



**HAL**  
open science

# Apprendre un mot L2 absent dans ma langue maternelle : une étude auprès d'apprenants francophones du chinois dans la compréhension écrite

Enhao Leger-Zheng

## ► To cite this version:

Enhao Leger-Zheng. Apprendre un mot L2 absent dans ma langue maternelle : une étude auprès d'apprenants francophones du chinois dans la compréhension écrite. Manon Boucharéchas, Iris Fabry, Marie Peuzin, Roxanne Comotti, Rim Abouwarda, Alexis Ladreyt. Sciences du langage : enjeux théoriques et pratiques méthodologiques. Actes du colloque CEDIL22, , 2024, 10.5281/zenodo.13623199 . hal-04700261

**HAL Id: hal-04700261**

<https://hal.univ-grenoble-alpes.fr/hal-04700261v1>

Submitted on 17 Sep 2024

**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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

---

# APPRENDRE UN MOT L2 ABSENT DANS MA LANGUE MATERNELLE : UNE ETUDE AUPRES D'APPRENANTS FRANCOPHONES DU CHINOIS DANS LA COMPREHENSION ECRITE

Enhao LEGER-ZHENG

[enhao.leger-zheng@univ-tlse2.fr](mailto:enhao.leger-zheng@univ-tlse2.fr)

*Laboratoire de NeuroPsychoLinguistique (EA4156), Université de Toulouse - Jean Jaurès, Toulouse*

---

## 1. Lire, inférer et apprendre le vocabulaire

Le vocabulaire joue un rôle fondamental dans la compréhension d'un texte, il est pourtant impossible de maîtriser de manière exhaustive les mots d'une langue. Il n'est pas rare qu'on tombe sur un mot non familier pendant une lecture. À la rencontre d'un mot inconnu, l'inférence lexicale est une stratégie qu'un lecteur peut employer afin de déduire le sens de ce mot, à la lumière de ses connaissances et des indices contextuels (Haastrup, 2008).

Depuis longtemps, les enseignants et chercheurs remarquent que les enfants natifs sont capables de déduire le sens des mots inconnus rencontrés dans une lecture, sans aides extérieures (Werner & Kaplan, 1952). Selon Carnine et al. (1984), l'avantage de cette stratégie est que son utilisation est indépendante de la disponibilité d'un dictionnaire et des instructions spécifiques d'autrui. D'ailleurs, cela permet à l'apprenant de maintenir son attention sur le texte sans être interrompu. En langue seconde/étrangère (L2), l'inférence lexicale est la stratégie la plus utilisée par les apprenants lorsqu'ils rencontrent un mot inconnu dans la compréhension écrite (Qian, 2004). De nombreuses études ont été réalisées pour observer le processus de l'inférence lexicale (Nassaji, 2004) et examiner l'effet des éléments divers sur le processus et sur le résultat de l'inférence lexicale, tels que le niveau linguistique en L2 de l'apprenant (Haastrup, 1991), la disponibilité des indices nécessaires dans le contexte (Mori, 2003) et la connaissance du thème (Atef-Vahid *et al.*, 2013).

L'inférence lexicale est également considérée comme une stratégie importante de l'acquisition du vocabulaire. Des chercheurs ont constaté qu'en langue maternelle (L1), les enfants parviennent à déduire le sens d'un mot inconnu grâce au contexte, puis à retenir ce mot inféré (Nagy *et al.*, 1985, 1987). D'ailleurs, ces études montrent que les enfants et les jeunes apprenants natifs développent leur vocabulaire principalement par le biais de la lecture extensive hors du parcours scolaire. En L2, comme en L1, l'inférence du sens d'un mot inconnu peut aboutir à des gains lexicaux pour ce mot dans le lexique mental de l'apprenant (Fraser, 1999 ; Ghobadi *et al.*, 2016 ; Shao & Li, 2019). Selon l'hypothèse de De Bot *et al.* (1997), le processus d'inférer le sens d'un mot renvoie à l'activation d'un concept correspondant dans la mémoire à long terme de l'apprenant grâce aux indices repérés. L'attention portée à ce mot peut aboutir à la création d'un lemme vide dans le lexique mental de l'apprenant. Pendant l'inférence de ce mot, ce nouveau lemme est progressivement comblé par des informations lexicales repérées, ce qui conduit à l'acquisition de ce mot.

L'acquisition du vocabulaire subséquente à l'inférence lexicale représente un intérêt particulier en L2. Le développement lexical est un processus complexe et long. L'étude d'une L2 a majoritairement lieu en classe dans un contexte d'apprentissage hétéroglotte, ce qui signifie un nombre d'heures pédagogiques très limité. Le contexte hétéroglotte implique le manque de stimuli en dehors de la classe, ce qui complique le problème d'exposition insuffisante à la langue cible. De ce fait, l'acquisition du vocabulaire à la suite de l'inférence fournit une piste « auxiliaire » qui peut aider les apprenants à enrichir leur vocabulaire en L2.

## 2. Lexicalisation et acquisition du vocabulaire à la suite de l'inférence

De nombreux éléments peuvent intervenir pendant l'inférence lexicale et avoir un impact sur le résultat de l'acquisition subséquente. Parmi ces éléments, la distance typologique, ou la similarité structurale, entre la L1 de l'apprenant et la L2 (Koda, 2007), est non négligeable. Un apprenant L2 adulte possède généralement de riches connaissances linguistiques et encyclopédiques grâce au développement de sa langue maternelle. Les connaissances acquises en L1 restent actives pendant les activités en L2 (Kroll *et al.*, 2010). La similarité L1-L2 implique la disponibilité des connaissances « réutilisables » dans l'apprentissage de la L2. Par exemple, dans l'étude de Paribakht & Tréville (2007) réalisée auprès de deux groupes linguistiques, pour inférer les mots anglais inconnus, le groupe francophone s'appuie plus souvent sur les connaissances linguistiques en L1 que le groupe persanophone. En particulier, seulement les apprenants francophones ont utilisé l'homonymie en L1 comme source de connaissance pour inférer les mots anglais.

En d'autres termes, quand on apprend une L2 typologiquement lointaine de sa L1, par rapport à une L2 similaire, moins de connaissances pré-acquises sont « transférables ». D'autant plus, l'apprenant risque davantage de tomber sur des mots L2 n'ayant pas d'équivalent lexical attesté dans sa L1, ou des mots non lexicalisés (Paribakht, 2005). Selon l'hypothèse de De Bot *et al.* (1997, cf. supra), l'inférence du sens d'un mot inconnu renvoie à l'activation d'un concept correspondant dans la mémoire à long terme de l'apprenant. Autrement dit, l'existence d'un concept pour ce mot a une importance fondamentale dans le résultat de la compréhension de ce dernier. Quand un mot L2 n'est pas lexicalisé dans la L1 de l'apprenant, son équivalent L1 est manquant, ainsi que le concept correspondant dans la mémoire à long terme de l'apprenant pour ce mot L2. Par conséquent, l'inférence du sens d'un mot L2 non lexicalisé est cognitivement plus coûteuse et plus difficile. Cela implique également que la nature de lexicalisation d'un mot L2 peut avoir des effets sur le processus et sur le résultat de l'inférence lexicale. Comme l'indique Paribakht (2005), au lieu d'activer un lemme lié à un concept correspondant dans le cas de l'inférence d'un mot L2 lexicalisé, le processus d'inférer un mot L2 non lexicalisé renvoie plutôt à la construction d'un lemme nouveau. Les résultats de l'étude Paribakht (2005) menée à ce sujet ont fourni une preuve : les apprenants persanophones se sont appuyés sur des sources de connaissances plus variées pour inférer le sens des mots anglais non lexicalisés en persan ; d'ailleurs, ils ont eu un taux de réussite presque trois fois plus faible pour l'inférence des mots anglais non lexicalisés que pour les mots lexicalisés. Les apprenants francophones dans l'étude de Paribakht & Tréville (2007) ont également eu un degré de réussite inférieur pour l'inférence des mots anglais non lexicalisés en français (28 % vs 34 % pour les mots anglais lexicalisés). En accord avec ces deux études, les résultats de l'étude de Fan & Zhang (2014) révèlent une plus grande difficulté chez les apprenants sinophones à inférer les mots anglais non lexicalisés en chinois (seulement 18 % des réponses sont entièrement correctes), par rapport à ceux lexicalisés (48 % des réponses sont entièrement correctes).

Les mots L2 non lexicalisés semblent également plus difficiles à retenir pour les apprenants. Dans l'étude de Paribakht (2005), les résultats du post-test immédiat après l'inférence lexicale montrent que les apprenants persanophones ont eu un meilleur score de rétention pour les mots anglais lexicalisés en persan (2,16/5) que pour les mots non lexicalisés (2,03/5). Dans l'étude de Chen & Truscott (2010) sur l'acquisition du vocabulaire via la lecture, les résultats du posttest immédiat après la lecture montrent qu'un apprenant sinophone peut obtenir un gain sémantique pour un mot anglais lexicalisé en chinois au bout de trois fois d'exposition, tandis que, pour un mot anglais non lexicalisé, il n'a quasiment aucun gain sémantique même au bout de sept fois d'exposition. Les chercheurs concluent alors qu'il est nécessaire de fournir davantage d'instructions explicites pour les mots L2 non lexicalisés. Dans le cadre des travaux de Heidari-Shahreza & Tavakoli (2016) concernant également l'acquisition du vocabulaire via la lecture, les apprenants persanophones de l'anglais L2 ont eu un score significativement meilleur pour la rétention immédiate des mots anglais lexicalisés en persan, par rapport à ceux non lexicalisés. En outre, deux semaines plus tard, lors

du post-test, les apprenants ont eu une diminution de score significative sur le plan sémantique seulement pour les mots anglais non lexicalisés en persan.

### 3. Questions de recherche

Malgré l'importance de la lexicalisation dans l'acquisition du vocabulaire à la suite de l'inférence en L2, il existe jusqu'à présent peu d'études dans ce domaine. Nous aimerions alors explorer davantage l'effet de la lexicalisation sur l'inférence et l'apprentissage du vocabulaire en L2 dans la compréhension écrite. Nous avons visé le cas où un francophone apprend le chinois L2. La langue chinoise est non alphabétique, et l'unité d'écriture de cette langue est le « caractère (字) ». Les caractères représentent un grand défi pour les apprenants dont la L1 est alphabétique, par exemple la langue française, et ce, dès le début de l'apprentissage. La langue chinoise est distante de la langue française non seulement au niveau linguistique, mais également au niveau socio-culturel. Cela signifie qu'un apprenant francophone risque de rencontrer plus souvent des mots chinois dont le concept est manquant en français, par rapport à une L2 proche comme l'espagnol ou l'italien.

Nous avons mené la présente étude dans le but d'examiner la possibilité pour un apprenant francophone d'apprendre des mots chinois inconnus rencontrés dans un texte. Précisément, nous nous sommes posés les questions suivantes autour de l'effet de la nature de lexicalisation des mots chinois :

- i) Un apprenant francophone procède-t-il différemment l'inférence pour un mot chinois non lexicalisé en français et celle pour un mot lexicalisé ?
- ii) Un apprenant francophone parvient-il à comprendre un mot chinois non lexicalisé en français présent dans un texte ? Si oui, l'apprenant a-t-il davantage de difficulté à comprendre un mot non lexicalisé qu'un mot lexicalisé ?
- iii) Un apprenant francophone parvient-il à retenir un mot chinois inféré qui est non lexicalisé en français ? Si oui, l'apprenant a-t-il davantage de difficulté à retenir un mot non lexicalisé qu'un mot lexicalisé ?

Au fil de notre lecture, nous avons émis les hypothèses suivantes : un apprenant francophone procède différemment au niveau de l'inférence du sens pour un mot chinois inconnu selon la nature de lexicalisation de ce mot ; un apprenant francophone parvient à comprendre un mot chinois non lexicalisé présent dans un texte, mais avec un degré de réussite plus faible par rapport à un mot chinois lexicalisé ; l'apprenant arrive, par la suite, à retenir à court et à long terme le mot chinois inféré qui n'est pas lexicalisé en français, mais avec plus de difficulté par rapport à un mot chinois lexicalisé.

## 4. Protocole de l'expérience

### 4.1. *Public visé*

Nous avons sollicité trente-trois étudiants universitaires francophones dont l'âge moyen est de 22 ans. Ayant un niveau intermédiaire, les participants avaient fait en moyenne 4 ans d'études en langue chinoise au moment de l'expérience, et ils connaissaient à ce moment-là plus de 600 mots chinois.

### 4.2. *Mots cibles*

En ce qui concerne le choix des mots cibles, nous avons consulté le Dictionnaire chinois-français (« 汉法词典 », 2001) ainsi que les avis des enseignants francophones du chinois L2. Nous avons enfin sélectionné 10 mots chinois lexicalisés (désormais L) en français et 10 autres mots non lexicalisés (désormais NL) (cf.

Tableau 1 pour les exemples des mots cibles de différente nature de lexicalisation). Nous avons ensuite rédigé deux textes narratifs, dans chacun desquels nous avons intégré 5 mots L et 5 mots NL. Les deux textes contiennent au total 639 caractères.

Lexicalisation	Mot cible	Signification
Lexicalisé	风筝	cerf-volant
	五颜六色	de diverses couleurs, multicolore
	在乎	prendre qch. à cœur, faire attention à
Non lexicalisé	白茫茫	d'une blancheur éclatante (immaculée) ; une immensité argentée
	还礼	faire un présent en retour, offrir un cadeau à son tour
	飘零	tomber en voltigeant

Tableau 1 - Exemples de mots cibles

### 4.3. Test VKS

Nous avons adapté le test VKS (« Vocabulary Knowledge Scale », Paribakht & Wesche, 1993) pour évaluer les connaissances lexicales chez nos participants sur les mots cibles. Nous avons utilisé une échelle de six niveaux afin d'examiner les connaissances lexicales sur les plans de la forme, du sens et de l'emploi pour un mot donné (Nation, 2014) chez nos participants. Lors d'un test VKS, parmi les six niveaux présentés ci-dessous, les participants devaient noter pour chaque mot cible le niveau correspondant à leur connaissance.

- Connaissance zéro : le niveau 1 correspond à la connaissance « jamais vu » du mot.
- Connaissance « forme » : les niveaux 2 et 3 correspondent à la connaissance seulement au niveau de la forme du mot, vague (niveau 2) ou certaine (niveau 3).
- Connaissance « forme + sens » : les niveaux 4 et 5 correspondent à la connaissance sémantique partielle (niveau 4) ou entière (niveau 5) pour un mot, en dehors de la connaissance formelle du mot. Si le participant note « niveau 4 » ou « niveau 5 » pour un mot cible, il doit donner le sens du mot pour justifier sa réponse.
- Connaissance « forme + sens + emploi » : le niveau 6 correspond à la connaissance intégrale du mot dans un contexte spécifique. Si le participant note « niveau 6 » pour un mot cible, il doit donner le sens du mot puis formuler une phrase en employant le mot en question.

### 4.4. Procédure

Une semaine avant la séance de l'inférence lexicale, nous avons fait passer un test VKS (VKS1) afin d'examiner les connaissances lexicales sur les mots cibles chez nos participants.

La séance de l'inférence lexicale a eu lieu sous forme d'entretien individuel. Le participant devait lire les textes testés, puis déduire le sens des mots cibles surlignés dans ces textes l'un après l'autre. Il a été demandé au participant d'inférer les mots cibles avec la méthode « penser à voix haute » exigeant que le participant verbalise simultanément tout ce qu'il pense pendant l'inférence d'un mot (Ericsson, 2006). À l'issue de la

séance, le participant a été demandé d'accomplir un exercice lacunaire concernant les mots cibles. L'objectif de cet exercice est de donner l'accès au participant aux sens corrects des mots inférés.

Respectivement deux semaines et quatre semaines après la séance de l'inférence lexicale, nous avons fait passer deux autres tests VKS (VKS2 et VKS3) auprès de nos participants sur les mots cibles.

## 5. Résultats

Selon le résultat du VKS1, certains participants connaissaient déjà quelques mots cibles. Ces cas ont été exclus de nos observations. Nous avons donc examiné 324 inférences pour les mots L et 330 inférences pour les mots NL, soit 654 inférences au total.

### 5.1. *Processus de l'inférence pour les mots L et NL*

Les entretiens d'inférence lexicale ont tous été enregistrés puis transcrits. Nous avons opté pour la typologie de Paribakht (2005) pour identifier les sources de connaissance utilisées par nos participants.

La typologie de Paribakht (2005) contient trois catégories, les sources non linguistiques, les sources linguistiques propres à la L1 ou à la Lx (une autre langue non L2 que l'apprenant connaît et/ou pratique), et les sources linguistiques propres à la L2. Cette dernière catégorie comporte trois sous-catégories de sources respectivement au niveau du discours, de la phrase contenant le mot inféré et du mot inféré. Il est à noter que, dans notre expérience, aucun participant n'a eu recours aux sources linguistiques propres à la L1/Lx. Les définitions des autres sources de connaissance, accompagnées des exemples extraits des entretiens, sont présentées ci-dessous.

- Sources non linguistiques : les connaissances encyclopédiques de l'apprenant, incluant sa connaissance du thème du texte.  
Ex. « 白茫茫 (d'une blancheur éclatante [immaculée] ; une immensité argentée) », mot NL : *Puisqu'on parle de « beaucoup de neige », ça serait la transformation de la ville en une ville « toute blanche ».*
- Sources linguistiques L2, niveau du discours : le sens général du texte, du paragraphe ou des phrases voisines de la phrase contenant le mot inféré.  
Ex. « 飘零 (tomber en voltigeant) », mot NL : *(Dans le paragraphe,) Ça parle de l'automne... la température commence à être douce...*
- Sources linguistiques L2, niveau de la phrase – ponctuation : les règles de ponctuation et leur signification.  
Ex. « 还礼 (faire un présent en retour, offrir un cadeau à son tour) », mot NL : *Le contexte me dit que c'est un résultat puisqu'on introduit quelque chose tout début (avec un) deux-points.*
- Sources linguistiques L2, niveau de la phrase – contexte proche : le sens général de la phrase contenant le mot inféré.  
Ex. « 风筝 (cerf-volant) », mot L : *Hm... (dans la phrase,) « s'il y a beaucoup de vent », Hm... Il y a peut-être quelque chose en rapport avec le vent.*
- Sources linguistiques L2, niveau de la phrase – grammaire : les propriétés syntaxiques ou les catégories grammaticales du mot inféré.  
Ex. « 五颜六色 (multicolore) », mot L : *Je pense que c'est un adjectif pour qualifier... (après) le « de » [的, particule marquant un rapport déterminant/déterminé], de petites fleurs.*

- Sources linguistiques L2, niveau du mot – association/collocation : l'association entre le mot inféré et un autre mot familier ou une expression.

Ex. « 在乎 (prendre qch. à cœur, faire attention à) », mot L : *Hm j'ai vu « ye bu » (dans la phrase), et ça me directement fait penser à « pas du tout » en chinois (qui est « yi dian'r ye bu »).*

En outre, nous avons identifié des indices spécifiquement liés aux caractéristiques de la langue chinoise. Nous avons alors ajouté des sources de connaissance concernant les particularités linguistiques du chinois :

- Sources linguistiques L2, niveau du mot – sens d'un caractère constitutif du mot : déduction du sens d'un mot à partir du sens d'un caractère au sein de ce mot.

Ex. « 还礼 (faire un présent en retour) », mot NL : *Peut-être ils utilisent cet argent pour donner aux autres d'autres cadeaux, parce que c'est le « li » [le deuxième caractère 礼] de « liwu » [礼物 cadeau].*

- Sources linguistiques L2, niveau du mot – forme d'un caractère constitutif du mot : l'identification d'un caractère constitutif et/ou la déduction du sens de ce caractère à partir de son composant.

Ex. « 飘零 (tomber en voltigeant) », mot NL : *Je pense que je connais le caractère séparément. (À droite il y a) le « vent 风 » et (à gauche il) y a le « xi 西 »... ou le caractère qui semble un peu à « yao 要 ».*

- Sources linguistiques L2, niveau du mot – son d'un caractère constitutif du mot : l'identification de la prononciation d'un caractère par le biais de son composant.

Ex. « 飘零 (tomber en voltigeant) », mot NL : *Le premier caractère « feng 风 »... J'ai un doute sur le premier caractère. Mais la partie droite, vu que c'est « feng 风 »... ça donne la prononciation du caractère, dans beaucoup de cas.*

- Sources linguistiques L2, niveau du mot – mode de constitution du mot : la façon à laquelle est constitué un mot ou la relation entre les caractères constitutifs.

Ex. « 五颜六色 (multicolore) », mot L : *Et donc pour moi, ça serait un « chengyu », en quatre caractères, une expression figée, pour décrire les couleurs des fleurs et le printemps.*

Selon les entretiens transcrits, les participants ont utilisé en moyenne plus de deux sources de connaissance différentes pour inférer un mot chinois inconnu. L'ensemble des inférences de mot, ayant engagé deux sources de connaissance et au-delà, représente 82,1 % des inférences observées. Les sources de connaissance non linguistiques étant peu présentes (4,8 %), les sources linguistiques propres à la L2 au niveau phrastique sont les plus utilisées (54,8 %), suivies par celles au niveau du mot (37,3 %) (Figure 1). Les sources les plus fréquemment utilisées pour inférer les mots chinois inconnus sont le contexte proche (36,9 %), le sens d'un caractère constitutif (19,5 %), la grammaire (17,6 %) et la forme d'un caractère constitutif (14 %).

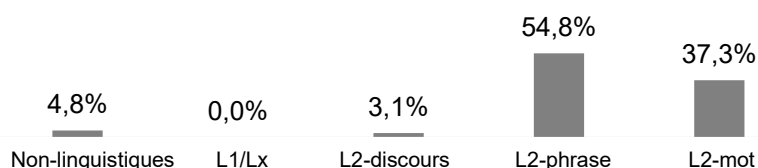


Figure 1 - Poids des sources de connaissance diverses utilisées dans l'inférence des mots chinois

Nos participants ont utilisé en moyenne 2,3 sources différentes pour inférer les mots cibles L et 2,4 sources pour inférer les mots NL. La différence à ce sujet entre les deux catégories de mot n'est pas significative.

Les poids des quatre sources de connaissance principalement utilisées dans l'inférence pour les mots L et les mots NL sont montrés dans la Figure 2 ci-dessous. Nous pouvons constater que le contexte proche est la source la plus fréquemment utilisée dans l'inférence pour les mots L, autant pour les mots NL. Selon les analyses statistiques avec le modèle linéaire général univarié, le poids de l'utilisation de la grammaire est significativement plus important dans l'inférence pour les mots L que dans celle pour les mots NL ( $p = 0,00$ ). En revanche, l'utilisation du sens d'un caractère au sein du mot inféré est plus prononcée dans l'inférence pour les mots NL que pour les mots L ( $p = 0,00$ ). Quant au poids de la source formelle d'un caractère constitutif du mot inféré, la différence d'utilisation entre les inférences des mots L et des mots NL est relativement importante (13 % pour les mots L vs 15 % pour les mots NL). Mais cette différence n'est pas statistiquement significative ( $p = 0,08$ ).

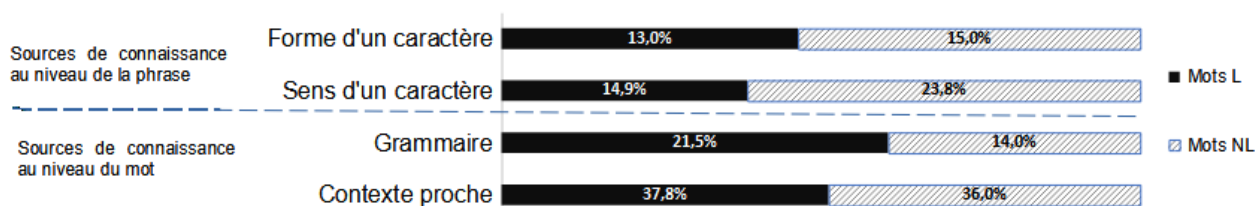


Figure 2 - Poids des sources de connaissance principales utilisées dans l'inférence des mots chinois L et NL

Les résultats révèlent que globalement, pour inférer les mots L, les participants ont préféré les indices au niveau phrastique (59,5 % vs 50,2 % pour les mots NL), tandis qu'ils se sont plus inspirés par les indices au sein du mot pour déduire les sens des mots NL (42,6 % vs 31,7 % pour les mots L). Cela semble valider notre hypothèse sur une différence de processus entre l'inférence d'un mot chinois lexicalisé et celle d'un mot non lexicalisé.

## 5.2. Scores de l'inférence pour les mots L et NL

Afin de pouvoir réaliser des analyses quantitatives, nous avons utilisé une échelle de cinq niveaux pour noter les réponses sur le sens des mots cibles obtenues lors de l'inférence lexicale, de 0 point à 1 point selon le rapprochement au sens correct du mot inféré.

Les réponses erronées sont notées 0. Outre ceci, 133 inférences, soit 20,3 % des 654 inférences observées, se sont trouvées abandonnées (c'est-à-dire que les participants n'ont pas pu donner une réponse sur le sens d'un mot inféré). Ces inférences abandonnées sont également notées 0 point. Les réponses notées 0 représentent 62,1 % de la totalité des 654 réponses d'inférence. Le poids des réponses partiellement correctes et celui des réponses entièrement correctes sont respectivement de 29,2 % et de 8,7 %. (Figure 3)

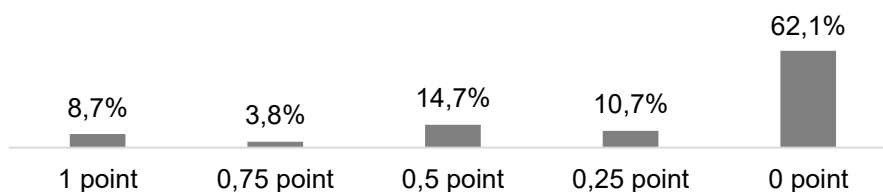


Figure 3 - Répartition des réponses d'inférence pour les mots chinois selon le score

Les participants ont eu un moyen de 0,22 point pour les 654 inférences de mot observées, le score maximal étant 1 point. Le score moyen pour l'inférence des mots L est de 0,19 point, mais de 0,25 point pour l'inférence des mots NL. La différence de score entre les mots de différente nature de lexicalisation est significative selon les analyses statistiques avec T-test bilatéral ( $p = 0,01$ ). La répartition des réponses d'inférences selon le score pour les mots L et NL est montrée dans la Figure 4. Proportionnellement, nos participants ont donné plus de réponses partiellement correctes pour les mots chinois NL (47,9 % vs 10,2 %



pour les mots L), mais davantage de réponses entièrement correctes pour les mots L (13,9 % vs 3,6 % pour les mots NL).

Parmi les sources de connaissance utilisées par nos participants pour inférer les mots cibles, nous avons observé un lien étroit entre l'utilisation du sens d'un caractère constitutif du mot inféré et le degré de réussite de l'inférence de ce mot. Cette source de connaissance a été utilisée dans 63,9 % des inférences entièrement réussies. Ce poids se révèle encore plus important dans les inférences partiellement réussies, soit de 75,4 %.

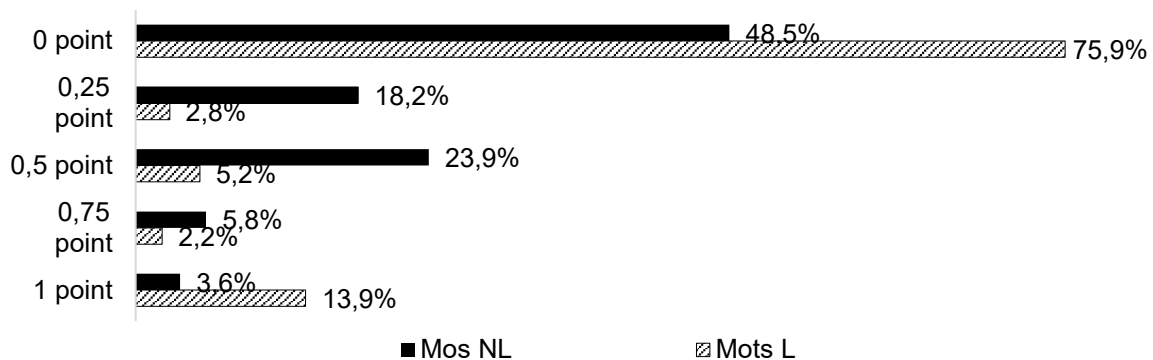


Figure 4 - Répartition des réponses d'inférence pour les mots chinois L et NL selon le score

En résumé, nos participants parviennent à comprendre le sens d'un mot chinois inconnu présent dans un texte, quelle que soit la nature de lexicalisation de celui-ci. Cependant, le taux de réussite de l'inférence est globalement médiocre. Selon les scores d'inférence pour les mots L et les mots NL, nos participants ont eu une meilleure performance pour inférer les mots chinois non lexicalisés en français que ceux lexicalisés. Nous avons émis une hypothèse sur l'effet négatif de la non-lexicalisation sur la compréhension d'un mot chinois grâce aux contextes. Les résultats présentés ci-dessus n'ont donc pas validé notre hypothèse.

### 5.3. Gains lexicaux à court et à long terme pour les mots inférés L et NL

Nous examinons maintenant les résultats des post-tests ayant lieu respectivement deux semaines et quatre semaines après la séance de l'inférence lexicale. Les résultats sur la rétention subséquente des mots inférés se révèlent-ils aussi « inattendus » ?

Selon les réponses de nos participants, du VKS1 (prétest, 7 jours avant la séance de l'inférence) au VKS3 (posttest, 28 jours après l'inférence), la proportion des mots, sur lesquels les participants avaient une connaissance zéro, a diminué de 62,7 % à 11,9 % ; le poids des mots sur lesquels les participants avaient une connaissance formelle et sémantique a augmenté de 1,5 % à 21,8 %. Enfin, au stade pré-inférence, aucun participant ne savait employer l'un des mots cibles dans un contexte. Au VKS3, des participants ont réussi à produire une phrase correcte avec certains mots (Figure 5). Selon les analyses statistiques réalisées conformément au modèle mixte pour les variables ordinales, l'évolution des connaissances lexicales sur divers plans chez nos participants sur les mots cibles est significative ( $p = 0,00$ ).

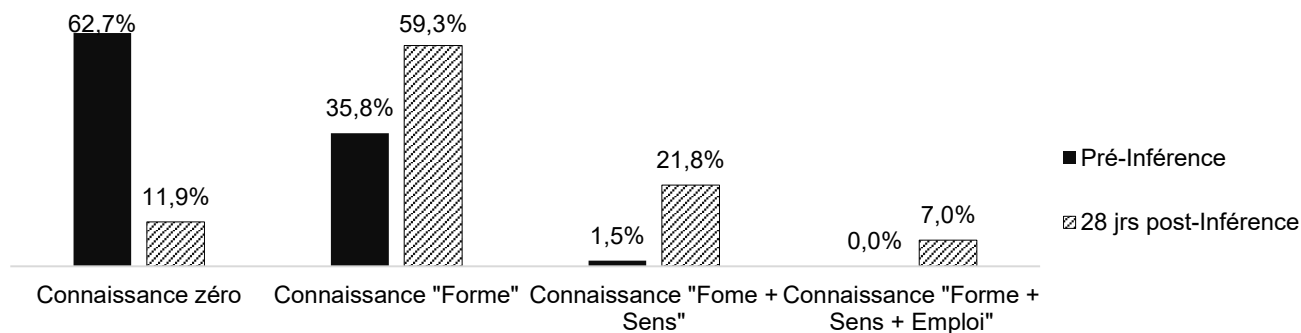


Figure 5 - Évolution des connaissances lexicales sur les mots chinois

Sur le plan sémantique, afin d'évaluer la rétention des mots cibles chez nos participants, nous avons noté les réponses obtenues lors des post-tests (VKS2 et VKS3) selon le même barème appliqué pour la notation des réponses d'inférence lexicale.

Selon les résultats des deux post-tests, bien que nos participants aient mieux réussi l'inférence pour les mots cibles NL (0,25 point), le score de rétention en moyenne pour les mots de cette catégorie a connu une chute de 24 % à la phase post-inférence (0,19 point au VKS2 et au VKS3). Quant aux mots cibles L, le score de rétention sémantique a augmenté respectivement de 15,8 % au VKS2 (0,22 point) et de 42,1 % au VKS3 (0,27 point), par rapport au score de l'inférence (0,19 point). L'évolution de la rétention sémantique ultérieure pour les mots L et celle pour les mots NL s'avèrent opposées, selon les courbes de tendance. (Figure 6)

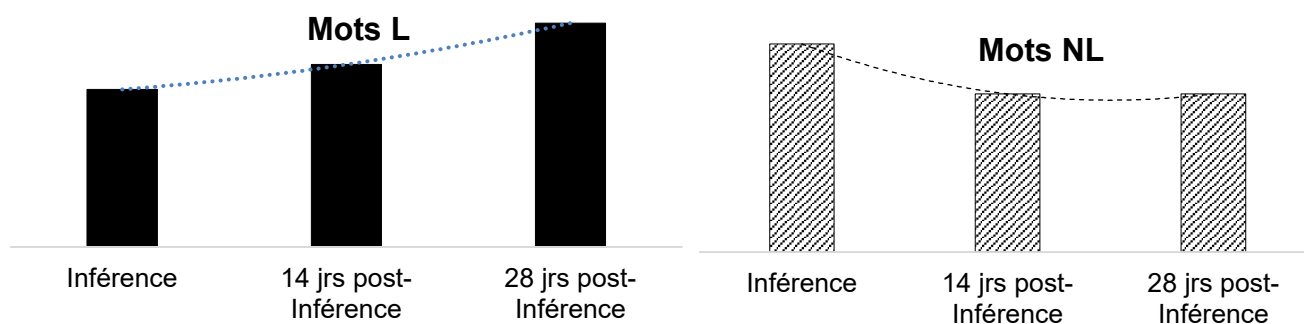


Figure 6 - Évolution de la connaissance sémantique pour les mots chinois L et NL

L'évolution de la connaissance sémantique sur les mots L chez nos participants est statistiquement différente de celle sur les mots NL, selon les analyses avec le modèle mixte linéaire ( $p = 0,01$ ). Nos participants ont donc pu retenir les sens des mots chinois non lexicalisés en français, mais avec plus de difficulté par rapport à ceux lexicalisés. Ces résultats valident donc notre troisième hypothèse sur l'effet négatif de la non-lexicalisation d'un mot chinois sur sa rétention ultérieure.

## 6. Discussion

En termes de processus de l'inférence, nos résultats montrent que nos participants ont eu recours à un panel de sources de connaissance variées pour inférer le sens d'un mot chinois inconnu rencontré dans une lecture, indépendamment de la nature de lexicalisation de ce mot. La différence de diversité au niveau des sources de connaissance utilisées pendant l'inférence pour les mots L et les mots NL n'est pas significative. Les indices au niveau phrastique sont importants pour l'inférence des mots chinois des deux catégories, lexicalisés et non lexicalisés, notamment le sens de la phrase comportant le mot inféré. Cependant, le poids de cette catégorie de sources est très prononcé par rapport aux autres catégories dans l'inférence pour les mots L

(presque 60 %) ; pour inférer les mots NL, nos participants ont eu un appui important sur cette catégorie (50 %), aussi sur celle au niveau du mot (43 %). Cela est en accord avec les résultats des études en anglais L2 focalisant sur la lexicalisation : les apprenants ont un choix de sources de connaissance plus variées pour inférer les mots L2 non lexicalisés dans leur L1, par rapport aux mots lexicalisés (Paribakht & Tréville, 2007 ; Fan & Zhang, 2014). Nous pouvons dire que cela est probablement lié à l'absence d'un équivalent en français, donc d'un concept correspondant dans la mémoire de nos participants, pour les mots chinois NL. Il est possible que la phrase comportant un mot chinois NL, la piste d'indice principale, ne fournisse pas suffisamment d'informations pour que les participants puissent comprendre ce mot. Nos participants doivent alors s'inspirer d'autres pistes, à savoir un contexte plus large (au niveau du discours) ou plus petit (au niveau 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. »

Au niveau du résultat de la compréhension des mots chinois inconnus grâce aux contextes, globalement, nos participants du niveau intermédiaire en chinois L2 ont montré une grande difficulté : 60 % des réponses sont notées 0 point, dont plus de 20 % concernent des inférences abandonnées. Nos participants ont mieux réussi l'inférence pour les mots NL que pour les mots L, un résultat opposé à ceux des études susmentionnées réalisées en anglais L2. Afin de comprendre cela, nous avons examiné les 20 mots cibles sélectionnés, dont 19 sont composés de deux caractères (par exemple, 风筝, *cerf-volant*), voire plus (par exemple, 白茫茫, *d'une blancheur argentée immense* et 五颜六色, *multicolore*). Nous avons constaté que les mots cibles NL comportent plus de caractères connus par nos participants que les mots L sélectionnés. En chinois contemporain, la grande majorité des mots sont constitués de plusieurs caractères. Il n'est pas rare qu'un caractère constitutif d'un mot chinois ait un apport sémantique important au sens global de ce mot. Dans le cas où l'apprenant connaît un ou plusieurs caractères constitutifs d'un mot chinois nouveau qui est sémantiquement transparent (c'est-à-dire la signification du mot est étroitement liée au sens de ses caractères constitutifs), l'apprenant parvient à comprendre au moins partiellement ce mot nouveau. À titre d'exemple, dans le mot cible « 白茫茫, *d'une blancheur argentée immense* », non lexicalisé en français, « 白, *blanc* », est un caractère courant et connu par nos participants. Beaucoup d'entre eux ont alors proposé, lors de l'inférence, une réponse liée à la couleur blanche pour le sens de ce mot.

Cela peut également expliquer que nos participants ont eu davantage de réponses partiellement correctes pour les mots NL que pour les mots L. Paribakht (2005) souligne dans ses conclusions que, bien qu'un concept existant dans la mémoire à long terme de l'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. La chercheuse explique que :

« la familiarité des apprenants avec les composants conceptuels constitutifs de l'item lexical (c'est-à-dire lorsque les concepts de composants existent dans le système conceptuel) peut leur permettre de paraphraser mentalement le sens du mot dans leur L1 et de comprendre au moins partiellement son sens. Dans ce cas-là, l'élément lexical inconnu est traité à un degré suffisant pour la compréhension » (Paribakht, 2005 : 730).

Dans le cas de notre étude, la connaissance sémantique sur les caractères constitutifs des mots NL semble avoir permis à nos participants de comprendre ces mots, même si dans la majorité des cas il s'agit d'une compréhension partielle. En revanche, nos résultats montrent que le degré de réussite entière pour l'inférence des mots L est presque quatre fois plus important que celui pour les mots NL. Dans l'étude de Paribakht (2005), l'écart du degré de réussite entière entre les mots L et NL est également remarquable (respectivement

17 % et 6 %). Cela confirme que la disponibilité d'un équivalent lexical en L1 de l'apprenant pour un mot L2 a un effet positif sur la compréhension entière du mot L2.

Enfin, les résultats des trois tests d'évaluation révèlent globalement une évolution significativement positive des connaissances lexicales sur les 20 mots chinois chez les apprenants francophones, en particulier sur les plans formel et sémantique. Cette évolution concerne les mots chinois lexicalisés en français, ainsi que les mots non lexicalisés. Cela fournit une preuve solide au sujet de la possibilité pour les apprenants francophones d'apprendre de nouveaux mots chinois via la compréhension écrite.

Précisément, au niveau de l'évolution de la connaissance sémantique, le score de l'inférence en moyenne pour les mots NL est supérieur à celui pour les mots L de 33 % ; 14 jours plus tard, le score du rappel du sens des mots L devance celui des mots NL de 16 % ; au bout de 28 jours, l'écart entre les deux scores continue à croître jusqu'à ce que le score en moyenne des mots L soit supérieur à celui des mots NL de 42 %. Compte tenu des évolutions opposées de la connaissance sémantique sur les mots L et NL, il est raisonnable de conclure que, par rapport à un mot chinois lexicalisé en français, il est plus difficile pour un apprenant francophone de retenir le sens d'un mot chinois non lexicalisé à court et à long terme. Nous pensons que cette difficulté est liée au fait que le concept correspondant est absent dans la mémoire à long terme des apprenants francophones pour un mot chinois NL. Quand on apprend un mot L2, le lemme de ce mot L2 est lié à son équivalent en L1 qui correspond à un concept. Les informations lexicales du lemme L1 seront transférées vers le lemme L2 au fur et à mesure de l'utilisation de ce mot L2, jusqu'à ce qu'un lien direct soit établi entre le lemme L2 et le concept correspondant (Jiang, 2004). Dans le cas où un mot L2 non lexicalisé, ce nouveau lemme L2 est possiblement alimenté par plusieurs lemmes L1 correspondant chacun à un concept connexe. Cela représente une plus grande complexité par rapport à un mot L2 lexicalisé. Cette complexité se manifeste aussi au moment du rappel sémantique de ce mot non lexicalisé, parce que l'apprenant risque davantage d'activer partiellement les concepts connexes.

## 7. Conclusions

Nous avons mené cette étude auprès d'apprenants francophones du chinois L2, afin d'examiner la possibilité pour un apprenant francophone d'apprendre un mot chinois inconnu rencontré dans une lecture. Nous nous sommes penchées sur la nature de lexicalisation d'un mot chinois en français, qui signifie l'existence d'un équivalent lexical pour un mot L2 dans la L1 des apprenants. Notre analyse s'est focalisée sur l'effet de la lexicalisation d'un mot chinois sur le processus et le résultat de l'inférence du sens de ce mot, ainsi que sur les gains lexicaux retenus pour ce mot à court et à long terme chez les apprenants francophones.

Nos résultats montrent qu'un apprenant francophone puise des sources de connaissance variées pour comprendre un mot chinois inconnu, lexicalisé ou non en français, présent dans un texte. Le sens de la phrase comportant ce mot est la piste d'indice principale. Toutefois, l'apprenant a besoin d'un ensemble de sources d'indice plus variées pour inférer un mot chinois non lexicalisé en français qu'un mot lexicalisé, probablement à cause de l'absence d'un équivalent en français, c'est-à-dire le manque d'un concept correspondant. La nature de lexicalisation peut donc influencer le processus de l'inférence du sens d'un mot chinois en termes de sources de connaissance utilisées.

Quant au résultat de l'inférence, la nature de lexicalisation d'un mot chinois n'est pas le facteur définitif dans la compréhension de ce mot. Les apprenants francophones parviennent à comprendre le sens d'un mot chinois non lexicalisé grâce aux connaissances de différente nature, notamment les connaissances sémantiques des caractères constitutifs du mot. Ceci dit, il ne faut pas ignorer l'effet facilitateur de la disponibilité d'un équivalent en L1, dans le cas d'un mot L2 lexicalisé, sur une compréhension entière du mot L2.

Les apprenants francophones parviennent à retenir ultérieurement un mot chinois inféré qui n'est pas lexicalisé en français. Cependant, leur faible performance au niveau des mots chinois non lexicalisés, par rapport à ceux lexicalisés, lors du rappel du sens, montre un effet négatif de l'absence d'un équivalent L1 sur la rétention sémantique à court et à long terme. Le lemme d'un nouveau mot L2 non lexicalisé est possiblement alimenté par plusieurs mots L1 liés chacun à un concept connexe. Il semble que cela exige davantage d'efforts cognitifs pour les apprenants par rapport à un transfert lexical un à un dans le cas d'un mot L2 lexicalisé.

En définitive, nous suggérons aux enseignants d'une L2 d'adopter la compréhension écrite comme piste auxiliaire dans l'enseignement du vocabulaire, en classe et hors de l'école. Afin que cela puisse porter ses fruits, l'enseignant doit préparer des supports de lecture compréhensibles, avec un contexte riche d'indices nécessaires pour inférer le sens des mots nouveaux qui y sont intégrés. Concernant un mot L2 pour lequel on ne trouve pas d'équivalent dans la L1 de l'apprenant, il est nécessaire pour l'enseignant de fournir davantage d'instructions explicites et de proposer plus d'activités lexicales autour de ce mot.

### Remerciement

Nous tenons à remercier M. Saïd Jmel, appui en statistiques aux projets scientifiques à l'Université de Toulouse Jean Jaurès, pour son aide généreuse et précieuse pendant les analyses statistiques des données de notre étude.

### Références

- ANDERS ERICSSON, K. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. ANDERS ERICSSON K. AK., CHARNNESS, N., FELTOVICH, P. J. & Hoffman R. R. (Eds.) *The Cambridge handbook of expertise and expert performance*, (p.223-242). Cambridge University Press.
- 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. <https://consortiacademia.org/10-5861jrsll-2012-216/>.
- CARNINE, D., KAMEENUI, E. & COYLE, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly*, 19(2): 188-204.
- CHEN, C. & TRUSCOTT, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, 31(5): 693-713. <https://doi.org/10.1093/applin/amq031>.
- DE BOT, K., PARIBAKHT, T. S. & WESCHE, M. (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. doi:10.1017/S0272263197003021.
- Dictionnaire chinois-français 汉法词典. (2001). 商务印书馆 The Commercial Press.
- 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.
- FRASER, C. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21(2): 225-241. doi:10.1017/S0272263199002041.
- 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(4), 212-218. <https://doi.org/10.7575/aiac.ijalel.v.5n.4p.212>.
- HAASTRUP, K. (1991). *Lexical inferencing procedures, or, talking about words: Receptive procedures in foreign language learning with special reference to English*. Gunter Narr Verlag.

- HAASTRUP, K. (2008). Lexical inferencing procedures in two languages. In ALBRECHTSEN, D., HAASTRUP, K. & HENRIKSEN, B. (Eds.) *Vocabulary and writing in a first and second language: Processes and development* (p.67-111). Palgrave Macmillan, London.
- HEIDARI-SHAHREZA, M. A. & TAVAKOLI, M. (2016). The effect of repetition and L1 lexicalization on incidental vocabulary acquisition by Iranian EFL learners. *The Language Learning Journal*, 44(1), 17-32. <https://doi.org/10.1080/09571736.2012.708051>.
- JIANG, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, 88(3), 416-432. <https://doi.org/10.1111/j.0026-7902.2004.00238.x>.
- KODA, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57(2), 1-44. <https://doi.org/10.1111/j.1467-9922.2007.00411.x>.
- KROLL, J. F., VAN HELL, J. G., TOKOWICZ, N. & GREEN, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: language and cognition*, 13(3), 373-381. DOI : 10.1017/S136672891000009X.
- MORI, Y. (2003). The roles of context and word morphology in learning new Kanji words. *The Modern Language Journal*, 87(3), 404-420. <https://doi.org/10.1111/1540-4781.00198>.
- NAGY, W., ANDERSON, R. & HERMAN, P. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270. <https://doi.org/10.3102/0002831202400>.
- NAGY, W., HERMAN, P. & ANDERSON, R. (1985). Learning words from context. *Reading Research Quarterly*, 20(2), 233-253.
- 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. <https://doi.org/10.3138/cmlr.61.1.107>.
- 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. <https://nflrc.hawaii.edu/rfl/item/303>.
- 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. <https://doi.org/10.1111/j.0023-8333.2005.00321.x>.
- PARIBAKHT, T. S. & TREVILLE, 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. <https://doi.org/10.3138/C271-Q11L-6084-7848>.
- PARIBAKHT, T. S. & WESCHE, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11(1), 09-29. <https://doi.org/10.18806/tesl.v11i1.623>.
- QIAN, D. D. (2004). Second language lexical inferencing: Preferences, perceptions and practices. In BOGAARDS, P. & LAUFER, B. (Éds.) *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (p.155-169). Benjamins.
- 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.
- WERNER, H. & KAPLAN, E. (1952). The acquisition of word meanings : A developmental study. *Monographs of the Society for Research in Child Development*, 15(1).