, le poids relatif des indices morphologiques est plus important dans l'inférence pour les mots non lexicalisés (7%) que pour les mots lexicalisés (4,4%). Globalement, les différences de préférence des sources de connaissance pour inférer les mots lexicalisés et non lexicalisés sont insignifiantes. Quant au taux des inférences entièrement réussies, le taux pour les mots lexicalisés (17%) est presque trois fois plus élevé que celui pour les mots non lexicalisés (6%).

Paribakht conclut donc que globalement, la nature de lexicalisation d'un mot inconnu n'a pas d'effet significatif sur le choix des sources de connaissance pendant

l'inférence lexicale. Celui-ci dépend plutôt du contexte dans lequel se trouve le mot, des caractéristiques du mot (la forme et la classe grammaticale) et des connaissances linguistiques et encyclopédiques de l'apprenant. Les résultats révèlent que les mots L2 non lexicalisés sont plus difficiles à comprendre que ceux lexicalisés. Cependant, le score d'inférence pour les mots non lexicalisés n'est pas nul. La chercheuse explique que « *...l'absence d'un équivalent lexical L1 peut compliquer, mais n'entrave pas nécessairement la compréhension du sens des mots lorsque les concepts constitutifs sont familiers.* (p. 730)²⁰ ». Elle souligne que le fait que les apprenants ont plus de difficultés à comprendre les mots non lexicalisés signifie que dans la pratique pédagogique, davantage d'instructions explicites préparées avec soin sont nécessaires dans l'enseignement des mots L2 non lexicalisés en général.

L. Fan & Zhang (2014) ont adopté le protocole de Paribakht (2005) pour examiner l'effet de la lexicalisation auprès des apprenants sinophones de l'anglais L2. Selon les résultats, les apprenants sinophones se sont appuyés davantage sur la phrase comportant un mot inconnu pour inférer le sens du mot, notamment le sens général de la phrase ; l'utilisation des connaissances morphologiques est plus fréquente dans l'inférence pour les mots non lexicalisés que pour les mots lexicalisés (le poids relatif de cette source de connaissance est 34,8% dans l'inférence pour les mots lexicalisés et 43% pour les mots non lexicalisés). Quant au taux de réussite, les apprenants sinophones ont réussi remarquablement mieux les inférences des mots lexicalisés que celles des mots non lexicalisés, respectivement 48% vs 18% pour les inférences entièrement correctes et 14% vs 22% pour les inférences partiellement correctes.

Une autre étude au sujet de la lexicalisation, menée par Paribakht & Tréville (2007), a examiné à la fois l'effet provenant de la distance typologique L1-L2 sur l'inférence lexicale et celui de la lexicalisation. Dans cette étude, Paribakht & Tréville ont comparé les résultats des apprenants persanophones et ceux des apprenants francophones, la langue française étant considérée plus similaire avec l'anglais que la langue farsi. Les questions de recherche principales portent sur l'effet de la L1 de l'apprenant sur le processus et sur le résultat de l'inférence lexicale pour les mots L2

²⁰ « ...the lack of an L1 lexical equivalent may complicate but does not necessarily impair word meaning comprehension when the constituent concepts are familiar. »

lexicalisés et non lexicalisés. Paribakht & Tréville (2007) ont adapté la typologie utilisée dans l'étude de Paribakht (2005) pour identifier les indices repérés par les participants (Tableau 1.2, p. 42).

Les résultats sur le choix des sources de connaissance pendant l'inférence lexicale révèlent que :

- 1) La phrase comportant le mot inconnu étant la première source importante pour tous les apprenants, les apprenants francophones cherchent plus d'indices au niveau du mot tandis que les apprenants persanophones s'appuient davantage sur les indices au niveau du discours. La source d'homonymie en L1 est exclusivement utilisée par les apprenants francophones.
- 2) L'ordre de priorité des sources de connaissance reste identique pour le groupe persan afin d'inférer les mots lexicalisés et non lexicalisés (l'indice du niveau de la phrase > l'indice du niveau du discours > l'indice du niveau du mot), alors que le groupe français puise plus dans les connaissances non linguistiques pour inférer les mots lexicalisés.

Pour Paribakht & Tréville (2007), le fait que les apprenants francophones s'appuient plus lourdement sur leurs connaissances linguistiques au niveau du mot est attribué à la similarité graphique et à la parenté linguistique entre les langues française et anglaise.

Par ailleurs, il est à noter que le groupe francophone est constitué des apprenants de collège et universitaires dans la région d'Ottawa au Canada et le groupe persanophone des étudiants universitaires à Téhéran en Iran. Autrement dit, les apprenants persanophones apprennent l'anglais dans un contexte d'apprentissage hétéroglotte (c'est-à-dire l'apprentissage d'une langue n'a pas lieu dans un milieu où la langue cible est la langue parlée, opposé au contexte d'apprentissage homoglotte). En revanche, les apprenants francophones sont dans un milieu bilingue, voire anglophone. Selon ces chercheuses, cela expliquerait l'utilisation plus importante des connaissances non linguistiques par les apprenants francophones (notamment pour inférer les mots lexicalisés) que les apprenants persanophones : au niveau de l'arrière-plan socio-culturel, le français est plus proche de l'anglais ; d'autant plus que géographiquement parlant, les apprenants francophones ont probablement plus

d'occasions de s'exposer à la culture et aux média anglophones.

Sources de nature linguistique

1) Sources propres à L2

– Niveau du mot

Association de mots : le sujet fait l'association entre le mot cible et un autre mot qui lui est familier.

Collocation : le sujet connaît une expression de L2 dont il croit que le mot cible fait partie.

Morphologie : le sujet analyse les composants du mot cible en se basant sur ses connaissances de désinence, affixes et racines.

Homonymie : le sujet confond le mot cible avec un autre mot L2 de forme (sonore ou graphique) ressemblante.

– Niveau de la phrase

Sens de la phrase : le sujet s'appuie sur le sens général de la phrase contenant le mot cible.

Grammaire : le sujet s'appuie sur les propriétés syntaxiques du mot cible, sur l'ordre des mots ou sur les catégories grammaticales.

Ponctuation : le sujet s'appuie sur les connaissances des règles de ponctuation et de leur signification.

– Niveau du discours

Sens : le sujet s'appuie sur le sens général du texte, paragraphe ou phrases entourant le mot cible.

Structure : le sujet s'appuie sur les connaissances sur la macrostructure du texte et des règles d'organisation des divers types de textes.

2) Sources propres à L1

Collocation en L1 : le sujet reconnaît des mots qui accompagnent généralement l'équivalent du mot cible en L1 et présume que la même relation existe en L2.

Homonymie en L1 : le sujet confond le mot cible avec un mot de L1 de forme (sonore ou graphique) ressemblante.

Sources de nature non linguistiques

Connaissances générales : le sujet puise dans les connaissances non linguistiques qu'il a en mémoire incluant sa connaissance du thème traité dans le texte.

Tableau 1.2 – Typologie des sources de connaissance utilisées dans l'inférence lexicale en L2 (Paribakht & Tréville, 2007, p. 407)

Les résultats au niveau du degré de réussite d'inférence lexicale montrent que les taux de réussite pour les apprenants francophones sont respectivement de 55,3% pour les mots lexicalisés et de 47,6% pour les mots non lexicalisés, tandis que pour les apprenants persanophones, ils sont respectivement de 28% pour les mots lexicalisés et de 17% pour les mots non lexicalisés. Autrement dit, 1) les deux groupes ont mieux réussi l'inférence pour les mots lexicalisés que pour les mots non lexicalisés ; 2) les apprenants francophones ont produit plus d'inférences correctes, partiellement et entièrement, que les apprenants persanophones.

Si nous récapitulons les trois études réalisées en anglais L2 présentées ci-dessus, l'étude de Paribakht (2005) auprès des apprenants persanophones, l'étude de Paribakht & Tréville (2007) auprès des apprenants persanophones et francophones et celle de L. Fan & Zhang (2014) auprès des apprenants sinophones, nous pouvons voir que, pendant l'inférence lexicale,

- 1) les apprenants de diverses L1 s'appuient sur un ensemble de différentes sources de connaissances pour inférer le sens d'un mot L2, lexicalisé ou non dans la L1 de l'apprenant ;
- 2) la distance typologique entre la L2 et la L1 de l'apprenant semble avoir un impact sur le choix des sources de connaissance pendant l'inférence lexicale :
 - dans l'étude de Paribakht & Tréville (2007), par rapport aux apprenants persanophones, les apprenants francophones s'inspirent davantage de leurs connaissances en L2 au niveau du mot pour inférer un mot anglais inconnu ;
 - dans les trois études, les sources linguistiques propres à la L1 sont peu utilisées par les apprenants sinophones et ceux persanophones, mais bien plus fréquemment par les apprenants francophones.

La raison est probablement liée à la distance typologique entre les langues chinoise, française, persane (L1) et la langue anglaise (L2) : par rapport au chinois et au persan, le français est plus similaire à l'anglais et les apprenants francophones ont ainsi pu tirer davantage de profit des connaissances transférées depuis la L1 pour inférer les mots anglais inconnus, par rapport aux apprenants chinois et aux apprenants persanophones.

- 3) il ne faut pas ignorer l'influence du contexte socio-culturel de l'apprenant. Les

apprenants francophones ont plus fréquemment recouru à leurs connaissances non linguistiques pour inférer les mots anglais inconnus, par rapport au groupe persan dans l'étude de Paribakht & Tréville (2007) ; dans l'étude de L. Fan & Zhang (2014), les apprenants chinois ont peu utilisé cette source de connaissance. Cette différence est probablement due à la fois au rapprochement socio-culturel entre les communautés francophone et anglophone et à l'effet bénéfique du contexte d'apprentissage homoglotte dans lequel les apprenants francophones étudient l'anglais. Un contexte d'apprentissage homoglotte donne la possibilité aux apprenants de s'exposer plus fréquemment aux stimuli linguistiques et culturels de la langue cible, ce qui peut favoriser le développement du système conceptuel de l'apprenant lié à la langue cible.

- 4) Quant au taux de réussite d'inférence lexicale, les trois études arrivent à la conclusion commune que les apprenants de diverses L1 ont plus de difficultés à déduire le sens d'un mot L2 qui n'est pas lexicalisé dans leur L1, par rapport à un mot lexicalisé.

En outre, nous avons remarqué que, comparés avec les apprenants persanophones dans l'étude de Paribakht (2005), les apprenants chinois dans l'étude de L. Fan & Zhang (2014) ont davantage exploré la forme d'un mot anglais inconnu pour inférer son sens : les poids relatifs des indices morphologiques et d'homonymie en L2 repérés par les apprenants chinois sont respectivement de 18,3% pour les mots lexicalisés et de 33,7% pour les mots non lexicalisés, tandis que ces poids sont respectivement de 6,1% pour les mots lexicalisés et de 8% pour les mots non lexicalisés pour les apprenants persanophones dans l'étude de Paribakht (2005). Cet écart est probablement lié aux matériels expérimentaux différents utilisés dans les deux études. Selon nous, ce fait est aussi possiblement lié aux différentes compétences de décodage demandées dans les L1 des apprenants : la langue chinoise étant une langue non alphabétique à l'écrit, le lien graphie-son pour les caractères (l'unité minimale à l'écrit) est relativement faible par rapport à une langue alphabétique. Par contre, la forme d'un caractère peut délivrer des informations lexicales (cf. le Chapitre III). Il en résulte que l'analyse de la forme d'un caractère est

très importante pour la reconnaissance du caractère.

A partir de ces résultats et conclusions, nous nous sommes posé des questions sur le cas où la L2 est non alphabétique et la L1 de l'apprenant est alphabétique, plus précisément, un apprenant francophone du chinois L2, pour les deux raisons suivantes :

- 1) La langue chinoise est bien différente de la langue française, avant tout, au niveau du système d'écriture : la langue française est alphabétique à l'écrit et la correspondance graphèmes-phonèmes est plus consistante par rapport à la langue chinoise, une langue non alphabétique à l'écrit et phonétiquement plus opaque. Cela signifie que, pour décoder un mot dans chaque langue, les compétences nécessaires ne sont pas les mêmes. D'ailleurs, le sens d'un mot chinois est souvent fortement lié aux sens individuels de ses composants (cf. le Chapitre III). Dans ce cas-là, pour inférer le sens d'un mot chinois inconnu, on peut s'interroger si un apprenant francophone s'appuie davantage sur les indices au niveau du mot, notamment sur ceux liés à la forme, parce qu'ils sont importants dans le processus de décodage en chinois, ou bien s'il recourt plutôt aux sources de connaissance d'autres niveaux parce que peu de liens forme-concept dans le lexique mental en français peuvent être « réutilisés » pour décoder un mot chinois.
- 2) Ces deux langues sont très différentes l'une de l'autre, non seulement au niveau linguistique mais également sur l'aspect socio-culturel. Autrement dit, par rapport à un apprenant japonais en chinois L2, un apprenant francophone risquerait plus de tomber sur des mots chinois non lexicalisés en français. Par exemple, pour le mot chinois « 盒饭 », un apprenant japonais le comprendrait facilement grâce au mot « 弁当 », l'homologue du mot chinois « 盒饭 » en japonais ; mais pour un apprenant francophone, il faudrait lui expliquer que « 盒饭 » signifie un ensemble de plats, accompagné de riz, soit vendu en boîte en magasin, soit préparé à la maison. Sans une expérience de la vie en Chine ou dans d'autres pays asiatiques comme le Japon, il est peu évident que l'apprenant francophone se fasse rapidement une idée explicite sur ce que « 盒

饭 » représente, même si on lui fournit l'explication du mot. Connaissant cette difficulté supplémentaire, le fait que l'apprentissage du chinois L2 pour les apprenants francophones ait généralement lieu en France, au moins au stade débutant (comme dans beaucoup d'autres cas d'apprentissage d'une L2), est un autre désavantage. Le manque de stimuli linguistiques et culturels dans un contexte hétéroglotte est constant.

Dans ce cas-là, que se passe-t-il quand un apprenant francophone, qui apprend le chinois en France, doit inférer le sens d'un mot chinois non lexicalisé ? Le processus est-il différent de celui pour inférer un mot chinois lexicalisé en français ? L'apprenant réussirait-il mieux l'inférence pour les mots chinois inconnus lexicalisés en français que pour ceux non lexicalisés ?

Nous n'avons malheureusement pas de réponses aux questions posées ci-dessus. Jusqu'à présent, la recherche sur l'effet de la lexicalisation sur l'inférence lexicale en L2 est très rare.

1.4 Récapitulation et questions de recherche sur l'inférence lexicale

Dans ce chapitre, nous avons discuté l'inférence lexicale qui est une stratégie de lecture importante utilisée par les lecteurs en L2 comme en L1. L'inférence lexicale permet aux apprenants d'établir une cohésion locale pour pallier à un manque sémantique afin d'assurer une compréhension fluide d'un texte.

Pour De Bot et al. (1997), l'inférence lexicale est un processus qui sert à créer un lien entre la forme d'un mot et un concept disponible dans la mémoire de l'apprenant. Le lexème et le lemme dans le lexique mental de l'apprenant jouent les médiateurs dans ce lien. L'apprenant doit collecter des informations nécessaires pour pouvoir activer le médiateur suivant afin d'avancer dans la création du lien, à la lumière des indices fournis par le texte et ses connaissances pré-acquises, linguistiques et non linguistiques. Les résultats des études réalisées montrent qu'en L2 comme en L1, les éléments tels que la richesse d'informations fournies par le contexte, le niveau linguistique et la familiarité du thème du texte de la part de l'apprenant peuvent intervenir dans le processus et le résultat de l'inférence lexicale.

Toutefois, par rapport à l'inférence lexicale en L1, la distance typologique entre la L2 et la L1 de l'apprenant est un facteur supplémentaire à prendre en compte dans l'inférence lexicale en L2, car cette dernière concerne aussi les connaissances transférées depuis la L1. Plus la L2 est typologiquement proche de la L1 de l'apprenant, plus de connaissances pré-acquises, linguistiques ou non linguistiques en L1 sont transférables et prêtes à utiliser dans l'inférence lexicale en L2. Quand les transferts de connaissances se situent au niveau du lexique, cela représente la disponibilité d'un concept dans le lexique mental de l'apprenant, avec lequel l'apprenant peut tenter d'établir un lien avec le mot L2 correspondant. Dans le cas où la L2 est typologiquement lointaine de la L1, l'apprenant risque davantage de tomber sur des mots L2 dont le concept n'existe pas dans le lexique mental de la L1 de l'apprenant, c'est-à-dire des mots L2 non lexicalisés. Le fait qu'aucun concept dans la mémoire de l'apprenant n'est disponible pour établir le lien rendrait l'inférence du sens d'un mot L2 non lexicalisé plus difficile.

Cela dit, jusqu'à présent, la recherche sur la lexicalisation est rare et la langue cible se limite à l'anglais, pour autant que nous le sachions. Visant le cas où un apprenant francophone apprend le chinois, une langue linguistiquement et culturellement lointaine de la langue française, nous nous sommes posé les questions de recherche suivantes :

QR 1 : Lors de la compréhension en lecture en chinois L2, de quelle manière un apprenant francophone s'appuie-t-il sur les différentes sources de connaissance pour inférer les mots chinois inconnus présents dans la lecture ? La nature des sources de connaissance utilisées pour inférer le sens des mots chinois inconnus diffère-t-elle selon que les mots chinois sont lexicalisés ou non en français ?

QR 2 : Avec quel degré de réussite un apprenant francophone parvient-il à inférer le sens des mots inconnus, lexicalisés ou non, présents dans les textes chinois ?

Les résultats des études réalisées étant en désaccord, les lectures que nous avons effectuées ne nous permettent pas de construire une hypothèse pour la première question de recherche. Quant à la deuxième question, nous avons émis l'hypothèse suivante à la lumière des conclusions des études réalisées : les apprenants en chinois

L2 réussissent mieux l'inférence des mots chinois lexicalisés que les mots non lexicalisés.

**CHAPITRE II – INFERENCE LEXICALE DANS L'APPRENTISSAGE
DU VOCABULAIRE**

Introduction

Nous avons présenté dans le Chapitre I l'inférence lexicale en tant que stratégie de lecture. Dans ce chapitre, nous présenterons le rôle de la stratégie d'inférence lexicale dans l'apprentissage du vocabulaire.

Nous commencerons par une revue des études réalisées sur le sujet de l'apprentissage du vocabulaire via la lecture en L1 et en L2. Ces études démontrent qu'un apprenant peut obtenir des gains lexicaux sur un mot inconnu présent dans un texte en tentant de déduire son sens grâce au contexte. Nous examinerons alors les éléments qui peuvent influencer le résultat d'apprentissage via la lecture pendant laquelle l'objectif principal de l'apprenant est la compréhension globale du texte, et non la mémorisation du vocabulaire. Par la suite, nous présenterons des études sur l'effet des tâches supplémentaires à la suite de l'inférence lexicale sur l'apprentissage des mots inférés.

Nous terminerons ce chapitre avec nos questions de recherche concernant l'apprentissage du vocabulaire en L2 subséquent à l'inférence lexicale, en insistant sur l'effet de la lexicalisation.

2.1 Inférence lexicale, une stratégie importante d'apprentissage du vocabulaire

Dans le chapitre précédent, nous avons discuté la stratégie d'inférence lexicale qu'un apprenant peut utiliser en cas de lacune lexicale pendant sa lecture. A la lumière des indices fournis par le texte et des connaissances de l'apprenant, l'inférence lexicale peut aider l'apprenant à remédier à un manque de connaissance au niveau du lexique afin d'obtenir une compréhension fluide et globale du texte.

En dehors de cela, à force d'avoir recours à des informations de toutes sortes pour déduire le sens d'un mot inconnu, l'apprenant peut arriver à mémoriser les informations lexicales concernant ce mot. Autrement dit, l'utilisation de l'inférence lexicale peut aboutir à l'acquisition du mot et, par conséquent, contribuer au

développement du lexique mental de l'apprenant.

L'inférence lexicale n'est pas seulement une stratégie de lecture, mais également une stratégie d'apprentissage du vocabulaire.

2.1.1 Apprentissage du vocabulaire via la lecture en L1

Les chercheurs observent chez les enfants et les jeunes apprenants que la croissance de leur vocabulaire n'est pas proportionnelle à la quantité de mots enseignés à l'école. Par exemple, les élèves anglophones entre 9 et 18 ans peuvent avoir une croissance annuelle d'environ 3000 mots, dont seulement 200-300 relèvent de l'intervention pédagogique (Jenkins & Dixon, 1983; Nagy et al., 1987). La seule explication de ce fait est que les enfants et les jeunes apprenants disposent d'autres ressources que celles scolaires pour développer leur lexique.

Comme disent Gray & Holmes (1938, p. 28, cité par; Nagy et al., 1985, p. 234), « *[Nous] savons par expérience que pratiquement tous les élèves acquièrent de nombreux mots grâce au contexte avec peu ou pas d'aide des enseignants.*²¹ » Chall (1987, cité par Hu & Nation, 2000) précise à ce sujet que, pour un enfant au début de son parcours scolaire, le vocabulaire soutient la compréhension écrite ; en revanche, la compréhension écrite peut être une ressource de développement lexical après que l'enfant maîtrise les compétences de lecture au bout de trois ou quatre ans d'études. Etant d'avis que les enfants acquièrent la majorité du vocabulaire via la lecture, Jenkins & Dixon (1983) soulignent que « *... l'argument selon lequel les enfants apprennent (et se souviennent de) le sens de nouveaux mots qu'ils ont rencontrés uniquement dans un contexte verbal repose sur l'hypothèse qu'ils sont capables de déduire le sens des mots dans ces conditions* (p. 248)²² » ; en d'autres termes, cela dépend de leur capacité d'inférer le sens d'un mot inconnu contextualisé.

Nguyen-Xuan (1995) précise que, selon l'approche cognitive en psychologie, afin de comprendre le sens d'un mot inconnu présent dans un texte, l'apprenant doit

²¹ « [W]e know from experience that practically all pupils acquire many meanings from the context with little or no help from teachers. »

²² « ... the argument that children learn (and remember) meanings for new words which they have encountered only in verbal context rests on the assumption that they are able to derive word meanings under these conditions. »

trier et vérifier les informations recueillies, rejeter et relancer les actes précédents si nécessaire, avant d'enfin décider le sens du mot qui est logique dans son contexte. Ce genre d'opérations mentales, c'est-à-dire raisonnements complexes, peuvent conduire à l'acquisition des informations lexicales.

De nombreuses études ont été menées pour examiner l'hypothèse sur l'apprentissage du vocabulaire via la lecture. Par exemple, dans l'étude de Carroll & Drum (1983) auprès de lycéens, on a d'abord demandé à ces jeunes apprenants de définir les sens de 40 mots qu'ils ne connaissaient pas, puis de lire des passages comportant la moitié de ces mots cibles ; à l'issue de la lecture, on a demandé aux jeunes apprenants de donner à nouveau les définitions des 40 mots cibles. Les résultats montrent que les jeunes apprenants ont eu de meilleurs scores pour les 20 mots cibles rencontrés dans des passages que les 20 autres. Selon les chercheurs, les apprenants, sans aide externe, ont pu tirer profit des informations fournies par le contexte pour comprendre les mots cibles et ont fini par en retenir certains.

Nagy et al. (1985) ont étudié pour leur part les gains lexicaux grâce au contexte dans un cadre naturel de lecture. Ils ont sollicité 57 collégiens natifs anglophones de la classe de 3^{ème} pour lire deux textes. 15 mots les plus difficiles contenus dans chaque texte sont sélectionnés comme mots cibles. A l'issue de la tâche de lecture, on a évalué les gains lexicaux chez les apprenants à travers différentes tâches. Les résultats montrent que, pour les mots cibles dans les deux textes, les gains lexicaux chez ces jeunes apprenants sont significatifs après la lecture. Dans une autre étude d'une plus grande ampleur de Nagy et ses collègues (1987), les chercheurs ont sollicité 352 apprenants de CE2, CM2 et de 5^{ème}. L'objectif de cette étude était de rechercher les éléments qui peuvent intervenir dans le résultat d'apprentissage du vocabulaire grâce au contexte sur le plan du mot inconnu (le nombre d'occurrence du mot dans le texte, la longueur du mot, la transparence morphologique et le niveau de la difficulté conceptuelle du mot) et sur le plan du texte (la richesse d'informations utiles fournies par le texte afin de deviner le sens d'un mot inconnu, la lisibilité du texte et la densité de mots difficiles présents dans le texte). Les apprenants devaient lire des textes, soit argumentatifs soit narratifs, pour la compréhension et passer un entretien individuel immédiatement après. Sans être prévenus en amont, les apprenants ont reçu une

semaine plus tard un questionnaire à choix multiples qui évaluait leurs connaissances sur les mots cibles. Les résultats montrent que les gains lexicaux sur les mots cibles sont petits, mais significatifs chez les apprenants de tous niveaux scolaires.

Les deux études ont conclu communément que pour les apprenants, enfants et adolescents, « ... *l'apprentissage grâce au contexte pendant la lecture pour le plaisir sans intervention pédagogique est le mode majeur de l'acquisition du vocabulaire durant les années scolaires et le volume d'expérience avec la langue écrite, en interaction avec la capacité de compréhension en lecture, est le déterminant principal de la croissance du vocabulaire* (Nagy et al., 1985, p. 234)²³ ». En outre, dans l'étude de Nagy et al. (1987), parmi les éléments examinés, la seule propriété qui est significativement interactive avec le résultat d'apprentissage est le niveau de difficulté conceptuelle. Concrètement, les enfants ont plus de difficultés à comprendre les mots inconnus dont le concept est difficile ou nouveau pour eux. Le taux d'apprentissage d'un mot dans une lecture pour le plaisir sans instructions est de 0,05, à peu près un tiers du résultat obtenu par l'étude de Nagy et al. (1985), une différence liée aux différents délais du post test de vocabulaire selon les chercheurs. Ces résultats confirment la conclusion de Nagy et al. (1985) que l'apprentissage d'un mot est un processus graduel. Il est donc nécessaire pour les enseignants d'encourager les apprenants à lire largement et régulièrement pour avoir un meilleur développement lexical à long terme et aussi de les assister dans leur restructuration de concepts.

Shu et al. (1995) ont adapté le protocole de Nagy et al. (1987) pour observer des enfants de différentes L1 dans l'apprentissage des mots inconnus présents dans des textes. 317 enfants chinois et 170 enfants américains, de niveau CE1 et CM2, devaient lire des textes et ensuite répondre à des questionnaires à choix multiple. Selon les résultats, tous les enfants, quel que soit leur niveau scolaire et leur langue maternelle, ont obtenu des gains lexicaux significatifs après la lecture. Les résultats montrent aussi qu'un contexte riche en informations utiles afin de deviner le sens

²³ « ...learning from context during free reading is the major mode of vocabulary acquisition during the school years and the volume of experience with written language, interacting with reading comprehension ability, is the major determinant of vocabulary growth. »

d'un mot promeut l'acquisition de ce mot ; toutefois, les enfants ont plus de difficultés avec des mots qui renvoient à des concepts difficiles (ou aux nouveaux concepts). En d'autres termes, ils apprennent plus facilement les mots dont le concept est déjà connu, un résultat en accord avec celui de Nagy et al. (1987). La différence entre les deux groupes linguistiques se trouve au niveau de l'importance de la transparence morphologique dans l'inférence des mots cibles : aucune interaction n'est trouvée entre cette source de connaissance et le résultat d'inférence chez les enfants américains, mais chez les enfants chinois de la classe CM2, cette interaction est significative. L'explication possible de cette différence est que la forme d'un caractère chinois peut fournir des informations sémantiques et le sens d'un mot chinois constitué de plusieurs caractères est souvent lié fortement aux sens individuels de ces derniers (nous reviendrons sur ce point dans le Chapitre III).

Les résultats des études présentées ici fournissent de solides preuves pour l'hypothèse que les apprenants natifs, enfants et adolescents, parviennent à inférer le sens des mots inconnus grâce au contexte pendant la lecture, puis à les retenir. Leurs résultats sont en accord avec ceux d'autres études semblables à ce sujet, tels que : un contexte fournissant de riches informations sur le sens d'un mot inconnu peut favoriser l'apprentissage du mot (Jenkins et al., 1984) ; l'apprentissage du vocabulaire est graduel, une lecture extrascolaire extensive et régulière est donc nécessaire pour augmenter la fréquence d'exposition aux mots nouveaux dans divers contextes (Jenkins et al., 1984; Jenkins & Dixon, 1983; Krashen, 1989) ; les mots conceptuellement difficiles méritent davantage d'attention (Herman et al., 1987) et il est nécessaire d'encourager et aider les enfants à intégrer les mots nouveaux avec leurs connaissances déjà acquises et à restructurer leur système conceptuel ; afin de consolider les mots nouveaux appris par cette piste, les apprenants doivent pratiquer concrètement les mots nouveaux autant que possible avec d'autres moyens (Nagy, 1988).

On peut ici se demander si tel est le cas en L2. C'est ce qu'on va examiner dans la partie suivante.

2.1.2 Apprentissage du vocabulaire via la lecture en L2

En L2, les chercheurs ont étudié la possibilité de développer le vocabulaire via la lecture en se basant sur les résultats obtenus en L1.

De Bot et al. (1997) ont essayé d'expliquer le processus d'inférence lexicale sous l'angle cognitif à partir de l'hypothèse lexicale (« lexical hypothesis ») de Levelt (1993) (cf. la Section 1.3.1). Allant plus loin, les chercheurs suggèrent que l'attention initiale que l'apprenant porte à un mot inconnu puisse aboutir à la mise en place d'une structure de connaissance vide pour un nouveau lemme. Pendant l'inférence du sens du mot inconnu, l'apprenant comble progressivement la nouvelle structure de lemme d'informations lexicales repérées à partir du contexte. De cette façon, l'inférence du sens d'un mot inconnu peut alors conduire à la création d'un nouveau lemme comportant des informations lexicales concernant ce mot dans le lexique mental de l'apprenant, ce qui expliquerait la possibilité d'apprendre un mot inconnu rencontré pendant la lecture à la suite de l'inférence du sens du mot. Cela dit, De Bot et ses collègues (1997, p. 317) précisent que les informations syntaxiques peuvent être inférées par le biais des co-textes, mais que les informations sémantiques dans un lemme dérivent seulement des concepts existants dans le système conceptuel de l'apprenant. Le processus d'apprentissage du sens d'un mot doit alors impliquer l'acte de copier-coller des informations sémantiques depuis le système conceptuel vers le lemme.

De nombreuses études menées à ce sujet ont généralement validé l'idée qu'en L2, comme en L1, l'inférence du sens d'un mot inconnu en lecture peut aboutir à des gains lexicaux chez l'apprenant pour ce mot, ce qui signifie que la lecture extensive peut également aider les apprenants L2 à développer leur lexique, autant pour une langue alphabétique (ex. (Fraser, 1999; Ghobadi et al., 2016; Hulstijn, 1992; Pigada & Schmitt, 2006; Reynolds & Wible, 2014; X. Zhang & Lu, 2015) que pour une langue non alphabétique comme le chinois (X. J. Qian, 2003; Shao & Li, 2019; X. Y. Zhu & Zhou, 2007).

Selon les résultats des études réalisées sur ce sujet, les gains lexicaux subséquents à l'inférence lexicale dépendent d'un ensemble d'éléments, par exemple :

- Au niveau du mot, Vidal (2011) a remarqué que les apprenants ont moins de difficultés à apprendre les mots L2 pour lesquels on peut trouver des mots L1 correspondants grâce aux congénères ; Laufer (1990) a constaté pour sa part que les apprenants de bas niveau en L2 ont plus de difficultés à apprendre les mots L2 longs.
- Un contexte plus riche d'informations utiles (B. Chen et al., 2017; Li, 1988) ou une densité plus importante du vocabulaire visuel présent dans le texte (Pulido, 2004, 2007) peut aboutir à un meilleur résultat d'apprentissage des mots L2 inconnus qui y sont présents.
- Sur le plan de l'exposition à un mot, une exposition plus fréquente à un mot L2 inconnu dans différents contextes contribue à l'apprentissage de ce mot (Hulstijn et al., 1996; Laufer & Rozovski-Roitblat, 2011; Rott, 1999; Teng, 2016; Waring & Takaki, 2003).
- Le niveau linguistique en L2 et la capacité de compréhension en lecture de l'apprenant (Peters et al., 2009; Pulido, 2007; A. Zhao et al., 2016), ainsi que ses connaissances métacognitives (Tseng & Schmitt, 2008; A. Zhao et al., 2016), peuvent avoir une influence importante : les apprenants ayant un meilleur niveau linguistique en L2 parviennent à mieux retenir les mots L2 inconnus inférés ; par ailleurs, une grande motivation à comprendre des mots inconnus rencontrés en lecture et une meilleure maîtrise de stratégies pourraient aider l'apprenant à obtenir de meilleurs gains lexicaux sur les mots.

En outre, au sujet des facteurs concernant l'apprenant, Pulido (2004, 2007) se penche spécifiquement sur l'effet de l'arrière-plan culturel et la familiarité du thème du texte sur l'apprentissage des mots L2 inconnus rencontrés en lecture. Les deux études de Pulido ont été réalisées auprès d'apprenants anglophones de l'espagnol L2 aux Etats-Unis. Dans l'étude de 2004, les apprenants ont lu pour la compréhension des textes suivant un même scénario ayant lieu dans différents arrière-plans culturels (par exemple l'inscription à l'Université d'Illinois et à l'Université de Paraguay). Les résultats du test de rétention immédiate montrent que les apprenants ont reconnu plus de non-mots intégrés dans les textes culturellement plus familiers que ceux moins familiers. Dans l'étude de 2007, les apprenants ont lu pour la compréhension des

textes aux thèmes plus ou moins familiers (par exemple, « Un rendez-vous avec un docteur » est considéré comme un thème plus familier et « Publication d'un article » comme un thème moins familier). Selon les résultats, deux jours après la lecture, les apprenants ont reconnu significativement plus de mots cibles intégrés dans les textes au thème familier que ceux dans les textes au thème non familier.

Pulido (2004, 2007) en conclut qu'une meilleure connaissance du contexte conceptuel facilite la compréhension du texte dans lequel se situe un mot inconnu, ce qui aide par la suite la construction d'une nouvelle représentation pour le mot inconnu dans le lexique mental de l'apprenant. Conjointement avec les résultats obtenus en L1 sur le fait que les apprenants parviennent moins bien à obtenir des gains lexicaux sur les mots inconnus conceptuellement difficiles (ou nouveaux) rencontrés en lecture (Herman et al., 1987; Nagy et al., 1985, 1987), ces résultats révèlent la grande importance des connaissances conceptuelles dans l'apprentissage des mots inconnus via la lecture.

En cas de manque conceptuel pour un mot L2, un apprenant arriverait-il à apprendre le mot grâce au contexte ?

2.1.3 Effet de la lexicalisation sur l'apprentissage du vocabulaire grâce au contexte

Nous avons discuté dans le Chapitre I la lexicalisation d'un mot L2 dans la L1 de l'apprenant. Les résultats des études réalisées ont démontré communément que par rapport aux mots L2 non lexicalisés dans la L1, les apprenants comprennent plus facilement les mots L2 lexicalisés (cf. la Section 1.3.3.3). Selon l'hypothèse de De Bot et al. (1997), un nouveau lemme est créé dans le lexique mental de l'apprenant à partir de la forme d'un mot inconnu pendant l'inférence lexicale et il est comblé par des informations lexicales au fur et à mesure ; cependant, les informations sémantiques ne peuvent pas être dérivées du co-texte qui entoure le mot, mais doivent être transférés d'un concept déjà existant dans le système conceptuel de l'apprenant. Il serait logique d'en induire que la disponibilité d'un concept pour être transféré vers le nouveau lemme en facilite l'accomplissement. Effectivement, dans l'apprentissage

du vocabulaire en L2, des chercheurs ont remarqué qu'un mot L2 accompagné par son équivalent en L1 facilite l'apprentissage (J. Liu, 2008; Lotto & Groot, 1998), notamment pour les apprenants ayant un niveau faible en L2 (Prince, 1996).

Par conséquent, dans le cas où un mot L2 n'est pas lexicalisé dans la L1 de l'apprenant, c'est-à-dire que le concept correspondant à ce mot L2 est absent dans le lexique mental en L1 de l'apprenant, serait-il plus difficile pour l'apprenant d'apprendre ce mot ?

L'étude importante de Wesche & Paribakht (2009) a examiné auprès des apprenants persanophones et francophones l'effet de la lexicalisation sur les gains lexicaux sur les mots inconnus rencontrés en lecture. Un post test à l'insu des apprenants a été administré immédiatement après la tâche d'inférence lexicale. Les résultats montrent que, pour les apprenants de diverses L1, l'inférence du sens des mots inconnus a contribué aux gains lexicaux sur ces mots, lexicalisés ou non, mais que les gains sont globalement au niveau de la forme des mots inférés ; le taux de gain pour les mots lexicalisés est peu différent de celui pour les mots non lexicalisés pour les apprenants des deux groupes. Autrement dit, la nature de la lexicalisation des mots inférés n'a pas d'effet sur le taux de gain lexical immédiat. Ces résultats sont généralement en accord avec ceux obtenus par l'étude de L. Fan & Zhang (2014). Wesche & Paribakht ont souligné dans leurs conclusions finales l'importance du facteur socioculturel : une exposition plus fréquente aux stimuli linguistiques et culturels en L2 favorise le développement lexical chez un apprenant dans divers aspects, notamment en termes de concepts.

C. Chen & Truscott (2010) se sont focalisés sur le rôle de deux variables dans l'apprentissage du vocabulaire via la lecture, le nombre d'exposition et la lexicalisation. L'étude a sollicité 72 apprenants sinophones de l'anglais L2 qui ont été aléatoirement classés en trois groupes respectivement exposés une fois, trois fois et sept fois aux mots cibles intégrés dans des textes. Deux post tests, adaptés des tests de Webb (2007), ont été administrés auprès des apprenants, l'un immédiatement après la lecture et l'autre deux semaines plus tard, pour évaluer leurs gains lexicaux sur les mots cibles sous divers aspects. Selon les résultats, les apprenants ont obtenu

un score de rétention sémantique immédiate significativement meilleur pour les mots lexicalisés, sans différence significative au niveau du gain sur l'association syntagmatique entre les mots lexicalisés et non lexicalisés. Au deuxième post test, les chercheurs n'ont observé aucune différence significative entre les mots lexicalisés et non lexicalisés. Les deux tests post-lecture montrent que les apprenants n'ont eu quasiment aucun gain lexical pour les mots non lexicalisés, quelle que soit la fréquence d'exposition, à l'exception du mot « troubadour ». Selon les chercheurs, les apprenants ont eu des gains lexicaux sur ce mot non lexicalisé parce qu'il fait partie de l'idée principale du texte. Autrement dit, ce mot était important pour la compréhension du texte. Apprendre un mot L2 inconnu via la lecture dépend donc en partie de l'importance du mot dans la compréhension du texte concerné. C. Chen & Truscott (2010) indiquent que, par rapport aux mots L2 lexicalisés, les apprenants ont de grandes difficultés à apprendre les mots non lexicalisés dans leur L1 : dans cette étude, les apprenants ont généralement obtenu un gain sémantique pour les mots lexicalisés à partir de trois expositions, mais quasiment aucun gain sémantique pour la grande majorité des mots non lexicalisés, même au bout de sept expositions. Les chercheurs suggèrent alors que l'apprentissage des mots L2 non lexicalisés nécessite davantage d'instructions explicites.

L'étude de Heidari-Shahreza & Tavakoli (2012) auprès des étudiants universitaires persans qui apprennent l'anglais a adopté le protocole de C. Chen & Truscott (2010). 10 mots anglais utilisés dans l'étude de Paribakht (2005) ont été sélectionnés comme mots cibles, dont 5 mots lexicalisés et 5 autres non lexicalisés en persan. Les résultats des deux post tests, l'un immédiatement après la lecture et l'autre deux semaines plus tard, montrent que les participants ont obtenu des gains significativement meilleurs pour les mots lexicalisés, par rapport aux mots non lexicalisés, dans l'aspect sémantique. En outre, au deuxième post test, la seule diminution de score significative se situe dans l'aspect sémantique (à savoir le sens et l'association) pour les mots non lexicalisés. Les chercheurs sont arrivés à la conclusion que pour ce qui est de l'apprentissage immédiat et de la rétention, les apprenants persanophones ont plus de difficultés pour les mots anglais non lexicalisés.

Il est à noter une différence méthodologique entre les deux études présentées

ci-dessus : dans l'étude de Wesche & Paribakht (2009), on a demandé aux apprenants d'inférer l'un après l'autre les sens des 50 mots cibles (25 mots lexicalisés et 25 non lexicalisés) présents dans les textes et les apprenants ont mieux réussi l'inférence pour les mots lexicalisés que pour les mots non lexicalisés ; dans l'étude de C. Chen & Truscott (2010), l'inférence du sens des 10 non-mots (5 lexicalisés et 5 non lexicalisés) a eu lieu implicitement pendant la lecture et, immédiatement après la lecture, les apprenants ont eu une meilleure performance sur le rappel du sens pour les mots lexicalisés que pour ceux non lexicalisés.

On peut conclure de ces études que l'utilisation de la stratégie d'inférence lexicale pendant une lecture peut avoir un effet bénéfique sur l'apprentissage des mots inconnus présents dans le texte, lexicalisés ou non, mais les gains lexicaux semblent rester essentiellement au niveau de la reconnaissance de la forme des mots inférés, sauf si ces derniers sont de petite quantité ; les apprenants ont généralement plus de difficultés à apprendre des mots L2 non lexicalisés. La recherche concernant l'effet de la lexicalisation sur l'apprentissage du vocabulaire²⁴ demeure rare jusqu'à présent et, à notre connaissance, la langue cible se limite à l'anglais.

2.2 Connaître un mot

Les résultats des études que nous avons présentés plus haut montrent que les gains lexicaux subséquents à l'inférence du sens d'un mot L2 inconnu concernent diverses dimensions, par exemple la forme et le sens. En effet, la connaissance d'un mot est multidimensionnelle.

Concrètement, qu'est-ce qu'on entend par « connaître un mot » ?

Il existe traditionnellement deux axes principaux pour mesurer la compétence lexicale d'une personne, l'étendue (en bref, la quantité des mots connus) et la profondeur (la qualité, ou les connaissances sous divers aspects pour un mot donné).

²⁴ Reynolds (2016) a réalisé une étude dans laquelle la lexicalisation est l'une des sous-variables étudiées sur les caractéristiques des mots L2 dans l'apprentissage du vocabulaire en L2 via la lecture. Le matériel expérimental utilisé dans cette étude est un roman et les mots cibles sélectionnés sont des non-mots anglais. En outre, la proportion des mots cibles lexicalisés et non lexicalisés est considérablement déséquilibré, 84% des mots cibles considérés lexicalisés vs 16% non lexicalisés. Comme la lexicalisation n'est pas l'une des questions de recherche principales de cette étude, nous ne présentons pas ici ses résultats.

On peut également mesurer la connaissance d'un mot selon deux critères, réceptif et productif, qui se réfèrent respectivement à la capacité de reconnaître un mot et à la capacité de l'utiliser (Henriksen, 1999).

Read (2004) résume à trois niveaux les principales propositions de définition sur la connaissance d'un mot en profondeur (Tableau 2.1) : la connaissance du sens d'un mot donné, les connaissances lexicales sur divers plans sur un mot et les liens plus ou moins étroits entre un mot donné et d'autres mots. Les deux premières définitions examinent un mot plutôt isolément et la troisième s'appuyant sur le réseau lexical lié à un mot spécifique (cf. Théophanous, 2019, p. 10–14).

1. Précision sémantique	La différence entre connaissance vague et connaissance précise du sens du mot
2. Connaissance holistique du mot	La connaissance non seulement des traits sémantiques mais aussi phonologique, orthographique, morphologique, syntaxique, collocationnel et pragmatique du mot
3. Connaissance associationnelle ou organisation lexicale	L'intégration du mot dans un réseau avec la capacité de le lier à d'autres mots comparables et de l'en distinguer

Tableau 2.1 – Définitions de la profondeur des connaissances lexicales (Read 2004, adapté par Théophanous, 2019, p. 12)

Hulstijn (2001) définit la connaissance d'un mot depuis l'angle de la reconnaissance. En accord avec De Bot et al. (1997), il allègue que la reconnaissance d'un mot est le processus qui consiste à activer l'accès à ce mot dans le lexique mental. Pour un adulte natif instruit, l'accès à un mot dans son lexique mental comprend plusieurs caractéristiques – sémantique, pragmatique, stylistique, collocationnel, syntaxique, catégoriel, morphologique, phonologique, articulatoire et orthographique. Une reconnaissance qui échoue peut être liée à une activation inférieure au seuil minimal (c'est-à-dire que les informations perçues et extraites sont insuffisantes) ou alors au fait que les informations enregistrées sont incomplètes, voire vides.

Nation (2001), pour sa part, divise la connaissance d'un mot en trois aspects,

la forme (écrite et orale), le sens et l'emploi, dont chacun comprend plusieurs sous-aspects à double dimension – la réception et la production (Tableau 2.2). Selon lui, une connaissance complète d'un mot comprendrait donc la maîtrise de l'intégralité des aspects d'un mot dans un contexte spécifique.

Forme	Orale	R	Puis-je le reconnaître à l'oreille ?	
		P	Puis-je le prononcer correctement ?	
	Écrite	R	Puis-je le reconnaître à l'écrit ?	
		P	Puis-je l'orthographier ?	
	Parties du mot	R	Puis-je déceler ses parties ?	
		P	Puis-je dire lesquelles de ses parties sont nécessaires pour exprimer son sens ?	
Sens	Forme et sens	R	Quel est le sens de cette forme lexicale ?	
		P	Quelle forme lexicale puis-je utiliser pour exprimer ce sens ?	
	Concepts et références	R	Qu'est-ce que ce concept inclut ?	
		P	A quels items fait référence ce concept ?	
	Associations	R	A quels autres mots me fait-il penser ?	
		P	Puis-je employer un autre mot pour à sa place ?	
	Emploi	Fonctions grammaticales	R	Dans quelle structure apparaît-il ?
			P	Dans quelle structure dois-je l'employer ?
		Collocations	R	Quels mots ou types de mots apparaissent avec ce mot ?
			P	Quels mots ou types de mots dois-je employer avec ce mot ?
		Contraintes d'emploi (registre, fréquence...)	R	Où, quand et avec quelle fréquence dois-je m'attendre à le rencontrer ?
			P	Où, quand et avec quelle fréquence dois-je l'employer ?

R – Réception, P – Production

Tableau 2.2 – Connaissance d'un mot (Nation, 2001, adapté par Théophanous, 2019, p. 14)

Le fait que la connaissance d'un mot concerne des informations lexicales sous divers aspects expliquerait que l'apprentissage d'un mot se déroule de façon graduelle

(ex. Nagy et al., 1985). A tout moment pendant le développement du vocabulaire d'un apprenant, les connaissances concernant un mot ne sont pas en état « tout-ou-rien » (Laufer, 1998).

Sous l'angle de vue de la production, Théophanous (2019) indique que « *[S]i tous les aspects de la connaissance lexicale sont maîtrisés pour un mot donné, l'apprenant sera alors capable d'utiliser le mot avec assurance, aisance et précision à la manière des locuteurs natifs.* (p. 13) ». Cela dit, des études en anglais L2 démontrent que, chez un apprenant, le développement du vocabulaire réceptif est plus rapide que celui du vocabulaire productif et que le volume de vocabulaire réceptif est, par conséquent, plus important que le vocabulaire productif (Laufer, 1998; Zhong & Hirsh, 2009). En d'autres termes, les connaissances productives sont plus difficiles par rapport aux connaissances réceptives et sont acquises normalement après les connaissances réceptives (Melka, 1997). (Cf. Zhong, 2011, p. 120)

2.3 Apprentissage incident vs intentionnel du vocabulaire

Différentes conditions liées à l'input peuvent conduire à des gains lexicaux à divers niveaux (Barcroft, 2015). Dans une discussion sur les activités d'apprentissage du vocabulaire, Nation (2014, p. 96) indique que les gains lexicaux à la suite de l'inférence du sens grâce au contexte (« *guessing from context* ») sont au niveau de la forme, du lien forme-sens et de la collocation.

Toutefois, en L2 comme en L1, les résultats des études réalisées sur l'apprentissage du vocabulaire grâce au contexte pendant la lecture s'avèrent modestes sur le plan du gain formel, encore plus faibles sur le plan des gains sémantiques et autres. Cela est indépendant de l'objectif de recherche et de la méthode expérimentale utilisée : des études imitent une condition naturelle de lecture extensive, c'est-à-dire que les apprenants lisent un texte à fin de compréhension sans effectuer la tâche d'inférer l'un après l'autre les mots cibles présents (par exemple les études de Nagy et al. (1985) et de Waring & Takaki (2003)) ; d'autres études visent une exploration profonde de l'inférence lexicale. Dans ce cas-là, les apprenants doivent rapporter oralement, de façon introspective ou rétrospective, leurs processus

et les résultats d'inférence des mots surlignés ou jugés inconnus après une lecture pour la compréhension (par exemple les études de Sternberg (1987) et Li (1988)).

Cela suscite des doutes chez certains chercheurs sur l'apprentissage du vocabulaire sous la condition d'apprentissage incident, c'est-à-dire l'apprentissage des mots nouveaux pendant une activité dont l'objectif principal est autre chose que l'apprentissage du vocabulaire, par rapport à l'approche traditionnelle scolaire qui est catégorisée en apprentissage intentionnel.

Les termes « apprentissage incident » et « apprentissage intentionnel » sont originellement employés en psychologie depuis la seconde moitié du 20^{ème} siècle, et sont largement utilisés aujourd'hui dans le domaine de l'apprentissage du vocabulaire pour discerner ces deux conditions d'apprentissage (Hulstijn, 2013). La distinction entre les deux conditions d'apprentissage demeure toujours débattue de nos jours.²⁵

Certains chercheurs insistent sur la conscience de l'apprenant au moment de l'input (ex. N. Ellis, 1994; Krashen & Terrell, 1983) pour baliser la différence entre les deux conditions d'apprentissage. Schmidt (1990) résume à grands traits trois niveaux de conscience – « intention », « connaissance (implicite) » (« implicit knowledge ») et « prise de conscience » (« awareness »). Pour Schmidt (2010), l'apprentissage incident se réfère au fait qu'on peut apprendre quelque chose sans intention particulière. Selon lui, l'enregistrement conscient des éléments linguistiques spécifiques, auxquels l'apprenant accorde l'attention, est nécessaire pour que l'input se transforme en saisie (« l'intake »). La conscience est donc toujours présente, quelle que soit la condition d'apprentissage.

Hulstijn (2013), pour sa part, souligne l'aspect délibéré (ou « engagement » selon ses propres termes) de l'apprenant au moment de l'input. Il indique que, dans la linguistique appliquée, « *L'apprentissage incident ... se réfère à l'acquisition d'un mot ou d'une expression sans l'intention consciente de mémoriser des éléments.* (p.

²⁵ « Acquisition incidente » et « acquisition intentionnelle » sont deux termes également couramment utilisés dans la littérature du domaine. Nous utilisons dans notre thèse le terme « apprentissage », à l'exception des citations des autres auteurs.

2632)²⁶ » Par exemple, on parvient à se souvenir des objets et de l'agencement d'une chambre qu'on visite fréquemment, sans se forcer à la mémoriser. « *L'apprentissage intentionnel ... se réfère à une tentative délibérée de mémoriser des informations factuelles* (p. 2632)²⁷ », par exemple réviser les leçons pour préparer un examen à l'école. Selon ces définitions, **l'apprentissage incident peut se produire sans ou avec des instructions, pourvu que la tentative de mémoriser des mots soit absente chez l'apprenant.** Au niveau opérationnel, Hulstijn (2001) cite Eysenck (1982, p. 198) pour distinguer l'apprentissage incident de l'intentionnel : « *En termes opérationnels, l'apprentissage incident et intentionnel peuvent être simplement distingués en termes de l'utilisation d'instructions avant l'apprentissage qui préviennent, ou pas, les sujets de l'existence d'un test de rétention subséquente.* (p. 268)²⁸ » Par ailleurs, à propos de deux autres termes couramment utilisés dans le domaine de l'acquisition des langues secondes/étrangères, l'apprentissage explicite et l'apprentissage implicite, Hulstijn (2013) précise que ces deux termes renvoient plutôt à l'apprentissage conscient ou inconscient des informations factuelles ou des régularités dans les matériels d'input auxquels les apprenants s'exposent dans les expériences empiriques. Les régularités concernées dans les études du domaine se rapportent souvent aux phénomènes grammaticaux, par exemple les règles morphosyntaxiques.

Sok (2014), dans sa revue des propositions sur la distinction des deux conditions d'apprentissage, classe les définitions de Hulstijn dans la catégorie « méthode-orientée » et ajoute que cette dernière provient du fait que la distinction entre les apprentissages incident et intentionnel, en ce qui concerne l'attention ou la conscience de l'apprenant au moment de l'input, est difficile et complexe à maintenir sur le plan théorique. Le concept d'« engagement » est plus important et plus pertinent et mérite davantage de discussions. En ce sens, pour Barcroft (2015, p. 45), en réalité, il n'existe pas de condition d'apprentissage du vocabulaire purement incidente ou

²⁶ «... incidental learning... refer[s] to the acquisition of a word or expression without the conscious intention to commit the element to memory...»

²⁷ « The intentional learning ... refers to a deliberate attempt to commit factual information to memory...»

²⁸ « In operational terms, incidental and intentional learning can simply be distinguished in terms of the use of prelearning instructions that either do, or do not, forewarn subjects about the existence of a subsequent retention test. »

intentionnelle, parce que l'attention de l'apprenant pendant les activités n'est pas dichotomique (Gass, 1999). De toute façon, peu de moyens peuvent nous permettre d'observer à un moment précis à quoi l'apprenant prête attention et dans quelle mesure. Barcroft propose donc un continuum de condition d'apprentissage dont les deux extrémités sont caractérisées hautement « incident » ou « intentionnel », et il indique qu'il faudrait tenir compte de la co-existence « incidente-intentionnelle » de nature continue dans la recherche sur l'apprentissage du vocabulaire.

Malgré la polémique sur la distinction entre les deux conditions d'apprentissage, l'apprentissage du vocabulaire via une lecture pour la compréhension est toujours caractérisé « incident » puisque l'objectif de la lecture est de comprendre le sens global du texte, non la mémorisation des informations lexicales sur les mots inconnus qui y sont présents.

2.4 Interrogations sur l'apprentissage du vocabulaire via la lecture

Puisque l'obtention de gains lexicaux sur les mots inconnus présents dans un texte est fortuite, comme l'indique le terme « apprentissage incident », par rapport à l'apprentissage du vocabulaire avec instructions pédagogiques traditionnelles, des chercheurs ont mis en question l'efficacité de cette approche pour apprendre des mots nouveaux (Horst et al., 1998 ; Raptis, 1997), même pour les mots fréquents en L2 (Cobb, 2010). Les interrogations principales se situent à trois niveaux, a) la probabilité de remarquer un mot inconnu présent dans un texte, b) le degré de réussite une inférence du sens d'un mot inconnu grâce au contexte et c) la fréquence d'exposition nécessaire.

2.4.1 Négligence et ignorance des mots inconnus

Puisque l'objectif des apprenants pendant la lecture est la compréhension du texte, non la mémorisation des informations lexicales pour un mot inconnu présent, le gain lexical subséquent à l'inférence du sens du mot grâce au contexte n'est qu'un sous-produit (« by-product », Gass, 1999) de la lecture. Cela dit, les apprenants

n'entreprennent pas pour tous les mots inconnus présents dans un texte une inférence du sens (Laufer & Hulstijn, 2001; Schmidt, 1990, 2001; Sternberg & Powell, 1983).

Par exemple, dans l'étude de Laufer & Yano (2001, cité par Laufer, 2003), les participants de l'anglais L2 ont d'abord lu un texte pour la compréhension et passé ensuite deux tests évaluant leurs connaissances sur les mots cibles intégrés dans le texte, l'un en auto-évaluation et l'autre où le sens des mots cibles sont demandés. Le résultat s'avère surprenant : le score d'auto-évaluation est en moyenne de plus de 60% supérieur à celui de l'autre test. Laufer (2003) explique que cette surestimation peut provenir du fait que les apprenants comprennent globalement le texte sans faire attention aux sens précis des mots individuels et/ou que les apprenants confondent des mots inconnus avec ceux connus par la forme ou à cause des faux congénères, ou encore de la transparence trompeuse (c'est-à-dire que la forme d'un mot fournit de faux indices sur le sens du mot (« deceptive transparency », Laufer, 1989, p. 13). Les participants n'ont en fait pas toujours pris conscience de la présence de ces mots inconnus dans le texte.

En outre, les apprenants n'entreprennent pas une inférence de sens pour tous les mots inconnus constatés. Dans l'étude de Fraser (1999), ignorer le mot représente 24% des stratégies utilisées par ses participants face aux mots inconnus. Dans l'étude de Paribakht & Wesche (1999), plus de la moitié des mots que les participants jugent inconnus ont été ignorés pendant l'accomplissement des tâches (respectivement 52% pour la tâche de compréhension et 56% pour le résumé du texte). En chinois L2, X. J. Qian (2003) remarque que les participants japonais ont ignoré en moyenne 43% des mots inconnus présents dans les textes. Y. Huang (2008) constate chez ses participants indonésiens que l'utilisation de la stratégie d'inférence lexicale n'est pas automatique et que seuls les mots inconnus jugés importants pour la compréhension du texte attirent leur attention.

2.4.2 Disponibilité des informations lexicales dans le contexte

Même pour les mots pour lesquels une inférence du sens est entreprise, les apprenants ne parviennent pas toujours à obtenir un résultat satisfaisant. Nous avons

discuté dans le Chapitre I que beaucoup d'éléments peuvent intervenir sur le résultat de l'inférence lexicale, par exemple la richesse d'informations nécessaire fournies par le contexte qui peuvent aider l'apprenant à déduire correctement le sens d'un mot inconnu dans le contexte (cf. la Section 1.3.3).

Selon le modèle de De Bot et al. (1997, cf. la Section 1.3.1), le processus d'inférence d'un mot inconnu est une collecte d'informations pour créer un lien forme-lexème-lemme-concept. L'absence ou l'insuffisance d'informations pour activer un ou plusieurs maillons dans ce lien peut conduire à une tentative de compréhension vaine ou incomplète. En réalité, les mots inconnus dans une lecture ne sont pas tous entourés d'informations nécessaires pour que l'apprenant puisse s'en servir et procéder à une inférence. Même si cette condition est satisfaite, l'apprenant doit être suffisamment compétent pour pouvoir en bénéficier. Cela signifie qu'une bonne partie des tentatives d'inférences finissent par être erronées, voire abandonnées. Même dans les études dont les paramètres de lecture sont généralement « contrôlés », le taux de réussite d'inférence n'est pas encourageant. Dans l'étude de D. D. Qian (2005), par exemple, les apprenants de l'anglais L2 devaient inférer les mots cibles soulignés dans un texte ; 41% des tentatives ont été jugées invalides parce qu'elles se sont terminées sans réponse ou ont obtenu une mauvaise réponse.

2.4.3 Fréquence d'exposition

La connaissance d'un mot étant multidimensionnelle (cf. la Section 2.2), un seul contexte peut difficilement fournir toutes les informations lexicales concernant un mot donné. D'ailleurs, les informations lexicales de différentes natures exigent des traitements à différents niveaux de profondeur. Par exemple, le traitement des informations sémantiques a lieu à un niveau plus profond que le traitement des informations concernant la forme phonétique (Craik & Lockhart, 1972, cité par Laufer & Hulstijn, 2001). Comme l'indique Schmitt (2008), « *Les mots devront être rencontrés dans de nombreux contextes différents pour développer la maîtrise des différents types de connaissance des mots, ce qui implique une approche récursive à*

long terme de l'apprentissage du vocabulaire. (p. 335)²⁹ » « Nombreux contextes différents » implique diverses informations lexicales et un bon nombre d'expositions à un même mot.

Il est communément accepté aujourd'hui que l'apprentissage du vocabulaire se déroule de façon graduelle et que le nombre d'expositions à un mot est donc un bon prédicteur du résultat d'apprentissage de ce mot dans le cadre d'apprentissage incident (Webb, 2007). Dans l'examen de Uchihara et al. (2019, p. 562) des études sur l'effet des répétitions d'exposition dans l'apprentissage incident du vocabulaire via la lecture en L2, les résultats montrent une grande différence entre les gains lexicaux à divers niveaux en fonction du nombre d'expositions :

- Afin qu'un apprenant puisse reconnaître la forme d'un mot, une seule exposition peut suffire : dans l'étude de C. Chen & Truscott (2010), les participants ont reconnu la forme de 43% des mots cibles présentés une seule fois dans les textes ; ce taux est de 67% dans l'étude de Webb (2007) ;
- Le gain au niveau de la reconnaissance du sens nécessite plus d'expositions : 2 expositions pour 45% des mots cibles dans l'étude de Rott (1999), 2 à 4 expositions pour 33% des mots cibles dans l'étude de Pellicer-Sánchez & Schmitt (2010), 8 à 10 expositions pour 54% dans l'étude de Waring & Takaki (2003) et 10 fois et au-delà pour 36% des mots cibles dans Pigada & Schmitt (2006) ;
- Plus d'expositions sont nécessaires pour qu'un apprenant puisse se souvenir de la forme et du sens d'un mot nouveau : 7 expositions pour le rappel de la forme de 27% des mots cibles dans l'étude de C. Chen & Truscott (2010) et 10 expositions pour le rappel du sens de 29% des mots cibles dans l'étude de Webb (2007) ; selon les résultats de Pellicer-Sánchez & Schmitt (2010), 10 à 17 expositions sont nécessaires (48% des mots cibles).
- Les gains des autres informations lexicales telles que la collocation et la syntaxe exigeraient encore plus d'expositions (ex. Pellicer-Sánchez & Schmitt, 2010; Webb, 2007; Webb et al., 2013).

²⁹ « Words will have to be met in many different contexts in order to develop mastery of the different word knowledge types, and this entails a long-term recursive approach to vocabulary learning. »

Ces résultats montrent qu'un apprenant ne retient pas les informations de différents aspects lexicaux d'un mot avec la même facilité. En d'autres termes, les efforts exigés pour obtenir des gains sur les différents aspects lexicaux ne sont pas les mêmes (N. Ellis, 1997; VanPatten, 1994). Le gain au niveau sémantique, par exemple, nécessite bien plus d'expositions par rapport au gain formel pour un mot uniquement rencontré en lecture.

En outre, Nagy et al. (1985) remarquent qu'un mot inconnu peut être acquis grâce à une seule exposition seulement si le concept du mot est bien connu par l'apprenant. Cela veut dire que, dans le cas opposé où le concept d'un mot est difficile ou nouveau, voire absent, l'apprenant aura plus de difficultés à apprendre le mot. En effet, les résultats de l'étude de C. Chen & Truscott (2010) montrent que, dans le cas de l'apprentissage des mots L2 inconnus non lexicalisés (dont les concepts sont absents dans le lexique mental de l'apprenant), jusqu'à sept expositions n'apportent pas de différence significative au niveau du gain sémantique tandis que pour les mots lexicalisés, le gain sémantique s'avère significativement différent à partir de trois expositions. A notre connaissance, il n'existe à l'heure actuelle pas d'études empiriques sur le seuil d'exposition minimal pour que l'acquisition sémantique d'un mot L2 non lexicalisé soit réalisée. Toutefois, il serait raisonnable d'en déduire que, dans le cadre naturel de la lecture faite pour le plaisir, l'apprentissage du sens des mots L2 non lexicalisés exige une exposition plus fréquente que pour celui des mots lexicalisés.

2.4.4 Volume de lecture nécessaire

Après la discussion sur la fréquence d'exposition pour apprendre un mot nouveau via la lecture, il est temps de nous interroger sur l'ampleur d'exposition nécessaire pour qu'un apprenant acquière une étendue lexicale satisfaisante.

Combien de mots faudrait-il alors connaître ? Selon Nation (2001), une connaissance d'environ 10 000 mots est nécessaire afin d'assurer une communication normale dans la vie courante en anglais. Pour Coady (1996), pour un apprenant en anglais L2, une bonne connaissance d'au moins 5000 mots, complétée par une bonne

compétence de lecture, est nécessaire pour comprendre des textes académiques et authentiques de niveau avancé.

Ces seuils semblent faciles à atteindre pour un apprenant natif en milieu homoglotte : Nagy & Anderson (1984) estiment que les écoliers américains, du niveau CE1 jusqu'à la classe 3^{ème}, peuvent être exposés à environ 88 000 mots par le biais des matériels scolaires imprimés ; Goulden et al. (1990) estiment qu'un adulte natif instruit connaît autour de 17 000 mots de base (« base word » qui correspond à « une entrée principale dans un dictionnaire » (Goulden et al., 1990, p. 345)).

En revanche, pour un apprenant L2, il s'agit d'une tâche beaucoup plus difficile à réaliser. Hill & Laufer (2003) estiment qu'un apprenant L2 doit lire plus de 8 millions de mots, ce qui représente grosso modo 420 œuvres, afin d'enrichir son vocabulaire de façon incidente de 2000 mots. Selon Nation (2014), si un apprenant d'anglais L2 veut acquérir 1000 mots par an via la lecture comme les natifs, il devrait lire annuellement de 3 millions de mots pour apprendre les 9 000 mots anglais les plus fréquents (Tableau 2.3). En outre, la lecture doit être réalisée dans un cadre contrôlé, c'est-à-dire que le seuil des mots non familiers dans les matériels de lecture ne dépasse pas 2%.

Famille de 1000-mot	Quantité de lecture (mot imprimé)	Minutes par semaine (200 mots par minute)
2 ^{ème} 1000	171 411	21 minutes
3 ^{ème} 1000	300 219	38 minutes
4 ^{ème} 1000	534 697	1 heure 5 minutes
5 ^{ème} 1000	1 061 382	2 heures 12 minutes
6 ^{ème} 1000	1 450 068	3 heures
7 ^{ème} 1000	2 035 809	4 heures 5 minutes
8 ^{ème} 1000	2 427 807	5 heures 3 minutes
9 ^{ème} 1000	2 956 908	6 heures 10 minutes

Tableau 2.3 – Volume de lecture en mot et en durée de temps hebdomadaire pour s'exposer autour de 12 fois aux familles de 1000-mots en anglais (Nation, 2014, p. 7)

Les chiffres mentionnés ci-dessus représentent un travail d'une ampleur considérable, une tâche ardue à accomplir pour les apprenants de L2. Notamment, pour la grande majorité des apprenants de L2, l'apprentissage a lieu dans un cadre hétéroglotte. Le fait d'apprendre une langue dans un milieu dont la langue parlée n'est pas la langue cible amplifie le problème du manque d'expositions régulières aux stimuli quotidiens. Sous un autre angle de vue, cela signifie que le développement du lexique en L2 est un processus encore plus long et complexe par rapport à celui en L1.

2.5 Éléments intervenant dans l'apprentissage incident du vocabulaire

Comme le résume Laufer (2017), « *Les apprenants ne font pas nécessairement attention aux mots non familiers dans l'input. Quand ils les remarquent, ils n'infèrent pas nécessairement leurs sens. Quand ils le font, ils ne retiennent pas nécessairement les mots inférés. Enfin, l'exposition massive à la L2, une condition nécessaire pour 'repérer et sélectionner' des mots (pour travailler dessus), on ne peut pas s'attendre à ce qu'elle se produise dans un contexte d'apprentissage en classe.* (p. 344)³⁰ » Sternberg (1987), le grand partisan de l'approche d'apprentissage du vocabulaire grâce au contexte, a reconnu pour sa part qu'il serait imprudent de conclure que l'apprentissage incident est le moyen le plus rapide ou le plus efficace pour apprendre tout type de vocabulaire.

Cependant, nous ne devrions pas nous passer de cette approche d'apprentissage du vocabulaire. La raison est simple : il n'est pas question d'apprendre tous les mots en classe en utilisant les méthodes pédagogiques traditionnelles.

Des chercheurs se sont alors mis à explorer des éléments qui peuvent intervenir dans l'apprentissage du vocabulaire par cette piste.

³⁰ « Learners do not necessarily notice unfamiliar words in the input. When they do, they do not necessarily infer their meaning. When they do, they do not necessarily retain the guessed words. Finally, a necessary condition for 'picking up' words, massive exposure to the foreign language, cannot be expected to occur in classroom learning context. »

2.5.1 Attention portée à un mot

Selon le résumé de Laufer (2017) mentionné plus haut, afin qu'un mot nouveau soit appris, il faudrait avant tout que l'apprenant ait conscience de la présence du mot dans un texte puis qu'il porte suffisamment d'attention à ce mot. Autrement dit, l'attention que l'apprenant accorde à un mot inconnu est primordiale dans l'apprentissage incident de celui-ci (ex. N. Ellis, 1994 ; Gass, 1988).

Dans sa discussion sur la conscience de l'apprenant dans l'input, un grand sujet dans l'apprentissage incident, Schmidt (1990) distingue à grands traits trois niveaux de conscience – l'intention, la connaissance (implicite) et l'attention³¹. L'attention comprend trois niveaux – percevoir, constater (« (focal) noticing ») et réfléchir pour comprendre (« understanding »). Schmidt explicite ainsi la différence entre « percevoir » et « constater » : pendant une lecture, on se rend généralement compte de ce qu'on lit (« constater »), sans faire attention à la mise en page du texte même si on la voit (« percevoir »). Quant à l'autre niveau de l'attention, réfléchir pour comprendre qui comporte les actes tels que les analyses d'informations et la compréhension de l'importance d'un élément constaté, il renvoie à un acte volontaire de la part de l'apprenant fortement liée à l'objectif à atteindre ou à la motivation.

Dans sa revue de littérature, Schmidt (2001) note que l'attention d'un apprenant est limitée, donc allouée de façon sélective et stratégique aux stimuli (VanPatten, 1994) et soumise à un contrôle volontaire (LaBerge, 1995) ; peu de gain, voire aucun, ne peut résulter d'un apprentissage sans attention (Baars, 1996) parce que les stimuli auxquels on ne fait pas attention ou peu attention³² se préservent au mieux seulement quelques secondes dans la mémoire à court terme ; l'attention est donc la condition nécessaire et suffisante pour que s'opère l'enregistrement des informations dans la mémoire à long terme (Gass, 1988; Schmidt, 1995; VanPatten, 1994).

Selon l'hypothèse de Noticing (« Noticing hypothesis ») de Schmidt (1990,

³¹ Les trois termes utilisés par Schmidt (1990) sont « intention », « knowledge » et « awareness ». Schmidt (2001) remplace « awareness » par le terme psychologique « attention », le mécanisme qui contrôle l'accès à « awareness ».

³² Le terme utilisé par Schmidt (2001) est « unattended stimuli ».

2010), si l'apprenant ne prend pas conscience d'un élément, l'input n'aboutit pas à la saisie. L'étude de Godfroid & Schmidtke (2013) auprès d'apprenants du niveau avancé en anglais L2 a fourni une preuve solide. Dans cette étude, Godfroid & Schmidtke (2013, p. 183) définissent trois niveaux d'attention³³ : l'attention auto-noétique (« auto-noetic awareness », c'est-à-dire l'extraction d'une mémoire épisodique³⁴), l'attention noétique (« noetic awareness », c'est-à-dire un sentiment de familiarité³⁵) et non-attention (« no awareness »). L'oculométrie et la « pensée à voix haute » (à savoir le rapport verbal simultané des pensées)³⁶ ont été employées dans l'étude de Godfroid & Schmidtke pour observer l'attention des apprenants sur les mots L2 inconnus rencontrés en lecture. Surveillés par l'oculométrie, les participants, néerlandais L1, ont lu 20 paragraphes en anglais à fin de compréhension, dont 12 contiennent chacun un mot cible. Lors d'un post test sous forme d'exercice lacunaire à leur insu, on a demandé aux participants de choisir, parmi 18 mots proposés, le bon mot cible pour compléter la même phrase dans laquelle ils ont rencontré le mot cible. A l'issue de ce test, les participants ont effectué une tâche de reconnaissance des mots cibles avec la méthode de la « pensée à voix haute ». Cette tâche avait pour but de tester si les participants avaient remarqué les mots cibles pendant la lecture des 20 passages. Les résultats manifestent une corrélation positive entre la durée du regard et le score de reconnaissance des mots cibles. D'autant plus, les participants passent en moyenne 306 millisecondes de plus pour lire les mots cibles qu'ils déclarent plus tard une reconnaissance auto-noétique. Godfroid & Schmidtke ont conclu alors que pour les apprenants du niveau avancé, leur attention aux mots inconnus a un effet positif sur l'apprentissage incident de ces mots.

Schmidt (2001) met en avant que le fait que l'apprenant constate un élément linguistique, par exemple un mot inconnu dans un texte, est la première étape dans la construction linguistique (« language buiding »), mais pas la dernière. En s'appuyant sur Baars (1986), Schmidt (2001) indique qu'au lieu de chercher à savoir si l'apprentissage peut avoir lieu sans attention ou non, il est plus important de savoir si

³³ Godfroid & Schmidtke (2013) utilisent le terme « awareness ».

³⁴ « retrieval of an episodic memory »

³⁵ « a sense of familiarity »

³⁶ Nous présenterons ces deux méthodes expérimentales dans la Section 4.3 du Chapitre IV.

plus de gains d'apprentissage peuvent résulter d'une attention plus soutenue. Puisque les objectifs et la motivation sont des déterminants importants de l'attention (Baars, 1986, cité par Schmidt, 2001), selon Schmidt, l'accord d'une attention délibérée aux aspects moins saillants ou redondants dans l'input L2 est peut-être une nécessité pratique. Cela justifie le besoin des activités ou des tâches guidées dont les exigences orienteraient le focus attentionnel de l'apprenant vers des choses auxquelles il est moins susceptible de faire attention ou de remarquer par lui-même (par exemple les propriétés linguistiques de la L2 qui n'existent pas dans la L1, ou bien les informations lexicales qui doivent être traitées différemment en L2 et en L1).

2.5.2 Engagement de l'apprenant

Le facteur lié à la motivation de l'apprenant et aux traitements d'informations mentionné dans l'hypothèse de Schmidt est accentué et développé dans l'hypothèse de la charge d'engagement (« Involvement load hypothesis ») proposée par Laufer & Hulstijn (2001, p. 15). Selon cette hypothèse, trois facteurs sont déterminants dans l'apprentissage incident du vocabulaire :

- 1) Le besoin (« need ») : la nécessité de connaître le sens d'un mot inconnu pour comprendre correctement le texte.
- 2) La recherche d'informations (« search ») : la recherche des informations concernant la signification du mot.
- 3) L'évaluation (« evaluation ») : la vérification des sens possibles du mot d'après le contexte afin de choisir celui le plus approprié.

Laufer & Hulstijn nomment l'intégration de ces trois facteurs « engagement » (« involvement »), un concept motivationnel-cognitif. Ils définissent la charge d'engagement (« involvement load ») comme la combinaison de la présence ou de l'absence des trois facteurs.³⁷ Les chercheurs expliquent que le besoin est la dimension motivationnelle dans l'engagement. Il s'agit de la volonté de l'apprenant de se conformer aux exigences d'une tâche qui peuvent être imposées par un agent externe ou par l'apprenant lui-même. Les deux autres facteurs, la recherche

³⁷ Schmitt (2008) critique le fait que Laufer & Hulstijn (2001) n'ont pas pris en compte la motivation et l'attitude de l'apprenant et ses comportements stratégiques durant sa lecture.

d'informations et la vérification du sens du mot dans un contexte donné, relèvent de la dimension cognitive (qui se réfèrent ici strictement au traitement d'informations, à l'exclusion de l'aspect affectif de la cognition) de l'engagement.

Selon ces chercheurs, les deux facteurs de la dimension cognitive sont concernés dans le cas où l'apprenant remarque la présence d'un mot inconnu et accorde délibérément de l'attention au lien forme-sens pour ce mot. La dimension cognitive concerne la profondeur du traitement d'informations. Selon l'hypothèse de Craik & Lockhart (1972, cité par Laufer & Hulstijn, 2001) concernant la profondeur de traitement (« depth of processing »), la chance pour qu'une nouvelle information soit enregistrée dans la mémoire à long terme ne dépend pas de la durée de sa présence dans la mémoire à court terme, mais de la profondeur du traitement initial de cette information. En d'autres termes, plus l'apprenant traite les informations de façon profonde, plus il est possible qu'il mémorise ces informations. En effet, des études montrent que les apprenants retiennent mieux le vocabulaire dont le sens a été inféré par eux-mêmes que lorsqu'il est fourni à l'avance (ex. Hulstijn, 1992; Mondria & Wit-De Boer, 1991). Selon cette hypothèse, afin de réaliser les inférences, l'apprenant est mieux engagé cognitivement, alors que son engagement est nul lorsque le sens des mots inconnus lui sont donnés à l'avance. En outre, Laufer & Hulstijn indiquent que la dimension cognitive a une plus grande importance que la dimension motivationnelle parce que « *[L]es performances de la mémoire sont déterminées beaucoup plus par la nature des activités de traitement engagées par l'apprenant que par la volonté d'apprendre en soi.* (Eysenck, 1982, p. 203, cité par Laufer & Hulstijn, 2001, p. 6)³⁸ »

Ainsi, Laufer & Hulstijn (2001) mettent en avant que la charge d'engagement de l'apprenant peut expliquer et prédire le résultat d'apprentissage d'un mot inconnu de façon incidente : les autres facteurs étant égaux, 1) les mots traités avec un plus grand engagement de la part l'apprenant seraient mieux retenus et 2) les tâches conçues par les enseignants ou les chercheurs qui exigent davantage d'engagement de l'apprenant seraient plus effectives dans la rétention du vocabulaire.

³⁸ « [M]emory performance is determined far more by the nature of the processing activities engaged in by the learner than it is by the intention to learn per se. »

Les résultats de l'étude de Hulstijn et al. (1996) confirment l'idée de Laufer & Hulstijn (2001). Hulstijn et ses collègues comparent trois groupes d'étudiants universitaires néerlandais qui travaillent sur des mots français dans trois conditions différentes, a) avec une liste de lexique imprimé sur la marge de page à droite (on considère donc que ces participants travaillent sur les mots cibles avec le moins d'engagement par rapport aux autres participants), b) avec l'accès à un dictionnaire et c) sans aide (le groupe d'inférence lexicale). Les résultats des post tests de vocabulaire immédiats sur les mots cibles, à l'insu des participants, semblent contredire l'idée de l'hypothèse de la charge d'engagement (Laufer & Hulstijn, 2001) à première vue : le groupe avec le lexique en marges du texte retient mieux les mots cibles que les deux autres groupes. Hulstijn et ses collègues expliquent que, dans leur étude, on n'a pas explicitement demandé aux participants du groupe sans aide d'inférer le sens des mots inconnus présents dans le texte, tandis que les participants avec accès à un dictionnaire ont peu utilisé le dictionnaire pendant la lecture. Il est possible que 1) les participants de ces deux groupes se concentrent tellement sur la compréhension globale du texte qu'ils n'ont pas remarqué tous les mots inconnus présents ; 2) dans le cas où ils repèrent des mots inconnus, ils n'éprouvent pas de nécessité de connaître les sens de ces mots ou ils préfèrent garder une continuité de lecture sans être interrompus. Autrement dit, ils ont moins besoin de connaître le sens des mots inconnus remarqués et, par conséquent, ils sont moins « motivés ». Ces résultats soutiennent les hypothèses de Schmidt (1990, 2001) et de Laufer & Hulstijn (2001). Hulstijn et al. (1996) ajoutent pareillement que beaucoup d'éléments peuvent influencer le résultat de l'inférence lexicale (cf. aussi la Section 1.3.3) et que dans le cas où les participants ont tenté d'inférer le sens des mots inconnus, ils ont possiblement produit des inférences erronées pour telle ou telle raison. Cela explique aussi pourquoi le groupe sans aide a eu un résultat inférieur à celui du groupe avec un lexique en marges.

Il est à noter que les participants de l'étude de Hulstijn et al. (1996) ayant fréquemment utilisé le dictionnaire ont obtenu de meilleurs résultats de rétention que ceux ayant travaillé avec le lexique en marges du texte. Cela est conforme à l'hypothèse de Laufer & Hulstijn (2001) et montre qu'un besoin plus important de

connaître le sens d'un mot donné et un traitement d'informations plus profond de ce mot contribuent à la rétention du mot. La même conclusion a été obtenue par l'étude de Fraser (1999). Dans cette étude, les participants francophones devaient lire des textes en anglais pour la compréhension et un dictionnaire bilingue leur était mis à disposition. Lors de l'entretien individuel immédiatement après chaque lecture, les participants devaient montrer les mots jugés inconnus dans le texte et rapporter oralement de façon rétrospective ce qu'ils faisaient de ces mots pendant la lecture. D'après les réponses des participants, l'inférence lexicale est la stratégie principale utilisée à la rencontre d'un mot inconnu et la consultation d'un dictionnaire est aussi courante. 45% des utilisations d'un dictionnaire se sont faites à la suite d'une inférence du sens d'un mot. Les résultats des tests de vocabulaire administrés une semaine après la lecture montrent que les participants se rappellent environ 30% des mots pour lesquels ils ont inféré le sens ou consulté le dictionnaire, mais 50% des mots pour lesquels ils ont inféré le sens et puis consulté le dictionnaire.

Des études visant le chinois L2 ont abouti à des résultats similaires. Par exemple, Y. Zhu (2004) a adapté le protocole de Hulstijn et al. (1996) pour comparer les résultats d'apprentissage incident du vocabulaire chinois L2. Dans cette étude, les participants du niveau intermédiaire en chinois doivent lire afin de comprendre un texte dans l'une des trois conditions, avec un lexique en marges du texte, avec l'accès à un dictionnaire et sans aide. Immédiatement après la lecture, on montre une liste de mots comportant les 5 mots cibles aux participants. Les participants doivent juger s'ils ont vu ces mots pendant la lecture et donner ensuite le sens des mots. Le même test est administré auprès des participants une semaine plus tard. Selon les résultats, les participants ont eu des gains lexicaux sur les mots cibles dans ces trois conditions d'apprentissage. Le groupe de lexique devance significativement les deux autres groupes sur l'acquisition immédiate (la différence entre le groupe dictionnaire et le groupe sans aide est anodine) ; cependant, les scores du deuxième test ne manifestent plus de différences significatives parmi les trois groupes : par rapport aux scores d'acquisition immédiate, les scores des groupes de lexique et de dictionnaire ont diminué tandis que le score du groupe sans aide a légèrement augmenté. Selon Zhu, les gains lexicaux subséquents à l'apprentissage grâce au contexte sont les mieux

retenus parce que cette condition d'apprentissage est celle, parmi les trois, qui implique le traitement d'informations le plus profond.

Également sur le chinois L2, l'étude de H. Fan (2012) a été menée auprès de Chinois adultes et d'apprenants L2 de diverses L1. Tous les participants doivent lire un texte pour la compréhension et passer immédiatement après un test de vocabulaire qui évalue leurs connaissances sur les 6 mots cibles intégrés dans le texte. Par la suite, on demande aux participants d'inférer le sens des mots cibles présents dans des textes et on montre aux participants le sens correct de ces mots après la tâche d'inférence. Une semaine plus tard, on évalue une deuxième fois les connaissances chez les participants pour les mots cibles présentés d'abord isolément, puis dans le même contexte de lecture. Les résultats montrent que l'apprentissage des mots chinois subséquent à l'inférence lexicale se produit en L2 comme en L1. Il est intéressant de voir qu'immédiatement après la première lecture, les apprenants L2 ont reconnu en moyenne la forme de 4 mots cibles sur 6 et ne connaissaient guère le sens de ces mots, tandis que les natifs ont déjà eu un gain sémantique significatif à ce stade ; toutefois, les participants L2 sont parvenus à produire des inférences correctes et à se rappeler le sens de certains mots cibles au bout d'une semaine. Autrement dit, la tâche d'inférence lexicale a orienté explicitement l'attention des participants L2 vers tous les mots inconnus et créé pour eux un « besoin » de connaître les sens de ces mots, ce qui a incité par la suite les participants à s'investir cognitivement dans les traitements d'informations lexicales.

2.6 Intégration des instructions dans l'apprentissage incident du vocabulaire

A partir des deux hypothèses présentées ci-dessus, nous pouvons synthétiser ainsi le processus cognitif d'apprendre de façon incidente un mot inconnu qu'un apprenant rencontre en lecture : l'apprenant perçoit un mot inconnu présent dans sa lecture et prend conscience qu'il ne connaît pas le mot ; l'apprenant est motivé pour connaître le sens du mot parce que ce mot est important pour la compréhension du texte ou la tâche à réaliser l'oblige à connaître le sens du mot ; par la suite, l'apprenant se met à la recherche d'indices, vérifie les informations repérées, et reprend le cycle

recherche-vérification si nécessaire, afin d’essayer de déduire le sens du mot ; ces opérations aboutissent enfin aux gains lexicaux dans la mémoire de l’apprenant (Figure 2.1). Les résultats des études présentées plus haut montrent communément qu’afin d’obtenir un meilleur résultat d’apprentissage incident du vocabulaire via la lecture, il est important d’établir une condition d’apprentissage qui incite l’apprenant à porter davantage d’attention aux mots inconnus et lui permette de travailler avec un investissement cognitif important sur les mots. Des chercheurs travaillant sur l’apprentissage du vocabulaire en L2 ont alors examiné les approches liées aux interventions pédagogiques.

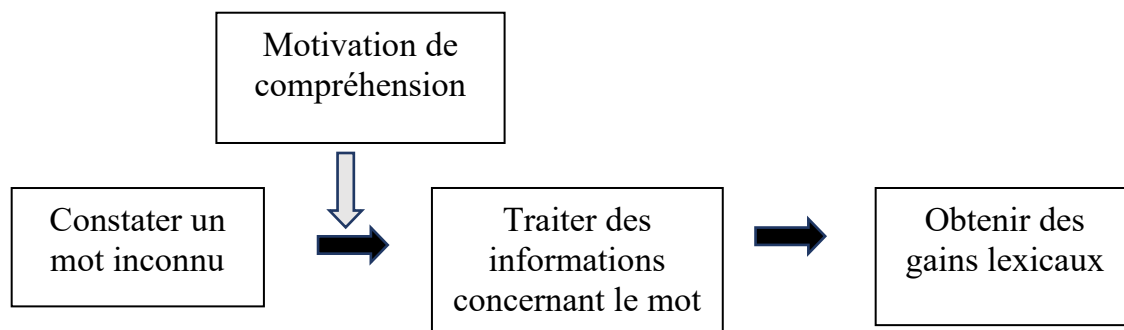


Figure 2.1 – Schéma théorique du processus d’apprentissage d’un mot via la lecture

2.6.1 Approche « Instruction centrée sur la forme »

Dans le domaine de la linguistique appliquée, Long (1988) a évoqué pour la première fois l’approche « centrée sur la forme » (« Focus on form »). Le terme « forme » comprend les aspects phonologique, lexical, grammatical et pragmalinguistique (R. Ellis, 2001, p. 1). Ce concept a été bien développé pendant les années suivantes au fil de l’avancée des recherches, en s’inspirant notamment de l’hypothèse de noticing de Schmidt (R. Ellis, 2016). Essayant de discerner l’approche « centrée sur la forme » et l’autre approche « centrée sur les formes » (« Focus on forms ») qui a émergé plus tard, Long (1998, p. 40) explique que, dans le contexte « centrée sur les formes », les connaissances linguistiques sont segmentées sous différentes formes, par exemple phonème, mot, collocation, morphème, phrases, et ensuite enseignées dans l’ordre de priorité déterminée par l’enseignant. L’apprenant

est censé tout apprendre d'un coup et il lui revient de synthétiser la/les partie(s) à utiliser dans la communication ; en revanche, l'approche « centrée sur la forme » se réfère à la méthode qui attire l'attention des apprenants vers les éléments linguistiques (mots, collocations, structures grammaticales, schémas pragmatiques, etc.) dans une communication, dans laquelle les changements temporaires d'attention focale des apprenants sont déclenchés par les problèmes de compréhension ou de production.³⁹ De ce point de vue, l'approche « centrée sur les formes » est impliquée dans l'enseignement structural, un moyen plutôt « traditionnel », et l'approche « centrée sur la forme » est orientée vers la tâche ou vers la résolution de problèmes. Un autre terme, « Instruction centrée sur la forme » (« Form-focused instruction ») est « ...le terme générique largement utilisé pour désigner toute technique pédagogique, proactive ou réactive, implicite ou explicite, utilisée pour attirer l'attention des apprenants sur la forme linguistique (Long, 1998, p. 41).⁴⁰ »

R. Ellis (2001) met en relief le rôle de l'apprenant pour distinguer ces deux approches : l'apprenant étudie la langue dans le contexte « centrée sur les formes », mais utilise la langue dans le contexte « centrée sur la forme ». Il étiquète respectivement ces deux approches « intentionnelle » et « incidente ». Selon Ellis, « *L'instruction centrée sur la forme comprend à la fois des approches traditionnelles des formes d'enseignement basées sur des programmes structurels et des approches plus communicatives, où l'attention à la forme découle d'activités qui sont principalement axées sur le sens.* (p. 2)⁴¹ »

2.6.2 Apprentissage incident du vocabulaire avec l'approche « instruction centrée sur la forme »

Adoptant la théorie que la mémorisation délibérée au moment de l'input est le facteur pivot pour distinguer l'apprentissage incident de l'apprentissage intentionnel

³⁹ Selon R. Ellis (2016), l'interprétation sur l'approche « centrée sur la forme » de Long renvoie bien à son concept de « l'enseignement des langues basé sur tâche » (« task-based language teaching », Long & Crookes, 1992).

⁴⁰ « ... an umbrella term widely used to refer to any pedagogical technique, proactive or reactive, implicit or explicit, used to draw students' attention to language form. »

⁴¹ « FFI [Form-Focused Instruction] includes both traditional approaches to teaching forms based on structural syllabi and more communicative approaches, where attention to form arises out of activities that are primarily meaning-focused. »

(Hulstijn, 2013; Laufer & Hulstijn, 2001), Laufer (2006) estime que les approches « centrée sur les formes » et « centrée sur la forme » peuvent avoir lieu toutes les deux dans l'apprentissage intentionnel et incident du vocabulaire. Elle a sollicité 158 lycéens israéliens pour travailler sur 12 mots cibles en anglais L2, la moitié avec l'approche « centrée sur les formes » (« FormeS » en abrégé) et l'autre avec l'approche « centrée sur la forme » (« Forme » en abrégé). Le groupe « FormeS » est muni d'une liste des mots cibles avec leur sens en anglais et le groupe « Forme » lit un texte avec l'accès à un dictionnaire. Tous les participants doivent ensuite accomplir deux exercices de vocabulaire – le choix du sens correct parmi quatre options en anglais et un exercice lacunaire. A l'issue de cette première partie que Laufer nomme « l'apprentissage incident », tous les participants reçoivent un cours de 90 minutes pour travailler sur ces mots. On demande aux participants de mémoriser les 12 mots en 15 minutes avant de passer un test de vocabulaire annoncé en amont. La deuxième partie est donc définie comme « l'apprentissage intentionnel ». Les résultats de la tâche de traduction des mots cibles, effectuée immédiatement après l'apprentissage incident, dévoilent que le groupe « FormeS » (72% des mots cibles dont le sens est retenu) devance le groupe « Forme » (47%). En revanche, cet écart diminue après l'apprentissage intentionnel, respectivement 68% (le groupe « FormeS ») vs 60% (le groupe « Forme ») pour le rappel du sens et 87% (le groupe « FormeS ») vs 88% (le groupe « Forme ») pour le rappel de la forme. Le deuxième post test deux semaines plus tard montre des scores similaires entre les deux approches, toutes les deux autour de 50% pour le rappel du sens et de 60% pour le rappel de la forme. Laufer est donc arrivé à la conclusion que les approches « centrée sur les formes » et « centrée sur la forme » sont bénéfiques toutes les deux pour l'apprentissage incident du vocabulaire. Elle souligne notamment la nécessité de l'approche « centrée sur les formes », qui permet de travailler uniquement sur le vocabulaire, sans qu'il soit dans un contexte communicatif, et qui promeut efficacement les compétences lexicales chez un apprenant.

Hong & Wang (2016), auprès d'apprenants de chinois L2, arrivent à la même conclusion sur la nécessité de l'approche « centrée sur les formes » dans l'apprentissage du vocabulaire. Hong & Wang comparent deux groupes travaillant

sur 10 mots cibles. Le groupe « centrée sur la forme » (« Forme » en abrégé) doit lire un texte comportant les mots cibles, avec l'accès à un dictionnaire (en papier ou en ligne), avant de répondre aux questions de compréhension, tandis que le groupe « centrée sur les formes » (« FormeS » en abrégé) doit travailler sur une liste du lexique avant d'accomplir les exercices de vocabulaire concernant les mots cibles. Deux post tests de vocabulaire, l'un immédiatement après et l'autre une semaine plus tard, ont été administrés pour évaluer les connaissances sur ces mots chez les apprenants. Aux deux post tests, le groupe « FormeS » a de meilleurs scores que le groupe « Forme », mais les différences entre les deux groupes au niveau des connaissances formelle (reconnaissance), syntaxique et sémantique ne sont pas significatives ; toutefois, le groupe « FormeS » devance significativement le groupe « Forme » au niveau du gain dans les aspects de collocation et d'association. Hong & Wang expliquent que les exercices dans le contexte « centrée sur les formes » permettent aux apprenants de se concentrer sur les informations lexicales des mots concernés sans partager leurs ressources cognitives, qui sont limitées, entre plusieurs tâches en parallèle. Avec l'approche « centrée sur les formes », les apprenants ont plus de possibilités d'acquérir des informations lexicales sur un mot exigeant des traitements profonds, par exemple la collocation. Dans l'étude de Laufer (2006), la différence au niveau du gain sémantique entre les deux groupes est significative, mais ce n'est pas le cas dans cette étude. Selon l'étude de Hong & Wang, pour répondre aux questions de compréhension, les apprenants doivent comprendre le sens des mots cibles concernés, ce qui n'est pas nécessaire dans l'étude de Laufer (2006). Cela démontre à nouveau que le besoin de comprendre le sens d'un mot inconnu est un facteur important dans l'apprentissage du mot dans la condition incidente.

L'effet bénéfique de l'approche « instruction centrée sur la forme » est aussi reconnu par l'étude de Laufer & Rozovski-Roitblat (2015) auprès de lycéens israéliens qui apprennent l'anglais. L'objectif de l'étude est d'observer l'effet individuel et superposé de deux paramètres – le type de la tâche à réaliser et la fréquence d'exposition à un mot inconnu. Les participants sont exposés à 30 mots cibles à trois fréquences différentes, respectivement dans trois conditions d'apprentissage – la lecture seule (le groupe de contrôle), la lecture avec accès à un

dictionnaire (le groupe « Forme ») et la lecture avec des exercices de vocabulaire post-lecture (le groupe « FormeS »). Deux semaines plus tard, tous les participants ont reçu les tests évaluant leurs connaissances passives et actives sur les mots cibles. Les résultats montrent un avantage évident du groupe « FormeS » par rapport aux deux autres groupes, quelle que soit la fréquence d'exposition aux mots cibles. Le groupe « Forme » devance le groupe de contrôle sur les gains lexicaux sur les mots cibles, mais les scores sont globalement décourageants par rapport au groupe « FormeS ». Les chercheuses concluent alors que le type de la tâche à réaliser a un effet plus important que la fréquence d'exposition dans l'apprentissage incident du vocabulaire en L2. Les activités dans le contexte « instruction centrée sur la forme », notamment les exercices de vocabulaire en particulier, doivent donc être prises en compte lors de la conception des matériels et des programmes pédagogiques.

Laufer & Hulstijn (2001) indiquent que « *Depuis des années, les pratiques d'instructions sont fondées sur l'opinion, soutenue à l'unanimité par les psychologues, qu'un travail bien élaboré sur les caractéristiques d'un mot nouveau favorise sa rétention.* (p. 1)⁴² » Avec l'approche « instruction centrée sur la forme », à force d'accomplir des tâches spécifiques orientant l'attention des apprenants vers divers aspects d'un mot nouveau, l'apprenant doit se concentrer davantage sur la forme du mot et ressent le besoin de connaître le sens du mot et de vérifier son utilisation dans différents contextes (ce qui augmente du même coup l'exposition de l'apprenant au même mot). Cela expliquerait que les résultats généraux des études réalisées montrent un effet positif de l'approche « instruction centrée sur la forme » dans l'apprentissage incident du vocabulaire. (Cf. aussi Laufer & Rozovski-Roitblat, 2011, et Peters et al., 2009)

2.7 Récapitulation et questions de recherche sur l'apprentissage subséquent du vocabulaire inféré

L'inférence lexicale n'est pas seulement une stratégie de lecture, mais

⁴² « For many years, instructional practice has been based on the view, unanimously supported by psychologist, that **elaboration** on features of new words promotes their retention. »

également une stratégie d'apprentissage du vocabulaire. En L1, des études ont démontré que les enfants parviennent à inférer le sens des mots nouveaux dans une lecture grâce au contexte et que la lecture extensive sans intervention pédagogique est le mode majeur pour eux de développer leur vocabulaire. En L2, de nombreuses études empiriques ont également montré que l'inférence du sens des mots inconnus présents dans un texte peut aboutir à des gains lexicaux sur ces mots.

Cependant, en L2 comme en L1, les gains lexicaux subséquents à la déduction du sens d'un mot grâce au contexte sont généralement modestes pour deux raisons principales : 1) pendant la lecture, l'apprenant n'accorde pas son attention à tous les mots inconnus présents dans le texte, ni parvient à comprendre tous les mots inconnus pour lesquels il tente de déduire le sens ; 2) la connaissance d'un mot est multi-dimensionnelle et l'acquisition de certains aspects lexicaux, tels que l'aspect sémantique, exige des traitements cognitifs profonds. L'acquisition des informations lexicales diverses d'un mot implique qu'une bonne fréquence d'exposition au mot dans différents contextes est nécessaire, tout comme est indispensable un volume considérable de lecture afin d'acquérir un vocabulaire satisfaisant. Le développement du lexique est donc un processus long et difficile et ce problème en L2 est accentué par le manque de stimuli, parce que l'apprentissage d'une L2 a lieu majoritairement dans un contexte hétéroglotte.

Malgré cela, l'apprentissage du vocabulaire grâce au contexte offre une autre piste aux apprenants L2 pour enrichir leur vocabulaire en dehors des classes traditionnelles, qui sont normalement limitées en nombre d'heures d'enseignement et donc non suffisantes pour une croissance du vocabulaire satisfaisante des apprenants.

Il est généralement admis de nos jours que l'attention qu'un apprenant porte à un mot inconnu et la profondeur du traitement des informations lexicales du mot sont deux facteurs importants dans l'apprentissage du vocabulaire. Des chercheurs mettent en avant qu'il serait nécessaire de créer une condition d'apprentissage plus renforcée afin d'améliorer l'apprentissage d'un mot de façon incidente. Des études ont démontré que, dans la condition d'apprentissage incident, l'utilisation de l'approche « instruction centrée sur la forme » peut orienter l'attention d'un apprenant sur un

mot donné et permettre à l'apprenant de traiter en profondeur des informations lexicales sur le mot ; en outre, le type de la tâche à réaliser concernant le mot a un effet important sur les gains lexicaux sur le mot.

Nous avons expliqué dans le Chapitre I que notre étude vise le chinois L2. En dehors de l'effet de la lexicalisation sur le processus et le résultat de l'inférence lexicale, nous nous sommes également penchée sur l'apprentissage subséquent des mots inférés, en gardant toujours notre intérêt sur l'effet de la lexicalisation. Focalisant sur les deux facteurs principaux dans l'apprentissage incident du vocabulaire, l'attention et la profondeur de traitement dans l'input, nous voulions savoir si une condition d'apprentissage cognitivement plus chargée aboutit à des meilleurs gains lexicaux sur un mot chinois L2 et si la nature de lexicalisation du mot a un effet sur l'attention qu'un apprenant porte sur le mot pendant la lecture et sur les gains lexicaux sur ce mot à court et à long terme.

Pour cela, la deuxième partie de notre étude a été menée pour chercher des réponses aux questions de recherche suivantes :

QR 3 : Pendant une lecture en chinois pour la compréhension, l'attention d'un apprenant francophone portée à un mot chinois inconnu présent dans la lecture diffère-t-elle selon que le mot est lexicalisé ou non en français ?

QR 4 : A l'issue de l'inférence du sens d'un mot chinois inconnu présent dans une lecture, une combinaison de deux tâches concernant le mot, par rapport à une seule tâche, aboutit-elle à de meilleurs gains lexicaux sur le mot à court et à long terme ? La nature de lexicalisation du mot a-t-elle un effet sur les gains lexicaux sur le mot à court et à long terme ?

Nous avons mentionné dans le Chapitre I que, jusqu'à présent, la recherche sur la lexicalisation est rare et que peu d'études ont été réalisées à ce sujet. A notre connaissance, au moins au moment de la réalisation de notre expérience, il n'y avait pas eu d'études examinant la différence de l'attention qu'un apprenant accorde aux mots inconnus selon leur nature de lexicalisation. La revue de littérature ne nous permettait pas d'avoir une hypothèse pour la QR 3. Quant à la QR4, voici nos hypothèses :

Hypothèses : L'inférence lexicale suivie d'une combinaison de deux tâches de vocabulaire post-inférence, par rapport à une seule tâche, aboutit à de meilleurs gains lexicaux sur les mots L2 inconnus, à court et à long terme. Un apprenant francophone obtient plus de gains lexicaux sur les mots chinois L2 lexicalisés en français, par rapport aux mots L2 non lexicalisés, à court et à long terme.

**CHAPITRE III – APPRENTISSAGE DU VOCABULAIRE EN
CHINOIS L2**

Introduction

Le centre d'intérêt de notre thèse est le rôle de la lexicalisation d'un mot L2 rencontré en lecture pendant le processus d'inférence du sens du mot grâce au contexte, ainsi que l'effet de la lexicalisation sur l'apprentissage subséquent du mot à court et à long terme. Pour cela, nous visons le cas où les apprenants francophones apprennent la langue chinoise principalement pour la distance linguistique et culturelle importante entre la langue chinoise et la langue maternelle des apprenants, en l'occurrence, la langue française.

Ce chapitre est dédié à la présentation des particularités de la langue chinoise, par rapport à une langue alphabétique telle que le français, et les influences de ces particularités sur l'apprentissage de chinois L2 dans la pratique et dans la recherche. Nous présenterons à la fin du chapitre les obstacles dans l'apprentissage du vocabulaire chinois que nous avons identifiés chez les apprenants francophones en France, en particulier les entraves liées à la lexicalisation des mots chinois. Nous terminerons ce chapitre avec nos réflexions à partir desquelles nous avons amené cette étude auprès d'apprenants francophones en chinois L2.

3.1 Enseignement du chinois L2 dans le monde et en France

Selon les chiffres du ministère de l'Éducation de la Chine⁴³, à la fin de l'année 2020, la langue chinoise est enseignée dans plus de 180 pays et régions du monde, parmi lesquels plus de 70 pays ont intégré la langue chinoise dans leur liste des langues officiellement enseignées dans les établissements scolaires. On estime aujourd'hui à plus de 20 millions, le nombre d'apprenants du chinois dans le monde (en dehors de la Chine). Au total, il existe de par le monde 525 Instituts Confucius (孔子学院, établissement public de l'enseignement de la langue chinoise), dont 17 en France.

⁴³ http://www.moe.gov.cn/fbh/live/2021/53486/mtbd/202106/t20210602_535133.html (consulté le 15 août 2021)

Suite à l'établissement de relations diplomatiques en 1964 entre la France et la Chine, les échanges multidimensionnels entre les deux pays sont fréquents depuis les années 1990. Selon les chiffres présentés dans le rapport de Joël Bellassen⁴⁴ (2015), alors que la Chine est la première destination des expatriés européens, la France a accueilli 2 millions de touristes chinois en 2014. Le chinois est la troisième langue, après le français et l'anglais, de la signalétique du Louvre. Comme l'indique Bellassen, l'émergence de la langue chinoise en France et dans le monde est inscrite dans le contexte de la dynamique économique de la Chine et la tendance à la mondialisation depuis ces dernières décennies. « *Hier langue exotique et lointaine, le chinois devient dans un tel contexte une langue dont la 'proximité' s'est sensiblement accrue et il convient au plan éducatif de prendre la mesure de ce phénomène inédit.* (p. 2) » Au sujet de l'enseignement du chinois en France, Bellassen (2015) résume que, jusqu'à fin 2015, il y a eu plus de 100 000 apprenants de cette langue en France dont environ 20% dans l'enseignement supérieur ; 28 départements universitaires dispensent la discipline LLCE (Langues, Littératures et Civilisations Etrangères) ou LEA (Langue Etrangère Appliquée) avec environ 7000 étudiants. Le nombre des élèves engagés dans le parcours réglementaire de chinois dans le Secondaire a connu une augmentation de 400% en 5 ans ; le chinois étant la 5^{ème} langue enseignée dans l'enseignement secondaire, le cours de chinois est proposé dans 663 collèges et lycées (seulement 203 en 2005) et la proportion du chinois LV1-LV2 est de 50%, alors que dans les années 1980, elle était seulement de 20%.

La popularisation de l'apprentissage de la langue chinoise en France, concrétisée par les chiffres impressionnants présentés ci-dessus, n'efface pas l'idée générale des Français sur le chinois que cette langue est difficile à apprendre (au point que l'expression « c'est du chinois pour moi » signifie « je n'y comprends rien »). En effet, les apprenants francophones de chinois rencontrent en général de grandes

⁴⁴ Joël Bellassen était, depuis 2006, premier inspecteur général de chinois au ministère de l'Éducation nationale jusqu'à sa retraite en 2016. Il est internationalement connu dans sa discipline depuis la publication de son livre *Méthode d'initiation à la langue et à l'écriture chinoises* (1989, 1^{ère} édition, La Compagnie), qui devint l'un des principaux ouvrages utilisés pour enseigner le chinois dans les établissements francophones, représentatif au niveau international de l'approche dite de l'entrée par le caractère. (Informations adaptées de https://fr.wikipedia.org/wiki/Jo%C3%ABl_Bellassen, consulté le 27 mai 2021.)

difficultés liées aux particularités de cette langue et de son écriture que nous allons présenter ci-dessous.

3.2 Caractéristiques de la langue chinoise

Le mot « langue chinoise » s'écrit 汉语 (*Hànyǔ*) ou 中文 (*Zhōngwén*) en chinois. Aujourd'hui, dans l'enseignement, le chinois renvoie au mandarin, langue officielle en Chine qui se base phonétiquement sur le dialecte pékinois et grammaticalement sur le style littéraire proche de celui oral du chinois moderne, 白话文 (*Báihuàwén*).

La langue chinoise n'est pas alphabétique à l'écrit et l'unité d'écriture minimale est le caractère, 字 (*zì*). Afin de faciliter l'apprentissage des caractères, le gouvernement chinois a lancé une campagne de simplification de caractères chinois dans les années 1950. Jusqu'au milieu des années 1980, 2 235 caractères au total ont été simplifiés au niveau de la forme (B. R. Huang & Liao, 2005). Par exemple, le caractère qui signifie « dragon » s'écrit 龍 (*lóng*) en version traditionnelle et 龙 (*lóng*) en version simplifiée. Les caractères traditionnels sont utilisés à Taiwan, à Macao et à Hongkong, tandis que les caractères simplifiés sont utilisés en Chine continentale et à Singapour. Aujourd'hui, l'enseignement du mandarin implique l'enseignement de cette langue avec des caractères simplifiés. (Cf. aussi W. Y. Song, 2017, pp. 25-28, pour la diachronie de la forme des caractères chinois)

3.2.1 Transcription phonétique

L'écriture en chinois n'est pas alphabétique, mais il existe un système de transcription phonétique en alphabet romain, 拼音 (*Pīnyīn*)⁴⁵, ou Pinyin. Adopté officiellement depuis la fin des années 1950 par le gouvernement chinois, Pinyin est un outil pour aider les apprenants à prononcer un caractère.

Les caractères chinois sont monosyllabiques. Le Pinyin d'un caractère chinois contient globalement une « initiale » (une ou deux consonnes), une « finale » (une

⁴⁵ Dans cette thèse, les Pinyin sont marqués en italique entre parenthèses juste derrière les caractères concernés.

seule voyelle ou une combinaison de voyelles ou de voyelle-consonne) et un ton (l'accent).

Ex. *zhōng* = « *zh* » (initiale) + « *ong* » (finale) + « - » (ton)
háo = « *h* » (initiale) + « *ao* » (finale) + « ' » (ton)

Certaines finales peuvent être utilisées seules pour transcrire la prononciation d'un caractère.

Ex. *áng* = « *ang* » (finale) + « ' » (ton)






Rarement, une ou plusieurs consonnes peuvent être utilisées pour transcrire la prononciation de certains caractères, par exemple 嗯 (*ńg*). Le ton est présent dans la grande majorité des transcriptions en Pinyin. Mais certains caractères se prononcent sans ton, par exemple 的 (*de*), 着 (*zhe*).

Les transcriptions phonétiques ne remplacent pas les caractères chinois dans les communications écrites.

3.2.2 Caractère

Un caractère, unité d'écriture minimale en chinois, est une symbiose unie de trois aspects, phonétique (音 *yīn*), graphique (形 *xíng*) et sémantique (意 *yì*) (S. X. Lü, 1980). Certains chercheurs sont donc d'avis qu'il serait plus pertinent de dire que la langue chinoise a une écriture « morpho-syllabique », qui met plus en saillance les trois propriétés, par rapport au terme « logographique »⁴⁶ (cf. Drocourt, 2015). Un caractère chinois n'a pas de genre et sa forme graphique n'envoie pas d'information morpho-grammaticale.

Les premiers caractères sont généralement pictographiques qui imitent ou symbolisent ce que les ancêtres voyaient.

Ex.  → 山 (*shān*, montagne)
 →  → (鳥) 鸟 (*niǎo*, oiseau)
 →  → (魚) 鱼 (*yú*, poisson)

⁴⁶ Dans notre thèse, nous employons le terme plus répandu de « logographique ».

Une partie des caractères chinois ont une forme mono-structurale (独体字 *dú tǐ zì*), un tronçon qu'on ne peut plus décomposer.

Ex. 在 山 日 木 人

Mais 80% des 7000 caractères chinois utilisés aujourd'hui sont des caractères multi-structurels (合体字 *hé tǐ zì*)⁴⁷, qui sont constitués de plusieurs unités composantes (*Liste des caractères communs du chinois moderne* (汉语通用字表), 2009, cité par Chang et al., 2014, p. 289).

Ex. 娘 (女 + 良) 尖 (小 + 大)
 赶 (走 + 干) 困 (口 + 木)

La majorité des caractères chinois utilisés aujourd'hui contiennent un bloc composant sémantique (形旁 *xíng páng*) et un bloc composant phonétique (声旁 *shēng páng*) (B. R. Huang & Liao, 2005, p. 186).

3.2.2.1 Bloc composant sémantique

Le bloc composant sémantique dans un caractère renvoie à la partie qui apporte des informations sur le sens du caractère, comme l'indique son nom. Par exemple, 水 (*shuǐ*, eau) peut être utilisé comme bloc composant dans 泉 (*quán*, source (d'eau)) ou sous une autre forme « 氵 » dans 湖 (*hú*, lac) et 河 (*hé*, fleuve). Autrement dit, le bloc composant « eau » est souvent présent dans les caractères dont le sens est en rapport avec le liquide. Le bloc composant sémantique est généralement la clé (部首 *bù shǒu*) dans un caractère multi-structurel par le biais de laquelle on cherche un caractère dans un dictionnaire, notamment dans le cas où on ne connaît pas la prononciation du caractère (Nous reviendrons sur ce point dans le Chapitre IV).

Cependant, les blocs composants sémantiques ne délivrent généralement qu'une partie du sens du caractère. Autrement dit, ils ne sont pas suffisamment informatifs. Par exemple, grâce au bloc composant 手 (*shǒu*, main) dans le caractère

⁴⁷ 合体字 (*hé tǐ zì*) est généralement traduit comme « caractère composé ». Afin d'éviter l'éventuelle confusion avec le terme « mot composé », présenté plus loin, nous utilisons dans cette thèse le terme « caractère multi-structurel » pour désigner la forme du caractère de cette catégorie.

掌 (*zhǎng*, paume), on comprendrait que ce caractère est en rapport avec la main ou une action faite avec les mains, mais sans plus. D'ailleurs, la transposition sémantique par les blocs composants n'est pas toujours fiable. Par exemple, 闻 (*wén*) contient le composant 耳 (*ěr*, oreille) et signifie « entendre » dans le chinois ancien. Aujourd'hui il signifie « flairer » dans le chinois moderne.

3.2.2.2 Bloc composant phonétique

Le composant phonétique apporte l'information sur la prononciation du caractère. Certains composants transposent la prononciation de façon entière,

Ex. 中 (*zhōng*) → 钟 (*zhōng*), 忠 (*zhōng*), 盅 (*zhōng*), 衷 (*zhōng*)

et certains plutôt de façon partielle.

Ex. 中 (*zhōng*) → 种 (*zhǒng, zhòng*), 肿 (*zhǒng*), 冲 (*chōng, chòng*)

Si on nomme les caractères composés d'une partie déterminant entièrement sa prononciation « caractères réguliers », ceux qui sont les plus couramment utilisés sont les caractères les moins réguliers (Shu & Anderson, 1999). (Cf. aussi Koda, 2005, pp. 79-84)

Il est à noter que, malgré le lien graphie-son relativement faible pour les caractères chinois, certaines études montrent que le son n'est pas complètement absent dans l'activation du lexique mental en chinois, notamment chez les enfants. Pour Song et al. (1995), les enfants chinois s'appuient davantage sur la voie phonétique pour reconnaître les caractères, tandis que les adultes passent plutôt par la voie graphique. Song et ses collègues expliquent que les enfants acquièrent leurs premiers mots notamment par les communications orales quotidiennes. Quand ils lisent, ils ont besoin de l'intermédiaire phonétique pour créer le lien graphie-sens. Plus ils sont compétents en langue et en lecture, plus le lien graphie-sens est renforcé jusqu'à ce que l'activation devienne quasi automatique sans aide phonétique. S. Zhang & Perfetti (1993) ajoutent pour leur part que le processus phonologique pour décoder des caractères est général dans la lecture en chinois pour un chinois adulte.

3.2.2.3 Caractères en cas d'homophonie, de polysémie et de polyphonie

Nous précisons que le terme « homophone » employé ici renvoie au cas dans lequel plusieurs caractères de différentes formes partagent une même prononciation, « polysémie » un caractère qui possède plusieurs sens et « hétérophonie » un caractère possédant plusieurs prononciations (qui correspondent généralement aux sens différents).

La relation graphie-son est loin d'être un à un en chinois. Une même prononciation peut correspondre à beaucoup de caractères différents. Par exemple, dans le *Dictionnaire chinois moderne* (现代汉语词典, 1984), « jùn » correspond à 13 caractères (ex. 郡, 俊, 骏, 隼) et « shān » à 18 caractères (ex. 山, 删, 衫, 芟). Par rapport à cela, un caractère peut avoir plusieurs sens et classes grammaticales différentes.

Ex. 点 (*diǎn*) n. : (1) point ; (2) aspect ; (3) heure ...
v. : (1) marquer d'un point (2) toucher (légèrement)
(3) commander ; (4) allumer ...

Un caractère peut également avoir plusieurs prononciations différentes qui correspondent aux différentes significations et souvent aux différentes classes grammaticales.

Ex. 冲 1. (*chōng*) n. : (1) centre de communication ; (2) conflit, collision
2. (*chōng*) v. : (1) rincer ; (2) infuser ; (3) franchir, heurter
3. (*chòng*) adj. : (1) vigoureux ; (2) fort (en odeur)
4. (*chòng*) prép. : vers, vis-à-vis de
没 1. (*méi*) adv. : ne...pas
2. (*mò*) v./n. : submerger, submersion

3.2.3 Mot

3.2.3.1 Caractère, morphème et mot

N'ayant pas de genre, la grande majorité des caractères chinois représentent chacun un morphème (语素 *yǔsù*), unité sémantique minimale dans la langue chinoise.

Ex. 书 (*shū*) : livre

Mais il existe des morphèmes constitués de plusieurs caractères qui n'ont pas de sens concret individuellement ou n'ont pas d'apport sémantique au mot. Les morphèmes constitués de plusieurs caractères sont minoritaires.

Ex. 什么 (*shénme*) pron. : quoi, que
踌躇 (*chóuchú*) v. : hésiter
巧克力 (*qiǎokèlì*) n. : chocolat

Comme ce que nous avons mentionné plus haut, un caractère est souvent polyvalent au niveau sémantique et grammatical. Il peut former seul un morphème ou former un autre morphème par association avec un autre caractère (cf. B. R. Huang & Liao, 2005, pp. 250-254). Dans l'exemple suivant, les deux caractères 东 et 西 peuvent être chacun un morphème avec un sens concret et également former ensemble un autre morphème 东西 avec un sens différent. C'est aussi le cas pour les caractères 马 et 虎.

Ex. 东 (*dōng*) n. : est 马 (*mǎ*) n. : cheval
西 (*xī*) n. : ouest 虎 (*hǔ*) n. : tigre
东西 (*dōngxi*) n. : chose, objet 马虎 (*mǎhu*) adj. : négligent, inattentif

Un mot chinois peut être constitué d'un seul ou de plusieurs morphèmes.

Ex. 书 (*shū*) : livre
书签 (*shūqiān*) : marque-page
图书馆 (*túshūguǎn*) : bibliothèque

L'évolution de la langue chinoise depuis plusieurs millénaires rend aujourd'hui les frontières entre « caractère » (字 *zì*), « morphème » (语素 *yǔsù*) et « mot » (词 *cí*) très vagues (Y. H. Liu et al., 2005).

3.2.3.2 Mot composé

La majorité des mots en chinois contemporain sont constitués de plusieurs morphèmes, nommés « mot composé » (复合词 *fùhécí*).⁴⁸ *Le vocabulaire commun*

⁴⁸ En chinois contemporain, la grande majorité des morphèmes ne contiennent qu'un caractère. Dans la partie suivante de notre thèse, nous utiliserons le terme « caractère » pour désigner l'unité constitutive d'un mot au lieu de « morphème » pour simplifier la compréhension.

pour l'instruction obligatoire (ébauche) (义务教育常用词表 (草案), 2019) contient plus de 17 000 items lexicaux dont presque 90% sont des mots composés. Les mots à deux caractères représentent autour de 75% des mots composés.

3.2.3.2.1 Sens d'un mot composé

Le sens d'un mot composé est parfois facile à élaborer à partir des sens des caractères constitutifs. Ce genre de mots composés est alors sémantiquement transparent (c'est-à-dire que le sens du mot est fortement lié aux sens des caractères constitutifs individuels).

Ex. 书 (*shū*, livre) + 包 (*bāo*, sac) → 书包 (*shūbāo*, cartable)
始 (*shǐ*, début) + 终 (*zhōng*, fin) → 始终 (*shǐzhōng*, toujours)

D'autres sont moins transparents sémantiquement.

Ex. 口 (*kǒu*, bouche) + 吃 (*chī*, manger) → 口吃 (*kǒuchī*, bégayer)
吹 (*chuī*, souffler) + 牛 (*niú*, bovin) → 吹牛 (*chuīniú*, hâbler, se vanter)

Le dernier exemple montre que les mots chinois ne sont pas toujours transparents sémantiquement et que le sens d'un mot composé n'est pas une simple addition des sens des caractères constitutifs.

En outre, du fait que les caractères possèdent souvent plusieurs significations, voire plusieurs prononciations, le décodage d'un caractère chinois s'opère au niveau du caractère et également au niveau du contexte immédiat. Voici un exemple :

Ex. 开 (*kāi*) + 会 (*huì*, réunion) → 开会 (*kāihuì*, avoir / tenir une réunion)
打 (*dǎ*, frapper) + 开 (*kāi*) → 打开 (*dǎkāi*, ouvrir)
开 (*kāi*) + 车 (*chē*, véhicule) → 开车 (*kāichē*, conduire (un véhicule))
开 (*kāi*) + 心 (*xīn*, cœur) → 开心 (*kāixīn*, (être) joyeux, ravi)
开 (*kāi*) + 水 (*shuǐ*, eau) → 开水 (*kāishuǐ*, l'eau bouillie)
开 (*kāi*) + 除 (*chú*, effacer) → 开除 (*kāichú*, licencier)

3.2.3.2.2 Mode de constitution d'un mot composé

La mode de constitution (构词类型 *gòucí lèixíng*) renvoie à la relation entre les caractères constitutifs d'un mot composé. Nous présentons ici quelques modes de constitution principales (cf. B. R. Huang & Liao, 2005, pp. 258-264) :

- Associatif (联合 *liánhé*) : le mot est constitué des caractères dont les sens sont équipotents.
 - Ex. 美好 *měihǎo* : 美 (*měi*, beau) + 好 (*hǎo*, bien) → beau, excellent,
 - 开关 *kāiguān* : 开 (*kāi*, allumer) + 关 (*guān*, éteindre) → interrupteur
- Déterminatif (偏正 *piānzhèng*) : le sens du caractère qui précède détermine ou qualifie le sens du caractère qui suit.
 - Ex. 微笑 (*wēixiào*): 微 (*wēi*, petit, léger) + 笑 (*hǎo*, rire) → sourire
 - 金黄 (*jīnhuáng*): 金 (*jīn*, or) + 黄 (*huáng*, jaune) → doré
- Complémentaire (补充 *bǔchōng*) : le sens du caractère qui suit a un apport explicatif ou complémentaire au sens du caractère qui précède.
 - Ex. 合成 (*héchéng*): 合 (*hé*, réunir) + 成 (*chéng*, devenir) → composer, synthétiser
 - 说服 (*shuōfú*): 说 (*shuō*, narrer) + 服 (*fú*, être convaincu) → persuader
- Sujet-prédicatif (主谓 *zhǔ wèi*)
 - Ex. 地震 (*dìzhèn*): 地 (*dì*, sol, terre) + 震 (*zhèn*, trembler) → séisme
 - 耳鸣 (*ěrmíng*): 耳 (*ěr*, oreille) + 鸣 (*míng*, sonner, siffler) → acouphène
- Verbe-complément d'objet (动宾 *dòng bīn*)
 - Ex. 司机 (*sījī*): 司 (*sī*, contrôler) + 机 (*jī*, mécanisme, appareil) → chauffeur
 - 失业 (*shīyè*): 失 (*shī*, perdre) + 业 (*yè*, emploi) → chômage
- Annexe (附加 *fùjiā*)
 - Ex. 第一 (*dìyī*): 第 (*dì*, préfixe ordinal) + 一 (*yī*, un) → premier
 - 鼻子 (*bízi*): 鼻 (*bí*, nez) + 子 (*zi*, suffixe) → nez

3.2.3.3 Etendue lexicale exigée pour un Chinois natif

En 1988, le Comité national de la langue et de l'écriture de Chine (国家语言文字委员会) classa en deux niveaux les 3500 caractères les plus utilisés, 2500 caractères au niveau A qui couvrent 97,97% des caractères utilisés dans les publications quotidiennes, et ce pourcentage est majoré à 99,48% avec les 1000 caractères du niveau B.

Selon *La norme du programme de cours de chinois dans l'instruction obligatoire* (义务教育语文课程标准, 2011), les écoliers des niveaux CP-CE1 doivent maîtriser autour de 1600 caractères à la lecture dont la moitié à l'écrit ; ceux des niveaux CE2-CM1, 2500 caractères à la lecture dont 1600 à l'écrit ; ceux des niveaux CM2-6^{ème}, 3000 caractères dont 2500 à l'écrit et enfin ceux des niveaux 5^{ème}-3^{ème}, environ 3500 caractères à la lecture.

Le vocabulaire commun pour l'instruction obligatoire (ébauche) (义务教育常用词表 (草案), 2019) réunit plus de 17 000 items lexicaux proposés pour l'enseignement de la langue chinoise auprès des écoliers natifs. Ces items sont classés en plusieurs niveaux qui correspondent aux diverses exigences pédagogiques, dont le niveau 1 (qui correspond aux niveaux CP-CE1 en France) qui contient 2001 items, le niveau 2 (CE2-CM1) 5503, le niveau 3 (CM2-6^{ème}) 5975 et le niveau 4 (5^{ème}-3^{ème}) 3613 items. En outre, jusqu'à la fin de la classe de 3^{ème}, les écoliers devraient accomplir en dehors des cours une lecture extensive de plus de 4 millions de caractères dans l'ensemble (*La norme du programme de cours de chinois dans l'instruction obligatoire* (义务教育语文课程标准), 2011).

3.2.4 Phrase

N'ayant pas de conjugaison ni d'accord, une phrase en chinois contient des caractères en juxtaposition.

Ex. 我去图书馆。

Par rapport à la phrase équivalente en français « Je vais à la bibliothèque. » dans laquelle un espace se présente entre deux mots voisins, dans la phrase en chinois, il n'y a pas d'espace balisant les frontières inter-mots. Afin de comprendre cette phrase, le lecteur doit segmenter cette phrase ainsi :

Ex. 我 (wǒ, je) / 去 (qù, aller) / 图书馆 (túshūguǎn, bibliothèque)。

En d'autres termes, le lecteur doit identifier lui-même les unités lexicales, qui peuvent être constituées d'un seul ou de plusieurs caractères, dans une phrase en chinois. Pour les apprenants qui ont l'habitude que les mots soient séparés dans une phrase avec un

espace, par exemple dans la langue française, l'identification des items lexicaux représente une tâche supplémentaire dans la compréhension écrite en chinois.

La compréhension ambiguë d'un énoncé en chinois peut résulter de différentes segmentations.

Ex. 南京市长江大桥

1. 南京市 (*Nánjīng shì*, la ville Nankin) / 长江 (*Chángjiāng*, la rivière Yangtze) / 大桥 (*dàqiáo*, (grand) pont)

→ Pont de la rivière Yangtze à Nankin

2. 南京 (*Nánjīng*, Nankin) / 市长 (*Shìzhǎng*, maire) / 江大桥 (*Jiāng Dàqiáo* (nom propre))

→ Maire de Nankin JIANG Daqiao

Nous rappelons que le décodage d'un caractère ne s'accomplit pas seulement au niveau du caractère, mais également au niveau du contexte immédiat. Dans cet exemple, le caractère « 长 » ne se prononce pas de la même façon dans les deux cas, *cháng* dans le cas 1, mais *zhǎng* dans le cas 2. « 长 » est donc un caractère polyphonétique et polysémique.

3.3 Distance L1-L2, élément mis en relief dans la recherche sur le chinois L2

Par rapport à une langue alphabétique comme le français, la langue chinoise est caractérisée avant tout par son système d'écriture. Tournadre (2014, pp. 278-279, cité par Song, 2017, p. 52) propose ainsi une gradation dans l'ordre croissant au niveau de la complexité des systèmes d'écriture : alphabet (ex. le français) < alphasyllabaires (ex. le tibétain) < syllabaires (ex. le Yi) < systèmes logographique (ex. le chinois) ; par ailleurs, il ajoute que « ...*parmi les systèmes logographiques, il semble que l'on puisse aussi faire une gradation en fonction de la complexité des graphismes et de la structure des logogrammes. (...) L'écriture chinoise est, sans aucun doute, du point de vue du nombre de signes, la plus difficile à maîtriser de toutes les écritures en usage...* »

La mémorisation des caractères est le premier grand obstacle dans

l'apprentissage de cette langue, même pour les enfants natifs. A l'école primaire en Chine, les élèves des niveaux CP-CE1 consacrent beaucoup de temps à l'apprentissage des caractères (Le cours de chinois représente environ 25% du programme pédagogique au niveau du nombre d'heures. Souvent un cours hebdomadaire exclusivement dédié à l'écriture de caractère est inclus dans le programme scolaire pour les écoliers du niveau CP, voire CE1). L'exercice de copie des caractères en répétition est indispensable, notamment pour les écoliers de bas niveaux afin de maîtriser les caractères au niveau de la reconnaissance et de la production.

Pour un apprenant de chinois L2 dont la L1 est alphabétique, la maîtrise des caractères représente également une tâche ardue et souvent anxiogène (Dong, 2007; Y. G. Zhao, 2009). Comme l'indique Song (2017), dans une langue alphabétique, un mot est écrit de façon linéaire et « *Pour un étranger limité à ses lettres alphabétiques, le caractère chinois est plus un dessin qu'une écriture.* » Par ailleurs, « *...un caractère chinois ne peut s'écrire sans mémorisation des gestes du tracé et ne peut se lire sans une indispensable mémorisation phonétique.* (p. 63) » Un apprenant francophone en chinois L2, surtout du niveau débutant, doit passer beaucoup de temps à mémoriser les formes des caractères chinois, ce qui ne serait pas le cas s'il débutait son apprentissage en anglais ou en espagnol L2.

Feng (2009) précise que, chez un Chinois natif adulte, l'accès au sens depuis une forme graphique est direct et rapide, parce que 1) la forme graphique d'un caractère est enregistrée en structure soudée dans son lexique mental (par exemple, le caractère 明 n'est pas enregistré de façon séparée en 日 et 月) et 2) pour les caractères polysémiques, les différents sens se présentent de façon indépendante l'un de l'autre dans le lexique mental, explicitement caractérisés. Quand un Chinois instruit voit un mot, pour chaque caractère constitutif du mot les sens correspondants sont rapidement activés (autrement dit, la décomposition d'un caractère constitutif afin de récupérer les informations lexicales correspondant n'est pas nécessaire (Yang et al., 2000, cité par Feng, 2009)). Les sens appropriés, en cas de polysémie pour un ou des caractères constitutifs, sont rapidement sélectionnés. Pour les apprenants

débutants du chinois L2, l'identification des caractères par leur forme graphique n'est pas facile parce que les formes graphiques des caractères ne sont pas encore explicites dans leur lexique mental. Pour cela, l'activation forme-sens est très longue et risque plus d'activer à la fois plusieurs caractères de forme similaire (par exemple, 日 et 旦, 九 et 几, 巳 et 己). En outre, l'opacité phonétique des caractères n'aide pas les apprenants dont la L1 est alphabétique, pour lesquels le son est important dans les traitements lexicaux dans leur langue maternelle, dans ce processus d'activation.

Afin de savoir quel aspect, phonétique ou graphique, a une présence plus importante dans le processus du traitement des caractères chez les apprenants du chinois L2, X. H. Lü (2007) a mené une étude auprès des apprenants de trois niveaux différents en chinois, du bas vers le haut, débutant 1, débutant 2 et intermédiaire. A chaque niveau, les participants sont regroupés en fonction de leur L1, un groupe japonais-coréen et un groupe d'autres L1 alphabétiques diverses. Tous les participants doivent lire et comprendre des textes dans lesquels certains caractères sont remplacés. Dans le cas « Son », les caractères cibles sont remplacés par ceux du même son (par exemple 到 (*dào*) → 道 (*dào*)), et dans le cas « Graphie », les caractères cibles sont remplacés par ceux similaires du point de vue de la forme (par exemple 愛 (*ài*) → 受 (*shòu*)). Si les participants remarquent des « erreurs » de caractère lors de leur lecture, ils doivent les souligner. Les résultats révèlent que tous les participants remarquent bien plus d'« erreurs » dans le cas « Son » que dans le cas « Graphie ». Lü conclut qu'en général les apprenants du chinois L2, quelle que soit leur L1 et leur niveau en chinois, s'appuient plus sur la forme graphique pour reconnaître les caractères, parce que les caractères comportent en général peu d'informations phonétiques. Dans les deux cas de reconnaissance de caractère, le groupe japonais-coréen devance largement le groupe d'autres L1, probablement grâce aux similarités entre le chinois et les langues japonaise et coréenne (cf. la note de bas de page 16, p. 35). D'ailleurs, plus les apprenants japonais et coréens avancent en chinois, plus ils s'appuient sur la forme pour reconnaître les caractères (ce qui correspond à l'évolution de la reconnaissance de caractère chez les Chinois natifs, cf. Song et al., 1995). De façon générale, la faible maîtrise des formes graphiques des

caractères est un grand obstacle pour tous les débutants. Surtout pour les apprenants dont la L1 est alphabétique et typologiquement lointaine du chinois, cette entrave demeure même s'ils ont un niveau intermédiaire en chinois.

Pour cette raison, l'arrière-plan linguistique et culturel des apprenants est généralement pris en compte dans la recherche en chinois L2 visant des apprenants de diverses L1, par exemple, en dehors de l'inférence lexicale (cf. la Section 1.3.3.2), dans le traitement lexical des caractères chinois (Ma, 2009; Pan, 2007; Y. H. Xu, 2014a), l'apprentissage de la syntaxe (H. C. Wu & Zhao, 2018) et du ton phonologique (H. Zhang, 2016). Opérationnellement, les apprenants dont la L1 est typologiquement lointaine de la langue chinoise sont classés dans un groupe d'observation différent de celui des apprenants japonais et coréens.

3.4 Récapitulation et objectif de cette étude

Par rapport à l'apprentissage d'une autre langue typologiquement proche telle que l'anglais ou l'espagnol, l'apprentissage de la langue chinoise exige davantage d'efforts pour un apprenant francophone afin de surmonter les obstacles liés aux particularités de cette langue présentées plus haut

Par ailleurs, la langue chinoise est typologiquement distante de la langue française, non seulement linguistiquement mais aussi culturellement. Cela signifie qu'un apprenant francophone peut moins s'appuyer sur ses connaissances pré-acquises en français et risquerait plus de tomber sur des mots chinois non lexicalisés en français. En effet, durant des années d'enseignement du chinois L2, nous avons constaté que les apprenants francophones ont plus de difficultés à comprendre les mots chinois qui n'ont pas d'homologue en français, par exemple 舍得 (*shědé*, abandonner (quelque chose de précieux) de bon gré) et 难道 (*nándào*, serait-il possible que (utilisé dans un ton interrogatif, exclamatif ou ironique pour insister sur son soupçon ou désaccord)). La difficulté de compréhension dans le cas de non lexicalisation se situe non seulement au niveau du mot lui-même, mais parfois également au niveau de l'ensemble de mots comportant un mot non lexicalisé, parce qu'il ne s'agit pas d'appariement « mot-à-mot ». Par exemple, certains chiffres en

chinois, tels que 万 (*wàn*, dix mille), posent problème aux apprenants francophones dans la compréhension : 三万 (三 (*sān*, trois) + 万 (*wàn*, dix mille)) en chinois, signifie « trente mille » en français.

Face à cela, le nombre d'heures en classe est limité, même pour les disciplines de spécialité, telles que LLCE et LEA à l'université. Prenons l'exemple des programmes pédagogiques de Licence LEA à l'Université de Toulouse Jean Jaurès. On peut compter en moyenne 9 heures de cours de chinois hebdomadaires, dont la moitié sont dédiées au lexique et à la compréhension écrite (sauf en Licence 2 où le taux est de moins de 30%). Selon le programme, les étudiants doivent maîtriser plus de 900 unités lexicales à la fin du parcours Licence. Cela représente une tâche laborieuse à accomplir pour la majorité des étudiants. En réalité, plus les apprenants avancent en chinois, plus longue est la liste de lexique dans leur manuel et moins les enseignants ont de temps pour enseigner chaque mot dans la liste. Il est courant que les étudiants ne retiennent pas tous les mots enseignés en classe, surtout les mots difficiles à comprendre et ceux que les apprenants ont vu seulement une ou deux fois en classe. Comme dans l'apprentissage de toutes les langues, une exposition fréquente est un facteur fondamental pour la rétention des informations lexicales en chinois L2 (X. Jiang, 2006). Or, le manque de stimuli est largement amplifié dans le contexte d'apprentissage hétéroglotte. Même si les étudiants parviennent à maîtriser tout ce qu'on leur enseigne en classe, leur étendue lexicale au bout de trois ans d'études est tout de même modeste.

Il est donc important d'aider les apprenants en chinois L2 dans leur développement du vocabulaire par d'autres stratégies, en dehors de l'apprentissage traditionnel en classe. C'est pour cela que nous nous intéressons à l'inférence lexicale et à l'apprentissage incident du vocabulaire inféré.

La connaissance d'un mot est multidimensionnelle (la forme, le sens et l'emploi) et la maîtrise d'un mot peut être catégorisée selon deux axes principaux, réceptif et productif (cf. la Section 2.2). Comme l'indique Webb (2005), « *la majorité du vocabulaire est apprise de manière réceptive par la lecture ou l'écoute* (p. 33)⁴⁹. »

⁴⁹ « The majority of vocabulary is learned receptively through reading or listening. »

En milieu hétéroglotte, l'apprentissage du vocabulaire L2 a lieu en général plus par le biais de la lecture que par l'écoute. Par ailleurs, en comparaison de l'écoute (ou la compréhension orale), un apprenant a plus de possibilités de faire attention à un mot nouveau à la lecture car, dans une communication orale, les messages sont délivrés vite avec peu de possibilité de revenir en arrière, ce qui incite l'auditeur à prioriser la compréhension globale du message sur les détails (J. Guo, 2015; Van Zeeland, 2014). Quand un mot nouveau est identifié dans un contexte, pour qu'il soit bien enregistré dans le lexique mental de l'apprenant, la création d'un lemme dans la mémoire de l'apprenant à partir de la forme (graphique ou phonétique) du mot est primordiale (voir supra p. 52 ; cf. aussi Théophanous, 2019, Chapitre IV, sur les connaissances réceptives du vocabulaire). Cela est particulièrement vrai dans le cas de la langue chinoise :

- 1) L'identification de la forme phonétique d'un caractère (ou d'un mot) est généralement difficile pour les apprenants francophones (J. Guo, 2015; H. Zhang, 2016). Entre autres, la prononciation d'un caractère chinois est accompagnée d'un ton (cf. la Section 3.2.1) et la maîtrise des tons est une autre grande entrave pour les apprenants francophones. Par exemple, dans son étude auprès d'apprenants francophones sur l'inférence lexicale dans la compréhension orale, Guo (2015) remarque qu'une identification de mot inefficace ou erronée dans un contexte oral résulte principalement d'une capacité de décodage auditif faible en chinois des apprenants (par exemple, un participant entendait « *jiù* (就, juste) » mais il croyait entendre « *jiǔ* (酒, alcool) » et cela n'est un cas isolé).
- 2) Les caractères chinois sont opaques sur le plan graphie-phonie (cf. la Section 3.2.2.3). L'identification correcte de la forme phonétique d'un caractère ne fournit pas à l'apprenant l'accès à la forme graphique du caractère, surtout quand il s'agit d'un caractère jamais rencontré.
- 3) La relation graphie-phonie est loin d'être un à un pour les caractères chinois (cf. la Section 3.2.2.3). Cela signifie que même si la forme phonétique d'un mot est correctement identifiée, celle-ci peut correspondre à plusieurs formes graphiques ayant des significations différentes, comme le montre l'exemple

ci-dessous. Autrement dit, en chinois, le lien forme phonétique–sens est moins fiable que le lien forme graphique-sens.

- Ex. *jiǔ* 1. 九 (neuf)
 2. 酒 (alcool)
 3. 久 (longtemps)
- dānjià* 1. 担架 (civière)
 2. 单价 (prix à unité)

En revanche, à la lecture, l'apprenant est exposé aux formes graphiques explicites des caractères, qui peuvent délivrer eux-mêmes des informations sémantiques (cf. la Section 3.2.2.1), autant de fois et aussi longtemps que l'apprenant le souhaiterait dans le cadre naturel de la lecture.

La recherche sur l'inférence lexicale et l'apprentissage incident du vocabulaire en chinois L2 dans la compréhension écrite a connu une croissance rapide depuis les vingt dernières années à mesure des échanges internationaux et de l'augmentation fulgurante des apprenants L2 de diverses L1. Les études réalisées ont essayé d'identifier les éléments qui peuvent intervenir dans le processus et le résultat de l'inférence lexicale ou/et d'apprentissage lexical, tels que les sources de connaissance impliquées (Gan, 2012; S. H. Liu, 2001; J. L. Zhang, 2010), le niveau en chinois L2 de l'apprenant (Dong, 2007; Gan, 2008; Y. Huang, 2008; S. H. Liu, 2001; Y. Zhu & Cui, 2005), le mode de constitution du mot inconnu (Ke & Koda, 2017), la classe grammaticale du mot (Gan, 2010), la transparence sémantique du mot (Gan, 2008; Y. H. Xu, 2014b; H. M. Zhang et al., 2019; X. Y. Zhu & Zhou, 2007), la richesse d'informations contextuelles fournies (Y. X. Fang & Jiang, 2012; Y. Zhu & Cui, 2005), l'emplacement de l'indice contextuel (Y. Fang, 2005 ; Gan, 2011, 2014) et la tâche à réaliser dans l'input (Hong & Wang, 2016; Y. Zhu, 2004). Il est à noter que la grande majorité des études réalisées ont été menées auprès des apprenants de chinois en Chine au moment de l'expérience empirique.

Après tout, malgré la prise en compte de l'effet de la distance L1-L2 dans l'inférence et l'apprentissage lexical en chinois L2, pour autant que nous le sachions, aucune étude n'a exploré le sujet sur la lexicalisation dans la L1 de l'apprenant jusqu'à présent. C'est pour cela que nous avons mené cette étude sur l'inférence

lexicale et l'apprentissage subséquent du vocabulaire chinois L2 dans la compréhension écrite auprès d'apprenants francophones. Notre centre d'intérêt est l'effet de la lexicalisation d'un mot chinois L2 inconnu dans l'apprentissage incident du mot à divers stades, c'est-à-dire celui de la lecture à des fins de compréhension (donc pré-inférence), l'inférence du sens du mot et la rétention postérieure du mot (donc post-inférence).

Nous présenterons dans le Chapitre IV le protocole expérimental de notre étude et les méthodes utilisées.

CHAPITRE IV – PROTOCOLE EXPERIMENTAL

Introduction

Pour rappel, nous nous sommes intéressée à l'inférence lexicale et l'apprentissage incident du vocabulaire en chinois L2, en nous penchant en particulier sur la lexicalisation dans la L1 de l'apprenant, en l'occurrence la langue française.

Nous avons émis les questions de recherche suivantes :

QR 1 : Lors de la compréhension en lecture en chinois L2, de quelle manière un apprenant francophone s'appuie-t-il sur les différentes sources de connaissance pour inférer les mots chinois inconnus présents dans la lecture ? La nature des sources de connaissance utilisées pour inférer le sens des mots chinois inconnus diffère-t-elle selon que les mots chinois sont lexicalisés ou non en français ?

QR 2 : Avec quel degré de réussite un apprenant francophone parvient-il à inférer le sens des mots inconnus, lexicalisés ou non, présents dans les textes chinois ?

QR 3 : Pendant une lecture en chinois pour la compréhension, l'attention d'un apprenant francophone sur un mot chinois inconnu présent dans la lecture diffère-t-elle selon que le mot est lexicalisé ou non en français ?

QR 4 : A l'issue de l'inférence du sens d'un mot chinois inconnu présent dans une lecture, une combinaison de deux tâches concernant le mot, par rapport à une seule tâche, aboutit-elle à de meilleurs gains lexicaux pour ce mot à court et à long terme ? La nature de lexicalisation a-t-elle un effet sur les gains lexicaux sur le mot à court et à long terme ?

Les lectures que nous avons effectuées ne nous permettent pas de construire une hypothèse pour la QR1 ni pour la QR2. Quant aux deux autres questions de recherche, nous avons émis les hypothèses suivantes à la lumière des conclusions des études réalisées :

Hypothèse 1 : Les apprenants en chinois L2 réussissent mieux l'inférence des mots chinois lexicalisés que celle des mots non lexicalisés.

Hypothèse 2 : L'inférence lexicale suivie d'une combinaison de deux tâches de post-inférence, par rapport à une seule tâche, aboutit à de meilleurs gains lexicaux sur les mots chinois inconnus, à court et à long terme. Un apprenant

francophone a plus de gains lexicaux sur les mots chinois L2 lexicalisés en français que sur les mots L2 non lexicalisés, à court et à long terme.

Dans ce chapitre, nous présenterons notre protocole de recherche afin de répondre à ces questions de recherche et de vérifier les hypothèses émises.

4.1 Public visé

Nous avons évoqué dans le Chapitre I qu'afin qu'un apprenant L2 puisse mettre en place les compétences de lecture acquises en L1, il doit atteindre préalablement un seuil lexical minimal en L2 (cf. la Section 1.1.2). Entre autres, dans l'inférence lexicale, un apprenant ayant un faible niveau en L2 risque de ne pas comprendre le co-texte d'un mot inconnu et de ne pas extraire suffisamment d'informations contextuelles qui sont une source d'indice très importante pour déduire son sens. Notamment en chinois, compte tenu de la difficulté à maîtriser les caractères, un apprenant ayant un faible niveau en chinois risque alors de ne pas pouvoir accomplir une tâche de compréhension écrite. Par exemple, dans l'étude de M. J. Wu (2005), les participants de chinois L2 du niveau débutant (estimé A1) n'ont effectué que la tâche d'inférence du sens d'un mot inconnu intégré dans une phrase courte parce qu'ils n'étaient pas capables de comprendre un passage plus long.

Pour notre expérience, nous avons visé les apprenants ayant une connaissance de 600 et 1200 mots chinois, ce qui correspond au niveau 3 du HSK (汉语水平考试 *Hànyǔ Shuǐpíng Kǎoshì*, un test d'évaluation du niveau de chinois, normalisé par le gouvernement chinois depuis 1992. Il comprend aujourd'hui six niveaux d'évaluation concernant la compréhension orale, la compréhension écrite et la production écrite, ainsi que trois niveaux d'évaluation de la production orale.)⁵⁰ Ce niveau correspond au niveau B1 dans le *Cadre européen commun de référence pour les langues* (CECR, 2001)⁵¹. Il est à noter que Bellassen (2011) remet en question la grille de niveaux en

⁵⁰ La grille de niveaux, avec les étendues lexicales exigées et les correspondances avec le CECR, est affichée sur le site officiel d'inscription aux tests d'évaluation de chinois : <http://www.chinesetest.cn/gosign.do?id=1&lid=0#> (consulté le 30 août 2021).

⁵¹ Le CECR a été conçu pour fournir une base transparente, cohérente et complète pour l'élaboration de programmes d'études et de lignes directrices en matière de langues, la conception de matériel d'enseignement

HSK confrontée au CECR et insiste sur la nécessité d'une nouvelle table de correspondances entre les niveaux de ces deux systèmes. Dans notre thèse, nous considérons tout de même le niveau de nos participants en chinois comme niveau B1, en accord avec la table de correspondance affichée sur le site officiel du HSK.

Nous avons sollicité 33 participants francophones dont le profil est résumé dans le Tableau 4.1 plus bas. 25 femmes et 8 hommes ont participé à l'étude. Au moment de l'expérience, les participants étaient en moyenne âgés de 21,6 ans. Ils avaient fait 4 ans d'études en moyenne de la langue chinoise. 32 participants étaient étudiants au deuxième semestre en troisième année de Licence à l'Université de Toulouse II Jean Jaurès (désormais UT2J), dont 27 étaient en Langue Etrangère Appliquée (option anglais-chinois) et 5 autres en Langue, Littérature, Civilisation Etrangère en chinois.

Détendeur du certificat HSK niveau 2, le participant P23 était auditeur libre en 3^{ème} année de LLCE. P23 avait passé quatre ans en Chine, dont un an pour ses études en Master dans une université chinoise (la langue d'enseignement était l'anglais) puis trois ans pour le travail, au cours desquels il a suivi des cours particuliers hebdomadaires.

Le participant P24 a étudié le chinois en Langue Vivante 3 au lycée pendant 3 ans et a obtenu 20 sur 20 au Baccalauréat. Il suivait, au moment de l'expérience, des cours de chinois collectif depuis trois ans une fois par semaine dans une association à Toulouse.

Participant	Âge	Sexe	Discipline	Nb. d'années d'étude en chinois
P01	21	F	LEA	6
P02	20	F	LEA	6
P03	20	F	LEA	3
P04	21	M	LEA	3
P05	21	F	LEA	4
P06	21	F	LEA	6
P07	21	F	LEA	5
P08	20	F	LEA	3
P09	20	F	LEA	3
P10	22	F	LEA	8
P11	20	F	LEA	3
P12	22	F	LEA	3
P13	21	F	LEA	3
P14	21	F	LEA	3
P15	25	M	LLCE	6
P16	20	M	LEA	6
P17	22	M	LEA	2
P18	20	F	LEA	10
P19	21	F	LEA	7
P20	21	F	LEA	3
P21	25	F	LEA	3
P22	21	F	LLCE	5
P23	27	M	LLCE (audition libre)	3
P24	21	F	Ecole d'architecture	6
P25	23	F	LEA	3
P26	22	M	LEA	3
P27	21	M	LEA	3
P28	20	F	LEA	3
P29	21	F	LEA	6
P30	20	F	LLCE	3
P31	30	M	LLCE	3
P32	23	F	LEA	3
P33	20	F	LEA	6

Tableau 4.1 – Informations concernant les 33 participants

4.2 Matériels conçus pour l'expérience

4.2.1 Mots cibles

Comme le centre d'intérêt de notre étude est la lexicalisation dans la L1 de l'apprenant, il nous fallait tester deux types de mots : des mots chinois lexicalisés en français et des mots non lexicalisés. Afin de choisir les mots, nous avons utilisé le *Guide hiérarchique des caractères et des mots chinois* (« 汉语水平词汇与汉字等级大纲 », 1992) qui comprend 2905 caractères et 8822 mots les plus fréquemment utilisés. Les mots sont classés selon quatre niveaux de fréquence d'utilisation dans l'ordre décroissant : classes A (1033 mots, les plus fréquemment utilisés), classe B (2018 mots), classe C (2202 mots) et classe D (3569 mots). Afin de minimiser le risque que les participants connaissent déjà les mots cibles ou qu'ils rencontrent ces mots avant la fin de l'expérience, nous avons cherché des mots exclusivement dans les classes C et D. En parallèle, nous avons également cherché des mots non lexicalisés en français dans des articles de journaux chinois concernant les coutumes traditionnelles chinoises.

Après avoir opéré une première sélection, nous avons vérifié dans le *Dictionnaire chinois-français* (« 汉法词典 », 2001) le sens des mots chinois pré-sélectionnés et gardé les mots chinois ayant relativement peu de variation de significations (Nous reviendrons sur ce point plus tard). Par la suite, nous avons consulté l'avis des enseignants de chinois à l'UT2J. Nous avons ensuite éliminé les mots qui étaient déjà enseignés, pratiqués en classe ou prévus d'être enseignés dans les programmes pédagogiques immédiats. Pendant la rédaction des textes expérimentaux, selon les réponses du premier test de vocabulaire mené auprès des participants, nous avons remplacé quelques mots cibles inappropriés. Nous avons enfin sélectionné 10 mots lexicalisés (L) et 10 mots non lexicalisés (NL), présentés dans le Tableau 4.2 plus bas. Les significations données dans la dernière colonne sont fournies par le *Dictionnaire chinois-français* (汉法词典, 2001).

Nous avons consulté l'avis de plusieurs enseignants de chinois, natifs français, sur les mots chinois non lexicalisés sélectionnés et ils nous ont confirmé que, pour

chacun de ces 10 mots cibles non lexicalisés, il n'y a pas d'équivalent attesté en français.

	Mot cible		Classe gramm.	Nb. de caractères	Lexicali-sation	Mot cible
1	彩虹	<i>cǎihóng</i>	nom	2	L	arc-en-ciel
2	蝉	<i>chán</i>	nom	1	L	cigale
3	乏味	<i>fáwèi</i>	adjectif	2	L	fade, insipide
4	风筝	<i>fēngzhēng</i>	nom	2	L	cerf-volant
5	节俭	<i>jiéjiǎn</i>	adjectif	2	L	économe, frugal
6	盼望	<i>pànwàng</i>	verbe	2	L	espérer, désirer
7	雀跃	<i>quèyuè</i>	verbe	2	L	sauter de joie
8	五颜六色	<i>wǔyánliùsè</i>	adjectif	4	L	de diverses couleurs, multicolore
9	闲聊	<i>xiánliáo</i>	verbe	2	L	bavarder
10	在乎	<i>zàihu</i>	verbe	2	L	prendre qch. à cœur, faire attention à
11	安适	<i>ānshì</i>	adjectif	2	NL	tranquille et confortable
12	白茫茫	<i>báimángmáng</i>	adjectif	3	NL	d'une blancheur éclatante (immaculée) ; une immensité argentée
13	窗花	<i>chuānghuā</i>	nom	2	NL	papier découpé pour la décoration d'une fenêtre
14	还礼	<i>huánlǐ</i>	verbe	2	NL	faire un présent en retour, offrir un cadeau à son tour
15	腊月	<i>làyuè</i>	nom	2	NL	le douzième mois du calendrier lunaire
16	明丽	<i>mínglì</i>	adjectif	2	NL	clair et beau
17	飘零	<i>piāolíng</i>	verbe	2	NL	tomber en voltigeant
18	送灶	<i>sòngzào</i>	verbe	2	NL	saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
19	踏青	<i>tàqīng</i>	verbe	2	NL	faire une excursion (se promener) à la campagne au printemps
20	汤圆	<i>tāngyuán</i>	nom	2	NL	boulette de riz glutineux farcie de viande (fourrée à la confiture)

Tableau 4.2 – Mots cibles

Parmi ces 20 mots, 17 mots sont constitués de deux caractères, un mot d'un seul caractère (蝉 *chán*, cigale), un mot de trois caractères (白茫茫 *báimángmáng*, d'une blancheur éclatante ; une immensité argentée) et un autre de quatre caractères (五颜六色 *wǔyánliùsè*, de diverses couleurs, multicolore). Ces 20 mots cibles sont de différentes classes grammaticales, soit 6 adjectifs, 6 noms et 8 verbes. Il est à noter que le mot 8, 五颜六色 (*wǔyánliùsè*, de diverses couleurs, multicolore), est un proverbe ou expression figée chinoise (成语 *chéngyǔ*). Nous l'avons classé dans la classe adjectivale selon son rôle grammatical dans le texte que nous allons présenter dans la Section 4.2.2.

4.2.2 Textes expérimentaux

Nous avons décidé de rédiger deux textes pour intégrer les 20 mots cibles pour deux raisons principales :

- Afin que les apprenants puissent comprendre un texte, le nombre des mots inconnus ne doivent pas dépasser un certain seuil. Il est généralement admis qu'une connaissance de 90%-95% des mots dans un texte est nécessaire pour comprendre ce dernier (Hirsch, 2003) (cf. supra p. 16). Cela signifie, par exemple pour le seuil de 95%, qu'un texte « théoriquement compréhensible », comportant 20 mots inconnus, doit être d'une longueur environ 400 mots. Cela signifie que ce texte peut comporter autour de 800 caractères si on compte deux caractères par mot. Compte tenu du niveau de chinois de nos participants, ce volume de lecture nous semble chargé.
- Pour répondre à la question de recherche sur l'attention que les apprenants portent aux mots inconnus pendant la lecture, nous avons décidé d'utiliser la technique d'oculométrie (présentée plus tard dans la Section 4.3.2). Afin d'éviter que le déplacement du texte sur l'écran perturbe la lecture, nous avons opté pour un affichage du texte expérimental dans son intégralité sur un même écran.

A partir de certains mots caractéristiques, tels que 腊月 (*làyuè*, le douzième mois du calendrier lunaire), 汤圆 (*tāngyuán*, boulette de riz glutineux farcie de viande

(fourrée à la confiture)), 白茫茫 (*báimángmáng*, d'une blancheur éclatante) et 飘零 (*piāolíng*, tomber en voltigeant), nous avons choisi deux thèmes, l'un sur les coutumes traditionnelles pendant la fête du Nouvel an chinois et l'autre sur la description de paysage des quatre saisons. Nous avons ensuite rédigé nous-même un texte pour chaque thème qui comporte 10 mots cibles, 5 mots lexicalisés et 5 autres non lexicalisés.

Les deux textes ont été rédigés avec grand soin parce que la mise en page du texte et les emplacements des mots cibles dans un texte ont un effet direct sur le recueil de données d'oculométrie (Nous reviendrons sur ce point dans la Section 4.3.2). Les deux textes expérimentaux sont présentés ci-dessous, les mots cibles dans chaque texte étant surlignés et numérotés avec leur traduction en français⁵².

Texte A

我小时候，过年真的很热闹。那时候的经济不太好，大家平时都非常节俭¹，只有过年时才会花很多钱。新年前一个多月，大人们就开始忙着为过年买各种东西。很多人会用红纸做一些窗花²，让家里更漂亮。新年前几天，每家都会做一些好吃的来送灶³。孩子们从腊月⁴第一天开始就每天想着新年快点儿到，因为过年时会有新衣服穿，还可以吃汤圆⁵和很多别的好吃的。当然，过年时最让孩子们雀跃⁶的事儿还是拿红包。很多孩子的红包最后不是自己的：父母没什么钱，别人给了孩子红包，孩子交给父母，父母再用这些钱去给其他人还礼⁷。不过，孩子们拿到红包时都高兴极了，也不在乎⁸钱最后是不是自己的。

现在日子好了，商店里的东西也多了，想什么时候买都可以，连发红包也能用手机。但大家都说过年越来越乏味⁹了。以前盼望¹⁰着过年穿新衣服、吃好吃的、拿红包，那是一种快乐，现在都没有了。

Traduction du Texte A

Quand j'étais petite, la Fête du Nouvel an était un grand événement. L'économie ne marchait pas bien à l'époque. On vivait de façon frugale¹ et on dépensait seulement à l'occasion de la Fête. Un mois avant le Nouvel an, les adultes

⁵² Afin que les traductions globales des textes soient plus fluides en français, les traductions des mots cibles dans les textes ne correspondent pas tous aux significations marquées dans le Tableau 4.2. Par exemple, MC02 窗花 (*chuānghuā*, papier découpé pour la décoration d'une fenêtre) du texte A est traduit par « découpe en jolie forme » dans le texte.

se mettaient à faire des achats pour la Fête. Beaucoup de gens faisaient des découpes en jolies formes² avec des papiers rouges pour égayer la maison. Quelques jours avant le Nouvel an, chaque famille préparait de bonnes choses à manger pour saluer le départ du Génie du foyer³. Les enfants, depuis le premier jour du douzième mois du calendrier lunaire⁴, espéraient tous les jours que le Nouvel an arriverait bientôt, parce qu'ils auraient de nouveaux habits, ainsi que des Tangyuan⁵ et d'autres délices à manger. Bien sûr, la chose qui leur apportait le plus de joie / les faisait sauter de joie⁶ était de recevoir des hongbao⁵³. Beaucoup d'enfants ne conservaient pas leurs hongbao : leurs parents n'étant pas riches, lorsqu'on donnait un hongbao à l'enfant, celui-ci le remettait à ses parents qui se servaient de cet argent pour faire un présent en retour⁷. Cependant, les enfants étaient extrêmement contents lorsqu'ils recevaient des hongbao, sans faire attention à⁸ qui serait finalement destiné l'argent.

Aujourd'hui la vie est plus aisée et les magasins sont plus fournis. On peut faire des courses à tout moment, et même envoyer des hongbao par téléphone mobile. Mais tout le monde dit que la Fête du Nouvel an devient de plus en plus fade⁹. C'était une joie d'attendre (avec impatience)¹⁰ la Fête pour avoir de nouveaux habits, se régaler et recevoir des hongbao. Tout cela s'est perdu maintenant.

N°	Mot cible				
MC01	节俭 <i>jiéjiǎn</i>	a.	L	économe, frugal	
MC02	窗花 <i>chuānghuā</i>	n.	NL	papier découpé pour la décoration d'une fenêtre	
MC03	送灶 <i>sòngzào</i>	v.	NL	saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune	
MC04	腊月 <i>làyuè</i>	n.	NL	le douzième mois du calendrier lunaire	
MC05	汤圆 <i>tāngyuán</i>	n.	NL	boulette de riz glutineux farcie de viande (fourrée à la confiture)	
MC06	雀跃 <i>quèyuè</i>	v.	L	sauter de joie	
MC07	还礼 <i>huánlǐ</i>	v.	NL	faire un présent en retour, offrir un cadeau à son tour	
MC08	在乎 <i>zàihu</i>	v.	L	prendre qch. à cœur ; faire attention à	
MC09	乏味 <i>fáwèi</i>	a.	L	fade, insipide	
MC10	盼望 <i>pànwàng</i>	v.	L	espérer ; désirer	

Tableau 4.3 – Mots cibles présents dans le Texte A

⁵³ « Hongbao », littéralement traduit « enveloppe rouge », est un don d'étrennes traditionnel dans une enveloppe rouge lors du Nouvel an chinois, généralement des adultes aux enfants.

Texte B

我的家乡是一个人口不多的小城市，交通很方便，一年四季都非常美，是一个很**舒适**¹¹的地方。家乡的春天很温暖，城里的树和草很早就换上了绿色的新衣，路边都开着**五颜六色**¹²的小花儿。春天也是一个多雨的季节。每场春雨过后，小城都会变得更**明丽**¹³。有时下完雨后，天上还会有漂亮的**彩虹**¹⁴。每到天气好的时候，人们会到城市以外的地方去**踏青**¹⁵。要是风比较大的话，还可以去小城旁边的山上放**风筝**¹⁶。夏天很热，树上的**蝉**¹⁷从早叫到晚。到了晚上，吃完了晚饭的人们都走出家门，大人们坐在一起**闲聊**¹⁸，孩子们就在一边玩儿着，那是一天中最热闹的时候。一到秋天，天气就开始转凉，鸟儿们都飞去温暖的南方过冬。草和树叶慢慢地变黄，秋风刮过，树叶**飘零**¹⁹。小城安静下来，等着冬天的到来。家乡的冬天常常下雪。大雪过后，城市**白茫茫**²⁰的，城里的草地上、树上、房子上和城外的山上都是雪，真是漂亮极了。

Traduction du Texte B

Je viens d'une petite ville pas très peuplée, mais bien desservie et qui offre de beaux paysages toute l'année. C'est un endroit **tranquille et confortable**¹¹. Il fait doux au printemps. Dans la ville, l'herbe et les arbres sont très tôt habillés de vert. Les rues sont toutes décorées de petites fleurs **de toutes couleurs**¹². Le printemps est également une saison où il pleut beaucoup et chaque averse rend la petite ville encore plus **claire et belle**¹³. Parfois, à l'arrêt d'une pluie, on voit un bel **arc-en-ciel**¹⁴. Quand il fait beau, on sort tous de la ville pour **se promener dans la campagne**¹⁵. Quand il y a du vent, on peut aller faire du **cerf-volant**¹⁶ dans les collines proches de la ville. En été il fait chaud. Les **cigales**¹⁷ chantent jour et nuit dans les arbres. A la tombée du soir, après le dîner, tout le monde sort de chez soi. Les adultes **bavardent**¹⁸ ensemble tandis que les enfants jouent à côté. C'est le meilleur moment de la journée. A l'arrivée de l'automne, le temps devient de plus en plus frais. Les oiseaux partent au sud pour passer l'hiver. L'herbe et les feuilles jaunissent petit à petit. Après un coup de vent saisonnier, les feuilles **tombent en voltigeant**¹⁹. La ville s'éteint, en attendant l'hiver. Là d'où je viens, il neige souvent en hiver. Une grosse chute de neige crée **une immensité argentée**²⁰ dans la ville. Les pelouses, les arbres, les toits d'immeuble et les montagnes autour de la ville sont toutes couvertes de neige. C'est vraiment splendide.

N°	Mot cible				
MC11	安逸	<i>ānshì</i>	a.	NL	tranquille et confortable
MC12	五颜六色	<i>wǔyánliùsè</i>	a.	L	de diverses couleurs, multicolore
MC13	明丽	<i>mínglì</i>	a.	NL	clair et beau
MC14	彩虹	<i>cǎihóng</i>	n.	L	arc-en-ciel
MC15	踏青	<i>tàqīng</i>	v.	NL	faire une excursion (se promener) à la campagne au printemps
MC16	风筝	<i>fēngzhēng</i>	n.	L	cerf-volant
MC17	蝉	<i>chán</i>	n.	L	cigale
MC18	闲聊	<i>xiánliáo</i>	v.	L	bavarder
MC19	飘零	<i>piāolíng</i>	v.	NL	tomber en voltigeant
MC20	白茫茫	<i>báimángmáng</i>	a.	NL	d'une blancheur éclatante (immaculée) ; une immensité argentée

Tableau 4.4 – Mots cibles présents dans le Texte B

Le texte A décrit les événements traditionnels lors de la fête du printemps en Chine et le texte B donne une description d'un village pendant les quatre saisons. Les deux textes sont du même niveau de difficulté de compréhension. Cependant, les mots non lexicalisés contenus dans le texte A exigent davantage de connaissances culturelles pour inférer leur sens.

Chaque mot cible n'apparaît qu'une seule fois dans les textes, intégré dans un contexte qui fournit des informations sur l'idée générale que le mot représente. Par exemple, la phrase comportant MC03 « 送灶 (*sòngzào*, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune) » signifie « Quelques jours avant le Nouvel an, chaque famille préparait de bonnes choses à manger pour MC03. » On peut en déduire que MC03 renvoie à quelque chose de traditionnel que les Chinois font en famille avec de bons plats juste avant la grande fête ; la phrase comportant MC14 « 彩虹 (*cǎihóng*, arc-en-ciel) » signifie « Parfois, à l'arrêt d'une pluie, on voit un(e) joli(e) MC14. » qui peut aider un apprenant à comprendre que MC14 est quelque chose parfois apparaissant au ciel après la pluie.

Les mots cibles représentent environ 5% de la totalité des mots contenus dans chaque texte. Les mots non cibles présents dans les textes ont été enseignés en classe selon les programmes pédagogiques à l'UT2J, à l'exception de « 日子 (la vie

(quotidienne)) » dans le texte A. Les participants connaissent les deux caractères constitutifs sans connaître le sens du mot dans le texte. Toutefois, la méconnaissance du mot « 日子 (la vie (quotidienne)) » n'empêche pas la compréhension globale du texte. Nous avons consulté l'avis de Chinois natifs qui nous ont confirmé que les deux textes sont explicites sémantiquement et corrects au niveau de la grammaire. Nous avons ensuite demandé l'avis des enseignants de la section de langue chinoise à l'UT2J en charge des cours de la compréhension écrite des Licence 3. Ils nous ont confirmé qu'au moment de l'expérience, les étudiants en Licence 3, des filières LLCE ou LEA, parviendraient à comprendre le sens global des deux textes. Nous avons ensuite sollicité quelques étudiants du même niveau de chinois que nos participants pour vérifier leur compréhension pour ces deux textes. Ces étudiants au prétest nous ont confirmé la compréhensibilité de ces deux textes.

4.2.3 Test « Vocabulary Knowledge Scale »

Afin d'évaluer l'évolution des connaissances lexicales des participants, il était nécessaire d'avoir un outil pour mesurer la connaissance des mots cibles avant et après l'inférence lexicale.

Nous avons expliqué dans la Section 2.2 que la connaissance d'un mot est multidimensionnelle et qu'il existe traditionnellement deux approches pour évaluer la compétence lexicale : « étendue-profondeur » et « réception-production ». L'étendue lexicale a trait à la quantité des mots qu'un apprenant connaît, alors que la profondeur renvoie à ce qu'un apprenant connaît sur un mot (Milton, 2009, p. 13), sur sa forme, son sens et son emploi (Nation, 2001). Dans chaque aspect lexical d'un mot, un apprenant peut avoir une connaissance qui s'étend de la reconnaissance passive (« réception ») vers la capacité de production (« production ») (cf. le Tableau 2.2, supra p. 63).

Pour notre étude, nous voulions observer l'évolution des connaissances lexicales pour les mots cibles chez nos participants avant et après les tâches. Nous avons besoin d'un outil pour mesurer la profondeur des connaissances. Nous avons choisi le test Vocabulary Knowledge Scale.

Le test Vocabulary Knowledge Scale (désormais VKS, Paribakht & Wesche, 1993, 1997) a été développé par Paribakht et Wesche pour évaluer les connaissances lexicales des participants dans leurs études. Ce test est aujourd’hui largement utilisé dans les études concernant l’acquisition du vocabulaire en L2.

Lors du test VKS, un apprenant doit estimer sa connaissance pour un mot donné selon une échelle de connaissances qui s’étend d’une connaissance zéro jusqu’à l’emploi approprié du mot dans un contexte. Le modèle VKS le plus couramment utilisé aujourd’hui est celui développé par Paribakht et Wesche (Paribakht & Wesche, 1993, 1996, 1997; Wesche & Paribakht, 1996) qui contient cinq niveaux (Tableau 4.5).

Catégorie d’auto-évaluation	Scores possibles	Significations des scores
I	→ 1	Le mot n’est pas du tout familier.
II	→ 2	Le mot m’est familier, mais son sens est inconnu.
III	→ 3	Un synonyme ou une traduction correcte du mot est donné.
IV	→ 4	Le mot est convenablement utilisé dans une phrase d’un point de vue sémantique.
V	→ 5	Le mot est convenablement utilisé dans une phrase d’un point de vue sémantique et ses caractéristiques grammaticales sont respectées.

Tableau 4.5 – Signification des scores du VKS (VKS scoring categories: Meaning of scores, d’après Paribakht & Wesche, 1997, p. 181)

Les questions correspondant aux cinq niveaux sont présentées dans le Tableau 4.6.

1. Je ne me souviens pas avoir déjà vu ce mot.
2. J’ai déjà vu ce mot, mais je ne sais pas ce que cela signifie.
3. J’ai déjà vu ce mot et je pense qu’il signifie _____. (Synonyme ou traduction)
4. Je connais ce mot. Il veut dire _____. (Synonyme ou traduction)
5. Je peux utiliser ce mot dans une phrase : _____. (Écrire une phrase)

Tableau 4.6 – Echelle d’éllicitation du VKS : la catégorie d’auto-évaluation (VKS elicitation scale: Self-report categories, d’après Paribakht & Wesche, 1997, p. 180)

Malgré le fait que le VKS est largement utilisé dans la recherche en L2, on interroge la capacité d'évaluation du VKS principalement sur la connaissance sémantique d'un mot par la production d'une phrase. Concrètement, si un participant produit une phrase avec le mot concerné à une place sémantiquement neutre, par exemple « J'aime *sauter*. » avec le mot « sauter », sa connaissance du sens du mot « sauter » ne peut guère être vérifiée grâce à cette phrase. Pour cela, Milton (2009, pp. 159-161) considère que le VKS dévoile plutôt l'étendue lexicale que la profondeur. Wesche & Paribakht (1996) elles-mêmes ont indiqué que le VKS n'est pas un test capable d'évaluer les connaissances lexicales de façon subtile et sophistiquée (par exemple, montrer le réseau du lexique mental auquel un mot appartient).

Nonobstant les faiblesses du VKS, nous avons choisi ce test pour évaluer les connaissances lexicales de nos participants pour les raisons suivantes :

- 1) Nous avons besoin d'évaluer l'évolution des connaissances de divers aspects des mots cibles chez les participants. En réalité, l'« *acquisition de la signification de n'importe quelle unité lexicale grâce à la lecture est quasiment toujours partielle dans la phase initiale* (Bruton, 2007, p. 164)⁵⁴ ». Il est probable que l'acquisition d'un mot inconnu à l'écrit après trois ou quatre expositions se cantonne à la dimension « forme » (cf. la Section 2.4.3), voire zéro. Le VKS utilise des scores en gradation incrémentale, de la connaissance nulle jusqu'à l'intégralité des trois dimensions d'un mot (à savoir la forme, le sens et l'emploi), pour mesurer les connaissances lexicales chez un apprenant pour un mot donné. Il permet donc de dévoiler, de manière multidimensionnelle et continue, l'évolution des connaissances pour les mots cibles chez les participants.
- 2) Le VKS se présente sous forme de questionnaire à réponses libres. Par rapport à d'autres méthodes comme le test à choix multiple (désormais QCM), par exemple « Vocabulary Level Test » (désormais VLT) version réceptive (Nation, 2001; Schmitt, 2000) et « Word-Associate Test » (Read, 1993) qui

⁵⁴ « The acquisition of the meaning of any new lexical items from reading is almost always initially only partial... »

demandent aux participants de cocher des cases d'options proposées, le VKS est plus flexible et plus apte à évaluer les connaissances partielles. En outre, le format de réponses libres réduit en quelque sorte le risque que les participants soient inspirés ou trompés par les options improprement choisies. En effet, Waring & Takaki (2003) ont utilisé deux types de test, le QCM et le test à réponses libres, pour évaluer la rétention des mots cibles rencontrés à différentes fréquences dans l'input. Ces chercheurs ont conclu que le type de test de vocabulaire a un grand effet sur les scores du gain lexical et insistent sur le fait que les chercheurs doivent être particulièrement prudents lorsqu'ils choisissent des tests à choix multiples pour valider l'apprentissage du vocabulaire (p. 153).

Dans notre étude, nous avons sollicité les participants ayant un niveau intermédiaire en chinois L2. Les apprenants de ce niveau ont en général un vocabulaire chinois plutôt limité. Autrement dit, nos participants risquent de ne pas connaître tous les mots chinois en option dans un QCM. Bruton (2009) conclut, d'après son analyse du VKS, qu'un apprenant prend plus de temps pour donner des réponses libres que pour répondre à un QCM. Le VKS serait donc plus adapté pour un nombre limité de mots cibles. Par ailleurs, le test VKS peut être utilisé auprès de participants de tous niveaux.

- 3) Dans un questionnaire de VKS, les mots cibles sont présentés de façon isolée. Par rapport aux méthodes qui intègrent les mots cibles dans un contexte, par exemple le VLT version productive (Laufer & Nation, 1995), les participants y ont moins de possibilité d'être inspirés ou biaisés par le contexte.⁵⁵

Le test VKS a été adapté et employé dans les études en L2 dans la compréhension écrite, en anglais (Nassaji, 2004), en allemand (Rott & Williams, 2003), en espagnol (Pulido, 2004; Rott et al., 2002), en chinois (W. Zhao, 2014) et en français (Morrison, 1996; Paribakht & Tréville, 2007). Il a même été adapté et utilisé dans les études sur l'oral (De la Fuente, 2002).

Le VKS demande à un apprenant de justifier sa réponse si seulement

⁵⁵ Pour davantage d'informations, cf. Cervatiuc (2007) et Bruton (2007) sur les méthodes évaluant les connaissances lexicales.

l'apprenant estime qu'il a la connaissance du sens et/ou de l'emploi du mot, ce qui correspond au niveau 3 et à ceux supérieurs dans le modèle de Paribakht & Wesche (1997). Afin de pouvoir observer plus en détail l'évolution des connaissances lexicales, y compris celle formelle, pour les mots cibles chez nos participants, nous avons adapté le modèle de Paribakht & Wesche (1997) en ajoutant le niveau 2 qui correspond à une connaissance incertaine de la forme d'un mot donné. Nous avons donc utilisé une échelle de connaissances à 6 niveaux (présentés ci-dessous) dans notre test VKS, dans l'ordre croissant de la connaissance nulle jusqu'à la connaissance intégrale d'un mot, c'est-à-dire la forme, le sens et l'emploi.

1. Je n'ai jamais vu ce mot.
2. Je pense avoir déjà vu ce mot, mais je ne sais pas ce qu'il veut dire.
3. J'ai déjà vu ce mot, mais je ne sais pas ce qu'il veut dire.
4. J'ai déjà vu ce mot. Il veut peut-être dire ..., mais je n'en suis pas sûr(e).
5. Je connais ce mot et je sais qu'il veut dire...
6. Je connais ce mot et je sais qu'il veut dire... Je sais écrire une phrase en chinois avec ce mot.

Le niveau 1 reflète la connaissance nulle d'un mot, ou « jamais-vu ». Les niveaux 2 et 3 correspondent à une connaissance de la forme du mot à différentes nuances de certitude. Les niveaux 4 et 5 impliquent la connaissance sémantique du mot concerné et le niveau 6 représente la capacité d'employer correctement le mot concerné dans un cadre contextualisé.

Dans notre expérience, le test VKS était administré sous format papier où les mots chinois étaient présentés isolément dans une liste (cf. l'Annexe 1). Tous les participants devaient choisir l'un des 6 niveaux de connaissances pour chaque mot, en marquant le numéro correspondant. Si un participant auto-évaluait les niveaux 4 ou 5 pour un mot, il devait donner la signification du mot en français ; s'il choisissait le niveau 6 pour un mot, il devait donner la signification du mot et produire une phrase en chinois avec ce mot à fin de justification.

4.2.4 Tâches post-inférence à réaliser

Nous avons présenté dans le Chapitre II les deux facteurs importants dans l'apprentissage incident du vocabulaire, l'attention et la profondeur de traitement

concernant un mot nouveau dans l'input. Des études ont montré que le type de la tâche à réaliser dans l'input ont un effet direct sur le résultat d'apprentissage des mots nouveaux rencontrés en lecture : plus la tâche oriente l'attention d'un apprenant vers un mot nouveau et incite l'apprenant à traiter en profondeur les informations lexicales concernant le mot, par exemple l'utilisation du mot dans divers contextes, plus l'apprenant obtient de meilleurs gains lexicaux pour ce mot.

Concernant notre étude, nous nous intéressons aux gains lexicaux sur les mots inconnus rencontrés en lecture à la suite de l'inférence lexicale. Nous cherchions à savoir si la combinaison de deux tâches post-inférence, par rapport à une seule tâche, aboutit à de meilleurs gains lexicaux sur les mots cibles à court et à long terme.

Nous avons opté pour la consultation d'un dictionnaire et l'exercice lacunaire comme tâche supplémentaire à l'issue de l'inférence des sens des mots cibles.

4.2.4.1 Exercice lacunaire

Il existe différents types d'exercices pour l'enseignement/apprentissage du lexique chinois, par exemple on peut former un mot en reliant deux caractères parmi ceux proposés. Par rapport aux exercices dans lesquels les mots sont présentés isolément, l'exercice lacunaire, c'est-à-dire compléter une phrase avec un mot approprié, s'appuie davantage sur la maîtrise d'un mot dans les aspects sémantique et grammatical dans un cadre contextualisé.

Nous avons conçu un exercice lacunaire avec 20 phrases. Les sens de certains mots inconnus par les participants dans les phrases sont fournis à côté. En accompagnement des phrases, une liste des 20 mots cibles est fournie, avec leur prononciation transcrite en Pinyin et leurs significations en français (Les significations en français de ces 20 mots cibles sont fournies par le *Dictionnaire chinois-français* (« 汉法词典 », 2001). Les mots sont numérotés. Les participants doivent choisir un mot parmi les 20 mots cibles afin de compléter chaque phrase en marquant le numéro correspondant.

Les 20 phrases, accompagnées de leur traduction en français et la bonne réponse, sont présentées ci-dessous.

1. “我们公司的销售额全国第一！”员工们欢呼(faire des ovations)着、
____着。
« Notre entreprise est classée numéro un national en chiffre d'affaires ! » les employé-es font des ovations en ____. → 雀跃 (quèyuè, sauter de joie)
2. 他很有钱，可是很____，每天骑着他的旧自行车去公司。
Il est riche, mais il est ____. *Il va au travail tous les jours avec son vieux vélo.*
→ 节俭 (jiéjiǎn, économe, frugal)
3. 中国人过年的时候都会吃____，北方人爱吃甜的，南方人爱吃咸(salé)的。
Les Chinois mangent souvent des ____ lors du Nouvel an. Au nord on les aime sucrés et au sud on les préfère salés. → 汤圆 (tāngyuán, boulette de riz glutineux farcie de viande (fourrée à la confiture))
4. 这家养老院(maison de retraite)很好，老人们在这里过着____的生活(vie)。
Elle est bien cette maison de retraite. Les personnes âgées y ont une vie ____.
→ 安适 (ānshì, tranquille et confortable)
5. 那条河边有很多____的卵石(galet), 红的、白的、绿的……非常漂亮。
Au bord de la rivière il y a beaucoup de galets ____, rouge, blanc, vert... C'est très joli. → 五颜六色 (wǔyánliùsè, de diverses couleurs, multicolore)
6. 我过生日的时候，他送了很贵的礼物，他生日的时候我应该买些什么____呢？
Elle m'a offert un cadeau cher lors de mon anniversaire. Pour le sien, que pourrais-je acheter pour ____ ? → 还礼 (huánlǐ, faire un présent en retour, offrir un cadeau à son tour)
7. ____是中国人过年前的一种传统(traditionnel)活动。
____ est une activité traditionnelle que les Chinois font avant la fête du Nouvel an. → 送灶(sòngzào, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune)
8. 只有在夏天我们才会听见____叫。
On entend les ____ chanter seulement en été. → 蝉 (chán, cigale)
9. ____是中国年的最后一个月。
____ est le dernier mois dans le calendrier chinois. → 腊月 (làyuè, le douzième mois du calendrier lunaire)
10. ____有七种颜色(couleur)。
____ a sept couleurs. → 彩虹(cǎihóng, arc-en-ciel)
11. 春天到了，老师带着孩子们去农村____。
En printemps, les instituteurs amènent les enfants à la campagne pour ____.
→ 踏青(tàqīng, faire une excursion (se promener) à la campagne au printemps)
12. 雪下得真大，到处(partout)都是____的。

- Il neige beaucoup. C'est ____ par tout. → 白茫茫(báimángmáng, d'une blancheur éclatante (immaculée))*
13. 她们是好朋友, 常常在一起喝茶____。
Elles sont bonnes amies. Elles se retrouvent souvent pour boire du thé et pour ____. → 闲聊(xiánliáo, bavarder)
14. 我不喜欢这本书, 很____。
Je n'aime pas ce livre. Il est ____. → 乏味(fáwèi, fade, insipide)
15. 这里春天风很大, 是放____最好的季节。
Ici il vente fort au printemps. C'est la meilleure saison pour faire du ____. → 风筝(fēngzhēng, cerf-volant)
16. 过年的时候, 妈妈会在家里的窗户(fenêtre)上贴(coller)很多____, 好看极了。
Lors du Nouvel an, maman colle beaucoup de ____ sur les fenêtres. C'est très joli. → 窗花(chuānghuā, papier découpé pour la décoration d'une fenêtre)
17. 天太热了! 人们都____着快点儿下雨。
Il fait trop chaud! Les gens ____ tous que la pluie tombe très prochainement. → 盼望(pànwàng, espérer, désirer)
18. 我很喜欢这份工作。能不能赚(gagner)很多钱, 我不____。
J'aime bien ce travail. Je ne ____ pas qu'il m'apporte beaucoup d'argent. → 在乎(zàihu, prendre qch. à cœur, faire attention à)
19. 那个地方的景色(paysage)____, 所以有很多人去那里玩儿。
Là-bas le paysage est ____. → 明丽(mínglì, clair et beau)
20. 看, 下雪了, 一片、两片、三片……雪花(flocon de neige)在风中____。
Regarde, il neige. Un, deux, trois... les flocons ____ avec le vent. → 飘零(piāolíng, tomber en voltigeant)

Nous avons consulté les enseignants de chinois de LEA et de LLCE à l'UT2J au sujet du niveau de difficulté de cet exercice. Ils nous ont confirmé que cet exercice correspondait au niveau de la Licence 3.

4.2.4.2 Consultation d'un dictionnaire

Nous avons également opté pour la consultation d'un dictionnaire bilingue comme tâche supplémentaire à l'issue de l'inférence lexicale particulièrement parce que le dictionnaire est l'un des principaux recours pour les apprenants lorsqu'ils rencontrent un mot inconnu, surtout dans un contexte extrascolaire. Notamment à

mesure de la banalisation de l'Internet, et la prolifération de sites de dictionnaires, la consultation d'un dictionnaire est suffisamment accessible et pratique pour que les apprenants puissent accomplir cette recherche en autonomie hors classe. En outre, lorsqu'un apprenant cherche un mot dans un dictionnaire, il doit prêter attention avant tout, si ce n'est uniquement, à la forme du mot, ce qui peut l'aider à mieux le mémoriser. C'est une étape primordiale dans la création d'un nouveau lemme dans le lexique mental de l'apprenant. Particulièrement dans l'apprentissage du chinois, il est difficile pour un apprenant dont la L1 est alphabétique de mémoriser la forme d'un caractère. La consultation d'un dictionnaire représente donc pour nous un grand intérêt dans l'apprentissage incident du vocabulaire en chinois L2. Concernant nos participants, dans les programmes pédagogiques de LLCE et LEA à l'UT2J, les étudiants doivent apprendre à consulter un dictionnaire monolingue chinois ou un dictionnaire chinois-français dès la première année de Licence. Ils maîtrisent donc cet outil. Les participants d'autres parcours nous ont également confirmé qu'ils maîtrisaient cet outil.

Il est à noter que, par rapport à une langue alphabétique, la consultation d'un dictionnaire chinois est plus complexe.

- 1) La mise en page peut varier d'un dictionnaire à l'autre. Les dictionnaires chinois sont classiquement indexés par ordre alphabétique du pinyin d'une syllabe donnée et par le composant-clé. Les deux accès principaux à un caractère dans un dictionnaire sont l'index Pinyin et l'index composant-clé.
- 2) Les mots composés de plus d'un caractère sont regroupés par leur premier caractère, dans l'ordre alphabétique de la syllabe phonétique du caractère suivant. Autrement dit, la procédure principale pour chercher un mot chinois dans un dictionnaire traditionnel commence par la recherche du premier caractère du mot. Par exemple, pour chercher le mot « 节俭 (*jiéjiǎn*, frugal) » ou le mot « 节日 (*jiérì*, fête) », on doit commencer par retrouver le caractère 节 (*jié*) dans le dictionnaire ; par la suite, on trouvera les deux mots concernés dans la sous-catégorie de 节 (*jié*), « 节俭 (*jiéjiǎn*, frugal) » se situant avant « 节日 (*jiérì*, fête) ».

Dans le *Dictionnaire du chinois moderne* (现代汉语词典, 1984), pour chercher le mot 节俭 (*jiéjiǎn*, frugal). Il existe deux pistes principales :

- 1) Par l'Index des syllabes en Pinyin (ou « 汉语拼音音节索引 ») si on connaît la prononciation du premier caractère, en l'occurrence 节 (*jié*), on cherche en premier dans l'index des syllabes en Pinyin pour trouver la page où commence le groupe de tous les caractères qui partagent la prononciation « *jié* », donc la page 576 (Figure 4.1, ①). On poursuit la recherche à partir de cette page jusqu'à ce qu'on trouve le bon caractère (② et ③) et puis dans sa sous-catégorie le mot qu'on cherche (④).

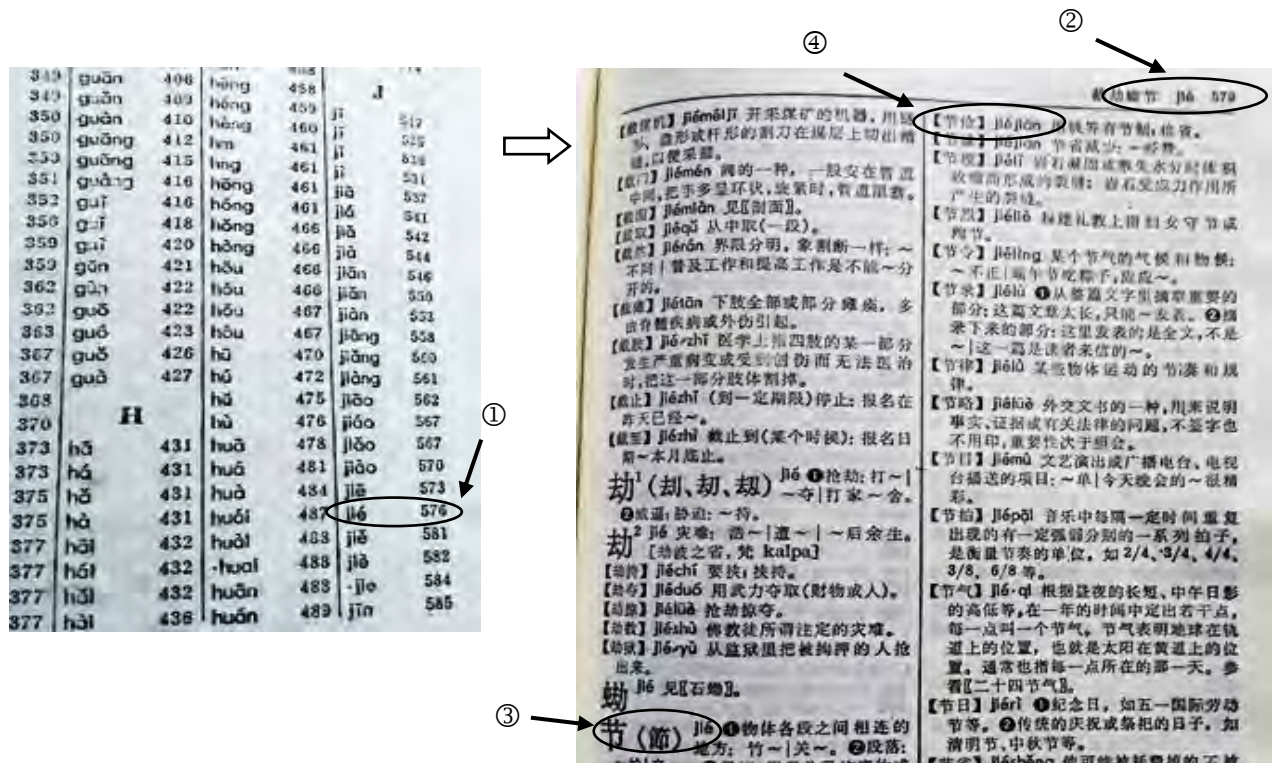


Figure 4.1 – Procédure pour consulter un mot chinois par l'accès phonétique dans un dictionnaire chinois monolingue

Il est à noter qu'on pourrait se passer de l'index général des syllabes et rechercher directement la syllabe du premier caractère dans un dictionnaire, puisque toutes les syllabes en Pinyin sont classées dans l'ordre alphabétique. Dans le *Dictionnaire chinois-français* (汉法词典, 2001), celui que nous avons

utilisé pour notre expérience, il n'y a pas d'index général des syllabes en Pinyin.

- 2) Par l'Index des composants-clés (ou « 部首目录 ») qu'on utilise lorsqu'on ne connaît pas la prononciation du premier caractère.

Le lien entre la forme écrite et la forme phonétique est relativement faible pour les caractères chinois. En d'autres termes, la prononciation d'un caractère chinois inconnu est souvent inaccessible par sa forme écrite. Pour cette raison, les composants-clés (部首) sont plus fréquemment utilisés pour chercher un caractère ou un mot inconnu en chinois. Concrètement, pour chercher le mot 节俭 (*jiéjiǎn*, frugal), on doit d'abord identifier le composant-clé du caractère « 节 (*jié*) », donc « 艹 » (la clé « herbe ») contenant 3 traits. Cette clé est donc classée sous la catégorie « 三画 (trois traits) » (Figure 4.2, p. 135, ①). On récupère ensuite la page 30 qui débute cette catégorie (② et ③). On compte ensuite le nombre des traits de l'autre composant du caractère. Dans le cas de « 节 (*jié*) », l'autre composant sous la clé « 艹 » contient 2 traits. On cherche alors dans la sous-catégorie « 一至二画 (1 à 2 traits) » (④) pour trouver « 节 » et son numéro de page (⑤ et ⑥). On trouve enfin « 节俭 (*jiéjiǎn*, frugal) » dans la sous-catégorie de « 节 (*jié*) » (⑦ et ⑧). (Deux numéros de page sont notés pour « 节 (*jié*) » dans ce dictionnaire parce que ce caractère possède deux prononciations, ce qui représente des sens différents et/ou des emplois variés. Dans notre étude, nous utilisons le « 节 (*jié*) » sur la page 579.)

Il est évident qu'avec la deuxième piste, les participants passent bien plus de temps à observer la forme des caractères composant les mots cibles. L'engagement cognitif est logiquement renforcé par rapport à la première piste. Revenons à notre étude. Compte tenu du temps nécessaire pour la lecture et l'inférence lexicale et de l'état de fatigue de nos participants, nous avons adopté la piste d'accès au mot par les syllabes phonétiques. Concrètement, nous avons fourni aux participants les prononciations transcrites en Pinyin des mots cibles pour les aider à retrouver plus rapidement ces mots dans le dictionnaire.



Figure 4.2 – Procédure pour consulter un mot chinois par l'accès composant-clé dans un dictionnaire chinois monolingue

4.3 Méthodes d'expérience

4.3.1 La méthode de la « pensée à voix haute »

Nous rappelons que l'un des focus de notre étude est de comprendre dans quelle mesure la lexicalisation dans la L1 intervient sur le processus de l'inférence du sens d'un mot L2 inconnu. Plus précisément, nous cherchons à savoir sur quelles sources de connaissance s'appuient nos participants pour inférer le sens des mots chinois inconnus et si leurs choix de sources de connaissance varient selon que le mot

chinois inconnu est lexicalisé ou non en français. Pour cela, nous avons utilisé la méthode ou protocole de la « pensée à voix haute » (« Think aloud »). La méthode de la « pensée à voix haute » (PVH en abrégé) demande aux participants de révéler verbalement leurs pensées durant une tâche spécifique (Figure 4.3).

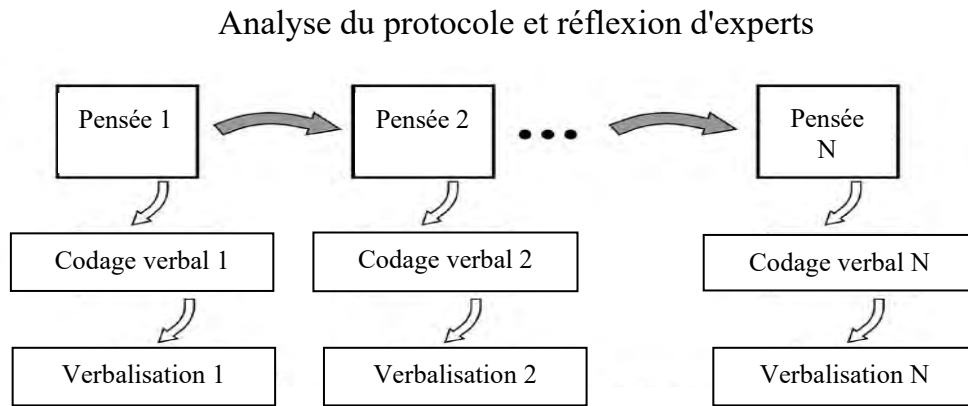


Figure 4.3 – Déroulement de la pensée à voix haute durant une tâche (Ericsson, 2006, p. 227)

Cette méthode donne la possibilité aux chercheurs d’observer sur quelles informations un participant se concentre et comment les informations sont structurées durant une tâche de résolution d’un problème (Fonteyn et al., 1993, p. 430). Selon Ericsson & Simon (1984), il existe différents types de verbalisations : 1) « parler à voix haute » (« talk-aloud », où on dit ce qu’on pense) ; 2) « penser à voix haute » (« think-aloud », où on décrit ou explique ce qu’on pense) et 3) « rétrospection », où on décrit et explique ce qu’on a fait. Classés introspectifs, les deux premiers types se réalisent pendant une tâche, tandis que le troisième, qui consiste en une récupération ultérieure des informations après la tâche, est par conséquent plus subjectif et biaisé (Ericsson, 2006; Ericsson & Simon, 1993). Cohen, (1987, p. 84, 2014, p. 99) fixe à 20 secondes après la production d’un évènement, le seuil de distinction entre le rapport introspectif et le rapport rétrospectif.

La méthode PVH, largement utilisée aujourd’hui dans la recherche, renvoie généralement au rapport verbal introspectif réalisé en parallèle avec une tâche à résoudre un certain problème (Bowles, 2010). L’intérêt de ce type de rapport verbal est qu’il « ...fournit des informations sur les comportements séquentiels (dans l’ordre

du temps) d'un sujet. (Crutcher, 1994, p. 241)⁵⁶ »

Dans les années 1950 et 1960, la PVH fut utilisée dans la recherche en psychologie pour sonder les processus cognitifs humains. Depuis les années 1970, les chercheurs travaillant dans le domaine de l'acquisition des langues secondes/étrangères ont adapté la PVH pour les explorations dans la compréhension écrite. Très répandue dans les études en L2, la PVH est utilisée dans la compréhension et la production écrite (Cohen, 1987; Hayes & Flower, 1983), l'apprentissage et la pratique de la L2 (Cohen, 2014), les stratégies de lecture en L1 et en L2 (Yamashita, 2001b), les stratégies de prise de test (Cohen, 2000; Storey, 1997), l'attention de l'apprenant dans la compréhension écrite en L2 (Leow, 2001) et l'inférence lexicale (L. Fan & Zhang, 2014; Fraser, 1999; Haastруп, 1991, 2008; Morrison, 1996; Nassaji, 2004; Théophanous & Dreyfus, 2004; Wesche & Paribakht, 2009) (cf. la Figure 4.4 pour le schéma synthétique de l'utilisation de la PVH dans l'inférence lexicale).

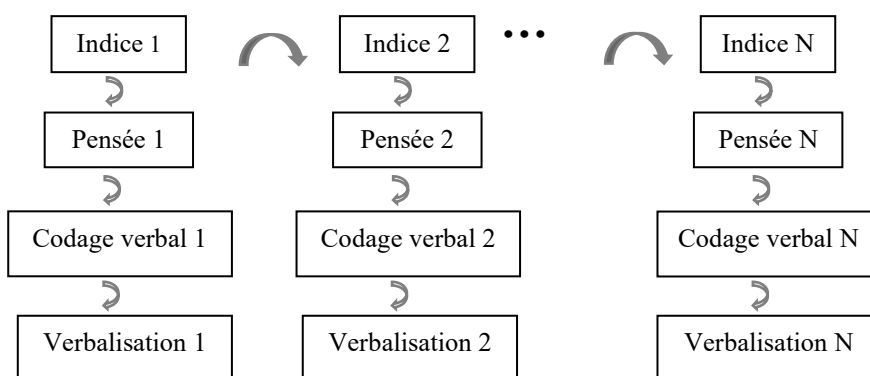


Figure 4.4 – Schéma synthétique de l'utilisation de la méthode de la pensée à voix haute dans l'inférence lexicale (basé sur le schéma d'Ericsson, 2006, p. 227)

A part son rôle de méthode d'investigation, la PVH est aussi employée comme une méthode d'instruction dans la compréhension écrite et orale. Elle engage les enseignants et les élèves dans des dialogues plus collaboratifs et interactifs, afin d'élaborer progressivement le sens d'un texte ou d'une conversation. (Cf. Kucan & Beck, 1997)

⁵⁶ « ..provide information on the sequential (time-ordered) behavior of subject. »

4.3.1.1 Interrogations sur la validité de la pensée à voix haute

Selon la Figure 4.3 (cf. supra p. 136) qui montre le prototype du déroulement de la PVH pendant une tâche, les pensées s'enchaînent les unes après les autres. Avant d'arriver au stade final de la verbalisation, le participant doit d'abord passer par le codage verbal dont la réalisation exige du temps et des ressources cognitives. Nous pouvons donc facilement imaginer qu'en réalité, les pensées ne se déroulent pas de façon séquentielle ; même si c'était le cas, la capacité de s'exprimer varie d'une personne à l'autre et il est possible que la pensée $N+1$ se déclenche avant la fin de la verbalisation de la pensée N . Cela signifie que le déroulement de la PVH pendant qu'un sujet résout un problème se réalise probablement avec les maillons décalés, de façon synthétique comme ceci :

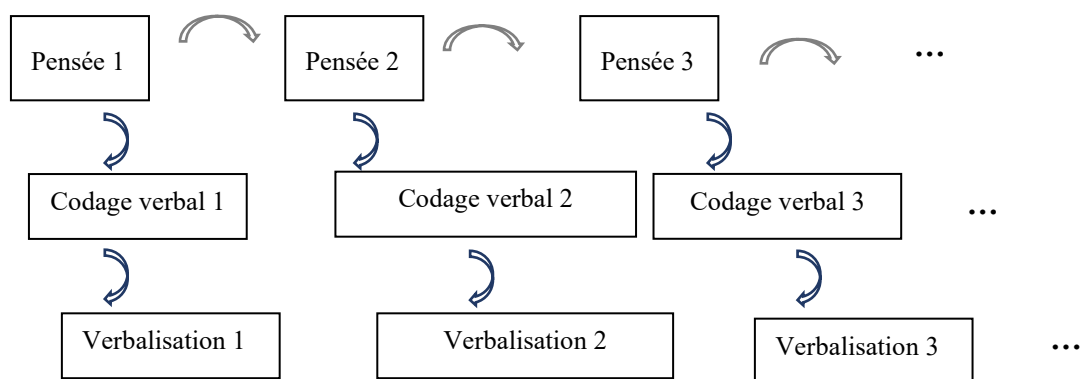


Figure 4.5 – Déroulement non séquentiel de la pensée à voix haute pendant une tâche (basée sur le schéma d'Ericsson,2006, p. 227)

Pour cela, malgré la grande popularité de la méthode PVH, les chercheurs en psychologie qui eux-mêmes ont utilisé la PVH, ont soulevé des questions depuis les années 1970 sur la validité de la PVH. Les critiques se concentrent généralement sur les deux aspects suivants :

- 1) L'apprenant risque de rapporter des informations incomplètes. Par exemple, Ericsson & Simon (1984) remarquent que les pensées se déroulent plus rapidement que les productions de paroles. Il est donc impossible pour un individu de verbaliser tous les détails de ses pensées.
- 2) Le fait que la verbalisation se fait en parallèle avec la résolution d'une tâche

surcharge cognitivement le participant. Par conséquent, la PVH crée de l'interférence avec la tâche principale (Ericsson & Simon, 1980, 1993; Storey, 1997). Le terme le plus courant pour l'interférence est la « réactivité » qui signifie que la PVH altère la performance de la tâche réalisée (Bowles, 2010, p. 113). Cohen (2000) indique que, par exemple, l'introspection pourrait inciter les participants à raconter ce qu'ils croient avoir fait ou ce qu'ils font habituellement au lieu de la réalité.

Des études ont été menées dans différents domaines avec l'objectif d'examiner la validité de la PVH. Ericsson & Simon (1993) résument, dans leur revue de littérature renvoyant à une dizaine d'études réalisées à ce sujet, qu'il n'y pas de preuve montrant que la PVH du type « talk-aloud » change la séquence des pensées d'un sujet durant la réalisation d'une tâche, par rapport à la condition dans laquelle il accomplit la tâche en silence. Cependant, les deux autres types de PVH sont influents sur la performance de la tâche principalement parce qu'ils surchargent le sujet avec des processus cognitifs supplémentaires à force de l'obliger à produire des explications et des descriptions verbales. En général, la PVH, quel que soit le type, risque de ralentir la réalisation d'une tâche. Selon Ericsson & Simon (1993), un adulte doit maîtriser les techniques permettant de verbaliser ses pensées de façon simultanée sans changement systématique de ses processus de pensées, après avoir reçu une brève instruction ou une formation de familiarisation à la PVH. Il est donc important de fournir une formation à la méthode PVH, par le biais de différents exercices, aux sujets testés avant une expérience.

Dans le domaine de l'acquisition d'une L2, selon Bowles (2010), l'étude de Leow & Morgan-Short (2004) est la première qui a porté sur le test de la validité de la méthode PVH. Les chercheurs ont sollicité des apprenants débutants en espagnol afin d'effectuer des tâches de compréhension écrite et d'observer leur acquisition subséquente de la grammaire dans la production écrite. Les résultats ne montrent pas de réactivité de la PVH. Bowles & Leow (2005), dans leur étude, distinguent deux catégories de verbalisation, une « non métalinguistique » qui se réfère à la verbalisation en soi et une « métalinguistique » à l'explication et la justification

verbale⁵⁷. Les chercheurs ont comparé trois groupes d'apprenants en espagnol L2, à savoir le non métalinguistique, le métalinguistique et le contrôle, afin d'observer la réactivité des deux types de verbalisation dans la compréhension et la production écrite. Leurs résultats montrent que, par rapport au groupe de contrôle ayant effectué la lecture en silence, les deux groupes de verbalisation ont passé bien plus de temps pour accomplir la tâche de compréhension écrite ; au niveau du score de compréhension, aucune différence significative ne s'est révélée entre le groupe de contrôle et l'un des deux groupes de verbalisation ; en revanche, le groupe de verbalisation non métalinguistique a un score significativement meilleur que celui du groupe métalinguistique. Les chercheurs sont arrivés à la conclusion que la verbalisation a avant tout une influence sur le temps de lecture, mais non sur le résultat de compréhension. La différence de compréhension significative entre les deux groupes est probablement liée au fait que le texte testé est long (plus de 800 mots) et la structure syntaxique est assez difficile pour les participants.

En résumé, des études investiguant la validité de la PVH montrent que globalement la réactivité de cette méthode est négligeable quand les participants rapportent leurs pensées de façon simultanée pendant une tâche sans explication ni description ; une formation sur la méthode PVH avant une expérience est nécessaire afin de minimiser l'interférence liée à la verbalisation.

Il ne faut pas non plus oublier qu'à force de verbaliser ses pensées, un apprenant a besoin de plus de temps pour accomplir une tâche de lecture. Conjointement au fait que la transcription des entretiens réalisés avec la PVH et les analyses qualitatives subséquentes sont également chronophages (Afflerbach & Johnston, 1984), l'utilisation de la méthode PVH est de préférence réservée pour les études recrutant un nombre de participants relativement restreint.

4.3.1.2 Variables importantes de la pensée à voix haute

Malgré les limites évoquées ci-dessus, la PVH est le seul moyen qui permet d'observer, chez un sujet testé, les informations impliquées pendant le déroulement

⁵⁷ Selon Bowles (2010, p. 113), « non métalinguistique » correspond au « talk-aloud » du classement d'Ericsson & Simon (1993), et « métalinguistique » au type « think-aloud » et type « retrospection ».

de ses pensées. Cela explique la grande popularité de cette méthode dans le domaine de l'acquisition des L2. Pratiquement dans toutes les branches de recherche dans le domaine de l'acquisition des langues secondes/étrangères, l'utilisation de la PVH est en hausse (Bowles, 2008).

Bowles (2010, p. 121) déclare à la fin de sa revue de littérature sur la PVH que l'utilisation de cette méthode n'est pas une question unidimensionnelle et qu'elle dépend d'un grand nombre de variables. A ce sujet, Afflerbach & Johnston (1984) mettent en avant plusieurs variables qui pourraient influencer les résultats de la PVH : 1) la formation préalable à la PVH ; 2) le type de la tâche à réaliser ; 3) les questions de relance lors du silence du participants ; 4) le caractère et le niveau d'engagement du participant et 5) la façon de traiter les données.

Dans notre expérience, nous avons opté pour la PVH parce qu'elle nous permet d'observer de façon explicite sur quels indices les participants s'appuient pour déduire le sens d'un mot chinois inconnu, information qui n'est pas accessible par d'autres moyens expérimentaux. Compte tenu des variables mises en avant par Afflerbach & Johnston (1984), afin de pouvoir étudier l'inférence lexicale à l'aide de la PVH dans une condition optimale, nous avons pris les mesures suivantes :

- 1) Une tâche trop difficile décourage les participants à la mener jusqu'au bout. Au contraire, la réalisation d'une tâche trop facile risque d'être automatisée pour les participants et il est donc possible que certains processus cognitifs leur soient indécélables (Nisbett & Wilson, 1977). Nous nous sommes arrangée pour que la compréhension des textes expérimentaux et l'inférence des mots cibles soient « faisables » (Afflerbach & Johnston, 1984, p. 314), à la portée de nos participants.
- 2) Tous les participants ont suivi une formation sur la PVH juste avant l'inférence lexicale pour qu'ils comprennent ce que nous attendions d'eux.
- 3) Nous avons préparé quelques questions neutres, telles que « Qu'est-ce que vous regardez ? », « A quoi pensez-vous ? », pour relancer les participants immergés dans leurs pensées ou qui n'étaient pas très prolixes.

4.3.2 Oculométrie⁵⁸

Afin de répondre à la question de recherche concernant l'attention qu'un apprenant francophone porte aux mots chinois inconnus pendant la lecture, nous avons utilisé l'oculométrie. En tant que méthode d'expérience, l'oculométrie est relativement récente et peu présente dans le domaine linguistique. A notre connaissance, jusqu'au moment de notre expérience, il n'y a pas eu d'étude qui utilise l'oculométrie pour mesurer l'attention qu'un sujet prête aux mots non familiers pendant la compréhension d'un texte en chinois L2. Cette partie de notre expérience est donc exploratoire.

Selon l'hypothèse de noticing (cf. la Section 2.5.1), l'attention de l'apprenant est primordiale pour que l'input puisse se transformer en saisie. Les lecteurs ne se tiennent pas toujours compte qu'ils ne connaissent pas certains mots lors de leur lecture. L'identification des mots inconnus est traditionnellement réalisée par l'intermédiaire des rapports verbaux, introspectifs ou rétrospectifs (Godfroid & Schmidtke, 2013). Concrètement, les expérimentateurs demandent aux participants venant de finir une lecture de marquer dans le texte les mots qu'ils jugent inconnus. Nous venons de voir dans la section précédente qu'on interroge la validité des rapports verbaux, notamment celle du rapport rétrospectif. Afin de pouvoir mesurer l'attention qu'un apprenant porte sur un mot inconnu pendant sa lecture de façon plus intégrale et objective, nous avons adopté la technique d'oculométrie (« Eye-tracking »).

L'oculométrie est une technique qui trace les mouvements des yeux de quelqu'un pendant une durée spécifique, par exemple lors de l'exécution d'une tâche intellectuelle. Just & Carpenter (1980) indiquent que les mouvements d'yeux d'une personne peuvent refléter ses processus cognitifs. Cela devient l'argument général des chercheurs qui emploient l'oculométrie afin d'explorer des activités cognitives humaines. Les premiers prototypes d'oculométrie datent de la fin du 19^{ème} siècle. A partir des années 1970 paraissaient les systèmes d'oculométrie dédiés à la recherche.

⁵⁸ Nous tenons à remercier chaleureusement Mme Emilie MASSA et M. Pierre-Vincent PAUBEL pour leur aide précieuse sur la conception et la mise en place de l'expérience par le biais de l'oculométrie.

L'utilisation de l'oculométrie a connu une croissance considérable depuis ces dernières décennies. Selon les chiffres de Holmqvist et ses collègues (2011), en 2009 il y avait environ 23 entreprises fabriquant les systèmes d'oculométrie, dont seulement deux avaient été fondées avant 1985. Plus de la moitié de ces 23 entreprises ont été fondées après 2000. Depuis les années 1990, le fonctionnement dominant des systèmes d'oculométrie consiste à enregistrer en vidéo (« video-based ») les réflexions des pupilles et des cornées des yeux d'un sujet testé. Un eye-tracker (c'est-à-dire l'appareil d'oculométrie) qui fonctionne en vidéo comprend généralement deux composants principaux : a) l'illuminateur qui projette des rayons infrarouges et b) la caméra qui capte les réflexions des pupilles et des cornées. (Cf. Holmqvist et al., 2011, pp. 9-12, 21-29)

4.3.2.1 Mesures d'oculométrie utilisées

L'oculométrie peut fournir une grande variété de données mesurant les informations comportementales d'un sujet testé dans différents aspects. Parmi toutes les mesures, le temps total de la fixation du regard (« Dwell time », désormais « temps total »), la durée de la première fixation du regard (« First Fixation Duration », désormais « première fixation ») et le nombre de la revisite du regard se trouvent aux premiers rangs d'utilisation.

Le temps total est sans doute la mesure la plus utilisée par les chercheurs parce qu'on la considère comme l'indicateur direct de l'attention d'une personne. Selon la définition, le temps total renvoie à la durée totale du regard d'un sujet restant sur une partie spécifique sur un stimulus visuel. Ce genre de partie délimitée est nommée « aire d'intérêt » (« Area Of Interest », désormais AOI), sur laquelle le système d'oculométrie recueille et traite des données. Le temps total peut refléter l'effort exigé pour traiter certaines informations ou la profondeur d'un processus cognitif (Holmqvist et al., 2011; Jegerski, 2014). Dans la compréhension écrite, par exemple, Rayner (1998) remarque que les mots les moins fréquents reçoivent des fixations du regard plus longues parce qu'ils exigent un processus d'activation plus long, ce qui corrèle avec les résultats obtenus avec ERP⁵⁹ de Dambacher & Kliegl (2007) (cf.

⁵⁹ ERP – « Event-Related Potential » (« Potentiel lié aux événements »)

Holmqvist et al., 2011, p. 382).

La première fixation renvoie à la durée de la première fixation du regard d'un sujet testé dans une AOI spécifique. Cette mesure reflète le temps nécessaire pour les traitements rapides, tels que la reconnaissance et l'identification. Juste après le temps total, elle est la deuxième mesure la plus fréquemment utilisée dans les études concernant la compréhension écrite : par le biais de cette mesure, on peut observer les facteurs qui peuvent intervenir dans le processus initial d'activation de l'accès lexical à un mot, par exemple la fréquence du mot, la complexité morphologique, la propriété orthographique et le degré polysémique (Clifton Jr. et al., 2007; Radach et al., 2004) (cf. Holmqvist et al., 2011, p. 385).

La revisite⁶⁰, c'est-à-dire les repassages du regard dans une AOI que les yeux ont déjà quitté, est aussi une mesure importante. Selon Holmqvist et al. (2011), les zones sémantiquement informatives et intéressantes reçoivent souvent de nombreuses revisites. Dans l'inférence lexicale en L2, la revisite pourrait être un indicateur qu'un apprenant cherche des informations contextuelles pour inférer le sens d'un mot nouveau (Godfroid et al., 2013). Dolgunsöz (2016) explique que les revisites sont plutôt interprétées comme la réanalyse ou la vérification dans un contexte discursif des informations sémantiques et syntaxiques d'un mot.

4.3.2.2 Dispositifs informatiques de l'oculométrie

On peut distinguer deux modèles principaux d'eye-tracker en termes de mobilité, a) le modèle fixe (c'est-à-dire l'illuminateur et la caméra sont fixés sur une table, sans ou avec un mentonnier qui limite les mouvements de la tête du participant) et b) le modèle mobile en casque ou en lunettes avec l'illuminateur et la caméra intégrée. Le premier modèle projette les stimuli visuels sur un écran, les participants doivent donc rester devant cet écran pendant les manipulations. En revanche, le modèle mobile est peu restreint au niveau du support d'affichage, les participants sont donc libres physiquement et il y a moins de contraintes en termes de local d'expérience. Cela dit, cette liberté physique compromet la précision de

⁶⁰ D'autres termes, par exemple « retour » (« return »), sont aussi utilisés pour cette mesure. Nous gardons le terme « revisite » dans cette thèse, utilisé par SMI Red250 que nous avons utilisé pour réaliser notre projet.

calibration.

Compte tenu de nos exigences sur la précision de calibration, nous avons d'abord exclu le modèle mobile. Nous avons ensuite décidé d'adopter un modèle sans mentonnier, parce que le fait de limiter les mouvements de la tête d'un participant avec un dispositif mécanique risque fort de le gêner. Cela peut, par conséquent, le perturber pendant ses réflexions.

Nous avons finalement choisi le system SMI Red250, à la fréquence 250 Hz, fixée au bord inférieur d'un écran de 22 pouces à résolution 1920x1080 (pixel) avec une fréquence de rafraîchissement de 60 Hertz. Ce dispositif a été jugé comme le meilleur compromis par rapport à nos exigences, à la fois sur la précision de calibration et le confort du participant lui permettant une lecture relativement naturelle.

L'oculométrie est une méthode d'expérience performante qui fournit des données comportementales objectives. Cependant, cet outil présente des aspects contraignants qui nous ont causé, pendant les manipulations, quelques inconvénients :

- 1) En termes de local d'expérience, le modèle SMI Red250 demande l'affichage des stimuli sur un écran. Afin de préserver les mêmes conditions environnementales en termes de luminosité et d'isolation acoustique lors des recueils de données, pour tous les participants, nous avons conduit les manipulations à la plateforme d'oculométrie de la Maison de la Recherche à UT2J. Il nous était donc impossible de mettre en place l'expérience en nous adaptant aux emplacements des participants.
- 2) Il fallait choisir des participants en évitant dans la mesure du possible les éléments pouvant perturber le recueil de données d'oculométrie : les lunettes de vue sont entre autres « perturbantes » pour les eye-trackers à cause des reflets des verres ; les produits de maquillage de couleur foncée ou ceux scintillants autour des yeux, peuvent tromper les eye-trackers et biaiser les résultats (cf. Holmqvist et al., 2011, pp. 116-128). Pour cela les données provenant de certains participants ont été exclues des analyses finales (cf. la Section 6.2 du Chapitre VI).

- 3) Concernant les stimuli visuels, il était nécessaire d'adapter les textes testés en fonction des exigences d'oculométrie. La distance (entre les yeux et les stimuli) et l'angle de vue sont deux paramètres fondamentaux à contrôler pour assurer le suivi des yeux d'un sujet testé. Qui plus est, les parcours d'yeux d'un lecteur durant une lecture ne sont pas linéaires. Notamment pour un eye-tracker sans mentonnier comme SMI Red250, le moindre mouvement, par exemple, un grattement de la tête, un changement de position assise ou même un toussotement, peut perturber, voire interrompre, le suivi des yeux (cf. Holmqvist et al., 2011, pp. 37-43). De ce fait, afin d'optimiser la calibration, nous avons espacé, horizontalement et verticalement, les caractères de nos textes pour pouvoir créer une AOI suffisamment grande. Cela a eu des répercussions au niveau du nombre de caractères affichés sur l'écran (cf. la Section 4.2.2 pour la conception des textes expérimentaux) et de la mise à l'écran des textes, que nous allons présenter ci-dessous.

4.3.2.3 Format d'affichage des textes testés

Dans les études en lecture réalisées à l'aide de l'oculométrie, souvent pour satisfaire la qualité de la calibration, les stimuli en texte se résument à quelques mots ou phrases affichées gros en plein centre d'un écran. Il existe peu d'expériences qui utilisent un texte entier. A notre connaissance, jusqu'au moment de notre expérience, il n'y a pas eu d'étude qui utilise l'oculométrie pour mesurer l'attention qu'un sujet prête aux mots non familiers pendant la compréhension d'un texte en chinois.

Afin de favoriser une lecture fluide et d'optimiser la précision des données oculométriques, nous avons conçu la mise en page des textes selon les règles suivantes :

- Les mots cibles ne se situent pas aux quatre bords de l'écran ;
- Tous les caractères constitutifs d'un mot se situent sur une même ligne ;
- Deux mots cibles ne se trouvent pas dans une même proposition ;
- Les caractères constitutifs des mots cibles se situant dans deux lignes consécutives ne se trouvent pas dans une même colonne.

Dans le respect de toutes ces règles, afin d'optimiser la calibration pour obtenir

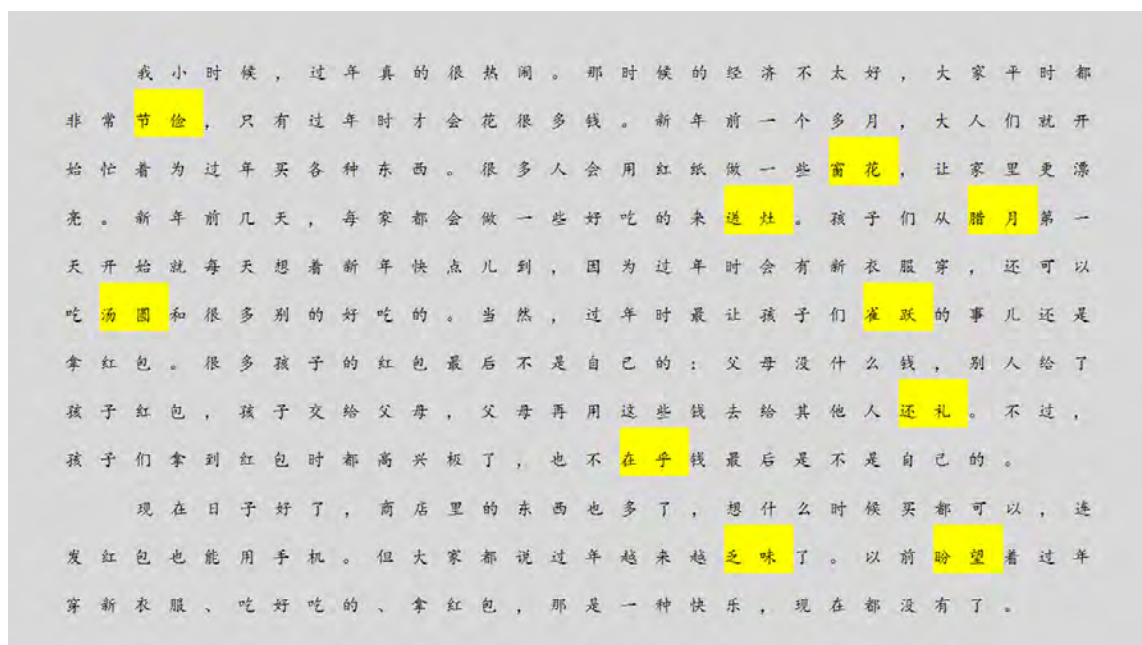
des données valides, nous avons conduit un bon nombre d'essais et plusieurs pré-tests avant de définir les paramètres d'affichage de texte comme ceci :

- Résolution : 1920 x 1080 (pixel)
- Police de caractère : Kaiti⁶¹
- Taille de police : 14 (point)
- Interligne : 38 (point)
- Espacement (entre caractères) : 13,4 (point)

Nous avons sollicité deux étudiants en chinois L2 pour faire des pré-tests. Selon les résultats des pré-tests, la partie sous la surveillance de l'oculométrie dure plus de 45 minutes. Compte tenu de la fatigue due à la fixation des yeux sur un écran pendant une durée importante, nous avons paramétré la couleur page en « Blanc, Arrière-plan 1, plus sombre 15% », afin de rendre l'affichage plus agréable.

Les deux textes ont été finalement affichés sur l'écran comme dans la Figure 4.6 ci-dessous avec les mots cibles surlignés en jaune⁶².

Texte A



我小时候，过年真的很热闹。那时候的经济不太好，大家平时都非常节俭，只有过年时才会花很多钱。新年前一个多月，大人们就开始忙着为过年买各种东西。很多人会用红纸做一些窗花，让家里更漂亮。新年前几天，每家都会做一些好吃的来送灶。孩子们从腊月第一天开始就每天想着新年快点儿到，因为过年时会有新衣服穿，还可以吃汤圆和很多别的好吃的。当然，过年时最让孩子们雀跃的事儿还是拿红包。很多孩子的红包最后不是自己的：父母没什么钱，别人给了孩子红包，孩子交给父母，父母再用这些钱去给其他人还礼。不过，孩子们拿到红包时都高兴极了，也不在乎钱最后是不是自己的。

现在日子好了，商店里的东西也多了，想什么时候买都可以，连发红包也能用手机。但大家都说过年越来越乏味了。以前盼望着过年穿新衣服、吃好吃的、拿红包，那是一种快乐，现在都没有了。

⁶¹ Kaiti est la police utilisée pour les textes et les exercices dans les manuels pédagogiques utilisés par la section de langue chinoise à UT2J, également la police la plus couramment utilisée dans les communications imprimées quotidiennes.

⁶² Lors du test, les mots cibles ne sont pas surlignés sur les stimuli pour la lecture.

Texte B

我的家乡是一个人口不多的小城市，交通很方便，一年四季都非常美，是一个很安逸的地方。家乡的春天很温暖，城里的树和草很早就换上了绿色的新衣，路边都开着五颜六色的小花儿。春天也是一个多雨的季节。每场春雨过后，小城都会变得更明丽。有时下完雨后，天上还会有漂亮的彩虹。每到天气好的时候，人们会到城市以外的地方去踏青。要是风比较大的话，还可以去小城旁边的山上放风筝。夏天很热，树上的蝉从早叫到晚。到了晚上，吃完了晚饭的人们都走出家门，大人们坐在一起闲聊，孩子们就在一边玩儿着，那是一天中最热闹的时候。一到秋天，天气就开始转凉，鸟儿们都飞去温暖的南方过冬。草和树叶慢慢地变黄，秋风刮过，树叶飘零。小城安静下来，等着冬天的到来。家乡的冬天常常下雪。大雪后的小城白茫茫的，城里的草地上、树上、房子上和城外的山上都是白雪，真是漂亮极了。

Figure 4.6 – Mise à l'écran des textes A et B et emplacement des mots cibles dans les textes

Dans notre cas, il était nécessaire de définir pour chaque caractère une AOI dans les deux textes. Toutes les AOI sont identiques en dimension. Plus précisément, il y a 318 AOI pour le texte A et 321 AOI pour le texte B (les signes de ponctuation ne sont pas pris en compte). Les AOI sont en général de dimension 80x57 (pixel)⁶³. Tous les caractères se trouvent au milieu de leur AOI qui est numérotée. La Figure 4.7 montre un exemple des AOI mises en place pour les caractères du Texte A.

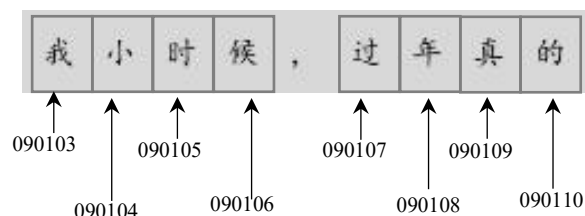


Figure 4.7 – Exemple des AOI mises en place pour les caractères

⁶³ Comme les dimensions des deux textes ne permettent pas que la hauteur totale (967 pixels) soit divisée de façon identique pour les 12 lignes du Texte A et du Texte B, la hauteur de certaines AOI est arrondie à 79 pixels.

Pour standardiser la procédure et les instructions de l'expérience, toutes les consignes pour les parties de la compréhension écrite et de l'inférence lexicale ont été affichées sur l'écran. Une image animée d'une croix de fixation situant en plein centre de l'écran est affichée au démarrage des stimuli de texte pour la lecture. La croix sert à guider les regards du participant au départ. L'affichage de la croix ne dure que 500 millisecondes, et au-delà elle disparaît, enchaînée par un stimulus de texte (cf. Annexe 2 pour les stimuli oculométriques utilisés dans notre expérience).

Afin d'éviter des variations de l'affichage liées au traitement des caractères chinois sous différents paramétrages de matériels informatiques, tous les stimuli visuels du test ont d'abord été créés sous format DOCX et ensuite transformés en image sous format JPG, à la résolution 1920 x 1080 (pixel).

4.4 Déroulement de l'expérience

4.4.1 Pré-test

Avant de lancer formellement l'expérience, nous avons sollicité deux étudiants en LEA Master 2 à l'UT2J pour faire des pré-tests. Les deux étudiants ont d'abord été testés par le biais du VKS sur leur niveau de connaissance des mots cibles. Une semaine plus tard, ils ont suivi une formation sur la méthode PVH et, sous la surveillance de l'oculométrie, ont ensuite lu les textes à fin de compréhension et effectué les inférences des sens des mots cibles présents dans les textes. A l'issue de la lecture, ils ont accompli les tâches post-inférence, l'exercice lacunaire suivi ou non de la consultation du dictionnaire bilingue afin de compléter l'exercice.

A la fin du pré-test, ces deux étudiants nous ont confirmé que toutes les indications étaient claires et ils n'ont pas eu de problème pour comprendre et accomplir les tâches post-inférence.

4.4.2 Procédure

Nous avons contacté les étudiants en Licence 3 à l'UT2J, par le biais de leurs enseignants. En parallèle, nous cherchions d'autres apprenants de niveau similaire en chinois L2 via les associations d'enseignement du chinois à Toulouse. Nous avons

recruté au total 33 participants (cf. la Section 4.1).

Notre expérience contient quatre parties. Elle s'est déroulée pendant plus de cinq semaines (Figure 4.8).

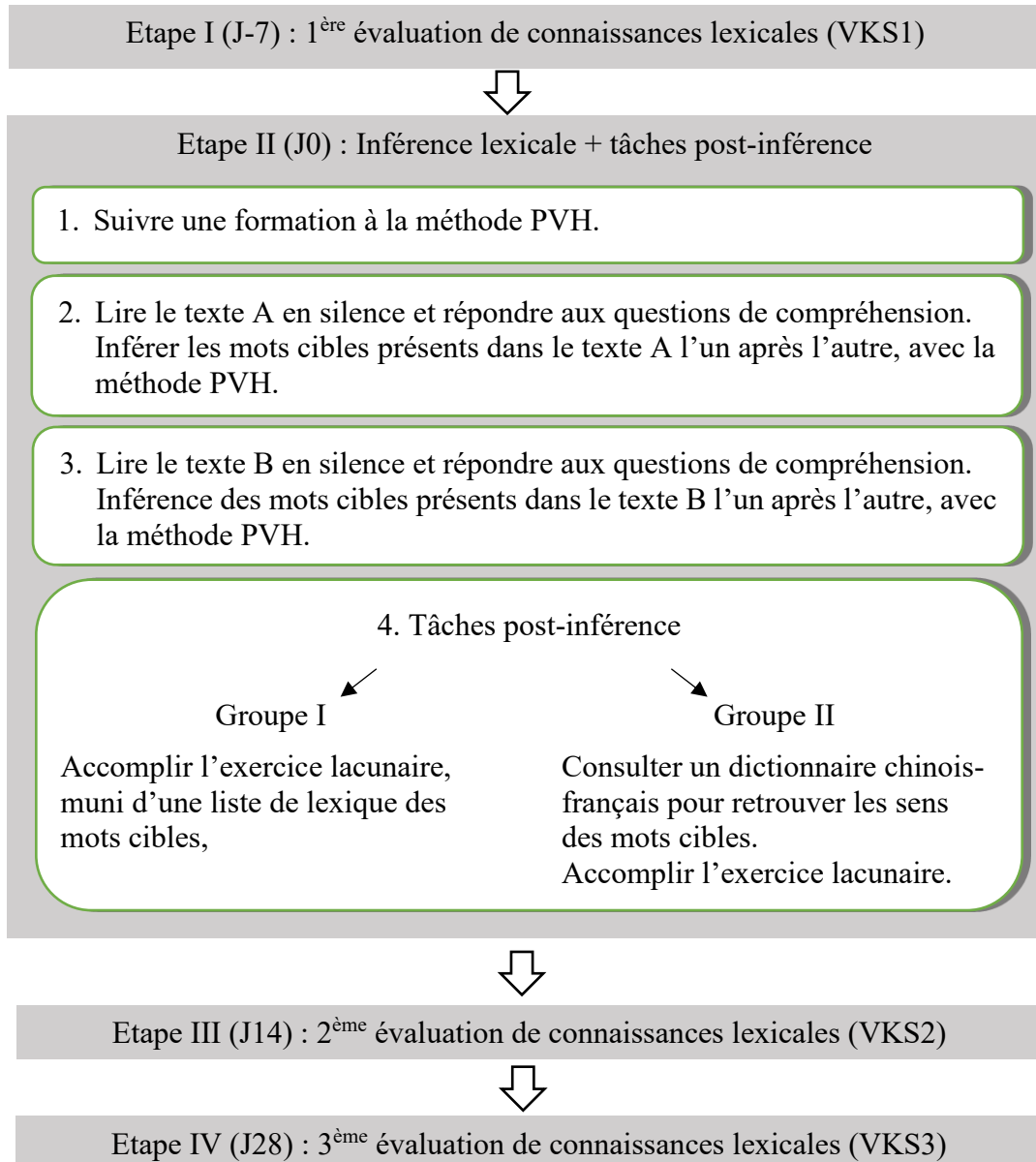


Figure 4.8 – Procédure de l'expérience

4.4.2.1 Étape I – Evaluation de connaissances lexicales avant l'inférence lexicale

Nous avons administré le premier VKS (désormais VKS1) au moins une semaine avant la séance de l'inférence lexicale (marqué « J-7 » dans la Figure 4.8). Les résultats du VKS1 étaient utiles pour nous assurer de la méconnaissance des mots

cibles chez les participants. La majorité des participants ont répondu au VKS1 en classe à la fin d'un cours de chinois. D'autres l'ont fait lors d'un rendez-vous individuel avec nous.

Les réponses des premiers participants au VKS1 montrent que certains participants connaissent quelques mots cibles. Nous avons donc remplacé ces mots cibles par des synonymes moins couramment utilisés. Du fait de ces modifications, certains participants ont accompli le VKS1 plusieurs fois. Le VKS1 en version définitive contient une liste de 60 mots chinois, comportant les 20 mots cibles.

A la fin du VKS1, nous avons fixé par échange de courriel un rendez-vous individuel d'une durée de deux heures⁶⁴ avec tous les participants pour la séance d'inférence lexicale. Dans le courriel, nous avons demandé aux participantes de sexe féminin de ne pas se maquiller les yeux le jour de la passation sans donner d'explication précise.

4.4.2.2 Etape II - Jour de l'inférence lexicale

Le jour de la séance d'inférence lexicale (marqué « J0 » dans la Figure 4.8 plus haut), chaque participant s'est rendu à la plateforme d'oculométrie dans la Maison de la Recherche à l'UT2J. Après avoir signé un consentement, il s'est installé devant l'eye-tracker SMI Red250, à une distance de 55 à 75 cm. Nous avons d'abord présenté au participant les fonctions des dispositifs informatiques et demandé son accord pour l'enregistrement audio de l'entretien.

Après avoir testé la calibration d'oculométrie (les angles entre l'œil du participant, les points réels de calibration affichés sur l'écran, et les points de fixation du regard du participant ne doivent pas être supérieur à $0,5^\circ$) et réglé l'emplacement et la position assise du participant si nécessaire. Nous avons ensuite lancé les stimuli de notre expérience. Nous avons commencé par la formation à la méthode PVH pour aider les participants à comprendre ce que nous attendions d'eux, c'est-à-dire qu'ils nous racontent tout ce qu'ils voient et ce qu'ils pensent et qu'il n'y avait pas de « bonne » ou « mauvaise » réponse. La formation comporte des exercices portant sur

⁶⁴ Certains participants n'ont consenti qu'à un créneau de moins de deux heures.

un calcul mental (adapté de Ericsson, 2006), la description d'une caricature⁶⁵, un exercice d'inférence d'un non mot français intégré dans une phrase et un exercice d'inférence d'un mot chinois intégré dans une phrase. La formation a duré entre 15 et 25 minutes (cf. l'Annexe 2).

A l'issue de la formation, nous avons demandé au participant de lire le texte A en silence et l'avons prévenu qu'à la fin de la lecture nous lui poserions quelques questions de compréhension. Nous avons également expliqué au participant que la surveillance de l'oculométrie peut être perturbée par des gestes brutaux. Nous lui avons alors demandé, dans la mesure du possible, de ne pas quitter l'écran des yeux et d'éviter de basculer la tête pendant la lecture du texte. Après la compréhension en lecture, nous avons demandé au participant d'inférer le sens des mots cibles surlignés en jaune dans le texte, un par un, avec la méthode PVH. A la fin de l'inférence des 10 mots cibles du texte A, le participant a repris la même procédure pour le texte B, de la compréhension du texte jusqu'à l'inférence des 10 mots cibles. Toutes les consignes pour les parties de la lecture et de l'inférence lexicale étaient affichées à l'écran. Nous contrôlions l'avancement des stimuli à l'aide d'un clavier. Nous interrogeons le participant seulement à la fin des lectures de compréhension et le relançons en cas de silence pendant les inférences avec des questions neutres (cf. la Section 4.3.1.2). Parfois nous avons demandé au participant de donner davantage d'explications s'il n'avait pas bien détaillé son raisonnement. Cette partie a duré entre 40 minutes et une heure et quart. Durant toute cette période, nous surveillions la position des yeux du participant sur un moniteur placé à côté. Nous faisons des remarques au participant pour qu'il se repositionne en veillant à interrompre le moins possible ses réflexions. Il nous est arrivé de repositionner nous-même le participant par rapport à l'écran pendant les pauses prévues pour cela.

A la fin de la compréhension et l'inférence lexicale, nous avons invité le participant à s'installer devant un bureau pour travailler sur les exercices post-inférence sous format papier. Le participant a été désigné de façon aléatoire pour le Groupe I ou le Groupe II (cf. l'Annexe 3 3 pour les tâches post-inférence que nous

⁶⁵ Cité de <http://lalettredejaures.over-blog.com/article-combien-d-augmentations-ou-de-creations-de-taxes-depuis-2007-99228045.html> (consulté le 1 juillet 2019).

avons demandé aux participants du Groupe I et du Groupe II d'accomplir). Cependant, si le participant se montrait fatigué ou il ne lui restait pas assez de temps pour accomplir les deux tâches, nous donnions seulement l'exercice lacunaire au participant, et ce dernier était donc classé dans le Groupe I. La durée des tâches post-inférence a varié entre 15 et 45 minutes.

Tous les participants ont été remerciés à la fin de cette séance avec des cadeaux souvenirs de Chine. Nous avons ensuite fixé avec le participant un rendez-vous d'une demi-heure pour la deuxième évaluation de connaissances lexicales (VKS2) deux semaines plus tard, sans expliquer au participant l'objectif du rendez-vous.

4.4.2.3 Etape III – Deuxième évaluation de connaissances lexicales (VKS2)

14 jours après la séance de l'inférence lexicale, nous avons revu chacun des participants et administré le VKS2 comportant seulement les 20 mots cibles (dans un ordre aléatoire) sous format papier.⁶⁶ Il est à noter que deux de nos participants n'ont pas pu se présenter au rendez-vous pour le VKS2. Nous avons ainsi administré le VKS2 à ces deux participants en ligne. Concrètement, le jour du rendez-vous, nous avons envoyé le lien du test VKS2 enregistré dans « Google Forms » au participant à l'heure convenue. Le participant répondait au questionnaire. A la réception des réponses du participant, nous avons invalidé immédiatement son accès au test. Nous avons procédé ainsi pour que le participant ne puisse plus avoir accès aux mots cibles avant la troisième évaluation (VKS3).

A l'issue du VKS2, nous avons pris à nouveau un autre rendez-vous individuel avec chaque participant. Selon le protocole de notre expérience, les participants ne devaient pas connaître en amont l'existence des post tests de vocabulaire (cf. la Section 2.3 concernant l'apprentissage incident). Pour cela, afin de minimiser les éventuels soupçons du participant sur l'objectif de ce rendez-vous, nous avons proposé le rendez-vous sous prétexte d'une enquête sur l'arrière-plan socio-linguistique du participant.

⁶⁶ A cause du confinement mis en place lors de la crise sanitaire due au Covid-19 en mars 2020, nous avons administré le VKS2 auprès d'un participant au bout de 10 jours au lieu de 14 jours afin d'éviter la perte de contact avec le participant pendant le confinement.

4.4.2.4 Etape IV – Troisième évaluation de connaissances lexicales (VKS3)

28 jours après l'inférence lexicale, nous avons à nouveau évalué les connaissances chez nos participants pour les mots cibles par le biais du VKS3 comportant seulement les mots cibles (dans un ordre aléatoire) sous format papier. Ne pouvant se rendre à l'UT2J, une dizaine de nos participants ont accompli le VKS3 à distance. Cette fois-ci nous avons transformé le VKS3 en PDF éditable pour que les participants puissent le compléter directement. Nous avons ensuite envoyé le fichier de VKS3 au participant par mail, en soulignant dans le mail qu'il nous était important de recevoir leurs réponses le jour même.

4.5 Récapitulation

Dans ce chapitre nous avons présenté les participants visés, les outils conçus et les méthodes que nous avons utilisées pour réaliser notre expérience.

Nous avons sollicité 33 apprenants francophones en chinois L2 du niveau intermédiaire. Nous avons sélectionné 20 mots cibles dont la moitié des mots lexicalisés en français et l'autre moitié non lexicalisés et ensuite rédigé deux textes pour intégrer dans chacun 10 mots cibles. Nous avons adopté la consultation d'un dictionnaire et l'exercice lacunaire comme tâches postérieures à l'inférence lexicale.

Nous avons opté pour la méthode de la pensée à voix haute afin de comprendre les sources de connaissance sur lesquelles nos participants s'appuient lorsqu'ils infèrent le sens des mots cibles. Nous avons également décidé d'utiliser l'oculométrie pour observer l'attention que nos participants portent aux mots cibles pendant la lecture pour la compréhension. Afin d'examiner l'évolution des connaissances lexicales chez nos participants pour les mots cibles, nous avons administré trois tests VKS, l'un avant la séance d'inférence lexicale et deux autres après la séance d'inférence lexicale.

Nous présenterons dans les trois chapitres suivants les résultats de notre expérience et les réponses aux questions de recherche.

**CHAPITRE V – SOURCES DE CONNAISSANCE UTILISEES DANS
L'INFERENCE LEXICALE**

Introduction

Dans ce chapitre, nous présenterons les résultats et les analyses des sources de connaissance dont les 33 participants se sont inspirés pour inférer les 20 mots cibles. Nous cherchons à comprendre si la lexicalisation d'un mot chinois inconnu a un effet sur le choix des sources de connaissance qu'un apprenant francophone utilise pour inférer le sens du mot.

Nous présenterons d'abord la typologie des sources de connaissance utilisées pour identifier les indices que nos participants ont repérés, suivie d'exemples extraits des entretiens d'inférence lexicale. Par la suite, nous présenterons l'utilisation des sources de connaissance lors de l'inférence pour les 20 mots cibles et respectivement pour les mots lexicalisés et les mots non lexicalisés. Nous analyserons les résultats au niveau du nombre des sources de connaissance que nos participants utilisent pour inférer un mot cible, également au niveau du poids relatif des sources de connaissance dans l'inférence des 20 mots cibles.⁶⁷ Nous discuterons ensuite quelques cas d'inférence lexicale qui nous ont intéressée plus particulièrement.

Nous concluons ce chapitre par nos réponses à la question de recherche mentionnée plus haut.

5.1 Analyse qualitative

Nous rappelons que la séance d'inférence lexicale s'est déroulée sous forme d'entretien individuel avec la méthode PVH (cf. la Section 4.3.1). Les entretiens ont été tous enregistrés et puis transcrits par un bilingue chinois-français. Toutes les transcriptions ont été vérifiées par un deuxième bilingue chinois-français (cf. le Tome II de la thèse pour les transcriptions).⁶⁸

⁶⁷ Nous avons utilisé le logiciel SPSS pour faire les analyses statistiques des données présentées dans ce chapitre. Nous remercions chaleureusement M. Saïd JMEL pour son aide précieuse sur les analyses statistiques présentées dans ce chapitre.

⁶⁸ Nous remercions sincèrement Mme Qiyuan JI, Mme Lin ZHANG et Mme Wen YU pour leur aide sur les transcriptions des entretiens d'inférence lexicale.

Afin d'identifier les sources de connaissance (« SC » en abrégé) sur lesquelles un apprenant peut s'appuyer pour inférer un mot chinois inconnu rencontré pendant la lecture, nous avons utilisé la typologie de Paribakht & Tréville (2007, cf. le Tableau 1.2, p. 42) adaptée. Plus précisément, un apprenant peut s'appuyer sur ses connaissances générales sur le monde (donc les sources de connaissance non linguistiques, **SNL**) et sur celles linguistiques. Ces dernières peuvent être propres à une langue que l'apprenant connaît, entre autres la L1 (en l'occurrence, la langue française) (donc les sources de connaissance en Ln, **SLn**). Les connaissances linguistiques propres à la L2 (en l'occurrence, la langue chinoise) peuvent se distinguer en fonction de l'emplacement dans un texte, aux niveaux d'un discours (**ND**), d'une phrase (**NP**) ou d'un mot (**NM**). Chacune de ces cinq catégories, présentées ci-dessus, comprend une ou plusieurs sources de connaissance.

Certains indices repérés par nos participants ne correspondent à aucune source de connaissance définie dans la typologie de Paribakht & Tréville (2007). Ces indices sont liés aux particularités de la langue chinoise. Nous avons vu dans le Chapitre III que la langue chinoise est non alphabétique. Un caractère chinois est l'union de la forme, du son et du sens. La majorité des caractères couramment utilisés aujourd'hui sont composés de plusieurs blocs composants qui peuvent délivrer des informations sémantiques et phonétiques. Au niveau du mot, en chinois contemporain, les mots sont majoritairement constitués de deux caractères ou plus qui sont liés d'une façon ou d'une autre. Sémantiquement, les caractères constitutifs d'un mot peuvent avoir un apport plus ou moins important au sens global du mot constitué. Si le sens d'un mot est en rapport étroit avec les sens de ses caractères constitutifs, ce mot est sémantiquement transparent. Dans le cas où l'apprenant connaît le sens des caractères constitutifs, même si l'apprenant n'a jamais vu auparavant ce mot, il parviendrait plus ou moins à comprendre son sens.

Lors de l'inférence des mots cibles, nos participants ont fréquemment repéré les indices dans les aspects présentés ci-dessus. Nous avons donc développé la catégorie NM de la typologie de Paribakht & Tréville (2007). Concrètement, quatre sources de connaissance y ont été ajoutées : le sens d'un caractère constitutif du mot

(SC08), la forme d'un caractère constitutif du mot (SC09), le son d'un caractère constitutif du mot (SC10) et le mode de constitution du mot (SC11).

Au total onze sources de connaissance ont été définies pour identifier les indices repérés par nos participants pour inférer les 20 mots cibles. La typologie utilisée dans notre étude et les définitions des onze sources de connaissance sont présentées dans le Tableau 5.1 plus bas.⁶⁹

5.1.1 Connaissances générales (SC01)

Certains participants ont puisé dans leurs connaissances générales pour déduire le sens d'un mot cible, comme le montre l'exemple suivant.

Ex. (MC14 « 彩虹 ») : *Ça doit sûrement être un « arc-en-ciel », parce qu'en général au printemps quand il pleut et parfois il fait soleil... (il y a un arc-en-ciel.)*

5.1.2 Sources propres à Ln (SC02)

Aucun indice de la SC02 n'a été repéré par nos participants pour inférer les 20 mots cibles. Certains participants ont dit des mots en anglais ou mentionné des figures de style en français lorsqu'ils analysaient les informations dans le texte.

Ex. (MC13 « 明麗 ») : *[S]ur la montagne pour voir le... alors, <((Le participant dit un mot anglais.))> « landscape ». ... Je crois que c'est « paysage », <((Le participant répète le mot anglais.))> « landscape ». Oui, « paysage ».*

Ex. (MC12 « 五颜六色 ») : *<((Le participant lit un mot dans le texte.))> « cao di », c'est « le sol », « le gazon ». ... et que les couleurs changent ... vers le vert. Euh, <((Le participant lit un mot dans le texte.))> « xin yi », c'est peut-être ... « du nouveau vêtement », enfin, dans le sens une nouvelle robe. En fait « la nature prend ... un nouveau vêtement ». Parce que, on dit « un manteau de neige », par exemple, très, très poétique.*

⁶⁹ Les définitions de la SC01 jusqu'à la SC07 sont adaptées de celles de Paribakht & Tréville (2007).

Sources de nature non linguistique (SNL)	
<i>SC01 Connaissances générales</i> : Le lecteur puise dans les connaissances non linguistiques qu'il a en mémoire incluant sa connaissance du thème traité dans le texte.	
Sources propres à la L1 et à une autre langue non L2 (SLn)	
<i>SC02 Sources propres à Ln</i> : Le lecteur puise dans ses connaissances linguistiques pré-acquises dans sa langue maternelle (français) ou d'autres langues non L2 pratiquées pour traiter les informations perçues.	
Sources propres à la L2	Niveau du discours (ND)
	<i>SC03 Contexte lointain</i> : Le lecteur s'appuie sur le sens général du texte, du paragraphe ou des phrases qui sont avant ou après la phrase contenant le mot inconnu.
	Niveau de la phrase du mot (NP)
	<i>SC04 Ponctuation</i> : Le lecteur s'appuie sur ses connaissances des règles de ponctuation et de leur signification.
	<i>SC05 Contexte proche</i> : Le lecteur s'appuie sur le sens général de la phrase contenant le mot.
	<i>SC06 Grammaire</i> : Le lecteur s'appuie sur les propriétés syntaxiques du mot, sur l'ordre des mots ou sur les catégories grammaticales.
	Niveau du mot (NM)
	<i>SC07 Association/Collocation</i> : Le lecteur fait l'association entre le mot et un autre mot qui lui est familier ou le lecteur connaît une expression de L2 dont il croit que le mot fait partie.
	<i>SC08 Sens d'un caractère constitutif du mot</i> : Le lecteur s'appuie sur le sens d'un caractère contenu dans un mot pour deviner le sens du dernier.
	<i>SC09 Forme d'un caractère constitutif du mot</i> : Le lecteur cherche à identifier un caractère contenu dans un mot à partir d'un bloc composant du caractère, ou bien puise dans la signification d'un bloc composant du caractère pour deviner le sens du caractère et du mot.
	<i>SC10 Son d'un caractère constitutif du mot</i> : Le lecteur cherche à identifier la prononciation d'un caractère contenu dans un mot par le biais d'un bloc composant du caractère.
<i>SC11 Mode de constitution du mot</i> : Le lecteur puise dans la façon dont est constitué un mot ou dans la relation entre les caractères constitutifs.	

Tableau 5.1 – Typologie des sources de connaissance utilisées pour inférer un mot chinois

Le participant du premier exemple parle français dans sa famille, mais il a grandi à l'étranger et suivi un parcours scolaire en anglais dans son enfance. Lors d'une inférence, il voulait proposer « paysage » pour le sens du mot cible concerné, mais le premier mot qui lui est venu à l'esprit est le mot anglais *landscape* (paysage). Il a ensuite retrouvé le mot français. Dans le deuxième exemple, le participant a associé une métaphore dans le contexte à une figure de style descriptive similaire en français, *un manteau de neige*.

Nous ne comptons pas de tels cas comme l'utilisation des connaissances linguistiques non L2 pour inférer un mot chinois inconnu.

5.1.3 Contexte lointain (SC03)

Dans l'exemple ci-dessous, le participant se rappelle le début du texte pour mieux comprendre le contexte du mot MC09 qui se trouve à la fin du texte.

Ex. (MC09 « 乏味 ») : *On parle de la situation générale de la Chine ... au début du texte ... Le narrateur ... compare les faits actuels, aujourd'hui, à son enfance.*

5.1.4 Ponctuation (SC04)

Dans l'exemple suivant, le participant repère un deux-points dans le contexte du mot MC07 qui, pour lui, introduit une explication pour l'élément devant le signe de ponctuation.

Ex. (MC07 « 还礼 ») : *Déjà, on a une explicitation juste après, puisque on a (un) deux-points.*

5.1.5 Contexte proche (SC05)

Les participants s'inspiraient souvent du sens global de la phrase comportant un mot cible pour déduire le sens du mot, comme le montre l'exemple suivant.

Ex. (MC19 « 飘零 ») : *[S]i je relis la phrase, ça parle encore de l'herbe et des feuilles ou des arbres. ... Oui. Là c'est l'« automne ». Euh... donc on parle de l'automne et des feuilles.*

5.1.6 Grammaire (SC06)

Dans les deux exemples suivants, les participants ont essayé de comprendre le rôle d'un mot cible dans une phrase ou la propriété grammaticale d'un mot cible par rapport à ses mots voisins.

Ex. (MC19 « 飘零 ») : *[J]e dirais plutôt un verbe avec le sujet, donc les feuilles des arbres. Les feuilles des arbres, peut-être qu'ils tournent ou avec du vent qu'ils tournent.*

Ex. (MC11 « 安适 ») : *[J]e pense que c'est un adjectif parce qu'il y a <((Le participant lit un mot dans le texte qui signifie « très »))> « hen » devant et y a <((Le participant lit un mot dans le texte qui est une particule.))> « de » derrière. Du coup c'est un qualifiant par rapport à <((Le participant lit un mot dans le texte.))> « di fang » qui est un « endroit ».*

5.1.7 Association/Collocation (SC07)

Voici deux exemples de l'utilisation de la source de connaissance « Association/Collocation ».

Ex. (MC13 « 明丽 ») : *<((Le participant lit le mot cible.))> « ming li », je me demande si je l'ai pas déjà vu. Il me semble que c'est un mot qui veut dire « joli ». Dans cet exemple, le mot cible MC13 « 明丽 (mínglì, clair et joli) » fait que le participant pense au mot « 美丽 (měilì, joli) ».*

Ex. (MC05 « 汤圆 ») : *Hm, ah, peut-être c'est quelque chose à boire. « tang », c'est la « soupe ». Mais c'est pas ça. Euh, ah oui, en même temps c'est pas <((Le participant dit un mot qui signifie « boire ».))> « he ». Donc c'est plutôt quelque chose à manger.*

En chinois, le verbe en collocation avec le nom « 汤 (tāng, soupe) » est « 喝 (hē, boire) ». Dans cet exemple ci-dessous, le participant a d'abord reconnu le premier caractère « 汤 (tāng, soupe, bouillon) » dans le mot cible MC05 « 汤圆 (tāngyuán, boulette de riz glutineux farcie de viande (fourrée à la confiture)) ». Il a ensuite remarqué que le verbe précédant MC05 n'est pas « 喝 (hē, boire) », mais

« 吃 (*chī*, manger) ». Cela a dissuadé le participant de s'orienter vers l'idée de « quelque chose de liquide » pour le sens du mot MC05.

5.1.8 Sens d'un caractère constitutif du mot (SC08)

Dans le chinois contemporain, le sens global d'un mot est souvent étroitement lié aux sens individuels des caractères constitutifs (cf. la Section 3.2.3.2.1). Dans l'exemple suivant, le participant a reconnu le premier caractère 窗 (fenêtre) du mot MC02. Il en déduit que le sens de MC02 a un rapport avec les fenêtres.

Ex. (MC02 « 窗花 ») : *Euh je pense que c'est quelque chose qui s'accroche à la fenêtre, puisqu'il me semble que le premier caractère, c'est pour « fenêtre ».*

5.1.9 Forme d'un caractère constitutif du mot (SC09)

En face d'un caractère inconnu, les participants se mettent souvent à décomposer le caractère afin d'en récupérer des informations lexicales (cf. la Section 3.2.2).

Ex. (MC19 « 飘零 ») : *[L]e deuxième caractère, tout en haut, ça fait penser à la « neige » ou, du moins, à la « pluie ». Et en bas, ça me faisait penser à <<(Le participant dit un caractère chinois)>> « leng », « froid ».*

Le caractère « 零 » (du mot MC19 « 飘零 (*piāolíng*, tomber en tourbillonnant) ») contient deux blocs composants, « 雨 » en haut et de « 令 » en bas. « 雨 » évoque au participant les caractères « 雪 (neige) » et « 雨 (pluie) » et « 令 » au caractère « 冷 (*lěng*, froid) ».

Ex. (MC03 « 送灶 ») : *Le deuxième caractère, c'est le « feu » avec la « terre » ... Des plats à emporter peut-être. Le « feu », peut-être le feu pour cuisiner, pour réchauffer les plats.*

Le participant a reconnu le bloc composant « 火 (feu) » dans « 灶 », le deuxième caractère de MC03 « 送灶 (*sòngzào*, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune) ». Cela donne l'idée de « plat (cuisiné) » au participant.

5.1.10 Son d'un caractère constitutif du mot (SC10)

Malgré l'opacité graphie-son des caractères chinois (cf. la Section 3.2.2), les participants ont parfois essayé de prononcer les caractères inconnus, comme le montre l'exemple suivant.

Ex. (MC03 « 送灶 ») : *[L]a clé du feu. ... Pourquoi pas quelque chose qui peut se prononcer en « du » ou en « tu » ?*

Le participant a reconnu le bloc composant « 土 (tǔ) » dans « 灶(zào) », le deuxième caractère de MC03 « 送灶 (sòngzào, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune) ». Il pense que la prononciation du bloc composant « 土 (tǔ) » peut être transposée au caractère « 灶 ». Comme un autre caractère « 肚 (dù) » contient également le bloc composant « 土 (tǔ) », il pense qu'il est aussi possible que le caractère « 灶 » se prononce « dù ».

5.1.11 Mode de constitution du mot (SC11)

Dans le chinois contemporain, les caractères constitutifs d'un mot sont liés l'un à l'autre d'une telle ou telle façon (cf. la Section 3.2.3.2.2). Voici trois exemples dans lesquels les participants ont fait attention à la façon dont est constitué un mot chinois inconnu.

Ex. (MC20 « 白茫茫 ») : *[Y] a la répétition des caractères. Et pour moi, quand il y a la répétition des caractères, c'est un peu si qu'on mettait « hen » devant un adjectif. C'est pour renforcer la chose.*

Le participant analyse le mot MC20 « 白茫茫 (báimángmáng, d'une blancheur éclatante) » dans lequel le caractère « 茫 » se répète. Le participant pense donc que la répétition représente un effet de renforcement.

Ex. (MC12 « 五颜六色 ») : *Et donc pour moi, ça serait un <((Le participant dit un mot chinois.))> « cheng yu » en quatre caractères du coup, une expression figée euh pour décrire les, les couleurs des fleurs.*

Le participant infère le mot MC12 « 五颜六色 (wǔyánliùsè, de diverses couleurs, multicolore) », un mot constitué de quatre caractères. La grande majorité des mots

chinois dans le chinois contemporain contiennent deux caractères. Le fait que MC12 contient quatre caractères fait penser au participant que ce mot est peut-être un « 成语 (*chéngyǔ*) », une sorte de tournure idiomatique, souvent traduit comme « proverbe », parce que les « 成语 (*chéngyǔ*) » sont majoritairement constitués de quatre caractères⁷⁰.

Ex. (MC02 « 窗花 ») : *[L]e deuxième c'est le nom et le premier c'est un qualifiant par rapport au deuxième caractère qui est pour moi « fleur ».*

Le participant analyse le mot MC02 « 窗花 (*chuānghuā*, découpage en papier pour décorer les fenêtres) ». Il pense que le premier caractère « 窗 (*chuāng*, fenêtre) », qui lui est inconnu, est l'attribut du deuxième caractère « 花 (*huā*, fleur) » qui est un nom.

5.2 Analyse quantitative

Nous avons observé au total 654 inférences lexicales (33 participants x 20 mots cibles = 660 observations – 6 observations = 654 observations). Nous avons exclu 6 cas parce que, d'après les réponses de nos participants aux trois tests VKS et pendant l'entretien d'inférence individuel, des participants connaissent certains mots cibles : le participant P15 connaît le mot cible MC12⁷¹ ; P22 connaît MC14⁷² ; P33 connaît MC08⁷³ et MC14⁷⁴ ; P31 et P32 ont respectivement noté niveau 1 et niveau 4 pour MC14 au premier test VKS, mais ces deux participants ont affirmé pendant la séance d'inférence qu'ils connaissaient ce mot. Ces deux participants ont également donné le sens correct du mot MC14 au deuxième test et au troisième test VKS.

C'est à partir de ces 654 inférences que nous avons relevé les indices sur lesquels les participants se sont basés pour comprendre le sens des mots inconnus (cf.

⁷⁰ Par exemple, « 井底之蛙 (*jǐngdǐzhīwā*) », littéralement « une grenouille au fond d'un puits (qui ne voit que la partie du ciel délimité par l'entrée du puits et qui croit donc que le ciel est tout petit) », signifie une personne avec une vue limitée.

⁷¹ MC12 « 五颜六色 (*wǔyánliùsè*, de diverses couleurs, multicolore) »

⁷² MC14 « 彩虹 (*cǎihóng*, arc-en-ciel) »

⁷³ MC08 « 在乎 (*zàihu*, prendre qch. à cœur ; faire attention à) »

⁷⁴ MC14 « 彩虹 (*cǎihóng*, arc-en-ciel) »

la Partie 1 du Volume II de cette thèse pour les transcriptions des entretiens d'inférence lexicale avec la méthode PVH).

Il est à noter que, si un participant repère deux indices de la même source de connaissance, par exemple deux indices grammaticaux, pour inférer un mot inconnu, une seule utilisation de la grammaire par le participant est comptée pour l'inférence de ce mot.

Au total, 1560 utilisations des sources de connaissance ont été identifiées pour les 654 inférences observées.

5.2.1 Diversité des sources de connaissance utilisées

Souvent les participants se sont appuyés sur plus d'une source de connaissance pour inférer un mot donné. Nous appelons le nombre total des sources de connaissance utilisées pour inférer un mot inconnu « diversité des sources de connaissance utilisées », ou plus simplement « diversité SC ».

Voici un exemple qui montre le raisonnement d'un participant pendant l'inférence du mot MC02 « 窗花 *chuānghuā*, papier découpé pour la décoration d'une fenêtre »).

Ex. : <((Le participant lit la phrase contenant le mot cible.))> « hen duo ren hui zuo yi xie rang jia li geng piao liang ». Ah le « hua », c'est « fleur ». Donc est-ce que ce serait des ... « couronnes de fleurs » ? Des « bouquets » ? ... Oui ça peut être ça, parce que beaucoup de gens vont utiliser ... des cartes rouges ? ... <((Le participant relit la phrase contenant le mot cible.))> « zuo yi xie » hm hm « hua » Je pense à des couronnes de fleurs ou quelque chose comme ça. En tout cas, c'est un nom parce qu'il y a <((Le participant lit quelque chose dans le texte.))> « yi xie » devant, « plusieurs », « plusieurs » quelque chose. Et puis il y a « faire », <((Le participant lit quelque chose dans le texte.))> « zuo » quelque chose. C'est un COD.

Ainsi, pour inférer le mot MC02 « 窗花 *chuānghuā*, papier découpé pour la décoration d'une fenêtre »), le participant a utilisé trois sources de connaissance :

- 1) le sens d'un caractère constitutif : *Ah le « hua », c'est « fleur ». Donc est-ce que ce serait des, des « couronnes de fleurs » ? Des « bouquets » ?*
- 2) le contexte proche : *Oui ça peut être ça, parce que euh beaucoup de gens vont utiliser ... des cartes rouges ? ... <((Le participant relit la phrase contenant le mot cible.))> « zuo yi xie » hm hm « hua » ... je sais pas. Je pense à des couronnes de fleurs ou quelque chose comme ça.*
- 3) la grammaire : *En tout cas c'est un nom parce qu'il y a <((Le participant lit quelque chose dans le texte.))> « yi xie » devant, « plusieurs », « plusieurs » quelque chose. Et puis il y a « faire », <((Le participant lit quelque chose dans le texte.))> « zuo » quelque chose. C'est un COD.*

Au total 1560 utilisations des sources de connaissance ont été identifiées pour 654 inférences de mot, soit en moyenne 2,4 sources de connaissance utilisées par inférence. Autrement dit, pour inférer un mot cible, en moyenne les participants se sont appuyés sur plus de 2 sources de connaissance différentes.

Le nombre d'inférences, classé selon la diversité SC, est présenté dans le Tableau 5.2. Parmi les 33 participants, 23 participants ont employé en moyenne plus de 2 sources de connaissance pour inférer tous les mots cibles, soit 70%. Les inférences dans lesquelles nos participants ont utilisé plus de 2 sources de connaissance représentent 82,1% des 654 inférences de mot observées.

Dans 8 inférences observées, les participants n'ont repéré aucun indice.

Diversité SC	Nb d'inférences	/654
0	8	1,2%
1	109	16,7%
2	274	41,9%
3	170	26,0%
4	74	11,3%
5	17	2,6%
6	2	0,3%
Total	654	100%

Tableau 5.2 – Nombre d'inférences classé selon la diversité des sources de connaissance utilisées

La Figure 5.1 ci-dessous présente les diversités SC en moyenne classées par mot. Selon la Figure 5.1, les trois mots pour lesquels les participants ont utilisé le moins de sources de connaissance sont MC06 « 雀跃 » (1,94 sources de connaissance), MC08 « 在乎 » (1,97 source de connaissance) et MC15 « 踏青 » (2,03 sources de connaissance).

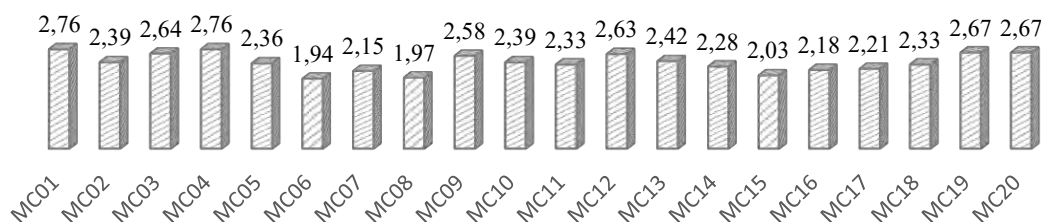


Figure 5.1 – Diversité en moyenne des sources de connaissance utilisées par mot

MC01	节俭	(<i>jiéjiǎn</i>) : économe, frugal
MC02	窗花	(<i>chuānghuā</i>) : papier découpé pour la décoration d'une fenêtre
MC03	送灶	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC04	腊月	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC05	汤圆	(<i>tāngyuán</i>) : boulette de riz glutineux farcie de viande (fourrée à la confiture)
MC06	雀跃	(<i>quèyuè</i>) : sauter de joie
MC07	还礼	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC08	在乎	(<i>zàihu</i>) : prendre qch. à cœur ; faire attention à
MC09	乏味	(<i>fáwèi</i>) : fade, insipide
MC10	盼望	(<i>pànwàng</i>) : espérer, désirer
MC11	安适	(<i>ānshì</i>) : tranquille et confortable
MC12	五颜六色	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC13	明丽	(<i>mínglì</i>) : clair et beau
MC14	彩虹	(<i>cǎihóng</i>) : arc-en-ciel
MC15	踏青	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps
MC16	风筝	(<i>fēngzhēng</i>) : cerf-volant
MC17	蝉	(<i>chán</i>) : cigale
MC18	闲聊	(<i>xiánliáo</i>) : bavarder
MC19	飘零	(<i>piāolíng</i>) : tomber en voltigeant
MC20	白茫茫	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

Parmi les 654 inférences lexicales observées, 324 inférences ont porté sur les mots lexicalisés et 330 sont sur les mots non lexicalisés. En moyenne 2,33 sources de

connaissance ont été utilisées pour inférer les mots lexicalisés et 2,44 sources de connaissance pour les mots non lexicalisés (Tableau 5.3).

	Nb d'inférences	Nb. d'utilisation des sources de connaiss.	Diversité SC en moyenne	Ecart-type
Mots L	324	754	2,3	1,0
Mots NL	330	806	2,4	1,1
Total	654	1560	2,4	1,0

Tableau 5.3 – Diversité des sources de connaissance utilisées pour inférer les mots lexicalisés (L) et non lexicalisés (NL)

Le nombre d'inférences pour les mots lexicalisés et pour les mots non lexicalisés, pour lesquelles les participants ont utilisé de zéro à six sources de connaissance, est montré dans la Figure 5.2.

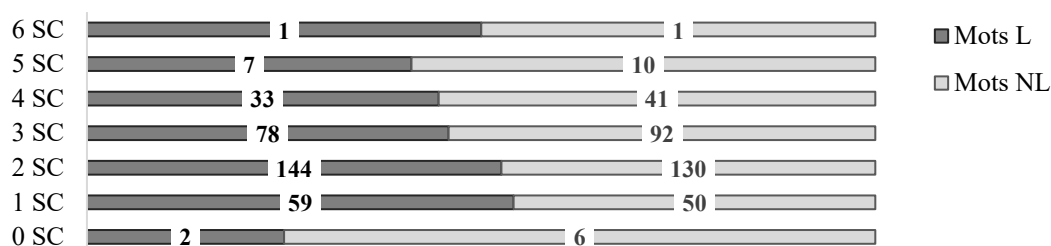


Figure 5.2 – Nombre d'inférences pour les mots lexicalisés (L) et non lexicalisés (NL), classées selon la diversité des sources de connaissance utilisées

Selon l'analyse univariée par modèle linéaire général⁷⁵, la différence de la diversité SC entre les deux catégories n'est pas significative ($p > 0,05$).

Variable dépendante : DiversitéSC

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	2,172a	1	2,172	2,044	,153
Ordonnée à l'origine	3719,138	1	3719,138	3500,482	,000
Lexicalisation	2,172	1	2,172	2,044	,153
Erreur	692,727	652	1,062		
Total	4416,000	654			
Total corrigé	694,899	653			

a. R deux = ,003 (R deux ajusté = ,002)

⁷⁵ Le modèle linéaire général univarié est un modèle de régression qui fournit une analyse de variance d'une variable dépendante par un facteur, en l'occurrence « Diversité SC » par « Lexicalisation » (cf. <https://www.ibm.com/docs/fr/spss-statistics/24.0.0?topic=option-glm-univariate-analysis>, consulté le 9 novembre 2020).

5.2.2 Fréquence d'utilisation des sources de connaissance

La fréquence d'utilisation d'une source de connaissance, désormais « fréquence SC », renvoie au nombre de fois où les participants ont utilisé une source de connaissance spécifique dans les 654 inférences lexicales observées.

La fréquence SC pour les onze sources de connaissance et les cinq catégories de source de connaissance est présentée dans le Tableau 5.4.

Sources de nature non linguistique (SNL)		75
	<i>SC01 Connaissances générales</i>	75
Sources propres à la L1 et à une autre langue non L2 (SLn)		0
	<i>SC02 Sources propres à Ln</i>	0
Sources propres à la L2	Niveau du discours (ND)	49
	<i>SC03 Contexte lointain</i>	49
	Niveau de la phrase du mot (NP)	854
	<i>SC04 Ponctuation</i>	4
	<i>SC05 Contexte proche</i>	575
	<i>SC06 Grammaire</i>	275
	Niveau du mot (NM)	582
	<i>SC07 Association/Collocation</i>	17
	<i>SC08 Sens d'un caractère constitutif du mot</i>	304
	<i>SC09 Forme d'un caractère constitutif du mot</i>	219
	<i>SC10 Son d'un caractère constitutif du mot</i>	15
<i>SC11 Constitution du mot</i>	27	
Total		1560

Tableau 5.4 – Fréquence d'utilisation des sources de connaissance pendant l'inférence lexicale

La Figure 5.3 ci-dessous présente le poids relatif des cinq catégories de source de connaissance dans les 654 inférences observées. Nous pouvons voir que les participants se sont appuyés le plus sur les sources de connaissance phrastiques propres à la L2 (NP).

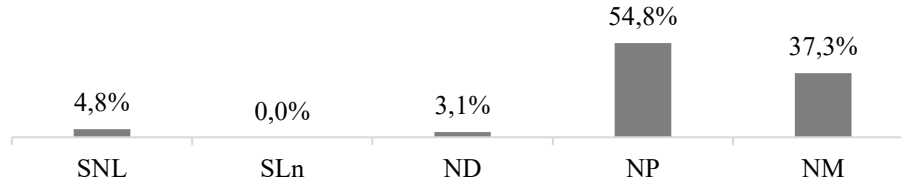


Figure 5.3 – Poids relatif des catégories de source de connaissance pendant l’inférence lexicale

Par ordre décroissant, le poids d’utilisation des cinq catégories est classé ainsi :

$$NP > NM > SNL > ND > SLn$$

Parmi les onze sources de connaissance, les participants se sont inspirés le plus du contexte proche (SC05) pour inférer les mots cibles. Les trois autres sources de connaissance importantes sont les connaissances grammaticales (SC06), le sens d’un caractère constitutif du mot (SC08) et la forme d’un caractère constitutif du mot (SC09). Les poids d’utilisation des sources de connaissance dans l’inférence des mots cibles sont présentés dans la Figure 5.4.

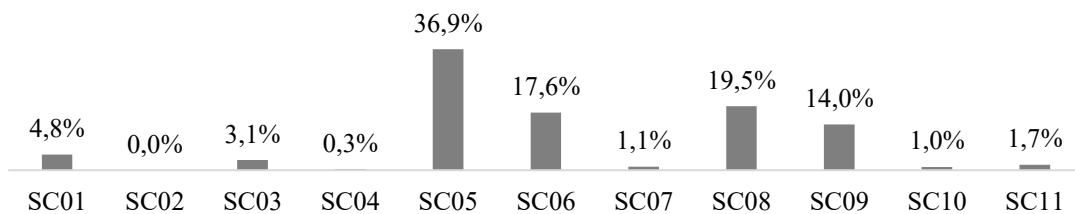


Figure 5.4 – Poids relatif des sources de connaissance pendant l’inférence lexicale

Par ordre décroissant, les quatre sources de connaissance les plus importantes sont classées ainsi selon leur poids d’utilisation :

$$\textit{Contexte proche (SC05)} > \textit{Sens d’un caractère constitutif du mot (SC08)} > \textit{Grammaire (SC06)} > \textit{Forme d’un caractère constitutif du mot (SC09)}$$

Les Figures 5.5, 5.6, 5.7 et 5.8 ci-dessous montrent respectivement la fréquence d’utilisation par mot des quatre sources de connaissance majeures, SC05, SC06, SC08 et SC09.

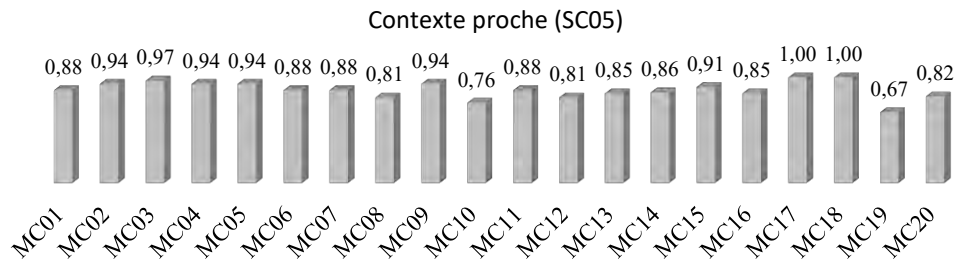


Figure 5.5 – Fréquence d'utilisation de la SC05 pendant l'inférence des mots cibles

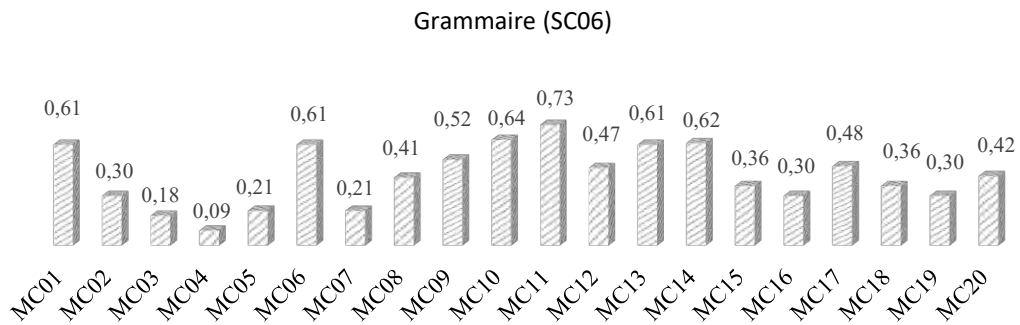


Figure 5.6 – Fréquence d'utilisation de la SC06 pendant l'inférence des mots cibles

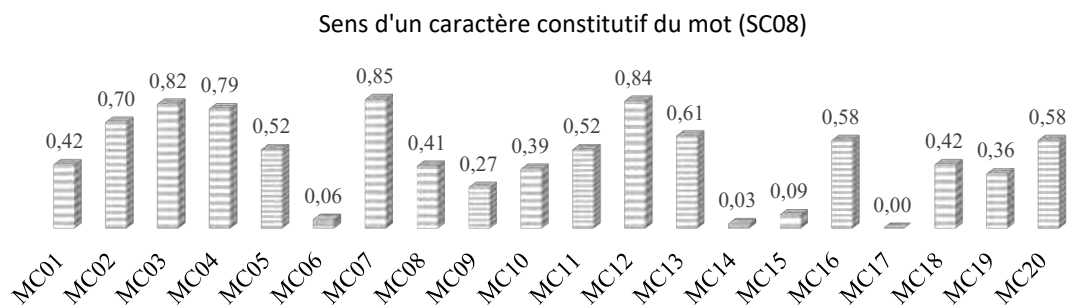


Figure 5.7 – Fréquence d'utilisation de la SC08 pendant l'inférence des mots cibles

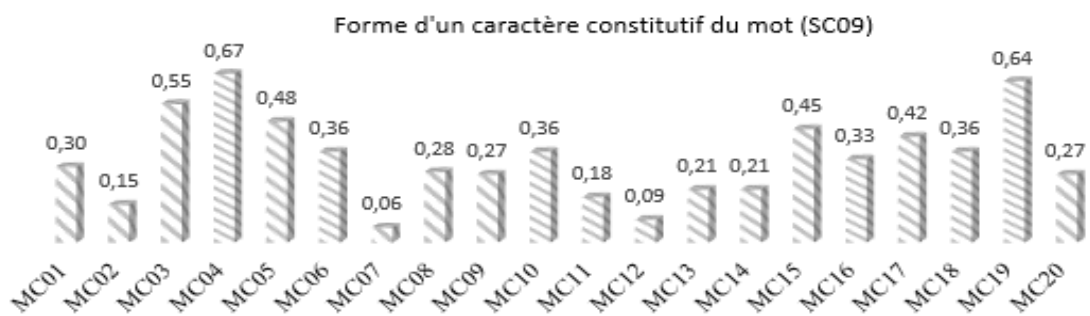


Figure 5.8 – Fréquence d'utilisation de la SC09 pendant l'inférence des mots cibles

5.2.2.1 Fréquence d'utilisation des sources de connaissance pour inférer les mots lexicalisés et non lexicalisés

La fréquence d'utilisation des sources de connaissance employées pour inférer les mots lexicalisés et non lexicalisés est montrée dans le Tableau 5.5 ci-dessous.

		Mots L	Mots NL	Total
Sources de nature non linguistique (SNL)		34	41	75
	<i>SC01 Connaissances générales</i>	<i>34</i>	<i>41</i>	
Sources propres à L1 et à une autre langue non L2 (SLn)		0	0	0
	<i>SC02 Sources propres à Ln</i>	<i>0</i>	<i>0</i>	
Sources propres à la L2	Niveau du discours (ND)	32	17	49
	<i>SC03 Contexte lointain</i>	<i>32</i>	<i>17</i>	
	Niveau de la phrase du mot (NP)	449	405	854
	<i>SC04 Ponctuation</i>	<i>2</i>	<i>2</i>	
	<i>SC05 Contexte proche</i>	<i>285</i>	<i>290</i>	
	<i>SC06 Grammaire</i>	<i>162</i>	<i>113</i>	
	Niveau du mot (NM)	239	345	582
	<i>SC07 Association/Collocation</i>	<i>8</i>	<i>9</i>	
	<i>SC08 Sens d'un caractère constitutif</i>	<i>112</i>	<i>192</i>	
	<i>SC09 Forme d'un caractère</i>	<i>98</i>	<i>121</i>	
	<i>SC10 Son d'un caractère constitutif</i>	<i>9</i>	<i>6</i>	
<i>SC11 Mode de constitution du mot</i>	<i>12</i>	<i>15</i>		
Total		754	806	1560

Tableau 5.5 – Fréquence d'utilisation des sources de connaissance pendant l'inférence des mots lexicalisés (L) et non lexicalisés (NL)

Les poids relatifs des catégories pour inférer les mots lexicalisés et non lexicalisés sont présentés dans la Figure 5.9 ci-dessous. Nous pouvons voir que les participants s'appuient davantage sur les sources de connaissance de la catégorie NP pour inférer les mots lexicalisés (59,5%), par rapport aux mots non lexicalisés (50,2%) ; mais ils recourent plus aux sources de connaissance de la catégorie NM pour inférer les mots non lexicalisés (42,6%), par rapport aux mots lexicalisés (31,7%).

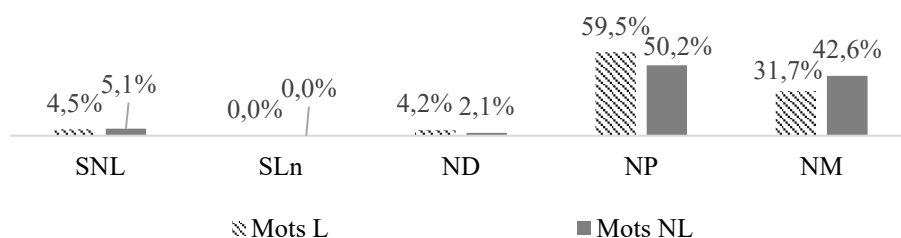


Figure 5.9 – Poids relatif des catégories de sources de connaissance pendant l'inférence des mots lexicalisés (L) et non lexicalisés (NL)

Les catégories sont classées ainsi en fonction de leur poids d'utilisation pour inférer les mots lexicalisés et non lexicalisés :

Mots lexicalisés : NP > NM > SNL > ND > SLn

Mots non lexicalisés : NP > NM > SNL > ND > SLn

Les poids des sources de connaissance utilisées pour inférer les mots lexicalisés et non lexicalisés sont présentés dans la Figure 5.10 ci-dessous. Les sources de connaissance les plus importantes que les participants ont utilisées pour inférer les mots lexicalisés et non lexicalisés sont le contexte proche (SC05), les connaissances grammaticales (SC06), le sens et la forme d'un caractère constitutif du mot (respectivement SC08 et SC09).

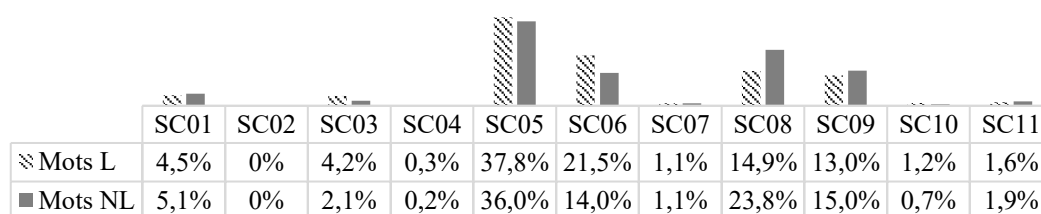


Figure 5.10 – Poids relatifs des sources de connaissance utilisées pendant l'inférence des mots lexicalisés (L) et non lexicalisés (NL)

Parmi toutes les sources de connaissance, le contexte proche (SC05) est la piste principale pour inférer les mots lexicalisés (37,8%) et pour les mots non lexicalisés (36%). Les différences se trouvent principalement au niveau de l'utilisation du sens d'un caractère constitutif du mot (SC08) (14,9% pour les mots lexicalisés vs 23,8% pour les mots non lexicalisés) et de la grammaire (SC06) (21,5% pour les mots lexicalisés vs 14% pour les mots non lexicalisés).

Mots lexicalisés :

Contexte proche (SC05) > Grammaire (SC06) > Sens d'un caractère constitutif du mot (SC08) > Forme d'un caractère constitutif du mot (SC09)

Mots non lexicalisés :

Contexte proche (SC05) > Sens d'un caractère constitutif du mot (SC08) > Forme d'un caractère constitutif du mot (SC09) > Grammaire (SC06)

Selon les analyses univariées par modèle linéaire général, parmi ces quatre sources de connaissance majeures, la lexicalisation n'a pas d'effet sur l'utilisation du contexte proche (SC05) ($p > 0,05$).

Variable dépendante : SC05

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	,000a	1	,000	,001	,974
Ordonnée à l'origine	505,505	1	505,505	4745,220	,000
Lexicalisation	,000	1	,000	,001	,974
Erreur	69,457	652	,107		
Total	575,000	654			
Total corrigé	69,457	653			

a. R deux = ,000 (R deux ajusté = -,002)

La différence entre les mots lexicalisés et non lexicalisés en termes de l'utilisation de la forme des caractères constitutifs (SC09) n'est pas statistiquement significative ($p > 0,05$), pourtant cette différence est assez importante (98 utilisations dans les 324 inférences pour les mots lexicalisés et 121 utilisations dans les 330 inférences pour les mots non lexicalisés).

Variable dépendante : SC09

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	,674a	1	,674	3,030	,082
Ordonnée à l'origine	73,200	1	73,200	329,166	,000
Lexicalisation	,674	1	,674	3,030	,082
Erreur	144,991	652	,222		
Total	219,000	654			
Total corrigé	145,665	653			

a. R deux = ,005 (R deux ajusté = ,003)

En revanche, pour inférer les mots lexicalisés et non lexicalisés, la différence au niveau de l'utilisation de la grammaire (SC06) est significative ($p = 0,00$). Également, la différence dans l'utilisation du sens d'un caractère constitutif du mot (SC08) est significative ($p = 0,00$).

Variable dépendante : SC06

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	4,059a	1	4,059	17,042	,000
Ordonnée à l'origine	116,023	1	116,023	487,082	,000
Lexicalisation	4,059	1	4,059	17,042	,000
Erreur	155,306	652	,238		
Total	275,000	654			
Total corrigé	159,365	653			

a. R deux = ,025 (R deux ajusté = ,024)

Variable dépendante : SC08

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	9,116a	1	9,116	38,703	,000
Ordonnée à l'origine	140,639	1	140,639	597,082	,000
Lexicalisation	9,116	1	9,116	38,703	,000
Erreur	153,575	652	,236		
Total	304,000	654			
Total corrigé	162,691	653			

a. R deux = ,056 (R deux ajusté = ,055)

La Figure 5.11 montre la fréquence d'utilisation en moyenne de la grammaire (SC06) et du sens d'un caractère constitutif du mot (SC08) pour inférer les 20 mots cibles, classée selon la lexicalisation.

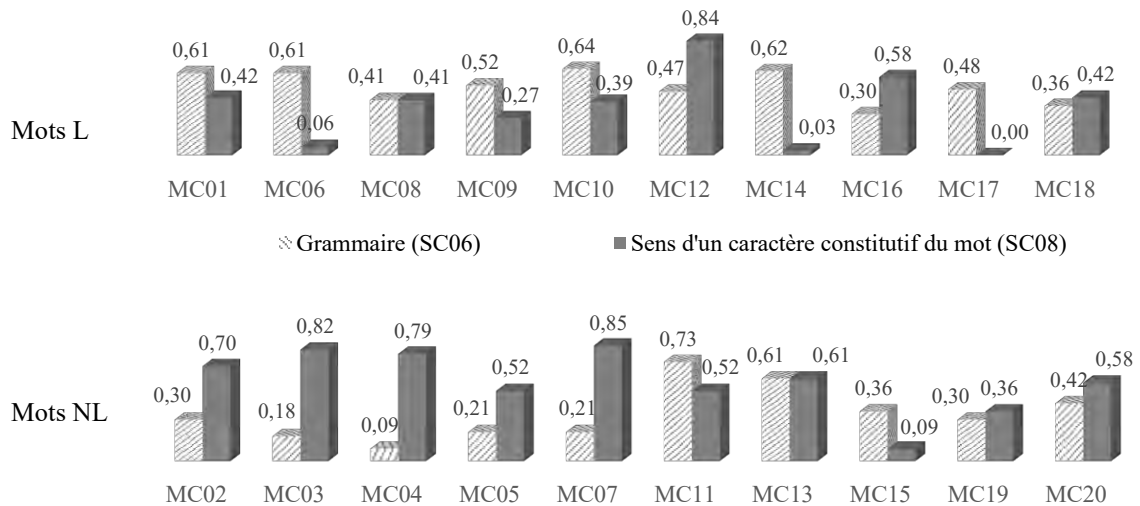


Figure 5.11 – Fréquence d'utilisation de la grammaire (SC06) et du sens d'un caractère constitutif du mot (SC08) dans l'inférence des mots lexicalisés (L) et non lexicalisés (NL)

MC01	节俭	L	(<i>jiéjiǎn</i>) : économe, frugal
MC06	雀跃	L	(<i>quèyuè</i>) : sauter de joie
MC08	在乎	L	(<i>zàihu</i>) : prendre qch. à cœur ; faire attention à

MC09	乏味	L	(<i>fáwèi</i>) : fade, insipide
MC10	盼望	L	(<i>pànwàng</i>) : espérer, désirer
MC12	五颜六色	L	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC14	彩虹	L	(<i>cǎihóng</i>) : arc-en-ciel
MC16	风筝	L	(<i>fēngzhēng</i>) : cerf-volant
MC17	蝉	L	(<i>chán</i>) : cigale
MC18	闲聊	L	(<i>xiánliáo</i>) : bavarder
MC02	窗花	NL	(<i>chuānghuā</i>) : papier découpé pour la décoration d'une fenêtre
MC03	送灶	NL	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC04	腊月	NL	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC05	汤圆	NL	(<i>tāngyuán</i>) : boulette de riz glutineux farcie de viande (fourrée à la confiture)
MC07	还礼	NL	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC11	安适	NL	(<i>ānshì</i>) : tranquille et confortable
MC13	明丽	NL	(<i>mínglì</i>) : clair et beau
MC15	踏青	NL	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps
MC19	飘零	NL	(<i>piāolíng</i>) : tomber en voltigeant
MC20	白茫茫	NL	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

Le poids du contexte lointain (SC03) dans l'inférence pour les mots lexicalisés et celui pour les mots non lexicalisés sont significativement différents ($p < 0,05$).

Variable dépendante : SC03

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	,365a	1	,365	5,293	,022
Ordonnée à l'origine	3,692	1	3,692	53,539	,000
Lexicalisation	,365	1	,365	5,293	,022
Erreur	44,964	652	,069		
Total	49,000	654			
Total corrigé	45,329	653			

a. R deux = ,008 (R deux ajusté = ,007)

Nous pensons que cela est liée à l'inférence de MC09 « 乏味 (*fáwèi*, insipide, fade) », un mot lexicalisé. Parmi les 49 utilisations de la SC03, 14 utilisations, soit 28,6%, sont pour inférer le mot MC09. La phrase contenant le mot est « 但大家都说过年越来越乏味了。Mais tout le monde dit que la Fête du Nouvel an devient de plus en plus MC09. » Cette phrase ne fournit pas suffisamment d'informations pour que les participants puissent comprendre le mot. Il leur est donc nécessaire de lire les phrases

voisines afin d'obtenir des informations complémentaires. Par exemple, la phrase suivante compare la situation à l'enfance du narrateur et la situation actuelle – « 以前盼望着过年穿新衣服、吃好吃的、拿红包，那是一种快乐，现在都没有了。C'était une joie d'attendre (avec impatience) la Fête pour avoir de nouveaux habits, se régaler et recevoir des *hongbao*. Tout s'est perdu maintenant. » Pour cette raison-là, nous ne pensons pas que la lexicalisation ait causé la différence d'utilisation du contexte lointain.

5.2.2.2 Fréquence d'utilisation des sources de connaissance pour inférer les mots de différentes classes grammaticales

Dans le Tableau 5.6 sont présentées les fréquences SC pour inférer les mots cibles des trois classes grammaticales.

		Adj /505	Nom /459	Verbe /596	Total /1560
Sources de nature non linguistique (SNL)		12 2,4%	41 8,9%	22 3,7%	75 4,8%
	<i>SC01 Connaissances générales</i>	12 2,4%	41 8,9%	22 3,7%	75 4,8%
Sources propres à L1 et à une autre langue non L2 (SLn)		0 0,0%	0 0,0%	0 0,0%	0 0%
	<i>SC02 Sources propres à Ln</i>	0 0,0%	0 0,0%	0 0,0%	0 0%
Sources propres à la L2	Niveau du discours (ND)	33 6,5%	7 1,5%	9 1,5%	49 3,1%
	<i>SC03 Contexte lointain</i>	33 6,5%	7 1,5%	9 1,5%	49 3,1%
	Niveau de la phrase du mot (NP)	280 36,9%	243 52,9%	331 55,5%	854 54,8%
	<i>SC04 Ponctuation</i>	0 0,0%	0 0,0%	4 0,7%	4 0,3%
	<i>SC05 Contexte proche</i>	170 33,7%	179 39,0%	226 37,9%	575 36,9%
	<i>SC06 Grammaire</i>	110 21,8%	64 13,9%	101 16,9%	275 17,6%
	Niveau du mot (NM)	180 35,6%	168 36,6%	234 39,3%	582 37,3%
	<i>SC07 Association/Collocation</i>	6 1,2%	3 0,7%	8 1,3%	17 1,1%
	<i>SC08 Sens d'un caractère constitutif du mot</i>	106 21,0%	86 18,7%	112 18,7%	304 19,5%
	<i>SC09 Forme d'un caractère constitutif du mot</i>	44 8,7%	74 16,1%	101 16,9%	219 14,0%
	<i>SC10 Son d'un caractère constitutif du mot</i>	4 0,8%	1 0,2%	10 1,7%	15 1,0%
<i>SC11 Mode de constitution du mot</i>	20 4,0%	4 0,9%	3 0,5%	27 1,7%	
Total	505 100%	459 100%	596 100%	1560 100%	

Tableau 5.6 – Fréquence d'utilisation des sources de connaissance pour inférer les mots cibles de différentes classes grammaticales

Les fréquences SC sont classées ainsi pour les mots cibles de différentes classes selon les poids relatifs pendant l'inférence lexicale :

Adjectif : NP (36,9%) > NM (35,6%) > ND (6,5%) > SNL (2,4%) > SLn (0%)

Nom : NP (52,9%) > NM (36,6%) > SNL (8,9%) > ND (1,5%) > SLn (0%)

Verbe : NP (55,5%) > NM (39,3%) > SNL (3,7%) > ND (1,5%) > SLn (0%)

L'importance de la catégorie ND est plus prononcée dans l'inférence pour les noms, par rapport aux adjectifs et aux verbes. Une cause plausible à cette différence est le mot MC09 « 乏味 (*fáwèi*, fade, insipide) » (cf. supra p. 179).

Pour inférer les mots des trois classes grammaticales, les catégories NP et NM sont les plus utilisées. Les sources de connaissance majeures pour inférer les mots des trois classes grammaticales sont le contexte proche (SC05), la grammaire (SC06), le sens et la forme d'un caractère constitutif du mot (SC08 et SC09). Elles sont classées ainsi pour les trois classes grammaticales par ordre décroissant :

Adjectif :

Contexte proche (33,7%) > Grammaire (21,8%) > Sens d'un caractère constitutif du mot (21%) > Forme d'un caractère constitutif du mot (8,7%)

Nom :

Contexte proche (39%) > Sens d'un caractère constitutif du mot (18,7%) > Forme d'un caractère constitutif du mot (16,1%) > Grammaire (13,9%)

Verbe :

Contexte proche (37,9%) > Sens d'un caractère constitutif du mot (18,7%) > Forme d'un caractère constitutif du mot (16,9%) > Grammaire (16,9%)

L'utilisation des connaissances grammaticales est plus prononcée dans l'inférence pour les adjectifs (21,8% vs 13,9% pour les noms vs 16,9% pour les verbes).

5.3 Discussion

Nous avons défini au total onze sources de connaissance pour identifier les indices repérés par les participants pour inférer les mots cibles. Ces onze sources de connaissance sont regroupées sous cinq catégories : SNL (les connaissances non linguistiques), SLn (les connaissances propres aux langues non L2), ND (les

connaissances propres à la L2 au niveau d'un discours), NP (les connaissances propres à L2 au niveau d'une phrase) et NM (les connaissances propres à L2 au niveau d'un mot).

5.3.1 Diversité des sources de connaissance utilisées

Les participants se sont appuyés sur des sources de connaissance diverses pour inférer les mots cibles qui leur sont inconnus, en moyenne 2,4 sources de connaissance par mot. Les inférences lexicales dans lesquelles les participants ont utilisé deux sources de connaissance et plus représentent plus de 80% des 654 inférences observées. Seulement 16,7% des 654 inférences observées ont engagé une seule source de connaissance. Cela signifie qu'un apprenant puise en général dans des connaissances de diverses natures pour inférer un mot chinois inconnu.

Néanmoins, dans 8 inférences observées, aucun indice n'a été repéré. Cela signifie qu'il est possible qu'un apprenant ne trouve aucune information aidante pour comprendre le sens d'un mot inconnu, même s'il entreprend une inférence du sens pour ce mot.

En moyenne, la diversité des sources de connaissance utilisées pour inférer les mots lexicalisés est quasiment identique à celle pour les mots non lexicalisés.

5.3.2 Fréquence d'utilisation des sources de connaissance

5.3.2.1 Mots lexicalisés vs non lexicalisés

L'ordre de préférence des cinq catégories de source de connaissance est le même dans l'inférence des mots lexicalisés et non lexicalisés, mais avec des poids d'utilisation relatifs différents. Par ordre décroissant, les poids relatifs des cinq catégories peuvent être schématisés comme suit :

Mots lexicalisés :

NP > NM > SNL > ND > SLn

Mots non lexicalisés :

NP > NM > SNL > ND > SLn

Par rapport à d'autres études réalisées en anglais L2, l'utilisation de la catégorie NM est plus prononcée dans notre étude. En effet, dans les études réalisées en chinois L2 (ex. X. Jiang & Fang, 2012; S. H. Liu, 2001; Y. H. Xu, 2014b) et en japonais L2 (Mori, 2003), langues entièrement ou partiellement non alphabétiques à l'écrit, les sources de connaissance au niveau du mot ont généralement une grande importance dans l'inférence lexicale. Dans l'étude en chinois L2 de Dong (2007), le sens d'un caractère constitutif du mot inconnu est la source de connaissance principale pour les apprenants coréens pour inférer un mot chinois inconnu.

Concernant la préférence des catégories de source de connaissance dans l'inférence pour les mots lexicalisés et non lexicalisés, nous avons remarqué que le poids de la catégorie NP est plus important dans l'inférence pour les mots lexicalisés (59,5%) que pour les mots non lexicalisés (50,2%). En revanche, l'importance de la catégorie NM est plus prononcée dans l'inférence pour les mots non lexicalisés (31,7% pour les mots lexicalisés vs 42,6% pour les mots non lexicalisés).

Dans d'autres études en anglais L2 focalisant sur la lexicalisation, on a également trouvé des différences au niveau du poids d'utilisation des catégories de source de connaissance : pour les participants iraniens dans l'étude de Paribakht & Tréville (2007), le poids de NP est de 75% dans l'inférence pour les mots lexicalisés et de 72% pour les mots non lexicalisés ; le poids de ND est de 13% dans l'inférence pour les mots lexicalisés et de 15,3% pour les mots non lexicalisés ; L. Fan & Zhang (2014) ont également remarqué que, pour les apprenants chinois, les poids d'utilisation des catégories NP et NM dans les inférences des mots anglais sont respectivement de 58,1% et de 34,8% pour les mots lexicalisés, mais de 50,1% et de 43% pour les mots non lexicalisés.

Nous en déduisons que, comme un mot non lexicalisé n'a pas d'équivalent dans le lexique en L1 de l'apprenant, il est plausible que le contexte proche, la piste principale, ne fournisse pas suffisamment d'informations pour que l'apprenant puisse comprendre le mot. Dans ce cas-là, l'apprenant doit s'inspirer d'autres pistes et porter son attention soit vers un contexte plus large, soit vers l'intérieur du mot. Comme l'indiquent Paribakht & Tréville (2007, p. 417), « *[L]es sujets qui se sont davantage*

fiés aux indices parsemés dans le discours pour inférer le sens des mots non lexicalisés ont manifesté leur besoin d'un bassin d'information plus large pour soutenir leur effort de déduction du sens des mots non lexicalisés. »

Les participants se sont servi des connaissances non linguistiques (SNL) dans l'inférence pour les mots lexicalisés comme pour les mots non lexicalisés. Mais globalement cette catégorie ne s'est pas montrée d'une grande importance dans notre étude.

Parmi les connaissances linguistiques propres à la L2, nos participants ont peu utilisé la source de connaissance au niveau du discours (ND). Rappelons ici que, dans notre étude, nous avons utilisé deux textes courts (d'environ 300 caractères chacun) de structure simple. Le contenu des textes est présenté de façon linéaire. Les participants ont donc moins besoin de puiser dans la structure globale du texte ou de chercher loin les informations pour comprendre un mot cible.

Nos participants n'ont pas puisé dans les connaissances linguistiques propres aux langues non L2 (SLn) pour inférer les mots cibles. Paribakht & Tréville (2007) ont remarqué dans leur étude en anglais L2 que seuls les participants français se fient à l'homonymie en L1 (c'est-à-dire que le sujet confond le mot cible avec un mot de L1 de forme, sonore ou graphique, ressemblante) dans l'inférence des mots anglais L2, mais non les participants iraniens ; l'utilisation des connaissances linguistiques en L1 est beaucoup moins importante chez les participants iraniens (0,3%), par rapport aux participants français (4%). Les apprenants chinois de l'anglais L2 dans l'étude de L. Fan & Zhang (2014) n'ont pas utilisé leurs connaissances linguistiques en chinois pour inférer les mots anglais non lexicalisés, et le poids de cette source de connaissance dans l'inférence des mots lexicalisés est seulement de 0,1%. La langue chinoise est typologiquement lointaine de la langue française (la L1 de nos participants), mais aussi de l'anglais, une autre langue étrangère que nos participants pratiquent. Cela est probablement la raison pour laquelle nos participants n'ont pas pu tirer profit de leurs connaissances linguistiques déjà acquises dans ces deux langues pour inférer les mots chinois inconnus.

Les quatre sources de connaissance majeures que nos participants ont utilisées

pour inférer les mots cibles sont le contexte proche (ou le sens de la phrase contenant le mot inconnu), la grammaire, le sens d'un caractère constitutif et la forme d'un caractère constitutif du mot.

Le contexte proche est la source de connaissance principale. Cela est en accord avec beaucoup d'autres études sur l'inférence lexicale (ex. H. Fan, 2012; L. Fan & Zhang, 2014; Nassaji, 2003; D. D. Qian, 2004; Wesche & Paribakht, 2009). En effet, il s'agit d'une stratégie typique de lecture de relire la phrase qui comporte le mot inconnu. Wesche & Paribakht (2009) ont sollicité des natifs anglophones et des apprenants français et persans de l'anglais L2 pour effectuer l'inférence lexicale dans la L1 des apprenants (donc l'anglais, le français et le farsi) et en anglais L2. Le résultat montre que le sens de la phrase contenant un mot inconnu est la source de connaissance principale pour inférer le sens du mot, en L2 comme en L1, quelle que soit la L1 du participant.

Les quatre sources de connaissance les plus utilisées pour inférer les mots chinois inconnus sont classées ainsi par ordre décroissant dans l'inférence pour les mots lexicalisés et pour les mots non lexicalisés :

Mots lexicalisés :

Contexte proche > Grammaire > Sens d'un caractère constitutif du mot > Forme d'un caractère constitutif du mot

Mots non lexicalisés :

Contexte proche > Sens d'un caractère constitutif du mot > Forme d'un caractère constitutif du mot > Grammaire

Une différence significative se situe au niveau du poids de la grammaire (21,5% pour les mots lexicalisés vs 14% pour les mots non lexicalisés) et du sens d'un caractère constitutif du mot (14,9% pour les mots lexicalisés vs 23,8% pour les mots non lexicalisés). La différence au niveau de la forme d'un caractère constitutif n'est pas significative, mais assez importante ($0,05 < p < 0,1$).

Nos résultats montrent une plus grande importance de l'utilisation du sens des caractères constitutifs dans l'inférence pour les mots chinois non lexicalisés inconnus. Nous pensons que cela est lié aux faits que 1) en chinois, le sens d'un caractère constitutif d'un mot est souvent en rapport étroit avec le sens global du mot et 2) les

mots non lexicalisés utilisés dans notre expérience contiennent plus de caractères courants et connus par nos participants. L'exploitation plus importante du sens des caractères constitutifs pour inférer les mots non lexicalisés démontre, selon nous, le fait que nos participants, ayant un niveau intermédiaire en chinois L2, sont conscients de l'apport sémantique d'un caractère constitutif au sens global d'un mot chinois.

Le fait que la source de connaissance de la forme des caractères constitutifs fait partie des sources de connaissance majeures pour inférer les mots inconnus, lexicalisés et non lexicalisés, montre également que nos apprenants sont conscients qu'au sein d'un caractère chinois, les blocs composants peuvent délivrer des informations lexicales.

Nous avons constaté que nos participants s'appuient souvent sur les particules pour identifier la classe grammaticale d'un mot inconnu. Le fait que les mots lexicalisés sont davantage entourés d'une particule dans les textes (par exemple, la particule « 的 (*de*) » derrière MC06, MC12, MC17 et devant MC14 ; la particule « 着 » derrière MC10) expliquerait que le poids d'utilisation de la grammaire plus important dans l'inférence des mots lexicalisés.

5.3.2.2 Adjectif vs Nom vs Verbe

Pour inférer les mots cibles de différentes classes grammaticales, les préférences des catégories de source de connaissance sont les mêmes, à l'exception de l'importance des sources de connaissance au niveau du discours (ND) plus grande pour inférer les adjectifs inconnus. Cela est vraisemblablement dû au contexte de MC09 « 乏味 (*fǎwèi*, fade, insipide) » (cf. supra p. 179). Nous concluons donc que l'ordre d'importance des catégories est identique pour l'inférence des mots inconnus des trois classes grammaticales.

Parmi les sources de connaissance, l'utilisation des connaissances grammaticales est plus prononcée dans l'inférence pour les adjectifs (21,8% vs 13,9% pour les noms vs 16,9% pour les verbes). Nous pensons que cette inclination vers les indices grammaticaux est liée au fait que les mots cibles adjectivaux dans nos textes testés sont davantage « balisés » : les adjectifs tels que MC11 « 安逸 (*ānshì*,

tranquille et confortable) », MC01 « 节俭 (jiéjiǎn, économe, frugal) », MC13 « 明丽 (mínglì, clair et beau) » et MC09 « 乏味 (fáwèi, fade, insipide) » sont les mots pour l'inférence desquels les participants ont recouru relativement plus à la piste grammaticale.

Ex. 1 : *je pense que c'est un adjectif, puisqu'il est placé après <((Le participant lit un mot dans le texte.))> « hen », avant <((Le participant lit un mot dans le texte.))> « de ». Je pense que c'est pour donner une précision sur le nom qui est après le « de ».*

Dans cet exemple, le participant pense que MC11 « 安逸 (ānshì, tranquille et confortable) » est un adjectif à cause de deux indices grammaticaux, l'adverbe « 很 (hěn, très) » juste devant MC11 et la particule « 的 (de) » immédiatement derrière. En chinois, un adverbe se situe généralement devant un adjectif dans une phrase et la particule « 的 (de) » peut se placer entre un adjectif déterminant et un nom déterminé.

Ex. 2 : *<((Le participant lit quelque chose dans le texte))> « ping shi dou fei chang ». Du coup, je pense que c'est un adjectif, parce qu'il est après <((Le participant lit quelque chose dans le texte.))> « fei chang ».*

Dans le deuxième exemple, le participant a repéré l'adverbe « 非常 (fēicháng, très, extrêmement) » placé juste devant MC01 « 节俭 (jiéjiǎn, économe, frugal) ».

Ex. 3 : *Donc, <((Le participant lit le mot cible.))> « ming li » c'est un adjectif. Vu qu'il y a <((Le participant lit un mot dans le texte.))> « geng » juste avant, qui fait la référence à « un peu plus »*

Ex. 4 : *A cause de <((Le participant lit un mot dans le texte.))> « yue lai yue », ... il me semble qu'on utilise un adjectif après ça, dans le sens « plus ça allait, plus c'était quelque chose ».*

Dans le troisième exemple, le participant a remarqué l'adverbe « 更 (gèng, (encore) plus) » qui se trouve devant le mot cible MC13 « 明丽 (mínglì, clair et beau) » ; dans le quatrième exemple, le participant juge que MC09 « 乏味 (fáwèi, fade, insipide) » est un adjectif parce qu'il se situe derrière « 越来越 (yuèlái yuè, de plus en plus) ».

Pour cela, nous pensons que, dans notre expérience, les différences de choix des sources de connaissance pendant les inférences des mots de différentes classes grammaticales dépendent plutôt du contexte dans lequel se trouve le mot inféré. Nous ne pouvons donc pas donner une conclusion définitive sur les différences de préférence des sources de connaissance pendant les inférences des mots inconnus de différentes classes grammaticales.

5.3.3 Interprétation de quelques cas intéressants

5.3.3.1 Contexte proche

Il est courant que nos participants relisent la phrase contenant le mot à inférer à plusieurs reprises pour essayer de trouver plus d'informations.

Ex. (MC07 « 还礼 ») : *Alors, je connais la signification de « hai ». Je connais la signification de « li », mais la combinaison je la connais pas. Donc, pour essayer de comprendre je relis la phrase.*

Il est à noter que l'utilisation du contexte proche est souvent combinée à d'autres sources de connaissance dans notre étude. Parmi les 654 inférences observées, seules 72 inférences concernent les cas où les participants se sont servis uniquement du contexte proche, soit 11%. Ce résultat est différent de celui de l'étude de Paribakht (2005) en anglais L2, dans laquelle l'utilisation unique de cette source représente 85% des inférences observées.

La langue chinoise n'est pas alphabétique à l'écrit. La grande majorité des caractères chinois contiennent plusieurs blocs composants qui peuvent délivrer des informations lexicales, sémantiques ou phonétiques. En plus, les mots chinois, constitués d'un ou de plusieurs caractères, sont juxtaposés dans une phrase. Il n'y a pas d'espace supplémentaire entre les mots pour baliser les frontières. Pour un apprenant dont la L1 est alphabétique comme le français, la reconnaissance des caractères et l'identification des unités lexicales sont les premiers obstacles dans la compréhension écrite en chinois. En effet, nous avons remarqué que les caractères inconnus rendaient nos participants anxieux. Lorsqu'on demandait à un participant d'inférer le sens d'un mot surligné, il était courant qu'il fasse tout de suite une

remarque sur sa méconnaissance des caractères impliqués.

Ex. : *[L]e caractère ... souligné, je le connais pas du tout ... Le premier caractère, je l'ai jamais vu. Le second c'est possible, mais... les deux ensemble, je vois pas du tout ce que ça veut dire.*

De ce fait, le décryptage des caractères constitutifs et l'identification du rôle d'un mot inconnu représentent une grande importance dans l'inférence lexicale en chinois L2. L'apprenant est donc obligé d'explorer des pistes variées, y compris le contexte proche, pour chercher des informations utiles. Cela explique l'utilisation importante de la forme et du sens d'un caractère constitutif du mot (33,5%) et de la grammaire (17,6%) dans les 654 inférences observées.

5.3.3.2 Sens d'un caractère constitutif d'un mot inconnu

Le sens d'un caractère constitutif est une piste importante pour nos participants pour inférer un mot chinois inconnu. Quand un mot contient un ou des caractères connus, tels que MC02⁷⁶ (qui contient le caractère « 花 » que nos participants connaissent), MC03⁷⁷ (« 送 »), MC04⁷⁸ (« 月 »), MC05⁷⁹ (« 汤 »), MC07⁸⁰ (« 还 » et « 礼 ») et MC12⁸¹ (« 五 », « 六 » et « 色 »), la majorité des participants ont recouru à cette source de connaissance. En revanche, bien logiquement, elle est quasiment absente dans l'inférence pour les mots dont les caractères constitutifs sont inconnus pour nos participants, tels que MC06⁸² (« 雀 » et « 跃 »), MC14⁸³ (« 彩 » et « 虹 »), MC15⁸⁴ (« 踏 » et « 青 ») et MC17⁸⁵ (« 蝉 »).

Nous rappelons qu'un grand nombre de caractères chinois possèdent plusieurs sens différents (cf. la Section 3.2.2.3). Il n'est donc pas toujours fiable de compter

⁷⁶ MC02 « 窗花 (*chuānghuā*, papier découpé pour la décoration d'une fenêtre) »

⁷⁷ MC03 « 送灶 (*sòngzào*, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune) »

⁷⁸ MC04 « 腊月 (*làyuè*, le douzième mois du calendrier lunaire) »

⁷⁹ MC05 « 汤圆 (*tāngyuán*, boulette de riz glutineux farcie de viande (fourrée à la confiture)) »

⁸⁰ MC07 « 还礼 (*huánlǐ*, faire un présent en retour, offrir un cadeau à son tour) »

⁸¹ MC12 « 五颜六色 (*wǔyánliùsè*, de diverses couleurs, multicolore) »

⁸² MC06 « 雀跃 (*quèyuè*, sauter de joie) »

⁸³ MC14 « 彩虹 (*cǎihóng*, arc-en-ciel) »

⁸⁴ MC15 « 踏青 (*tàqīng*, faire une excursion (se promener) à la campagne au printemps) »

⁸⁵ MC17 « 蝉 (*chán*, cigale) »

seulement sur les sens des caractères constitutifs pour inférer un mot inconnu, notamment pour les apprenants d'un bas niveau qui ont en général une faible maîtrise de caractères. (Cf. aussi Gan, 2008 et Y. Wang et al., 2017)

Dans notre étude, nous avons remarqué l'entrave des caractères polysémiques dans l'inférence pour les mots cibles, entre autres pour les mots MC07 et MC08.

– MC07 « 还礼 (*huánlǐ*, faire un présent en retour, offrir un cadeau à son tour) »

Le premier caractère du mot, « 还 », est un caractère polyphonétique et polysémique. Plus précisément, ce caractère peut être un adverbe « *hái* (encore) » ou un verbe qui se prononce « *huán* (rendre (un emprunt), faire un retour) ». Le « 还 » dans MC07 « 还礼 » a un sens verbal. Nos participants connaissent ce caractère dans les deux cas, le premier étant plus couramment utilisé par nos participants.

Lors de l'inférence de ce mot, seuls 2 participants sur 33 ont correctement identifié le caractère « 还 (*huán*) », grâce aux autres indices.

a)

P26 : Du coup, là le caractère je pense le connaître. C'est pas <((Il prononce le premier caractère du mot surligné.))> « hai » comme on le croirait au premier coup. C'est <((Il prononce le même caractère autrement.))> « huan », euh qui est « échanger ». Et la partie, euh le deuxième caractère, c'est le « li » de « li wu », euh qui est le « cadeau ». ...

Expérimentateur : Et pourquoi vous avez dit que c'est pas « hai », mais que c'est « huan » ?

P26 : C'est le cadeau qui me fait euh penser à ça.

b)

P32 : Euh je pense que ça se prononce « huan li », pour dire « échanger des cadeaux ». Donc le premier en fait c'est euh... il peut se- c'est un caractère qui peut se prononcer différemment. Euh... je pense que ça peut se prononce soit « hai » pour dire « et aussi » ou alors « huan » pour dire « échanger ». Et « li » euh c'est un caractère qu'on retrouve dans le <((Elle dit un mot chinois.))> « li wu », pour le « cadeau ». ...

Expérimentateur : Et pourquoi vous pensez que c'est « huan » au lieu de « hai » ?

P32 : Euh « hai », « hai li », euh <((Elle lit quelque chose dans le texte.))> « fu mu zai yong zhe xie qian qu gei qi ta ren » Euh, parce que là euh... j- je sais pas si « li » c'est un verbe. Parce que si on met « hai », il faut un verbe après,

Certains participants ont compris le sens global de la phrase contenant MC07. Toutefois, l'identification erronée de « 还 (huán, rendre (un emprunt), faire un retour) » ne leur permet pas d'avoir une réponse complètement correcte.

Ex. (MC07 « 还礼 ») : *Je pense que c'est un parent qui réutilise l'argent euh pour offrir des, des cadeaux à d'autres gens. Je, je mets « hai » avec « li », je pense que ça veut dire « cadeau » ou « présent », quelque chose comme ça.*

– MC08 « 在乎 (zàihu, prendre quelque chose à cœur ; faire attention à) »
« 在 (zài) » est un caractère courant que les participants connaissent bien pour sa fonction de préposition pour préciser un emplacement, par exemple « 我(je)在法国(France)。 (Wǒ zài Fǎguó. Je suis en France.) ». Le fait que les participants ne connaissent « 在 (zài) » qu'en tant que préposition les a perturbés pour analyser la phrase contenant MC08.

Ex. (MC08 « 在乎 ») : *En plus, dans ma lecture j'avais associé le deuxième caractère avec <((Le participant lit le caractère immédiatement après MC08 qui signifie « argent ».))> « qian », et je me disais « zai » ... peut-être c'est un lieu où y a de l'argent ?... du coup je sais pas du tout ce que ça veut dire si on prend que la deuxième partie de la phrase, « C'est pas notre propre argent ». Je comprends pas trop.*

Certains participants doutaient même que les deux caractères de MC08, « 在 (zài) » et « 乎 (hu) » constituent ensemble un mot.

Ex. (MC08 « 在乎 ») : *[L]es deux mots⁸⁶ sont surlignés ensemble mais... je sais pas. C'est le... « zai » qui me... perturbe. ... Celui-là c'est une préposition. ... je pense que je l'ai vu qu'en tant que caractère seul qui veut dire un lieu ... du coup je sais pas s'il est associé au second caractère ou s'ils sont tous les deux euh indépendants... Mais je ne vois pas.*

⁸⁶ Les participants confondaient facilement les termes « caractère » et « mot ». En l'occurrence, le participant voulait dire « deux caractères ».

En effet, la polysémie des caractères chinois (cf. Z. Chen, 2017) est un grand obstacle pour les apprenants du chinois L2 de tous niveaux, même pour les apprenants japonais dont la L1 est très proche du chinois (X. J. Qian, 2005; Y. Wang et al., 2017).

Y. Wang et ses collègues (2017) ont mené une étude auprès d'apprenants de niveau intermédiaire en chinois L2 sur la compréhension des mots sémantiquement transparents (c'est-à-dire ceux dont le sens est bien étroitement lié aux sens des caractères constitutifs). Ces mots contiennent des caractères polysémiques. Les résultats montrent que, quand les sens des caractères constitutifs ne sont pas ceux les plus fréquemment utilisés, leurs participants ont plus de difficultés à comprendre le sens du mot. Par exemple, le caractère « 黑 (*hēi*) » peut signifier « noir » et « clandestin ». Il serait plus probable pour un apprenant du chinois L2 d'utiliser le sens « noir » que l'autre sens. Dans ce cas-là, même si l'apprenant connaît le sens de « 话 (*huà*, parole) », il n'est pas évident pour lui de comprendre le mot « 黑话 (*hēihuà*, argot) ».

Nous pensons que nos participants ayant un niveau intermédiaire ont une maîtrise faible des caractères chinois au niveau du sens. En face d'un caractère polysémique, 1) quand l'apprenant connaît plusieurs sens du caractère (comme « 还 » dans MC07 « 还礼 (*huánlǐ*, faire un présent en retour, offrir un cadeau à son tour) »), le sens le plus fréquemment utilisé serait activé le premier ; 2) quand la connaissance de l'apprenant sur un caractère n'est pas appropriée pour la compréhension d'un mot constitutif de ce caractère (comme « 在 » dans MC08 « 在乎 (*zàihu*, prendre quelque chose à cœur ; faire attention à) »), ce caractère est en réalité « inconnu » pour l'apprenant. Dans ce cas-là, la connaissance préalable du caractère peut même désorienter l'apprenant dans sa compréhension du mot et/ou de la phrase.

Nous pensons donc qu'un apprenant de niveau intermédiaire en chinois L2 n'est pas encore conscient de la nécessité d'identifier le sens approprié d'un caractère par le biais d'autres indices dans le contexte.

5.3.3.3 Connaissances grammaticales

La grammaire est une autre source de connaissance phrastique importante

utilisée par nos participants. Les indices grammaticaux ne fournissent pas d'informations sémantiques concrètes d'un mot inconnu, mais ils peuvent aider un apprenant à comprendre le rôle d'un mot inconnu dans la phrase.

Ex. (MC17 « 蝉 ») : <((Le participant lit quelque chose dans le texte.))> « *shu shang de* » ... « *shu shang de shen me shen me cong zao jiao dao wan* » ... Donc là ça serait plutôt un nom, ... en fait ça fait partie ... du groupe nominal, du sujet de la phrase.

Dans cet exemple, le participant infère MC17 « 蝉 (*chán*, cigale) ». La phrase contenant le mot est « 树上的蝉从早叫到晚。 *Shù shàng de chán cóng zǎo jiào dào wǎn*. Les MC17 dans les arbres chantent du matin au soir. ». Le participant a conclu que MC17 fait partie du groupe nominal « 树上的蝉 (*Shù shàng de chán*, les cigales dans les arbres) » qui est le sujet de la phrase.

Il n'est pas rare qu'un mot chinois puisse appartenir à plusieurs catégories grammaticales. Par exemple, « 认识 (*rènshi*) » peut être un verbe (« connaître ») ou un nom (« connaissance »). Autrement dit, l'identification de la classe grammaticale d'un mot nécessite parfois de faire attention au contexte dans lequel on l'emploie.

Nous avons constaté un grand appui sur les particules chez nos participants afin de déceler la classe grammaticale d'un mot donné.

Ex. (MC10 « 盼望 ») : [J]e pense que c'est peut-être un adjectif ...puisque'il y a un <((Le participant lit quelque chose dans le texte.))> « *zhe* » après. Non, du coup c'est un, un verbe.

Dans cet exemple, le participant infère MC10 « 盼望 (*pànwàng*, espérer, désirer) ». Il a remarqué la présence de la particule « 着 (*zhe*) » derrière MC10. Cette particule peut être placée immédiatement derrière un verbe pour signifier la durée d'une action ou d'un état. Il a donc jugé que MC10 « 盼望 (*pànwàng*, espérer, désirer) » est un verbe.

Ex. (MC11 « 安适 ») : [L]e <((Le participant lit un mot dans le texte.))> « *de* », je sais que ... il y a l'histoire de déterminé, déterminant ... Donc en gros, il y a un nom, un adjectif, une histoire comme ça.

Dans l'exemple ci-dessus, le participant a repéré la particule « 的 (de) » qui peut se placer entre un adjectif déterminant et un nom déterminé. Le participant pense donc que MC11 est un adjectif qui détermine le nom derrière la particule « 的 (de) ».

Cependant, la même particule « 的 (de) » a créé une entrave à beaucoup de participants dans l'analyse de la phrase contenant le mot MC06 « 雀跃 (*quèyuè*, sauter de joie) ». La phrase contenant MC06 est « 当然, 过年时最让孩子们雀跃的事儿还是拿红包。 Bien sûr, ce qui MC06 le plus les enfants était de recevoir des *hongbao*. »⁸⁷. La particule « 的 (de) » se situe entre MC06 « 雀跃 (*quèyuè*, sauter de joie) » et un nom « 事儿 (*shì'r*, chose, affaire) ». Mais dans l'exemple ci-dessous, le participant est persuadé que MC06 est un adjectif qui qualifie le nom « 事儿 (*shì'r*, chose) ».

Ex. 2 (MC06 « 雀跃 ») : *Je pense que c'est un adjectif, parce qu'il est placé avant le « de » pour définir le mot <((Le participant lit quelque chose dans le texte.)> « shì'r » ... qui veut dire « chose », je pense.*

Or, le déterminant introduit par la particule « 的 (de) » peut être un nom, un adjectif, un pronom ou une proposition. En l'occurrence, le déterminant de « 事儿 (*shì'r*, chose) » est une proposition, « 让孩子们雀跃 (faire les enfants sauter de joie) ». La traduction de « 让孩子们雀跃的事儿 » est « la chose **qui** MC06 le plus les enfants ».

Visiblement, pour nos participants, la particule « 的 (de) » peut les aider à analyser une phrase quand elle introduit un mot déterminant. Mais elle désoriente nos participants dans les cas où le déterminant est une proposition. En français, une proposition subordonnée relative est introduite derrière le nom qualifié à l'aide d'un pronom relatif (par exemple, « la chose **qui** faisait sauter de joie les enfants »). Nous pensons qu'un apprenant ayant un niveau intermédiaire n'a pas encore une bonne maîtrise de la grammaire du chinois.

Pour cette raison-là, même si certains participants ont compris le sens global

⁸⁷ Littéralement traduit, cette phrase signifie « Bien sûr, **la chose** qui MC06 le plus les enfants était de recevoir des *hongbao*. »

du mot MC06 « 雀跃 (*quèyuè*, sauter de joie) » grâce au contexte, ils n'ont pas pu identifier la classe grammaticale du mot ; c'est le cas dans l'exemple suivant.

Ex. (MC06 « 雀跃) : *Euh, donc là après le mot surligné, on a <((Le participant lit quelque chose dans le texte.))> « de shi er ». Euh je pense que le mot surligné, ... il va avoir la fonction d'adjectif ... pour qualifier le mot « shi er »... Et comme toujours c'est par rapport au Nouvel an. C'est quelque chose de positif. ... c'est quelque chose de joyeux. ... Donc je pense que c'est un adjectif positif euh pour dire qu'ils sont contents ...*

5.3.3.4 Forme d'un caractère constitutif d'un mot inconnu

La forme des caractères constitutifs représente 14% de la totalité des utilisations de sources de connaissance. Beaucoup de participants ont utilisé cette source de connaissance pour inférer MC03 « 送灶 »⁸⁸, MC04 « 腊月 »⁸⁹ et MC19 « 飘零 »⁹⁰. Ces trois mots contiennent des caractères inconnus pour les participants, mais ces caractères sont composés des blocs composants courants et connus par les participants. Nos participants sont donc conscients qu'ils peuvent se fier sur la forme d'un caractère inconnu pour chercher des indices.

Or, du fait de l'évolution de la langue chinoise au fil des siècles, la forme d'un caractère ne délivre pas toujours des informations pertinentes.

Ex. (MC17 « 蝉 ») : *[L]a première partie du caractère c'est « chong »⁹¹ ... comme dans ... <((Le participant mentionne le mot MC14.))> « cai hong ». ... [Ç]a représente les arbres. Donc peut-être les branches d'arbre ? Parce que <((Le participant mentionne le mot MC14.))> « cai hong », c'est l'idée de quelque chose de long et de droit. Peut-être ça représente ... les branches.*

Dans cet exemple, le participant infère le mot cible MC17 « 蝉 »⁹². Il a reconnu la clé « 虫 » dans ce caractère et a ensuite pensé au deuxième caractère « 虹 » dans

⁸⁸ MC03 « 送灶 (*sòngzào*, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune) »

⁸⁹ MC04 « 腊月 (*làyuè*, le douzième mois du calendrier lunaire) »

⁹⁰ MC19 « 飘零 (*piāolíng*, tomber en voltigeant) »

⁹¹ Le participant voulait dire que la partie à gauche du caractère « 蝉 ». « 虫 (*chóng*) » est à la fois un caractère.

⁹² MC17 « 蝉 (*chán*, cigale) »

MC14 « 彩虹 »⁹³. Selon le participant, la clé « 虫 » symbolise quelque chose de long et droit. Tenant compte du contexte de MC17 qui parle d'arbres, il a donc proposé « branche d'arbre » pour le sens de MC17. Effectivement, la forme originale de la clé « 虫 » est « 𧈧 », qui symbolise le serpent. Aujourd'hui on retrouve cette clé généralement dans les caractères qui sont en rapport avec les vers et les insectes, comme MC17 « 蝉 (*chán*, cigale) ».

Quand la clé et/ou le bloc composant d'un caractère inconnu est moins familier, il semble que les participants ont été vite dissuadés d'explorer davantage le caractère, par exemple « 颜 » dans M12 « 五颜六色 »⁹⁴ et « 窗 » dans MC02 « 窗花 »⁹⁵.

MC07 « 还礼 »⁹⁶ contient deux caractères dont la forme est connue par les participants. Dans l'inférence de MC07, cette source de connaissance est peu utilisée. Selon les résultats de l'étude de (Y. G. Zhao, 2009), chez les apprenants du chinois L2 de niveau avancé, lors de la reconnaissance d'un caractère de haute fréquence, les activations de la forme et du sens dans la mémoire se font presque simultanément. Nous en déduisons que, pour nos participants ayant un niveau intermédiaire, l'activation du lien forme-sens pour certains caractères courants est rapide. Sous un autre angle, les formes de ces caractères sont enregistrées en « unité soudée » (cf. Feng, 2009) dans la mémoire des participants qui n'ont pas besoin de « rassembler » des blocs composants au moment de la reconnaissance du caractère. Plus ils utilisent ces caractères, plus l'activation du lien forme-sens de ces caractères est automatique. Ils n'ont alors pas besoin de décomposer ces caractères.

En outre, il est intéressant de voir qu'un apprenant peut vérifier une idée sur le sens du mot inféré par le biais de la forme de caractère.

Ex. (MC17 « 蝉 ») : *[N]on, c'est pas des fruits, parce que ... je sais ce que c'est le caractère de « fruit ».*

Ayant compris du contexte que le mot MC17 « 蝉 »⁹⁷ a un rapport avec les arbres, le

⁹³ MC14 « 彩虹 (*cǎihóng*, arc-en-ciel) »

⁹⁴ MC12 « 五颜六色 (*wǔyánliùsè*, de diverses couleurs, multicolore) »

⁹⁵ MC02 « 窗花 (*chuānghuā*, papier découpé pour la décoration d'une fenêtre) »

⁹⁶ MC07 « 还礼 (*huánlǐ*, faire un présent en retour, offrir un cadeau à son tour) »

⁹⁷ MC17 « 蝉 (*chán*, cigale) »

participant a voulu proposer « fruit » pour le sens de MC17. Mais il a ensuite renoncé à cette réponse parce que le caractère chinois correspondant à « fruit » est « 果 », qui ne ressemble pas à MC17 « 蝉 » à la forme.

Mais le plus souvent, nos participants confondent les caractères ayant une forme similaire.

Ex. 1 (MC10 « 盼望 ») : *Le premier ... pourrait me faire penser à <((Le participant dit deux caractères chinois.))> « yue », « fen ».*

Le participant analyse le premier caractère « 盼 » dans MC10 « 盼望 »⁹⁸ et il le décompose en « 目 (mù) » et « 分 (fēn) ». Mais il a confondu « 目(mù) » avec « 月 (yuè) ».

Ex. 2 (MC06 « 雀跃 ») : *[L]e premier caractère me fait penser aux deux caractères pour dire « préparer », <((Le participant dit un mot chinois qui signifie « préparer ».))> « zhun bei ».*

Dans cet exemple, le participant a confondu « 雀 », le premier caractère de MC06 « 雀跃 »⁹⁹ avec « 准 », le premier caractère du « 准备 (zhǔnbèi, préparer) ». « 雀 » et « 准 » partagent en effet un même bloc composant « 隹 ».

Nous pensons que, pour un apprenant du chinois L2 ayant un niveau intermédiaire, la maîtrise de la forme des caractères chinois reste relativement faible.

5.3.3.5 Information phonétique

Le lien graphie-son dans un caractère est faible (cf. la Section 3.2.2). Néanmoins, certains participants ont tenté d'identifier la prononciation des mots inférés.

Ex. (MC09 « 乏味 ») : *Alors je me demande si ça se prononce « zhi wei ». Je suis pas sûre, mais comme ... y a le même caractère que <((Le participant dit un caractère chinois.))> « zhi » dans le premier caractère.*

Le participant ne connaît pas le premier caractère « 乏 » dans MC09 « 乏 »

⁹⁸ MC10 « 盼望 (pànwàng, espérer, désirer) »

⁹⁹ MC06 « 雀跃 (quèyuè, sauter de joie) »

味 »¹⁰⁰. Mais il a remarqué qu'une partie de « 乏 » est identique au caractère « 之 (zhī) ». Le participant a alors voulu transposer la prononciation du caractère « 之 (zhī) » sur « 乏 ».

Dong (2007) a remarqué dans son étude que les apprenants coréens du chinois L2 cherchent souvent à identifier la prononciation des mots chinois inconnus. Dong explique cela comme une habitude que les Coréens transfèrent de leur L1, le coréen, qui est une langue alphabétique et pour laquelle la prononciation est importante dans la reconnaissance des mots. Nous pensons que, pour un apprenant francophone de niveau intermédiaire en chinois L2, le son est toujours un élément aidant pour activer dans sa mémoire les informations sur un caractère. Une autre raison plausible est le fait que nous ayons utilisé la méthode PVH pendant l'entretien d'inférence. Cela a pu pousser nos participants à prononcer les caractères présents dans le texte à voix haute (même s'ils ont été informés préalablement que, lorsqu'ils rapportent oralement leurs pensées pendant les inférences lexicales, ils pourraient sauter les caractères inconnus ou les remplacer par « quelque chose je ne sais pas quoi », « hm hm » etc.).

5.3.3.6 Mode de constitution d'un mot

L'appellation « mode de constitution d'un mot » utilisée dans notre étude se réfère à toutes les informations que nos participants ont repérées au sujet de la façon dont est constitué un mot.

Cette source de connaissance est particulière dans l'inférence lexicale en chinois et d'autres langues linguistiquement influencées par le chinois (par exemple la langue japonaise). Dans notre étude, nos participants ont surtout utilisé cette source de connaissance dans l'inférence pour les mots contenant plus de deux caractères, MC12 « 五颜六色 »¹⁰¹ et MC20 « 白茫茫 »¹⁰². La raison principale, selon nous, est que la majorité des mots chinois que nos participants connaissent sont constitués de deux caractères. Les mots contenant plus de deux caractères sont visuellement remarquables et cela a incité nos participants à s'interroger davantage sur la relation

¹⁰⁰ MC09 « 乏味 (fǎwèi, fade, insipide) »

¹⁰¹ MC12 « 五颜六色 (wǔyánliùsè, de diverses couleurs, multicolore) »

¹⁰² MC20 « 白茫茫 (báimángmáng, d'une blancheur éclatante (immaculée) ; une immensité argentée) »

entre les caractères constitutifs.

Nous constatons que nos participants n'ont pas beaucoup utilisé cette source de connaissance. Cependant, il est intéressant de voir que nos participants en ont la connaissance et qu'ils s'en servent pour analyser un mot chinois inconnu.

5.4 Conclusions

Nous voulions comprendre les sources de connaissance sur lesquelles un apprenant francophone s'appuie pour inférer un mot chinois inconnu rencontré dans une lecture. Plus particulièrement, nous voulions savoir si le statut de lexicalisation du mot chinois inconnu a un effet sur le choix des sources de connaissance pendant l'inférence lexicale.

Nous avons adopté et développé la typologie de sources de connaissance de Paribakht & Tréville (2007). Nous avons défini onze sources de connaissance, regroupées sous cinq catégories en fonction de la propriété des sources de connaissance.

Après avoir observé 654 inférences lexicales produites par nos 33 apprenants francophones, nous concluons qu'un apprenant francophone s'appuie en général sur diverses sources de connaissance pour inférer le sens d'un mot chinois inconnu. Le poids relatif des sources de connaissance n'est pas le même dans l'inférence du sens d'un mot chinois inconnu. **Les sources de connaissance propres à la L2 (en l'occurrence, la langue chinoise) se situant au niveau de la phrase ont une importance considérable, ainsi que celles au niveau du mot. La source de connaissance principale pour inférer un mot chinois inconnu est le sens de la phrase contenant le mot inconnu. Les sources de connaissance au niveau du mot ont une importance plus prononcée dans l'inférence lexicale en chinois L2, par rapport à d'autres L2 alphabétiques, par exemple l'anglais.** Cela est probablement lié aux faits que 1) la forme d'un caractère peut délivrer de riches informations lexicales et que 2) le sens d'un mot est souvent en fort lien avec les sens des caractères constitutifs. Les autres sources de connaissance majeures sont la grammaire, le sens des caractères constitutifs et la forme des caractères constitutifs du mot.

Pour inférer des mots chinois inconnus lexicalisés et non lexicalisés en français, un apprenant francophone utilise les mêmes sources de connaissance. Quelle que soit la lexicalisation du mot chinois inféré, l'apprenant s'appuie le plus sur la phrase contenant le mot inféré pour chercher des informations. Toutefois, par

rapport aux mots non lexicalisés, dans l'inférence des mots lexicalisés, l'importance des sources de connaissance au niveau de la phrase est plus prononcée. A contrario, les sources de connaissance au niveau du mot ont une plus grande importance dans l'inférence pour les mots non lexicalisés. Nous concluons donc que **la nature de lexicalisation d'un mot chinois a une influence sur le choix des sources de connaissance utilisées dans la déduction du sens du mot. Le choix des sources de connaissance est plus parsemé dans l'inférence d'un mot chinois non lexicalisé que dans celle d'un mot lexicalisé.**

Précisément, dans notre étude, l'utilisation des sens des caractères constitutifs est significativement plus importante dans l'inférence des mots non lexicalisés par rapport aux mots lexicalisés. Cela est probablement lié au fait que les mots chinois non lexicalisés contiennent plus de caractères connus par les participants. La forme des caractères constitutifs du mot a une plus grande importance dans l'inférence pour les mots non lexicalisés que pour les mots lexicalisés ($0,05 < p < 0,1$). Pour un mot chinois inconnu non lexicalisé, un apprenant francophone ne possède pas dans son lexique en français une unité lexicale équivalente liée à un concept. Cette lacune exige logiquement des informations plus riches et plus variées pour que l'apprenant puisse comprendre le sens du mot. Dans le cas où la piste principale, le sens de la phrase contenant le mot, ne fournit pas suffisamment d'informations, l'apprenant doit explorer d'autres pistes et s'appuyer sur des indices parsemés pour élaborer une idée sur le sens du mot. Un caractère chinois peut délivrer lui-même de riches informations lexicales et il a souvent un apport sémantique important au mot qu'il constitue. Il est donc logique que l'apprenant se penche davantage sur le mot lui-même pour chercher des informations.

Nous avons constaté que, pendant l'inférence d'un mot chinois, nos apprenants ayant un niveau intermédiaire en chinois L2 ont intentionnellement exploré au sein de ce mot et de ses caractères constitutifs pour trouver des informations utiles. Cela signifie qu'ils sont conscients que les blocs composants d'un caractère sont porteurs d'informations lexicales et que le sens d'un mot est souvent basé sur les sens de ses caractères constitutifs.

Néanmoins, les apprenants du chinois L2 ayant un niveau intermédiaire ont globalement une faible maîtrise des caractères, aux niveaux de la forme et du sens, ainsi que de la grammaire. Entre autres, **les caractères polysémiques représentent une grande entrave pour les apprenants de niveau intermédiaire dans l'inférence lexicale en chinois L2.** Les apprenants de ce niveau ne tiennent généralement pas compte de la polysémie des caractères. Ainsi, ils se fient généralement au sens le plus courant du caractère activé, sans le vérifier par rapport au contexte où le caractère est utilisé.

Dans notre étude, les apprenants ne se sont pas basés sur les connaissances propres à leur L1 pour inférer un mot chinois inconnu. La grande distance typologique entre le chinois et le français, la L1 des apprenants, implique sans surprise peu de correspondances linguistiques entre ces deux langues, donc peu de connaissances linguistiques acquises en français pouvant servir pour inférer le sens d'un mot chinois inconnu.

**CHAPITRE VI – DEGRE DE REUSSITE DE L'INFERENCE
DES MOTS CIBLES ET ATTENTION PORTEE AUX MOTS CIBLES
PENDANT LA LECTURE**

Introduction

Dans ce chapitre, nous présenterons les résultats de l'inférence pour les 20 mots cibles, ainsi que les résultats d'oculométrie à partir desquels nous examinerons l'attention que nos participants ont portée aux mots cibles lors de leur lecture. Nous cherchons à comprendre si la nature de lexicalisation d'un mot L2 inconnu a un effet 1) sur le degré de réussite d'inférence lexicale du mot et 2) sur l'attention qu'un apprenant francophone porte à ce mot pendant la lecture pour la compréhension. Nous avons émis l'hypothèse qu'un apprenant francophone réussirait mieux l'inférence d'un mot chinois inconnu lexicalisé en français, par rapport à un mot chinois non lexicalisé.

Nous montrerons les scores des inférences observées et explorerons ensuite les facteurs qui pourraient intervenir dans le résultat de l'inférence lexicale, tels que la diversité des sources de connaissance utilisées et l'utilisation d'une source de connaissance spécifique. Nous examinerons, entre autres, l'effet de la lexicalisation des mots cibles sur le degré de la réussite de l'inférence lexicale. Par la suite, nous présenterons les résultats de l'oculométrie et ferons des comparaisons entre les mots cibles et non cibles, ainsi qu'entre les mots lexicalisés et les mots non lexicalisés. Pour finir, nous examinerons le rapport entre les résultats d'oculométrie et les scores d'inférence lexicale.¹⁰³

Nous concluons ce chapitre avec nos réponses aux questions de recherche sur le degré de réussite de l'inférence lexicale, ainsi que sur l'effet de la lexicalisation sur l'attention des apprenants portée aux mots inconnus pendant la lecture.

¹⁰³ Nous avons utilisé le logiciel SPSS pour faire les analyses statistiques des données présentées dans ce chapitre. Nous remercions chaleureusement M. Saïd Jmel pour les aides et les conseils précieux sur les analyses faites avec le logiciel SPSS.

6.1 Résultats d'inférence des mots cibles

6.1.1 Notation des réponses sur le sens

Nous rappelons que nous avons observé au total 654 inférences lexicales. Les réponses des inférences ont été notées selon une échelle de cinq niveaux, 0, 0,25, 0,5, 0,75 et 1 point. Plus précisément, une réponse complètement erronée est notée 0 et une réponse complètement correcte est notée 1 ; une réponse partiellement correcte est notée 0,25, 0,5 ou 0,75 point, en fonction du rapprochement entre la réponse et le bon sens. Par exemple, pour le mot MC04 « 腊月 (*làyuè*, le douzième mois du calendrier lunaire) », la réponse « une certaine période » est notée 0,25 point ; la réponse « un nom d'un mois particulier » est notée 0,5 point et la réponse « un mois lié au Nouvel an chinois » est notée 0,75 point.

Nous avons opté pour une échelle de notation en cinq niveaux afin de montrer la finesse de la compréhension d'un participant pour le sens d'un mot chinois inconnu grâce au contexte, notamment dans le cas d'un mot non lexicalisé en français. Un mot L2 non lexicalisé ne correspond pas à un concept dans le lexique mental en L1. Afin d'expliquer un mot L2 non lexicalisé, il est nécessaire de recourir à la combinaison de plusieurs mots en L1, voire une phrase, comme MC04 montré ci-dessus. Ceci dit, le concept d'un mot L2 non lexicalisé peut comporter plusieurs concepts dans le lexique mental en L1. Par exemple, MC04 « 腊月 (*làyuè*, le douzième mois du calendrier lunaire) » comprend au moins les concepts tels que « mois », « dernier » ou « douzième », « calendrier chinois » ou « calendrier lunaire ». Pour cela, par rapport à la réponse « une certaine période », la réponse « un mois lié au Nouvel an chinois » est bien plus pertinente car elle délivre davantage d'informations étroitement liées au sens du mot. Nous pensions nécessaire de révéler en score le fait que les participants ont compris les sens des mots cibles, notamment ceux non lexicalisés, à différents degrés.

Nous avons sollicité deux natifs chinois bilingues en français pour effectuer une double correction. En cas de désaccord sur le score de certains mots, nous avons opté pour le score le plus voté.

6.1.2 Degré de réussite d'inférence pour les mots cibles

Le score en moyenne pour les 654 inférences observées est de 0,22 point (Tableau 6.1), bien loin du score maximal (1 point). Cela montre que, pour un apprenant de niveau intermédiaire en chinois L2, il est plutôt difficile de deviner le sens d'un mot chinois inconnu présent dans un texte.

Nb d'inférence	Score total	Score moyenne*	Ecart type
654	141,25	0,22	0,32

* Score maximal = 1

Tableau 6.1 – Score d'inférence lexicale

Les scores en moyenne par mot sont détaillés dans la Figure 6.1 ci-dessous.

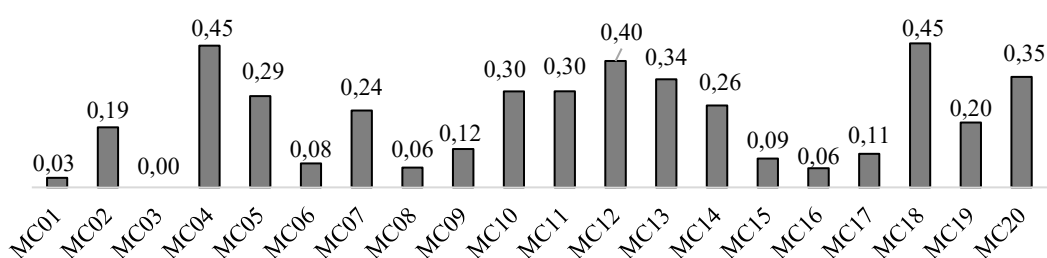


Figure 6.1 – Score d'inférence en moyenne pour les mots cibles

MC01	节俭	(<i>jiéjiǎn</i>) : économe, frugal
MC02	窗花	(<i>chuānghuā</i>) : papier découpé pour la décoration d'une fenêtre
MC03	送灶	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12 ^e lune
MC04	腊月	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC05	汤圆	(<i>tāngyuán</i>) : boulette de riz glutineux farcie de viande (fourrée à la confiture)
MC06	雀跃	(<i>quèyuè</i>) : sauter de joie
MC07	还礼	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC08	在乎	(<i>zàihu</i>) : prendre qch. à cœur ; faire attention à
MC09	乏味	(<i>fáwèi</i>) : fade, insipide
MC10	盼望	(<i>pànwàng</i>) : espérer, désirer
MC11	安逸	(<i>ānshì</i>) : tranquille et confortable
MC12	五颜六色	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC13	明丽	(<i>mínglì</i>) : clair et beau
MC14	彩虹	(<i>cǎihóng</i>) : arc-en-ciel
MC15	踏青	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps

MC16	风筝	(<i>fēngzhēng</i>) : cerf-volant
MC17	蝉	(<i>chán</i>) : cigale
MC18	闲聊	(<i>xiánliáo</i>) : bavarder
MC19	飘零	(<i>piāolíng</i>) : tomber en voltigeant
MC20	白茫茫	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

Les mots pour lesquels les participants ont le meilleur score d'inférence sont

MC04	腊月	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC12	五颜六色	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC13	明丽	(<i>mínglì</i>) : clair et beau
MC18	闲聊	(<i>xiánliáo</i>) : bavarder
MC20	白茫茫	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

Les mots pour lesquels les participants ont le score le plus faible sont

MC01	节俭	(<i>jiéjiǎn</i>) : économe, frugal
MC03	送灶	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC06	雀跃	(<i>quèyuè</i>) : sauter de joie
MC08	在乎	(<i>zàihu</i>) : prendre qch. à cœur ; faire attention à
MC16	风筝	(<i>fēngzhēng</i>) : cerf-volant

Par rapport aux scores du test VKS1, les scores d'inférence lexicale pour les mots MC02, MC04, MC05, MC07, MC10, MC11, MC12, MC13, MC14, MC18, MC19 et MC20 ont eu une augmentation significative ($p < 0,05$)¹⁰⁴. Les 654 inférences lexicales observées concernent les mots inconnus pour les participants avant la séance d'inférence lexicale. La majoration significative du score à l'inférence pour les douze mots cibles signifie que les participants ont réussi, complètement ou partiellement, à comprendre ces mots. En d'autres termes, l'inférence lexicale a un effet positif sur la compréhension de ces douze mots inconnus dans un texte.

En revanche, par rapport aux scores au test VKS1, les scores d'inférence des huit autres (MC01, MC03, MC06, MC08, MC09, MC15, MC16 et MC17) mots sont

¹⁰⁴ Nous avons utilisé le modèle mixte linéaire (logiciel SPSS) pour examiner les effets des mesures répétée (le temps, le groupe et la lexicalisation) sur les scores du sens des mots cibles. Les résultats de comparaison du score du sens entre VKS1 et l'inférence présentés ci-dessous sont extraits des résultats d'analyse par modèle mixte linéaire que nous présenterons dans le Chapitre VII..

peu différents. Cela signifie que, malgré une inférence entreprise, les apprenants n'ont pas pu comprendre ce que ces huit mots voulaient dire dans les textes.

Comparaisons appariées^a

Mot	(I) Phase	(J) Phase	Différence moyenne (I-J)	Erreur standard	ddl	Sig.c	Intervalle de confiance à 95 % pour la différence	
							Borne inférieure	Borne supérieure
MC01	VKS1	Inférence	-0,029	0,065	2424,842	1,000	-0,202	0,143
MC02	VKS1	Inférence	-,176*	0,065	2424,825	0,042	-0,349	-0,004
MC03	VKS1	Inférence	1,67E-16	0,065	2424,859	1,000	-0,172	0,172
MC04	VKS1	Inférence	-,449*	0,065	2424,842	0,000	-0,621	-0,276
MC05	VKS1	Inférence	-,281*	0,065	2424,842	0,000	-0,453	-0,108
MC06	VKS1	Inférence	-0,078	0,065	2424,825	1,000	-0,25	0,095
MC07	VKS1	Inférence	-,244*	0,065	2424,842	0,001	-0,416	-0,072
MC08	VKS1	Inférence	-0,047	0,066	2424,844	1,000	-0,222	0,128
MC09	VKS1	Inférence	-0,123	0,065	2424,842	0,357	-0,296	0,049
MC10	VKS1	Inférence	-,246*	0,065	2424,859	0,001	-0,418	-0,073
MC11	VKS1	Inférence	-,289*	0,065	2424,842	0,000	-0,461	-0,116
MC12	VKS1	Inférence	-,377*	0,066	2424,848	0,000	-0,552	-0,202
MC13	VKS1	Inférence	-,321*	0,065	2424,859	0,000	-0,493	-0,148
MC14	VKS1	Inférence	-,260*	0,07	2424,834	0,001	-0,444	-0,075
MC15	VKS1	Inférence	-0,09	0,065	2424,842	1,000	-0,262	0,083
MC16	VKS1	Inférence	-0,061	0,065	2424,842	1,000	-0,233	0,112
MC17	VKS1	Inférence	-0,099	0,065	2424,825	0,783	-0,271	0,074
MC18	VKS1	Inférence	-,413*	0,065	2424,842	0,000	-0,585	-0,24
MC19	VKS1	Inférence	-,205*	0,065	2424,863	0,010	-0,377	-0,033
MC20	VKS1	Inférence	-,341*	0,065	2424,833	0,000	-0,514	-0,169

Basées sur les moyennes marginales estimées

*. La différence moyenne est significative au niveau ,05.

Variante dépendante : Inférence Score.

c. Ajustement pour les comparaisons multiples : Bonferroni.

6.1.3 Rapport entre le degré de réussite et les sources de connaissance utilisées

Le Tableau 6.2 plus bas présente le nombre d'inférences classé selon la diversité des sources de connaissance que nos participants ont utilisées. Par exemple, dans 109 inférences parmi les 654 inférences observées, les participants ont utilisé une seule source de connaissance et le score en moyenne pour ces 109 inférences est de 0,11 ; les participants ont utilisé trois sources de connaissance différentes dans 170 inférences observées et le score en moyenne pour ces derniers est de 0,29.

Selon le Tableau 6.2, globalement les scores d'inférence en moyenne augmentent en fonction de la diversité des sources de connaissance utilisées. En observant la relation entre les scores d'inférence et les sources de connaissance que les participants ont utilisées pour inférer les mots cibles, nous avons trouvé un effet statistiquement significatif de la diversité des sources de connaissance utilisées sur

les scores d'inférence ($p = 0,00$), selon l'analyse univariée par modèle linéaire général.

Variable dépendante : Inférence_Score

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	4,729a	1	4,729	48,022	,000
Ordonnée à l'origine	,038	1	,038	,386	,535
DiversitéSC	4,729	1	4,729	48,022	,000
Erreur	64,202	652	,098		
Total	99,438	654			
Total corrigé	68,931	653			

a. R deux = ,069 (R deux ajusté = ,067)

Cela montre que la diversité des sources de connaissance utilisées a un effet positif sur le degré de réussite de l'inférence. Un grand écart du score se trouve entre les inférences utilisant deux sources de connaissance (0,17 point) et celles utilisant trois sources de connaissance (0,29 point) : le score en moyenne augmente de plus de 70%.

Diversité SC	Nb d'observ.	Score d'inférence lexicale		
		Somme	Moyenne*	Ecart-type
0	8	0,00	0,00	0,00
1	109	12,75	0,12	0,24
2	274	46,50	0,17	0,30
3	170	49,75	0,29	0,35
4	74	23,00	0,31	0,38
5	17	8,25	0,49	0,41
6	2	1,00	0,50	0,00
Total	654	141,25	0,22	0,32

* Score maximal = 1

Tableau 6.2 – Nombre des inférences observées et score d'inférence, classés selon la diversité des sources de connaissance utilisées

Selon les analyses univariées par modèle linéaire général, parmi les sources de connaissance majeures employées par nos participants pour inférer les mots cibles, c'est-à-dire SC05 (« Contexte proche »), SC06 (« Grammaire »), SC08 (« Sens d'un caractère constitutif du mot ») et SC09 (« Forme d'un caractère constitutif d'un mot ») (cf. la Figure 5.4, p. 173), seul l'effet principal de l'utilisation de SC08 (« Sens d'un caractère constitutif du mot ») est significatif ($p = 0,00$) sur les scores d'inférence.

Aucune interaction n'est significative parmi les quatre sources de connaissance majeures.

Variable dépendante : Inférence_Score

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	9,514a	15	,634	6,810	,000
Ordonnée à l'origine	5,342	1	5,342	57,364	,000
SC05	,221	1	,221	2,377	,124
SC06	,092	1	,092	,984	,322
SC08	2,232	1	2,232	23,962	,000
SC09	,206	1	,206	2,214	,137
SC05 * SC06	,007	1	,007	,079	,779
SC05 * SC08	,120	1	,120	1,293	,256
SC05 * SC09	,046	1	,046	,490	,484
SC06 * SC08	,008	1	,008	,081	,777
SC06 * SC09	,159	1	,159	1,707	,192
SC08 * SC09	,147	1	,147	1,575	,210
SC05 * SC06 * SC08	,091	1	,091	,978	,323
SC05 * SC06 * SC09	,149	1	,149	1,599	,206
SC06 * SC08 * SC09	,147	1	,147	1,577	,210
SC05 * SC06 * SC08 * SC09	,169	2	,085	,908	,404
Erreur	59,417	638	,093		
Total	99,438	654			
Total corrigé	68,931	653			

a. R deux = ,138 (R deux ajusté = ,118)

Plus précisément, parmi les 654 inférences observées, les participants ont utilisé SC08 dans 304 inférences dont le score en moyenne est de 0,33, supérieur à celui pour les 350 autres inférences n'ayant pas engagé SC08 dont le score en moyenne est de 0,07.

Variable dépendante : Inférence_Score

SC08	Moyenne	Erreur standard	Intervalle de confiance à 95%	
			Borne inférieure	Limite supérieure
Non utilisée	,070	,026	,020	,121
Utilisée	,328	,046	,238	,417

La Figure 6.2 ci-dessous montre, selon le score obtenu, l'utilisation de SC08 (« Sens d'un caractère constitutif d'un mot ») dans les inférences observées. Nous pouvons voir que l'utilisation de SC08 est proportionnellement plus importante dans les inférences dont la réponse est notée 0,5 point, 0,75 point et 1 point, mais moins

importante dans les inférences dont la réponse est notée 0. L'utilisation de SC08 a donc un effet positif sur le résultat de l'inférence des mots cibles.



Figure 6.2 – Utilisation du sens d'un caractère constitutif (SC08) dans les inférences observées, classée selon le score d'inférence

La Figure 6.3 montre la fréquence d'utilisation de la SC08 pour les 20 mots et les scores de ces derniers.

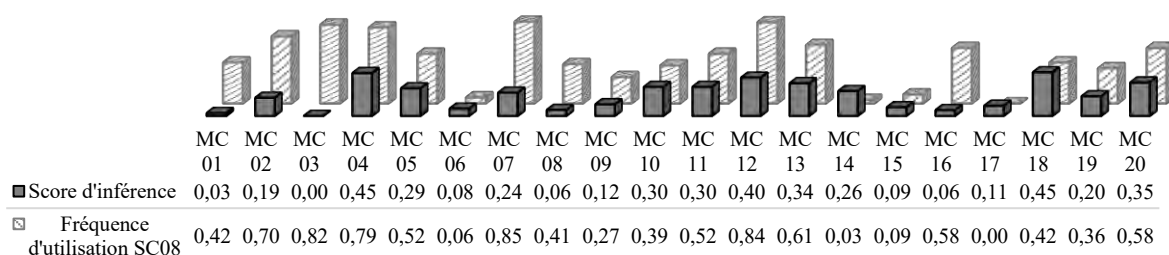


Figure 6.3 – Fréquence d'utilisation du sens d'un caractère constitutif du mot (SC08) et score d'inférence par mot

6.1.4 Comparaison du score entre les mots lexicalisés et non lexicalisés

Le Tableau 6.3 montre le nombre des inférences échouées (« Score 0 »), des inférences partiellement réussies (« Score 0~1 ») et des inférences réussies (« Score 1 »), classés selon la lexicalisation.

	Nb d'observ.	Score 0					Score 0~1		Score 1	
		Cas 1	Cas 2	Cas 3	Total	%	Total	%	Total	%
Mots L	324	24	57	165	246	75,9%	33	10,2%	45	13,9%
Mots NL	330	26	26	108	160	48,5%	158	47,9%	12	3,6%
Total	654	50	83	273	406	62,1%	191	29,2%	57	8,7%

Cas 1 : Abandon Cas 2 : Informations grammaticales seules Cas 3 : Réponse erronée

Tableau 6.3 – Nombre et répartition des inférences observées pour les mots lexicalisés (L) et non lexicalisés (NL), classés selon le score

Les inférences échouées (scores 0) concernent trois types de réponse : 1) le participant a essayé de chercher des indices pour deviner le sens d'un mot, mais il a fini par y renoncer ; 2) le participant a essayé de chercher des indices pour deviner le sens d'un mot ; il a identifié la classe grammaticale du mot sans proposer de sens pour le mot ; 3) le participant a cherché des indices et puis proposé un sens pour le mot, mais la réponse est incorrecte. Proportionnellement, pour les mots lexicalisés, les participants ont fait plus d'inférences échouées (score 0). Plus précisément, la différence entre le nombre d'abandon pour les mots lexicalisés et celui pour les mots non lexicalisés est minime. Mais les participants ont pu élaborer une idée sur le sens du mot inféré pour les mots lexicalisés, même si cette idée est erronée.

En dehors de cela, les participants ont fait plus d'inférences partiellement réussies (score 0~1) pour les mots non lexicalisés (47,6%) que pour les mots lexicalisés (10,8%), mais plus d'inférences réussies (score 1) pour les mots lexicalisés (13,3%) que pour les mots non lexicalisés (3,6%).

La comparaison des scores en moyenne montre que les participants ont plus réussi l'inférence pour les mots non lexicalisés que pour les mots lexicalisés (Tableau 6.4). Le score en moyenne pour les mots lexicalisés est inférieur à celui pour les mots non lexicalisés de 0,06 point.

	Nb d'observ.	Somme	Moyenne *	Ecart type
Mots L	324	60,25	0,19	0,36
Mots NL	330	81	0,25	0,28
Total	654	141,25	0,22	0,32

* Score maximal = 1

Tableau 6.4 – Score d'inférence pour les mots lexicalisés (L) et non lexicalisés (NL)

Le Tableau 6.5 montre les scores d'inférence en moyenne pour les mots lexicalisés et les mots non lexicalisés pour les 33 participants.

Participant	Score d'inférence en moyenne*	
	Mots L	Mots NL
P01	0,43	0,43
P02	0,20	0,35
P03	0,00	0,13

P04	0,00	0,18
P05	0,00	0,05
P06	0,55	0,10
P07	0,25	0,33
P08	0,18	0,33
P09	0,05	0,03
P10	0,40	0,40
P11	0,03	0,08
P12	0,00	0,13
P13	0,18	0,35
P14	0,00	0,10
P15	0,25	0,35
P16	0,20	0,20
P17	0,00	0,03
P18	0,18	0,20
P19	0,20	0,23
P20	0,00	0,13
P21	0,00	0,03
P22	0,22	0,30
P23	0,13	0,23
P24	0,00	0,28
P25	0,05	0,20
P26	0,33	0,35
P27	0,00	0,03
P28	0,25	0,40
P29	0,40	0,33
P30	0,35	0,53
P31	0,36	0,45
P32	0,44	0,38
P33	0,69	0,55

* Score maximal = 1

Tableau 6.5 – Score d’inférence classé par lexicalisation pour les participants

L’analyse par Test-T apparié (bilatéral)¹⁰⁵ des scores d’inférence pour les mots lexicalisés et non lexicalisés montre que la différence entre les deux catégories est significative ($p < 0,05$).

	Moyenne	Ecart-type	Erreur standard moyenne	Intervalle de confiance 95% de la différence		t	ddl	Sig. (bilatérale)
				Inférieure	Supérieure			
L - NL	-,05667	,12477	,02172	-,10091	-,01243	-2,609	32	,014

¹⁰⁵ Le test-T pour échantillons appariés compare la moyenne de deux variables pour un seul groupe, dans notre cas les scores d’inférence pour les mots lexicalisés et non lexicalisés obtenus par les mêmes participants (cf. <https://www.ibm.com/docs/fr/spss-statistics/24.0.0?topic=tests-paired-samples-test>, consulté le 9 novembre 2020).

Les scores d'inférence pour les 20 mots cibles, classés selon la lexicalisation, sont présentés dans la Figure 6.4.

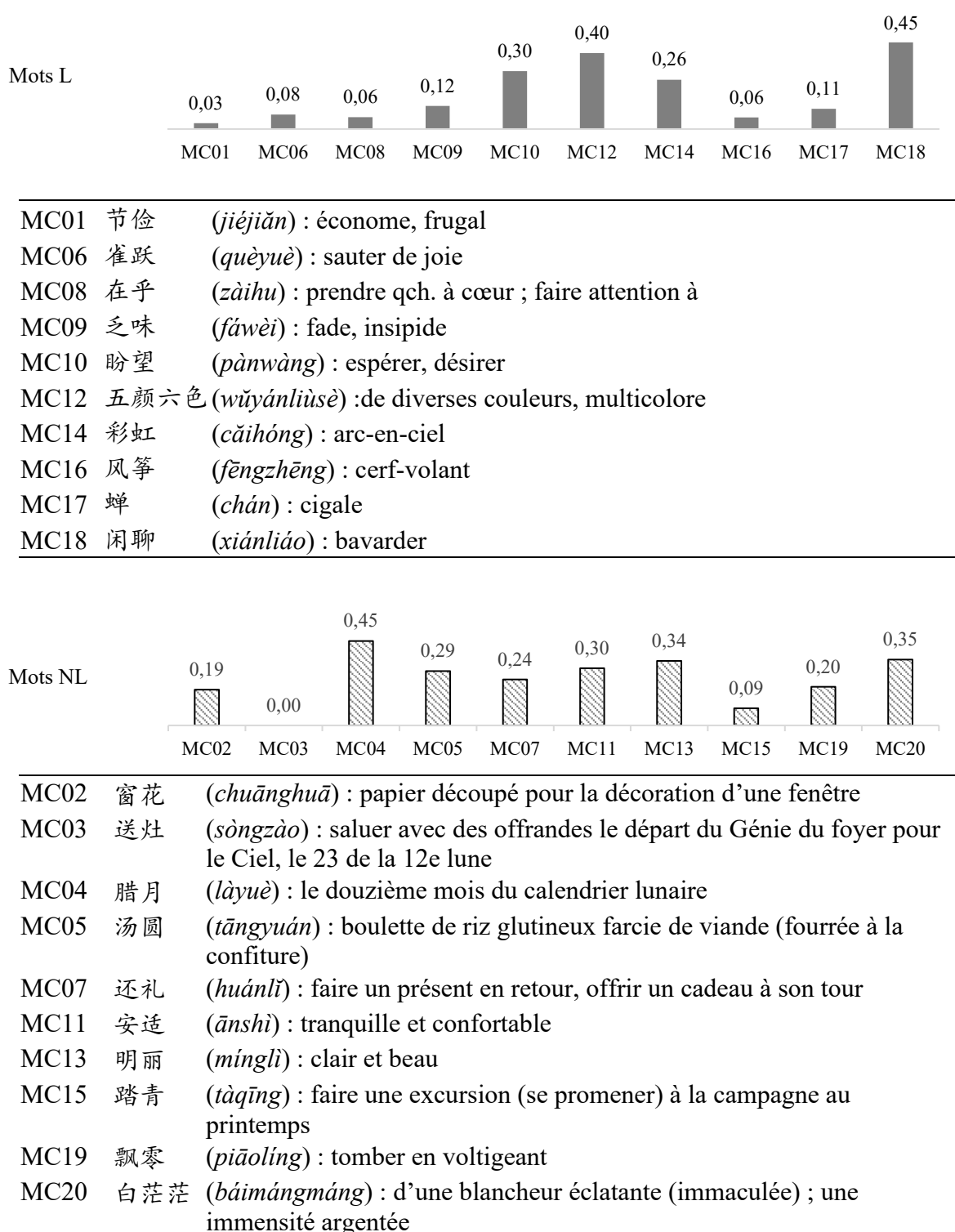


Figure 6.4 – Score d'inférence pour les mots lexicalisés (L) et non lexicalisés (NL)

6.1.5 Discussion

6.1.5.1 Degré de réussite d'inférence

En moyenne, nos participants ont obtenu un score de 0,22 pour les 654 inférences lexicales observées, un score modeste qui montre que pour les apprenants du niveau intermédiaire en chinois L2, la déduction du sens d'un mot chinois inconnu grâce au contexte n'est pas facile.

Nos participants ont réussi à inférer le sens des 20 mots cibles à des degrés différents. Les participants ont réussi à comprendre, complètement ou partiellement, les sens des mots MC02, MC04, MC05, MC07, MC10, MC11, MC12, MC13, MC14, MC18, MC19 et MC20 dans les textes. En revanche, les participants n'ont pas réussi à comprendre les mots cibles MC01, MC03, MC06, MC08, MC09, MC15, MC16 et MC17, malgré une tentative d'inférence.

6.1.5.1.1 Inférences les mieux réussies

Les participants ont eu les meilleurs scores d'inférence pour les mots MC04, MC12, MC13, MC18 et MC20. Nous remarquons que ces mots contiennent chacun des caractères sémantiquement connus par nos participants. En outre, chacun des mots se trouvait dans un contexte riche d'informations propice à l'inférence du sens.

- MC04 « 腊月 (*làyuè*, le douzième mois du calendrier lunaire) »

Ce mot contient le caractère « 月 (*yuè*, mois) » qui est l'un des caractères chinois les plus courants. La phrase contenant le mot est « A partir du premier jour de MC04 ». « A partir de » et « le premier jour » peuvent aussi faire facilement comprendre les participants que ce mot concerne une période spécifique.

Ex. (MC04 « 腊月 ») : *Le premier caractère, je le connais pas, et le deuxième, c'est « yue », de ... « mois », <((Le participant lit quelque chose dans le texte.))> « di yi tian », donc ... les enfants du euh premier jour. Mais je pense que c'est en rapport avec le mois, ... parce que c'est le calendrier lunaire. ... donc peut-être c'est calculé en fonction du calendrier lunaire. ... je pense que c'est en rapport avec ... une date, un mois ou ... avec le temps, je pense.*

- MC13 « 明丽 (*mínglì*, clair et beau) »

Les deux caractères de MC13 sont connus des participants. Beaucoup de participants se rappellent de « 丽 (*lì*, beau) » dans le mot « 美丽 (*měilì*, beau, joli) ». Dans le contexte est décrit le beau paysage et le printemps pluvieux du village natal du narrateur.

Ex. (MC13 « 明丽 ») : *Parce que le « li », il veut dire « beau » ... « ming », ... on le connaît en tant que « demain ». Mais c'est aussi l'association du « soleil » et de la « lune ». Donc ça veut potentiellement dire la clarté, ... Alors, en lisant la partie, la phrase. ... [L]e caractère suivant pour moi qui veut dire « changer ». ... dans le sens « modifier », « transformer ». Donc pour moi, ça voudrait dire que chaque fois qu'il pleut, le paysage se transforme à quelque chose de toujours plus beau.*

- MC18 « 闲聊 (*xiánliáo*, bavarder) »

Ce mot contient le caractère « 聊 (*liáo*, causer, communiquer) » dont certains participants connaissent le sens. Dans le contexte est décrit l'ambiance conviviale dans le quartier où a grandi le narrateur : « Après dîner, les gens sortaient de chez eux. Les adultes s'asseyaient ensemble MC18 et les enfants jouaient à côté. ». Pour des participants qui ne connaissent pas le caractère « 聊 (*liáo*, causer, communiquer) », ce contexte a fourni un indice aidant à deviner le sens de MC18.

Ex. (MC18 « 闲聊 ») : *[J]e reconnais le deuxième caractère, c'est dans un mot qui veut dire « bavarder »... [C]'est quelque chose que les adultes font euh... enfin ils s'assoient et ils font de choses ensemble. Je pense que c'est « discuter » ... quelque chose comme ça.*

- MC12 « 五颜六色 (*wǔyánliùsè*, de diverses couleurs, multicolore) »

Ce mot contient des caractères « 五 (*wǔ*, cinq) », « 六 (*liù*, six) » et « 色 (*sè*, couleur) » qui sont connus des participants. Dans le contexte est décrit le paysage au printemps du village natal du narrateur. La phrase contenant MC12 est « Aux bords des routes il y a de petites fleurs MC12. » MC12 se trouve devant la particule « 的 (de) » et le groupe nominal « 小花儿 (*xiǎohuār*, les petites fleurs) ». Beaucoup de

participants ont indiqué que le rôle de MC12 est de qualifier « 小花儿 (xiǎohuā'r, petites fleurs) ».

Ex. (MC12 « 五颜六色 ») : *[J]e reconnaissais le mot « wu », donc « cinq ». Euh, le ... je dirais « liu ». Je crois que c'est « six ». Et celui entre les deux, je ne connais pas. Et là ... je crois qu'il se prononce « se ». Je sais que c'est « les couleurs ». Et alors je sais qu'à ce moment-là, ... il décrivait le printemps. ... Il décrivait un peu le paysage en gros. ... [L]es fleurs donc ont repoussé. Des petites fleurs qui repoussent, et euh... avaient des, plusieurs couleurs.*

- MC20 « 白茫茫 (báimángmáng, d'une blancheur éclatante (immaculée) ; une immensité argentée) »

Ce mot contient le caractère « 白 (bái, blanc) ». Dans le contexte, le narrateur décrit le paysage enneigé du village natal en hiver. Beaucoup de participants ont deviné que le MC20 signifie « beaucoup de blanc » ou « le blanc qui recouvre ».

Ex. (MC20 « 白茫茫 ») : *Alors, « bai », ça me fait penser à « blanc ». ... il y avait ... au-dessus des arbres, sur l'herbe, ... de plein de choses. Il y avait de la neige partout, même la montagne. ... Donc, pour moi il décrit la neige. En plus, il y a le « de », qui justement renforce l'idée qui décrit comment est la neige. Pour moi, elle est très, très blanche.*

6.1.5.1.2 Inférences les moins réussies

Les cinq mots pour lesquels l'inférence est la moins réussie sont MC01, MC03, MC06, MC08 et MC16. Nous avons présenté dans le Chapitre V (cf. la Section 5.3.3.2) le problème de la polysémie des caractères chinois qui ont entravé l'inférence des mots cibles : beaucoup de caractères chinois sont polysémiques, parfois polyphonétiques (par exemple, « 还 » dans MC07 « 还礼 (huánlǐ, faire un présent en retour, offrir un cadeau à son tour) » peut aussi prononcer « hái » pour le sens de « encore »), mais nos participants ayant un niveau intermédiaire n'en sont pas tous conscients et ne pensent pas systématiquement à vérifier s'ils ont correctement identifié le sens adéquat du caractère. MC01 « 节俭 (jiéjiǎn, économe, frugal) » et MC08 « 在乎 (zàihu, prendre qch. à cœur ; faire attention à) » contiennent

respectivement « 节 » et « 在 », les caractères dont nos participants connaissent la forme, mais non le sens approprié. Nous pensons que cela a gêné nos participants dans leur compréhension de ces deux mots. (Nous reviendrons sur le cas de MC01 dans la Section 6.1.5.2.1.)

Quant à MC06 « 雀跃 (*quèyuè*, sauter de joie) », ce mot contient deux caractères complexes dans leur forme, inconnus pour nos participants. Il se trouve dans une phrase qui ne fournit pas beaucoup d'informations propices à la compréhension du mot (« 当然, 过年时最让孩子们雀跃的事儿还是拿红包。 Bien sûr, ce qui MC06 le plus les enfants était de recevoir des hongbao. »). D'autant que, ainsi que nous l'avons mentionné dans le Chapitre V (cf. la Section 5.3.3.3) le fait que MC06 se trouve dans une proposition relative introduite par la particule « 的 (de) » a perturbé nos participants dans leur analyse de la phrase contenant le mot.

Dans le cas de MC03 « 送灶 » (*sòngzào*, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12^e lune), des connaissances socio-culturelles sont nécessaires pour la compréhension du mot.

Nous avons constaté, à notre surprise, que nos participants ont peu réussi l'inférence pour le mot MC 16 « 风筝 (*fēngzhēng*, cerf-volant) ». Il y a deux indices explicites pour MC16, 1) le premier caractère « 风 (*fēng*, vent) » que les participants connaissent bien et 2) la phrase contenant le mot (« 要是风比较大的话, 还可以去小城旁边的山上放风筝。 Quand il y a du vent, on peut aller faire du MC16 dans les collines proches de la ville. ») met l'accent sur le contexte – « quand il y a du vent ». Certains participants ont compris le contexte, pourtant cet indice ne les a pas suffisamment aidés pour comprendre ce mot.

Ex. (MC16 « 风筝 ») : *Je sais pas trop ce que ça peut être, mais je pense que c'est un endroit ou... puisqu'il parle des gens qui y vont quand il y a beaucoup de vent... là ce que j'ai compris, c'est la montagne à côté de la ville. ... Je pense que c'est un endroit où i-... peut-être encore une activité. Je sais pas trop.*

Il est plausible que le cerf-volant soit une activité moins répandue en France et que nos participants ne soient pas bien familiers avec cette activité. Une autre raison

possible est liée au verbe « 放 (fàng) » immédiatement devant MC16. Beaucoup de participants ont reconnu et fait attention à ce verbe lors de l'inférence. Ils connaissent ce verbe pour le sens « mettre, déposer », par exemple « 把书放在桌上 (mettre/déposer le/les livre(s) sur la table) ». En fait, ce caractère fait partie d'un autre verbe « 放心 (fàngxīn, se tranquilliser, se rassurer) » que les participants ont appris dans leurs manuels. Dans le mot « 放心 (fàngxīn, se tranquilliser, se rassurer) », le sens de « 放 (fàng) » est plutôt « relâcher, relaxer ; libérer »¹⁰⁶. Lors de l'inférence, un seul participant a mentionné le mot « 放心 (fàngxīn, se tranquilliser, se rassurer) ». Ex. (MC16 « 风筝 ») : *Alors, « fang », je le connais le sens de « poser ». ... je sais que par exemple il y a euh <((Le participant dit un mot chinois.))> « fang xin », c'est euh « poser le cœur », littéralement ça veut dire « se relaxer », enfin « se calmer ».*

Nous pensons que, pour nos participants, le sens « mettre, déposer » du verbe « 放 (fàng) » est mieux maîtrisé. Par conséquent, ce sens a été activé le premier. En outre, « 放 (fàng) » forme une collocation avec « 风筝 (fēngzhēng, cerf-volant) » en chinois, qui correspondant à la collocation française « faire du cerf-volant ». Le fait qu'en français le verbe usuel est « faire » pourrait aussi être une entrave dans la compréhension de MC16 pour nos participants.

6.1.5.2 Facteurs intervenant dans la réussite d'inférence

6.1.5.2.1 Choix de sources de connaissance

Les résultats de notre étude montrent que les participants ont réussi mieux les inférences pour lesquelles ils ont utilisé plusieurs sources de connaissance plutôt qu'une seule. Cependant, le rapport entre le degré de réussite et la diversité des sources de connaissance utilisées n'est pas étroitement linéaire. Une grande différence du score d'inférence est observée entre d'une part les inférences pour lesquelles 2 sources de connaissance ont été utilisées (0,17 point), et d'autre part les inférences pour lesquelles 3 sources de connaissance ont été utilisées (0,29 point),

¹⁰⁶ 放 (fàng, relâcher, relaxer ; libérer) + 心 (xīn, cœur) → 放心 (fàngxīn, se tranquilliser, se rassurer)

soit une majoration de 70,6%.

Le sens d'un caractère constitutif du mot (SC08) est la seule source de connaissance ayant un effet positif significatif sur le degré de réussite d'inférence.

Les indices provenant d'une seule source sont souvent partiels et parfois incorrects. La combinaison de divers indices peut fournir à l'apprenant une vue plus globale sur les aspects lexicaux d'un mot inconnu. Grâce à cela, l'apprenant a plus de possibilités de vérifier les informations repérées. Nous avons montré dans le Chapitre V que, grâce au contexte et aux indices grammaticaux, deux participants ont pu correctement identifier le premier caractère du mot MC07 « 还礼 (*huánlǐ*, faire un présent en retour, offrir un cadeau à son tour) », « 还 » qui peut aussi signifier « encore » (cf. supra p. 190).

Toutefois, la combinaison de plusieurs sources d'indices ne conduit pas toujours à un meilleur résultat, surtout quand certaines informations repérées sont fausses. Voici deux exemples dans lesquels les participants infèrent incorrectement MC01 « 节俭 (*jiéjiǎn*, économe, frugal) » :

Ex. 1 (MC01 « 节俭 ») : *Alors, dans la phrase, déjà on parle du Nouvel an chinois.*

Dans la phrase, la partie après ..., ils disent que les gens dépensent beaucoup d'argent pendant le Nouvel an. ... dans la première partie de la phrase, ... <((Le participant lit quelque chose en chinois.))> « bu tai hao », donc ça veut dire que « c'est pas très bien ». Je pense que, ... peut-être. ... pendant la préparation du Nouvel an, ... ils vont devoir dépenser beaucoup de sous pour faire la nourriture et tout ça. ... Après, je pense à « festif » parce qu'il y a le « jie » dedans. ... Donc je pense que là il fait une comparaison avec les adultes, ... ils sont « occupés » par les préparatifs quoi.

Ex. 2 (MC01 « 节俭 ») : *<((Le participant lit quelque chose dans le texte.))> « da jia ping shi dou fei chang » ... Oui, le premier je le connais ... celui qui est dans le <((Le participant dit le mot chinois pour « la fête du Printemps ».))> « chun jie », ... enfin c'est pour les « fêtes » ... peut-être pour les « saisons » aussi. ... Non, c'est pour les « fêtes ». Euh donc il désigne quelque chose, <((Le participant lit un mot dans le texte.))> « fei chang », « très ». ... J'aurais encore*

dit que c'est un adjectif, ..., vu que la famille elle est très ... « joyeuse » ou « festive ». ... à la première partie de la phrase ... elle est plutôt négative, parce qu'on dit que « en ce moment, l'économie était pas très bonne ». Donc toute la famille ... on a quand-même dépensé beaucoup d'argent. Donc finalement c'est ... pas aussi positif que ce que je le pensais. ... beaucoup dépense, l'économie n'était pas très bonne.

Le premier caractère « 节 » dans MC01 peut aussi signifier « fête » ou « saison ». Le thème du texte porte sur les coutumes lors du Nouvel an chinois en Chine. En quelque sorte, le contexte décrivant l'ambiance du Nouvel an a fait penser aux deux participants que le sens de MC01 « 节俭 (*jiéjiǎn*, économe, frugal) » est lié au sens « fête ». Cependant, un autre indice contextuel, « 不太好 (*bú tài hǎo*, pas très bien) », semble contradictoire à l'ambiance festive. Les participants dans ces deux cas ont alors été désorientés. Ils ont terminé l'inférence pour MC01 par une réponse erronée ou par l'abandon.

Nous avons constaté chez nos participants une entrave importante dans la polysémie des caractères chinois. La reconnaissance de la forme d'un caractère chinois polysémique n'aide pas les participants à extraire des informations correctes pour inférer le sens d'un mot. Bien au contraire, une identification erronée d'un caractère désoriente les participants dans leurs collectes, tris et vérifications d'information. Cela conduit souvent à une réponse incorrecte ou à un abandon de la stratégie de l'inférence lexicale.

6.1.5.2.2 Effet de la lexicalisation

Nous avons vu que parmi les douze mots cibles pour lesquels l'inférence est complètement ou partiellement réussie, huit mots sont non lexicalisés. Parmi les cinq mots cibles pour lesquels l'inférence est la mieux réussie, trois mots sont non lexicalisés et deux autres sont lexicalisés ; parmi les cinq mots cibles pour lesquels l'inférence est la moins réussie, quatre mots sont lexicalisés et un seul non lexicalisé.

Le taux d'abandon d'inférence respectivement pour les mots lexicalisés (7,4%) et pour les mots non lexicalisés (7,8%) sont proches.

En moyenne, nos participants ont eu un meilleur score d'inférence pour les mots non lexicalisés que pour les mots lexicalisés (0,25 vs 0,19). Cette différence est statistiquement significative. Ce résultat est opposé aux résultats obtenus par d'autres études réalisées en anglais L2 (L. Fan & Zhang, 2014; Paribakht, 2005; Paribakht & Tréville, 2007) qui ont montré communément qu'il est plus difficile pour un apprenant L2 de comprendre un mot qui n'est pas lexicalisé dans la L1 de l'apprenant.

Afin de comprendre ces résultats discordants, nous avons cherché des explications au niveau 1) des mots cibles utilisés dans notre étude et 2) du choix de sources de connaissance de nos participants pour inférer les mots cibles.

Nous rappelons que, par rapport aux mots lexicalisés, les mots non lexicalisés sélectionnés dans notre expérience contiennent plus de caractères sémantiquement connus par les participants. En effet, les participants ont bénéficié, lors de l'inférence, d'un appui plus important sur le sens des caractères constitutifs (SC08) pour inférer les mots non lexicalisés (23,8%), par rapport aux mots lexicalisés (14,9%) (cf. la Section 5.3.2). Parmi toutes les sources de connaissance utilisées pour inférer les mots cibles, le sens des caractères constitutifs d'un mot (SC08) a un effet positif sur le degré de réussite de l'inférence. En d'autres termes, plus un apprenant connaît le sens des caractères constitutifs d'un mot inconnu, mieux il réussit l'inférence du sens du mot.

Un mot chinois est souvent constitué de plusieurs caractères. Si le mot est sémantiquement transparent (c'est-à-dire si le sens d'un mot est étroitement lié aux sens de ses caractères constitutifs) et l'apprenant connaît les sens appropriés de ces caractères constitutifs, il parvient à comprendre, complètement ou partiellement, le sens du mot, indépendamment du statut lexicalisé ou non du mot en question dans sa L1. Cela expliquerait le fait que parmi les cinq mots cibles dont l'inférence a été la mieux réussie (MC04, MC12, MC13, MC18 et MC20), trois sont non lexicalisés (MC04, MC13 et MC20).

Voici deux autres exemples où les participants infèrent deux mots non lexicalisés, MC02 « 窗花 (*chuānghuā*, papier découpé pour la décoration d'une fenêtre) » et MC05 « 湯圓 (*tāngyuán*, boulette de riz glutineux farcie de viande

(fourrée à la confiture)) » :

Ex. (MC02 « 窗花 ») : *Donc beaucoup de gens utilisent du papier rouge pour faire plusieurs, je ne sais pas, ... à l'intérieur de maison d'être très beau. Donc, je pense que c'est quelque chose liée à la décoration. ... le deuxième caractère ... je crois que c'est « une fleur », « hua ». Et ... le premier caractère, ... c'est la clé du « cave » ou du « grenier ». ... et après il y a <<(Le participant dit un mot dans le texte.)> « piao ling¹⁰⁷ ». Donc c'est « joli ». Donc je pense c'est des décorations.*

Ex. (MC05 « 汤圆 ») : *Euh... du coup on parle de nourriture. Il y a le verbe <<(Le participant dit un mot chinois qui signifie « manger ».)> « chi » juste devant. Euh le premier c'est « tang ». C'est « la soupe ». Et juste après c'est « yuan » pour ce qui est « rond ». ... et après on dit euh ... encore d'autres choses à manger. Euh...<<(Le participant lit le mot cible ».)> « tang yuan » ça pourrait peut-être... « tang », ça pourrait être le contenant, et « yuan » peut-être...*

Dans ces deux exemples, les participants se sont inspirés à la fois des caractères constitutifs du mot et du contexte. Ils ont réussi à la fin à avoir une compréhension partielle de ces deux mots qui ne sont pas lexicalisés en français L1.

Dans l'étude de Paribakht (2005), les apprenants iraniens n'ont pas complètement échoué l'inférence pour les mots anglais non lexicalisés. La chercheuse souligne alors dans ses conclusions que, bien qu'un concept existant dans la mémoire d'un apprenant puisse faciliter l'inférence d'un mot inconnu, il n'est pas toujours essentiel pour comprendre le sens d'un mot non lexicalisé qui est conceptuellement familier. Selon Paribakht (2005, p. 730), s'il existe, dans le système conceptuel d'un apprenant, des concepts de composants pour un mot L2 non lexicalisé, ces derniers peuvent permettre à l'apprenant de paraphraser mentalement le sens du mot dans sa L1. Autrement dit, si l'unité lexicale inconnue est traitée à un degré suffisant, cela peut aboutir à la compréhension partielle, voire complète, de ce mot L2 qui n'a pas d'équivalent dans la langue maternelle de l'apprenant.

Or, en comparaison des mots cibles utilisés dans notre étude, qui sont majoritairement des mots chinois composés, les mots cibles anglais utilisés dans les

¹⁰⁷ Le participant s'est trompé sur la prononciation du mot « piao liang ».

études de Paribakht (2005) et de Paribakht & Tréville (2007) sont de types variés (simples, dérivés et composés). Notamment pour les mots anglais simples, tels que « snoop », « sleet » et « tow » (cf. Paribakht, 2005, p. 740), il est difficile pour un apprenant d'en repérer des informations sémantiques. Cela peut expliquer le fait que nos participants ont mieux réussi l'inférence pour les mots chinois inconnus non lexicalisés que ceux lexicalisés, alors que les études en anglais L2 ont donné des résultats opposés.

D'un autre point de vue, puisque la connaissance des caractères constitutifs est une piste d'indice importante dans la compréhension d'un mot chinois inconnu, il serait difficile pour un apprenant L2 de comprendre un mot chinois, lexicalisé ou non, si ce dernier est constitué des caractères inconnus et intégré dans un contexte moins informatif (par exemple MC06 « 雀跃 (*quèyuè*, sauter de joie) », un mot lexicalisé). Cela explique le fait que, dans notre étude, parmi les cinq mots pour lesquels l'inférence est la moins réussie, quatre mots sont lexicalisés.

Cela dit, l'effet de la lexicalisation dans la L1 de l'apprenant n'est pas sans importance. Au sujet des réponses partiellement correctes, les participants en ont eu beaucoup plus pour les mots non lexicalisés que pour les mots lexicalisés (47,9% vs 10,2%). Toutefois, ils ont eu davantage de réponses entièrement correctes pour les mots lexicalisés que pour les mots non lexicalisés (13,9% vs 3,6%). Paribakht (2005) a observé le même phénomène dans son étude auprès d'apprenants iraniens de l'anglais L2 : le degré de la réussite entière de l'inférence pour les mots lexicalisés est presque 3 fois plus important que celui pour les mots non lexicalisés (17% vs 6%). Cela révèle qu'un concept existant dans la L1 de l'apprenant a plutôt favorisé la compréhension entière d'un mot inconnu.

Notre système de notation a éventuellement amplifié la différence entre les scores d'inférence pour les mots lexicalisés et non lexicalisés. Dans les autres études (ex. Paribakht, 2005; Paribakht & Tréville, 2007), une échelle à trois niveaux est utilisée pour noter les réponses, c'est-à-dire « réponse erronée », « réponse partiellement correcte » et « réponse complètement correcte ». Nous avons noté les réponses selon cinq niveaux. Cela permet de noter en détail les compréhensions

partielles à différents niveaux, et une compréhension partielle est plus souvent le cas pour les mots non lexicalisés dans notre étude.

6.1.5.2.3 Connaissances socio-culturelles

Les dix mots non lexicalisés dans notre étude sont intégrés dans deux textes. Le texte A porte sur le Nouvel an chinois et les coutumes traditionnelles. La compréhension du texte A et des mots non lexicalisés y figurant, listés ci-dessous, demande des connaissances socio-culturelles chinoises.

- MC02 « 窗花 (*chuānghuā*, papier découpé pour la décoration d'une fenêtre) »
- MC03 « 送灶 (*sòngzào*, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune) »
- MC04 « 腊月 (*làyuè*, le douzième mois du calendrier lunaire) »
- MC05 « 汤圆 (*tāngyuán*, boulette de riz glutineux farcie de viande (fourrée à la confiture)) »
- MC07 « 还礼 (*huánlǐ*, faire un présent en retour, offrir un cadeau à son tour) »,

Le texte B décrit le paysage en quatre saisons du village natal du narrateur. Les mots non lexicalisés intégrés sont les suivants :

- MC11 « 安适 » (*ānshì*, tranquille et confortable) »
- MC13 « 明丽 » (*mínglì*, clair et beau) »
- MC15 « 踏青 » (*tàqīng*, faire une excursion (se promener) à la campagne au printemps) »
- MC19 « 飘零 » (*piāolíng*, tomber en voltigeant) »
- MC20 « 白茫茫 » (*báimángmáng*, d'une blancheur éclatante (immaculée) ; une immensité argentée) »

Le score d'inférence en moyenne pour les mots non lexicalisés intégrés dans le texte A est inférieur à celui pour les mots non lexicalisés dans le texte B (0,23 point vs 0,26 point). Cela reflète une plus grande difficulté chez nos participants pour comprendre les mots non lexicalisés qui exigent davantage de connaissances non linguistiques.

Entre autres, MC03 « 送灶 (*sòngzào*, saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune) » est le mot pour lequel aucun participant n'a réussi l'inférence. Le deuxième caractère « 灶 (*zào*, fourneau (de cuisine)) » symbolise « 灶王爷 (*Zàowángyé*, génie du foyer) ». Ce dernier, dans les

mythes chinois, surveille les faits et les comportements d'une maisonnée et rapporte une fois par an au Grand Dieu. Aucun de nos participants n'ont eu ces connaissances. Sans explications explicites dans le texte, la compréhension de ce mot a donc été quasiment impossible pour nos participants.

6.2 Attention portée aux mots cibles

Le mouvement des yeux d'un lecteur peut refléter son processus cognitif pendant sa lecture. C'est pour cela que nous avons opté pour l'oculométrie pour mesurer l'attention que nos participants ont portée aux mots cibles pendant leur lecture. Jusqu'à présent, nous n'avons pas trouvé d'articles au sujet de l'utilisation de l'oculométrie dans l'inférence lexicale en chinois L2. Cette partie de notre étude est donc exploratoire.

Nous rappelons que les participants ont lu deux textes, le texte A (318 caractères) et le texte B (321 caractères). Tous les caractères dans les deux textes que les participants ont lus sont définis chacun comme une AOI (« Area of Interest », aire d'intérêt), l'unité de mesure dans l'oculométrie. Nous avons observé, pour chaque AOI, le temps total (« Dwell time », le temps total du regard), le nombre de revisite du regard (« Revisit ») et la première fixation (« First fixation duration », la durée de la première fixation du regard) (cf. la Section 4.3.2.1).

Afin d'interpréter les données de façon plus explicite, nous avons transformé le nombre de revisite du regard au nombre de visite du regard total. Par exemple, pour AOI X, si le nombre de revisite du regard est de 2, le nombre de visite du regard total dans AOI X est alors de 3. Si aucun regard ne s'est passé dans AOI X, le nombre de visite est de 0.

Avant la lecture de chaque texte, nous avons informé les participants que 1) ils pouvaient passer autant de temps qu'ils voulaient pour lire le texte et 2) ils devaient répondre à des questions de compréhension à la fin de la lecture. Le recueil de données oculométriques exige un contrôle minutieux. Un bon nombre d'éléments, par exemple le port de lunettes (y compris les lentilles de contact), le maquillage autour des yeux, les mouvements de la tête ou corporels, peuvent biaiser le suivi du regard. Nous avons donc demandé aux participants de ne pas basculer la tête dans la mesure du possible pendant la lecture, afin d'optimiser le suivi du regard des participants.

SMI Red 250, ce que nous avons utilisé pour recueillir les données, effectuée

un suivi du regard en mode binoculaire. Autrement dit, les mouvements des deux yeux d'un sujet sont suivis individuellement. Par le biais de la fonction « Gaze replay » du logiciel SMI BeGaze, nous avons examiné les parcours des deux yeux de chaque participant pendant la lecture en silence. Parmi les 132 lectures (2 textes x 33 participants x 2 yeux), les données des participants P13, P27, P29 et P33 ont été exclues des données à analyser à cause de la mauvaise qualité. Pour les 29 participants observés, en fonction de la qualité du suivi, nous avons choisi les données de l'œil gauche pour analyser.¹⁰⁸

6.2.1 Vitesse de lecture

Pour les 29 participants observés, le temps de lecture pour le texte A est en moyenne de 4,3 minutes, soit 75 caractères/minute, et de 4,1 minutes pour le texte B, soit 78 caractères/minute. (Tableau 6.6 plus bas)

Selon le *Programme d'enseignement du chinois pour les étudiants étrangers dans les établissements d'enseignement supérieur : études avancées à long terme* (« 高等学校外国留学生汉语教学大纲：长期进修 », 2002), pour un étudiant étranger ayant accompli en Chine le programme d'apprentissage du chinois du niveau débutant, la vitesse de lecture est de 90~110 caractères/minute. Nos participants ont eu une vitesse de lecture plus lente parce qu'ils ont été informés avant la lecture qu'ils pouvaient passer autant de temps qu'ils voulaient pour lire les textes. Quelques participants ont relu une ou plusieurs fois les textes (par exemple P17). En outre, il est possible que la mise en page espacée des textes sur l'écran, c'est-à-dire les caractères et les lignes plus écartées (dans l'objectif d'obtenir des données du suivi plus précises, cf. la Section 4.3.2.3) ait également ralenti la lecture des participants (Bai et al., 2009; Liang & Bai, 2010)¹⁰⁹.

¹⁰⁸ Nous remercions ici chaleureusement M. Pierre-Vincent PAUBEL et Mme Emilie MASSA pour leur aide précieuse et leurs conseils lors de la mise en place et la sélection de données oculométriques.

¹⁰⁹ Il est à noter que ces deux études ont été conduites auprès des adultes chinois.

Participant	Texte A	Texte B
P01	2,5	2,2
P02	2,0	2,5
P03	3,2	4,6
P04	4,3	5,5
P05	3,5	3,9
P06	3,8	3,7
P07	5,4	6,8
P08	4,6	6,6
P09	3,2	3,9
P10	3,4	5,0
P11	3,3	2,5
P12	7,0	4,6
P14	3,4	4,1
P15	1,7	1,8
P16	3,6	2,7
P17	14,0	8,0
P18	2,7	3,2
P19	4,2	5,3
P20	8,5	3,3
P21	4,7	4,5
P22	3,1	2,1
P23	3,2	4,6
P24	4,7	4,5
P25	3,3	3,4
P26	3,2	2,5
P28	4,5	6,9
P30	4,9	3,8
P31	4,3	2,4
P32	3,1	4,4
Moyenne	4,3	4,1
Ecart-type	2,3	1,6

Tableau 6.6 – Temps de lecture (minute) en moyenne pour les deux textes par participant

6.2.2 Caractères sautés pendant la lecture

Il existe un décalage entre la perception oculaire d'un mot et l'activation des

informations lexicales concernant ce mot dans le cerveau du lecteur. Des études en anglais montrent que, lors d'une lecture, ce décalage chez un natif anglais est autour de 50 millisecondes (Rayner, 1998; Rayner et al., 2003).

En chinois, les conclusions sur le décalage entre la perception oculaire et l'activation d'informations lexicales varient d'une étude à l'autre, mais il est communément admis que, dans un caractère qui est l'union de la forme, du son et du sens (cf. la Section 3.2.2), les trois aspects lexicaux ne s'activent pas simultanément. Selon Chen et al. (2003), la forme du caractère s'active la première chez un lecteur natif à partir de 43 millisecondes après la perception du caractère. W. Wang (2007) indique que, pour un lecteur adulte chinois, une fixation de 60 millisecondes sur un caractère suffit à extraire des informations lexicales de ce dernier afin d'accomplir la compréhension d'un texte. Il est à noter qu'en chinois, le traitement des informations lexicales se produit non seulement au niveau du caractère, mais aussi au niveau du mot, constitué souvent de plusieurs caractères dans le chinois contemporain (cf. la Section 3.2.3). En d'autres termes, le sens des caractères constitutifs d'un mot et le sens de ce mot sont activés tous les deux dans la compréhension écrite (Peng et al., 1999, cité par W. Wang, 2007). Les résultats de W. Wang (2007) montrent qu'une fixation du regard de 80 millisecondes sur un mot chinois (constitué de deux caractères) suffit à accomplir la compréhension d'un texte pour un adulte chinois. Ce seuil est de 50 millisecondes selon la conclusion de Y. H. Liu et al. (2005).

Pour notre étude, nous avons mis le seuil à 50 millisecondes lors de l'extrait de données oculométriques par le biais de la fonction « Metrics export » du logiciel SMI BeGaze. Cela veut dire que, si la durée de fixation sur un caractère est inférieure à 50 millisecondes, nous considérons que le regard de l'apprenant sur ce caractère est trop court afin que l'apprenant puisse entreprendre le traitement d'informations concernant ce caractère. Opérationnellement, toutes les fixations dont la durée est inférieure à 50 millisecondes sont rectifiées à 0 milliseconde.

Selon nos résultats, en moyenne 21% des caractères dans les textes ont reçu un regard de moins de 50 millisecondes. Cela est cohérent avec la conclusion des études qui montrent que les lecteurs ne lisent pas tous les mots présents dans un texte.

Autrement dit, certains mots sont sautés pendant une lecture (« skipping » ou « saut », Rayner et al., 2011). Un lecteur expert en anglais saute au moins 30% des mots présents dans un texte (Angele & Rayner, 2013, p. cité par Zang et al., 2018). Le taux de saut peut être plus important pendant une lecture en chinois pour les raisons suivantes : 1) à l'écrit, les mots chinois sont plus compacts en longueur par rapport à leur équivalent en langue alphabétique (par exemple, « 电视 » en chinois par rapport à « télévision » en français) et 2) comme il n'y a pas d'espace supplémentaire séparant les mots dans une phrase, une phrase en chinois est physiquement plus courte (par exemple, « 他看电视。 » par rapport à « Il regarde la télévision. »). Un lecteur chinois adulte peut sauter jusqu'à 50% des mots présents dans un texte (Zang et al., 2017).

Il n'est pas étonnant que nos participants, ayant un niveau intermédiaire en chinois, aient un taux de saut considérablement moins important par rapport à un Chinois natif. Comme l'indiquent Zang et al., (2018), quand un lecteur a plus de difficultés à traiter les informations lexicales des caractères/mots présentés dans un texte, il passe plus de temps pour lire et le taux du saut de mots diminue. Les résultats de Fu (2008) montrent que, pendant une lecture en chinois, la distance de saccade (ou « saccade amplitude », la distance entre deux points de fixation du regard consécutifs) d'un Chinois natif est presque 2,5 fois plus importante que celle d'un apprenant du chinois de niveau avancé¹¹⁰, parce qu'à chaque fixation du regard, un Chinois natif parvient à traiter plus d'informations lexicales des caractères se situant à droite du point de fixation du regard, par rapport à un apprenant du chinois. Cela permet à un lecteur natif de ne pas lire chaque caractère présent dans le texte. En dehors du facteur en termes de la performance linguistique, nos participants ont lu les textes entièrement affichés sur l'écran, avec les caractères horizontalement et verticalement écartés (cf. la Section 4.3.2.3). Cela aurait pu inciter les participants à lire plus attentivement les caractères.

¹¹⁰ HSK 7~8 correspondant au niveau C1+ selon CECR (Bellassen, 2011)

6.2.3 AOI observées

Parmi nos participants, P23 et P24 n'ont pas suivi les mêmes cours de chinois ni utilisé les mêmes manuels pédagogiques que les autres participants ayant suivi le parcours de Licence à l'UT2J. Il est possible que ces deux participants n'aient pas le même niveau de connaissance de tous les mots présents dans les deux textes. Afin d'homogénéiser les données oculométriques à analyser, nous avons exclu celles de P23 et de P24 des analyses. Nous rappelons que les données des participants P13, P27, P29 et P33 sont déjà exclues à cause de la mauvaise qualité de suivi. Les analyses que nous présentons dans les sections 6.2.4 et 6.2.5 sont donc basées sur les données oculométriques de 27 participants.

Pour rappel, nous avons défini au total 639 AOI pour les deux textes. Parmi ces AOI, 42 AOI concernent les caractères constituant les mots cibles (2 AOI x 17 mots + 4 AOI x 1 mot + 3 AOI x 1 mot + 1 AOI x 1 mot)¹¹¹ et 597 autres AOI (639 AOI – 42 AOI) concernent les mots non cibles dans les deux textes.

Tous les mots non cibles présents dans les deux textes avaient été enseignés et/ou pratiqués en classe au moment de l'expérience, selon le programme pédagogique à l'UT2J, à l'exception du mot « 日子 (la vie (quotidienne)) » dans le texte A. Nous avons donc également exclu les données de ce mot et avons finalement observé pour chaque participant 595 AOI concernant les mots non cibles.

Nous avons ensuite exclu 10 AOI concernant les mots cibles déjà connus par certains participants (4 AOI x 1 participant + 2 AOI x 3 participants)¹¹².

Ainsi, nous avons observé au total 17189 AOI (cf. le Tableau 6.7), parmi lesquelles 16065 AOI concernent les mots non cibles (« AOI non MC », 595 AOI x 27 participants) et 1124 AOI pour les mots cibles (« AOI MC », 42 AOI x 27 participants – 10 AOI). Parmi les 1124 AOI pour les mots cibles, 557 AOI concernent les mots cibles lexicalisés (« AOI MC_L ») et 567 autres AOI concernant les mots cibles non lexicalisés (« AOI MC_NL »). Les résultats et les analyses que nous

¹¹¹ MC12 est constitué de 4 caractères, MC17 d'un seul caractère et MC20 de 3 caractères. Les autres mots cibles contiennent chacun 2 caractères.

¹¹² Nous rappelons que le mot MC12 (constitué de 4 caractères, donc 4 AOI) est connu par le participant P15 et le mot MC14 (constitué de 2 caractères, donc 2 AOI) est connu par les participants P22, P31 et P32.

présenterons dans les sections 6.2.4 et 6.2.5 sont basées sur les données de ces 17189 AOI (cf. la Partie 2 du Volume II de cette thèse pour les données oculométriques en détail).

	AOI MC_L	557
	AOI MC_NL	567
AOI MC		1124
AOI non MC		16065
Total		17189

Tableau 6.7 – Nombre des AOI observées

6.2.4 Comparaisons entre les AOI concernant les mots cibles et non cibles

Selon nos résultats, en moyenne, les participants ont fixé leur regard 1097 millisecondes (désormais ms) sur les AOI contenant les mots cibles et 762 ms sur les AOI contenant les mots non cibles, soit 44% de plus (Figure 6.5).



Figure 6.5 – Temps total (ms) en moyenne pour les AOI contenant les mots cibles et non cibles

Les yeux des participants ont visité en moyenne 2,6 fois les AOI contenant les mots cibles, par rapport à 1,6 fois pour les mots non cibles, soit 25% de plus (Figure 6.6).



Figure 6.6 – Nombre de visite du regard en moyenne pour les AOI contenant les mots cibles et non cibles

Quant à la première fixation, la durée en moyenne pour les AOI des mots cibles est de 453 ms et celle pour les mots non cibles est de 366 ms, soit 24% de plus (Figure 6.7).



Figure 6.7 – Première fixation (ms) en moyenne pour les AOI contenant les mots cibles et non cibles

Le Tableau 6.8 ci-dessous présente, pour les 27 participants observés, le temps total, le nombre de visite et la première fixation en moyenne pour les AOI concernant les mots cibles (« AOI MC ») et non cibles (« AOI non MC »). Selon le Tableau 6.8, les participants n’ont pas porté leur attention de façon équivalente d’une part aux AOI contenant les mots cibles et d’autre part à celles contenant les mots non cibles. Pour la majorité des participants, les résultats des trois mesures pour les mots cibles sont supérieurs à ceux pour les mots non cibles. Mais pour les participants P21, P22 et P26, il y a peu de différence entre les mots cibles et non cibles. Pour les participants P09 et P20, le temps total pour les mots cibles est inférieur à celui pour les mots non cibles.

Participant	Temps total *		Nombre de visite		Première fixation *	
	AOI MC	AOI non MC	AOI MC	AOI non MC	AOI MC	AOI non MC
P01	652	422	1,6	1,3	326	256
P02	614	410	0,9	0,8	479	302
P03	986	715	2,1	1,8	486	308
P04	1066	909	1,0	1,2	694	554
P05	823	686	1,7	1,4	379	355
P06	911	686	1,9	1,5	399	334
P07	1997	1082	2,7	1,7	741	521
P08	2363	962	3,7	2,2	733	390
P09	585	673	1,1	1,1	339	411
P10	1435	748	1,9	1,2	832	525
P11	866	519	1,7	1,2	400	308

P12	1233	1070	2,9	2,4	386	414
P14	858	689	1,1	1,0	572	471
P15	486	317	0,9	0,7	346	261
P16	752	582	1,6	1,4	473	316
P17	2173	2062	3,9	4,3	508	402
P18	635	549	1,0	1,1	439	341
P19	2148	800	3,9	2,1	581	318
P20	992	1112	2,2	2,4	356	409
P21	929	865	2,0	1,9	326	395
P22	492	495	1,4	1,3	291	259
P25	834	607	1,6	1,3	410	319
P26	599	522	1,4	1,4	345	272
P28	1600	1028	3,3	2,5	345	381
P30	1243	780	2,4	2,0	345	362
P31	867	615	1,5	1,2	345	359
P32	1473	661	2,6	1,7	345	349

* En milliseconde

Tableau 6.8 – Temps total, nombre de visite et première fixation en moyenne pour les AOI contenant les mots cibles (AOI MC) et non cibles (AOI non MC) par participants

Les analyses statistiques par Test T pour échantillons appariés montrent qu'entre les AOI concernant les mots cibles et non cibles, la différence est significative au sujet du temps total ($p = 0,00$) et au sujet du nombre de visite ($p = 0,00$) ; la différence sur la première fixation est également significative entre les deux groupes d'AOI ($p = 0,00$).

	Moyenne	Ecart-type	Erreur standard moyenne	Intervalle de confiance 95% de la différence		t	ddl	Sig. (bilatérale)
				Inférieure	Supérieure			
Temps total MC – Non MC	334,96370	388,24239	74,71728	181,38013	488,54728	4,483	26	,000
Visite MC – Non MC	,36333	,50622	,09742	,16308	,56359	3,730	26	,001
Première fixation MC – Non MC	106,13963	106,45193	20,48668	64,02865	148,25061	5,181	26	,000

6.2.5 Résultats pour les mots cibles

Nous rappelons que les mots cibles ne sont pas constitués du même nombre de caractères. Par exemple, MC12 contient quatre caractères, MC17 en contient un seul et MC20 contient trois caractères. Pour cette raison, le temps total sur un mot cible est la moyenne des durées pour les caractères constitutifs de ce mot. Par exemple, dans le cas où un mot cible M est constitué de deux caractères, si le temps total sur son premier caractère est de 100 ms et celui sur le deuxième caractère est de 200 ms, le temps total en moyenne passé sur le mot M est alors de 150 ms.

Les résultats en moyenne pour les mots cibles en termes du temps total (Figure 6.8), du nombre de visite (Figure 6.9) et de la première fixation (Figure 6.10) sont montrés ci-dessous.

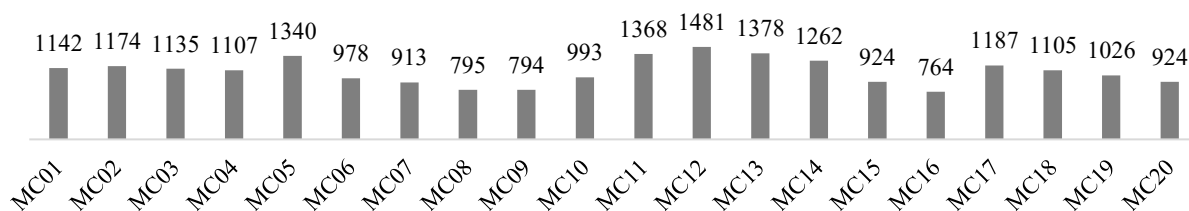


Figure 6.8 – Temps total (ms) en moyenne par mot

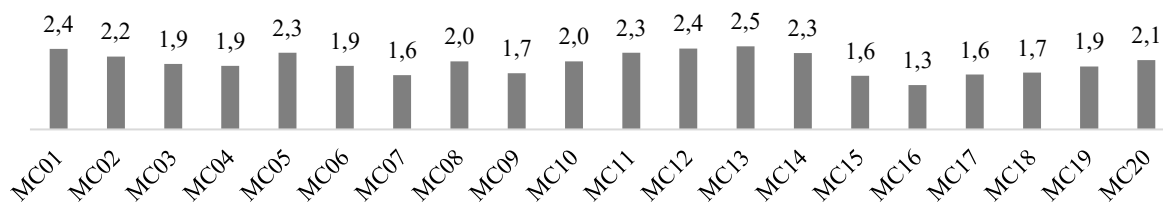


Figure 6.9 – Nombre de visite du regard en moyenne par mot

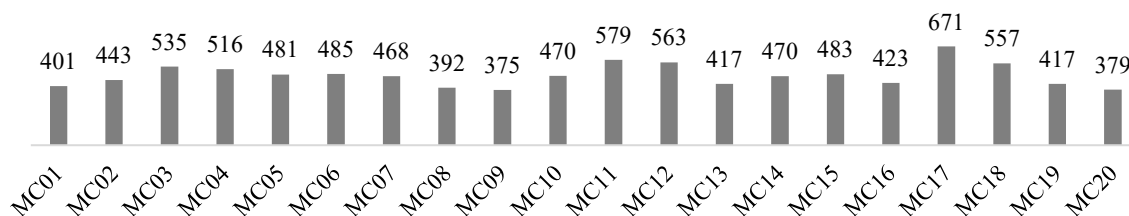


Figure 6.10 – Première fixation (ms) en moyenne par mot

MC01	节俭	(<i>jiéjiǎn</i>) : économe, frugal
MC02	窗花	(<i>chuānghuā</i>) : papier découpé pour la décoration d'une fenêtre
MC03	送灶	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune

MC04	腊月	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC05	汤圆	(<i>tāngyuán</i>) : boulette de riz glutineux farcie de viande (fourrée à la confiture)
MC06	雀跃	(<i>quèyuè</i>) : sauter de joie
MC07	还礼	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC08	在乎	(<i>zàihu</i>) : prendre qch. à cœur ; faire attention à
MC09	乏味	(<i>fáwèi</i>) : fade, insipide
MC10	盼望	(<i>pànwàng</i>) : espérer, désirer
MC11	安适	(<i>ānshì</i>) : tranquille et confortable
MC12	五颜六色	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC13	明丽	(<i>mínglì</i>) : clair et beau
MC14	彩虹	(<i>cǎihóng</i>) : arc-en-ciel
MC15	踏青	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps
MC16	风筝	(<i>fēngzhēng</i>) : cerf-volant
MC17	蝉	(<i>chán</i>) : cigale
MC18	闲聊	(<i>xiánliáo</i>) : bavarder
MC19	飘零	(<i>piāolíng</i>) : tomber en voltigeant
MC20	白茫茫	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

Selon les Figures 6.8, 6.9 et 6.10 ci-dessus, nos participants n'ont pas lu les mots cibles de façon équivalente : parmi les 20 mots inconnus, en moyenne les participants ont passé le plus de temps pour lire le mot MC12 ; les mots cibles les plus visités sont MC01 et MC12 ; le mot cible dont la première fixation est la plus longue est MC17.

6.2.5.1 Comparaisons entre les AOI des mots lexicalisés et non lexicalisés

Selon nos résultats, le temps total en moyenne pour les AOI contenant les mots cibles lexicalisés est de 1073 ms et celui pour les mots cibles non lexicalisés est de 1119 ms. (Figure 6.11)



Figure 6.11 – Temps total (ms) en moyenne pour les AOI des mots cibles lexicalisés (L) et non lexicalisés (NL)

Au niveau du nombre de visite, les résultats pour les AOI contenant les mots cibles lexicalisés et non lexicalisés sont pareils, soit 2 visites (Figure 6.12).



Figure 6.12 – Nombre de visite en moyenne pour les AOI des mots cibles lexicalisés (L) et non lexicalisés (NL)

Quant à la première fixation, les résultats obtenus pour les deux groupes sont respectivement de 477 ms en moyenne pour les mots cibles lexicalisés et de 467 ms en moyenne pour les mots non lexicalisés (Figure 6.13).



Figure 6.13 – Première fixation (ms) en moyenne pour les AOI des mots cibles lexicalisés (L) et non lexicalisés (NL)

Le Tableau 6.9 ci-dessous présente, pour les 27 participants observés, le temps total, le nombre de visite et la première fixation en moyenne pour les AOI contenant les mots cibles lexicalisés (« AOI MC_L ») et non lexicalisés (« AOI MC_NL »).

Participant	Temps total *		Nb de visite		Première fixation *	
	AOI MC_L	AOI MC_NL	AOI MC_L	AOI MC_NL	AOI MC_L	AOI MC_NL
P01	651	654	1,5	1,8	355	298
P02	408	820	0,8	1,0	367	591
P03	1037	935	2,3	2,0	466	505
P04	1094	1039	0,9	1,1	772	616
P05	860	785	1,5	1,9	416	341
P06	1036	787	2,3	1,5	371	427
P07	2227	1766	3,1	2,3	725	758
P08	2703	2023	4,4	2,9	731	736
P09	590	579	0,9	1,3	355	324
P10	1374	1496	1,6	2,1	857	807

P11	698	1033	1,6	1,8	329	472
P12	1304	1163	3,1	2,6	451	321
P14	693	1023	1,0	1,2	516	629
P15	377	574	0,8	1,0	214	453
P16	718	785	1,5	1,7	507	439
P17	1745	2602	3,3	4,5	561	455
P18	585	686	0,8	1,3	463	415
P19	2058	2237	3,7	4,1	550	611
P20	1072	912	2,2	2,2	451	261
P21	772	1086	1,6	2,3	261	392
P22	466	516	1,4	1,3	309	274
P25	747	921	1,5	1,6	422	398
P26	584	613	1,5	1,2	303	387
P28	1608	1591	3,6	3,0	435	344
P30	1254	1232	2,4	2,3	610	421
P31	729	992	1,2	1,8	456	439
P32	1590	1367	2,7	2,6	626	507

* En milliseconde

Tableau 6.9 – Temps total, nombre de visite et première fixation en moyenne pour les AOI concernant les mots lexicalisés (L) et non lexicalisés (NL) par participants

Les analyses par Test T pour échantillons appariés montrent qu'entre les AOI contenant les mots cibles lexicalisés et non lexicalisés, la différence n'est pas significative au sujet du temps total ($p > 0,05$), ni au sujet du nombre de visite du regard ($p > 0,05$) ; la différence n'est pas non plus significative au niveau de la première fixation ($p > 0,05$).

	Moyenne	Ecart-type	Erreur standard moyenne	Intervalle de confiance 95% de la différence		t	ddl	Sig. (bilatérale)
				Inférieure	Supérieure			
Temps total MC L – MC NL	-45,73296	292,52027	56,29555	-161,45013	69,98420	-,812	26	,424
Visite MC L - MC_NL	-,04593	,54988	,10583	-,26345	,17160	-,434	26	,668
Première fixation MC L - MC_NL	9,54815	114,39651	22,01562	-35,70560	54,80190	,434	26	,668

6.2.5.2 Rapport entre l'attention portée aux mots et le degré de réussite d'inférence

Le Tableau 6.10 ci-dessous présente, pour les 27 participants observés, les résultats des trois mesures oculométriques et les scores d'inférence correspondants.

Participant	Temps total *	Nb de visite	Première fixation *	Score d'inférence**
P01	652	1,6	326	0,43
P02	614	0,9	479	0,28
P03	986	2,1	486	0,06
P04	1066	1,0	694	0,09
P05	823	1,7	379	0,03
P06	911	1,9	399	0,33
P07	1997	2,7	741	0,29
P08	2363	3,7	733	0,25
P09	585	1,1	339	0,04
P10	1435	1,9	832	0,40
P11	866	1,7	400	0,05
P12	1233	2,9	386	0,06
P14	858	1,1	572	0,05
P15	486	0,9	346	0,30
P16	752	1,6	473	0,20
P17	2173	3,9	508	0,01
P18	635	1,0	439	0,19
P19	2148	3,9	581	0,21
P20	992	2,2	356	0,06
P21	929	2,0	326	0,01
P22	492	1,4	291	0,26
P25	834	1,6	410	0,13
P26	599	1,4	345	0,34
P28	1600	3,3	389	0,33
P30	1243	2,4	515	0,44
P31	867	1,5	447	0,41
P32	1473	2,6	563	0,41

* En milliseconde

** Score maximal = 1

Tableau 6.10 – Temps total, nombre de visite, première fixation et score d'inférence en moyenne par participant

Selon les analyses par modèle linéaire général, le temps total n'a pas d'effet significatif sur le score d'inférence lexicale pour les mots cibles ($p > 0,05$) ; le nombre de visite du regard n'a pas non plus d'effet significatif sur le score d'inférence ($p > 0,05$), ni la première fixation ($p > 0,05$).

Variable dépendante : Inférence_Score

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	,001 ^a	1	,001	,042	,839
Ordonnée à l'origine	,200	1	,200	8,824	,006
TempsTotal	,001	1	,001	,042	,839
Erreur	,567	25	,023		
Total	1,754	27			
Total corrigé	,568	26			

a. R deux = ,002 (R deux ajusté = -,038)

Variable dépendante : Inférence_Score

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	4,053E-6 ^a	1	4,053E-6	,000	,989
Ordonnée à l'origine	,197	1	,197	8,661	,007
Visite	4,053E-6	1	4,053E-6	,000	,989
Erreur	,568	25	,023		
Total	1,754	27			
Total corrigé	,568	26			

a. R deux = ,000 (R deux ajusté = -,040)

Variable dépendante : Inférence_Score

Source	Somme des carrés de type III	ddl	Moyenne des carrés	D	Sig.
Modèle corrigé	,019 ^a	1	,019	,867	,361
Ordonnée à l'origine	,032	1	,032	1,438	,242
PremièreFixation	,019	1	,019	,867	,361
Erreur	,549	25	,022		
Total	1,754	27			
Total corrigé	,568	26			

a. R deux = ,034 (R deux ajusté = -,005)

6.2.6 Discussion

Afin d'observer l'attention que les participants ont portée aux mots cibles pendant la lecture, nous avons examiné les mouvements des yeux de nos participants par le biais de trois mesures oculométriques, le temps total, le nombre de visite et la première fixation. Rayner (1998) a remarqué que, par rapport aux mots fréquents, les

mots moins fréquents reçoivent des regards de plus longue durée. Au sujet de la visite du regard, Godfroid et al. (2013) indiquent que, dans l'inférence lexicale en L2, la revisite du regard pourrait être un indicateur qu'un apprenant cherche des informations contextuelles pour inférer le sens d'un mot nouveau. Liversedge et al. (1998) indiquent, pour leur part, que la première fixation est un bon indicateur de la difficulté du traitement d'informations immédiatement après avoir rencontré un mot.

6.2.6.1 Attention portée aux mots inconnus rencontrés en lecture

Selon nos résultats, pendant la lecture, les participants ont passé globalement plus de temps à déchiffrer les caractères constitutifs des mots inconnus. Cela est en accord avec le résultat de l'étude de Godfroid & Schmidtke (2013), dans laquelle les apprenants ayant un niveau avancé en anglais L2 ont passé plus de temps pour lire les mots inconnus présents dans les textes. Les regards de nos participants sont également retournés plus fréquemment sur les AOI contenant les mots cibles. Quant à la première fixation, la durée en moyenne est plus longue pour les AOI contenant les mots cibles que pour les mots non cibles. Cela révèle que globalement nos participants ont pris conscience de la présence des mots inconnus dans les textes. Les participants ont porté plus d'attention aux mots cibles qui leur étaient inconnus. Ces mots inconnus ont donc exigé d'eux des traitements cognitifs plus appuyés

Il est à noter qu'en chinois, un mot est souvent constitué de plusieurs caractères. Dans notre étude, la revisite du regard est mesurée par caractère. Cela signifie que nos participants ont probablement relu les caractères non seulement pour chercher des indices contextuels afin de déduire le sens d'un mot cible (Godfroid et al., 2013), mais également pour identifier les unités lexicales. Comme dans une phrase en chinois, les caractères sont en juxtaposition sans espaces supplémentaires qui balisent la frontière entre les mots, c'est au lecteur lui-même de segmenter les caractères afin d'identifier les unités lexicales en lisant la phrase (cf. la Section 3.2.4). Jusqu'à présent, il n'y a pas de conclusion sur l'élément critique pour un lecteur, natif ou non, dans l'identification d'un mot dans la compréhension écrite en chinois (Xiang et al., 2018).

En y regardant de plus près, nous avons constaté que les participants n'ont pas

porté leur attention de façon homogène aux mots cibles :

- La forme d'un mot inconnu semble avoir un effet sur l'attention de nos participants. En termes de temps total, le mot MC12 « 五颜六色 (*wǔyánliùsè*, de diverses couleurs, multicolore) » a reçu en moyenne 1507 ms de regard, presque deux fois plus que le temps total en moyenne pour les mots non cibles (763 ms). La raison plausible est que ce mot contient deux caractères de chiffre bien courants et connus par les participants, « 五 (*wǔ*, cinq) » et « 六 (*liù*, six) ». Les participants connaissent aussi le quatrième caractère, « 色 (*sè*, couleur) ». La majorité des mots en chinois contemporain sont constitués de deux caractères. Or, les deux chiffres pourraient facilement faire penser aux participants qu'il y a un lien particulier entre les quatre caractères, « 五*六色 (cinq * six couleurs) ». Il n'est donc pas étonnant de voir que, parmi tous les caractères constitutifs des mots cibles, les trois premiers caractères de MC12, « 五 », « 颜 » et « 六 », sont parmi les plus visités (soit 2,5 fois). D'autant plus, le caractère « 颜 », inconnu pour les participants, est celui ayant reçu le regard à la plus longue durée parmi tous les caractères constitutifs des 20 mots cibles, soit 2206 ms.
- L'emplacement d'un mot inconnu dans le texte influence plausiblement l'attention de nos participants. En termes de nombre de visite du regard, MC01 « 节俭 (*jiéjiǎn*, économe, frugal) » a reçu en moyenne 2,5 visites, bien supérieur à la moyenne pour les mots non cibles (1,6 visites). Le mot MC01 est le premier mot inconnu que les participants ont rencontré dans le texte. Cela pourrait être la raison pour laquelle les participants ont prêté particulièrement attention à ce mot pour le déchiffrer : le deuxième caractère de MC01, « 俭 » qui est inconnu pour les participants, a reçu au total 1488 ms de regard, presque deux fois plus que la moyenne pour les caractères des mots non MC. Ce caractère est aussi celui le plus visité (soit 2,8 visites) parmi tous les caractères constitutifs des 20 mots cibles. Pareillement, MC11 « 安逸 (*ānshì*, tranquille et confortable) » est le premier mot cible que les participants ont rencontré dans le texte B. Le temps total et le nombre de visite pour MC11

sont aussi relativement importants (1402 ms et 2,4 visites). Le deuxième caractère de MC11, « 适 » qui est inconnu pour les participants, a reçu en moyenne 1641 ms de regard et 2,5 visites du regard. Nous en déduisons donc que l'emplacement d'un mot inconnu a une influence sur l'attention portée sur lui.

Parmi les 42 caractères constitutifs des mots cibles, la première fixation pour les caractères inconnus (par exemple « 颜 » de MC12, « 跃 » de MC06, « 适 » de MC11, « 闲 » de MC18 et « 蝉 » de MC17) est plus longue que celle pour les caractères connus par les participants (par exemple « 五 » et « 六 » de MC12 « 五颜六色 »). Cela signifie que les participants ont repéré les caractères inconnus dans les mots cibles et ont alors passé plus de temps pour déchiffrer ces caractères.

Par ailleurs, le temps total en moyenne respectivement pour les mots MC08 « 在乎 (*zàihu*, prendre qch. à cœur ; faire attention à) » (778 ms), MC09 « 乏味 (*fáwèi*, fade, insipide) » (773 ms) et MC16 « 风筝 (*fēngzhēng*, cerf-volant) » (782 ms) est peu différent de la moyenne pour les autres mots non cibles (763 ms). En termes de la visite du regard, MC16 « 风筝 (*fēngzhēng*, cerf-volant) » (1,3 visite) est inférieur à la moyenne des mots non cibles (1,6 visites).

Nos résultats sur l'attention hétérogène portée aux mots cibles par nos participants fournissent une preuve pour les résultats des autres études (ex. Dong, 2007; Fraser, 1999; Paribakht & Wesche, 1999; X. J. Qian, 2005) qui montrent communément qu'un apprenant peut ignorer ou négliger des mots inconnus présents dans sa lecture.

6.2.6.2 Comparaison entre les mots inconnus lexicalisés et non lexicalisés

Au niveau du temps total et du nombre de visite, les moyennes pour les mots non lexicalisés sont supérieures à celles pour les mots lexicalisés. Au niveau de la première fixation, la moyenne pour les mots lexicalisés est supérieure à celle des mots non lexicalisés. Toutefois, ces différences ne sont pas significatives.

Nous rappelons que, parmi les trois mesures, la première fixation du regard est

un indicateur spécifiquement lié à la difficulté rencontrée immédiatement après avoir lu un mot dans le texte. Pour un apprenant du chinois L2, cette difficulté est généralement liée au déchiffrement des caractères. Par rapport aux mots non lexicalisés, les mots cibles lexicalisés sélectionnés dans notre étude contiennent plus de caractères inconnus par nos participants. Cela est une raison plausible pour laquelle les participants ont eu une première fixation du regard plus longue pour les mots lexicalisés.

Nous pensons que, pendant la lecture, malgré la prise de conscience de la présence des mots inconnus dans les textes, nos participants ne faisaient pas de distinction entre les mots inconnus en fonction de leur nature de lexicalisation. En d'autres termes, la lexicalisation d'un mot inconnu n'est pas un élément influençant l'attention de l'apprenant portée à ce mot lors d'une lecture.

6.2.6.3 Rapport entre l'attention portée aux mots cibles et le degré de réussite d'inférence

Nous n'avons pas trouvé d'effet significatif des trois mesures, le temps total, le nombre de visites et la première fixation, sur le score d'inférence lexicale. Autrement dit, la réussite de l'inférence du sens d'un mot inconnu ne dépend pas de l'attention de l'apprenant portée sur ce mot lors de la lecture, possiblement parce que, selon nos résultats, l'attention qu'un apprenant porte à un mot chinois inconnu peut se diversifier pour les causes telles que la forme du mot (plus précisément, le nombre des caractères constitutifs), la connaissance des caractères constitutifs et l'emplacement du mot dans le texte. Un mot inconnu peut être ignoré ou négligé par l'apprenant pendant la lecture (Dong, 2007; Fraser, 1999; Paribakht & Wesche, 1999; X. J. Qian, 2005). En revanche, quand l'apprenant entreprend une inférence du sens pour un mot inconnu, les facteurs qui peuvent intervenir dans le résultat sont la richesse d'information dans le contexte, les connaissances linguistiques et non linguistiques préalables, etc. (cf. la Section 1.3).

6.3 Conclusions

Dans la première partie du chapitre, nous avons présenté et analysé les scores d'inférence des mots cibles et les éléments qui peuvent intervenir dans le résultat de l'inférence lexicale.

Selon les scores d'inférence obtenus dans notre étude, un apprenant francophone réussit mieux à comprendre le sens d'un mot chinois non lexicalisé en français. Ce résultat est différent des résultats obtenus par les études à ce sujet réalisées en anglais L2. Il est aussi opposé à notre hypothèse.

Nous avons trouvé une corrélation positive entre l'utilisation des sens des caractères constitutifs d'un mot et le degré de réussite de l'inférence du mot. En d'autres termes, si un mot inconnu est sémantiquement transparent et si l'apprenant connaît les sens de ses caractères constitutifs, l'apprenant parvient à comprendre, ne serait-ce que partiellement, le sens du mot, qu'il soit lexicalisé ou non dans la L1 de l'apprenant. En revanche, lors de l'inférence d'un mot chinois lexicalisé, si les informations repérées par l'apprenant ne suffisent pas à activer l'accès au concept disponible dans sa L1, la tentative de la compréhension du mot inconnu n'aboutira pas, même si le mot est lexicalisé. Nous concluons alors que **la lexicalisation d'un mot chinois inconnu n'est pas l'élément concluant dans la compréhension du sens du mot.**

Néanmoins, nos résultats montrent qu'**un concept équivalent disponible dans la L1 de l'apprenant donne plus de possibilités à l'apprenant de comprendre entièrement un mot L2 inconnu.** Des concepts connexes dans la L1 peuvent aider l'apprenant à comprendre un mot L2 non lexicalisé, mais le plus souvent partiellement.

En dehors de cela, nous avons constaté que **l'inférence des mots chinois dont la compréhension exige des connaissances socio-culturelles est la moins réussie.** Les mots L2 de ce type sont généralement non lexicalisés dans la L1 de l'apprenant. Dans le cadre de l'apprentissage hétéroglotte, les apprenants ont largement moins d'occasions d'obtenir des connaissances sur cet aspect.

Il est à noter que nous avons également trouvé une corrélation positive entre la diversité des sources de connaissance utilisées dans l'inférence d'un mot et le degré de réussite de l'inférence du mot. La connaissance d'un mot est multi-dimensionnelle. Un ensemble d'indices de différentes sources fournit à l'apprenant des informations lexicales de divers aspects sur le mot. Cela donne à l'apprenant plus de possibilités de vérifier toutes les informations repérées afin d'avoir une meilleure analyse sur le sens du mot concerné. Notamment dans le cas d'un caractère polysémique, l'identification du sens approprié en fonction du contexte a une grande importance.

Dans la deuxième partie du chapitre, nous avons ensuite montré les résultats oculométriques qui peuvent refléter l'attention de nos participants portée aux mots cibles pendant leur lecture.

Selon nos résultats, **un apprenant francophone porte globalement plus d'attention aux mots chinois inconnus présents dans un texte qu'aux mots connus**. En d'autres termes, il prend conscience de la présence des mots inconnus présents dans sa lecture. Cependant, l'apprenant ne prête pas la même attention à tous les mots inconnus. Les éléments tels que la forme du mot, l'emplacement du mot et la connaissance des caractères constitutifs peuvent influencer l'attention de l'apprenant. **La nature de lexicalisation du mot n'a pas d'influence significative sur l'attention de l'apprenant lors de la lecture**. Autrement dit, au moment de la lecture, l'apprenant ne traite pas les mots inconnus de façon distincte en fonction de la lexicalisation ou non du mot.

L'attention qu'un apprenant francophone porte à un mot chinois inconnu lors d'une lecture n'a pas d'effet sur le degré de la réussite de l'inférence du mot. L'attention de l'apprenant portée à un mot inconnu peut varier en fonction des éléments tels que la forme de constitution du mot et l'emplacement du mot. L'apprenant peut même ignorer un mot inconnu présent dans un texte. La réussite de l'inférence d'un mot chinois inconnu dépend d'éléments tels que la connaissance des caractères constitutifs, la richesse informative du contexte, la familiarité socio-culturelle, etc.

CHAPITRE VII – EVOLUTION DES CONNAISSANCES LEXICALES

Introduction

Dans ce chapitre, nous présenterons les résultats obtenus à quatre phases d'apprentissage, c'est-à-dire le VKS1, l'inférence lexicale, le VKS2 et le VKS3. A partir de ces résultats, nous analyserons l'évolution de la connaissance par rapport aux mots cibles chez les participants sur une période de cinq semaines, afin de savoir si 1) à l'issue de l'inférence du sens d'un mot chinois inconnu présent dans une lecture, une combinaison de deux tâches concernant le mot, par rapport à une seule tâche, aboutit à de meilleurs gains lexicaux à court et à long terme et si 2) la nature de lexicalisation du mot a un effet sur les gains lexicaux à court et à long terme.

Nous examinerons d'abord l'évolution du niveau de connaissance des mots cibles par le biais de l'échelle de connaissance à six niveaux du test VKS. Par la suite, nous nous focaliserons sur l'évolution de la connaissance sémantique pour les mots cibles chez nos participants. Nous ferons des analyses statistiques, à partir desquelles nous observerons, pendant les évolutions, les différences entre les deux groupes de participants ayant accompli différentes tâches après la tâche d'inférence lexicale et celles entre les mots cibles lexicalisés et non lexicalisés.

Nous concluons ce chapitre avec nos réponses aux questions énoncées plus haut.

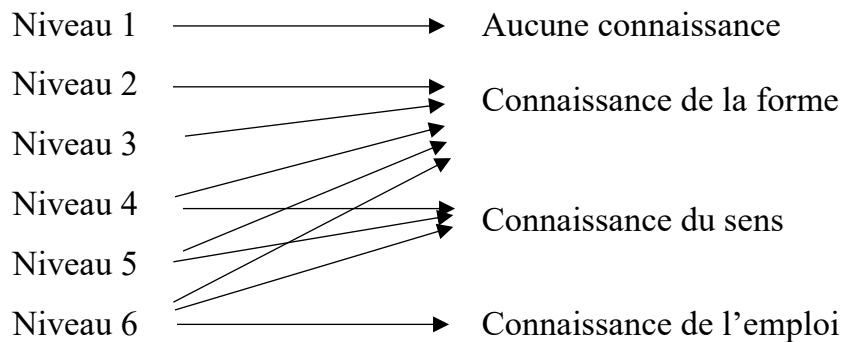
7.1 Evolution des connaissances lexicales multidimensionnelles pour les mots cibles

Nous voulions observer l'évolution de la connaissance lexicale sous divers aspects chez nos participants pour les mots cibles. Nous avons ainsi analysé les réponses des participants aux trois tests VKS sur leur niveau de connaissance pour les mots cibles.

Pour rappel, nous avons utilisé dans notre expérience une échelle de connaissances à six niveaux pour le test VKS, montrées ci-dessous :

1. Je n'ai jamais vu ce mot.
2. Je pense avoir déjà vu ce mot, mais je ne sais pas ce qu'il veut dire.
3. J'ai déjà vu ce mot, mais je ne sais pas ce qu'il veut dire.
4. J'ai déjà vu ce mot. Il veut peut-être dire ..., mais je n'en suis pas sûr(e).
5. Je connais ce mot et je sais qu'il veut dire...
6. Je connais ce mot et je sais qu'il veut dire... Je sais écrire une phrase en chinois avec ce mot.

Les connaissances lexicales correspondantes sont les suivantes :



Certaines réponses données par nos participants ne correspondent pas aux niveaux indiqués. Nous avons alors rectifié les réponses en respectant les règles suivantes :

- 1) Dans le cas où un participant a noté niveau 6 pour un mot en donnant le bon sens du mot et produisant une phrase avec le mot, si le mot est improprement utilisé grammaticalement dans la phrase, nous rectifions la réponse au niveau 5. Les erreurs d'écriture ou de grammaire dans les autres composants de la phrase ne sont pas comptées. Nous gardons, dans ce cas-là, le niveau 6 pour la connaissance de ce participant pour le mot. Une phrase produite par un participant est considérée correcte si l'usage sémantico-grammatical du mot cible concerné est correct. Les erreurs orthographiques ou syntaxiques n'ont pas été prises en compte. Par exemple, un participant a produit la phrase « 这个孩子求他的风筝。 *Zhè gè háizǐ qiú tā de fēngzhēng。* » pour justifier qu'il a la connaissance du mot cible « 风筝 (*fēngzhēng*, cerf-volant) » au niveau 6. Il a marqué à côté le sens de la phrase en français « Cet enfant a perdu son cerf-volant. » pour laquelle la phrase correcte en chinois est « 这个孩子弄丢了 他的风筝。 *Zhè gè háizi nòngdiū le tā de fēngzhēng。* » Ce participant n'a

pas utilisé le bon verbe et s'est trompé d'aspect du verbe. Cette phrase a été tout de même validée.

- 2) Si un participant a noté le niveau 5 ou 6 pour un mot en fournissant un sens partiellement correct, nous rectifions la réponse au niveau 4 ;
- 3) Si un participant a noté le niveau 4, 5 ou 6 pour un mot en fournissant un sens erroné, nous rectifions la réponse au niveau 3.

L'échelle de différents niveaux de connaissance est une mesure ordinale. Nous avons donc utilisé le modèle mixte à lien cumulatif (CLMM, Cumulative Link Mixed Model) du logiciel R¹¹³ afin d'effectuer des analyses statistiques que nous présentons dans la section 7.1.¹¹⁴

7.1.1 Evolution globale du niveau de connaissance pour les mots cibles

Nous rappelons que six cas ont été exclus des observations parce que quelques participants connaissaient certains mots cibles. En outre, au VKS1, des participants ont donné des réponses correctes ou partiellement correctes sur le sens de certains mots cibles. Toutefois, lors de l'entretien d'inférence lexicale, les rapports verbaux de ces participants ne démontraient pas une connaissance entière de ces derniers pour les mots concernés. Par exemple, une participante a donné une réponse correcte pour MC18 « 闲聊 (*xiánliáo*, bavarder) » au VKS1. Lors de l'inférence lexicale, nous avons découvert qu'elle connaissait un autre mot « 聊天 (discuter) », un synonyme de MC18, les deux mots comportant le caractère « 聊 ». Toutefois, elle ne connaissait pas ou ne se rappelait pas le premier caractère « 闲 » de MC18. Il est possible que ces participants aient rencontré ces mots cibles auparavant sans bien les retenir, ou bien qu'ils aient simplement confondu le mot cible avec un autre mot similaire à la forme qui, par exemple, partage un même caractère constitutif. Dans ce cas-là, nous avons gardé ces mots dans nos observations en rectifiant le niveau de connaissance à

¹¹³ Le modèle mixte à lien cumulatif (Cumulative Link Mixed Model) analyse par régression les effets des facteurs en mesure répétée (dans notre cas, les facteurs « Phase », « Groupe » et « Lexicalisation ») sur une variable ordinale (dans notre cas, les réponses de différents niveaux de connaissance) (cf. <https://rdrr.io/cran/ordinal/>, consulté le 21 avril 2021).

¹¹⁴ Nous remercions sincèrement M. Saïd JMEL et Mme Lyanne AHUMADA-EBRATT pour leur aide sur les analyses faites avec le logiciel R.

niveau 4 (une connaissance incertaine sur le sens du mot) pour les mots.

L'évolution globale des connaissances multidimensionnelles pour les mots cibles est présentée dans le Tableau 7.1 (cf. les Annexe 4, Annexe 5 et Annexe 6 pour les réponses de divers niveaux de connaissance obtenues lors des trois tests VKS).

	Niveau 1		Niveau 2		Niveau 3		Niveau 4		Niveau 5		Niveau 6		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	N	%	Nb	%	Nb	%
VKS1	410	63%	113	17%	121	18%	10	2%	0	0%	0	0%	654	100%
VKS2	110	17%	104	16%	272	41%	65	10%	71	11%	32	5%	654	100%
VKS3	78	12%	93	14%	295	45%	71	11%	72	11%	45	7%	654	100%

Tableau 7.1 – Nombre et répartition des réponses de différents niveaux aux trois tests VKS

Nous constatons qu'au VKS1, les participants jugent qu'ils n'ont jamais vu la majorité des mots cibles (c'est-à-dire les réponses du niveau 1 qui représentent 63% de la totalité des réponses). La proportion des réponses des niveaux 2 et 3, correspondant à une connaissance plausible ou certaine de la forme du mot, représentent 35% de la totalité. Au total, 98% des mots observés, qui correspondent aux réponses des niveaux 1, 2 et 3, sont jugés sémantiquement inconnus par les participants. Aucun mot cible n'a été jugé niveau 6 (c'est-à-dire une connaissance de l'emploi du mot) dans une phrase lors du VKS1.

Par rapport au VKS1, la proportion des réponses du niveau 1, donc niveau « jamais-vu », a diminué à 17% au VKS2 et à 12% au VKS3 (vs 63% au VKS1); l'ensemble des réponses des niveaux 2 et 3, correspondant à une connaissance de la forme, montre une augmentation (35% au VKS1 vs 58% au VKS2 et 59% au VKS3) ; le nombre des réponses du niveau 4 et au-delà, correspondant à une connaissance du sens, a également connu une augmentation (2% au VKS1 vs 25% au VKS2 et 29% au VKS3). Au VKS1, aucun participant n'est parvenu à employer correctement un mot cible pour produire une phrase, correspondant au niveau 6. Les proportions des réponses du niveau 6 sont respectivement de 5% au VKS2 et de 7% au VKS3.

Les courbes d'évolution des réponses de divers niveaux de connaissance sont présentées dans la Figure 7.1.

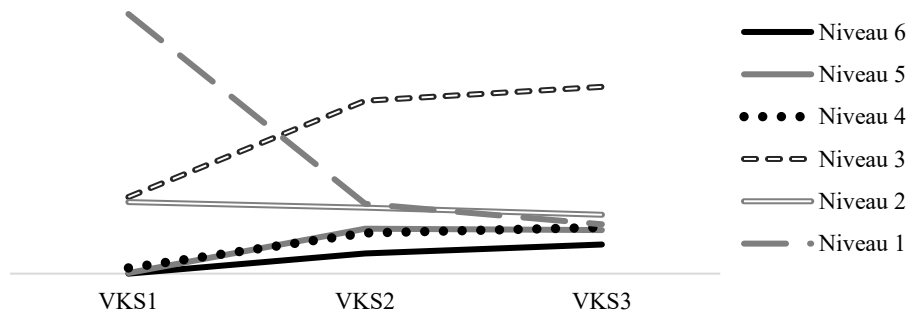


Figure 7.1 – Evolution de la répartition des réponses de différents niveaux de connaissance

Voici les résultats d’analyse statistique concernant l’effet du facteur « Phase » et l’effet de deux facteurs, « Mot » et « Phase », sur les réponses de différents niveaux de connaissance.

	LR Chisq	Df	Pr(>Chisq)
Phase	839,16	2	< 2,2e-16
Mot:Phase	136,84	38	4,394e-13

Selon les résultats, l’effet du facteur « Phase », qui renvoie aux trois tests VKS, est significatif ($p < 0,000$). Cela signifie qu’en général, les niveaux de connaissance sur les mots cibles chez nos participants aux trois tests VKS sont statistiquement différents. Les résultats post-hoc montrent que le niveau de connaissance au VKS1 est significativement inférieur à celui au VKS2, également à celui au VKS3. Cela signifie qu’aux VKS2 et VKS3, nos participants ont eu des gains lexicaux pour les mots cibles. La différence entre le VKS2 et le VKS3 est aussi significative, mais l’écart est moins important.

contrast	estimate	SE	df	z.ratio	p.value
VKS1 - VKS2	-3,078	0,136	Inf	-22,552	< ,0001
VKS1 - VKS3	-3,436	0,140	Inf	-24,551	< ,0001
VKS2 - VKS3	-0,358	0,104	Inf	-3,425	0,0018

L’effet des deux facteurs, « Mot » et « Phase », est également significatif ($p < 0,000$). Les résultats post-hoc montrent que, du VKS1 au VKS3, à l’exception de MC01 ($p > 0,1$), le niveau de connaissance de tous les mots cibles a eu une évolution significative positive ($p < 0,05$). Cela signifie qu’au VKS3, nos participants ont eu des gains lexicaux pour 19 mots cibles, non pour MC01.

	contrast	estimate	SE	df	z.ratio	p.value
MC01	VKS1 – VKS2	-1,68	0,47	Inf	-3,60	0,23
	VKS1 – VKS3	-1,60	0,48	Inf	-3,37	0,39
	VKS2 – VKS3	0,08	0,44	Inf	0,18	1,00
MC02	VKS1 – VKS2	-2,88	0,51	Inf	-5,62	< 0,0001
	VKS1 – VKS3	-3,19	0,51	Inf	-6,26	< 0,0001
	VKS2 – VKS3	-0,31	0,46	Inf	-0,67	1,00
MC03	VKS1 – VKS2	-2,32	0,48	Inf	-4,80	0,00
	VKS1 – VKS3	-2,76	0,49	Inf	-5,63	< 0,0001
	VKS2 – VKS3	-0,44	0,43	Inf	-1,03	1,00
MC04	VKS1 – VKS2	-3,08	0,52	Inf	-5,92	< 0,0001
	VKS1 – VKS3	-3,42	0,52	Inf	-6,63	< 0,0001
	VKS2 – VKS3	-0,34	0,46	Inf	-0,74	1,00
MC05	VKS1 – VKS2	-2,14	0,46	Inf	-4,64	0,01
	VKS1 – VKS3	-2,36	0,46	Inf	-5,11	0,00
	VKS2 – VKS3	-0,22	0,44	Inf	-0,50	1,00
MC06	VKS1 – VKS2	-2,40	0,68	Inf	-3,55	0,26
	VKS1 – VKS3	-3,03	0,67	Inf	-4,56	0,01
	VKS2 – VKS3	-0,63	0,48	Inf	-1,31	1,00
MC07	VKS1 – VKS2	-2,53	0,46	Inf	-5,50	0,00
	VKS1 – VKS3	-2,71	0,47	Inf	-5,83	< 0,0001
	VKS2 – VKS3	-0,19	0,46	Inf	-0,41	1,00
MC08	VKS1 – VKS2	-1,80	0,48	Inf	-3,78	0,14
	VKS1 – VKS3	-2,16	0,48	Inf	-4,50	0,01
	VKS2 – VKS3	-0,37	0,47	Inf	-0,78	1,00
MC09	VKS1 – VKS2	-2,97	0,52	Inf	-5,70	< 0,0001
	VKS1 – VKS3	-3,66	0,53	Inf	-6,89	< 0,0001
	VKS2 – VKS3	-0,68	0,46	Inf	-1,49	1,00
MC10	VKS1 – VKS2	-2,30	0,51	Inf	-4,54	0,01
	VKS1 – VKS3	-2,98	0,52	Inf	-5,78	< 0,0001
	VKS2 – VKS3	-0,69	0,50	Inf	-1,37	1,00
MC11	VKS1 – VKS2	-2,17	0,47	Inf	-4,63	0,01
	VKS1 – VKS3	-2,13	0,48	Inf	-4,45	0,01
	VKS2 – VKS3	0,04	0,44	Inf	0,08	1,00
MC12	VKS1 – VKS2	-5,47	0,53	Inf	-10,40	< 0,0001
	VKS1 – VKS3	-6,42	0,54	Inf	-11,99	< 0,0001
	VKS2 – VKS3	-0,95	0,46	Inf	-2,06	1,00
MC13	VKS1 – VKS2	-2,18	0,47	Inf	-4,67	0,00
	VKS1 – VKS3	-2,28	0,46	Inf	-4,93	0,00
	VKS2 – VKS3	-0,09	0,42	Inf	-0,22	1,00
MC14	VKS1 – VKS2	-3,56	0,74	Inf	-4,84	0,00
	VKS1 – VKS3	-4,32	0,74	Inf	-5,86	< 0,0001
	VKS2 – VKS3	-0,76	0,51	Inf	-1,48	1,00
MC15	VKS1 – VKS2	-3,30	0,74	Inf	-4,47	0,01
	VKS1 – VKS3	-3,64	0,74	Inf	-4,95	0,00
	VKS2 – VKS3	-0,34	0,48	Inf	-0,70	1,00
MC16	VKS1 – VKS2	-3,92	0,54	Inf	-7,21	< 0,0001
	VKS1 – VKS3	-4,60	0,55	Inf	-8,33	< 0,0001
	VKS2 – VKS3	-0,68	0,49	Inf	-1,40	1,00
MC17	VKS1 – VKS2	-5,14	0,60	Inf	-8,50	< 0,0001
	VKS1 – VKS3	-5,06	0,60	Inf	-8,38	< 0,0001
	VKS2 – VKS3	0,08	0,51	Inf	0,16	1,00
MC18	VKS1 – VKS2	-2,85	0,54	Inf	-5,29	0,00
	VKS1 – VKS3	-3,29	0,53	Inf	-6,18	< 0,0001

MC19	VKS2 – VKS3	-0,44	0,49	Inf	-0,90	1,00
	VKS1 – VKS2	-2,89	0,54	Inf	-5,37	0,00
	VKS1 – VKS3	-3,24	0,53	Inf	-6,09	< 0,0001
MC20	VKS2 – VKS3	-0,35	0,46	Inf	-0,76	1,00
	VKS1 – VKS2	-5,99	0,67	Inf	-8,92	< 0,0001
	VKS1 – VKS3	-5,86	0,67	Inf	-8,79	< 0,0001
	VKS2 – VKS3	0,13	0,44	Inf	0,29	1,00

7.1.2 Comparaison du niveau de connaissance entre le Groupe I et le Groupe II

Nous rappelons qu'à l'issue de l'inférence lexicale, les participants ont été désignés de façon aléatoire dans le Groupe I pour effectuer la tâche lacunaire avec une liste des mots cibles, ou bien dans le Groupe II pour retrouver d'abord le sens des mots cibles dans un dictionnaire bilingue chinois-français avant d'effectuer la même tâche lacunaire. Au total, le Groupe I comporte 17 participants et le Groupe II 16 participants. Nous avons eu 338 observations pour le Groupe I et 316 observations pour le Groupe II.

Les résultats du Groupe I et du Groupe II aux trois tests VKS, classés sur l'échelle à six niveaux de connaissance, sont donnés dans le Tableau 7.2.

Groupe I	Niveau 1		Niveau 2		Niveau 3		Niveau 4		Niveau 5		Niveau 6		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	207	61%	66	20%	62	18%	3	1%	0	0%	0	0%	338	100%
VKS2	69	20%	72	21%	128	38%	25	7%	32	10%	12	4%	338	100%
VKS3	43	13%	63	19%	153	45%	27	8%	31	9%	21	6%	338	100%

Groupe II	Niveau 1		Niveau 2		Niveau 3		Niveau 4		Niveau 5		Niveau 6		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	203	64%	47	15%	59	19%	7	2%	0	0%	0	0%	316	100%
VKS2	41	13%	32	10%	144	46%	40	13%	39	12%	20	6%	316	100%
VKS3	35	11%	30	9%	142	45%	44	14%	41	13%	24	8%	316	100%

Tableau 7.2 – Nombre et répartition des réponses de différents niveaux de connaissance aux trois tests VKS pour le Groupe I et le Groupe II

Selon les résultats d'analyse, l'effet du facteur « Groupe » n'est pas significatif ($p > 0,05$), mais l'effet des deux facteurs, « Groupe » et « Phase », est significative

($p < 0,00$). Cela signifie que les connaissances lexicales des mots cibles chez les participants des deux groupes n'ont pas évolué au même niveau.

	LR Chisq	Df	Pr(>Chisq)
Groupe	2,03	1	0,1547193
Groupe:Phase	15,94	2	0,0003455

Les résultats post-hoc montrent que, pour les deux groupes, le niveau de connaissances en général au VKS1 est significativement inférieur à ceux aux VKS2 et VKS3 ($p < 0,00$). Les participants des deux groupes ont donc eu des gains lexicaux pour les mots cibles à la phase post-inférence. L'écart entre les niveaux de connaissance à la phase post-inférence (donc VKS2 et VKS3) et celui à la phase pré-inférence (VKS1) est plus important pour le Groupe II (respectivement 3,371 et 3,586) que pour le Groupe I (respectivement 2,464 et 2,936).

	contrast	estimate	SE	df	z,ratio	p.value
Groupe I	VKS1 – VKS2	-2,464	0,166	Inf	-14,847	< 0,000
	VKS1 – VKS3	-2,936	0,169	Inf	-17,342	< 0,000
	VKS2 – VKS3	-0,472	0,143	Inf	-3,295	0,013
Groupe II	VKS1 – VKS2	-3,371	0,181	Inf	-18,643	< 0,000
	VKS1 – VKS3	-3,586	0,183	Inf	-19,581	< 0,000
	VKS2 – VKS3	-0,214	0,149	Inf	-1,438	0,704

En comparant les deux groupes aux trois tests VKS, nous pouvons voir qu'au VKS1, le niveau de connaissance du Groupe I est légèrement supérieur à celui du Groupe II (0,047), mais aux VKS2 et VKS3, le niveau du Groupe I est inférieur à celui du Groupe II (respectivement -0,861 et -0,603). Cela signifie que, par rapport aux participants du Groupe I, les participants du Groupe II ont eu plus de gains lexicaux pour les mots cibles à la phase post-inférence.

	contrast	estimate	SE	df	z,ratio	p.value
VKS1	Groupe I – Groupe II	0,047	0,407	Inf	0,115	1,000
VKS2	Groupe I – Groupe II	-0,861	0,394	Inf	-2,185	0,245
VKS3	Groupe I – Groupe II	-0,603	0,394	Inf	-1,531	0,644

En observant de près le Tableau 7.2 plus haut, nous pouvons voir que la différence d'évolution entre les deux groupes se manifeste essentiellement au niveau des réponses des niveaux 1, 2 et 3 :

- Etant 61% pour le Groupe I et 64% pour le Groupe II au VKS1, le poids des réponses du niveau 1 a connu une diminution pour les deux groupes au VKS2

(20% pour le Groupe I vs 13% pour le Groupe II) et au VKS3 (13% pour le Groupe I vs 11% pour le Groupe II). Les diminutions sont plus importantes pour le Groupe II que pour le Groupe I.

- Etant 20% pour le Groupe I et 15% pour le Groupe II au VKS1, le poids des réponses du niveau 2 reste peu varié pour le Groupe I au VKS2 (21%) et au VKS3 (19%), mais diminue pour le Groupe II aux VKS2 (10%) et VKS3 (9%).
- Etant 18% pour le Groupe I et 19% pour le Groupe II au VKS1, le poids des réponses du niveau 3 a augmenté pour le Groupe I aux VKS 2 et VKS3 (respectivement 38% et 45%), également pour le Groupe II (46% au VKS2 et 45% au VKS3). L'augmentation est plus importante au VKS2 pour le Groupe II que pour le Groupe I.

Pour rappel, les niveaux 2 et 3 représentent la connaissance, vague ou certaine, sur la forme d'un mot. Pour cela, les participants du Groupe II semblent avoir eu plus de gains lexicaux sur le plan formel pour les mots cibles, par rapport aux participants du Groupe I.

L'évolution des réponses de divers niveaux de connaissance pour les deux groupes est présentée dans la Figure 7.2 ci-dessous.

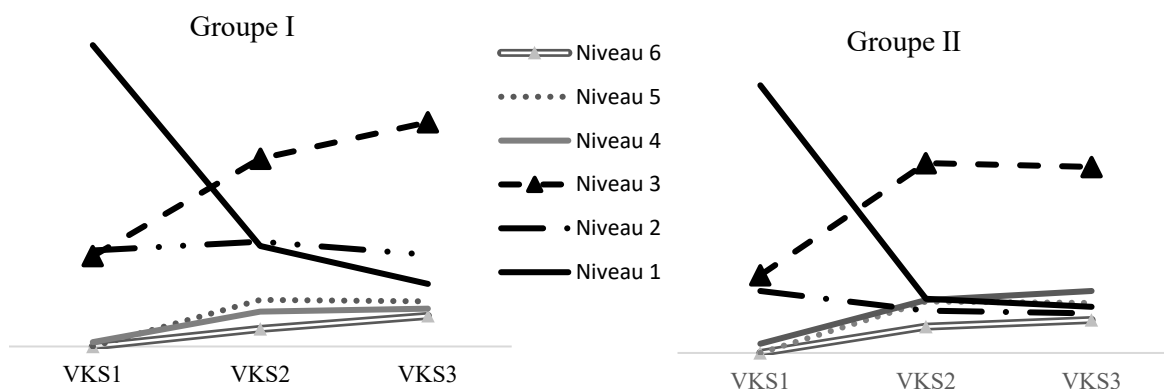


Figure 7.2 – Evolution de la répartition des réponses de différents niveaux de connaissance pour le Groupe I et le Groupe II

7.1.3 Comparaison du niveau de connaissance entre les mots lexicalisés et les mots non lexicalisés

Pour rappel, les six cas exclus de nos analyses concernent les mots lexicalisés.

Nous avons donc eu 324 observations pour les mots lexicalisés et 330 observations pour les mots non lexicalisés. Les résultats pour les mots lexicalisés et les mots non lexicalisés aux trois tests VKS, classés selon l'échelle à six niveaux de connaissance, sont montrés dans le Tableau 7.3.

Mots L	Niveau 1		Niveau 2		Niveau 3		Niveau 4		Niveau 5		Niveau 6		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	218	67%	56	17%	44	14%	6	2%	0	0%	0	0%	324	100%
VKS2	63	19%	60	19%	126	39%	7	2%	42	13%	26	8%	324	100%
VKS3	45	14%	52	16%	136	42%	6	2%	51	16%	34	10%	324	100%

Mots NL	Niveau 1		Niveau 2		Niveau 3		Niveau 4		Niveau 5		Niveau 6		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	192	58%	57	17%	77	24%	4	1%	0	0%	0	0%	330	100%
VKS2	47	14%	44	13%	146	44%	58	18%	29	9%	6	2%	330	100%
VKS3	33	10%	41	12%	159	48%	65	20%	21	7%	11	3%	330	100%

Tableau 7.3 – Nombre et répartition des réponses de différents niveaux aux trois tests VKS pour les mots lexicalisés (L) et non lexicalisés (NL)

Les analyses statistiques montrent que l'effet du facteur « Lexicalisation » seul n'est pas significatif ($p > 0,05$), mais l'effet des deux facteurs, « Lexicalisation » et « Phase », est significatif sur les réponses de divers niveaux de connaissance ($p < 0,05$). Cela signifie que les connaissances lexicales des mots lexicalisés et des mots non lexicalisés ont évolué différemment.

	LR Chisq	Df	Pr(>Chisq)
Lexicalisation	0,19	1	0,6614889
Lexicalisation:Phase	7,39	2	0,0248558

Pour les mots des deux catégories de lexicalisation, le niveau de connaissance en général au VKS1 est significativement inférieur aux niveaux au VKS2 et au VKS3 ($p < 0,000$). Cela signifie que nos participants ont eu des gains lexicaux pour les mots cibles, lexicalisés et non lexicalisés, à la phase post-inférence. L'écart VKS1-VKS2 et celui de VKS1-VKS3 sont plus importants pour les mots lexicalisés (respectivement 3,093 et 3,573) que pour les mots non lexicalisés (respectivement 2,742 et 2,948).

	contrast	estimate	SE	df	z.ratio	p.value
Mots L	VKS1 – VKS2	-3,093	0,180	Inf	-17,195	< 0,000
	VKS1 – VKS3	-3,573	0,184	Inf	-19,378	< 0,000
	VKS2 – VKS3	-0,480	0,152	Inf	-3,168	0,019
Mots NL	VKS1 – VKS2	-2,742	0,166	Inf	-16,489	< 0,000
	VKS1 – VKS3	-2,948	0,168	Inf	-17,577	< 0,000
	VKS2 – VKS3	-0,206	0,141	Inf	-1,467	0,686

En comparant les niveaux de connaissance pour les mots lexicalisés et non lexicalisés aux trois tests VKS, nous pouvons voir qu’au VKS1, les participants ont montré une meilleure connaissance pour les mots non lexicalisés (0,551 supérieur aux mots lexicalisés) ; cet écart a diminué au VKS2 (0,2) ; au VKS3, le niveau de connaissance pour les mots non lexicalisés est inférieur à celui pour les mots lexicalisés (-0,074). Mais ces différences ne sont pas significatives ($p > 0,05$).

	contrast	estimate	SE	df	z.ratio	p.value
VKS1	L - NL	-0,551	0,443	Inf	-1,244	0,815
VKS2	L - NL	-0,200	0,431	Inf	-0,465	0,997
VKS3	L - NL	0,074	0,431	Inf	0,171	1,000

Selon le Tableau 7.3 plus haut, la différence d’évolution de connaissance entre les mots lexicalisés et les mots non lexicalisés se trouve principalement au niveau des réponses des niveaux 4, 5 et 6 :

- Etant respectivement 2% et 1% au VKS1 pour les mots lexicalisés et non lexicalisés, la proportion des réponses du niveau 4 pour les mots lexicalisés demeure peu variée au VKS2 (2%) et au VKS3 (2%), mais celle pour les mots non lexicalisés a une augmentation importante au VKS2 (18%) et au VKS3 (20%) ;
- Etant 0% pour tous les mots au VKS1, la proportion des réponses du niveau 5 a connu une augmentation plus importante pour les mots lexicalisés à la phase post-inférence (13% au VKS2 et 16% au VKS3) que pour les mots non lexicalisés (9% au VKS2 et 6% au VKS3).
- Etant 0% pour tous les mots au VKS1, la proportion des réponses du niveau 6 a connu une augmentation plus importante pour les mots lexicalisés à la phase post-inférence (8% au VKS2 et 10% au VKS3) que pour les mots non lexicalisés (au 2% VKS2 et au 3% VKS3).

La différence d’évolution entre les mots lexicalisés et non lexicalisés est aussi

importante en termes des réponses du niveau 3. Nous discuterons la cause de cette différence dans la Section 7.3.1.2.

L'évolution des réponses de divers niveaux de connaissance pour les mots lexicalisés et non lexicalisés est présentée dans la Figure 7.3 ci-dessous.

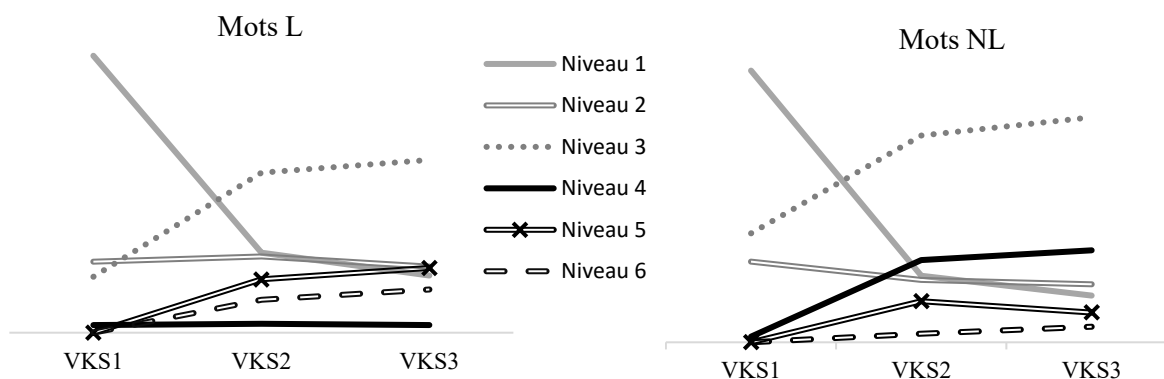


Figure 7.3 – Evolution de la répartition des réponses de différents niveaux de connaissance pour les mots lexicalisés (L) et les mots non lexicalisés (NL)

7.1.4 Comparaison du niveau de connaissance entre les mots lexicalisés et les mots non lexicalisés pour les deux groupes

Les résultats de différents niveaux de connaissance aux trois tests VKS, classés selon la lexicalisation des mots et le groupe d'apprentissage, sont présentés dans le Tableau 7.4.

Groupe I	Niveau 1		Niveau 2		Niveau 3		Niveau 4		Niveau 5		Niveau 6		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
Mots L														
VKS1	111	66%	30	18%	25	15%	2	1%	0	0%	0	0%	168	100%
VKS2	38	23%	39	23%	58	34%	3	2%	19	11%	11	7%	168	100%
VKS3	23	14%	35	21%	68	40%	2	1%	24	14%	16	10%	168	100%
Groupe I														
Mots NL														
VKS1	96	56%	36	21%	37	22%	1	1%	0	0%	0	0%	170	100%
VKS2	31	18%	33	19%	70	41%	22	13%	13	8%	1	1%	170	100%
VKS3	20	12%	28	16%	85	50%	25	15%	7	4%	5	3%	170	100%
Groupe II														
Mots L														
VKS1	107	69%	26	17%	19	12%	4	2%	0	0%	0	0%	156	100%
VKS2	25	16%	21	13%	68	44%	4	2%	23	15%	15	10%	156	100%
VKS3	22	14%	17	11%	68	44%	4	2%	27	17%	18	12%	156	100%

Groupe II Mots NL	Niveau 1		Niveau 2		Niveau 3		Niveau 4		Niveau 5		Niveau 6		Total	
	Nb	%	Nb	%	Nb	%	Nb	Mot	Nb	%	Nb	%	Nb	%
VKS1	96	60%	21	13%	40	25%	3	2%	0	0%	0	0%	160	100%
VKS2	16	10%	11	7%	76	48%	36	22%	16	10%	5	3%	160	100%
VKS3	13	8%	13	8%	74	46%	40	25%	14	9%	6	4%	160	100%

Tableau 7.4 – Nombre et répartition des réponses de différents niveaux de connaissance aux trois tests VKS pour les mots lexicalisés (L) et non lexicalisés (NL) des deux groupes d'apprentissage

Les résultats d'analyse statistique montrent aussi que l'effet des trois facteurs, « Groupe », « Lexicalisation » et « Phase », n'est pas significatif sur les réponses de divers niveaux de connaissance ($p > 0,05$). Cela signifie qu'il n'y a pas de différence significative entre l'évolution des connaissances lexicales pour les mots lexicalisés et non lexicalisés chez les participants du Groupe I et ceux du Groupe II.

	LR Chisq	Df	Pr(>Chisq)
Groupe:Lexicalisation:Phase	0,04	2	0,9812511

Les courbes d'évolution pour les deux groupes, respectivement pour les mots cibles lexicalisés et non lexicalisés, sont présentées dans la Figure 7.4 ci-dessous.

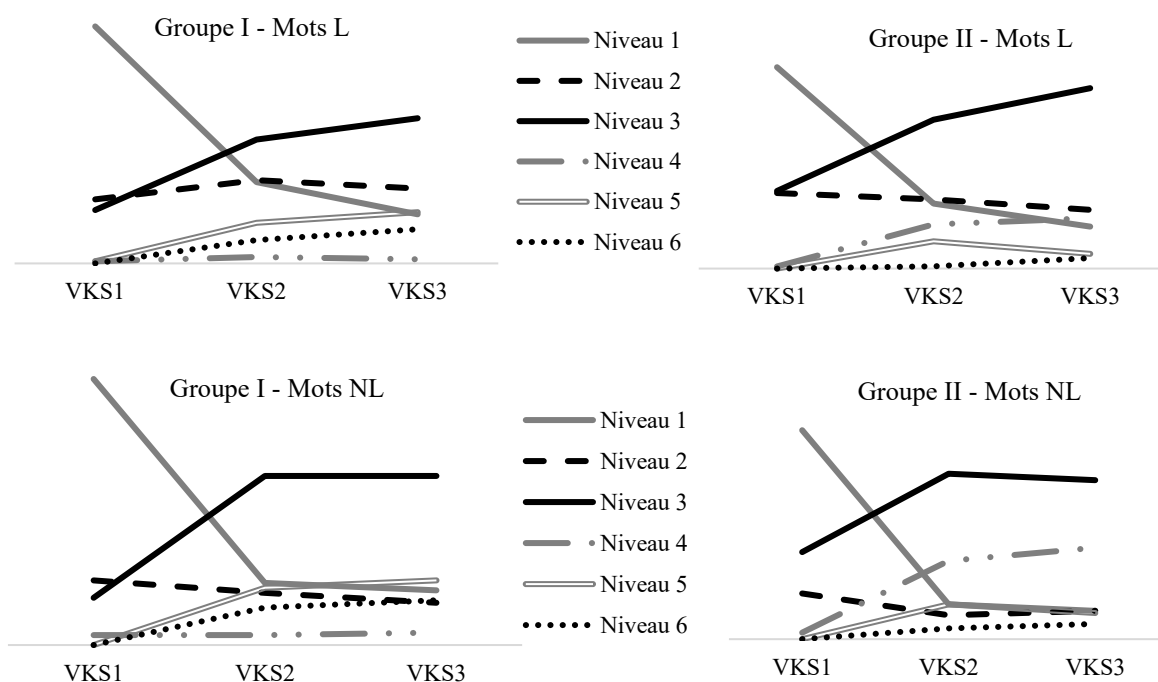


Figure 7.4 – Evolution de la répartition des réponses de différents niveaux de connaissance du Groupe I et du Groupe II pour les mots lexicalisés (L) et non lexicalisés (NL)

7.2 Evolution de la connaissance sémantique pour les mots cibles

Nous avons ensuite observé l'évolution de la connaissance sémantique des mots cibles chez nos participants. Nous avons ainsi analysé les réponses sur le sens des mots cibles obtenues lors du VKS1, de l'inférence lexicale, du VKS2 et du VKS3 (cf. la Section 4.4.2 pour la procédure de notre expérience).

Nous rappelons que, lors d'un test VKS, si un participant note niveau 4, 5 ou 6 pour un mot cible, il doit fournir le sens du mot. Les réponses sur le sens des mots cibles obtenues aux trois tests VKS sont notées selon le même barème que nous avons utilisé pour noter les réponses d'inférence lexicale : une réponse entièrement fautive est notée 0 ; une réponse entièrement correcte est notée 1 ; une réponse partiellement correcte est notée 0,25, 0,5 ou 0,75 selon le rapprochement au sens correct du mot. Nous avons sollicité deux natifs chinois bilingues en français pour effectuer une double correction de la notation. En cas de désaccord sur la notation de certains mots, nous avons opté pour le score le plus plébiscité. Toutes les réponses des niveaux 1, 2 et 3 obtenues lors des trois tests VKS sont notées 0, puisque les participants estiment qu'ils n'ont pas de connaissance sur le sens du mot en question.

Pour rappel, au VKS1, quelques participants ont donné des réponses correctes pour certains mots cibles, mais leurs rapports verbaux lors de l'inférence lexicale ne pouvaient pas justifier leur connaissance complète pour ces mots. Nous avons alors gardé ces mots dans nos observations et, par conséquent, les réponses concernant ces mots ont été rectifiées à 0,5 point.

Nous avons utilisé le modèle mixte linéaire du logiciel SPSS¹¹⁵ pour effectuer les analyses statistiques que nous présentons dans la Section 7.2.¹¹⁶

7.2.1 Evolution globale du score sur le sens des mots cibles

Nous rappelons que nous avons exclu six cas de nos observations parce que

¹¹⁵ Le modèle mixte linéaire permet une analyse par régression des effets des facteurs en mesure répétée (dans notre cas, les facteurs « Phase », « Groupe » et « Lexicalisation ») sur une variable continue (dans notre cas, le score du sens des mots cibles) (cf. <https://www.ibm.com/docs/fr/spss-statistics/version-missing?topic=statistics-linear-mixed-models>, consulté le 9 novembre 2020).

¹¹⁶ Nous remercions sincèrement M. Saïd JMEL et Mme Lucille SOULIER pour leur aide et conseil précieux sur toutes les analyses statistiques faites avec le logiciel SPSS.

certaines mots sont déjà connus par des participants. Nos analyses sont donc basées sur 654 observations (cf. les Annexe 7, 8, 9 et 10 pour les scores sur le sens des mots cibles obtenues respectivement au VKS1, à l'inférence lexicale, au VKS2 et au VKS3).

Les scores sur le sens des mots cibles aux quatre phases sont présentés dans le Tableau 7.5.

	VKS1 (J-7)	Inférence (J0)	VKS2 (J14)	VKS3 (J28)
Nb d'observation	654	654	654	654
Score total	4,25	141,25	134,75	151,5
Score moyen*	0,01	0,22	0,21	0,23
Ecart-type	0,05	0,32	0,38	0,39

* Score maximal = 1

Tableau 7.5 – Score sur le sens de mots cibles aux quatre phases d'apprentissage

Selon les résultats d'analyse statistique, l'effet du facteur « Phase » est significatif sur les scores ($p = 0,00$). Cela signifie que les scores sur le sens des mots cibles aux quatre phases d'apprentissage sont significativement différents.

Tests des effets fixes de type III ^a				
Source	Numérateur ddl	Dénominateur dll	F	Sig.
Phase	3	217,000	53,890	,000

a. Variable dépendante : Score.

Les différences sont significatives respectivement entre le VKS1 et l'inférence ($p = 0,00$), entre le VKS1 et le VKS2 ($p = 0,00$) et entre le VKS1 et le VKS3 ($p = 0,00$). Cela signifie que statistiquement, les participants ont eu une meilleure connaissance du sens des mots cibles à l'inférence et aux VKS2 et VKS3 par rapport au VKS1.

La différence de score entre l'inférence et VKS2 (0,009) et celle entre l'inférence et le VKS3 (-0,016) ne sont pas significatives ($p > 0,05$). Cela signifie que les scores sur le sens des mots cibles aux VKS2 et VKS3 sont statistiquement indifférents de ceux à l'inférence.

Comparaisons par paire^b

(I) Phase	(J) Phase	Différence moyenne (I-J)	Erreur standard	ddl	Sig. ^a	95% Intervalle de confiance pour la différence ^a	
						Borne inférieure	Borne supérieure
VKS1	Inférence	-,205*	,016	2579,841	,000	-,248	-,161
	VKS2	-,196*	,016	2579,841	,000	-,239	-,153
	VKS3	-,221*	,016	2579,841	,000	-,264	-,178
Inférence	VKS2	,009	,016	2579,841	1,000	-,034	,052
	VKS3	-,016	,016	2579,841	1,000	-,060	,027
VKS2	VKS3	-,025	,016	2579,841	,733	-,068	,018

Basé sur les moyennes marginales estimées

*. La différence moyenne est significative au niveau ,05.

a. Ajustement pour des comparaisons multiples : Bonferroni.

b. Variable dépendante : Score.

Le Tableau 7.6 présente le nombre et la répartition des réponses de différents scores sur le sens des mots cibles. Selon le tableau, par rapport à l'inférence (62,1%), la proportion des réponses erronées (score 0) montre une augmentation au VKS2 (74,3%) et au VKS3 (71,3%) ; à la phase post-inférence, la proportion des réponses partiellement correctes (score 0,25~0,75) diminue (10% au VKS2 et 10,8% au VKS3), par rapport à l'inférence (29,2%) ; le nombre des réponses entièrement correctes (score 1) a également augmenté proportionnellement au VKS2 (15,7%) et au VKS3 (17,9%) au VKS3, par rapport à l'inférence (8,7%).

	Score 0		Score 0,25		Score 0,5		Score 0,75		Score 0,25~0,75		Score 1		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	644	98,5%	3	0,4%	7	1,1%	0	0,0%	10	1,5%	0	0,0%	654	100
Inférenc	406	62,1%	70	10,7%	96	14,7%	25	3,8%	19	29,2	57	8,7%	654	100
VKS2	486	74,3%	18	2,8%	32	4,9%	15	2,3%	65	10,0	103	15,7	654	100
VKS3	466	71,3%	21	3,2%	33	5,0%	17	2,6%	71	10,8	117	17,9	654	100

Tableau 7.6 – Nombre et répartition des réponses par échelon de score aux quatre phases d'apprentissage

Les scores moyens pour les 20 mots cibles aux quatre phases sont montrés dans le Tableau 7.7 plus bas. Selon le tableau, à la phase post-inférence (VKS2 et VKS3), les cinq mots ayant les meilleurs scores sont MC07, MC12, MC16, MC17 et MC20 et ceux ayant les scores les plus faibles sont MC01, MC03, MC06, MC15 et MC19. Parmi les 20 mots cibles, MC12 a les meilleurs scores et MC03 a les scores

les plus faibles aux VKS2 et VK3.

	VKS1*	Inférence*	VKS2*	VKS3*
MC01	0,00	0,03	0,03	0,06
MC02	0,00	0,19	0,14	0,16
MC03	0,00	0,00	0,02	0,03
MC04	0,00	0,45	0,22	0,23
MC05	0,01	0,29	0,15	0,21
MC06	0,00	0,08	0,03	0,03
MC07	0,00	0,24	0,39	0,39
MC08	0,02	0,06	0,09	0,16
MC09	0,00	0,12	0,12	0,15
MC10	0,03	0,30	0,24	0,30
MC11	0,02	0,30	0,15	0,12
MC12	0,02	0,40	0,63	0,78
MC13	0,02	0,34	0,28	0,30
MC14	0,00	0,26	0,14	0,17
MC15	0,00	0,09	0,08	0,08
MC16	0,00	0,06	0,27	0,36
MC17	0,01	0,11	0,42	0,44
MC18	0,02	0,45	0,21	0,24
MC19	0,00	0,20	0,08	0,09
MC20	0,01	0,35	0,41	0,32

* Score maximal = 1

Tableau 7.7 – Score sur le sens des mots cibles en moyen aux quatre phases d'apprentissage, classé par mot

Selon les résultats statistiques, L'effet des deux facteurs, « Mot » et « Phase », est significatif sur les scores ($p = 0,00$). Cela signifie que les scores du sens des mots cibles ont évolué différemment du VKS1 au VKS3.

Tests des effets fixes de type III ^a				
Source	Ddl du numérateur	Ddl du dénominateur	F	Sig.
Mot * Phase		57	2424,838	4,755 ,000

a. Variable dépendante : VKS.

Les scores sur le sens des mots MC02, MC04, MC05, MC07, MC10, MC11, MC12, MC13, MC14, MC16, MC17, MC18, MC19 et MC20 ont une évolution significative ($p < 0,05$), du VKS1 au VKS3.

Tests univariés ^a				
Mot	Ddl du numérateur	Ddl du dénominateur	F	Sig.
MC01	3	2424,839	,288	,834

MC02	3	2424,834	3,197	,023
MC03	3	2424,839	,114	,952
MC04	3	2424,836	15,741	,000
MC05	3	2424,839	6,621	,000
MC06	3	2424,834	,485	,693
MC07	3	2424,839	15,991	,000
MC08	3	2424,842	1,583	,191
MC09	3	2424,839	2,119	,096
MC10	3	2424,839	6,613	,000
MC11	3	2424,839	6,667	,000
MC12	3	2424,844	49,262	,000
MC13	3	2424,839	9,549	,000
MC14	3	2424,837	4,796	,002
MC15	3	2424,839	,869	,457
MC16	3	2424,839	13,881	,000
MC17	3	2424,834	22,458	,000
MC18	3	2424,839	13,391	,000
MC19	3	2424,840	3,354	,018
MC20	3	2424,835	15,292	,000

Les tests de F permettent de tester les effets simples de Phase dans chaque combinaison de niveaux des autres effets affichés. Ils s'appuient sur les comparaisons appariées (indépendantes) linéaires parmi les moyennes marginales estimées.^a

a. Variable dépendante : Score.

Par rapport aux scores d'inférence lexicale, aux VKS2 et VKS3, les scores sur le sens des mots MC12, MC16 et MC17 augmentent significativement ($p < 0,05$) ; les scores des mots MC04, MC11 et MC18 montrent une diminution significative ($p < 0,05$).

Comparaisons appariées^a

Mot	(I) Phase	(J) Phase	Différence moyenne (I-J)	Erreur standard	ddl	Sig. ^c	Intervalle de confiance à 95 % pour la différence ^c	
							Borne inférieure	Borne supérieure
MC01	Inférence	VKS1	0,029	0,065	2424,8	I	-0,143	0,202
		VKS2	-3,33E-16	0,065	2424,8	I	-0,172	0,172
		VKS3	-0,031	0,065	2424,8	I	-0,204	0,141
MC02	Inférence	VKS1	,176*	0,065	2424,8	0,04	0,004	0,349
		VKS2	0,032	0,065	2424,8	I	-0,141	0,204
		VKS3	0,009	0,065	2424,8	I	-0,164	0,181
MC03	Inférence	VKS1	-1,67E-16	0,065	2424,8	I	-0,172	0,172
		VKS2	-0,023	0,065	2424,8	I	-0,195	0,15
		VKS3	-0,03	0,065	2424,8	I	-0,203	0,142
MC04	Inférence	VKS1	,449*	0,065	2424,8	0	0,276	0,621
		VKS2	,234*	0,065	2424,8	0,00	0,062	0,407
		VKS3	,218*	0,065	2424,8	0,00	0,045	0,39
MC05	Inférence	VKS1	,281*	0,065	2424,8	0	0,108	0,453
		VKS2	0,136	0,065	2424,8	0,22	-0,037	0,308
		VKS3	0,076	0,065	2424,8	I	-0,096	0,249
MC06	Inférence	VKS1	0,078	0,065	2424,8	I	-0,095	0,25
		VKS2	0,048	0,065	2424,8	I	-0,124	0,221
		VKS3	0,048	0,065	2424,8	I	-0,124	0,221
		VKS1	,244*	0,065	2424,8	0,00	0,072	0,416

MC07	Inférence	VKS2	-0,151	0,065	2424,8	0,12	-0,324	0,021
		VKS3	-0,143	0,065	2424,8	0,16	-0,316	0,029
MC08	Inférence	VKS1	0,047	0,066	2424,8	I	-0,128	0,222
		VKS2	-0,031	0,066	2424,8	I	-0,206	0,144
MC09	Inférence	VKS3	-0,094	0,066	2424,8	0,94	-0,269	0,081
		VKS1	0,123	0,065	2424,8	0,35	-0,049	0,296
MC10	Inférence	VKS2	0,002	0,065	2424,8	I	-0,171	0,174
		VKS3	-0,028	0,065	2424,8	I	-0,2	0,145
MC11	Inférence	VKS1	,246*	0,065	2424,8	0,00	0,073	0,418
		VKS2	0,047	0,065	2424,8	I	-0,125	0,22
MC12	Inférence	VKS3	-0,007	0,065	2424,8	I	-0,18	0,165
		VKS1	,289*	0,065	2424,8	0	0,116	0,461
MC13	Inférence	VKS2	0,152	0,065	2424,8	0,12	-0,021	0,324
		VKS3	,183*	0,065	2424,8	0,03	0,01	0,355
MC14	Inférence	VKS1	,377*	0,066	2424,8	0	0,202	0,552
		VKS2	-,231*	0,066	2424,8	0,00	-0,406	-0,055
MC15	Inférence	VKS3	-,380*	0,066	2424,8	0	-0,555	-0,205
		VKS1	,321*	0,065	2424,8	0	0,148	0,493
MC16	Inférence	VKS2	0,074	0,065	2424,8	I	-0,099	0,246
		VKS3	0,053	0,065	2424,8	I	-0,12	0,225
MC17	Inférence	VKS1	,260*	0,07	2424,8	0,00	0,075	0,444
		VKS2	0,12	0,07	2424,8	0,51	-0,065	0,305
MC18	Inférence	VKS3	0,082	0,07	2424,8	I	-0,103	0,267
		VKS1	0,09	0,065	2424,8	I	-0,083	0,262
MC19	Inférence	VKS2	0,006	0,065	2424,8	I	-0,166	0,178
		VKS3	0,005	0,065	2424,8	I	-0,167	0,178
MC20	Inférence	VKS1	0,061	0,065	2424,8	I	-0,112	0,233
		VKS2	-,211*	0,065	2424,8	0,00	-0,384	-0,039
MC07	Inférence	VKS3	-,303*	0,065	2424,8	0	-0,476	-0,131
		VKS1	0,099	0,065	2424,8	0,78	-0,074	0,271
MC08	Inférence	VKS2	-,316*	0,065	2424,8	0	-0,489	-0,144
		VKS3	-,330*	0,065	2424,8	0	-0,502	-0,158
MC09	Inférence	VKS1	,413*	0,065	2424,8	0	0,24	0,585
		VKS2	,229*	0,065	2424,8	0,00	0,056	0,401
MC10	Inférence	VKS3	,198*	0,065	2424,8	0,01	0,025	0,37
		VKS1	,205*	0,065	2424,8	0,01	0,033	0,377
MC11	Inférence	VKS2	0,129	0,065	2424,8	0,29	-0,044	0,301
		VKS3	0,113	0,065	2424,8	0,50	-0,059	0,286
MC12	Inférence	VKS1	,341*	0,065	2424,8	0	0,169	0,514
		VKS2	-0,065	0,065	2424,8	I	-0,237	0,108
MC13	Inférence	VKS3	0,034	0,065	2424,8	I	-0,139	0,206

Basées sur les moyennes marginales estimées

*. La différence moyenne est significative au niveau ,05.

a. Variable dépendante :Score.

c. Ajustement pour les comparaisons multiples : Bonferroni.

7.2.2 Comparaison du score sur le sens des mots cibles entre le Groupe I et le Groupe II

Les scores pour les deux groupes à quatre phases d'apprentissage sont montrés dans le Tableau 7.8 plus bas. Selon le tableau, les scores moyens du Groupe II aux trois dernières phases sont respectivement supérieurs à ceux pour le Groupe I.

		VKS1	Inférence	VKS2	VKS3
Groupe I	Nb d'observation	338	338	338	338
	Score total	1,5	59	55,25	65,5
	Score en	0,00	0,17	0,16	0,19
Groupe II	Nb d'observation	316	316	316	316
	Score total	2,75	82,25	79,5	86
	Score en	0,01	0,26	0,25	0,27
Total	Nb d'observation	654	654	654	654
	Score total	4,25	141,25	134,75	151,5
	Score en	0,01	0,22	0,21	0,23

* Score maximal = 1

Tableau 7.8 – Score sur le sens des mots cibles du Groupe I et du Groupe II aux quatre phases d'apprentissage

Le score d'inférence lexicale en moyenne du Groupe II devance celui du Groupe I de 0,09 point. Selon nous, il s'agit d'un fait aléatoire : A l'occasion du test VKS1, selon la durée d'expérience aux pré-tests et l'emploi du temps de nos participants, nous avons proposé aux participants un rendez-vous individuel, d'une durée de deux heures. Le jour du rendez-vous, à la fin des lectures et de l'inférence pour les 20 mots cibles, en fonction du temps qu'il restait et de l'état de fatigue du participant, nous avons alors affecté généralement au Groupe I les participants ayant passé plus de temps pour finir les premières tâches ou ceux étant fatigués à ce moment-là. Les autres participants ayant plus rapidement terminé les premières tâches ont alors accompli les tâches post-inférence du Groupe II, la consultation d'un dictionnaire bilingue chinois-français pour retrouver les sens des mots cibles et puis accomplir la tâche lacunaire. Il est possible que les participants ayant terminé plus rapidement les tâches de lecture et d'inférence lexicale aient une meilleure maîtrise de caractères chinois et/ou un meilleur niveau de compréhension en chinois L2. Une autre raison possible est l'état de fatigue de certains participants du Groupe I qui les a empêchés de donner de leur mieux. Il peut ainsi en résulter que le score moyen d'inférence lexicale du Groupe II est supérieur à celui du Groupe I.

Selon les résultats statistiques, l'effet du facteur « Groupe » sur les scores n'est pas significatif ($p > 0,05$). L'effet des deux facteurs, « Groupe » et « Phase », n'est pas significatif non plus ($p > 0,05$). Cela signifie que l'évolution des scores sur le

sens des mots cibles du Groupe I n'est statistiquement pas différente de celle du Groupe II.

Tests des effets fixes de type III ^a				
Source	Numérateur ddl	Dénominateur ddl	F	Sig.
Groupe	1	31,000	1,918	,176
Groupe * Phase	3	217,000	1,679	,172

a. Variable dépendante : Score.

Les courbes d'évolution des scores sur le sens des mots cibles en moyenne pour les deux groupes sont montrées dans la Figure 7.5.

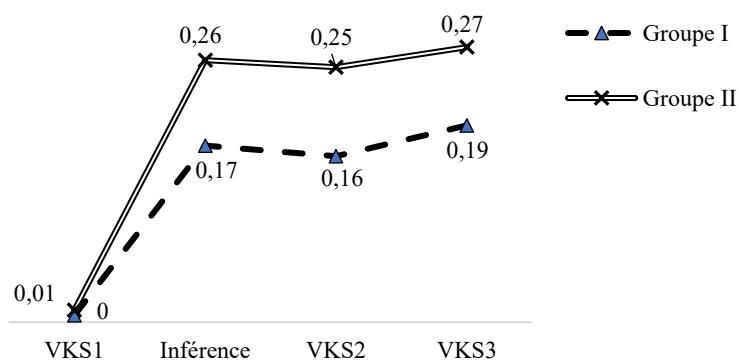


Figure 7.5 – Evolutions du score moyen sur le sens des mots cibles pour le Groupe I et le Groupe II

Le Tableau 7.9 présente le nombre et la répartition des réponses de différents scores du Groupe I et du Groupe II.

Groupe I	Score 0		Score 0,25		Score 0,5		Score 0,75		Score 1		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	335	99%	0	0%	3	1%	0	0%	0	0%	338	100%
Inférence	233	69%	31	9%	40	12%	11	3%	23	7%	338	100%
VKS2	269	79%	9	3%	12	4%	4	1%	44	13%	338	100%
VKS3	259	77%	7	2%	13	4%	7	2%	52	15%	338	100%

Groupe II	Score 0		Score 0,25		Score 0,5		Score 0,75		Score 1		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	309	98%	3	1%	4	1%	0	0%	0	0%	316	100%
Inférence	173	55%	39	12%	56	18%	14	4%	34	11%	316	100%
VKS2	217	69%	9	3%	20	6%	11	3%	59	19%	316	100%
VKS3	207	66%	14	4%	20	6%	10	3%	65	21%	316	100%

Tableau 7.9 – Nombre et répartition des réponses de différents scores du Groupe I et du Groupe II aux quatre phases d'apprentissage

7.2.3 Comparaison du score sur le sens entre les mots lexicalisés et les mots non lexicalisés

Les scores moyens pour les mots lexicalisés et non lexicalisés pour les participants aux quatre phases d'apprentissage sont détaillés dans le Tableau 7.10 ci-dessous.

Participant	VKS1*		Inférence*		VKS2*		VKS3*	
	Mots L	Mots NL	Mots L	Mots NL	Mots L	Mots NL	Mots L	Mots NL
P01	0,05	0,03	0,43	0,43	0,70	0,35	0,65	0,30
P02	0,00	0,00	0,20	0,35	0,45	0,33	0,45	0,38
P03	0,00	0,05	0,00	0,13	0,20	0,28	0,20	0,03
P04	0,00	0,00	0,00	0,18	0,20	0,05	0,23	0,00
P05	0,00	0,00	0,00	0,05	0,23	0,10	0,20	0,10
P06	0,00	0,00	0,55	0,10	0,40	0,23	0,40	0,20
P07	0,05	0,00	0,25	0,33	0,20	0,23	0,40	0,20
P08	0,00	0,00	0,18	0,33	0,20	0,20	0,30	0,20
P09	0,00	0,00	0,05	0,03	0,10	0,00	0,10	0,00
P10	0,00	0,05	0,40	0,40	0,10	0,40	0,10	0,40
P11	0,00	0,00	0,03	0,08	0,10	0,13	0,20	0,10
P12	0,00	0,00	0,00	0,13	0,00	0,00	0,00	0,00
P13	0,00	0,00	0,18	0,35	0,40	0,18	0,40	0,18
P14	0,00	0,00	0,00	0,10	0,10	0,18	0,20	0,18
P15	0,08	0,03	0,25	0,35	0,14	0,23	0,14	0,23
P16	0,00	0,00	0,20	0,20	0,20	0,15	0,20	0,20
P17	0,00	0,00	0,00	0,03	0,00	0,00	0,10	0,00
P18	0,00	0,00	0,18	0,20	0,00	0,00	0,10	0,00
P19	0,00	0,00	0,20	0,23	0,05	0,28	0,20	0,25
P20	0,00	0,00	0,00	0,13	0,00	0,03	0,00	0,03
P21	0,00	0,00	0,00	0,03	0,00	0,00	0,60	0,13
P22	0,00	0,00	0,22	0,30	0,25	0,25	0,25	0,33
P23	0,00	0,00	0,13	0,23	0,20	0,00	0,10	0,00
P24	0,00	0,00	0,00	0,28	0,00	0,00	0,00	0,00
P25	0,00	0,00	0,05	0,20	0,00	0,28	0,10	0,28
P26	0,00	0,00	0,33	0,35	0,50	0,18	0,60	0,25
P27	0,00	0,00	0,00	0,03	0,10	0,20	0,10	0,13
P28	0,00	0,00	0,25	0,40	0,38	0,40	0,28	0,38
P29	0,00	0,00	0,40	0,33	0,10	0,00	0,10	0,00
P30	0,05	0,00	0,35	0,53	0,15	0,13	0,30	0,33
P31	0,00	0,00	0,36	0,45	0,33	0,20	0,33	0,23
P32	0,00	0,00	0,44	0,38	0,67	0,68	0,78	0,68
P33	0,06	0,00	0,69	0,55	1,00	0,78	1,00	0,75

* Score maximal = 1

Tableau 7.10 – Score sur le sens des mots lexicalisés (L) et non lexicalisés (NL) en moyenne aux quatre phases d'apprentissage, classé par participant

Les scores moyens des mots cibles lexicalisés et non lexicalisés aux quatre phrases d'apprentissage sont montrés dans le Tableau 7.11. Selon les résultats, le score d'inférence lexicale en moyenne pour les mots lexicalisés (0,19) est inférieur à celui pour les mots non lexicalisés (0,25), mais supérieur au VKS2 (0,22 pour les mots lexicalisés vs 0,19 pour les mots non lexicalisés) et au VKS3 (0,27 pour les mots lexicalisés vs 0,19 pour les mots non lexicalisés).

		VKS 1	Inférence	VKS 2	VKS 3
Mots L	Nb d'observation	324	324	324	324
	Score total	2,75	60,25	71	87,5
	Score moyen*	0,01	0,19	0,22	0,27
Mots NL	Nb d'observation	330	330	330	330
	Score total	1,5	81	63,75	64
	Score moyen*	0,01	0,25	0,19	0,19
Total	Nb d'observation	654	654	654	654
	Score total	4,25	141,25	134,75	151,5
	Score moyen*	0,01	0,22	0,21	0,23

* Score maximal = 1

Tableau 7.11 – Score sur le sens pour les mots lexicalisés (L) et non lexicalisés (NL) aux quatre phases d'apprentissage

Selon les analyses statistiques, l'effet du facteur « Lexicalisation » n'est pas significatif sur les scores ($p > 0,05$). Cependant, l'effet des deux facteurs, « Lexicalisation » et « Phase », est significatif ($p = 0,01$). Cela signifie que l'évolution des scores sur le sens des mots cibles lexicalisés est différente de celle des mots non lexicalisés.

Tests des effets fixes de type III ^a				
Source	Numérateur ddl	Dénominateur dll	F	Sig.
Lexicalisation	1	217,000	,913	,340
Lexicalisation * Phase	3	217,000	3,598	,014

a. Variable dépendante : Score.

Les analyses par Test-T apparié¹¹⁷ montrent que la différence des scores moyens des mots lexicalisés et non lexicalisés est significative à l'inférence ($p < 0,05$) et au VKS3 ($p < 0,01$).

¹¹⁷ Test-T pour échantillons appariés (bilatéral) du logiciel SPSS

		Intervalle de confiance 95% de la différence							
		Moyenne	Ecart- type	Erreur standard moyenne	Inférieure	Supérieure	t	ddl	Sig.
Inférence	L - NL	-,05667	,12477	,02172	-,10091	-,01243	-2,609	32	,014
VKS2	L - NL	,03030	,14888	,02592	-,02249	,08309	1,169	32	,251
VKS3	L - NL	,08000	,15660	,02726	,02447	,13553	2,935	32	,006

Les courbes d'évolution des scores moyens aux quatre phases pour les mots lexicalisés et non lexicalisés sont montrées dans la Figure 7.6.

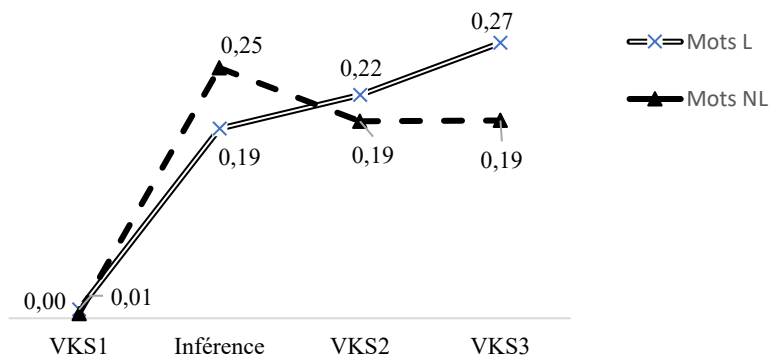


Figure 7.6 – Evolutions du score moyen sur le sens pour les mots lexicalisés (L) et non lexicalisés (NL)

En observant de plus près les scores des mots cibles (Tableau 7.12 plus bas), nous pouvons voir qu'à l'inférence lexicale, il y a eu plus de mauvaises réponses (score 0) pour les mots lexicalisés que pour les mots non lexicalisés (76% vs 48%) ; aux VKS2 et VKS3, la proportion des mauvaises réponses pour les mots lexicalisés change peu (respectivement 77% et 72%), alors que pour les mots non lexicalisés, la proportion des mauvaises réponses a connu une augmentation importante (respectivement 72% et 71%). Quant aux réponses partiellement correctes (scores 0,25~0,75), pour les mots des deux catégories, aux VKS2 et VKS3, la proportion diminue. En termes des réponses entièrement correctes (score 1), la proportion augmente aux VKS2 et VKS3 pour les mots lexicalisés et non lexicalisés.

Les participants ont donné plus de réponses partiellement correctes (score 0,25~0,75) pour les mots non lexicalisés que pour les mots lexicalisés à l'inférence (48% vs 10%), au VKS2 (17% vs 2%) et au VKS3 (19% vs 2%). Toutefois, ils ont donné plus de réponses entièrement correctes (score 1) pour les mots lexicalisés que

pour les mots non lexicalisés à l'inférence (14% vs 4%), au VKS2 (21% vs 11%) et au VKS3 (26% vs 10%).

Mots L	Score 0		Score 0,25		Score 0,5		Score 0,75		Score 1		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	318	98%	1	0%	5	2%	0	0%	0	0%	324	100%
Inférence	246	76%	11	3%	16	5%	6	2%	45	14%	324	100%
VKS2	249	77%	3	1%	3	1%	1	0%	68	21%	324	100%
VKS3	233	72%	3	1%	2	1%	1	0%	85	26%	324	100%

Mots NL	Score 0		Score 0,25		Score 0,5		Score 0,75		Score 1		Total	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
VKS1	326	98%	2	1%	2	1%	0	0%	0	0%	330	100%
Inférence	160	48%	59	18%	80	24%	19	6%	12	4%	330	100%
VKS2	237	72%	15	4%	29	9%	14	4%	35	11%	330	100%
VKS3	233	71%	18	5%	31	9%	16	5%	32	10%	330	100%

Tableau 7.12 – Nombre et répartition des réponses de différents scores pour les mots lexicalisés (L) et non lexicalisés (NL) aux quatre phases d'apprentissage

L'évolution des scores sur le sens aux quatre phases d'apprentissage pour les mots lexicalisés est détaillée dans le Tableau 7.13 ci-dessous. Parmi les dix mots lexicalisés, l'augmentation des réponses entièrement correctes (score 1) est plus importante pour les mots MC12, MC16 et MC17 que pour les autres mots.

	Total	Nb. de réponse selon l'échelle de score														
		Inférence					VKS2					VKS3				
		0	0,25	0,5	0,75	1	0	0,25	0,5	0,75	1	0	0,25	0,5	0,75	1
MC01	33	32	0	0	0	1	32	0	0	0	1	31	0	0	0	2
MC06	33	28	1	3	1	0	32	0	0	0	1	32	0	0	0	1
MC08	32	26	5	0	1	0	29	0	0	0	3	27	0	0	0	5
MC09	33	29	0	0	0	4	29	0	0	0	4	28	0	0	0	5
MC10	33	23	0	0	0	10	25	0	0	0	8	23	0	0	0	10
MC12	32	13	2	9	1	7	10	1	2	1	18	6	1	0	1	24
MC14	29	21	0	1	0	7	25	0	0	0	4	24	0	0	0	5
MC16	33	31	0	0	0	2	24	0	0	0	9	21	0	0	0	12
MC17	33	26	3	2	1	1	17	2	1	0	13	16	2	2	0	13
MC18	33	17	0	1	2	13	26	0	0	0	7	25	0	0	0	8
Total	324	246	11	16	6	45	249	3	3	1	68	233	3	2	1	85
%	100%	76%	3%	5%	2%	14%	77%	1%	1%	0%	21%	72%	1%	1%	0%	26%

Tableau 7.13 – Nombre et répartition des réponses de différents scores pour les mots lexicalisés (L)

MC01	节俭	L	(<i>jiéjiǎn</i>) : économe, frugal
MC06	雀跃	L	(<i>quèyuè</i>) : sauter de joie
MC08	在乎	L	(<i>zàihu</i>) : prendre qch. à cœur ; faire attention à
MC09	乏味	L	(<i>fáwèi</i>) : fade, insipide
MC10	盼望	L	(<i>pànwàng</i>) : espérer, désirer
MC12	五颜六色	L	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC14	彩虹	L	(<i>cǎihóng</i>) : arc-en-ciel
MC16	风筝	L	(<i>fēngzhēng</i>) : cerf-volant
MC17	蝉	L	(<i>chán</i>) : cigale
MC18	闲聊	L	(<i>xiánliáo</i>) : bavarder

L'évolution des scores sur le sens aux quatre phases d'apprentissage pour les mots non lexicalisés est détaillée dans le Tableau 7.14. Parmi les dix mots non lexicalisés, l'augmentation des réponses entièrement correctes (score 1) est plus importante pour le mot MC07 que pour les autres mots non lexicalisés.

	Total	Nb. de réponse selon l'échelle de score														
		Inférence					VKS2					VKS3				
		0	0,25	0,5	0,75	1	0	0,25	0,5	0,75	1	0	0,25	0,5	0,75	1
MC02	33	14	15	3	0	1	24	3	4	0	2	23	4	3	1	2
MC03	33	33	0	0	0	0	30	3	0	0	0	30	2	1	0	0
MC04	33	10	1	11	8	3	23	1	2	4	3	24	0	1	3	5
MC05	33	1	29	0	3	0	25	1	3	3	1	22	3	1	5	2
MC07	33	19	0	11	2	1	19	0	1	2	11	20	0	0	1	12
MC11	33	13	0	20	0	0	24	0	8	0	1	25	0	8	0	0
MC13	33	13	4	10	3	3	19	4	3	1	6	16	4	7	2	4
MC15	33	26	3	3	1	0	29	0	1	3	0	29	0	1	3	0
MC19	33	21	1	8	2	1	30	0	1	0	2	29	0	2	0	2
MC20	33	10	6	14	0	3	14	3	6	1	9	15	5	7	1	5
Total	330	160	59	80	19	12	237	15	29	14	35	233	18	31	16	32
%	100%	48%	18%	24%	6%	4%	72%	5%	9%	4%	11%	71%	5%	9%	5%	10%

Tableau 7.14 – Nombre et répartition des réponses de différents scores pour les mots non lexicalisés (NL)

MC02	窗花	NL	(<i>chuānghuā</i>) : papier découpé pour la décoration d'une fenêtre
MC03	送灶	NL	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC04	腊月	NL	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC05	汤圆	NL	(<i>tāngyuán</i>) : boulette de riz glutineux farcie de viande (fourrée à la confiture)
MC07	还礼	NL	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC11	安适	NL	(<i>ānshì</i>) : tranquille et confortable

MC13	明丽	NL	(<i>mínglì</i>) : clair et beau
MC15	踏青	NL	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps
MC19	飘零	NL	(<i>piāolíng</i>) : tomber en voltigeant
MC20	白茫茫	NL	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

7.2.4 Comparaison du score sur le sens des mots lexicalisés et non lexicalisés pour les deux groupes

L'effet des trois facteurs, « Phase », « Groupe » et « Lexicalisation », n'est pas significatif ($p > 0,05$) sur les scores. Cela signifie que les évolutions des scores du sens pour les mots lexicalisés et non lexicalisés, respectivement du Groupe I et du Groupe II, ne sont pas statiquement différentes.

Tests des effets fixes de type III^a

Source	Numérateur ddl	Dénominateur dll	F	Sig.
Groupe * Lexicalisation * Phase	4	217,000	,586	,673

a. Variable dépendante : Score.

Les courbes d'évolution sont montrées dans la Figure 7.7. Malgré la non significativité, nous pouvons constater qu'au VKS2, par rapport au Groupe II, le Groupe I a eu une diminution plus importante pour le score moyen des mots non lexicalisés.

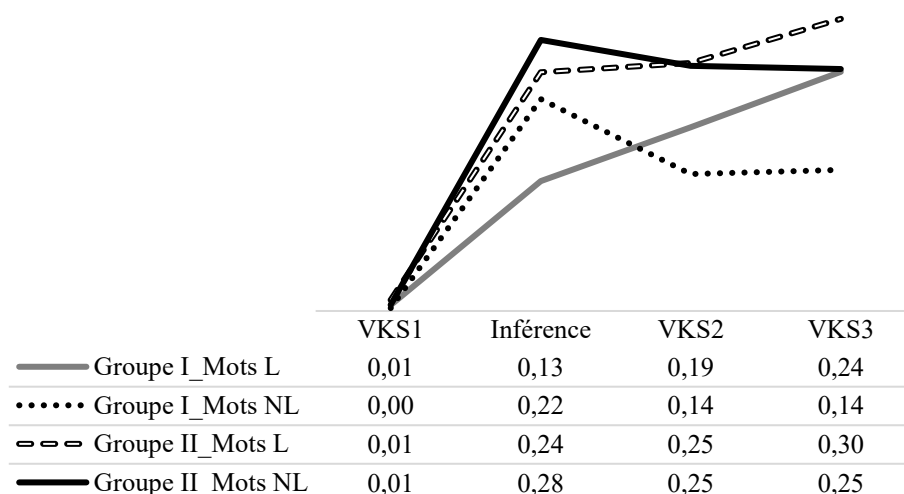


Figure 7.7 – Evolution du score moyen sur le sens des mots lexicalisés (L) et des mots non lexicalisés (NL) pour le Groupe I et le Groupe II

7.3 Discussion

7.3.1 Evolution des connaissances multidimensionnelles des mots cibles

Nous avons analysé l'évolution des connaissances lexicales pour les mots cibles chez nos participants, par le biais de l'échelle à six niveaux de connaissance du test VKS. Les niveaux de connaissance reflètent les connaissances lexicales sur la forme, sur le sens et sur l'emploi d'un mot.

Par rapport au VKS1, les connaissances lexicales chez nos participants pour les mots cibles ont positivement évolué du VKS1 au VKS3. Les niveaux de connaissance des mots cibles chez nos participants au VKS2 et au VKS3 sont significativement supérieurs à celui au VKS1. Cela signifie que l'inférence lexicale et les tâches post-inférence ont un effet positif sur les gains lexicaux pour les mots cibles.

7.3.1.1 Différence d'évolution du niveau de connaissance entre le Groupe I et le Groupe II

Pour les deux groupes, les niveaux de connaissance au VKS2 et au VKS3 sont supérieurs à celui au VKS1. Cela signifie que, dans notre étude, quelle que soit la condition d'apprentissage à laquelle nous avons soumis les participants, ces derniers en ont globalement tiré profit (à savoir l'inférence lexicale et les exercices de vocabulaire) et ont obtenu des gains lexicaux pour les mots cibles.

L'évolution du niveau de connaissance des mots cibles chez les participants du Groupe I est différente de celle chez les participants du Groupe II. Le niveau global du Groupe I est inférieur à celui du Groupe II aux VKS2 et VKS3. Cela signifie que les participants du Groupe II ont eu davantage de gains lexicaux pour les mots cibles que ceux du Groupe I.

La différence entre les deux groupes se situe principalement sur l'évolution des réponses du niveau 1 (« jamais-vu ») et des niveaux 2 et 3 (la connaissance vague ou certaine de la forme d'un mot). Par rapport au Groupe I, le Groupe II a une diminution plus importante pour les réponses du niveau 1 et du niveau 2, et une

augmentation plus importante pour les réponses du niveau 3. Les participants du Groupe II semblent donc avoir eu plus de gains lexicaux au niveau de la forme pour les mots cibles, par rapport à ceux du Groupe I.

7.3.1.2 Différence d'évolution du niveau de connaissance entre les mots lexicalisés et les mots non lexicalisés

Pour tous les mots cibles, lexicalisés et non lexicalisés, les niveaux de connaissance globaux aux VKS2 et VKS3 sont significativement supérieurs à celui au VKS1. Nos participants ont donc eu des gains lexicaux pour les mots des deux catégories de lexicalisation à la suite de l'inférence lexicale et des tâches de vocabulaire.

Au VKS1, le niveau de connaissance global estimé par nos participants pour les mots cibles non lexicalisés est supérieur à celui pour les mots lexicalisés. L'écart de niveau entre les deux catégories diminue au fur et à mesure. Au VKS3, le niveau pour les mots non lexicalisés est légèrement inférieur à celui pour les mots lexicalisés. Les différences de niveaux de connaissance globaux aux VKS2 et VKS3 entre les mots lexicalisés et non lexicalisés ne sont pas statistiquement significatives. Toutefois, aux VKS2 et VKS3, nos participants ont eu d'une part plus de réponses du niveau 4 (la connaissance partielle sur le sens) pour les mots non lexicalisés que pour les mots lexicalisés, et d'autre part plus de réponses du niveau 5 (la connaissance complète sur le sens) et du niveau 6 (la connaissance sur l'emploi du mot) pour les mots lexicalisés que pour les mots non lexicalisés.

L'évolution du niveau de connaissance pour les mots lexicalisés est significativement différente de celle pour les mots non lexicalisés. Cependant, nous pensons que nos participants ont eu une surestimation de leur connaissance sur les mots cibles non lexicalisés au VKS1. Le test VKS demande à un apprenant d'estimer lui-même son niveau de connaissance pour un mot donné présenté isolément. L'auto-évaluation ne reflète pas toujours le niveau réel de l'apprenant (cf. Laufer, 2006). Notamment pour la connaissance sur la forme du mot (qui correspond aux niveaux 2 et 3 sur l'échelle de connaissances utilisée dans notre étude), il n'y a pas de moyen pour l'apprenant de justifier sa réponse. Or, il n'est pas rare qu'un apprenant du

chinois L2, notamment de bas niveau, confonde les caractères dont les formes sont similaires et/ou confonde les mots constitués d'un même caractère (cf. Feng, 2009). Dans notre étude, par exemple, dans la liste de 60 mots chinois du VKS1, nous avons inclus le mot « 欢喜 (*huānxǐ*, content) », un mot non courant et inconnu pour les participants ; cependant, seulement 6 participants sur 33 ont noté niveau 1 (« jamais-vu ») pour ce mot. 19 participants ont noté niveau 6 (« Je connais ce mot et je sais qu'il veut dire... Je sais écrire une phrase en chinois avec ce mot. »), 1 participant a noté niveau 5 (« Je connais ce mot et je sais qu'il veut dire... ») et 2 participants ont noté niveau 4 (« J'ai déjà vu ce mot. Il veut peut-être dire ..., mais je n'en suis pas sûr(e). ») pour ce mot. Pour justifier leur réponse, ces participants ont fourni le sens d'un mot courant « 喜欢 (*xǐhuan*, aimer, aimer bien) », constitué de deux mêmes caractères mais dans l'ordre inverse de « 欢喜 ». Les participants ayant indiqué un niveau 6 pour ce mot ont produit une phrase avec le mot « 欢喜 (*huānxǐ*, content) » en prenant le sens de « 喜欢 (*xǐhuan*, aimer, aimer bien) ». Certains ont même directement produit une phrase avec le mot « 喜欢 ». Cela démontre que, pour un apprenant ayant un niveau intermédiaire en chinois L2, la reconnaissance d'un mot chinois peut être facilement biaisée.

Le Tableau 7.15 plus bas présente les réponses du niveau de connaissance données par nos participants lors du test VKS1. Nous pouvons voir que les participants estimaient une meilleure connaissance sur les mots MC07 « 还礼 », M08 « 在乎 » et MC13 « 明丽 » : Parmi 33 participants, 17 participants ont noté niveau 3 (« J'ai déjà vu ce mot, mais je ne sais pas ce qu'il veut dire. ») pour MC07 « 还礼 », 6 participants ont noté niveau 3 pour MC08 « 在乎 » et 16 participants pour MC13 « 明丽 ». Ces trois mots ne font pas partie des mots chinois les plus courants. Selon les programmes pédagogiques des participants, il est peu probable que les participants aient déjà vu ces trois mots. Les rapports verbaux lors de l'inférence lexicale nous ont également confirmé la non connaissance de ces trois mots par nos participants. Nous pensons que cette surestimation du niveau de connaissance chez nos participants est liée à la connaissance des caractères constitutifs (« 明 » et « 丽 » de

MC13, « 还 » et « 礼 » de MC07 et « 在 » de MC08). Certains participants ont marqué niveau 4 (« J'ai déjà vu ce mot. Il veut peut-être dire ..., mais je n'en suis pas sûr(e). ») pour MC13 « 明丽 (*mínglì*, clair et beau) » en fournissant « lumineux » comme sens du mot. En fait, « lumineux » est l'un des sens du premier caractère « 明 » de MC13. Certains participants ont noté niveau 4 pour MC08 « 在乎 (*zàihu*, prendre quelque chose à cœur ; faire attention à) ») en donnant « se trouver » pour la signification de MC08. Ils ont en réalité transposé le sens du premier caractère « 在 » au mot qui leur est réellement inconnu.

		Niveau 1		Niveau 2		Niveau 3		Niveau 4		Total	%
		Nb	%	Nb	%	Nb	%	Nb	%		
MC01	L	13	39%	9	27%	11	33%	0	0%	33	100%
MC06	L	29	88%	3	9%	1	3%	0	0%	33	100%
MC08	L	13	41%	10	31%	8	25%	1	3%	32	100%
MC09	L	24	73%	5	15%	4	12%	0	0%	33	100%
MC10	L	19	58%	7	21%	5	15%	2	6%	33	100%
MC12	L	21	66%	6	19%	4	13%	1	3%	32	100%
MC14	L	26	90%	1	3%	2	7%	0	0%	29	100%
MC16	L	24	73%	7	21%	2	6%	0	0%	33	100%
MC17	L	26	79%	3	9%	3	9%	1	3%	33	100%
MC18	L	23	70%	5	15%	4	12%	1	3%	33	100%
		218	67%	56	17%	44	14%	6	2%	324	100%
		Niveau 1		Niveau 2		Niveau 3		Niveau 4		Total	%
		Nb	%	Nb	%	Nb	%	Nb	%		
MC02	NL	22	67%	4	12%	7	21%	0	0%	33	100%
MC03	NL	19	58%	8	24%	6	18%	0	0%	33	100%
MC04	NL	22	67%	6	18%	5	15%	0	0%	33	100%
MC05	NL	14	42%	11	33%	7	21%	1	3%	33	100%
MC07	NL	7	21%	9	27%	17	52%	0	0%	33	100%
MC11	NL	16	48%	6	18%	10	30%	1	3%	33	100%
MC13	NL	9	27%	5	15%	18	55%	1	3%	33	100%
MC15	NL	30	91%	1	3%	2	6%	0	0%	33	100%
MC19	NL	24	73%	5	15%	4	12%	0	0%	33	100%
MC20	NL	29	88%	2	6%	1	3%	1	3%	33	100%
		192	58%	57	17%	77	23%	4	1%	330	100%

Tableau 7.15 – Nombre et répartition des réponses de différents niveaux de connaissance au VKS1 pour les mots lexicalisés (L) et non lexicalisés (NL)

Les mots non lexicalisés utilisés dans notre expérience contiennent plus de caractères connus par les participants. Nous pensons que cela peut expliquer qu'au VKS1, le niveau de connaissance global pour les mots non lexicalisés est supérieur à celui pour les mots lexicalisés : les participants ont estimé niveau 3 pour 14% des mots lexicalisés, mais 23% des mots non lexicalisés. Au VKS2, les réponses du niveau 2 et du niveau 3 augmentent plus nettement pour les mots lexicalisés que pour les mots non lexicalisés. Cette différence peut être liée au fait que, grâce à l'inférence lexicale et les tâches post-inférence sur les mots cibles, les participants ont eu une estimation plus juste à la phase post-inférence sur leur connaissance pour les mots cibles, lexicalisés et non lexicalisés.

En résumé, il semble que nos participants surestiment au VKS1 leur connaissance sur les mots cibles non lexicalisés à la phase pré-inférence, ce qui peut biaiser les analyses statistiques sur l'effet de la lexicalisation sur l'évolution du niveau de connaissance des mots cibles. Pour cette raison, nous ne pouvons pas tirer de conclusion nette sur ce sujet.

7.3.2 Evolution de la connaissance sémantique

Toutes les réponses sur le sens des mots cibles, obtenues au VKS1, à l'inférence lexicale, au VKS2 et au VKS3, sont notées sur une échelle de 5 niveaux, de 0 à 1 point. Par rapport au score moyen au VKS1, les scores à l'inférence, au VKS2 et au VKS3 ont connu une augmentation significative. Cela signifie que, grâce à l'inférence lexicale et aux tâches qui l'ont suivie, nos participants ont eu un gain sémantique pour les mots cibles qui leur étaient inconnus au VK1.

Par rapport au score moyen à l'inférence, le score moyen varie peu à la phase post-inférence (VKS2 et VKS3) et ces différences d'interphase ne sont pas significatives. Statistiquement, cela signifie que les tâches post-inférence que nos participants ont accomplies n'ont pas porté leurs fruits en gain sémantique pour les mots cibles chez les participants. Cependant, selon les résultats des études précédentes sur la rétention subséquente à l'inférence lexicale (ex. Fraser, 1999; Hu & Nassaji, 2012; Kaivanpanah & Rahimi, 2017), sans interventions supplémentaires,

les apprenants se rappellent moins le sens des mots inférés aux post tests. Nous pensons que la stabilité du score moyen sur le sens des mots cibles aux VKS2 et VKS3 dans notre étude est liée aux tâches que nos participants ont accomplies à l'issue de l'inférence lexicale. Dans le détail, par rapport à l'inférence lexicale, le nombre des réponses partiellement correctes a connu des diminutions importantes (-66% au VKS2 et -63% au VKS3), tandis que le nombre des réponses entièrement correctes ont considérablement augmenté (81% au VKS2 et 105% au VKS3). Nous en déduisons que les tâches de vocabulaire à la suite de l'inférence lexicale ont favorisé l'apprentissage des mots cibles chez nos participants, parce que a) toutes les inférences ne sont pas réussies et certaines sont partiellement correctes. Grace (1998) indique qu'un apprenant, qui donne une mauvaise réponse à un mot inconnu dont il essaie de deviner le sens, risque de garder la mauvaise réponse comme sens du mot dans sa mémoire. Les tâches post-inférence ont donné à nos participants accès au sens correct des mots inférés. Cela explique la diminution des réponses partiellement correctes et l'augmentation des réponses entièrement correctes aux VKS2 et VKS3 ; b) les tâches post-inférence ont demandé aux participants de retravailler sur les mots inférés. Cela a donc augmenté la répétition d'exposition aux mots et conduit l'attention des participants sur les mots inférés, particulièrement à la forme qui est élémentaire dans la création d'un nouveau lemme pour un mot dans le lexique mental de l'apprenant (De Bot et al., 1997 ; Théophanous, 2019).

Pour cela, malgré la non significativité pour les différences entre le score moyen à l'inférence et ceux aux tests VKS postérieurs, nous pensons que les tâches post-inférence ont favorisé l'apprentissage des mots cibles au niveau sémantique à court et à long terme.

A notre surprise, le score moyen au VKS3 est supérieur à celui au VKS2. Le nombre des réponses partiellement correctes et celui des réponses entièrement correctes continuent à augmenter au VKS3, par rapport au VKS2. Cette amélioration est possiblement liée à l'exposition répétitive de nos participants aux 20 mots cibles sous forme de post test.¹¹⁸ Prenant les résultats de l'étude de Mason & Krashen (2004)

¹¹⁸ Pour rappel, le test VKS1 contient 60 mots dont 40 sont des mots de distraction. Les tests VKS2 et VKS3 ne contiennent que les mots sur lesquels nous voulions évaluer les connaissances lexicales de nos participants.

comme exemple, Schmitt (2008, p. 339) indique que, « *Dans l'ensemble, il semble que pratiquement tout ce qui conduit à plus d'expositions, d'attention, de manipulations ou de temps passé sur des éléments lexicaux ajoute à l'apprentissage de ces derniers. En fait, même le processus de test sur des items lexicaux semble faciliter une meilleure rétention, car les conceptions de recherche qui incluent plusieurs post tests conduisent généralement à de meilleurs résultats sur le post-test final que des conceptions similaires avec moins ou pas de post-tests intermédiaires.*¹¹⁹ ».

7.3.2.1 Evolution du score du sens des mots cibles

La connaissance sémantique chez nos participants n'a pas évolué de la même façon pour les 20 mots cibles.

Du VKS1 au VKS3, les scores des mots MC01, MC03, MC06, MC08, MC09 et MC15 n'ont pas eu d'évolution significative. Ces mots sont les moins réussis lors de l'inférence lexicale et ils ont des scores les plus faibles au VKS2 et au VKS3. Cela signifie que l'inférence du sens et les tâches de vocabulaire post-inférence n'ont pas eu d'effet sur l'apprentissage du sens pour ces six mots chez nos participants.

Aux VKS2 et VKS3, parmi les 20 mots, MC12 a eu les meilleurs scores sur le sens ; les quatre autres mots ayant les meilleurs scores sont MC07, MC16, MC17 et MC20. Ces cinq mots contiennent les caractères sémantiquement connus par les participants, et les sens des caractères constitutifs ont de grands apports sémantiques aux sens de ces mots. Les scores sur le sens de MC12, de MC16 et de MC17 aux VKS2 et VKS3 connaissent des augmentations significatives par rapport aux scores d'inférence lexicale de ces trois mots. Cela signifie que les participants ont nettement tiré profit des tâches post-inférence dans l'apprentissage du sens de ces trois mots.

Au VKS2 et au VKS3, MC03 a eu les scores sur le sens les plus faibles ; MC01, MC06, MC15 et MC19 sont les quatre autres mots ayant les scores les plus faibles.

¹¹⁹ « Overall, it seems that virtually anything that leads to more exposure, attention, manipulation, or time spent on lexical items adds to their learning. In fact, even the process of being tested on lexical items appears to facilitate better retention, as research designs that include multiple post-tests usually lead to better results on the final delayed post-test than similar designs with fewer or no intermediate post-tests. »

Ces cinq mots contiennent essentiellement des caractères inconnus pour nos participants. D'ailleurs, MC01, MC03 et MC06 font partie des mots les moins réussis lors de l'inférence lexicale.

Statistiquement, les scores sur le sens aux VKS2 et VKS3 de MC04, de MC11 et de MC18 diminuent significativement par rapport à leurs scores d'inférence lexicale. MC18 est d'ailleurs l'un des mots les plus réussis lors de l'inférence lexicale. Ces mots contiennent des caractères inconnus par nos participants. Nous pensons que les participants se sont bien appuyés sur le contexte pour inférer les sens de ces trois mots. Le fait que les mots sont présentés de façon isolée dans les post tests VKS n'a possiblement pas facilité le rappel du sens pour ces mots.

MC01	节俭	L	(<i>jiéjiǎn</i>) : économe, frugal
MC02	窗花	NL	(<i>chuānghuā</i>) : papier découpé pour la décoration d'une fenêtre
MC03	送灶	NL	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC04	腊月	NL	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC05	汤圆	NL	(<i>tāngyuán</i>) : boulette de riz glutineux farcie de viande (fourrée à la confiture)
MC06	雀跃	L	(<i>quèyuè</i>) : sauter de joie
MC07	还礼	NL	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC08	在乎	L	(<i>zàihu</i>) : prendre qch. à cœur ; faire attention à
MC09	乏味	L	(<i>fáwèi</i>) : fade, insipide
MC10	盼望	L	(<i>pànwàng</i>) : espérer, désirer
MC11	安适	NL	(<i>ānshì</i>) : tranquille et confortable
MC12	五颜六色	L	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC13	明丽	NL	(<i>mínglì</i>) : clair et beau
MC14	彩虹	L	(<i>cǎihóng</i>) : arc-en-ciel
MC15	踏青	NL	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps
MC16	风筝	L	(<i>fēngzhēng</i>) : cerf-volant
MC17	蝉	L	(<i>chán</i>) : cigale
MC18	闲聊	L	(<i>xiánliáo</i>) : bavarder
MC19	飘零	NL	(<i>piāolíng</i>) : tomber en voltigeant
MC20	白茫茫	NL	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

7.3.2.2 Différence d'évolution du score sur le sens des mots cibles entre le Groupe I et le Groupe II

Pour les deux groupes, par rapport à l'inférence, les scores moyens demeurent quasiment inchangés au VKS2 et ont légèrement augmenté au VKS3. Toutefois, la proportion des réponses entièrement correctes augmente pour les deux groupes. Cela signifie que les participants des deux groupes ont tiré profit des tâches post-inférence pour apprendre les bons sens des mots cibles. Les augmentations sont légèrement plus importantes pour le Groupe II que pour le Groupe I au VKS2 (8% vs 6%) et au VKS3 (10% vs 9%).

Par rapport à la phase d'inférence, les deux groupes ont un gain sémantique relativement important pour les mots MC07, MC12, MC16 et MC17 au VKS2 et au VKS3.

MC0	还礼	NL	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC1	五颜六色	L	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC1	风筝	L	(<i>fēngzhēng</i>) : cerf-volant
MC1	蝉	L	(<i>chán</i>) : cigale

La différence entre les deux groupes en termes d'évolution de la connaissance sémantique des mots cibles n'est pas significative. En d'autres termes, la consultation d'un dictionnaire bilingue n'a pas abouti à des résultats différents d'apprentissage du sens des mots cibles pour les participants du Groupe II, à court et à long terme. Cela n'est pas en accord avec notre hypothèse de recherche. Nous pensons que les raisons possibles sont les suivantes :

- 1) L'apprentissage du sens d'un mot inconnu exige normalement plus d'effort que l'apprentissage de sa forme (cf. Nation, 2001 ; Teng, 2016). Les participants du Groupe II ont passé environ 15 minutes de plus que ceux du Groupe I pour travailler sur les 20 mots cibles par le biais d'un dictionnaire bilingue. Cela est probablement non suffisant pour obtenir un gain sémantique important, particulièrement pour les mots dont le concept est manquant (par exemple MC03) et ceux constitués des caractères inconnus (par exemple MC06 et MC15).

MC03	送灶	NL	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC06	雀跃	L	(<i>quèyuè</i>) : sauter de joie
MC15	踏青	NL	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps

- 2) Contrairement aux autres études (Laufer & Rozovski-Roitblat, 2015; Zhu, 2004), la consultation d'un dictionnaire bilingue dans notre étude n'a pas été un outil nécessaire pour la compréhension des textes. Dans notre étude, nous avons demandé aux participants du Groupe II de trouver les sens des mots inférés pour accomplir ensuite la tâche lacunaire. Le « besoin » d'apprendre le sens des mots cibles chez nos participants était sans doute relativement faible. En d'autres termes, nos participants étaient moins engagés cognitivement quand ils consultaient le dictionnaire pour trouver le sens des mots (cf. Hulstijn & Laufer, 2001).
- 3) La fatigue pourrait également être un élément à prendre en considération pour les participants du Groupe II qui ont passé plus de temps à accomplir les tâches post-inférence.

La différence non significative en termes de l'évolution de la connaissance sémantique des mots cibles entre les deux groupes confirme, sous un autre angle, qu'au niveau de l'évolution du niveau de connaissance lexicale sous divers aspects (la forme, le sens et l'emploi), la différence principale entre les deux groupes se situe au niveau de la connaissance sur la forme (cf. la Section 7.3.1.1).

7.3.2.3 Différence d'évolution du score sur le sens entre les mots lexicalisés et les mots non lexicalisés

Le score d'inférence lexicale en moyenne des mots non lexicalisés est significativement supérieur à celui des mots lexicalisés ; au VKS2, le score en moyenne pour les mots non lexicalisés a une diminution de 21%, tandis que celui des mots lexicalisés a une augmentation de 18%. A ce stade, le score moyen pour les mots lexicalisés devance celui pour les mots non lexicalisés ; au VKS3, l'écart du score entre les mots lexicalisés et non lexicalisés continue à croître et le score moyen pour les mots lexicalisés est significativement supérieur à celui des mots non

lexicalisés. L'évolution du score sur le sens des mots lexicalisés (L) et non lexicalisés (NL) peut être schématisée comme suit :

$$\begin{array}{ccccc} L < NL & \rightarrow & L > NL & \rightarrow & L > NL \\ (\text{Inférence}) & & (\text{VKS2}) & & (\text{VKS3}) \end{array}$$

Autrement dit, nos participants ont eu un meilleur gain sémantique pour les mots lexicalisés que pour les mots non lexicalisés à court et à long terme. Nos résultats sont en accord avec ceux de l'étude de Heidari-Shahreza & Tavakoli (2012).

L'évolution de la connaissance sémantique des mots cibles chez nos participants, pour les mots lexicalisés d'une part et pour les mots non lexicalisés d'autre part, est significativement différente. Compte tenu des évolutions opposées pour les mots lexicalisés et les mots non lexicalisés, nous concluons que, par rapport à un mot chinois lexicalisé, un apprenant francophone a plus de difficultés à retenir le sens d'un mot chinois non lexicalisé à long terme.

Dans notre étude, lors de l'inférence, les participants ont notamment compris partiellement les mots non lexicalisés dans une large mesure grâce à leur connaissance sémantique des caractères constitutifs (donc les concepts connexes disponibles dans leur mémoire). Les tâches post-inférence ont donné aux participants accès aux sens corrects de tous les mots cibles, pourtant les scores du sens obtenus aux post tests présentent une diminution pour les mots non lexicalisés. Plus précisément, au VKS2 et au VKS3, par rapport à la phase d'inférence, le nombre des réponses erronées augmente d'environ 48% pour les mots non lexicalisés et nos participants ont eu plus de réponses partiellement correctes pour les mots non lexicalisés que pour les mots lexicalisés. En comparaison, les participants ont moins réussi l'inférence lexicale pour les mots lexicalisés, mais les scores postérieurs pour les mots lexicalisés montrent une évolution positive : le nombre des réponses erronées évoluant peu, les participants ont donné plus des réponses entièrement correctes pour les mots lexicalisés que pour les mots non lexicalisés au VKS2 (21% vs 11%) et au VKS3 (26% vs 10%).

Par rapport aux scores à l'inférence, les scores du sens obtenus aux post tests pour les mots MC12, MC16 et MC17 augmentent significativement. Ces trois mots

sont lexicalisés. Les scores du sens postérieurs pour les mots MC04, MC11 et MC18 diminuent significativement. Parmi ces trois mots, MC04 et MC11 sont non lexicalisés. Les tâches post-inférence semblent avoir eu un effet favorable particulièrement pour les mots lexicalisés.

MC04	腊月	NL	(<i>làyuè</i>) : le douzième mois du calendrier lunaire
MC11	安适	NL	(<i>ānshì</i>) : tranquille et confortable
MC12	五颜六色	L	(<i>wǔyánliùsè</i>) : de diverses couleurs, multicolore
MC16	风筝	L	(<i>fēngzhēng</i>) : cerf-volant
MC17	蝉	L	(<i>chán</i>) : cigale
MC18	闲聊	L	(<i>xiánliáo</i>) : bavarder

Parmi les dix mots non lexicalisés, MC07 « 还礼 » est le seul ayant vu son score augmenter en continuité consécutivement à l'inférence. Ce mot est sémantiquement transparent et constitué des caractères connus par nos participants. Lors de l'inférence, nos participants ont eu des difficultés à déduire le sens de ce mot notamment à cause de la nature polysémique du premier caractère « 还 », mais les tâches post-inférence ont fourni à nos participants l'accès au bon sens du mot. En d'autres termes, les concepts connexes de MC07 sont déjà disponibles dans la mémoire de nos participants pour qu'ils puissent rapidement « créer » un nouveau concept pour MC07.

En revanche, MC03 est le mot ayant donné le moins de réussite à l'inférence lexicale, au VKS2 et au VKS3. La compréhension de ce mot nécessite des connaissances des mythes et coutumes chinois, un thème peu abordé dans l'enseignement dans un contexte hétéroglotte auprès des apprenants du niveau intermédiaire.

MC03	送灶	NL	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC07	还礼	NL	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour

Nous pensons que l'évolution opposée de la connaissance sémantique, pour les mots lexicalisés d'une part et pour les mots non lexicalisés d'autre part, révèle l'influence de la disponibilité/absence d'un concept dans le lexique mental de l'apprenant sur la création d'un lien forme-concept pour un mot L2 nouveau.

N. Jiang (2004, p. 417) met en avant que l'acquisition d'un mot L2 peut comprendre trois stades :

- L'association (« Word association stage », où l'entrée d'un mot L2 nouvellement créée dans le lexique mental, qui contient souvent seulement les informations lexicales formelles, et est fortement liée à sa traduction en L1. Le mot L2 est activé à l'accompagnement de sa correspondance en L1) ;
- La médiation (« L1 lemma mediation stage », où les informations sémantiques et syntaxiques du lemme L1 sont transférées vers le lemme L2. Le lien direct entre le concept et le lemme L2 est consolidé tandis que la présence du lemme L1 s'affaiblit lors de l'activation du lemme L2) ;
- L'intégration (« Full integration stade »), où les informations spécifiques du mot L2 sont intégrées dans le lemme L2 et l'apprenant peut théoriquement utiliser le mot L2 sans passer par sa traduction en L1).

Ainsi, le lexique mental en L1 de l'apprenant est activement impliqué dans l'acquisition du vocabulaire L2, au moins aux premiers stades qui peuvent être schématisée comme suit :

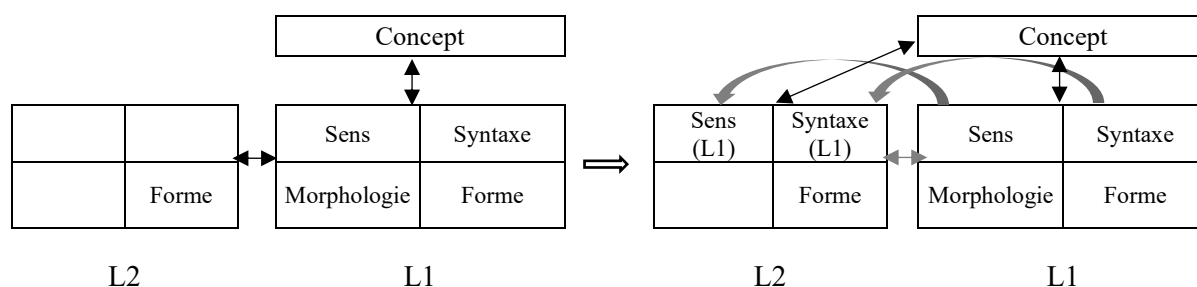


Figure 7.8 – Transfert d'informations lexicales d'un lemme L1 vers un lemme L2 (adapté de N. Jiang, 2004, p. 418)

Les informations sémantiques dans un lemme L2 sont donc transférées d'un lemme L1 correspondant, ou un lemme L2 est lié à un concept apparié avec un lemme L1, selon l'hypothèse du transfert sémantique (« Semantic transfer hypothesis », N. Jiang, 2004). Le transfert d'informations sémantiques du lemme L1 correspondant vers le lemme L2 s'accomplit au fur et à mesure des co-activations des deux lemmes.

Pour un mot L2 n'ayant pas de concept correspondant dans la mémoire de l'apprenant, ce dernier doit comprendre le sens du mot par le biais d'une combinaison

de plusieurs concepts connexes existants (Paribakht, 2005). Par exemple, la compréhension de MC04 腊月 (*làyuè*, le douzième mois du calendrier lunaire) doit mobiliser les concepts « mois », « dernier » (ou « douzième ») et « calendrier lunaire ». Selon l'hypothèse de N. Jiang (2004), il serait logique d'en déduire qu'afin de retenir le mot MC04, il aurait fallu que l'apprenant crée dans son lexique mental un nouveau lemme à partir de la forme 腊月 et qu'il alimente ensuite ce nouveau lemme L2 avec les informations conceptuelles des lemmes L1 concernés, par exemple ceux des mots « mois », « calendrier », etc. En d'autres termes, l'apprenant doit établir un lien « forme-conceptS » (Figure 7.9).

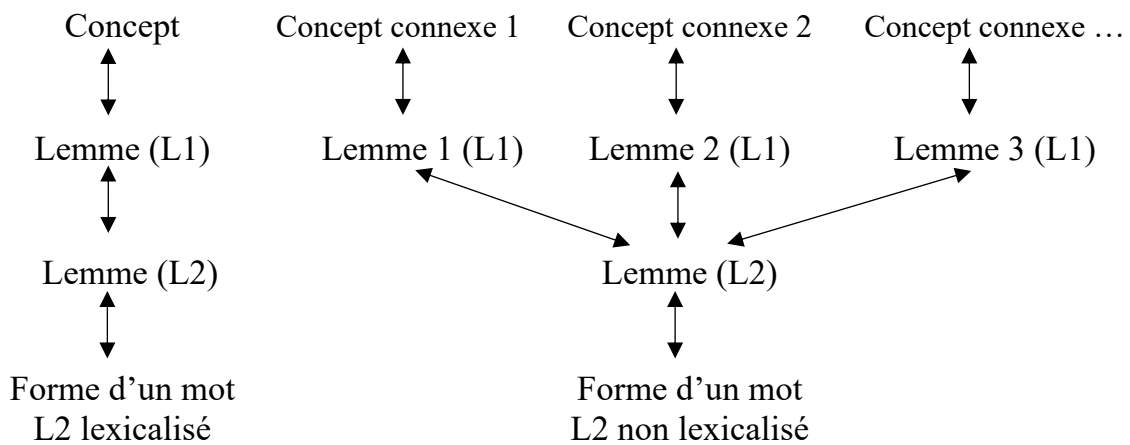


Figure 7.9 – Lien schématique « forme-concept » pour un mot L2 lexicalisé et « forme-conceptS » pour un mot L2 non lexicalisé

Plus de concepts connexes sont concernés pour un mot L2 non lexicalisé, plus le lien forme-conceptS à établir pour ce mot dans le lexique mental de l'apprenant est complexe.

Même si certains liens forme-conceptS ont été créés pour des mots cibles non lexicalisés lors des tâches post-inférence (cela est démontré par la croissance des réponses entièrement correctes pour les mots non lexicalisés au VKS2), ces liens semblent plus difficiles à maintenir par rapport aux liens forme-concept en relation un-à-un créés pour les mots lexicalisés : quand un apprenant tente de reconnaître un mot L2 non lexicalisé, il doit activer tous les concepts connexes concernés à partir de la forme du mot. Par rapport à l'activation d'un lien forme-concept pour un mot L2

lexicalisé, l'apprenant risquerait davantage de récupérer partiellement des informations conceptuelles pour un mot L2 non lexicalisé. Nos résultats montrent qu'au VKS3, par rapport au VKS2, les réponses entièrement correctes pour les mots non lexicalisés connaissent une décroissance, tandis que celles pour les mots lexicalisés, elles croissent. Par comparaison, les participants persanophones dans l'étude de Heidari-Shahreza & Tavakoli (2012) ont moins réussi au post test la traduction des mots cibles anglais non lexicalisés vers le persan que les mots cibles lexicalisés. Selon Heidari-Shahreza & Tavakoli, cela est peut-être lié au fait que les mots non lexicalisés concernent plus de composants sémantiques que les mots lexicalisés.

Ceci étant dit, si le concept d'un mot L2 non lexicalisé concerne relativement peu de concepts connexes existants et, en outre, le(s) caractère(s) constitutif(s) est/sont connu(s) par l'apprenant sur les plans de la forme et du sens (dans le cas d'un caractère polysémique, il doit alors s'agir du sens approprié), il sera plus facile pour l'apprenant de retenir ce mot, par rapport à un mot non lexicalisé constitué des caractères inconnus et/ou avec un concept plus complexe. Cela peut expliquer que, parmi les 10 mots cibles non lexicalisés, MC07, MC13 et MC20 ont eu les meilleurs scores de rétention,

MC07	还礼	NL	(<i>huánlǐ</i>) : faire un présent en retour, offrir un cadeau à son tour
MC13	明丽	NL	(<i>mínglì</i>) : clair et beau
MC20	白茫茫	NL	(<i>báimángmáng</i>) : d'une blancheur éclatante (immaculée) ; une immensité argentée

tandis que MC03, MC15 et MC19 ont eu les scores de rétention les plus faibles.

MC03	送灶	NL	(<i>sòngzào</i>) : saluer avec des offrandes le départ du Génie du foyer pour le Ciel, le 23 de la 12e lune
MC15	踏青	NL	(<i>tàqīng</i>) : faire une excursion (se promener) à la campagne au printemps
MC19	飘零	NL	(<i>piāolíng</i>) : tomber en voltigeant

Le lien forme-concept S pour un mot non lexicalisé exigerait donc davantage d'efforts à la création et plus de travaux de renforcement de la rétention à la phase postérieure, particulièrement pour ceux avec un concept complexe et/ou constitués de caractère(s) inconnu(s).

7.4 Conclusions

Dans ce chapitre, nous avons présenté l'analyse des résultats d'apprentissage des mots cibles obtenus respectivement deux semaines (VKS2) et quatre semaines (VKS3) après la séance de l'inférence lexicale et des tâches supplémentaires sur les mots cibles.

Nous avons observé l'évolution des connaissances lexicales pour les mots cibles chez nos participants (la forme, le sens et l'emploi) et ensuite l'évolution sémantique des mots cibles.

Selon nos résultats, par rapport au VKS1, l'évolution est significativement positive au VKS2 et au VKS3 au niveau des connaissances lexicales, également positive au niveau de la connaissance sémantique pour les mots cibles. Cela signifie que, pour un apprenant francophone, la combinaison de l'inférence lexicale et des tâches post-inférence concernant un mot chinois inconnu a un effet positif sur l'apprentissage du mot, à court et à long terme.

Entre les participants ayant respectivement accompli une et deux tâches post-inférence, l'évolution est significativement différente sur le plan de la connaissance formelle. Aux post tests, les participants ayant accompli deux tâches ont eu une augmentation plus importante de réponses entièrement correctes sur le sens des mots inconnus que ceux ayant accompli une seule tâche. Toutefois, la différence entre les deux groupes n'est pas significative sur le plan de la connaissance sémantique. Nous concluons donc, dans notre étude, que **la consultation d'un dictionnaire à l'issue de l'inférence lexicale pour accomplir un exercice lacunaire a porté ses fruits notamment pour l'apprentissage incident d'un mot chinois inconnu au niveau de la forme**. Il semble que la réalisation de cette tâche ne suffit pas pour obtenir un gain important sur le plan sémantique pour un mot inféré parce que les informations sémantiques exigent en général un traitement plus profond.

Entre les mots cibles lexicalisés et non lexicalisés, la différence d'évolution concerne la connaissance du sens. Ayant mieux réussi l'inférence pour les mots non lexicalisés grâce à leur connaissance des caractères constitutifs de ces derniers, les

participants ont par contre mieux retenu le sens des mots lexicalisés à la phase post-inférence. En d'autres termes, les participants ont plus de difficultés à retenir les sens des mots non lexicalisés par rapport aux mots lexicalisés. Globalement, à la phase post-inférence, les participants ont donné plus de réponses entièrement correctes pour les mots lexicalisés que pour les mots non lexicalisés

Nous concluons que, **pour un apprenant francophone, la non lexicalisation d'un mot chinois L2 rencontré dans la lecture a un effet négatif sur l'apprentissage incident de ce mot sur le plan sémantique, à court et à long terme.** Selon nous, après la création d'un lemme pour un mot L2 non lexicalisé dans le lexique mental de l'apprenant, le manque du concept correspondant à ce mot nécessite le transfert d'informations conceptuelles depuis plusieurs concepts connexes existants afin de combler ce nouveau lemme. Autrement dit, il faudrait établir **un lien forme-conceptS** pour ce mot non lexicalisé. Par rapport à un lien forme-concept pour un mot L2 lexicalisé, un lien forme-conceptS pour un mot L2 non lexicalisé semble plus complexe à mettre en place et, par la suite, à maintenir. Lors de la reconnaissance d'un mot non lexicalisé, l'apprenant risque de n'activer que partiellement les informations conceptuelles concernées. **De ce fait, par rapport à un mot chinois lexicalisé, l'apprentissage incident d'un mot chinois non lexicalisé en français nécessite davantage d'effort cognitif pour un apprenant francophone.** Des activités qui focalisent sur les mots L2 non lexicalisés et donnent des explications explicites sur ces derniers sont donc indispensables.

CHAPITRE VIII – CONCLUSIONS GENERALES

8.1 Réflexions pédagogiques

8.1.1 Apprentissage du lexique via la lecture en chinois L2

Dans l'objectif de comprendre a) l'effet de la lexicalisation dans la L1 de l'apprenant sur l'inférence et la rétention subséquente des mots L2 inconnus rencontrés en lecture et b) l'effet de la charge cognitive de la condition d'apprentissage sur l'acquisition incidente de ces mots, nous avons mené une étude auprès d'apprenants francophones en chinois L2. Nous avons examiné les sources de connaissance que les apprenants utilisaient pour inférer le sens des mots chinois inconnus, lexicalisés et non lexicalisés en français, présents dans des textes et le degré de réussite de l'inférence, ainsi que les gains lexicaux chez les apprenants francophones. Nous avons également comparé les gains lexicaux pour ces mots inconnus, à court et à long terme, à la suite d'une seule tâche post-inférence, avec ceux obtenus à la suite de deux tâches. Rappelons ici les résultats principaux obtenus de notre étude :

- Un apprenant ayant un niveau intermédiaire en chinois L2 parvient à comprendre le sens d'un mot inconnu grâce au contexte ;
- Le contexte, notamment la phrase comportant le mot inconnu, est la source principale que l'apprenant utilise pour chercher des indices permettant de comprendre ce mot ;
- La connaissance des caractères constitutifs d'un mot chinois inconnu est une source très importante dans la compréhension de ce mot et la rétention subséquente du mot ;
- Les interventions pédagogiques, telles que les exercices de vocabulaire, peuvent favoriser l'apprentissage des mots inférés ;
- La lexicalisation d'un mot chinois inconnu n'est pas l'élément déterminant pour la compréhension du sens de ce mot, mais un mot chinois non lexicalisé représente davantage de difficultés afin d'être retenu sémantiquement par l'apprenant à court et à long terme.

A partir de ces résultats, nous concluons donc qu'il est possible que les apprenants ayant un niveau intermédiaire en chinois L2 puissent apprendre de

nouveaux mots chinois via la lecture avec des interventions pédagogiques. Nous proposons donc qu'en dehors des programmes pédagogiques concernant l'enseignement du lexique, les enseignants de chinois mettent en pratique l'activité de lecture afin d'ouvrir une autre piste d'apprentissage du lexique pour les apprenants. L'activité de lecture peut être conduite sous forme d'exercice de compréhension écrite en classe ou de devoir à accomplir en période extrascolaire.

Nation (2013, p. 380) met en avant trois éléments dans le cadre d'apprentissage de mots L2 via la lecture : sur le plan de la lecture, la densité des mots non familiers pour les apprenants ne doit pas être trop importante afin que les indices contextuels soient compréhensibles pour eux ; les apprenants, pour leur part, doivent être suffisamment compétents en L2 et en lecture afin de parvenir à traiter toutes les informations lexicales concernant les mots inconnus. Nation insiste que la déduction du sens d'un mot inconnu grâce au contexte est une activité complexe et les apprenants doivent s'appuyer sur un ensemble de compétences et de connaissances variées. Pour cela, l'inférence lexicale ne semble pas être une stratégie adéquate pour les apprenants ayant un faible niveau en L2 et en lecture.

En conséquent, nous pensons que 1) l'apprentissage du lexique chinois L2 via la lecture est conseillé aux apprenants ayant un niveau intermédiaire et au-delà en chinois et 2) il est nécessaire que l'enseignant contrôle les matériels de lecture dans les aspects suivants :

- La lecture doit être globalement compréhensible pour les apprenants. La densité des mots inconnus ne doit pas être supérieure à 5% de la totalité des mots présents dans un texte, selon Nation (2013).
- L'enseignant peut, dans la mesure du possible, remplacer certains mots non familiers dans le texte par des mots sur lesquels les apprenants ont déjà travaillé, avec l'objectif de créer davantage d'expositions qui peuvent favoriser la consolidation de ces mots dans le lexique mental des apprenants.
- En accord avec beaucoup d'autres études, nos résultats montrent la grande importance du contexte dans l'inférence lexicale. Il est donc indispensable que l'enseignant contrôle au niveau de la compréhensibilité du contexte,

notamment la phrase comportant le mot inconnu en question, afin que les apprenants aient suffisamment d'informations utiles pour déduire le sens de ce mot.

En parallèle, l'enseignant doit également aider les apprenants à « se tenir prêts » à apprendre de mots nouveaux via la lecture. Nos résultats montrent que les apprenants peuvent ignorer qu'ils ne connaissent pas tel ou tel mot, et qu'ils ne portent pas leur attention de façon homogène à tous les mots inconnus présents dans un texte pendant la compréhension écrite. Il est donc important pour l'enseignant d'aider les apprenants à orienter leur attention sur les mots inconnus et de les encourager à utiliser l'inférence lexicale en cas de lacune lexicale. En d'autres termes, l'enseignant doit former les apprenants afin qu'ils soient plus autonomes dans l'apprentissage du lexique via la lecture.

Nos résultats révèlent également l'importance des exercices de vocabulaire consécutifs à la lecture dans l'appropriation des gains lexicaux sur les mots inférés : les exercices de vocabulaire donnent l'accès aux apprenants au sens correct des mots inférés, ce qui réduit donc le risque que les participants retiennent de fausses informations lexicales concernant un mot donné en cas d'une inférence erronée ; en outre, les exercices obligent les apprenants à porter à nouveau leur attention sur les mots inférés et leur permettent d'y travailler dans différents contextes, deux facteurs fondamentaux intervenant dans le résultat d'apprentissage incident du vocabulaire (Schmidt, 1990 ; Laufer & Hulstijn, 2001).

En discutant les stratégies d'apprentissage du vocabulaire, Clark & Nation (1980, cité par Nation, 2013) mettent en avant la procédure opérationnelle pour apprendre un mot nouveau grâce au contexte, comme suit :

- Etape 1 : Délimiter la partie du texte comportant le mot inconnu ;
- Etape 2 : Chercher des indices dans le contexte immédiat du mot ;
- Etape 3 : Chercher des indices dans les propositions/phrases voisines ;
- Etape 4 : Déduire le sens du mot ;
- Etape 5 : Vérifier le sens déduit.

A partir de l'idée de Clarke & Nation (1980), nous suggérons que, pendant la mise en pratique de l'apprentissage du lexique via la lecture en chinois L2, l'enseignant

doit :

- guider les apprenants dans l'identification des mots inconnus présents dans un texte. Selon Nation (2013), la première étape dans la procédure de Clarke & Nation (1980) consiste à aider les apprenants à cibler le mot inconnu sur lequel ils travailleront. Spécifiquement en chinois, la connaissance d'un/des caractère(s) constitutif(s) d'un mot peut être une source importante dans l'inférence du sens de ce mot. Toutefois, cela peut également être un élément trompeur dans l'identification de mots inconnus. L'enseignant doit rappeler régulièrement aux apprenants le rôle flexible des caractères dans le vocabulaire chinois et inciter les apprenants à s'appuyer sur diverses sources pour identifier plus efficacement les mots inconnus dans la lecture, notamment ceux comportant des caractères polysémiques.
- générer des besoins pour les apprenants de connaître le sens des mots inconnus repérés, par exemple sous forme de demande d'exercice.
- apprendre aux apprenants à inférer le sens d'un mot inconnu. Tous les apprenants du chinois L2 n'utilisent pas aisément la stratégie d'inférence lexicale (Sun, 2004). L'enseignant doit donc y former les apprenants en leur expliquant les pistes d'indices à leur disposition, à l'intérieur et à l'extérieur d'un mot à inférer.
- donner aux apprenants l'accès au bon sens des mots inférés par le biais des exercices de vocabulaire. Une inférence lexicale entreprise ne donne pas toujours de réponse correcte. Afin de favoriser l'apprentissage subséquent des mots inférés, il est indispensable de donner le bon sens des mots inférés aux apprenants à travers des exercices de vocabulaire post-inférence tels que la consultation d'un dictionnaire ou l'exercice lacunaire.

8.1.2 Accent sur les caractères dans l'enseignement du lexique chinois L2

Nous discutons dans cette partie le rôle des caractères en nous penchant sur la relation entre un caractère et un mot, ce premier étant un morphème de ce dernier (cf. la Section 3.2.3.1).

Nos participants en chinois L2 se sont beaucoup appuyés sur les sources de

connaissance au sein du mot afin de déduire le sens des mots chinois inconnus. Cela est lié au fait que les caractères chinois sont morphosémantiques, ainsi qu'à l'apport sémantique important qu'un caractère constitutif d'un mot peut avoir au sens global de ce mot. Cela démontre aussi que les apprenants en chinois L2 ayant un niveau intermédiaire sont conscients de ces particularités.

Il est intéressant de remarquer que, lors de l'inférence lexicale, les participants essayaient de reconnaître un caractère par le biais d'un autre mot comportant ce caractère. Comme dans l'exemple suivant, le participant a reconnu le deuxième caractère de MC13 « 明丽 *mínglì* », « 丽 *lì* », qui peut aussi constituer le mot « 美丽 *měilì*, joli ».

Ex. (MC13 « 明丽 ») : *Et juste après, « li » comme dans « mei li » pour dire quelque chose de très beau.*

Parfois des participants confondent les deux caractères constitutifs d'un mot. Dans l'exemple suivant, le participant a reconnu le deuxième caractère de MC13 « 明丽 *mínglì* » par le biais du mot « 美丽 *měilì* », mais il a confondu « 美 *měi* » et « 丽 *lì* ».

Ex. (MC13 « 明丽 ») : *En plus il y a ... le « mei » de « mei li », si je (ne) me trompe pas. Donc « mei li » c'est « joli ».*

Il semble que, pour nos participants, un sens donné est représenté plus par un mot chinois¹²⁰ que par un caractère. En fait, « 丽 » seul peut signifier « joli ». Nous pensons que cela résulte du fait qu'en général, dans l'enseignement du chinois L2, c'est le « mot » l'unité d'enseignement du lexique, non le « caractère » (B. S. Lü, 2011). Plus précisément, dans les manuels, les textes sont accompagnés d'une liste de mots chinois, accompagnés de leur signification en français, qui donnent aux élèves une voie rapide vers la compréhension du texte. Les élèves travaillent sur les caractères individuels essentiellement dans le cours d'écriture pour se familiariser avec la forme des caractères. Leur connaissance sémantique des caractères se fait en général à travers des mots. Prenons comme exemple *Le chinois contemporain* (当代中文), le manuel utilisé en LEA à l'UT2J. Dans une phrase chinoise transcrite en

¹²⁰ Le terme « mot » que nous citons ici renvoie à un mot chinois constitué de plusieurs caractères (morphèmes).

Pinyin, les syllabes des caractères constitutifs d'un mot sont regroupées pour faciliter la compréhension des apprenants français qui ont l'habitude, dans la langue française, d'avoir un marqueur qui sépare les mots. Par exemple, « 我去图书馆。(Je vais à la bibliothèque.) » serait transcrite phonétiquement « *Wǒ qù túshūguǎn* ». La transcription phonétique du mot 图书馆 est marquée « *túshūguǎn* » au lieu de « *tú shū guǎn* ». Pour autant, le lien entre un caractère et un mot comportant ce caractère est indissociable pour les apprenants, notamment chez les débutants. En d'autres termes, par rapport au « mot », le rôle du « caractère », en tant que morphème d'un mot, est relativement peu souligné dans l'enseignement du lexique chinois (Y. H. Xu, 2014b).

En effet, dans la pédagogie du chinois L2, il existe depuis quelques décennies deux théories : la « primauté au caractère (字本位 *zì běnwèi*) » et la « primauté au mot (词本位 *cí běnwèi*) ». Pour résumer, la « primauté au caractère » insiste sur le fait que le caractère est le porteur sémantique élémentaire dans la langue chinoise, tandis que la « primauté au mot » considère plutôt que c'est le mot qui joue ce rôle. Selon T. Q. Xu (2005), grand partisan de la théorie de la « primauté au caractère », le terme « primauté au caractère » est apparu pour la première fois dans l'article « La flexibilité dans le lexique chinois (中国词语的弹性作用) » de S. Y. Guo (1938), où il comparait le fait de s'appuyer sur les caractères à l'écrit et sur les mots à l'oral. Aujourd'hui ce terme attire à nouveau l'attention des linguistes de la langue chinoise pour dire qu'il serait plus efficace d'étudier les structures linguistiques de cette langue en prenant le « caractère » comme unité structurale de base (T. Q. Xu, 2005).

La théorie de la « primauté au caractère » intéresse également les pédagogues et les chercheurs en chinois L2. A partir des idées de T. Q. Xu, B. S. Lü (2011), chercheur et pédagogue influent dans l'enseignement du chinois L2, met en avant sur son blog¹²¹ que le « caractère » est l'unité linguistique de base en chinois et qu'il est logique de prendre le « caractère » comme unité d'enseignement en chinois L2, à l'oral et à l'écrit. Il ajoute que la caractéristique fondamentale de la langue chinoise

¹²¹ http://blog.sina.com.cn/s/blog_5bbada9d0100x15r.html

est « l'assemblage (组合生成) » : les traits et les blocs composants forment ensemble un caractère et ce dernier, combiné avec d'autres caractères, permet de constituer des mots¹²², avec lesquels on peut produire des phrases. En tant que l'unité linguistique de base, un caractère est l'union de forme, son et sens, apte à constituer des mots l'associant à d'autres caractères. Cela permet une grande liberté pour générer des mots et produire des phrases à partir d'une quantité limitée de caractères. Bellassen (2018), pionnier de la théorie « primauté au caractère » dans l'enseignement du chinois L2 en France, est d'avis que les caractères ont un grand poids dans l'apprentissage de mots et qu'il est économique d'avoir deux appuis en parallèle respectivement sur les caractères et sur les mots dans l'enseignement du lexique chinois. Déjà en 1985, en se basant sur la fréquence d'utilisation et sur la capacité de constitution de mots (构词能力, la flexibilité et la productivité de constitution de mots avec d'autres caractères), Bellassen a proposé la liste SMIC (Seuil Minimum Indispensable de Caractères)¹²³ contenant 400 caractères à apprendre pour un débutant en chinois L2.

Les résultats de notre étude ont confirmé que la connaissance préalable d'un caractère constitutif d'un mot L2 nouveau facilite la compréhension de ce mot et elle favorise par ailleurs la rétention du mot, que ce dernier soit lexicalisé ou non dans la L1 de l'apprenant. Selon Zheng (2006, cité par Y. H. Xu, 2014b), 80% des mots inclus dans le *Dictionnaire chinois moderne* (« 现代汉语词典 », 2002) sont plus ou moins sémantiquement transparents. Autrement dit, le sens de ces mots est en rapport avec le sens de leurs caractères constitutifs. En outre, il ne faut pas oublier que, pour un apprenant dont la L1 est alphabétique, la mémorisation de la forme d'un caractère n'est pas facile. Il est donc cognitivement moins coûteux pour les apprenants d'apprendre un mot chinois nouveau comportant un/des caractère(s) connu(s), par rapport à un mot constitué de caractères inconnus. A partir de cela, nous suggérons qu'il soit nécessaire de mettre plus en valeur les caractères dans l'enseignement du lexique chinois, notamment auprès des apprenants maîtrisant déjà une certaine

¹²² Cf. la Section 3.2.3.2 « Mot composé ».

¹²³ <https://www.afpc.asso.fr/Caracteres>, consulté le 5 septembre 2021.

quantité de caractères. L'enseignant doit aider les apprenants à être sensibilisés au rôle flexible du caractère dans le lexique du chinois, en tant que mot seul ou morphème dans un mot composé. Par ailleurs, il est également important pour l'enseignant d'inciter les apprenants à établir dans leur lexique mental, à partir d'un caractère commun, des réseaux de mots qui se lient par le biais des mots communs (Figure 8.1) (cf. Y. H. Xu, 2014b).

Sur le plan de l'apprentissage du lexique via la lecture, opérationnellement, l'enseignant peut intégrer dans les matériels de lecture des mots nouveaux constitués de caractères connus par les apprenants, entourés de contextes facilitateurs, puis demander aux apprenants d'inférer le sens de ces mots.

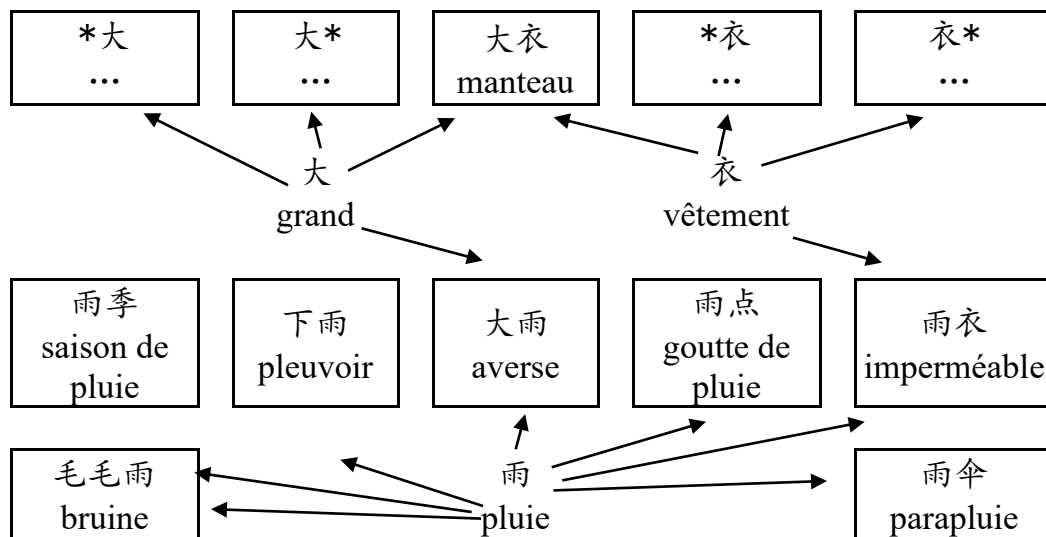


Figure 8.1 – Schéma des réseaux de mots établis à partir des caractères (basé sur les modèles de Y. H. Xu, 2014, p. 36)

8.1.3 Apprentissage plus soutenu des mots L2 non lexicalisés

Nation (2013, p. 374) indique que, pour un apprenant L2, son système conceptuel continue à se développer au fur et à mesure de l'apprentissage de mots L2, notamment des mots L2 dont le concept est non familier à l'apprenant (par exemple les termes spécifiques). Dans le cadre de l'apprentissage du lexique en L2 via la lecture, tous les mots L2 correspondant à un concept non familier pour l'apprenant

ne sont pas « inférables ». Ces mots « non inférables » méritent donc davantage d'attention des enseignants.

Un mot L2 non lexicalisé dans la L1 de l'apprenant signifie l'absence d'un concept correspondant dans la mémoire de l'apprenant. Selon nos résultats, la lexicalisation n'est pas l'élément concluant dans la compréhension d'un mot chinois inconnu. Toutefois, l'absence d'un concept correspondant signifie qu'afin de retenir un mot L2 non lexicalisé, le lemme créé pour ce mot doit être lié à plusieurs concepts connexes disponibles. Ces concepts connexes alimentent ce nouveau lemme au fur et à mesure de l'utilisation de ce mot (Jiang, 2004). Il est raisonnable d'en déduire que, par rapport à un mot L2 lexicalisé dont le lemme est alimenté par un concept, il est plus complexe d'établir, au premier abord, les liens entre ce nouveau lemme avec tous les concepts connexes et que, par la suite, davantage d'efforts cognitifs sont nécessaires pour combler ce nouveau lemme et puis consolider ce dernier dans le lexique mental de l'apprenant.

Pour cela, en accord aussi avec les conclusions de Paribakht (2005) et de Heidari-Shahreza & Tavakoli (2012), nous pensons que les mots non lexicalisés méritent davantage d'attention dans l'enseignement du lexique en L2. Afin d'aider les apprenants à mieux apprendre les mots L2 non lexicalisés, nous suggérons d'intégrer dans les programmes pédagogiques des activités qui incitent les apprenants à porter plus d'attention aux mots de cette nature et à s'investir davantage cognitivement pour ces mots ; notamment pour les mots dont la compréhension exige des connaissances socio-culturelles, des explications explicites et détaillées sont indispensables.

8.2 Originalité et limites de cette étude

A notre connaissance, notre étude a investigué pour la première fois l'effet de la lexicalisation dans l'inférence lexicale et l'apprentissage subséquent, à court et à long terme, du vocabulaire en chinois L2. Afin de mieux comprendre l'évolution des connaissances lexicales chez les apprenants pour les mots chinois inconnus appris de façon incidente, nous avons examiné l'évolution des connaissances lexicales sur le

plan multi-dimensionnel (la forme, le sens et l'emploi) et, par la suite, quantifié les réponses des apprenants sur le sens des mots inconnus afin d'examiner l'évolution de la connaissance sémantique.

Il est également inédit dans notre étude d'avoir observé l'attention portée par les apprenants aux mots inconnus, lexicalisés et non lexicalisés, lors d'une lecture.

Cependant, notre étude présente comme toute étude des limites que nous présenterons dans les sections ci-après.

8.2.1 Sélection de mots cibles

Les résultats de notre étude montrent que la connaissance du sens d'un ou des caractères constitutifs d'un mot chinois inconnu a un effet important sur la déduction du sens du mot, notamment quand il s'agit d'un mot sémantiquement transparent. Pour cela, un apprenant en chinois L2 s'appuie souvent sur cette piste pour inférer le sens global d'un mot chinois inconnu.

Or, pour notre étude, l'un des objectifs principaux est d'observer l'effet de la nature de lexicalisation d'un mot chinois L2 inconnu dans l'inférence et l'apprentissage incident du mot. Nous avons sélectionné les 20 mots cibles selon la définition de Paribakht (2005, 2010). Plus précisément, nous avons opté pour des mots chinois qui n'ont pas d'équivalent lexical en français, la langue maternelle de nos participants. Nous n'avons pas contrôlé la connaissance préalable des caractères constitutifs des mots cibles chez nos participants. Cela peut avoir influencé les résultats obtenus de notre étude qui montrent qu'un apprenant francophone s'appuie volontiers sur le sens des caractères constitutifs pendant l'inférence des mots chinois inconnus non lexicalisés en français.

8.2.2 Groupe de contrôle

Dans notre étude, les scores moyens du sens des mots chinois inconnus à la phase post-inférence ne sont pas significativement différents du score d'inférence en moyenne. Autrement dit, statistiquement, les tâches que nos participants ont accomplies à l'issue de l'inférence lexicale ne sont pas bénéfiques sur le plan du gain

sémantique pour les mots inférés. Toutefois, à la phase post-inférence, le nombre des réponses partiellement correctes a connu une diminution importante et le nombre des réponses entièrement correctes a considérablement augmenté. Nous pensons donc que les tâches post-inférence concernant les mots inférés ont porté leurs fruits en termes de gain sémantique chez les participants pour ces mots.

Le fait que nous n'ayons pas eu un troisième groupe de contrôle n'ayant effectué que l'inférence lexicale ne nous permet pas de comprendre précisément dans quelle mesure les tâches post-inférence ont pu influencer le gain lexical.

8.2.3 Observation de l'attention portée aux mots inconnus

Nous avons utilisé la technique de l'oculométrie pour observer l'attention que les apprenants francophones ont portée aux mots chinois inconnus rencontrés en lecture, une expérience exploratoire.

Nous avons expliqué plus haut que, lors du choix des mots cibles pour notre étude, nous n'avons pas pris en compte les éléments lexicaux autres que la nature de lexicalisation des mots cibles. Cependant, les éléments ci-dessous ont pu intervenir sur les résultats :

- Longueur des mots cibles. Nous avons expliqué dans le Chapitre III qu'en chinois, dans une phrase, tous les caractères se mettent en juxtaposition sans espace supplémentaire balisant les frontières entre les mots. Un apprenant doit donc identifier lui-même les unités lexicales en lisant une phrase. La majorité des mots dans le chinois contemporain sont constitués de deux caractères. Un mot constitué de plus ou moins de deux caractères, comme MC12 « 五颜六色 (*wǔyánliùsè*, de diverses couleurs, multicolore) », peut avoir perturbé l'apprenant lors de l'identification de ce mot pendant la lecture.
- Connaissance préalable pour les mots non cibles présents dans les textes. Lors de la rédaction des textes, nous avons utilisé, dans la mesure du possible, des mots déjà enseignés et/ou pratiqués en cours par les participants en Licence à l'UT2J. Nous avons ensuite exclu les données oculométriques de deux participants de parcours universitaires différents afin d'homogénéiser au

maximum les données oculométriques à analyser. Autrement dit, les mots non cibles présents dans les textes étaient supposés connus par les participants. Cependant, lors de l'entretien d'inférence lexicale, nous avons remarqué que la connaissance des mots non cibles chez les participants n'était pas homogène. Des participants ne se souvenaient pas des mots censés être maîtrisés selon leur programme pédagogique. Cela peut avoir biaisé nos résultats oculométriques.

En complément, nous cherchions à savoir si l'attention portée à un mot inconnu pendant la lecture a un effet sur le résultat de l'inférence lexicale. Lors de l'expérience, nous avons demandé aux participants d'inférer l'un après l'autre les mots cibles soulignés dans les textes après la lecture pour la compréhension. Selon les résultats oculométriques, pendant la lecture, les participants n'ont pas porté leur attention à tous les mots cibles de façon homogène. Le fait que nous ayons indiqué aux participants les mots cibles présents dans les textes et leur ayons demandé de déduire les sens de ces mots a obligé les participants, en quelque sorte, à porter leur attention, pour la première fois ou bien à nouveau, aux mots cibles.

Afin de comprendre de façon plus pertinente, dans un cadre naturel de lecture, la relation entre l'attention portée à un mot inconnu présent dans une lecture et le degré de réussite de la déduction du sens du mot grâce au contexte, il aurait fallu administrer un test VKS supplémentaire immédiatement après la lecture pour la compréhension.

8.3 Futures pistes de recherche

La lexicalisation dans la L1 de l'apprenant demeure un domaine de recherche peu exploré dans le domaine de l'acquisition des langues secondes/étrangères jusqu'à présent. Nous avons mené cette étude en visant le chinois L2 et en nous focalisant sur l'effet de la nature de lexicalisation d'un mot L2 inconnu sur l'inférence lexicale et l'apprentissage incident du vocabulaire. Lors de la sélection des mots cibles, nous n'avons pas pris en compte l'effet d'autres éléments lexicaux tels que le nombre de caractères constitutifs et la connaissance préalable chez les participants visés pour les

caractères constitutifs. Il serait intéressant d'observer, en prolongement de cette étude, l'effet de la lexicalisation sur le processus et sur le résultat de l'inférence lexicale par le biais des mots L2 inconnus plus homogènes sur le plan lexical, par exemple ayant le même nombre de caractères constitutifs connus ou inconnus par les participants visés.

La fréquence d'exposition à un même mot inconnu est un facteur majeur dans l'apprentissage incident du vocabulaire (cf. Uchihara et al., 2019). A notre connaissance, il n'y a pas encore eu d'études réalisées en chinois L2 qui observent l'effet des deux variables, la lexicalisation et la fréquence d'exposition, sur l'apprentissage incident du vocabulaire. Cela peut être un futur sujet de recherche intéressant.

Enfin, l'une des questions de recherche posées dans notre étude concerne l'effet de la charge cognitive de la condition d'apprentissage sur les gains lexicaux d'un mot L2 inconnu rencontré en lecture. Selon nos résultats, la consultation d'un dictionnaire bilingue pour vérifier le sens d'un mot chinois inféré n'a pas eu d'effet significatif sur le gain sémantique pour ce mot. Il semble que cette tâche à l'issue de l'inférence lexicale ne suffit pas à avoir un gain sémantique important pour un mot chinois dans la condition d'apprentissage incident. Nos résultats ouvrent de nouvelles perspectives de recherche sur l'apprentissage du vocabulaire chinois L2, notamment sur le plan sémantique, dans le cadre de l'apprentissage incident.

Bibliographie

- Afflerbach, P., & Johnston, P. (1984). Research methodology on the use of verbal reports in reading research. *Journal of Reading Behavior*, 16(4), 307-322.
- Akamatsu, N. (1999). The effects of first language orthographic features on word recognition processing in English as a second language. *Reading and Writing: An Interdisciplinary Journal*, 11(4), 381-403.
- Alderson, J. C. (1984). Reading in a foreign language : A reading problem or a language problem ? In J. C. Alderson & A. H. Urquhart (Éds.), *Reading in a Foreign Language*. Longman.
- Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, 33(2), 206-219.
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson (Éd.), *Handbook of Reading Research* (p. 255-291). Longman.
- Angele, B., & Rayner, K. (2013). Processing the in the parafovea : Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 649-662.
- Atef-Vahid, S., Maftoon, P., & Zahedi, K. (2013). Topic familiarity, passage sight vocabulary, and L2 lexical inferencing : An investigation in the Iranian EFL context. *International Journal of Research Studies in Language Learning*, 2(4), 79-99.
- Baars, B. (1986). *The Cognitive Revolution in Psychology*. Guilford Press.
- Baars, B. (1996). *In the Theater of Consciousness : The Workspace of the Mind*. Oxford University Press.
- Badian, N. A. (2005). Does a visual-orthographic deficit contribute to reading disability? *Annals of Dyslexia*, 55(1), 28-52.
- Bai, X., Tian, J., Yan, G., & Wang, T. (2009). 词切分对美国大学生汉语阅读影响的眼动研究 The effect of word segmentation on Chinese text reading : Evidence from American students' eye movement. *南开语言学刊 Nankai Linguistics*, 8, 140-153.
- Barcroft, J. (2015). *Lexical Input Processing and Vocabulary Learning*. John Benjamins Publishing Company.
- Bellassen, J. (2011). Is Chinese europcompatible ? Is the Common European Framework common ? The Common European Framework of References for Languages facing distant language. *New Prospect for Foreign Language Teaching in Higher Education — Exploring The Possibilities of Application of CECR, Tokyo, World Language and Society Education Center (WoLSEC)*, 23-31.
- Bellassen, J. (2015). *Le chinois, langue émergente : État de l'enseignement du chinois en 2015-2016 [en ligne, <http://www.fcae.fr/pdf/etat-du-chinois-dec2015.pdf>, consulté le 11 juillet 2020]*.

- Bellassen, J. (2018). 一元论抑或二元论：汉语二语教学本体认识论的根本分歧与障碍
Monism vs dualism: The fundamental divergency and obstacle in the ontological epistemology of Chinese second language teaching. *华文教学与研究 TCSOL Studies*, 4, 1-24.
- Bengeleil, N. F., & Paribakht, T. S. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *The Canadian Modern Language Review*, 61(2), 225-249.
- Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, 7(1), 15-32.
- Bowles, M. A. (2008). Task type and reactivity of verbal reports in SLA: A first look at a L2 task other than reading. *Studies in Second Language Acquisition*, 30(3), 359-387. JSTOR.
- Bowles, M. A. (2010). *The Think-Aloud Controversy in Second Language Research*. Routledge.
- Bowles, M. A., & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition*, 27(3), 415-440.
- Bruton, A. (2007). Partial lexical knowledge in tests of incidental vocabulary learning from L2 reading. *The Canadian Modern Language Review / La Revue Canadienne des Langues Vivantes*, 64(1), 163-180.
- Bruton, A. (2009). The vocabulary knowledge scale: A critical analysis. *Language Assessment Quarterly*, 6(4), 288-297.
- Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique*, 114(4), 647-662.
- Carnine, D., Kameenui, E. J., & Coyle, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly*, 19(2), 188-204.
- Carrell, P. L. (1998). Can reading strategies be successfully taught. *Australian Review of Applied Linguistics*, 21(1), 1-20.
- Carroll, B., & Drum, P. A. (1983). Definitional gains for explicit and implicit context clue. In J. A. Niles & L. A. Harris (Éds.), *New Inquiries in Reading Research and Instruction* (p. 158-162). Rochester, NY: National Reading Conference.
- Carton, A. (1971). Inferencing: A process in using and learning language. In P. Pimsleur & T. Quinn (Éds.), *The Psychology of Second Language Learning* (p. 45-58). Cambridge University Press.
- Chall, J. (1987). Two vocabularies for reading: Recognition and meaning. In M. G. McKeown & M. E. Curtis (Éds.), *The Nature of Vocabulary Acquisition* (p. 7-17). Lawrence Erlbaum Associates, Inc.
- Chang, L.-Y., Xu, Y., Perfetti, C. A., Zhang, J., & Chen, H.-C. (2014). Supporting orthographic learning at the beginning stage of learning to read Chinese as a second language. *International Journal of Disability, Development and Education*, 61(3), 288-305.

- Chen, B., Ma, T., Liang, L., & Liu, H. (2017). Rapid L2 word learning through high constraint sentence context: An event-related potential study. *Frontiers in Psychology*, 8, 2285.
- Chen B., Wang L., & Peng D. (2003). 汉字识别中形音义激活时间进程的研究 (II) The time course of graphic, phonological and semantic information processing in Chinese character recognition (II). *心理学报 Acta Psychologica Sinica*, 35(05), 576-581.
- Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31(5), 693-713.
- Chen, Z. (2017). 《国际汉语教学通用课程大纲》多音字分析及国际汉语教学建议 « *International Curriculum for Chinese Language Education* » *Chinese polyphonic word analysis and suggestions of Chiense language eduction* [Master Thesis]. 郑州大学 Zhenzhou University.
- Chernyshova, E., & Traverso, V. (2017). Inférences et processus d'intercompréhension dans les interactions quotidiennes : Quelques questions méthodologiques. *Cahiers de Praxématique*, 68.
- Clark, D., & Nation, I. P. (1980). Guessing the meanings of words from context : Strategy and techniques. *System*, 8(3), 211-220.
- Clifton Jr., C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Éds.), *Eye Movements : A Window on Mind and Brain* (p. 341-371). Elsevier.
- Coady, J. (1979). A psycholinguistic model of the ESL reader. In M. Ronald & B. Bruce (Éds.), *Reading in a Second Language* (p. 5-12). Newbury House Publishers.
- Coady, J. (1996). L2 vocabulary acquisition : A synthesis of the research. In J. Coady & T. Huckin (Éds.), *Second Language Vocabulary Acquisition : A Rationale for Pedagogy* (p. 273-290). Cambridge University Press.
- Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language*, 22(1), 181-200.
- Cohen, A. D. (1987). Using verbal reports in research on language learning. In C. Færch & G. Kasper (Éds.), *Introspection in Second Language Research* (Vol. 30, p. 82-95). Multilingual Matters.
- Cohen, A. D. (2000). Exploring strategies in test-taking : Fine-tuning verbal reports from respondents. In G. Ekbatani & H. D. Pierson (Éds.), *Learner-Directed Assessment in ESL* (p. 127-150). Lawrence Erlbaum Associates.
- Cohen, A. D. (2014). *Strategies in Learning and Using a Second Language* (Second Edition). Routledge.
- Comer, W. J. (2012). Lexical inferencing in reading L2 Russian. *Reading in a Foreign Language*, 24(2), 209-230.
- Craik, F., & Lockhart, R. (1972). Levels of processing : A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684.
- Crutcher, R. J. (1994). Telling what we know : The use of verbal report methodologies in psychological research. *Psychological Science*, 5(5), 241-241.

- Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines : Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research, 1155*, 147-162.
- De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition : Evidence from ESL reading. *Studies in Second Language Acquisition, 19*(3), 309-329.
- De la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary : The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition, 24*(1), 81-112.
- Dictionnaire chinois moderne 现代汉语词典* (2^e éd.). (1984). 商务印书馆 The Commercial Press.
- Dictionnaire chinois moderne 现代汉语词典* (4^e éd.). (2002). 商务印书馆 The Commercial Press.
- Dictionnaire chinois-français 汉法词典*. (2001). 商务印书馆 The Commercial Press.
- Dolgunsöz, E. (2016). Using eye-tracking to measure lexical inferences and its effects on reading rate during EFL reading. *Journal of Language and Linguistic Studies, 12*(1), 63-78.
- Dombey, H. (2009). The simple view of reading. *ITE English: Readings for Discussion, 2*-13.
- Dong, R. Y. (2007). 韩国留学生汉语阅读猜词所用知识及其使用顺序的考察 *A study on the knowledge and order of the knowledge used in word-guessing in the course of Chinese reading by Korean students* [Master Thesis]. 北京语言大学 Beijing Language and Culture University.
- Drocourt, Z. (2015). L'écriture chinoise : Entre universaux et spécificités. *Les Langues Modernes, 15*-26.
- Ellis, N. (1994). Implicit AND explicit language learning : Their dynamic interface and complexity. In P. Rebuschat (Éd.), *Implicit and Explicit Learning of Languages*.
- Ellis, N. (1997). Vocabulary acquisition : Word structure, collocation, word-class, and meaning. In M. McCarthy & R. Schmidt (Éds.), *Vocabulary : Description, Acquisition and Pedagogy* (1 edition, p. 122-139). Cambridge University Press.
- Ellis, R. (2001). Introduction : Investigating form-focused instruction. *Language Learning, 51*, 1-46.
- Ellis, R. (2016). Focus on form : A critical review. *Language Teaching Research, 20*(3), 405-428.
- Ericsson, K. A. (2006). Protocol analysis and expert thought : Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Éds.), *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis : Verbal Reports as Data*. MIT Press.

- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis : Verbal reports as data* (Revised Edition). The MIT Press.
- Eysenck, M. (1982). Incidental learning and orienting tasks. In C. R. Puff (Éd.), *Handbook of Research Methods in Human Memory and Cognition* (p. 197-228). Academic Press.
- Fan, H. (2012). 汉语第二语言学习者与母语者在猜词过程中的差别研究 *Research on difference of lexical inferencing process between non-native and native Chinese speakers* [Master Thesis]. 北京大学 Peking University.
- Fan, L., & Zhang, J. (2014). 汉语词汇化对英语词汇推理加工过程影响的研究 A study on the influence of Chinese lexicalization on lexical inferencing in English. *外语与外语教学 Foreign Languages and Their Teaching*, 6, 50-55.
- Fang, Y. (2005). 留学生利用词汇语境学习策略之探析 A discussion of the oversea students' learning strategy related to the vocabulary context. *云南师范大学学报 Journal of Yunan University*, 3(2), 6-11.
- Fang, Y. X., & Jiang, X. (2012). 外国学生利用语境和构词法猜测汉语词义的个体差异研究 A study on individual differences of L2 Chinese learners in inferring novel word meanings by integrating information from context and word formation. *世界汉语教学 Chinese Teaching in the World*, 3, 367-378.
- Fatih, Y. (2015). Turkish as a foreign language learners awareness and use of morphology in guessing the meanings of unknown words from context: A case study. *Educational Research and Reviews*, 10(5), 675-683.
- Feng, L. P. (2009). 外国学生汉语词素的形音义加工与心理词典的建构模式研究 A study of the orthographic, phonological and semantic processing of Chinese morphemes by foreign students and the constructive pattern of their mental lexicon. *世界汉语教学 Chinese Teaching in the World*, 23(1), 101-110.
- Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research*, 3(4), 430-441.
- Fraser, C. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21(2), 225-241.
- Fu, Y. (2008). 以汉语为第二语言的留学生高级阶段阅读眼动研究 *Investigation to the advanced Chinese reading patterns of foreign students by eye tracker* [Master Thesis]. 首都师范大学 Capital Normal University.
- Gan, H. M. (2008). 语义透明度对中级汉语阅读中词汇学习的影响 The effects of semantic transparency on vocabulary learning in intermediate Chinese reading. *语言文字应用 Applied Linguistics*, 1, 82-90.
- Gan, H. M. (2010). 词性及其识别对汉语伴随性词汇学习的影响 The effects of word-class on incidental vocabulary learning. *汉语学习 Chinese Language Learning*, 3, 105-112.
- Gan, H. M. (2011). 上下文语境对汉语阅读中词汇学习的影响——一项基于自然阅读的调查报告 The effects of context on Chinese vocabulary learning during reading : An investigation based on natural reading. *语言教学与研究 Language Teaching and Linguistic Studies*, 3, 10-16.

- Gan, H. M. (2012). 中级汉语学习者猜词过程和阅读模式分析 On the word-meaning inferring process and reading model of intermediate Chinese learners. *华文教学与研究 TCSOL Studies*, 2, 30-38.
- Gan, H. M. (2014). 语境对汉语阅读过程中词汇学习的影响——一项基于眼动技术的实验研究 The effects of context of vocabulary learning during Chinese reading—Evidence from eye-movement studies. *汉语学习 Chinese Language Learning*, 2, 88-96.
- Garner, R. (1987). *Metacognition and Reading Comprehension*. Norwood, NJ: Ablex.
- Gass, S. (1988). Integrating research areas: A framework for second language studies1. *Applied Linguistics*, 9(2), 198-217.
- Gass, S. (1999). DISCUSSION: Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21(2), 319-333.
- Genesee, F., & Nicoladis, E. (2006). Bilingual first language acquisition. In E. Hoff & M. Shatz (Éds.), *Blackwell Handbook of Language Development* (p. 153-183). Oxford.
- Ghobadi, M., Shahriar, M., & Azizi, A. (2016). Short-term and long-term effects of incidental vocabulary acquisition and instructed vocabulary teaching. *International Journal of Applied Linguistics and English Literature*, 5, 212-218.
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35(3), 483-517.
- Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Éds.), *Noticing and Second Language Acquisition: Studies in Honor of Richard Schmidt* (p. 183-205). University of Hawai'i, National Foreign Language Resource Center.
- Goigoux, R., Cèbe, S., & Pironom, J. (2016). Les facteurs explicatifs des performances en lecture-compréhension à la fin du cours préparatoire. *Revue Francaise de Pedagogie*, 196(3), 67-84.
- Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6(4), 126-135.
- Gough, P. B. (1972). One second of reading. In J. F. Kavanagh & I. G. Mattingly (Éds.), *Language by Ear and by Eye: The Relationships between Speech and Reading* (p. 331-358). MIT Press.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6-10.
- Goulden, R., Nation, I. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341-363.
- Grace, C. A. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *The Modern Language Journal*, 82(4), 533-544.
- Graesser, A., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.

- Gray, W., & Holmes, E. (1938). *The development of meaning vocabularies in reading: An experimental study: Vol. (No. 6)* (p. xii, 140). University of Chicago Publications of Lab Schools.
- Guide hiérarchique des caractères et des mots chinois 汉语水平词汇与汉字等级大纲*. (1992). 国家对外汉语教学领导小组, 汉语水平考试部 Groupe directeur national pour l'enseignement du chinois comme langue étrangère, Section HSK. 北京语言大学出版社 Beijing Language and Culture University Press.
- Guo, J. (2015). Inference-making and linguistic skills in listening comprehension: An observation of French students learning Chinese. *Electronic Journal of Foreign Language Teaching*, 12, Suppl. 1, 318-331.
- Guo, S. Y. (1938). 中国语词的弹性作用 La flexibilité dans le lexique chinois. 燕京大学哈佛燕京学社 Harvard-Yenching Institute of Nanjing University.
- Haastруп, K. (1991). *Lexical Inferencing Procedures, or, Talking about Words: Receptive Procedures in Foreign Language Learning with Special Reference to English*. Gunter Narr Verlag.
- Haastруп, K. (2008). Lexical inferencing procedures in two languages. In D. Albrechtsen, K. Haastруп, & B. Henriksen (Éds.), *Vocabulary and Writing in a First and Second Language: Processes and Development* (p. 67-111). Palgrave Macmillan, London.
- Hamada, M., & Koda, K. (2010). The role of phonological decoding in second language word-meaning inference. *Applied Linguistics*, 31(4), 513-531.
- Hamel, C., & Joannette, Y. (2007). Communication et inférences: Hémisphère droit ou gauche? *Fréquences*, 19(3), 12-14.
- Harris, R. J., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General*, 107(1), 1-22.
- Hayes, J. R., & Flower, L. S. (1983). Uncovering cognitive processes in writing: An introduction to protocol analysis. In P. Mosenal & L. T. Walnsle S. (Éds.), *Research on Writing: Principles and Methods* (p. 206-219). Longman Inc.
- Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. N. Huckin, M. Haynes, & J. Coady (Éds.), *Second Language Reading and Vocabulary Learning* (p. 46-64). Ablex Pub. Corp.
- Heidari-Shahreza, M. A., & Tavakoli, M. (2012). The effect of repetition and L1 lexicalization on incidental vocabulary acquisition by Iranian EFL learners. *The Language Learning Journal*, 1-16.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Herman, P. A., Anderson, R. C., Pearson, P. D., & Nagy, W. E. (1987). Incidental acquisition of word meaning from expositions with varied text features. *Reading Research Quarterly*, 22(3), 263-284.
- Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching*, 41(2), 87-106.

- Hirsch, E. D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator*, 27(1), 10, 12-13,16-22,28-29,48.
- Holmes, V. M. (2009). Bottom-up processing and reading comprehension in experienced adult readers. *Journal of Research in Reading*, 32(3), 309-326.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- Hong, W., & Wang, L. J. (2016). Focuson Form 和 Focuson Forms 两种教学法对汉语二语词汇学习的影响 The role of Fom-Focused Instruction in CSL vocabulary learning : A comparative study of « Focus on Form » versus « Focus on Forms ». *世界汉语教学 Chinese Teaching in the World*, 30(2), 264-275.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange : Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.
- Hu, H. M., & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies, and their relationship with the retention of word meanings inferred from context. *The Canadian Modern Language Review*, 68(1), 54-77.
- Hu, H.-C. M., & Nation, I. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Huang, B. R., & Liao, X. D. (Éds.). (2005). *现代汉语 (上册) Modern Chines (Tom I)*. 高等教育出版社 Higher Education Press.
- Huang, Y. (2008). *印尼汉语学习者二语阅读中猜词策略调查与分析 Word-guessing strategies in L2 reading by Indonesian learners of Chinese : An investigation and analysis* [Master Thesis]. 厦门大学 Xiamen University.
- Hudson, T. (1991). A content comprehension approach to reading English for science and technology. *TESOL Quarterly*, 25(1), 77-104.
- Hudson, T. (2007). *Teaching Second Language Reading*. Oxford : Oxford University Press.
- Hulstijn, J. (1992). Retention of inferred and given word meanings : Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Béjoint (Éds.), *Vocabulary and Applied Linguistics* (p. 113-125). Macmillan.
- Hulstijn, J. (2001). Intentional and incidental second-language vocabulary learning : A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Éd.), *Cognition and Second Language Instruction* (p. 258-286). Cambridge University Press.
- Hulstijn, J. (2013). Incidental learning in second language acquisition. In C. Chapelle (Éd.), *The Encyclopedia of Applied Linguistics* (Vol. 5, p. 2632-2640). Chichester: Wiley-Blackwell.
- Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students : The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3), 327-339.
- Jegerski, J. (2014). Self-paced reading. In B. VanPatten & B. VanPatten (Éds.), *Research Methods in Second Language Psycholinguistics* (p. 20-49). Routledge.

- Jelić, A.-B. (2007). Lexical inferencing strategy use by Croatian foreign-language learners. *Upr 2007: Empirical Studies in English Applied Linguistics*, 245-254.
- Jenkins, J., & Dixon, R. (1983). Vocabulary learning. *Contemporary Educational Psychology*, 8(3), 237-260.
- Jenkins, J., Stein, M., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, 21(4), 767-787.
- Jiang, N. (2002). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, 88, 416-432.
- Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, 88, 416-432.
- Jiang, X. (2006). 汉字频率和构词数对非汉字圈学生汉字学习的影响 The effects of frequency and productivity on character learning by elementary L2 learners from alphabetic language backgrounds. *心理学报 Acta Psychologica Sinica*, 38(4), 489-496.
- Jiang, X., & Fang, Y. X. (2012). 语境和构词法线索对外国学生汉语词义猜测的作用 The effects of context and word morphology on interpreting unknown words by learners of Chinese as a second language: The effects of context and word morphology on interpreting unknown words by learners of Chinese as a second language. *心理学报 Acta Psychologica Sinica*, 44(1), 76-86.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329-354.
- Kaivanpanah, S., & Rahimi, N. (2017). The effect of contextual clues and topic familiarity on L2 lexical inferencing and retention. *Porta Linguarum*, 27.
- Kaye, D. B., Sternberg, R. J., & Fonseca, L. (1987). Verbal comprehension: The lexical decomposition strategy to define unfamiliar words. *Intelligence*, 11(1), 1-20.
- Ke, S. H., & Koda, K. (2017). Contributions of morphological awareness to adult L2 Chinese word meaning inferencing. *The Modern Language Journal*, 101(4), 742-755.
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition* (p. xvi, 461). Cambridge University Press.
- Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes*, 39(2-3), 125-128.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Koda, K. (2005). *Insights into Second Language Reading*. Cambridge University Press.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development: reading and language learning. *Language Learning*, 57, 1-44.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440-464.
- Krashen, S., & Terrell, T. (1983). *The Natural Approach: Language Acquisition in the Classroom*. Pergamon Press.

- Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model : A critical review and assessment. *Bilingualism (Cambridge, England)*, 13(3), 373-381.
- Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research : Inquiry, instruction, and social interaction. *Review of Educational Research*, 67(3), 271-299.
- LaBerge, D. (1995). *Attentional Processing : The Brain's Art of Mindfulness*. Harvard University Press.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Laurén & M. Nordman (Éds.), *Special Language : From Humans Thinking to Thinking Machines* (p. 316-323). Multilingual Matters.
- Laufer, B. (1990). Ease and difficulty in vocabulary learning : Some teaching implications. *Foreign Language Annals*, 23(2), 147-155.
- Laufer, B. (1992). Reading in a foreign language : How does L2 lexical knowledge interact with the reader's general academic ability. *Journal of Research in Reading*, 15(2), 95-103.
- Laufer, B. (1994). The lexical profile of second language writing : Does it change over time? *RELC Journal*.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language : Same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. (2003). Vocabulary acquisition in a second language : Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review*, 59(4), 567-587.
- Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review*, 63(1), 149-166.
- Laufer, B. (2017). The three "I"s of second language vocabulary learning—Input, Instruction, Involvement. In E. Hinkel (Éd.), *Handbook of research in second language teaching and learning : Volume III* (p. 343-354). Routledge.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language : The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Laufer, B., & Nation, I. P. (1995). Vocabulary size and use : Lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition : The effects of task type, word occurrence and their combination. *Language Teaching Research*, 15, 391-411.
- Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words : Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, 19, 687-711.
- Laufer, B., & Yano, Y. (2001). Understanding unfamiliar words in a text : Do L2 learners understand how much they don't understand? *Reading in a Foreign Language*, 13(2), 549-566.

- Le vocabulaire commun pour l'instruction obligatoire (ébauche)* 义务教育常用词表 (草案). (2019). 教育部语言文字信息管理司 Département de l'application linguistique et de la gestion de l'information du Ministère de l'Éducation. 苏新春 Su X. C. (Éd.). 商务印书馆 The Commercial Press.
- Lee, J.-W., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance : A test of the threshold hypothesis in an EFL context. *TESOL Quarterly*, 31(4), 713-739.
- Leow, R. P. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania*, 84, 496.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud : The issue of reactivity in sla research methodology. *Studies in Second Language Acquisition*, 26(1), 35-57.
- Levelt, W. (1989). *Speaking : From Intention to Articulation*. MIT Press.
- Levelt, W. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C.-W. Wallesch (Éds.), *Linguistic Disorders and Pathologies : An International Handbook* (p. 1-15). De Gruyter.
- Li, X. L. (1988). Effects of contextual cues on inferring and remembering meanings of new words. *Applied Linguistics*, 9(4), 402-413.
- Liang, F., & Bai, X. (2010). 切分空间和切分方式对中文阅读绩效影响的眼动研究 The effects of segmentation space and type in Chinese reading : Evidence from eye movements. *心理研究 Psychological Research*, 3(1), 21-28.
- Liste des caractères communs du chinois moderne* 汉语通用字表. (2009). 国家语言文字委员会 Comité national de la langue et du caractère. 商务印书馆 The Commercial Press.
- Liu, J. (2008). L1 use in L2 vocabulary learning : Facilitator or barrier. *International Education Studies*, 1(2), p65.
- Liu, S. H. (2001). 关于在语境中猜测词义的调查 Guessing the meaning of unfamiliar words from context : An investigation. *汉语学习 Chinese Language Learning*, 1, 45-49.
- Liu, Y. H., Pan, W. Y., & Gu, W. (2005). *实用现代汉语语法 Practical Contemporary Chinese Grammar*. 外语教学与研究出版社 Commercial Press.
- Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Éd.), *Eye Guidance in Reading And Scene Perception* (2nd edition, p. 55-100). Elsevier Science Ltd.
- Long, M. (1988). Instructed interlanguage development. In L. Beebe (Éd.), *Issues in Second Language Acquisition : Multiple Perspectives* (p. 115-141). Rowley, MA: Newbury House.
- Long, M. (1998). Focus on form in task-based language teaching. *Working Papers in ESL*, 16(2), 35-49.
- Long, M., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, 26(1), 27-56.

- Lotto, L., & Groot, A. M. B. de. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, 48(1), 31-69.
- Lü, S. X. (1980). *语文常谈 Parlons du chinois*. 生活·读书·新知三联出版社 Sino United Publishing.
- Lü, X. H. (2007). 留学生在汉语阅读中利用字形和语音信息的研究 A study on the use of graphic and phonetic information by foreign students in Chinese reading. *云南师范大学学报(对外汉语教学与研究版) Journal of Yunan University (Teaching and Reserach of Chinese Foreign Language)*, 5(1), 34-38.
- Ma, S. M. (2009). *外国留学生汉字字形分解识别能力实验研究 A study on the ability of foreign students of Chinese characters decomposition and recognition* [Master Thesis]. 上海外国语大学 Shanghai International Studies University.
- Mason, B., & Krashen, S. (2004). Is form-focused vocabulary instruction worthwhile? *RELC Journal*, 35(2), 179-185.
- Melka, F. (1997). Receptive versus productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Éds.), *Vocabulary : Description, Acquisition, and Pedagogy* (p. 84-102). Cambridge University Press.
- Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.
- Moescheler, J. (2001). Pragmatique : État de l'art et perspectives. *Marges Linguistiques*, 1, 87-102.
- Mokhtari, K., & Kouider, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94(2), 249-259.
- Mondria, J.-A., & Wit-De Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language1. *Applied Linguistics*, 12(3), 249-267.
- Mori, Y. (2003). The roles of context and word morphology in learning new Kanji words. *The Modern Language Journal*, 87(3), 404-420.
- Morrison, L. (1996). Talking about words : A study of French as a second language learners' lexical inferencing procedures. *The Canadian Modern Language Review*, 53(1), 41-75.
- Nagy, W. E. (1988). *Teaching Vocabulary to Improve Reading Comprehension*. ERIC Clearinghouse on Reading et Communication Skills; National Council of Teachers of English; International Reading Association.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school english? *Reading Research Quarterly*, 19(3), 304-330.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270.
- Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253.
- Nagy, W. E., McClure, E. F., & Montserrat, M. (1997). Linguistic transfer and the use of context by Spanish-English bilinguals. *Applied Psycholinguistics*, 18(4), 431-452.

- Nassaji, H. (2003). L2 vocabulary learning from context : Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4), 645-670.
- Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review*, 61(1), 107-135.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (Second edition). Cambridge University Press.
- Nation, I. S. P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, 26(2), 1-16.
- Nguyen-Xuan, A. (1995). Les mécanismes cognitifs d'apprentissage. *Revue Française de Pédagogie*, 112, 57-67.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know : Verbal reports on mental processes. *Psychological Review*, 84(3), 231-259.
- Noordman, L. G. M., & Vonk, W. (2015). Inferences in discourse, psychology of. In *International Encyclopedia of the Social & Behavioral Sciences* (2nd éd., Vol. 12, p. 37-44). Elsevier.
- Pan, J. J. (2007). 韩国和欧美学习者汉字识别中的加工策略研究 *A study on processing strategies in Chinese character recognition of Korean and European and American learners* [Master Thesis]. 北京语言大学 Beijing Language and Culture University.
- Paribakht, T. S. (2005). The influence of first language lexicalization on second language lexical inferencing : A study of Farsi-speaking learners of English as a foreign language. *Language Learning*, 55(4), 701-748.
- Paribakht, T. S. (2010). The effect of lexicalization in the native language on second language lexical inferencing : A cross-language study. In D. R. Chacón-Beltrán, C. Abello-Contesse, & M. del M. Torreblanca-López (Éds.), *Further Insights into Non-Native Vocabulary Teaching and Learning*. Multilingual Matters.
- Paribakht, T. S., & Tréville, M.-C. (2007). L'inférence lexicale chez des locuteurs de français et des locuteurs de persan lors de la lecture de textes anglais : Effet de la lexicalisation en première langue. *The Canadian Modern Language Review*, 63(3), 399-428.
- Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 09-29.
- Paribakht, T. S., & Wesche, M. (1996). *Enhancing vocabulary acquisition through reading : A hierarchy of text-related exercise types*. 52(2), 155-175.
- Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Éds.), *Second Language Vocabulary Acquisition : A Rationale for Pedagogy* (p. 174-202). Cambridge University Press.

- Paribakht, T. S., & Wesche, M. (1999). Reading and « incidental » L2 vocabulary acquisition : An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, 21(2), 195-224.
- Paris, S. G., Wasik, B., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Éds.), *Handbook of Reading Research* (Vol. 2, p. 609-640). Lawrence Erlbaum Associates, Inc.
- Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel : Do things fall apart?. *Reading in a Foreign Language*, 22(1), 31-55.
- Peng, D., Li, W., Li, R., & Liu, Y. (1999). Information processing of Chinese compound words. *Psychologica (Japanese Psychological Society)*, 42(4), 252-266.
- Peters, E., Hulstijn, J., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German Vocabulary Through Reading : The Effect of Three Enhancement Techniques Compared. *Language Learning*, 59(1), 113-151.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading : A case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Prince, P. (1996). Second language vocabulary learning : The role of context versus translations as a function of proficiency. *The Modern Language Journal*, 80(4), 478-493.
- Prior, A., Goldina, A., Shany, M., Geva, E., & Katzir, T. (2014). Lexical inference in L2 : Predictive roles of vocabulary knowledge and reading skill beyond reading comprehension. *Reading and Writing*, 27(8), 1467-1484.
- Programme d'enseignement du chinois pour les étudiants étrangers dans les établissements d'enseignement supérieur : Études avancées à long terme (高等学校外国留学生汉语教学大纲：长期进修)*. (2002). 国家对外汉语教学领导小组办公室 (Bureau de la direction nationale pour l'enseignement du chinois comme langue étrangère). 北京语言文化大学出版社 Beijing Language and Culture University Press.
- Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition : A matter of topic familiarity? *Language Learning*, 54(3), 469-523.
- Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, 28(1), 66-86.
- Qian, D. D. (2004). Second language lexical inferencing : Preferences, perceptions and practices. In P. Bogaards & B. Laufer (Éds.), *Vocabulary in a Second Language : Selection, Acquisition, and Testing* (p. 155-169). Benjamins.
- Qian, D. D. (2005). Demystifying lexical inferencing : The role of aspects of vocabulary knowledge. *TESL Canada Journal*, 22(2), 34-54.
- Qian, X. J. (2003). 汉语阅读中的伴随性词汇学习研究 On incidental vocabulary learning in Chinese L2 reading. *北京大学学报 (哲学社会科学版) Journal of Peking University (Philosophy and Social Sciences)*, 40(4), 135-142.
- Qian, X. J. (2005). 词义猜测的过程和猜测所用的知识—伴随性词语学习的个案研究 Word meaning inferring and types of knowledge contributing to the process of inferring : A case study on incidental vocabulary acquisition. *世界汉语教学 Chinese Teaching in the World*, 1, 87-96.

- Radach, R., Inhoff, A., & Heller, D. (2004). Orthographic regularity gradually modulates saccade amplitudes in reading. *European Journal of Cognitive Psychology*, *16*, 27-51.
- Raptis, H. (1997). Is second language reading vocabulary best learned by reading? *Canadian Modern Language Review*, *53*(3), 566-580.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.
- Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science*, *14*(4), 385-388.
- Rayner, K., Slattery, T., Drieghe, D., & Liversedge, S. (2011). Eye Movements and Word Skipping During Reading: Effects of Word Length and Predictability. *Journal of experimental psychology. Human perception and performance*, *37*, 514-528.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*, 355-371.
- Read, J. (2004). *Assessing Vocabulary*. Cambridge University Press.
- Reynolds, B. L. (2016). The effects of target word properties on the incidental acquisition of vocabulary through reading. *The Electronic Journal for English as a Second Language*, *20*(1-31).
- Reynolds, B. L., & Wible, D. (2014). Frequency in incidental vocabulary acquisition research: An undefined concept and some consequences. *TESOL Quarterly*, *48*(4), 843-861.
- Ridgway, T. (2000). Listening strategies—I beg your pardon? *ELT Journal*, *54*(2), 179-185.
- Robert, P., Rey, A., & Morvan, D. (2001). *Le grand Robert de la langue française deuxième édition du dictionnaire alphabétique et analogique de la langue française de Paul Robert* (2e édition augmentée en 6 volumes nouveau format ed. 2001. Print.).
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*(4), 589-619.
- Rott, S., & Williams, J. (2003). Making form-meaning connections while reading: A qualitative analysis of word processing. *Reading in a Foreign Language*, *15*(1).
- Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language teaching research*, *6*(3), 183-222.
- Rousoulioti, T., & Mouti, A. (2016). Dealing with unknown words in L2 reading: Vocabulary and lexical inferencing strategies. *Colombian Applied Linguistics Journal*, *18*(1), 56-70.
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornič (Éd.), *Attention and Performance VI* (p. 573-603). L. Erlbaum Associates.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158.

- Schmidt, R. (1995). Consciousness and foreign language learning : A tutorial on the role of attention and awareness in learning. In R. Schmidt (Éd.), *Attention and Awareness in Foreign Language Learning* (p. 1-63). Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center.
- Schmidt, R. (2001). Attention. In P. Robinson (Éd.), *Cognition and Second Language Instruction* (p. 3-32). Cambridge University Press.
- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Éds.), *CLaSIC 2010: Vol. December, 2-4* (p. 721-737). Singapore: National University of Singapore, Centre for Language Studies.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge University Press.
- Schmitt, N. (2008). Review article : Instructed second language vocabulary learning. *Language Teaching Research*.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43.
- Shao, J., & Li, B. (2019). 汉语作为第二语言阅读中伴随性词汇习得的方式及成效 Assessing the effectiveness of incidental lexical processing strategies in L2 Chinese reading. *华文教学与研究 TCSOL Studies*, 2, 71-80.
- Shen, M. (2017). The role of text type and strategy use in L2 lexical inferencing. *International Review of Applied Linguistics in Language Teaching*, 56, 231-252.
- Shu, H., & Anderson, R. C. (1999). Learning to read Chinese : The development of metalinguistic awareness. In J. Wang, A. W. Inhoff, & H.-C. Chen (Éds.), *Reading Chinese Script : A Cognitive Analysis* (p. 1-18). Lawrence Erlbaum Associates Publishers.
- Shu, H., Anderson, R. C., & Zhang, H. (1995). Incidental Learning of Word Meanings While Reading : A Chinese and American Cross-Cultural Study. *Reading Research Quarterly*, 30(1), 76-95.
- Smith, F. (1994). *Understanding Reading* (5th éd.). Routledge.
- Sok, S. Y. (2014). Deconstructing the concept of 'incidental' L2 vocabulary learning. *Working Papers in TESOL & Applied Linguistics*, 14(2), 21-37.
- Song, H., Zhang, H., & Shu, H. (1995). The developmental shift of the role of graphic code and phonetic code in Chinese reading. *Acta Psychologica Sinica*, 27(02), 139-144.
- Song, W. Y. (2017). *L'intersinographie : Une étude de l'acquisition de l'écriture des caractères chinois par des apprenants francophones* [These de doctorat]. Aix-Marseille.
- Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Éds.), *The Nature of Vocabulary Acquisition* (p. 89-105). Lawrence Erlbaum Associates, Inc.
- Sternberg, R. J., & Powell, J. S. (1983). Comprehending verbal comprehension. *American Psychologist*, 38(8), 878-893.
- Sternberg, R. J., Powell, J. S., & Kaye, D. B. (1982). The nature of verbal comprehension. *Poetics*, 11(2), 155-187.

- Storey, P. (1997). Examining the test-taking process : A cognitive perspective on the discourse cloze test. *Language Testing*, 14(2), 214-231.
- Taillefer, G. F. (1996). L2 reading ability : Further insight into the short-circuit hypothesis. *Modern Language Journal*, 80(4), 461-477.
- Teng, F. (2016). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *The Language Learning Journal*, 1-14.
- Théophanous, O. (2001). Identification des congénères interlinguaux : Quels facteurs sont en jeu ? Un état de la question. *ITL Review of Applied Linguistics*, 131-132, 85-105.
- Théophanous, O. (2019). *Apprendre le vocabulaire dans une langue étrangère : Du mot isolé à la séquence préfabriquée* [Mémoire pour l'Habilitation à Diriger des Recherches en Sciences du langage]. Université de Toulouse II -Jean Jaurès.
- Théophanous, O., & Dreyfus, M. (2004). Attitudes d'inférence lexicale lors d'une tâche de lecture en FLS. *Les cahiers de l'association de didactique du français langue étrangère*, 15, 173-192.
- Théophanous, O., & Perez-Bettan, A. (2021). Séquences préfabriquées et usages chez des apprenants de français L2 : Le point de vue interne au locuteur. In I. Saddour & C. Gunnarsson-Largy (Éds.), *Réflexions sur les usages et les utilisateurs du français : Aspects acquisitionnels et didactiques: Vol. Collection GRAMM-R Etudes de Linguistique française* (p. 125-150). Bruxelles : Peter Lang.
- Tournadre, N. (2014). *Le Prisme des langues : Essai sur la diversité linguistique et les difficultés des langues*. Asiathèque.
- Tseng, W.-T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning : A structural equation modeling approach. *Language Learning*, 58(2), 357-400.
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning : A meta-analysis of correlational studies. *Language Learning*, 69(3), 559-599.
- Van Zeeland, H. (2014). Lexical inferencing in first and second language listening. *The Modern Language Journal*, 98(4), 1006-1021.
- VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition : Terms, linguistic features & research methodology. *AILA Review*, 11, 27-36.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219-258.
- Wang, L. Y. (2011). *日语汉字词及中国学习者习得研究 A study on Japanese Kanji and Chinese learners' acquisition* [PhD Thesis]. 南京大学 Nanjing University.
- Wang, W. (2007). *中文阅读过程中信息提取时间及词频效应的眼动研究 Eye movement study on information acquisition and frequency effect in Chinese reading* [Master Thesis]. 天津师范大学 Tianjin Normal University.
- Wang, Y., Song, B., & Chen, L. (2017). 语素义常用度影响留学生语义透明词习得的实证研究 An empirical study on effects of morpheme meaning frequency on semantic transparent words acquisition for CSL students. *语言文字应用 Applied Linguistics*, 103(03), 134-143.

- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2), 130-163.
- Webb, S. (2005). Receptive and productive vocabulary learning : The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33-52.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46-65.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91-120.
- Werner, H., & Kaplan, E. (1952). The acquisition of word meanings : A developmental study. *Monographs of the Society for Research in Child Development*, 15.
- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge : Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.
- Wesche, M., & Paribakht, T. S. (2009). *Lexical Inferencing in a First and Second Language : Cross-Linguistic Dimensions*. Multilingual Matters.
- Wu, H. C., & Zhao, Y. (2018). 英语、韩语母语者汉语否定标记习得研究 L2 Chinese acquisition of negative markers by English-and Korean-speaking learners. *世界汉语教学 Chinese Teaching in the World*, 32(2).
- Wu, M. J. (2005). 通过猜词策略看欧美韩日学生汉语阅读能力发展过程 *A study on the developmental process of Chinese reading ability : Evidence from investigations on Chinese word-guessing strategies used by learners with alphabetic language, Korean, Japanese backgrounds* [PhD Thesis]. 中山大学 Sun Yat-Sen University.
- Xi, M. Z. (2018). 韩语汉字词对韩国留学生学习汉语词汇的迁移影响 The transfer influence of Chinese characters and words in Korean language on Korean students' Chinese vocabulary learning. *新西部 New West*, 12, 88, 112.
- Xiang, H., Hu, X., Yang, R., & Sui, X. (2018). 中文词切分的依据及眼动控制模型的解释 The basis of Chinese word segmentation and explanation of eye movement control model. *心理科学 Journal of Psychological Science*, 4(15), 1097-1102.
- Xu, T. Q. (2005). “字本位”和语言研究 « Zi » as basic structural unit and linguistic studies. *语言教学与研究*, 6, 1-11.
- Xu, Y. H. (2014a). 复合词结构类型对词义猜测的影响 The influence of compound word structure on words meaning guessing. *语言教学与研究 Language Teaching and Linguistic Studies*, 4.
- Xu, Y. H. (2014b). 面向汉语二语教学的常用复合词语义透明度研究 *Semantic transparency research on frequently used Chinese compound words for second language teaching*. 北京师范大学 Beijing Normal University.
- Yamashita, J. (2001a). Transfer of L1 reading ability to L2 reading : An elaboration of the linguistic threshold. *言語文化論集 Studies in language and culture*, 23(1), 189-200.
- Yamashita, J. (2001b). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*, 25(1), 81-95.

- Yang, H., Peng, D. L., Perfetti, C. A., & Tan, L. H. (2000). 汉字阅读中语音的通达与表征 (I)—字水平与亚字水平的语音及其交互作用 Phonological activation and representation of Chinese characters (I)—the character-level and sub-character-level phonologies and their interaction. *心理学报 Acta Psychologica Sinica*, 32(2).
- Yorio, C. A. (1971). Some sources of reading problems for foreign-language learners. *Language Learning*, 21(1), 107-115.
- Zang, C., Lu, Z., Zhang, Y., & Zhang, M. (2018). 阅读过程中的词跳读及其产生的认知机制 How do readers skip a word during reading? *心理与行为研究 Studies of Psychology and Behavior*, 16(4), 477-483.
- Zang, C., Zhang, M., Bai, X., Yan, G., Angele, B., & Liversedge, S. P. (2017). Skipping of the very high frequency structural particle de (的) in Chinese reading. *Quarterly Journal of Experimental Psychology*, 1-10.
- Zhang, H. (2016). Dissimilation in the second language acquisition of Mandarin Chinese tones. *Second Language Research*, 32(3), 427-451.
- Zhang, H. M., Koda, K., Han, Y. T., & Lin, J. X. (2019). Word-specific and word-general knowledge in L2 Chinese lexical inference: An exploration of word learning strategies. *System*, 87, 102-146.
- Zhang, J. L. (2010). 词义与语素义之间的关系对词义猜测的影响 Effects of the relationship between the meaning of words and morphemes on inferring the meaning of Chinese words. *语言教学与研究 Language Teaching and Linguistic Studies*, 3, 16-22.
- Zhang, S., & Perfetti, C. A. (1993). The tongue-twister effect in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1082-1093.
- Zhang, X., & Lu, X. (2015). The relationship between vocabulary learning strategies and breadth and depth of vocabulary knowledge. *The Modern Language Journal*, 99(4), 740-753.
- Zhao, A., Guo, Y., Biales, C., & Olszewski, A. (2016). Exploring learner factors in second language (L2) incidental vocabulary acquisition through reading. *Undefined*.
- Zhao, W. (2014). 浅议"词汇水平测试"(VLT)和"词汇知识量表"(VKS)在国内二语词汇习得实证研究中的应用 A study on the use of « Vocabulary Proficiency Test » (VLT) and « Vocabulary Knowledge Scale » (VKS) in vocabulary acquisition empirical studies in Chinese as Second Language. *外语测试与教学 Foreign Language Testing and Teaching*, 1, 52-60.
- Zhao, Y. G. (2009). 留学生高频汉字识别过程中形音义激活的时间进程 *Graphic, phonological and semantic activation process of Chinese high frequency characters by beginning and advanced foreign students* [Master Thesis]. 暨南大学 Jinan University.
- Zheng, H. Y. (2006). 汉语双音复合词的词义与语素义关系研究 *A research on the relationship between word meaning and morphemic meaning of bi-syllabic compounds in modern Chinese* [PhD Thesis]. 华中师范大学 Central China Normal University.
- Zhong, H. (2011). *Learning a word : From receptive to productive vocabulary use*. 117-126.

- Zhong, H., & Hirsh, D. (2009). Vocabulary growth in an English as a foreign language context. *University of Sydney Papers in TESOL*, 4, 85-113.
- Zhu, X. Y., & Zhou, J. (2007). 留学生阅读中复合词词义猜测研究 Inferring the meaning of compound words in Chinese reading. *语言文字应用 Applied Linguistics*, 4.
- Zhu, Y. (2004). 边注和查词典等输入调整对留学生伴随性词汇学习的作用 The effect of modified input methods of marginal glosses and dictionary-consulting on the incidental vocabulary learning by foreign students. *世界汉语教学 Chinese Teaching in the World*, 04, 74-84.
- Zhu, Y., & Cui, H. S. (2005). 汉语阅读中的伴随性词汇学习再探 Re-probing of incidental vocabulary learning in Chinese reading. *暨南大学华文学院学报 Journal of College of Chinese Language and Culture of Jinan University*, 2, 15-22.
- Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9(3), 341-368.