

Université de Montréal

**Les collocations du champ sémantique des émotions
en mandarin**

par Xiaoyu Zhao

Département de linguistique et de traduction
Faculté des arts et des sciences

Mémoire présenté
en vue de l'obtention du grade de Maître ès arts (M.A.)
en linguistique

Novembre 2018

© Xiaoyu Zhao, 2018

Résumé

Les collocations sont un phénomène récurrent à travers les langues. Il s'agit d'associations arbitraires de mots où l'un est choisi librement et l'autre est choisi en fonction du premier. En raison justement de leur aspect arbitraire, des collocations qui expriment un même sens peuvent être très différentes d'une langue à l'autre. La lexicologie explicative et combinatoire [=LEC] (Mel'čuk, Clas et Polguère, 1995) utilise un outil formel, les fonctions lexicales, pour décrire les collocations dans plusieurs langues. Cependant, il n'existe pas encore de dictionnaire en mandarin qui décrive systématiquement les collocations au moyen des fonctions lexicales.

Ce mémoire présente un modèle de dictionnaire des collocations en mandarin dans le cadre de la lexicologie explicative et combinatoire. À partir d'entrées du champ des émotions choisies dans le DiCoLiLex en français, nous établissons une nomenclature correspondante en mandarin et nous extrayons d'un corpus, à l'aide de diverses méthodes statistiques, les cooccurents des lexies à l'étude afin de trouver les collocations. Ensuite, nous adaptons la microstructure du dictionnaire aux caractéristiques du mandarin afin de faciliter la description lexicologique des lexies dans la nomenclature chinoise. Puis, nous discutons des cas difficiles rencontrés lors de la rédaction des articles. Nous montrons que certains phénomènes ne cadrent pas bien dans la théorie et proposons des pistes de solutions.

Mots-clés : lexicographie, lexicologie explicative et combinatoire, collocations, fonctions lexicales, émotions, mandarin

Abstract

Collocations are recurrent across languages. They are arbitrary associations of words where one is freely chosen, while the other is chosen as a function of the first. Because of their arbitrary aspect, collocations that express the same meaning can vary greatly from one language to another. Explanatory and combinatorial lexicology [=ECL] (Mel'čuk et al., 1995) uses a formal tool, lexical functions, to describe collocations in several languages. However, there is still no dictionary of Mandarin that describes collocations systematically by means of lexical functions.

This dissertation presents a model of a collocation dictionary of Mandarin within the framework of explanatory and combinatorial lexicology. From entries in the field of emotions picked from DiCoLiLex in French, we establish a corresponding nomenclature in Mandarin and extract from a corpus the cooccurents of the lexical units under study to find collocations, using various statistical methods. Then, we adapt the microstructure of the dictionary to the characteristics of Mandarin in order to facilitate the lexicological description of the lexical units in the Chinese nomenclature. Finally, we discuss difficult cases encountered during the process and show that some phenomena do not fit well in the theory and propose possible solutions.

Keywords : lexicography, explanatory and combinatorial lexicology, collocations, lexical functions, emotions, mandarin

Table des matières

Résumé	i
Abstract	ii
Table des matières.....	iii
Liste des tableaux.....	v
Liste des figures	vi
Liste des abréviations et des sigles	vii
Remerciements.....	viii
Chapitre 1 Introduction.....	1
1.1 Problématique	1
1.2 Objectifs du mémoire.....	5
1.3 Organisation du mémoire.....	5
Chapitre 2 Cadre théorique.....	6
2.1 La lexicologie explicative et combinatoire	6
2.2 La lexicologie du mandarin moderne	8
2.3 Typologie des phrasèmes	16
2.4 Les fonctions lexicales	20
2.4.1 Axe standard/non standard.....	22
2.4.2 Axe paradigmatique/syntagmatique.....	23
2.4.3 Axe simple/complexe/configuration	24
2.5 DEC et DiCoLiLex	25
Chapitre 3 Méthodologie de la recherche	28
3.1 Établissement de la nomenclature.....	28
3.2 Choix du corpus	33
3.3 Segmentation du corpus	35
3.4 Extraction automatique des cooccurents	36
3.4.1 Score z et score t	39
3.4.2 Information mutuelle (MI) et information mutuelle spécifique (PMI)	40
3.4.3 Variantes du PMI	42

Chapitre 4	Dictionnaire en mandarin.....	45
4.1	Structure XML du DiCoLiLex.....	45
4.2	Adaptation de la structure dans XML.....	48
4.3	Typage sémantique des lexies et des actants	51
4.4	Difficultés lexicographiques	53
4.4.1	Mots élastiques.....	53
4.4.2	Signes complexes de la forme « V?N ».....	58
4.4.3	Constructions verbales sérielles résultatives (= CVSR)	64
4.4.4	Adv ₁ Magn vs Magn postposé : le cas de 死 ^{sǐ}	73
4.4.5	有 ^{yǒu} comme Oper ₁ ou Manif.....	76
4.4.6	高兴 ^{gāo xìng} : une lexie ou deux ?	79
Chapitre 5	Conclusion	83
Bibliographie	87
Annexe 1.	Structure XML du DiCoLiLex-FR.....	i
Annexe 2.	Structure XML du DiCoLiLex-ZH	iii
Annexe 3.	Exemples de vocables du DiCoLiLex-ZH	vi

Liste des tableaux

Tableau I.	Typologie des phrasèmes (Mel'čuk, 2015).....	17
Tableau II.	Nomenclature des lemmes en mandarin et leurs équivalents en français ..	33
Tableau III.	Corpus en mandarin moderne	34
Tableau IV.	Scores z et t des principaux cooccurrents de 恐惧 (‘effroi; s’effrayer’)...	40
Tableau V.	Ratio observé/attendu et PMI de 恐惧 (‘effroi; s’effrayer’).....	42
Tableau VI.	Intervalles de variation de PMI, NPMI, PMI ² et PMI ³	43
Tableau VII.	NPMI et PMI ³ de 恐惧 (‘effroi; s’effrayer’).....	44
Tableau VIII.	Mots élastiques dans notre nomenclature chinoise.....	54

Liste des figures

Figure 1.	SSém et SSyntP de l'expression 她痛苦 ^{tā tòng kǔ} 'elle est douloureuse'	31
Figure 2.	Description du vocable HEUREUX dans le DiCoLiLex	46
Figure 3.	Description du vocable 担心 ^{dān xīn} 'inquiétude' dans le DiCoLiLex-ZH	50
Figure 4.	Description de la lexie HONTE ₁ dans le DiCoLiLex	51
Figure 5.	SSém et SSyntP de la collocation 爱上 ^{ài shàng}	71
Figure 6.	SSém, SSyntP et SSyntS de la phrase (58a)	77

Liste des abréviations et des sigles

Adj : Adjectif

CL : Classificateur

CVSR : Construction Verbale Sérielle Résultative

DEC : Dictionnaire Explicatif et Combinatoire

ERR : Erreur

FL : Fonction lexicale

LEC : Lexicologie Explicative et Combinatoire

N : Nom

NEG : Marqueur de négation

NPR : Nom propre

PERF : Marqueur de l'aspect accompli

PROG : Marqueur de l'aspect progressif

SSém : Structure Sémantique

ssi : si et seulement si

SSyntP : Structure Syntaxique Profonde

SSyntS : Structure Syntaxique de Surface

TST : Théorie Sens-Texte

V : Verbe

Remerciements

En préambule à ce mémoire, je tiens à remercier tous ceux et celles qui m'ont soutenue et encouragée au fil de mes trois ans d'études à l'Université de Montréal.

Pour commencer, je tiens à exprimer mes remerciements les plus Magn et Ver (profonds et sincères) à mon directeur de recherche, François Lareau, pour ses amples connaissances, sa grande disponibilité, son immense patience et ses encouragements constants tout au long de la rédaction de ce mémoire. Sans lui, ce mémoire serait purement « du chinois ». Merci également de m'avoir guidée dans le monde des fonctions lexicales depuis le tout début.

Je tiens également à remercier mon codirecteur, Jian-Yun Nie, pour avoir accepté à relire mon mémoire et m'avoir donné des conseils.

Je voudrais aussi remercier Benoît Robichaud, qui m'a généreusement aidée à créer une belle interface pour mon dictionnaire en ligne.

Je tiens à remercier les membres du jury Marie-Claude L'Homme et Monique Cormier pour leurs précieux commentaires.

Mes remerciements vont également aux professeurs et à mes collègues de l'Observatoire de Linguistique Sens-Texte pour les petits moments amicaux le midi remplis de bonne humeur et de rire. Je les remercie aussi de m'avoir nommée récipiendaire de la bourse OLST en 2016.

J'aimerais aussi remercier mon ami de cœur Alessandro Lago qui a enduré toutes mes angoisses et m'a réconfortée durant les périodes difficiles de mes études. Merci pour tes magies d'amour qui diminuent mon stress et me font voir la vie en rose.

Enfin et surtout, je remercie humblement ma famille en Chine, qui m'a toujours soutenue et m'a encouragée à poursuivre des études supérieures. Merci papa et maman, pour me faire découvrir la beauté de la langue chinoise depuis mon enfance. Merci de tout mon cœur !

Chapitre 1 Introduction

1.1 Problématique

Une collocation est une association arbitraire de lexies où l'une est choisie librement et l'autre est choisie en fonction de la première. En raison justement de leur aspect arbitraire, des collocations qui expriment un même sens peuvent être très différentes d'une langue à l'autre. Alors, on ne peut pas traduire mot à mot les collocations; il faut connaître l'expression collocationnelle équivalente dans la langue cible, si elle existe. Ainsi, il est important d'avoir un dictionnaire qui décrit les collocations de manière systématique et exhaustive pour qu'un utilisateur puisse les consulter afin de produire des phrases naturelles. Cependant, les dictionnaires actuels ne donnent généralement pas toutes les collocations dans les articles, et il est parfois difficile pour les utilisateurs de trouver les expressions appropriées.

Quant aux dictionnaires monolingues du mandarin, à part l'absence d'une bonne description des collocations, le problème de la description lexicographique se pose dès l'origine en raison de l'unité lexicale utilisée dans les dictionnaires traditionnels. En fait, les dictionnaires monolingues chinois existants peuvent se diviser en deux grandes catégories : 字典^{zì diǎn} 'dictionnaire des sinogrammes'¹, qui donne des définitions des morphes, et 词典^{cí diǎn} 'dictionnaire des lexies', qui donne des définitions des lexies. Néanmoins, peu importe leur type, « l'unité de base de la plupart des dictionnaires monolingues chinois n'est pas la lexie, mais le caractère chinois, ce qui est très gênant puisque le caractère est un signifiant graphique et non pas un signe » (Nguyen, 2006). Même pour un morphe dissyllabique, l'entrée serait le premier

¹ Nous utilisons le terme « sinogramme » pour désigner le graphème chinois. Pour une typologie des sinogrammes, voir la note 8 à la section §2.2, p. 8.

sinogramme du morphe. Par exemple, la lexie 葡萄 ^{pú táo} ‘raisin’ dans 《现代汉语词典》 (*Dictionnaire du mandarin moderne*)² est classée sous l’entrée du sinogramme 葡, qui n’a pas de sens en lui-même. Sous la même entrée, nous trouverons d’autres lexies dérivées de celle-ci, telles que 葡萄干 ^{pú táo gān} ‘raisin’, 葡萄酒 ^{pú táo jiǔ} ‘vin’ (litt. ‘raisin alcool’), 葡萄糖 ^{pú táo táng} ‘glucose’ (litt. ‘raisin sucre’), etc. Pour les locutions et les collocations, elles sont toujours présentées sous l’entrée de leur premier sinogramme. Ainsi, la locution 吃醋 ^{chī cù} ‘être jaloux’ (litt. ‘manger vinaigre’) est présentée sous l’entrée du sinogramme 吃 ‘manger’, alors que sous l’entrée de 醋 ^{cù}, nous pouvons trouver deux acceptions : ‘liquide acide pour assaisonner’ et ‘jalousie’. Deux exemples 醋意 ^{cù yì} et 吃醋 ^{chī cù} sont donnés tout de suite après la deuxième acception, mais sans aucune interprétation, et en dessous, il y a une sous-entrée pour la locution 醋意 ^{cù yì} ‘jalousie’ (litt. ‘vinaigre-sentiment’).

Nguyen (2006) a examiné les lacunes des dictionnaires chinois d’une manière détaillée en les regroupant en deux types : les lacunes typiques des dictionnaires chinois, par exemple dans les aspects de l’unité de base du dictionnaire, les parties du discours, le statut morphologique des lexies, le registre et le degré de liberté syntaxique; et les lacunes communes aux dictionnaires monolingues traditionnels occidentaux et chinois, telles que le problème des vocables, la définition lexicographique, le régime, les dérivés sémantiques et surtout les collocations. En réalité, il existe plusieurs dictionnaires qui traitent les collocations en mandarin, mais ils ont des lacunes importantes. Par exemple, 《现代汉语搭配词典》 (*Dictionnaire des collocations du mandarin moderne*, Mei, 1999) utilise comme entrées des morphes autonomes, ensuite comme sous-entrées des mots-formes composés, des locutions, des collocations, voire des syntagmes libres qui contiennent les morphes vedettes, mais il liste simplement les combinaisons possibles pour une lexie donnée sans préciser leur relation sémantique ou syntaxique avec le mot-clé; 《现代汉语实词搭配词典》 (*Dictionnaire des collocations des*

² Compilé par le bureau d’édition du dictionnaire, Institut de linguistique, Académie chinoise des sciences sociales, c’est le premier dictionnaire des lexies du mandarin standard publié depuis la naissance de la République populaire de Chine, et c’est aussi une référence très importante dans les écoles. Sa 7^e édition a été publiée en 2016 et contient 70 000 entrées (cf. <https://zh.wikipedia.org/wiki/现代汉语词典> [consulté le 4 sept. 2018]).

mots pleins du mandarin moderne, Zhang et Lin, 1992) possède les entrées des lexies dissyllabiques, et il donne les fonctions syntaxiques que peut jouer la lexie auprès d'autres lexies qui se combinent à elle, mais sans spécifier les relations sémantiques en jeu. Un autre problème important de ces dictionnaires est qu'ils ne traitent pas seulement les collocations et les locutions au sens strict, mais aussi des syntagmes libres qui sont fréquemment associés aux lexies. Par exemple, dans le *Dictionnaire des collocations du mandarin moderne* (Mei, 1999), sous la lexie 爱 'aimer', il y a des exemples tels que 爱钱 'aimer de l'argent', 爱祖国 'aimer la patrie', 爱人民 'aimer le peuple', etc. Ce genre de syntagmes libres occupe une grande place dans ces dictionnaires, mais ils ne sont pas vraiment intéressants pour les études des collocations en mandarin. Ces dictionnaires ont largement contribué aux recherches sur les collocations en mandarin, mais, comme le note Chen (2006), leur couverture des collocations ne répond pas aux besoins réels et ils doivent encore être améliorés.

En revanche, en français et d'autres langues, plusieurs dictionnaires des collocations sont élaborés à l'aide du formalisme des fonctions lexicales [=FL] dans le cadre de la théorie Sens-Texte [=TST]. Pour énumérer quelques exemples :

- En français, il existe le DEC (*Dictionnaire explicatif et combinatoire*) (Melčuk et al. 1982, 1984, 1992, 1999), la base des données lexicales DiCo (*Dictionnaire de combinatoire*)³ (Jousse et Polguère, 2005; Polguère et Mel'čuk, 2006), DiCoEnviro (*Dictionnaire fondamental de l'environnement*)⁴ (L'Homme et Laneville, 2009), DiCoInfo (*Dictionnaire fondamental de l'informatique et de l'Internet*)⁵ (L'Homme, 2008, 2009), DiCoLiLex (*Dictionnaire de liens lexicaux*) (Marengo et

³ Voir : <http://olst.ling.umontreal.ca/dicouebe/>

⁴ Voir : <http://olst.ling.umontreal.ca/dicoenviro>

⁵ Voir : <http://olst.ling.umontreal.ca/dicoinfo>

Robichaud, 2017), RL-fr (*Réseau lexical du français*)⁶ (Lux-Pogodalla et Polguère, 2011);

- En anglais et en russe, il y a respectivement le RL-en (*Réseau lexical de l'anglais*) (Gader, Ollinger et Polguère, 2014) et RL-ru (*Réseau lexical du russe*) (Krylosova, 2017; Krylosova, Polguère, Chepiga et Baïandina-Soujaeff, 2014);
- En espagnol, il existe le DiCE (*Diccionario de Colocaciones del Español*)⁷ (Alonso-Ramos, 2004);
- En coréen, il y a le dictionnaire électronique *Sejong*, dont la partie des collocations sont encodés avec les FL (National Institute of Korean Language, 2006), etc.

Dans ces dictionnaires, les collocations dans ces diverses langues sont décrites d'une manière formelle et systématique, alors qu'en mandarin, il n'existe pas encore de dictionnaire de ce genre, à notre connaissance.

Ainsi, pour mieux décrire les collocations chinoises dans un dictionnaire, nous proposons d'étudier les collocations d'un groupe de lexies du mandarin et nous essayons d'établir un modèle de dictionnaire des collocations en mandarin basé sur le DiCoLiLex, afin de combler certaines lacunes dans les dictionnaires existants en mandarin. D'une part, cela peut servir de modèle pour un dictionnaire chinois pédagogique à l'intention des apprenants. D'autre part, il est possible d'intégrer ce dictionnaire au système GenDR (Lambrey et Lareau, 2015; Lareau et al., 2018), un réalisateur profond multilingue, pour faire de la génération automatique de texte (GAT) en chinois.

⁶ Voir : ORTOLANG (Open Resources and TOols for LANGuage), www.ortolang.fr/, <https://hdl.handle.net/11403/lexical-system-fr/v1>

⁷ Voir : www.dicesp.com/

1.2 Objectifs du mémoire

Notre recherche vise à élaborer un modèle de dictionnaire des collocations en mandarin constitué de lexies du champ sémantique des émotions, à partir d'un sous-ensemble de la base de données lexicales DiCoLiLex. Les buts de cette recherche sont les suivants :

- 1) Décrire les caractéristiques des collocations en mandarin;
- 2) Adapter la microstructure du DiCoLiLex aux caractéristiques du mandarin;
- 3) Rédiger un dictionnaire des collocations du champ sémantique des émotions en mandarin.

1.3 Organisation du mémoire

Le mémoire est séparé en cinq chapitres. Dans le deuxième chapitre, nous présentons d'abord le cadre théorique de notre recherche : la lexicologie explicative et combinatoire. Nous clarifions également les concepts des unités lexicologiques du mandarin dans ce cadre. Ensuite, nous présentons la méthodologie de recherche dans le troisième chapitre. À partir des entrées choisies en français, nous établissons une nomenclature correspondante en mandarin et nous utilisons un corpus pour en extraire les cooccurrents des lexies à l'étude afin de trouver les collocations. Nous comparons dans ce chapitre diverses méthodes d'extraction. C'est le quatrième chapitre qui constitue le cœur de ce mémoire. D'abord, nous adaptons la microstructure du dictionnaire dans le XML aux caractéristiques du mandarin afin de faciliter la description lexicologique des lexies dans la nomenclature chinoise. En nous inspirant de l'interface du DiCoLiLex, nous créons également une interface en ligne pour faciliter la consultation du lecteur (<http://olst.ling.umontreal.ca/xiaoyu>). Puis, nous discutons des cas difficiles rencontrés lors de la rédaction des articles. Nous montrons que certains phénomènes ne cadrent pas bien dans la théorie et proposons des pistes de solutions. Enfin, la conclusion présente une synthèse de ce mémoire et des applications possibles.

Chapitre 2 Cadre théorique

Notre recherche est basée sur la lexicologie explicative et combinatoire [= LEC] (Mel'čuk et al., 1995) dans le cadre de la théorie Sens-Texte [= TST]. Cette théorie considère la lexie comme unité de base et son objet central. Ainsi, il est important de définir le terme **lexie** et les autres termes lexicologiques afin de les distinguer avec d'autres concepts tels que **mot** qui est vague. Dans la deuxième section, nous essayons de clarifier les notions de morphe, mot-forme et syntagme en mandarin dans le cadre de la TST.

2.1 La lexicologie explicative et combinatoire

Dans notre cadre théorique, une **lexie** ou **unité lexicale**, est un regroupement de mots-formes [= un lexème] ou de syntagmes [= une locution] pris dans une seule acception bien spécifique (Mel'čuk et al., 1995).

Pour une explication plus cohérente, commençons par la définition du **mot-forme**, concept central de la morphologie :

Définition : Mot-forme

Un signe linguistique segmental s'appelle mot-forme si et seulement s'il satisfait simultanément aux deux conditions suivantes :

- 1) Par rapport à son environnement externe, il possède une certaine autonomie (≈ il est utilisable tout seul dans le discours).
- 2) Par rapport à son organisation interne, il possède une certaine cohésivité. (≈ ses éléments constituants sont très unis). (Mel'čuk et Milićević, 2014b, p. 16)

Le concept de **mot-forme** et celui de **lexème** sont liés de façon très importante. Pour les langues synthétiques comme le français, un mot-forme est une forme flexionnelle d'un lexème et inversement, un lexème est, grosso modo, un ensemble de mots-formes dont les différences se réduisent à la flexion (Mel'čuk et Milićević, 2014b, p. 33), par exemple : POMME = {*pomme, pommes*}; MANGER = {*manger, mange, mangeais, mangeraient, a mangé, mangée, mangeant...*}.

Définition : Lexème

Un lexème est un ensemble de mots-formes et de syntagmes de type particulier (=formes analytiques) dont les signifiés ne se distinguent que par des significations flexionnelles et dont les signifiants incluent le signifiant du même radical, qui exprime leur signification partagée. (Mel'čuk et Milićević, 2014a, p. 250)

Cependant, pour les langues isolantes ou analytiques, dont le mandarin, il n'y a pas de flexion véritable (surtout pas de flexion morphologique). Donc, les mots sont en général invariables, quelle que soit leur fonction syntaxique, chaque lexie ne possède qu'une forme. Cependant, Mel'čuk et Milićević (2014b, p. 208) mentionnent que les langues isolantes peuvent avoir des processus de formation de mots et/ou de quasi-flexion, « ainsi, le chinois mandarin affiche la formation des mots par la composition₁, la reduplication et, dans une moindre mesure, la suffixation ».

Quant à la **locution**, c'est une unité lexicale non compositionnelle dont les signifiants sont des syntagmes qui ont les propriétés des mots-formes. Par exemple, les syntagmes *pomme de terre* et *pommes de terre* sont des formes d'une même locution 'POMME DE TERRE'. Les locutions, ainsi que les collocations et les clichés sont des sous-catégories des phrasèmes. Nous verrons la définition de la locution dans la section suivante sur la typologie des phrasèmes (§2.3, p. 17).

Maintenant, nous pouvons donner la définition de la lexie :

Définition : Lexie

Une lexie est soit un lexème, soit une locution. (Mel'čuk et Milićević, 2014a, p. 251)

Dans le cadre de TST, un signe linguistique tel qu'un mot-forme est un triplet: signe linguistique = <signifié, signifiant, syntactique> (Mel'čuk et Milićević, 2014a, p. 33-34). Ainsi, une lexie est une entité à trois composantes, dont un sens (le *signifié* saussurien), une forme phonique/graphique (le *signifiant* saussurien) ainsi qu'un ensemble de traits de combinatoire (le *syntactique*). Par conséquent, l'étude lexicologique englobe la sémantique et la syntaxique. En

même temps, cette étude doit aboutir à l'élaboration d'un dictionnaire, la cible et le produit final de la lexicologie.

2.2 La lexicologie du mandarin moderne

Contrairement à la plupart des langues indo-européennes, le mandarin possède un système d'écriture qui tente de représenter le signifié plutôt que le signifiant. De plus, il n'y a pas d'espaces entre les caractères. Chaque graphème chinois (ou sinogramme⁸) représente, hormis de rares exceptions⁹, une seule syllabe. Une même syllabe chinoise peut se représenter

⁸ Tous les sinogrammes sont des logogrammes, mais en fonction de leur étymologie, nous pouvons les classer en quatre catégories : pictogrammes (cf. 日^{rì} 'soleil', 月^{yuè} 'lune', 木^{mù} 'arbre'), indicateurs ou idéopictogrammes (cf. 一^{yī} 'un', 二^{èr} 'deux', 三^{sān} 'trois'), idéogrammes proprement dits (cf. 林^{lín} 'bois' = deux arbres, 森^{sēn} 'forêt' = trois arbres) et idéo-phonogrammes (cf. 清^{qīng} 'clair', 晴^{qíng} 'ensoleillé', 请^{qǐng} 'inviter'). Ce système de classification est premièrement expliqué dans 《说文解字》 (*Explication des pictogrammes et des caractères composés*) de Xu Shen (57—147). C'est le premier dictionnaire de sinogrammes à proposer une analyse de leur composition et à les classer à l'aide des « clés », éléments graphiques pour indiquer de quoi parle le sinogramme. Les sinogrammes d'aujourd'hui ont une variation en forme simplifiée en usage en Chine continentale et à Singapour et une autre variation en forme traditionnelle en usage à Taïwan et à Hong Kong. Dans ce mémoire et dans notre dictionnaire rédigé, nous utilisons les sinogrammes simplifiés. (cf. https://fr.wikipedia.org/wiki/Caractères_chinois [consulté le 23/08/2018])

⁹ Il existe en mandarin un suffixe 儿^{er} qui ne se prononce pas comme une syllabe mais modifie la voyelle de la syllabe précédente par une rétroflexion. Ainsi la syllabe modifiée s'écrit en deux graphèmes, par exemple, [hwaʐ] s'écrit comme 花儿^{huā er} 'fleur'. Par contre, il existe aussi certains sinogrammes dissyllabiques inventés à l'époque moderne. Par exemple, au 19^e siècle, des nouveaux sinogrammes composés de deux graphèmes sont inventés pour traduire les unités de mesure du système métrique, par exemple, 尅^{qiānkè} 'kilogramme' est une combinaison de deux graphème 千^{qiān} 'mille' et 克^{kè} 'gramme' et garde la prononciation de ces deux graphèmes. Cependant, ces sinogrammes dissyllabiques sont rejetés par la Commission d'État pour la langue et l'écriture chinoises depuis 1977 et ainsi ils sont rarement utilisés aujourd'hui (cf. <https://zh.wikipedia.org/wiki/計量用漢字> [consulté le 19/08/2018]).

par différents graphèmes. Par exemple, il y a une soixantaine de sinogrammes homophoniques qui représentent la syllabe /ɛī : 夕 ‘soir’, 汐 ‘marée’, 西 ‘ouest’, etc. À l’inverse, certains graphèmes peuvent représenter plusieurs monosyllabes, par exemple, le graphème 和 peut avoir cinq prononciations (/hé/, /hè/, /huó/, /huò/ et /hú/) selon le sens : ce sont des homographes (B. Huang et Liao, 2002). En chinois classique, la plupart des lexèmes étaient monosyllabiques (K. Zhao, 1994), donc chaque caractère représentait un mot-forme, d’où la polysémie de 字^{zì} ‘caractère/mot’, qui dénote à la fois un graphème et une unité lexicale. Comme Norman (1988, p. 154) le mentionne : « En chinois classique comme en chinois moderne, la grande majorité des morphèmes coïncident sur le plan phonologique avec une monosyllabe; inversement, presque chaque syllabe peut être analysée comme un morphème indépendant. » [notre traduction].

Néanmoins, en mandarin moderne, les mots dissyllabiques et polysyllabiques sont de plus en plus utilisés, c’est-à-dire que les mots-formes peuvent contenir deux caractères ou plus. Selon les statistiques de Zhou (1999), les mots dissyllabiques occupent 67,5% des entrées du 《现代汉语词典》 (*Dictionnaire du mandarin moderne, 1996*). Étant donné que chaque sinogramme en isolation possède souvent au moins un sens, il est parfois difficile de distinguer un mot-forme d’un morphe ou d’un syntagme. Avant de commencer notre description lexicographique, il est donc important de bien définir les unités morphologiques et lexicologiques en mandarin moderne dans notre cadre théorique.

En mandarin comme en français, les mots-formes sont constitués de signes plus petits que nous appelons les signes morphologiques. Le morphe est un type de signe morphologique le plus important parce qu’il est universel et omniprésent (Mel’čuk et Milićević, 2014b, p. 36). C’est la plus petite unité phonétique significative.

Définition : Morphe

|| Un signe linguistique segmental élémentaire s’appelle morphe. (Mel’čuk et Milićević, 2014b, p. 36)

La plupart des morphes en mandarin moderne sont monosyllabiques et se représentent par un seul graphème. Ainsi, DeFrancis (1986) suggère le terme « morphosyllabique » pour décrire la correspondance étroite entre l'unité syllabique phonologique et le morphème en mandarin. Cependant, il existe certains morphes polysyllabiques que nous pouvons généralement séparer en trois groupes (Wang, Lu, Fu, Ma et Su, 1993):

- 1) Morphes dissyllabiques issus du chinois classique : ce sont des morphes composés de deux graphèmes qui n'ont pas de sens tout seuls et qui représentent des sons. Ainsi, dans les livres anciens, il y avait parfois différentes variantes des graphèmes. On utilise le terme ^{liánmián cí} 连绵词 en phonologie chinoise pour désigner ce type de morphes, dont les deux syllabes font souvent une rime ou partagent la même attaque, par exemple, ^{línglóng} 玲珑 'exquis', ^{xiāo yáo} 逍遥 'libre', ^{dīngníng} 叮咛 'recommander', ^{jié jué} 孑孓 'larve de moustique' ainsi qu'une partie du terme lui-même, ^{liánmián} 连绵 'continu'.
- 2) Morphes empruntés phonétiquement : ce sont des morphes d'origine étrangère et rendus en mandarin par une prononciation proche, donc les graphèmes qui composent ce type de morphe perdent leur sens propre, par exemple, ^{kā fēi} 咖啡 'café', ^{shā fā} 沙发 'sofa', ^{sān míng zhì} 三明治 'sandwich', ^{jiā ná dà} 加拿大 'Canada'.
- 3) Morphes onomatopéiques : ce sont des morphes imitant des sons. Comme les morphes empruntés phonétiquement, les graphèmes servent à transcrire le son¹⁰, par exemple, ^{wēngwēng} 嗡嗡 'bourdonner', ^{dīngdōng} 叮咚 'tinter', ^{pīngpāng} 乒乒 'pingpong', ^{jī jī zhā zhā} 叽叽喳喳 'jacassement'.

Le nombre de morphes polysyllabiques est relativement limité et ce sont généralement des morphes lexicaux qui peuvent constituer des mots-formes par eux-mêmes. Quant aux

¹⁰ Les sinogrammes qui représentent ce type de morphes sont souvent des idéophonogrammes, dont plusieurs possèdent une « clé » 口 'bouche', pour désigner les sons produits par la bouche.

morphes monosyllabiques, la situation est plus complexe. Selon leur autonomie, Lü (1999) les divise en quatre catégories :

- 1) Autonomes : ce sont des morphes qui peuvent constituer des mots-formes et s'utiliser tout seuls dans un énoncé. Par exemple, dans la phrase «^{wǒ kàn shū}我看书» 'je lis un livre' : ^{wǒ}我 'moi', ^{kàn}看 'regarder/lire', et ^{shū}书 'livre' sont trois mot-formes constitués par trois morphes autonomes.
- 2) Non autonomes : ces morphes ne peuvent pas s'utiliser seuls dans un énoncé. Ils doivent se combiner avec d'autres pour former des mots-formes, mais ils sont associés à un sens, hérité du chinois ancien. Par exemple, ^{shēng chù}牲 et ^{chù}畜 dans le mot-forme ^{shēng chù}牲畜 'bétail'¹¹, ou ^{lǎo}老 dans ^{lǎo shǔ}老鼠 'souris' ou ^{lǎo hǔ}老虎 'tigre'¹².
- 3) Quasi-autonomes : ces morphes sont non autonomes en général, mais ils peuvent s'utiliser seuls comme mots-formes dans certains contextes. Par exemple, ^{shè}摄 'photographier' dans le syntagme verbal ^{běn bào jì zhě shè}本报记者摄 'le journaliste de notre journal photographie' ou ^{jīng}京 'Beijing', ^{jīn}津 'Tianjin', ^{tóng}唐 'Tangshan' dans le syntagme nominal ^{jīng jīn táng dì qū}京、津、唐地区 'région de Beijing, Tianjin, Tangshan' sont autonomes; cependant, ils ne sont pas autonomes quand ils sont dans des mots-formes tels que ^{shè yǐng}摄影 'photograohier', ^{běi jīng}北京 'Beijing', ^{tiān jīn}天津 'Tianjin', ^{tóngshān}唐山 'Tangshan'.
- 4) Semi-autonomes : ces morphes ne peuvent pas s'utiliser seuls dans un énoncé, mais ils peuvent se combiner avec d'autres morphes, des mots-formes, ou des syntagmes. Par exemple, ^{de}的 dans ^{bié de}别的 'autre', ^{wǒ de}我的 'le mien', ^{wǒ rèn shi de}我认识的 'ce que je connais' et

¹¹ Ce sont deux morphes qui ont le même sens 'animal domestique', hérité du chinois classique. En fait, ils étaient autonomes en chinois classique, mais ils ont perdu l'autonomie aujourd'hui, et peuvent seulement s'utiliser en se combinant à d'autres morphes, par exemple, ^{shēngkǒu}牲口 'bête de somme' ou ^{gēngchù}耕畜 'bête de labour'.

¹² Le sens propre du morphe en question ^{lǎo}老 est 'vieux', mais ici il est grammaticalisé en devenant un préfixe et son sens devient vide, ainsi, pour exprimer le sens d'un vieux tigre, on dit ^{lǎo lǎo hǔ}老老虎 'vieux tigre' dont le premier ^{lǎo}老 est autonome et le deuxième non autonome.

les marqueurs d'aspect 了^{le} 'accompli', 着^{zhe} 'non-accompli', etc. Selon Lü (1999), la plupart des mots-outils appartiennent à ce type.

Les deux premières catégories cadrent bien dans la TST. Les morphes autonomes sont tous des radicaux. Quant aux morphes non autonomes, Huang et Liao (2002) les appellent « morphes agglutinants », et les séparent selon que leur position est fixe ou non dans les mots-formes : ceux dont la position est fixe sont des affixes; ceux dont la position est libre sont des radicaux. Le troisième type ressemble beaucoup à un morphe autonome, mais avec un usage plutôt contraint, souvent à l'écrit. Nous pouvons décrire ces contraintes dans le syntactique du mot-forme constitué par ce type de morphe, et nous en discuterons plus en détail quand nous parlerons des « mots élastiques » (§4.4.1, p. 53). Les morphes de la quatrième catégorie ressemblent aux mots-formes avec une autonomie faible en français, tels les clitiques¹³. Nguyen (2006) propose plusieurs critères pour distinguer les statuts morphologiques de ces morphes, dont nous reparlerons de ces critères plus loin dans cette section.

Les morphes peuvent former une unité plus large : le mot-forme. Nguyen (2006) propose une définition du mot-forme chinois à partir de celle de Mel'čuk (1993) :

Définition du mot-forme chinois

Un signe segmental **X** du chinois est appelé mot-forme de la langue du chinois si et seulement si **X** est suffisamment autonome en chinois et **X** n'est pas représentable en termes d'au moins deux mots-formes du chinois (Nguyen, 2006, p. 50).

Les morphes monosyllabiques autonomes peuvent constituer un mot-forme à eux seuls, mais comme nous l'avons mentionné plus haut, les mots-formes dissyllabiques sont plus nombreux en mandarin moderne. Un mot-forme dissyllabique peut être composé de deux morphes autonomes, par exemple, 牛肉^{niú ròu} 'bœuf' (litt. 'bœuf-viande'), 生长^{shēngzhǎng} 'croître' (litt.

¹³ Mel'čuk et Milićević (2014b, p. 30) les appellent aussi les mots-formes primaires défectifs, qui se distinguent des mots-formes primaires parfaits et des mots-formes secondaires.

‘croître-grandir’); d’un morphe autonome et un morphe non autonome, par exemple, 人民^{rén mǐn} ‘peuple’ (litt. ‘gens-peuple’), 制造^{zhì zào} ‘fabriquer’ (litt. ‘fabriquer-faire’); ou de deux morphes non autonomes, par exemple, 植物^{zhí wù} ‘plante’ (litt. ‘plante-objet’) ou 研究^{yán jiū} ‘rechercher’ (litt. ‘étudier-rechercher’). De plus, plusieurs linguistes chinois parlent d’une tendance dissyllabique des mots-formes pour des raisons prosodiques (Dong, 2002; Duanmu, 2012; Huang et Liao, 2002; Lü, 1999). En conséquence, beaucoup de morphes autonomes en chinois classique deviennent des morphes non autonomes (B. Huang et Liao, 2002) de sorte qu’il faut leur ajouter un autre morphe pour former un mot-forme (dissyllabique). Par exemple, on doit ajouter le morphe autonome 国^{guó} ‘pays’ après les morphes monosyllabiques qui désignent certains pays, tels que 法^{fǎ} 国^{guó} ‘France’ ou 英^{yīng} 国^{guó} ‘Angleterre’, tandis que les morphes dissyllabiques peuvent constituer un mot-forme en soi, comme 日^{rì} 本^{běn} ‘Japon’ ou 印^{yìn} 度^{dù} ‘Inde’.

Un syntagme est, *grosso modo*, « une suite de mots-formes syntaxiquement liés dotée d’un contour prosodique » (Mel’čuk et Milićević, 2014a, p. 250). Comme en français, il existe en mandarin des syntagmes libres et contraints (c’est-à-dire des phrasèmes).

Pour les signes complexes en mandarin tels que ΨX ou $X\Psi$, où X est un signe segmental qui pourrait être soit un morphe soit un mot-forme et Ψ est un signe segmental autonome (ou une chaîne de tels signes), Nguyen (2006) utilise trois critères universels (séparabilité du signe; variabilité distributionnelle; transmutabilité) proposés par Mel’čuk (1993) pour distinguer les syntagmes évidents, dont les constituants possèdent une autonomie forte, et les mots-formes complexes évidents, dont les constituants ne possèdent aucune autonomie, ainsi qu’isoler les signes complexes « suspects », dont les constituants ne possèdent qu’une autonomie faible. Ces trois critères s’appliquent dans l’ordre pour vérifier l’autonomie du signe segmental X . Le critère de séparabilité est le plus important : si les constituants d’un signe complexe peuvent être séparés par un mot-forme évident, alors il s’agit d’un syntagme. Par exemple, dans les signes complexes 睡^{shuì} 觉^{jiào} ‘dormir’, 觉^{jiào} ‘acte.de.dormir’ peut être soit un mot-forme soit un morphe, mais puisqu’un adjectif 懒^{lǎn} ‘paresseux’ peut être inséré entre ces deux signes pour modifier 觉^{jiào} ‘acte.de.dormir’ et former ainsi l’expression 睡^{shuì} 懒^{lǎn} 觉^{jiào} ‘faire la grasse matinée’, nous constatons que 睡^{shuì} 觉^{jiào} ‘dormir’ est bien un syntagme. Si les constituants ne sont pas séparables, il faut vérifier

les deux autres critères : **X** a une autonomie faible si **X** est distributionnellement variable, ou s'il ne l'est pas mais est transmutable, nous avons affaire à un signe complexe « suspect ». Par exemple, les signes complexes de la forme « N?CL »¹⁴ constitués d'un nom et son classificateur pour exprimer un sens générique du nom, tels que 书本^{shū běn} 'livres en général', 船只^{chuán zhī} 'bateaux en général', ne sont pas séparables et les classificateurs ne sont pas distributionnellement variables mais ils sont transmutables : 一本书^{yī běn shū} 'un CL livre', 一只船^{yī zhī chuán} 'un CL bateau'. Il s'agit donc d'un cas de signe complexe « suspect », et il faut s'appuyer sur d'autres critères afin de déterminer son statut morphologique. En revanche, si le signe complexe ne remplit aucun de ces trois critères, il s'agit d'un mot-forme. Ainsi, à l'aide de ces critères, Nguyen (2006) montre que le signe 子^{zǐ} dans les signes complexes tels que 妻子^{qī zǐ} 'épouse' est un suffixe dérivationnel qui fait partie du mot-forme; 的^{de} (DE₁), le marqueur de modification du nom et 地^{dì} (DE₂), marqueur de modification du verbe, sont des clitiques; les marqueurs de phrase comme 了^{le} (LE₂) 'marqueur de changement de situation', 吗^{ma} 'est-ce que' et 吧^{ba} 'marqueur d'ordre' sont des clitiques.

Nguyen (2006) considère ces trois critères comme critères de niveau I, et il propose ensuite six critères de niveau II (cinq critères du mot-forme composé et un critère de l'existence de règles syntaxiques de surface) qui visent à démontrer l'autonomie ou la non-autonomie des constituants d'un signe complexe « suspect ». Ainsi, si un signe complexe « suspect » semble être un mot-forme, on lui appliquera les critères suivants de Nguyen :

- 1) Critère de non-productivité : un signe complexe **S₁?S₂** ne peut pas être un syntagme s'il n'est pas possible de produire librement d'autres signes de la forme « **S₁?S₂** » dans une langue L. Par exemple, ce ne sont pas tous les noms avec leurs classificateurs qui peuvent former les signes complexes « N?CL » pour exprimer un sens générique : si on peut dire 一支笔^{yī zhī bǐ} 'un CL stylo', 一条船^{yī tiáo chuán} 'un CL bateau', il

¹⁴ Nous utilisons la notation « X?Y » pour marquer l'incertitude du statut morphologique des signes en question (syntagme ou mot-forme).

n'existe pas *^{bǐ zhī}笔支 'stylo en général' ou *^{chuántiáo}船条 'bateau en général'. Les signes complexes de la forme « N?CL » sont ainsi des mots-formes composés.

- 2) Critère d'exocentricité : un signe complexe exocentrique¹⁵ est un mot-forme s'il n'existe pas de règles pour produire d'autres signes exocentriques de la forme « S₁?S₂ » dans une langue L. Par exemple, ^{sī xiǎng}思想 'pensée (litt. réfléchir-penser)' et ^{zhī shì}知识 'connaissance (litt. savoir-reconnaître)' sont des signes de la forme « V₁?V₂ », mais sémantiquement ils ne sont pas compositionnels et ce sont des signes nominaux plutôt que verbaux. Ce sont ainsi des mots-formes composés phraséologisés.
- 3) Perte de la capacité des constituants à admettre des modificateurs : les constituants d'un mot-forme composé ont perdu leur capacité à prendre des modificateurs. En revanche, si les constituants gardent leur capacité à prendre des modificateurs, alors le signe complexe en question est un syntagme. Par exemple, le constituant ^{bái}白 'être blanc' dans le signe complexe ^{bái xuě}白雪 'neige blanche' ne peut plus admettre le modificateur ^{hěn}很 'très', cf. *^{hěn bái xuě}很白雪 'neige très blanche', tandis qu'il le peut quand il est autonome, cf. ^{hěn bái de xuě}很白的雪 'neige très blanche (litt. très blanche DE₁ neige)' ou ^{xuě hěn bái}雪很白 'la neige est très blanche'. Ainsi ^{bái xuě}白雪 'neige blanche' est un mot-forme composé.
- 4) Critère de non-réduction conjonctive : un signe complexe est un mot-forme s'il n'est pas possible de factoriser un des constituants dans un syntagme coordonné. Par exemple, nous ne pouvons pas factoriser le mot-forme ^{zhé xué}哲学 'philosophie' dans un syntagme coordonné comme dans ^{zhé xué jiā hé zhé xué shēng}哲学家和哲学学生 'spécialistes de philosophie et étudiants de philosophie', car le syntagme *^{zhé xué jiā hé shēng}哲学家和生 'spécialistes et étudiants de philosophie' est agrammatical, donc ^{zhé xué jiā}哲学家 'spécialistes de philosophie' et ^{zhé xué shēng}哲学学生 'étudiants de philosophie' sont des mots-formes composés.

¹⁵ Un mot-forme composé est dit exocentrique si aucune de ses parties n'est la tête du mot-forme, par exemple, *sans-abri*, *brise-glace*, *rouge-gorge* (Cf. Bloomfield, 1933).

- 5) Critère syntactico-phonologique : si la pause prosodique et l'insertion d'une conjonction de coordination sont impossibles entre les constituants d'un signe complexe qui ressemble à un syntagme coordonné, alors ce signe est un mot-forme. Par exemple, ^{bēi shāng} 悲伤 'être triste (litt. être.triste-être.affligé)' et ^{qí guài} 奇怪 'être bizarre (litt. être.bizarre-être.bizarre)' peuvent ressembler à des syntagmes coordonnés sans marqueur de coordination, mais ce sont en fait des mots-formes, car il n'est pas possible de faire une pause ou d'insérer une conjonction de coordination entre les constituants.

Si le signe complexe en question satisfait au moins un de ces critères, alors S est un mot-forme. Sauf le troisième critère, si le signe complexe ne satisfait pas les quatre autres critères, cela ne peut pas prouver qu'il est un syntagme. On lui appliquera ensuite le critère de l'existence de règles syntaxiques de surface pour traiter la SSyntS de l'expression en question. Ce critère démontre qu'un signe complexe $S_1?S_2$ est un syntagme s'il est possible de le représenter sous forme d'un arbre syntaxique de surface régulier et de créer de manière libre et productive d'autres syntagmes de la forme « $S_1 S_2$ », sinon, il s'agit d'un mot-forme.

Nous utiliserons ces critères pour déterminer le statut morphologique des signes complexes dans notre nomenclature, et nous en reparlerons plus loin pour les cas spéciaux quand nous ferons la description lexicographique dans le deuxième chapitre.

2.3 Typologie des phrasèmes

Les collocations des lexies chinoises du champ sémantique des émotions sont notre objet d'étude principal. Nous avons dit que les collocations et les locutions (§2.1, p. 6) sont des sous-catégories de phrasèmes lexicaux dans le cadre de la TST. Avant de donner une définition de la collocation, il est important de parler de cette typologie des phrasèmes.

Selon Mel'čuk (2015), il existe trois grandes familles de phrasèmes dans les langues naturelles : les phrasèmes lexicaux, les phrasèmes morphologiques et les phrasèmes

syntaxiques. Notre recherche se concentre sur les phrasèmes lexicaux. Contrairement aux syntagmes libres, un phrasème lexical est un syntagme qui est contraint paradigmatiquement et syntagmatiquement, c'est-à-dire que le locuteur ne peut pas construire un phrasème lexical en choisissant chacun de ses composants lexémiques indépendamment et seulement selon les règles générales de la langue.

Selon les différentes contraintes sur ces syntagmes figés, Mel'čuk (2013) a proposé deux axes pour classer les phrasèmes : l'axe paradigmatique concerne les contraintes de sélection lexicale ou sémantico-lexicale de leurs composantes, tandis que l'axe syntagmatique implique les contraintes de combinaison de leurs composantes ou leur compositionnalité.

L'axe paradigmatique distingue les phrasèmes lexémiques des phrasèmes sémantico-lexémiques. Un phrasème lexémique possède un sens construit librement, mais le choix des composantes pour exprimer ce sens est contraint, par exemple, *porter son attention sur, au bout du rouleau*. Un phrasème sémantico-lexémique non seulement porte un sens qui n'est pas construit librement par le locuteur, mais en plus le choix de ses composantes est contraint, par exemple, *Vous dites ?, sauf imprévu*.

L'axe syntagmatique permet de distinguer les phrasèmes sémantiquement compositionnels et non compositionnels. La notion de compositionnalité est basée sur l'opération d'union linguistique, représentée par le symbole « \oplus » : elle réunit les signes et leurs composantes selon leur nature et leurs propriétés suivant les règles générales de la langue donnée (Mel'čuk, 2013). Ainsi, le sens d'un phrasème sémantiquement compositionnel est la composition des sens de ces composantes, par exemple, 'peinture fraîche' = 'peinture' \oplus 'frais', alors que le sens d'un phrasème non compositionnel n'est pas égal à l'union linguistique de ses composantes, par exemple, 'pomme de terre' \neq 'pomme' \oplus 'de' \oplus 'terre'.

Selon ces deux axes, Mel'čuk distingue quatre classes majeures de phrasèmes :

	Non compositionnel	Compositionnel
Contraintes lexémiques	locutions	collocations
Contraintes sémantico-lexémiques	nominèmes	clichés

Tableau I. Typologie des phrasèmes (Mel'čuk, 2015)

Dans ce mémoire, nous nous intéressons surtout aux phrasèmes lexémiques. À partir du tableau ci-haut, nous pouvons maintenant donner les définitions de locution et de collocation :

Définition : Locution

|| Une locution est un phrasème lexémique qui n'est pas compositionnel.

Définition : Collocation

|| Une collocation est un phrasème lexémique qui est compositionnel.

Les locutions peuvent être séparées en trois sous-types selon leur degré de transparence/opacité, c'est-à-dire, en fonction de l'inclusion du sens des composantes : les locutions fortes, les semi-locutions et les locutions faibles (Mel'čuk, 2013). Pour distinguer ces trois sous-types de locutions ainsi que les collocations, il faut d'abord préciser le concept de **pivot sémantique** proposé par Mel'čuk (2008) :

Définition : pivot sémantique

|| Le pivot sémantique d'un sens complexe 'X', composé de 'X₁' et 'X₂', est sa partie 'X₁' ssi 'X'='X₂ (X₁)'.

Autrement dit, si 'X₂' est un prédicat et 'X₁' est son argument, c'est 'X₁' qui est le pivot sémantique. Par exemple, dans 'café noir', 'café' est le pivot sémantique parce qu'il est un argument de 'noir.' Dans cet exemple, le pivot sémantique 'café' coïncide avec le nœud dominant, mais Mel'čuk (2008, p. 6) note que « le pivot sémantique d'un sens est en principe différent de l'élément communicativement dominant¹⁶ de ce sens (même si assez souvent les deux coïncident). Par exemple, dans 'réussir un examen', le nœud dominant est 'réussir', mais

¹⁶ La composante communicativement dominante 'X₁' du sens 'X' représente la paraphrase minimale de 'X' : 'X' peut être réduit à 'X₁' sans distorsion majeure de l'information, même si une partie d'information sera bien entendu perdue. Autrement dit, 'X₁' est la composante générique de 'X' (Mel'čuk, 2001)

le pivot sémantique est ‘examen’». Avec l’aide du pivot sémantique, nous pouvons maintenant définir les trois sous-types de locutions :

- 1) **Locutions fortes** ou complètes : ce type de locution est complètement non compositionnel, c’est-à-dire que le sens d’une locution forte n’inclut aucun des sens de ses composantes. Par exemple, le sens de l’expression ‘MANGER LA GRENOUILLE’ ‘s’approprier frauduleusement les fonds déposés’ n’inclut ni ‘manger’, ni ‘grenouille’; de même, le sens de l’expression chinoise ^{chī cù} 吃醋 ‘être jaloux (litt. manger vinaigre)’ n’inclut ni ‘manger’, ni ‘vinaigre’.
- 2) **Semi-locutions** : contrairement à la locution forte, la semi-locution inclut dans son sens celui d’une seule de ses composantes, tout en incluant encore un sens additionnel comme pivot sémantique. Par exemple, la locution ‘LOUP DE MER’ ‘homme dont la profession est de naviguer sur la mer et qui est très expérimenté’ inclut le sens ‘mer’, mais non le sens ‘loup’, tout en ajoutant le sens ‘homme’ en tant que pivot sémantique (le sémantème ‘homme’ est l’argument de ‘dont la profession est de naviguer sur mer’ et ‘être très expérimenté’).
- 3) **Locutions faibles** ou quasi-locutions : une locution faible inclut tous les sens de ses composantes, et y rajoute un sens qui sert de pivot sémantique. Par exemple, ‘ROUGE À LÈVRES’ ‘produit cosmétique (souvent rouge) pour les lèvres’ ou son équivalent en chinois ^{kǒu hóng} 口红 (litt. ‘bouche rouge’) inclut le sens de ses composantes ‘rouge’ et ‘lèvre’ (ou ‘bouche’), mais y ajoute le sens ‘produit’ comme pivot sémantique.

Les locutions faibles sont très proches des collocations parce que ces deux types sont compositionnels, et il existe certains cas où nous ne pouvons pas les distinguer avec certitude (§4.4.3). Cependant, il y a une différence importante quant à la propriété de sélection non contrainte (Mel’čuk, 2008). Dans une locution, la sélection de ses composantes est complètement contrainte, tandis que dans une collocation, la sélection est semi-contrainte : une de ses composantes est sélectionnée librement par le locuteur pour son sens (ce sens sert de pivot sémantique dans le sens complet), alors que l’autre composante est choisie en fonction de

la première composante et du sens à exprimer. Par exemple, dans l'expression *amour fou* 'amour intense', le sens de la première composante 'amour' est inclus comme pivot sémantique, alors que la deuxième composante *fou* est sélectionnée auprès de cette première pour exprimer un sens d'intensification. La première composante s'appelle la base de la collocation, et l'autre est le collocatif. Dans d'autres exemples avec un collocatif qui exprime le sens d'intensification, tels que *faim de loup*, *peur bleue*, *amour fou*, *pleurer comme une Madeleine*, nous pouvons clairement voir que le choix du collocatif est fait en fonction de la base (ombrée). Il est à noter que, en général, le collocatif n'a son sens qu'au sein de la collocation (par exemple, *bleue* porte le sens 'intense' au lieu de 'couleur' dans *peur bleue*), mais il reste que la collocation est sémantiquement compositionnelle.

En résumé, pour distinguer le type d'un syntagme *AB* 'AB' en fonction de la compositionnalité du sens :

- 1) *AB* est un syntagme libre, ssi 'AB'='A' ⊕ 'B';
- 2) *AB* est une locution forte, ssi 'AB'='C';
- 3) *AB* est une semi-locution, ssi 'AB'='A' ⊕ 'C' et 'C' est le pivot sémantique;
- 4) *AB* est une locution faible, ssi 'AB'='A' ⊕ 'B' ⊕ 'C' et 'C' est le pivot sémantique;
- 5) *AB* est une collocation, ssi 'AB'='A' ⊕ 'B', mais *B* est choisi en fonction de *A* et 'A' est le pivot sémantique.

2.4 Les fonctions lexicales

Comme nous avons discuté dans la section précédente, dans une collocation, le collocatif est sélectionné en fonction de la base, et il existe une certaine relation entre la base et le collocatif. Afin de modéliser formellement les collocations par les relations entre les bases et leurs collocatifs, le concept de **fonction lexicale** [=FL] a été proposé par Žolkovskij et Mel'čuk (1967) et a été discuté à plusieurs reprises dans la littérature (Mel'čuk, 1996, 1998, 2003; Mel'čuk et al., 1995). Le formalisme des fonctions lexicales permet de décrire les collocations

ainsi que les dérivations sémantiques de façon générale, homogène et systématique (Mel'čuk et Milićević, 2014a, p. 210).

Définition : Fonction lexicale

Une fonction lexicale **f** décrit la relation sémantico-lexicale entre la base d'une collocation et son collocatif : $f(\text{base}) = \text{collocatif}$ (Mel'čuk, 2013)¹⁷.

Du point de vue formel, une FL est une fonction au sens mathématique qui peut être représentée par la formule traditionnelle : $f(x)=y$, où x est l'argument de la fonction et y , sa valeur (Mel'čuk et al., 1995, p. 126). Les fonctions sont appelées *lexicales* parce qu'elles prennent seulement des lexies comme argument, et la valeur est un ensemble de lexies. L'argument d'une FL est souvent appelé son **mot-clé** pour éviter la confusion potentielle avec le terme *argument* (Mel'čuk et al., 1995, p. 126).

Ainsi, les collocations sont consignées sous forme de triplet: le mot-clé (=la base), la fonction lexicale en jeu et la valeur retournée par la fonction (=le collocatif). Par exemple, dans les expressions *peur bleue* et ^{shēnshēn de kǒng jù} 深深的恐惧 'profonde peur', la fonction **f** dénote une 'intensification'. Cette fonction s'appelle **Magn**. Ainsi, $x = \text{PEUR}$, ^{kǒng jù} 恐惧, et $y = \text{BLEU}$, ^{shēnshēn de} 深深的^{de}, donc nous avons :

$$\mathbf{Magn}(\text{PEUR}) = \text{BLEU} \qquad \mathbf{Magn}(\sup{kǒng jù} \text{ 恐惧}) = \sup{shēnshēn de} \text{ 深深的}$$

Nous pouvons classifier les FL en fonction des différents axes : fonctions standards/non standards, paradigmatiques/syntagmatiques, et simples/complexes/en configuration (Jousse, 2010; Mel'čuk et al., 1995).

¹⁷ Il est à noter que la définition proposée ici est réduite aux collocations. Les FL peuvent représenter d'autres relations, telles que les dérivés sémantiques, cf. §2.4.2, p. 23.

2.4.1 Axe standard/non standard

La classification des FL selon leur caractère standard ou non a été discutée dans plusieurs ouvrages (Jousse, 2010; Mel'čuk, 1996; Mel'čuk et al., 1995; Polguère, 2007). Les FL standards doivent répondre aux principes de cardinalité, de diversité et d'universalité. Mel'čuk et al. (1995, p. 127-128) donne quatre conditions que les FL standards doivent satisfaire:

- 1) Pour toute paire de lexies L_1 et L_2 , les lexies $f(L_1)$ et $f(L_2)$ montrent des relations sémantico-syntaxiques (presque) identiques à ces lexies.
- 2) En règle générale, $f(L_1)$ et $f(L_2)$ sont différents : $f(L_1) \neq f(L_2)$.
- 3) La fonction f a un nombre élevé d'arguments (= de mots-clés). [En d'autres mots, le sens 'f' est très abstrait et très général et s'applique à beaucoup d'autres sens].
- 4) La fonction f a un nombre élevé d'éléments dans sa valeur (= d'expressions).

Ainsi, la fonction **Magn** est standard. En revanche, une FL non standard a peu d'arguments et peu d'éléments de valeurs différents. Les FL non standard ont souvent un sens très spécifique et non généralisable et elles sont surtout typiques pour des mots concrets. Par exemple, *café noir* dénote un café sans ajout de produit laitier ni de sucre. Dans cette expression, le sens de *noir* est 'sans ajout de produit laitier ni de sucre', qui est très spécifique et ne peut pas s'appliquer à beaucoup d'autres mots-clés. Il est impossible de prévoir le nombre de FL non standards dans une langue donnée; elles doivent être établies empiriquement pour chaque langue (Mel'čuk et Milićević, 2014a, p. 221).

Jousse (2010) a une division plus fine en distinguant les FL semi-standards et les FL localement standards. Les FL semi-standards ajoutent une composante de sens qui n'est pas prise en compte dans les FL standards dont elles sont dérivées, par exemple, **complètement AntiReal₁(armoire) = vider** [ART ~ de N=Y]; **en échange de qqc. CausOper₁(silence) = acheter** [le ~ de qqn.]. Les FL localement standards sont proposées par Polguère (2007) pour les relations dont la nature universelle n'est pas démontrée, mais nous pouvons les encoder par des FL standards dans le contexte d'une langue donnée, par exemple, **De_nouveauFunc₀**

(*hostilité*) = *repandre*; **Essayer_deOper₂**(*augmentation*) = *réclamer* [ART ~]. Les FL localement standards et semi-standards sont des hybrides entre les FL standards et non standards.

2.4.2 Axe paradigmatique/syntagmatique

Les FL standards peuvent être séparées en deux groupes selon les relations sémantico-lexicales paradigmatiques ou syntagmatiques qu'elles modélisent (Alonso-Ramos et Tutin, 1996; Mel'čuk, 1996; Mel'čuk et al., 1995).

Les FL paradigmatiques expriment des relations de dérivation sémantique et décrivent des corrélats lexicaux paradigmatiques pour chaque lexie, tels que les (quasi-)synonymes (**Syn**, **QSyn**), les (quasi-)antonymes (**Anti**, **QAnti**), les (quasi-)conversifs (**Conv_{ijk}**) et les (quasi-)dérivés syntaxiques (**S₀**, **V₀**, **A₀**, **Adv₀**) ou sémantiques (**S_i**, **A_i**, **Adv_i**, **Able_i**, **Qual_i**). Ces FL ont trait à la substitution : normalement, un élément de la valeur d'une FL paradigmatique *remplace* le mot-clé dans le texte (Mel'čuk et Milićević, 2014a).

À l'inverse, les FL syntagmatiques renvoient à la combinatoire des lexies, et décrivent leurs corrélats lexicaux syntagmatiques, c'est-à-dire, les collocations. Ces FL permettent de décrire, en particulier, les modificateurs adjectivaux ou adverbiaux (**Magn**, **Ver**, **Bon**), les verbes supports (**Func_i**, **Oper_i**, **Labor_{ij}**), les verbes de réalisation (**Fact_i**, **Real_i**, **Labreal_{ij}**), les verbes phasiques (**Incep**, **Fin**, **Cont**) et les verbes causatifs (**Caus**, **Liqu**, **Perm**). Ces FL ont trait à la combinaison : normalement, un élément de la valeur d'une FL syntagmatique apparaît dans le texte *avec* le mot-clé (Mel'čuk et Milićević, 2014a).

Cependant, il existe certaines exceptions au caractère de substitution ou de combinaison des FL. Alonso-Ramos et Tutin (1996) ont remarqué que les **Figur**, **Gener** ou **A_i** sont traditionnellement classées parmi les FL paradigmatiques, mais leur valeur peut se combiner avec le mot-clé dans certains cas, par exemple **Gener**(*honte*) = *sentiment de ~*; **Figur**(*haine*) = *feu de ~*; **A₁**(*angoisse*) = *en proie à ~*.

En revanche, il existe aussi un phénomène de fusion où les FL normalement syntagmatiques deviennent paradigmatiques. Par exemple, dans **Magn**(*pluie*) = *//déluge*, la

valeur de **Magn** pour le mot-clé *pluie* renferme une expression lexicale *déluge* qui ne comprend pas formellement le mot-clé *pluie*, mais comprend à la fois le sens du mot-clé et celui de **Magn**, de telle sorte que ‘déluge’ = ‘pluie + **Magn**’ (Mel’čuk et al., 1995, p. 151). Un tel élément est appelé élément **fusionné**, et il est indiqué par le symbole « // » dans le DEC.

2.4.3 Axe simple/complexe/configuration

Les FL peuvent aussi être classées en fonction de leur décomposabilité. Une FL standard simple ne peut pas être représentée à l’aide d’autres FL (Mel’čuk, 1996), donc la relation qu’elle décrit est en quelque sorte non décomposable. Kahane et Polguère (2001) indiquent que l’encodage des FL dans le DEC est basé sur une soixante de relations « primitives » et certaines règles pour ensuite les combiner. Ce sont les FL simples.

Les FL complexes représentent une manière de combiner les FL simples. Une FL complexe est un enchaînement de FL simples syntaxiquement liées : cet enchaînement a une valeur globale cumulative et exprime le sens de l’enchaînement entier de façon indécomposable auprès du mot-clé (Mel’čuk et al., 1995, p. 148). Une FL complexe s’écrit sous forme d’une chaîne de FL simples constituantes, soit **fg(L)**, par exemple, **AntiMagn** (*angoisse*) = *légère*; **IncepOper₁**(*amoureux*) = *tomber* [~]. Mel’čuk et al. (1995) indiquent qu’une FL complexe n’est pas une composition de FL, soit **fg(L) ≠ f(g(L))** : si **IncepOper₁**(*amoureux*) = *tomber* [~] et **Oper₁**(*amoureux*) = *être* [~], **Incep**(*être*) ≠ *tomber*; ainsi, **IncepOper₁**(*amoureux*) ≠ **Incep**(**Oper₁**(*amoureux*)).

Contrairement aux fonctions complexes, la configuration de fonctions lexicales est une suite de FL (simples ou complexes) qui ne sont pas syntaxiquement liées entre elles, mais qui partagent le même argument : cette suite a une valeur cumulative et exprime de façon indécomposable le sens de la suite entière (Mel’čuk et al., 1995, p. 149). Les constituants d’une configuration de FL sont réunis par le symbole « + », et ils peuvent être des FL simples ou complexes, standards ou non standards (Mel’čuk et Milićević, 2014a). Par exemple, **Bon+Magn**(*joie*) = *paradisique* (*une joie paradisiaque* ≈ ‘une grande [= **Magn**] joie qui est

bonne [= **Bon**]); **trop.Magn^{quant}** + **IncepOper₁**(*marché*) = *inonder* [ART ~] (*inonder le marché* ≈ ‘commencer à se retrouver [= **IncepOper₁**] en trop grande quantité [= **trop.Magn^{quant}**] sur le marché’).

Dans le présent mémoire, notre recherche portera principalement sur les FL standards en mandarin, mais nous y ajouterons les autres types de FL le cas échéant.

2.5 DEC et DiCoLiLex

La lexicologie explicative et combinatoire tente de créer un dictionnaire idéalisé de la langue : un modèle de dictionnaire explicatif et combinatoire (dorénavant, DEC). Quatre volumes du dictionnaire français réalisés selon ce modèle sont déjà publiés (*Dictionnaire Explicatif et Combinatoire du Français Contemporain*, Melčuk et al. 1982, 1984, 1992, 1999). Un DEC a certaines particularités comparativement à d’autres dictionnaires (Mel’čuk, 2003, p. 29-30) :

- 1) Un DEC est un lexique théorique : ce n’est pas un dictionnaire commercial ou grand public; il est élaboré dans le cadre de TST, qui met l’accent sur le lexique et est particulièrement adaptée à la construction d’un dictionnaire; ses concepts de représentation sémantique, d’actant sémantique et syntaxique profond, et de régime sont utilisés de façon centrale dans le DEC.
- 2) Un DEC est un lexique actif : c’est un dictionnaire de synthèse plutôt que d’analyse, c’est-à-dire qu’il est orienté vers la production de texte.
- 3) Un DEC est un lexique sémantique : toutes les lexies vedettes sont munies de descriptions sémantiques élaborées en fonction de normes rigides, ce qui est naturel pour l’approche « sens ⇒ texte », parce que nous devons toujours partir d’un sens donné bien décrit.

- 4) Un DEC est un lexique combinatoire : il faut présenter toute la cooccurrence lexicale restreinte pour toute lexie vedette, c'est-à-dire ses collocations. Idéalement, nous trouvons tous les collocatifs d'une lexie sous son entrée.

De plus, la rédaction du DEC suit quatre principes rédactionnels : principe de formalité, principe de cohérence, principe d'uniformité et principe d'exhaustivité (Mel'čuk et al., 1995, p. 34).

Le **principe de formalité** demande une description strictement formelle dans le DEC. D'une part, un métalangage formel préétabli suffisamment expressif est utilisé pour toute description lexicologique : la définition analytique (par genre prochain et différences spécifiques) pour la description sémantique; le régime pour la description syntaxique ainsi que les fonctions lexicales pour la description de la combinatoire lexicale restreinte (=collocations). D'autre part, toute description lexicographique est explicite et rien n'est laissé à l'intuition de l'utilisateur.

Le **principe de cohérence** exige que, au niveau de la microstructure, les descriptions sémantique, syntaxique et de cooccurrence d'une lexie montrent un accord complet, c'est-à-dire une correspondance réciproque entre les composantes sémantiques dans la définition de la lexie, ses dépendants syntaxiques actanciels et ses cooccurrents lexicaux.

Le **principe d'uniformité** exige la cohérence à la macrostructure de dictionnaire, et affirme que les articles de dictionnaires des lexies d'un même champ sémantique, c'est-à-dire les lexies apparentées, doivent montrer un accord aussi complet que possible.

Le **principe d'exhaustivité** met l'accent sur une exhaustivité « interne » en vue d'inclure toutes les informations lexicologiques nécessaires dans l'article d'une lexie, y compris les références à toutes les lexies apparentées.

Selon Mel'čuk (2003), la meilleure version du DEC serait sur support informatique. Ainsi, sur la base du DEC, le DiCoLiLex (abréviation pour le Dictionnaire de liens lexicaux) est une ressource informatisée actuellement développée par l'Observatoire de linguistique Sens-Texte. Cette ressource englobe deux bases de données riches en liens lexicaux du français : la

première est le DiCouèbe¹⁸; la seconde est une base de données en format FileMaker cédée par Igor Mel'čuk à l'OLST. Ces deux bases de données ont été converties en format XML. À l'heure actuelle, la base de données du DiCoLiLex contient environ 5 000 lexies vedettes et 5 900 fonctions lexicales, pour un total d'environ 37 000 liens lexicaux (Marengo et Robichaud, 2017). Il existe aussi une interface en ligne qui est en développement. L'interface contient une nomenclature de vocables, et chaque vocable contient une liste des lexies distinguées à l'aide d'identifiants alphanumériques et d'étiquettes sémantiques et sous chaque lexie, nous pouvons trouver la forme propositionnelle ainsi que les fonctions lexicales qu'elle contrôle, ce qui permet une consultation classique, comme dans un dictionnaire traditionnel (Marengo et Robichaud, 2017).

Notre travail s'appuie sur ces dictionnaires : nous visons à élaborer un dictionnaire selon le modèle du DEC en mandarin, et nous nous inspirons en particulier de la structure du DiCoLiLex pour que ce dictionnaire puisse être intégré dans des applications de linguistique computationnelle.

¹⁸ Voir : <http://olst.ling.umontreal.ca/dicouebe>

Chapitre 3 Méthodologie de la recherche

3.1 Établissement de la nomenclature

Notre recherche est basée sur l'élaboration d'une nomenclature bilingue franco-mandarin à partir d'un sous-ensemble du DiCoLiLex. Ainsi, nous devons d'abord établir une nomenclature française dont les lexies appartiennent à un champ sémantique spécifique. Idéalement, les lexies de notre nomenclature doivent posséder le plus de fonctions lexicales standard possible, de sorte que nous ayons assez de matière pour mener une étude comparative sur les collocations des lexies dans la nomenclature correspondante en mandarin.

Dans le DiCoLiLex, un fichier contient un « mot », et il peut y avoir plusieurs vocables ayant la même forme, mais la plupart des fichiers contiennent seulement un vocable. Ensuite, chaque vocable englobe une ou plusieurs lexies copolysèmes. Pour choisir les lexies ayant le plus de fonctions lexicales dans le DiCoLiLex, nous avons trié les documents par taille en ordre décroissant et ainsi avons obtenu une liste des vocables à peu près par richesse des fonctions lexicales. Nous avons observé les lexies dans cette liste et dans le DiCoLiLex, ainsi nous avons constaté que plusieurs lexies dans le champ sémantique des émotions avaient un bon nombre de fonctions lexicales standard. Nous avons ainsi choisi les vingt lexies françaises qui ont le plus de fonctions lexicales dans le champ sémantique des émotions pour établir une nomenclature en français. À partir de cette nomenclature, nous avons établi une nomenclature correspondante des équivalents de ces lexies en mandarin, et nous y avons ajouté une vingtaine de lexies courantes dans ce champ, soit les lexies qui expriment les sentiments de base, tels que 'colère', 'joie', 'tristesse', 'peur' (Chen, Li et Hung 2009).

Il est à noter qu'en français, les lexies sous le même vocable possèdent la même partie du discours. Cependant, en mandarin, certains « mots » peuvent s'utiliser en tant que nom et verbe dans les phrases. Par exemple en (1), 爱^{ài} 'aimer' et 爱^{ài} 'amour' partagent le même signifiant et les mêmes actants sémantiques, mais ils ont différentes fonctions grammaticales

dans la phrase. Néanmoins, nous trouvons seulement une entrée verbale pour ces deux usages dans la plupart des dictionnaires chinois. Alors, il faut se demander si ces lexies appartiennent à deux vocables séparés.

- (1) a. 父^{fù}母^{mǔ} 无^{wú}私^{sī} 地^{de} 爱^{ài} 孩^{hái}子^{zi}¹⁹。
 père.mère être.désintéressé DE₂ aimer enfant
 ‘Les parents aiment leur enfant de façon désintéressée’
- b. 父^{fù}母^{mǔ} 对^{duì} 孩^{hái}子^{zi} 的^{de} 爱^{ài} 是^{shì} 无^{wú}私^{sī} 的^{de}。
 père.mère envers enfant DE₁ amour être être.désintéressé DE₁
 ‘L’amour des parents envers leur enfant est désintéressé’

En fait, la classification des catégories grammaticales a provoqué des discussions entre les linguistes depuis longtemps puisqu’il n’y pas vraiment de marqueur morphologique en mandarin pour distinguer les différentes parties du discours et les linguistes n’utilisent pas les mêmes critères pour distinguer les classes (R. Guo, 2002). Ma (1898) a premièrement proposé de classer les mots chinois en neuf classes grammaticales, en s’appuyant sur la grammaire occidentale et en la combinant avec un système de classement entre « mot plein » et « mot vide » déjà existant en mandarin, alors que Gao (1953) propose une hypothèse extrême qu’il n’existe pas de distinction de parties du discours pour les mots lexicaux, car il n’existe pas de morphologie en mandarin. Cette hypothèse est critiquée par beaucoup de linguistes chinois (Lü, 1955; Lu, 1953; Zhu, 1960). À part la discussion sur l’existence des parties du discours en mandarin, il n’existe pas de consensus pour traiter les mots qui peuvent appartenir à plus qu’une partie du discours. Lü et Zhu (1952) pensent que la partie du discours d’une lexie ne change pas selon sa fonction dans la phrase si le sens reste le même. Guo (2002) a fait une étude détaillée sur les classifications des mots, et il traite les verbes qui peuvent s’utiliser à la place des noms comme des verbes nominaux. Shen (2016) propose que le verbe soit une sous-catégorie du nom, car les verbes en mandarin peuvent servir de sujet; et les noms peuvent servir de prédicats.

¹⁹ La plupart des exemples en mandarin que nous utilisons dans ce mémoire proviennent du corpus BCC (<http://bcc.blcu.edu.cn>), avec des modifications mineures parfois afin d’illustrer les points les plus importants. Les autres sont inventés par nous-même.

Cependant, dans notre recherche, nous insistons pour bien séparer les parties du discours des lexies selon leur fonction syntaxique dans la phrase, parce que cette information fait partie de leur syntactique, qui est le troisième volet pour distinguer un signe linguistique d'un autre. De plus, si nous ne séparons pas clairement les parties du discours des lexies selon leur comportement syntaxique, cela va provoquer des problèmes au niveau de la syntaxe, et surtout nous empêcher de choisir les bonnes fonctions lexicales selon les relations syntaxiques. Par conséquent, nous supposons que la lexie qui est utilisée comme verbe et celle ayant la même forme, mais s'utilise dans une position nominale, sont bien deux lexies séparées : l'un est une lexie verbale, l'autre est une lexie nominale, parce qu'elles ont une combinatoire syntaxique différente, et elles contrôlent différentes collocations. Puisque ces deux lexies n'ont pas la même partie du discours, elles appartiennent à deux vocables. Ainsi, dans notre nomenclature, nous avons souvent un vocable verbal et un autre vocable nominal qui ont exactement le même signifié, ce qui peut avoir l'air un peu redondant, mais cette distinction va faciliter la description lexicale dans notre dictionnaire.

Il est aussi à noter que nous n'avons pas inclus la classe adjectivale dans notre dictionnaire en mandarin. En fait, les adjectifs en mandarin sont considérés comme une sous-catégorie des verbes par plusieurs linguistes (Duanmu, 1997; McCawley, 1992; Sproat et Shih, 1988), puisqu'ils peuvent s'utiliser indépendamment comme prédicat syntaxique. Nguyen (2006) a fait une étude sur les unités lexicales en mandarin dans le cadre de la TST, et a montré que la classe correspondant aux adjectifs en mandarin (qu'il appelle les qualificatifs) a plusieurs propriétés syntaxiques et morphologiques communes aux verbes : ce sont des unités lexicales prédicatives qui peuvent constituer le gouverneur syntaxique d'une unité lexicale qui exprime son argument sémantique, cf. Figure 1; elles peuvent également constituer le sommet syntaxique d'une phrase complète, comme les verbes, cf. (2)a; les qualificatifs peuvent porter les suffixes aspectuels des verbes, cf. (2)b; les qualificatifs peuvent être niés et interrogés de la même manière que les verbes à l'aide d'un adverbe de négation 不^{bù} 'NEG', cf. (2)c-d.

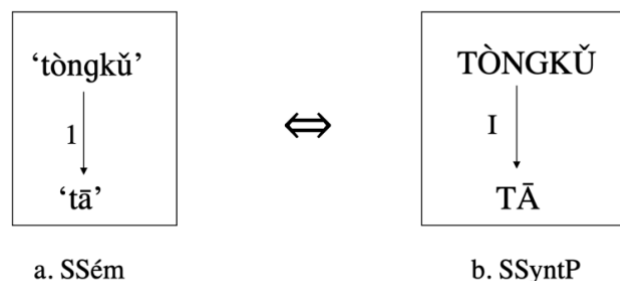


Figure 1. SSém et SSyntP de l'expression 她痛苦 'elle est douloureuse'

- (2) a. 她 很 痛苦。
 tā hěn tòng kǔ
 elle très être.douloureux
 'Elle est très douloureuse'
- b. 她 痛苦过。
 tā tòng kǔ guò
 elle être douloureux-PERF
 'Elle a été douloureuse.'
- c. 她 不 痛苦。
 tā bù tòng kǔ
 elle NEG être.douloureux
 'Elle est très douloureuse'
- c. 她 痛(苦) 不 痛苦?
 tā tòng kǔ bù tòng kǔ
 elle être douloureux NEG être douloureux
 'Elle est douloureuse ou pas douloureuse ?'
 ='Est-elle douloureuse?'

Ainsi, Nguyen (2006) affirme que les qualificatifs forment une sous-classe des verbes qu'il appelle les verbes qualificatifs. Nous adoptons sa proposition et traitons donc cette classe de lexies comme des verbes dans notre dictionnaire. Par conséquent, les lexies adjectivales dans notre nomenclature française vont correspondre à des verbes qualificatifs dans la nomenclature en mandarin.

Voici la nomenclature des 43 lemmes (78 vocables et 126 lexies au total) en mandarin que nous avons choisis de décrire ainsi que les vocables correspondants en français (pour les vocables qui n'ont pas d'équivalent français, nous avons mis leur glose) :

Vocable en mandarin	Vocables correspondants en français
爱	AMOUR; AIMER
爱河	'rivière d'amour'
悲伤	TRISTESSE; TRISTE
不安	TROUBLE; TROUBLÉ
不满	MÉCONTENTEMENT; MÉCONTENT
吃醋	JALOUX
耻辱	HONTE

dānxīn 担心	INQUIÉTUDE; S'INQUIÉTER
fánnǎo 烦恼	TRACAS; SE TRACASSER
fènnù 愤怒	FUREUR; FURIEUX
gǎnqíng 感情	SENTIMENT
gāoxìng 高兴	CONTENTEMENT; CONTENT
hàipà 害怕	PEUR; AVOIR PEUR
hàoqí 好奇	CURIOSITÉ; CURIEUX
hèn 恨	HAINE; HAÏR
huǒ 火	COLÈRE; SE FÂCHE
jiāoào 骄傲	ORGUEIL; ORGUEILLEUX
jiāolǜ 焦虑	ANXIÉTÉ; ANXIEUX
jīdòng 激动	AGITATION; AGITER
jīdù 嫉妒	JALOUSIE; JALOUX
jǔsàng 沮丧	FRUSTRATION; FRUSTRER
kāixīn 开心	JOIE; JOYEUX
kǒngjù 恐惧	EFFROI; S'EFFRAYER
kuàilè 快乐	JOIE; JOYEUX
liànài 恋爱	'relation amoureuse' ; AMOUREUX
nù 怒	COLÈRE; SE FÂCHER
pà 怕	PEUR; AVOIR PEUR
qì 气	COLÈRE; SE FÂCHER
qíngxù 情绪	ÉMOTION
rèqíng 热情	ENTHOUSIASME; ENTHOUSIASTE
tǎoyàn 讨厌	DÉGOÛT; DÉTESTER
tòngkǔ 痛苦	DOULEUR; DOULOUREUX
xǐhuan 喜欢	AFFECTION; AIMER
xīn 心	CŒUR
xīnshǎng 欣赏	ADMIRATION; ADMIRER
xīwàng 希望	ESPOIR; ESPÉRER

yàn wù 厌恶	AVERSION; DÉTESTER
yí hàn 遗憾	REGRET; REGRETTER
yōu chóu 忧愁	TRISTESSE; TRISTE
yōu lǜ 忧虑	SOUCIE; SOUCIEUX
yuàn hèn 怨恨	RANCUNE; 'EN VOULOIR'
yù wàng 欲望	DÉSIR
zì háo 自豪	FIERTÉ; FIER

Tableau II. Nomenclature des lemmes en mandarin et leurs équivalents en français

3.2 Choix du corpus

Une fois que nous avons obtenu la nomenclature en mandarin, nous pouvons commencer à travailler sur leurs collocations d'après le corpus mandarin avec l'aide du langage Python. Pour ce faire, il faut d'abord choisir un corpus assez grand et riche en mandarin. Nous avons trouvé plusieurs corpus informatisés développés par les universités ou les académies sur Internet, et dans le tableau ci-dessous, nous avons fait une comparaison de ces différents corpus selon leur nombre des tokens, des caractères et des lexies, etc. :

Corpus	Total des tokens	Total des caractères	Total des caractères mandarins	Total des lexies	Total des textes
BLCU Corpus Center (BCC) ²⁰	-	15 000 000 000	-	-	-
Centre for Chinese Linguistics PKU ²¹	-	581 794 456	509 913 589	10 645	-

²⁰ Voir : <http://bcc.blcu.edu.cn>

²¹ Voir : http://ccl.pku.edu.cn:8080/ccl_corpus/

Media Language Corpus ²²	135 767 884	241 316 530	200 071 896	-	34 039
Modern Chinese Corpus ²³	12 842 116	19 455 328	-	15 1300	9 487
Academia Sinica Balanced Corpus of Modern Chinese (version 4.0) ²⁴	11 245 330	-	17 554 089	239 598	19 247
Lancaster Corpus of Mandarin Chinese version 2 (LCMCv2) ²⁵	1,021,037	-	-	43,809	500
TORCH 2009 ²⁶	1,289,373	-	-	49,210	15
The UCLA Corpus of Written Chinese (2nd edition) ²⁷	1,097,113	-	-	43,320-	15

Tableau III. Corpus en mandarin moderne

Cependant, tous ces corpus ne peuvent s'utiliser qu'en ligne, et nous pouvons seulement télécharger les résultats de recherche au lieu de télécharger les corpus au complet, ce qui nous empêche d'extraire les cooccurrents dans le corpus complet avec l'aide de Python. Par conséquent, nous avons trouvé deux autres corpus téléchargeables : le premier est un corpus de Weibo qui inclut environ cinq millions de messages weibo (microblog chinois comme Twitter) fourni par Huaping Zhang²⁸; le deuxième est le corpus de Wikipédia chinois²⁹ qui est un grand corpus ouvert et qui est mis à jour chaque mois.

²² Voir : <http://ling.cuc.edu.cn/RawPub/>

²³ Voir : <http://www.aihanyu.org/cncorpus/CnCindex.aspx>

²⁴ Voir : http://asbc.iis.sinica.edu.tw/index_readme.htm

²⁵ Voir : <http://111.200.194.212/cqp/lcmc2/>

²⁶ Voir : <http://111.200.194.212/cqp/torch09/>

²⁷ Voir : <http://www.lancaster.ac.uk/fass/projects/corpus/UCLA/>; <http://111.200.194.212/cqp/ucla2/>

²⁸ Adresse de téléchargement : <http://www.nlpir.org/download/weibo.7z>

²⁹ Adresse de téléchargement : <https://dumps.wikimedia.org/zhwiki/latest/zhwiki-latest-pages-articles.xml.bz2>

3.3 Segmentation du corpus

À l'aide de Python, nous avons transformé les deux corpus de Weibo et de Wikipédia chinois que nous avons téléchargés en format texte. N'étant pas comme le français, l'anglais ou bien d'autres langues européennes où la plupart du temps les mots-formes sont bien segmentés par les espaces dans une phrase, il n'y a aucune espace entre les mots en mandarin. Ainsi, afin de pouvoir bien travailler sur notre corpus, il est important de segmenter les textes en premier lieu.

Nous trouvons deux modules Python pour segmenter les textes en mandarin. L'un est THULAC (Tsing-Hua University Lexical Analyzer for Chinese) développé par le laboratoire de traitement du langage naturel et de science sociale informatique de l'Université Tsing-Hua. Ce module permet aussi d'indiquer les parties du discours des mots-formes. L'autre est Jieba (litt. 'bégayer') qui existe dans plusieurs langages de programmation. Ce module de segmentation est basé sur les règles d'un dictionnaire et les statistiques de la fréquence des mots-formes³⁰. Nous avons fait quelques tests pour comparer ces deux modules en ce qui concerne la vitesse du traitement et l'exactitude des résultats, et avons constaté que THULAC prend beaucoup plus de temps à segmenter un texte que Jieba. Ce dernier a une vitesse de 400KB par seconde selon leur présentation. En raison de la moins bonne performance de THULAC, nous avons choisi le module Jieba pour segmenter nos deux corpus, et nous avons obtenu deux grands textes bien segmentés de Weibo et de Wikipédia chinois ayant respectivement la taille 0.98GB et 1.26GB. À partir de ces deux corpus, nous avons utilisé le langage Python pour nettoyer les bruits et ensuite chercher les collocations des lexies dans notre nomenclature. Toutefois, nous avons remarqué que le corpus de Weibo se rapproche plus du mandarin parlé, et contient plus d'occurrences des lexies à l'étude que celui de Wikipédia. Ainsi, nous utiliserons principalement le corpus de Weibo pour extraire les cooccurrents.

³⁰ Voir la partie d'algorithme sur le site : <https://github.com/fxsjy/jieba>

Voici deux exemples de la segmentation de Jieba dans le corpus de Weibo :

- (3) ^{hěn gāo xìng zài cì jiàn dào nǐ men}
 很 高兴 再次 见到 你们。
 très content encore-fois voir-arriver vous
 ‘Je suis très content de vous voir de nouveau.’
- (4) ^{xī wàng hé shī wàng zǒng shì qiānzhuóshǒu lái dào nǐ miànqián}
 希望 和 失望 总是 牵着手 来到 你 面前,
 espoir et désespoir toujours mener-PROG-main venir-arriver toi face-devant
 ‘L’espoir et le désespoir arrivent toujours devant toi la main dans la main’
- ^{rú guǒ nǐ hài pà shī wàng biàn de jù dà zuì hòu tūn shì le xī wàng}
 如果 你 害怕, 失望 变得 巨大, 最后 吞噬 了 希望;
 si toi avoir-peur, désespoir devenir énorme, finalement avaler PERF espoir
 ‘Si tu as peur, le désespoir deviendra énorme, et il avalera finalement l’espoir’
- ^{rú guǒ nǐ yǒng gǎn miàn duì jiù néng gǎi biàn kùn jìng}
 如果 你 勇敢 面对, 就 能 改变 困境。
 si toi brave affronter, alors pouvoir changer impasse
 ‘si tu l’affrontes bravement, et tu pourras changer l’impasse.’

Dans ces exemples, la plupart des mots-formes sont bien segmentés, sauf quelques imperfections : un syntagme libre ^{qiānzhuóshǒu} 牵着手 n’est pas segmenté; deux constructions verbales sérielles résultatives (§4.4.3) ^{jiàn dào lái dào} 见到,来到 sont considérées comme un mot-forme; le suffixe de l’aspect accompli ^{le} 了 est considéré comme un clitique.

3.4 Extraction automatique des cooccurrents

Pour trouver les collocations des lexies dans notre nomenclature, nous nous sommes inspirée de Barnbrook, Mason et Krishnamurthy (2013) pour calculer les différents scores de significations tels que le score z, le score t et quelques variantes de l’information mutuelle spécifique (PMI, NPMI, PMI² et PMI³). Nous avons adapté leur code (originellement écrit en awk) en Python, qui est plus simple et plus efficace pour de gros corpus.

Avant d’expliquer les procédures d’extraction, il faut d’abord clarifier la différence entre les collocations statistiques et les collocations linguistiques. Les collocations statistiques, aussi appelées cooccurrents, sont des unités linguistiques qui se présentent simultanément dans le même énoncé (la phrase, le paragraphe, l’extrait), mais elles ne sont pas nécessairement liées par des relations syntaxiques ou lexicales. Ce sont des associations purement statistiques, tandis que les collocations linguistiques possèdent forcément un lien linguistique, comme nous avons

expliqué dans le chapitre précédent. Dans le présent mémoire, nous utilisons le terme « cooccurrent » pour désigner les collocations statistiques et le terme « collocation » tout court pour les collocations linguistiques. Ainsi, nous devons d’abord extraire les cooccurrents dans le corpus, et nous pouvons ensuite chercher les collocations parmi eux.

Pour commencer le traitement des cooccurrents, nous devons d’abord trouver un contexte lexical pour chaque lexie dans notre nomenclature. Pour ce faire, nous avons relevé les mots-formes qui apparaissaient dans la fenêtre autour de la lexie en question qui fonctionne comme un nœud dans le contexte. Le choix de la largeur de la fenêtre va influencer notre résultat : si la fenêtre est trop grande, des mots dans les phrases avant ou après qui n’ont aucun lien syntaxique avec le nœud peuvent apparaître plus souvent, mais cela nous permet de trouver des dérivatés sémantiques; si la fenêtre est trop petite, nous risquons de perdre ces dérivations, mais on augmente les chances que les mots soient liés syntaxiquement. Ainsi, il est important de trouver une longueur optimale pour notre tâche en mandarin. Nous avons testé des fenêtres de 4 et de 5 mots-formes à gauche et à droite. Étant donné que les phrases courtes sont plus utilisées en mandarin, nous avons remarqué qu’il y avait plus de bruit dans la fenêtre de taille 5, donc nous avons opté pour la fenêtre de 4.

Ensuite, à l’aide de la fonction *nltk.FreqDist()* dans le module collocations de *Natural Language Toolkit* (NLTK), nous pouvons calculer les fréquences d’occurrence de chaque token dans le contexte lexical ainsi que dans le corpus au complet. Pour évaluer si une collocation est pertinente ou non, la façon la plus simple est de comparer sa fréquence observée par rapport à la fréquence à laquelle nous nous attendrions (Barnbrook et al., 2013). La fréquence observée est la fréquence de chaque token dans le contexte tandis que la fréquence attendue peut être calculée par la fréquence de chaque token dans le corpus au complet divisé par la taille du corpus et ensuite multiplié par la taille du contexte. Ainsi, nous pouvons diviser la fréquence observée par celle attendue pour calculer le ratio entre ces deux fréquences. Les mots-formes avec un score plus grand que 1 sont plus fréquents qu’attendus, et c’est le contraire pour ceux avec un score de moins de 1. Ainsi, les mots-formes qui sont lexicalement liés au mot-clé doivent avoir un ratio relativement élevé.

Cependant, il est à noter que les prénoms, les particules, la ponctuation et d'autres mots-outils ou symboles peuvent avoir un ratio assez élevé. Nous n'avons pas vraiment besoin de calculer les valeurs pour ces types de mots qui ne sont pas des collocations linguistiques, donc nous avons préparé une liste de « stopwords » en mandarin qui contient environ 1800 mots et symboles, par exemple, 他 'lui', 个 'CL', 这 'ce', 也 'aussi', 和 'et', 所以 'donc', 比如 'par exemple'. Nous les avons enlevés du dictionnaire de fréquences du contexte avant de faire les calculs.

Les hapax et les mots rares dans le corpus peuvent également entraîner des résultats peu fiables. Si ces mots se trouvent une ou deux fois dans le contexte du mot-clé par hasard, ils peuvent avoir un score élevé, donc nous avons aussi enlevé les mots qui apparaissent moins de trois fois pour éviter ce genre de problème.

Par conséquent, le ratio observé/attendu peut éliminer une partie du bruit dans la liste de candidats de collocations. Néanmoins, le ratio simple n'est pas très fiable en tant que mesure d'association parce qu'il contient au moins deux types de bruit (Barnbrook et al., 2013, p. 64). Le premier type de bruit vient des mots-formes avec des fréquences extrêmes, soit très haute soit très basse, mais nous pouvons déjà éliminer les mots-outils et les hapax pour éviter une partie de ce bruit comme ce que nous avons expliqué plus haut. Le deuxième type est provoqué par les mots-formes sans rapport linguistique qui sont attirés par les autres cooccurrents des lexies en question. Par exemple, dans une extraction de Barnbrook et al. (2013, p. 63), le mot-forme *severe* apparaît par hasard dans la liste des cooccurrents du mot-clé *eye*, mais la raison de son apparition est parce qu'il modifie un autre mot-forme *pain* mentionné ensemble avec *eye*, comme dans « *causing eye damage and severe pain* ». Cependant, ce mot-forme n'a pas vraiment de lien direct avec *eye*, et nous attendons plutôt sa cooccurrence avec *pain* qu'avec *eye*. Pour identifier ce type de bruit, il faut souvent retourner dans la concordance pour vérifier les relations entre les mots-formes, ce qui n'est pas très efficace quand nous traitons un gros corpus.

Par conséquent, nous avons utilisé d'autres scores plus complexes pour calculer la force des associations entre les mots. Dans les sections suivantes, nous présenterons brièvement les

formules des différents scores pour extraire les cooccurents ainsi que quelques exemples des résultats obtenus. Nous comparerons ensuite ces différents scores afin de trouver la meilleure mesure.

3.4.1 Score z et score t

La formule pour calculer le score z est : $z = \frac{(o-a)}{\sigma}$.

Il faut calculer d'abord la fréquence observée (o) moins la fréquence attendue (a) et ensuite diviser par la dérivation standard de la fréquence d'occurrence du mot dans l'ensemble du texte avec la formule suivante : $\sigma = \sqrt{N(p(1-p))}$, où N représente la taille du corpus au complet, et p représente la probabilité d'occurrence du token. Ainsi, nous pouvons obtenir le classement du score z pour chaque lexie dans son contexte lexical.

La formule pour calculer le score t ressemble beaucoup à celle du score z : $t = \frac{(o-a)}{\sqrt{o}}$.

La seule différence est de changer σ dans la formule du score z par la racine carrée de la fréquence observée. Ainsi, nous pouvons facilement calculer le score t.

Pour illustrer les différents classements obtenus par ces 2 scores, nous présentons ci-dessous les résultats que nous avons obtenus en calculant le score z et le score t du token ^{kǒng jù} 恐惧 'effroi; s'effrayer' :

Rang	Vocable	Score z	Vocable	Score t
1	恐惧 'effroi; s'effrayer'	47.98	恐惧 'effroi; s'effrayer'	49.36
2	克服 'surmonter'	1.92	内心 'cœur, intérieur'	8.27
3	密集 'dense'	1.55	未知 'inconnu'	7.77
4	战胜 'vaincre'	1.4	战胜 'vaincre'	7.58
5	事外 'en dehors des affaires'	1.34	克服 'surmonter'	7.47
6	未知 'inconnu'	1.32	密集 'dense'	7.32
7	免于 'exempt de'	1.12	死亡 'mourir; mort'	6.65
8	贪婪 'avide'	1	充满 'se remplir de'	5.87
9	幽闭 'claustration'	0.99	失败 'échouer; échec'	5.73
10	小丹 NPR	0.83	贪婪 'avide'	5.72

11	戒慎	ERR ³¹	0.82	害怕	‘craindre; craint’	5.63
12	忧虑	‘soucieux; souci’	0.75	未来	‘avenir’	5.25
13	内心	‘cœur; intérieur’	0.65	面对	‘face à; affronter’	5.11
14	不安	‘inquiet; inquiétude’	0.64	焦虑	‘anxieux; anxiété’	4.85
15	王莽	NPR	0.6	愤怒	‘furieux; colère’	4.79
16	焦虑	‘anxieux; anxiété’	0.6	心中	‘au cœur’	4.63
17	死亡	‘mourir; mort’	0.6	孤独	‘solitaire; solitude’	4.6
18	伴生	‘concomitance’	0.57	不安	‘inquiet; inquiétude’	4.55
19	驱走	‘chasser’	0.52	心里	‘dans le cœur’	4.31
20	能克服	ERR	0.52	一种	CL	4.29

Tableau IV. Scores z et t des principaux cooccurrents de 恐惧^{kǒng jù} (‘effroi; s’effrayer’)

Les deux listes semblent assez raisonnables, mais nous pouvons observer que certains noms propres et des tokens mal segmentés occupent un rang assez élevé selon le score z, alors que les mots plus communs, par exemple le classificateur [=CL], montent dans la liste du score t. En fait, dans l’extraction des cooccurrents de Barnbrook et al. (2013), les mots-outils ne sont pas enlevés avant de calculer les scores, donc ces types de mots sont montés dans la liste du score t, donc les auteurs pensent que le score t fonctionne moins bien que le score z pour filtrer les mots-outils.

3.4.2 Information mutuelle (MI) et information mutuelle spécifique (PMI)

Les formules pour calculer le MI et le PMI sont, respectivement :

$$MI(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad \text{et} \quad PMI(x, y) = \log \frac{p(x, y)}{p(x)p(y)}.$$

Dans la formule du PMI, $p(x, y)$ est la probabilité que x et y apparaissent ensemble dans le corpus, tandis que $p(x)$ et $p(y)$ sont respectivement la probabilité de trouver x et y (indépendamment) dans le corpus. Dans notre cas, la variable x est le mot-clé, et la variable y

³¹ Les ERR correspondent à des mots mal segmentés.

représente ses cooccurrents dans la concordance. Nous pouvons facilement calculer $p(x)$ et $p(y)$ en divisant leur nombre d'occurrences par la taille du corpus, et $p(x)$ multiplié par $p(y)$ représente la fréquence attendue de trouver x et y ensemble, alors que $p(x,y)$ représente la fréquence observée de x et y ensemble. En calculant le logarithme du ratio de ces deux fréquences, nous obtenons le PMI.

Quant au score MI, il est l'espérance mathématique du PMI, donc ce score n'est pas vraiment pertinent dans l'extraction des collocations ici. Alors, la façon de Barnbrook et al. (2013) pour calculer le soi-disant MI ($\frac{\log(\text{observé}/\text{attendu})}{\log(2)}$) est en fait celle pour calculer le PMI. Comme Bouma (2009) l'indique, dans la littérature de linguistique computationnelle, on fait souvent référence au score PMI en tant que MI tout court, ce qui amène parfois des confusions.

Il est à noter que le score PMI nous donne le même classement que le ratio observé/attendu. Étant donné que le logarithme est constant, les valeurs du score sont différentes, mais ce qui change, c'est l'ordre de grandeur des espaces entre les rangs, comme nous pouvons constater dans le tableau de comparaison ci-dessous, le 2^e rang a un PMI 10 fois plus élevé que le 15^e rang, mais un ratio seulement 1,3 fois plus élevé :

Rang	Vocable		Ratio observé/ attendu	Vocable		PMI
1	恐惧	'peur; avoir peur'	10.79	恐惧	'peur; avoir peur'	6038.35
2	事外	'en dehors des affaires'	9.82	事外	'en dehors des affaires'	2282.21
3	小丹	NPR	9.38	小丹	NPR	1474.66
4	戒慎	ERR	9.06	戒慎	ERR	1065.03
5	幽闭	'claustration'	8.76	幽闭	'claustration'	786.49
6	伴生	'concomitance'	8.62	伴生	'concomitance'	684.66
7	免于	'exempt de'	8.6	免于	'exempt de'	672.65
8	能克服	ERR	8.45	能克服	ERR	580.93
9	透骨	'qui pénètre jusqu'aux os'	8.42	透骨	'qui pénètre jusqu'aux os'	563.84
10	王莽	NPR	8.23	王莽	NPR	463.06
11	克服	'surmonter'	8.13	克服	'surmonter'	421.99
12	驱走	'chasser'	7.77	驱走	'chasser'	294.93

13	密集	‘dense’	7.75	密集	‘dense’	287.56
14	周公	NPR	7.6	周公	NPR	247.68
15	战胜	‘vaincre’	7.47	战胜	‘vaincre’	218.28
16	怯懦	‘lâche’	7.45	怯懦	‘lâche’	213.01
17	暗夜	‘nuit sombre’	7.44	暗夜	‘nuit sombre’	211.25
18	忧虑	‘soucieux; souci’	7.39	忧虑	‘soucieux; souci’	200.74
19	贪婪	‘avide’	7.37	贪婪	‘avide’	197.27
20	猜疑	‘soupçonner; soupçon’	7.36	猜疑	‘soupçonner; soupçon’	194.82

Tableau V. Ratio observé/attendu et PMI de 恐惧^{kǒng jù} (‘effroi; s’effrayer’)

Dans ce tableau, nous pouvons constater qu’il y a plus d’erreurs et de noms propres dans le haut des listes que dans celles du score z et du score t. Ce sont des mots-formes ayant une fréquence très basse dans le corpus. Étant donné que le dénominateur est très petit dans ces cas-là, ces mots rares obtiennent ainsi un score élevé. Il s’agit d’un problème bien connu du PMI. Pour éliminer ce genre de bruit, nous nous appuyerons sur les variantes du PMI.

3.4.3 Variantes du PMI

Afin de résoudre le problème que les mots rares ont souvent un score de PMI élevé dans le corpus, plusieurs variantes du PMI ont été proposées. Par exemple, la famille de PMI^k (Daille, 1994) est parmi les variantes les plus utilisées. Ces variantes consistent à élever $p(x,y)$, la probabilité de trouver x et y ensemble, à la puissance k à l’intérieur du logarithme pour corriger empiriquement le biais de PMI par rapport aux mots-formes ayant une fréquence basse (Role et Nadif, 2011). Voici les formules de PMI² et PMI³, qui sont couramment utilisées :

$$PMI^2(x; y) = \log \frac{p(x, y)^2}{p(x)p(y)} \quad \text{et} \quad PMI^3(x; y) = \log \frac{p(x, y)^3}{p(x)p(y)}$$

Contrairement aux scores de PMI, les scores de PMI² et PMI³ sont toujours des nombres négatifs : 0 représente une association parfaite et plus le score est proche de 0, plus l’association est forte.

En outre, Bouma (2009) a proposé une mesure de PMI normalisé entre $[-1, +1]$:

$$\text{NPMI}(x, y) = \frac{\text{PMI}(x, y)}{-\log p(x, y)}$$

Ainsi, les mots ayant un score -1 n'apparaissent jamais ensemble, 0 représente une indépendance et $+1$ signifie une cooccurrence complète.

Nous pouvons ainsi comparer les intervalles de variation des variantes du PMI dans le tableau suivant (Role et Nadif, 2011):

Variantes	Cooccurrence complète	Indépendance	Nul
PMI	$-\log p(x, y)$	0	$-\infty$
NPMI	1	0	-1
PMI ²	0	$\log p(x, y)$	$-\infty$
PMI ³	0	$2\log p(x, y)$	$-\infty$

Tableau VI. Intervalles de variation de PMI, NPMI, PMI² et PMI³

Dans notre extraction, nous avons remarqué que le PMI³ est plus efficace que le PMI² pour faire descendre les mots-formes ayant une fréquence basse, donc dans nos exemples, nous avons opté seulement pour les résultats du PMI³. Dans le tableau ci-dessous, nous comparons les classements entre le NPMI et le PMI³.

Rang	Vocable	NPMI	Vocable	PMI ³
1	恐惧 ‘effroi; s’effrayer’	0.99	恐惧 ‘effroi; s’effrayer’	-11.02
2	事外 ‘en dehors des affaires’	0.57	克服 ‘surmonter’	-21.23
3	克服 ‘surmonter’	0.55	密集 ‘dense’	-21.68
4	小丹 NPR	0.53	战胜 ‘vaincre’	-21.82
5	免于 ‘exempt de’	0.53	未知 ‘inconnu’	-21.88
6	幽闭 ‘claustration’	0.53	贪婪 ‘avide’	-23.05
7	密集 ‘dense’	0.53	内心 ‘cœur, intérieur’	-23.11
8	戒慎 ERR	0.52	死亡 ‘mourir; mort’	-23.72
9	战胜 ‘vaincre’	0.51	免于 ‘exempt de’	-23.84
10	未知 ‘inconnu’	0.5	忧虑 ‘soucieux; souci’	-24.24
11	伴生 ‘concomitance’	0.49	事外 ‘en dehors des affaires’	-24.37
12	贪婪 ‘avide’	0.48	焦虑 ‘anxieux; anxiété’	-24.38

13	王莽	NPR	0.48	不安	‘inquiet; inquiétude’	-24.4
14	能克服	ERR	0.48	害怕	‘craindre; crainte’	-24.43
15	透骨	‘qui pénètre jusqu’aux os’	0.48	幽闭	‘claustration’	-24.5
16	忧虑	‘soucieux; souci’	0.47	失败	‘échouer; échec’	-24.74
17	驱走	‘chasser’	0.46	愤怒	‘furieux; colère’	-25.12
18	周公	NPR	0.44	充满	‘se remplir de’	-25.19
19	不安	‘inquiet; inquiétude’	0.44	莫名	‘inconnu’	-25.49
20	猜疑	‘soupçonner; soupçon’	0.43	戒慎	ERR	-25.58

Tableau VII. NPMI et PMI³ de 恐惧 (kǒng jù) (‘effroi; s’effrayer’)

Dans ce tableau, nous pouvons remarquer que dans le classement du NPMI, plusieurs noms propres sont apparus dans le haut de la liste, alors qu’aucun de ces mots n’a eu un score élevé dans la liste du PMI³. Contrairement au PMI³, le NPMI n’est pas très efficace pour corriger le biais de PMI. En fait, le NPMI a été développé surtout pour pouvoir comparer les PMI issus de distributions différentes en les normalisant dans la même échelle (de -1 à 1), donc ce score n’est pas très utile pour calculer les cooccurrents qui viennent tous d’un même corpus.

En comparant les classements des différents scores des cooccurrents extraits des vocables dans notre nomenclature, le PMI³ a montré une meilleure performance en éliminant les mots ayant une fréquence trop basse. Par conséquent, nous utiliserons les classements du PMI³ comme matière première pour trouver les collocations des lexies dans notre nomenclature en mandarin.

Chapitre 4 Dictionnaire en mandarin

Dans ce chapitre, nous présenterons d'abord la structure XML du DiCoLiLex en français. Nous proposerons ensuite une adaptation de la structure dans XML ainsi qu'un typage des actants sémantiques avant de faire la description lexicographique des lexies en mandarin, en nous appuyant sur les méthodes d'extraction des cooccurrents en corpus introduites à la section §3.4, p. 36. À partir des fichiers XML rédigés, nous créerons une interface en ligne pour faciliter la consultation de notre dictionnaire. Nous présenterons en particulier les cas spéciaux pour l'encodage des collocations en mandarin en proposant une analyse adaptée au chinois.

4.1 Structure XML du DiCoLiLex

Le DiCoLiLex utilise le format XML³² pour encoder les informations du dictionnaire. Chaque fichier XML contient un vocable français. Dans le DEC et le DiCoLiLex, un vocable est un superarticle qui regroupe des lexies ayant les mêmes signifiants, une combinatoire similaire et des signifiés apparentés (Mel'čuk et al., 1995, p. 155). Les différentes informations concernant un vocable et les lexies qu'il contient sont encodées à l'aide de balises (des constructions entre deux chevrons). Les balises ouvrantes, de la forme `<élément attribut="valeur">`, contiennent nécessairement un élément et potentiellement un ou plusieurs attributs avec leur valeur, et les balises fermantes, de forme `</élément>`, ne possèdent pas d'attributs. Entre une balise ouvrante et sa balise fermante correspondante, nous pouvons insérer d'autres balises ou du texte. Il y a aussi des balises vides `<élément/>`

³² L'*Extensible Markup Language* (XML), ou « langage de balisage extensible » en français, est un métalangage de balisage générique dérivé du *Standard Generalized Markup Language* (SGML) qui représente explicitement la structure d'un document (cf. https://fr.wikipedia.org/wiki/Extensible_Markup_Language [consulté le 11 sept. 2018]).

terminant par une barre oblique, qu'on utilise ici notamment pour faire référence à la lexie vedette d'une fiche (<lexie-ref/>).

Pour donner un exemple, la figure ci-dessous montre la fiche du vocable HEUREUX. Pour commencer, il y a une balise ouvrante qui contient un élément <mot>, ensuite une autre balise ouvrante au deuxième niveau <vocable identificateur="heureux"> avec son élément « vocable » et un attribut « identificateur » et sa valeur « heureux ». Par la suite, il y a une balise vide au troisième niveau <caractéristiques/> qui est censé contenir les balises avec les éléments tels que la partie du discours, le genre, la définitude, le nombre, etc., mais ces informations ne sont pas encore complétées ici. Au même niveau se trouve la balise <lexie>, qui contient d'autres éléments : l'étiquette sémantique, la forme propositionnelle et les fonctions lexicales de cette lexie.

```

<mot>
  <vocable identificateur="heureux">
    <caractéristiques/>
    <lexie identificateur="1" numéro="heureux" date="2016-06-30"
      rédacteur="IM, LC, XYZ" statut="3">
      <étiquette-sémantique>qui éprouve un sentiment positif</étiquette-sémantique>
      <forme-propositionnelle>
        [L'<étiquette-sémantique>individu</étiquette-sémantique>
          <actant numéro="1">X</actant> EST] <lexie-ref/>
      </forme-propositionnelle>
      <fonction-lexicale>
        <fonction type="standard">Magn</fonction>
        <valeur statut="3"><élément>comme un pape</élément></valeur>
        <valeur statut="3"><élément>comme un poisson dans l'eau</élément></valeur>
        <valeur statut="3"><élément>comme un roi</élément></valeur>
      </fonction-lexicale>
      <fonction-lexicale>
        <fonction type="standard">Anti</fonction>
        <valeur statut="3"><élément>malheureux</élément></valeur>
      </fonction-lexicale>
      <fonction-lexicale>
        <fonction type="standard">S0</fonction>
        <valeur statut="3"><élément>bonheur</élément></valeur>
      </fonction-lexicale>
    </lexie>
  </vocable>
</mot>

```

Figure 2. Description du vocable HEUREUX dans le DiCoLiLex

Comme les fonctions lexicales sont la partie la plus importante de notre description lexicale, nous présentons leur encodage plus en détail. À l'intérieur des balises `<fonction-lexicale>...</fonction-lexicale>`³³, il y a d'abord une balise `<fonction>` qui prend le nom de la FL comme texte, et qui contient un attribut « type » pour indiquer si la FL en question est standard (la plupart des FL dans le DiCoLiLex sont de type standard, mais il existe également un petit nombre de FL non standard). Ensuite, nous avons une autre balise `<valeur>` au même niveau, avec un attribut de statut. Cet attribut est un entier de 0 à 3 indiquant le degré de complétude de la fiche. Un statut 0 indique que la valeur contient l'ensemble des éléments requis, à savoir un élément, un régime, un exemple annoté et une glose pour la FL à laquelle cette valeur correspond. Les valeurs de statut 0 ont toutes été révisées et peuvent être mises en ligne. Un statut 1 indique que la valeur contient l'ensemble des renseignements, mais qu'elle n'a pas encore été révisée. Un statut 2 correspond à une valeur à laquelle il manque un renseignement, tel que le régime ou la glose de FL. Finalement, un statut 3 renvoie à la « valeur qu'on a commencé à travailler, mais à laquelle il manque toujours des renseignements » (L'Homme, 2017). Cette balise peut aussi avoir un attribut `fusion="1"` s'il s'agit d'une FL fusionnée (§2.4.2). Enchâsser sous la valise ouvrante `<valeur>`, nous trouverons la balise `<élément>` qui contient un élément de valeur de la FL comme texte. Nous pouvons également trouver des balises pour le régime, les contraintes, les marques d'usage ou un exemple pour donner plus de précision à cette valeur.

Voir l'Annexe 1 pour le fichier original « `_flexidico.rnc` » qui représente un schéma validant la structure XML du DiCoLiLex.

³³ Une balise ouvrante a nécessairement une balise fermante correspondante. Pour alléger la description, nous indiquerons seulement la balise ouvrante dans le texte.

4.2 Adaptation de la structure dans XML

Pour la macrostructure du dictionnaire, dans le DiCoLiLex du français, le niveau de balisage le plus haut est <mot>, suivi de <vocable>, mais en général, un fichier contient un seul vocable regroupant les copolysèmes de ce vocable. Les lexies d'un vocable ont toutes la même partie du discours. Cependant, en mandarin, un sémantème peut correspondre à des lexies qui possèdent une même forme, mais appartiennent à des catégories grammaticales différentes, comme nous avons expliqué à la section §3.1, p. 28. Ainsi, nous avons mis les vocables ayant le même signifiant, mais de différentes parties du discours dans un même fichier.

Pour la microstructure du dictionnaire, d'une part, nous avons ajouté de nouveaux éléments, notamment la glose et le pinyin pour les sinogrammes afin de faciliter la lecture pour les locuteurs non natifs. Plus précisément, nous avons ajouté le pinyin comme attribut des balises <vocable> et <élément> pour indiquer la prononciation de ces éléments représentés par les sinogrammes. Nous avons également ajouté une glose comme attribut de la balise <élément> et comme balise <glose> à l'intérieur de la balise <lexie> afin de donner une traduction des lexies. D'autre part, nous avons supprimé les éléments non pertinents pour le mandarin, par exemple, le genre, la définitude et le nombre sous la balise <caractéristiques>, étant donné que les vocables chinois ne possèdent pas ces éléments de syntactique.

De plus, nous avons modifié les endroits où peuvent se trouver des balises <régime>, car en mandarin, les mots-clés ou leurs actants peuvent apparaître avant la valeur de la FL. Par exemple, le générique de la lexie HONTE est : **Gener**(*honte*) = *sentiment de ~*, mais son équivalent en mandarin est : **Gener**(^{chǐ rǔ}耻辱) = [^{gǎn}~]感. Par conséquent, nous adaptons la structure à l'intérieur de <valeur> pour que les paires de balises <régime>...</régime> puissent apparaître avant et après <élément>...</élément>.

En outre, le DEC utilise le tableau de régime pour décrire la correspondance entre les actants sémantiques et les actants syntaxiques profonds de la lexie en question : le nombre de colonnes correspond au nombre d'actants sémantiques, et le nombre de lignes correspond au

nombre maximal de possibilités d'expressions existantes pour les actants syntaxiques profonds (Mel'čuk, Clas et Polguère, 1995, p. 119-120). Les régimes des lexies sont très importants si nous voulons faire de la génération automatique de textes, puisqu'ils permettent de générer correctement les numéros des actants syntaxiques profonds dans la SSyntP à partir des actants sémantiques dans une SSém. Cependant, les informations de régime sont absentes dans le DiCoLiLex. Nous avons donc ajouté une balise <régime> à l'intérieur de <lexie> pour fournir le régime de la lexie en question.

Voici un exemple d'un vocable nominal encodé dans un fichier XML :

```
<vocable identificateur="担心_n" pinyin="dānxīn">
  <caractéristiques>
    <pdd>n.</pdd>
  </caractéristiques>
  <lexie identificateur="1" date="2018-05-11" statut="3">
    <glose>inquiétude (litt. 'porter coeur')</glose>
    <étiquette-sémantique pinyin="fùmiàn qínggǎn" glose="sentiment négatif">
      负面情感</étiquette-sémantique>
    <forme-propositionnelle>
      <étiquette-sémantique pinyin="gèrén" glose="individu">个人
        </étiquette-sémantique><actant numéro="1">X</actant>对
        <actant numéro="2">Y</actant>的<lexie-ref/>
    </forme-propositionnelle>
    <régime>
      <pdd>N</pdd>=<actant numéro="1">X</actant>对<pdd>N</pdd>=
        <actant numéro="2">Y</actant>的<lexie-ref/>
    </régime>
    <fonction-lexicale>
      <fonction type="standard">Magn</fonction>
      <valeur statut="3"><élément pinyin="nóngnóng" glose="dense">浓浓</élément>
        <régime>的</régime></valeur>
      <valeur statut="3"><élément pinyin="qiángliè" glose="fort">强烈</élément>
        <régime>(的)</régime></valeur>
    </fonction-lexicale>
    <fonction-lexicale>
      <fonction type="standard">AntiMagn</fonction>
      <valeur statut="3"><élément pinyin="yīsī" glose="un fil">一丝</élément>
        </valeur>
    </fonction-lexicale>
    <fonction-lexicale>
      <fonction type="standard">Oper1</fonction>
      <valeur statut="3"><élément pinyin="gǎndào" glose="ressentir">感到</élément>
        <régime><lexie-ref/></régime><contrainte>pas de l'actant Y sur
        <lexie-ref/></contrainte></valeur>
      <valeur statut="3"><élément pinyin="yǒu" glose="avoir">有</élément>
        <régime><lexie-ref/></régime><contrainte>pas de l'actant Y sur
        <lexie-ref/></contrainte></valeur>
    </fonction-lexicale>
    <fonction-lexicale>
      <fonction type="standard">Oper12</fonction>
      <valeur statut="3"><régime>对/为<pdd>N</pdd>=<actant>Y</actant>
```

```

        </régime><élément pinyin="gǎndào" glose="ressentir">感到
        </élément><régime><lexie-ref/></régime></valeur>
    <valeur statut="3"><régime>对<pdd>N</pdd>=<actant>Y</actant></régime>
        <élément pinyin="yǒu" glose="avoir">有</élément><régime>
        <lexie-ref/></régime></valeur>
</fonction-lexicale>
<fonction-lexicale>
    <fonction type="standard">Magn+Oper1</fonction>
    <valeur statut="3"><élément pinyin="chōngmǎn" glose="se remplir de">充满
        </élément><régime><lexie-ref/></régime></valeur>
    <valeur statut="3"><élément pinyin="shēngǎn" glose="ressentir profondément">
        深感</élément><régime><lexie-ref/></régime></valeur>
</fonction-lexicale>
<fonction-lexicale>
    <fonction type="standard">PermlManif</fonction>
    <valeur statut="3"><élément pinyin="liúlù chū" glose="révéler">流露出
        </élément><régime><lexie-ref/></régime></valeur>
    <valeur statut="3"><élément pinyin="lùchū" glose="montrer">露出
        </élément><régime><lexie-ref/></régime></valeur>
</fonction-lexicale>
<fonction-lexicale>
    <fonction type="standard">NonPermlManif</fonction>
    <valeur statut="3"><élément pinyin="yǎnshì" glose="dissimuler">掩饰
        </élément><régime><lexie-ref/></régime></valeur>
</fonction-lexicale>
<fonction-lexicale>
    <fonction type="standard">V0</fonction>
    <valeur statut="3"><élément pinyin="dānxīn" glose="s'inquiéter">担心_v#1
        </élément></valeur>
</fonction-lexicale>
<fonction-lexicale>
    <fonction type="standard">QSyn</fonction>
    <valeur statut="3"><élément pinyin="bù'ān" glose="trouble">不安_n
        </élément></valeur>
    <valeur statut="3"><élément pinyin="dānyōu" glose="inquiétude">担忧_n
        </élément></valeur>
    <valeur statut="3"><élément pinyin="fánnǎo" glose="tracas">烦恼_n#1
        </élément></valeur>
    <valeur statut="3"><élément pinyin="hàipà" glose="peur">害怕_n
        </élément></valeur>
    <valeur statut="3"><élément pinyin="jiāolù" glose="anxiété">焦虑_n
        </élément></valeur>
    <valeur statut="3"><élément pinyin="kǒngjù" glose="peur">恐惧_n
        </élément></valeur>
    <valeur statut="3"><élément pinyin="tǎntè" glose="inquiétude">忐忑_n
        </élément></valeur>
    <valeur statut="3"><élément pinyin="yōulù" glose="souci">忧虑_n
        </élément></valeur>
</fonction-lexicale>
</lexie>
</vocablé>

```

Figure 3. Description du vocable ^{dān xīn} 担心 'inquiétude' dans le DiCoLiLex-ZH

Voir l'Annexe 2 pour le fichier modifié « `_flexidico.rnc` » qui représente un schéma validant la structure XML de notre dictionnaire en mandarin.

4.3 Typage sémantique des lexies et des actants

Dans le DiCoLiLex, chaque lexie et ses actants sont annotés avec une étiquette pour indiquer leur sens. Ces **étiquettes sémantiques** sont des « outils formels qui permettent de décrire de façon approximative le sens des lexies et de leurs actants » (Milićević, 1997). Par exemple, sous le lexème HONTE₁, nous trouverons d’abord son étiquette sémantique ‘sentiment négatif’ et ensuite les étiquettes de ses trois actants dans la forme propositionnelle : « honte de la personne *x* à propos de *y* devant la personne *z* ». Les étiquettes de ses actants indiquent que le 1^{er} et le 3^e actant *x* et *z* doivent d’être une personne, alors que l’étiquette du 2^e actant *y* n’est pas donnée, c’est-à-dire que son sens n’est pas contraint et peut être autant une entité qu’un fait. Par exemple, nous pouvons dire « la honte de cette étudiante face au professeur », alors que « la honte de cette fleur face à l’arbre » serait absurde ou poétique.

```
<étiquette-sémantique>sentiment négatif</étiquette-sémantique>
<forme-propositionnelle>
  <lexie-ref/> DE LA <étiquette-sémantique>personne</étiquette sémantique>
  <actant numéro="1">X</actant> À PROPOS DE <actant numéro="2">Y</actant>
  DEVANT LA <étiquette-sémantique>personne</étiquette-sémantique><actant
  numéro="3">Z</actant>
</forme-propositionnelle>
```

Figure 4. Description de la lexie **honte₁** dans le DiCoLiLex

Notre dictionnaire se concentre sur le champ sémantique des émotions, donc la plupart des lexies vedettes portent l’étiquette ‘sentiment’. Dans la hiérarchie des étiquettes sémantiques du DiCoPop³⁴, l’étiquette ‘sentiment’ est classée sous l’étiquette ‘état psychique’, et elle est divisée en trois étiquettes plus spécifiques : ‘sentiment positif’, ‘sentiment négatif’ et ‘émotion’. Nous empruntons la même hiérarchie et nous utilisons les étiquettes équivalentes en mandarin pour noter les lexies dans notre dictionnaire. Par exemple, les trois lexies les plus vagues de notre nomenclature, 感情 ^{gǎn qíng} ‘sentiment’, 情绪 ^{qíng xù} ‘émotion’, et 心 ^{xīn} ‘cœur’, sont associées à l’étiquette

³⁴ DiCOPop est une interface Web pour naviguer dans les données du DiCo et les consulter. La présentation des données permet un accès vulgarisé et la navigation propose l’utilisation d’une hiérarchie d’étiquettes sémantiques. Cf. <http://olst.ling.umontreal.ca/dicopop/hes.php> [consulté le 30 oct. 2018]

xīn lǐ zhuàng tài ‘état psychique’, et les autres sont divisées en trois groupes : zhèngmiàn qíng gǎn ‘sentiment positif’, fù miàn qíng gǎn ‘sentiment négatif’, et qíng gǎn ‘sentiment’ tout court pour les lexies qui expriment un sentiment ni positif ni négatif en mandarin, comme jí dòng ‘agitation’ ou yù wàng ‘désir’. Dans le DiCoPop et le DiCoLiLex, l’étiquette ‘émotion’ est utilisée pour les lexies qui expriment un sentiment plus momentané et impulsif, telles que *coup de foudre*, *effroi*, *émoi*. Par contre, nous n’avons pas adopté cette étiquette équivalente en mandarin dans notre dictionnaire, parce que la nuance entre les lexies qíng gǎn ‘sentiment’, gǎn qíng ‘sentiment’ et qíng xù ‘émotion’ n’est pas évidente dans cette langue : elles peuvent s’interchanger dans plusieurs contextes, et elles peuvent être le générique des autres lexies de sentiments plus spécifiques. Par exemple, bēi shāng de qíng gǎn ‘sentiment de tristesse’, bēi shāng de gǎn qíng ‘sentiment de tristesse’ et bēi shāng de qíng xù ‘émotion de tristesse’ expriment grosso modo le même sens. Selon Chen, Li et Hung (2009), qíng gǎn est une notion assez large en mandarin qui englobe gǎn qíng ‘émotion’, tài du ‘attitude’ et pǐn xìng ‘trait de caractère’. En conséquence, nous choisissons seulement l’étiquette qíng gǎn ‘sentiment’ pour englober les sens de sentiment et d’émotion en mandarin, sans considérer leur différence dans le domaine psychologique.

Ces étiquettes nominales peuvent directement s’utiliser pour les lexies nominales. Pour les lexies verbales, nous distinguons trois sortes d’étiquettes selon leur sens :

- 1) gǎn dào yì zhǒng zhèngmiàn / fùmiàn qíng gǎn ‘éprouver un sentiment (positif/négatif)’
- 2) biǎoxiàn chū yì zhǒng zhèngmiàn / fùmiàn qíng gǎn ‘exprimer un sentiment (positif/négatif)’
- 3) yǐn qǐ yì zhǒng zhèngmiàn / fùmiàn qíng gǎn ‘provoquer un sentiment (positif/négatif)’

Dans notre dictionnaire, tous les vocables verbaux contiennent au moins une lexie avec la première étiquette, qui est un équivalent de l’étiquette sémantique nominale ‘sentiment’ formée par la FL **Oper₁** : **Oper₁**(*sentiment*) = éprouver (Milićević, 1997). Certains vocables ont également des lexies avec la deuxième et/ou la troisième étiquette, par exemple, les trois bēi shāng ‘être.triste’ en (5) correspondent respectivement aux trois lexies polysémiques avec les trois étiquettes verbales ci-dessus.

- (5) a. ^{bēi shāng} 悲伤 ^{de} 的 ^{rén} 人 b. ^{bēi shāng} 悲伤 ^{de} 的 ^{yǎn shén} 眼神 c. ^{bēi shāng} 悲伤 ^{de} 的 ^{gù shì} 故事
 être.triste DE₁ personne être.triste DE₁ regard être.triste DE₁ histoire
 ‘personne triste’ ‘regard triste’ ‘histoire triste’

La lexie ^{bēi shāng} 悲伤₁ au sens ‘éprouver un sentiment négatif’ en (5)a possède un actant *x* avec l’étiquette ^{gè tǐ} 个体 ‘individu’, tandis que l’actant *x* de la lexie ^{bēi shāng} 悲伤₃ ‘provoquer un sentiment négatif’ en (5)c est associé à l’étiquette ^{shì jiàn} 事件 ‘fait’. Quant à la lexie ^{bēi shāng} 悲伤₂ ‘exprimer un sentiment négatif’ en (5)b, il n’est pas facile d’étiqueter l’actant *x*, car cette lexie peut prendre comme 1^{er} actant les noms liés à une apparence ou un mode d’expression, tels que ^{miànróng} 面容 ‘mine’, ^{shén sè} 神色 ‘air’, ^{mù guāng} 目光 ‘regard’, ^{shēng yīn} 声音 ‘voix’, ^{wēi xiào} 微笑 ‘sourire’, ^{yǎn lèi} 眼泪 ‘larme’. Cependant, dans le DiCoLiLex et le DicoPop, nous n’avons pas trouvé d’étiquette sémantique qui puisse englober tous ces noms. Par conséquent, nous proposons l’étiquette ^{biǎoqíng} 表情 ‘expression’ pour ce type d’actant qui peut exprimer le sentiment d’une personne.

4.4 Difficultés lexicographiques

Pendant la rédaction de notre dictionnaire, nous avons rencontré des difficultés de divers types. Dans cette section, nous discuterons ces difficultés cas par cas.

4.4.1 Mots élastiques

En ce qui concerne la longueur des syllabes des mots-formes, il existe trois sortes de mots³⁵ en mandarin moderne, les mots purement monosyllabiques, par exemple, ^{shuǐ} 水 ‘eau’, ^{ài} 爱 ‘aimer’, les mots purement polysyllabiques qui sont constitués des morphes polysyllabiques, par exemple, ^{pú táo} 葡萄 ‘raisin’, ^{shā fā} 沙发 ‘sofa’, ainsi que les mots dits « élastiques », qui possèdent une forme courte monosyllabique et une forme longue polysyllabique (le plus souvent dissyllabique) ayant le même sens (Dong, 2015; Duanmu et Dong, 2016). L’existence des mots élastiques en mandarin est bien connue dans la littérature (Chao, 1948; Duanmu, 2007, 2013; S. Guo, 1938;

³⁵ Ici, nous utilisons volontairement le mot vague « mot » avant de déterminer le statut des « mots élastiques ».

Karlgren, 1923). Dans sa thèse, Dong (2015) liste tous les mots élastiques (7 387 au total) du 《现代汉语词典》 (*Dictionnaire du mandarin moderne*) et montre que plus que la moitié sont élastiques. À partir de sa liste, nous avons dressé la liste des mots élastiques qui apparaissent dans notre nomenclature chinoise en ajoutant des gloses littérales.³⁶

Forme courte	Forme longue	PdD	Autonome
傲 ‘orgueil’	骄傲 ‘fier-orgueil’	Adj	Oui
悲 ‘triste’	悲伤 ‘triste-affligé’	Adj	Non
耻 ‘honte’	耻辱 ‘honte-humiliation’	N	Non
耻 ‘honte’	羞耻 ‘honte-honte’	N	Non
醋 ‘vinaigre’	酸醋 ‘acide-vinaigre’	N	Oui
火 ‘colère’	发火 ‘lancer-colère’	V	Non
苦 ‘amer’	痛苦 ‘douloureux-amer’	Adj	Oui
恋 ‘aimer’	恋爱 ‘aimer-aimer’	N	Non
恼 ‘tracassé’	烦恼 ‘ennuyé-tracassé’	Adj	Non
恼 ‘contrarié’	恼怒 ‘contrarié-fâché’	V	Oui
怒 ‘fâché’	愤怒 ‘indignation-fâché’	V	Oui
怕 ‘peur’	害怕 ‘produire-peur’	V	Oui
气 ‘colère’	生气 ‘générer-colère’	V	Oui
情 ‘sentiment’	感情 ‘ressentir-sentiment’	N	Non
情 ‘sentiment’	爱情 ‘aimer-sentiment’	N	Non
辱 ‘humiliation’	耻辱 ‘honte-humiliation’	N	Non
辱 ‘humilier’	侮辱 ‘insulter-humilier’	V	Non
痛 ‘douloureux’	悲痛 ‘triste-douloureux’	Adj	Non
喜 ‘aimer’	喜爱 ‘aimer-aimer’	V	Non
欲 ‘désir’	欲望 ‘désir-espoir’	N	Non

Tableau VIII. Mots élastiques dans notre nomenclature chinoise

³⁶ Nous ne sommes pas d'accord avec toutes ces formes; par exemple, nous considérons les composantes verbales dans les formes longues telles que 发火, 害怕 et 生气 comme des verbes supports (cf. §4.4.2, p. 58).

Comme les deux formes partagent une même composante et possèdent le même sens, plusieurs linguistes les considèrent comme des variantes du même mot (Karlsgren 1918, 1923, Guo 1938, Chao 1948, cités par Duanmu et Dong, 2016). Mais, le terme « mot » lui-même est vague, alors qu'il est important de spécifier le statut morphologique de ces mots élastiques dans le cadre de TST. Pour les formes courtes, comme le montre la dernière colonne du tableau ci-dessus, les mots monosyllabiques peuvent être séparés en deux catégories : les morphes autonomes et les morphes non autonomes (*free/bound* dans la terminologie de Dong, 2015). Alors, il faut se poser deux questions :

- 1) Si la forme courte n'est pas autonome, est-ce qu'elle peut être un mot-forme ?
- 2) S'il s'agit d'un mot-forme court, ce mot-forme court et le mot-forme long ayant le même sens peuvent-ils appartenir à un seul lexème ?

Pour la première question, il faut vérifier si une forme courte d'un mot élastique peut être un mot-forme (§2.2). À part la catégorie autonome, les autres mots monosyllabiques du Tableau VIII ne peuvent pas être utilisés seuls dans un énoncé, c'est-à-dire entre deux pauses. Il faut ajouter un autre morphe pour former un mot-forme dissyllabique. Donc ce sont des morphes lexicaux. Cependant, les morphes monosyllabiques chinois sont très différents des affixes ou des radicaux en français et en anglais, parce que ces derniers ont besoin d'un morphe d'un côté spécifique (Duanmu, 2012). Par exemple, *théo-* a besoin d'un morphe à sa droite, et *-logie* a besoin d'un morphe à sa gauche. Par contre, en mandarin, il n'y a pas cette contrainte pour la plupart des radicaux lexicaux. Par exemple, le morphe 傲^{ào} 'orgueilleux' peut accepter un autre morphe de n'importe quel côté pour former un mot-forme : 骄傲^{jiāo ào} 'orgueilleux' (litt. 'fier-orgueilleux') et 傲慢^{ào màn} 'arrogant' (litt. 'orgueilleux-arrogant'). Par conséquent, Duanmu (2012) propose que la raison du manque d'autonomie des morphes monosyllabiques en mandarin est phonologique au lieu de morphologique, c'est-à-dire que les mots-formes ont besoin d'être dissyllabiques. Ainsi, il considère que les morphes courts sont comme des mots. Néanmoins, il existe bien des mots-formes monosyllabiques et il n'est pas toujours nécessaire de former un mot-forme dissyllabique. En conséquence, nous considérons que les formes courtes des « mots

élastiques » sont bien des morphes lexicaux, dont certains sont autonomes et peuvent s'utiliser comme un mot-forme.

Toutefois, parmi les morphes monosyllabiques non autonomes dans le Tableau VIII, ceux qui ont une partie du discours nominale peuvent avoir une certaine autonomie : même s'ils ne peuvent s'utiliser entre deux pauses, ils peuvent s'utiliser comme gouverneur dans un syntagme nominal avec un modificateur et la préposition 之^{zhī}, par exemple :

- | | | | | | | | | | | | |
|--------|-------------------------|-----|-------|----|--------------------|-----|-------|----|-------------------------------|-----|-----------|
| (6) a. | guó jiā | zhī | chǐ | b. | jīngshén | zhī | liàn | c. | gǎn jī | zhī | qíng |
| | 国家 | 之 | 耻 | | 精神 | 之 | 恋 | | 感激 | 之 | 情 |
| | nation | ZHI | honte | | esprit | ZHI | amour | | reconnaissance | ZHI | sentiment |
| | 'humiliation nationale' | | | | 'amour platonique' | | | | 'sentiment de reconnaissance' | | |

Ainsi, nous considérons ces signes comme des mots-formes au lieu de morphes dans ces constructions particulières. Par contre, ce sont bien des morphes dans les autres cas, y compris dans leurs formes longues correspondantes, cf. (7). Il est à noter que le mot-forme long d'un mot élastique est différent d'un vrai mot-forme composé, cf. (7), parce que le sens du morphe facultatif est souvent redondant, cf. (8), ou vide, cf. (9). Ainsi, la partie facultative du mot-forme long ne contribue pas vraiment au sens complet, c'est pourquoi Duanmu (2007) utilise le terme « pseudo-composés » (*pseudo-compounds*) pour ces mots-formes longs.

- | | | | | | |
|--------|-------------------|----|-----------------|----|---------------------|
| (7) a. | chǐ xiào | b. | liànqíng | c. | wú qíng |
| | 耻笑 | | 恋情 | | 无情 |
| | honte-rire | | amour-sentiment | | sans-sentiment |
| | 'se moquer' | | 'amour' | | 'sans-cœur' |
| (8) a. | chǐ rǔ | b. | liàn ài | c. | gǎn qíng |
| | 耻辱 | | 恋爱 | | 感情 |
| | honte-humiliation | | amour-amour | | ressentir-sentiment |
| | 'humiliation' | | 'amour' | | 'sentiment' |
| (9) a. | lǎo hǔ | b. | dà suàn | c. | xiāng yān |
| | 老虎 | | 大蒜 | | 香烟 |
| | vieux-tigre | | grand-ail | | parfumé-cigarette |
| | 'tigre' | | 'ail' | | 'cigarette' |

Quant à la deuxième question, à savoir si un mot-forme court et un mot-forme long appartiennent au même lexème, il y a deux raisons de croire que non. Premièrement, selon la définition du lexème (cf. §2.1, p. 6), les mots-formes qui appartiennent à un même lexème ne se distinguent que par la flexion. Cependant, comme nous avons expliqué plus haut, le morphe facultatif que le mot-forme long possède est souvent redondant ou vide au sens et il ne fait partie

d'aucune catégorie flexionnelle en mandarin, donc ce n'est pas un morphe flexionnel, donc les formes courtes et longues des mots élastiques ne se distinguent pas par la flexion.

Deuxièmement, même si les mots-formes de longueur différente sont interchangeables dans certains contextes, leur combinatoire n'est pas toujours identique. Le choix des formes courtes ou longues est souvent contraint par les autres mots-formes dans le contexte. Par exemple, les mots-formes courts 耻^{chǐ} et 辱^{rǔ} ainsi que leur combinaison 耻辱^{chǐ rǔ} ont tous le même sens 'honte, humiliation'. Pour exprimer le premier actant (par exemple, 'l'humiliation de la nation'), nous pouvons utiliser les trois mots-formes, cf. (10)a-d. Cependant, les deux expressions en (10)e-f sont agrammaticales parce que les deux mots-formes courts ne s'utilisent jamais avec la particule structurale 的^{de}, mais seulement avec 之^{zhī}, qui a la même fonction grammaticale (les deux lient un nom à un modificateur). La différence est que la particule 之^{zhī} donne un style plus formel et littéraire parce que c'est un mot-forme issu du chinois classique, alors que la particule 的^{de} provient du registre oral. Étant donné que ces mots-formes courts eux-mêmes sont aussi plus formels et gardent le style concis du chinois classique, ils ne peuvent s'utiliser qu'avec la particule 之^{zhī}. En outre, il existe aussi un « mot-valise », en (10)g, qui prend les deux formes courtes 国家^{guó jiā} et 耻辱^{chǐ rǔ} en supprimant les parties facultatives à droite, alors que (10)h est bizarre (nous avons trouvé seulement une occurrence dans le corpus de BCC), et (10)i, avec la forme longue, est agrammatical. Tous ces faits montrent que les mots-formes courts et longs ont des syntactiques différents.

- | | | |
|---|---|---|
| (10) a. 国(家) 之 耻
nation ZHI honte
'humiliation nationale' | b. 国家 之 辱
nation ZHI humiliation
'humiliation nationale' | c. 国家 的 耻辱
nation DE ₁ honte-humiliation
'humiliation nationale' |
| d. 国(家) 之 耻辱
nation ZHI honte-humiliation
'humiliation nationale' | e.* 国家 的 耻
nation DE ₁ honte
'humiliation nationale' | f.* 国家 的 辱
nation DE ₁ humiliation
'humiliation nationale' |
| g. 国耻
nation-honte
'humiliation nationale' | h.? 国辱
nation-honte
'humiliation nationale' | i.* 国耻辱
nation-honte-humiliation
'humiliation nationale' |

Par ailleurs, les formes courte et longue d'un même mot élastique ont souvent des collocations différentes. Par exemple, le mot-forme monosyllabique 怒^{nù} 'colère' a un verbe

support 发^{fā} ‘lancer’, tandis que le mot-forme dissyllabique 愤怒^{fèn nù} n’accepte pas ce même verbe support, cf. (11). Ces deux mots-formes possèdent aussi différents modificateurs pour exprimer l’intensification, cf. (12).

- | | | | |
|---------|--|----|--|
| (11) a. | 发 ^{fā} 怒 ^{nù}
lancer colère
‘être fâché’ | b. | *发 ^{fā} 愤怒 ^{fèn nù}
lancer indignation-colère
‘être fâché’ |
| (12) a. | 大 ^{dà} 怒 ^{nù}
grand colère
‘grande colère’ | b. | 强烈 ^{qiáng liè} 的 ^{de} 愤怒 ^{fèn nù}
fort-intense DE ₁ indignation-colère
‘grande colère’ |

Par conséquent, les formes courte et longue d’un même mot élastique sont des synonymes qui partagent le même signifié, mais qui ont des signifiants et syntactiques différents ainsi que des collocations différentes. Ils appartiennent à des lexèmes différents.

4.4.2 Signes complexes de la forme « V?N »

Dans notre nomenclature chinoise, il existe plusieurs expressions phraséologisées de la forme « Verbe?Nom » [= « V?N »], cf. (13). Étant donné qu’il n’y a pas d’espace entre les graphèmes chinois, parfois il n’est pas évident de distinguer si une expression phraséologisée est un mot-forme complexe ou un syntagme, ce que représente la notation « X?Y ». Toutefois, afin de bien décrire les formes propositionnelles et les collocations de ces expressions, il est important d’en déterminer le statut morphologique.

- | | | | | | |
|---------|--|----|---|----|---|
| (13) a. | 吃 ^{chī} 醋 ^{cù}
manger vinaigre
‘être jaloux’ | b. | 担 ^{dān} 心 ^{xīn}
porter cœur
‘s’inquiéter’ | c. | 害 ^{hài} 怕 ^{pà}
produire peur
‘avoir peur’ |
| d. | 发 ^{fā} 火 ^{huǒ}
lancer feu
‘être fâché’ | e. | 生 ^{shēng} 气 ^{qì}
générer air
‘être fâché’ | | |

Dans la littérature traditionnelle chinoise (Chao, 1965; C.-T. J. Huang, 1984; Li et Thompson, 1989; Packard, 2000), ces expressions sont souvent considérées comme des mots-formes composés, mais ce n’est pas toujours vrai, parce que les mots-formes sont par définition inséparables (Mel’čuk, 1993), tandis que plusieurs de ces expressions sont séparables. Par

exemple, ces cinq expressions peuvent toutes être séparées par les suffixes aspectuels 了^{le}, qui marque un accomplissement de l'action, ou 着^{zhe}, qui marque une continuation de l'action, cf. (14). De plus, nous pouvons insérer un mot-forme ou un syntagme dans certaines expressions, cf. (15). Par conséquent, ce sont clairement des syntagmes.

- (14) a. 他 吃了 醋。
 lui manger-PERF vinaigre
 'Il a été jaloux.'
- b. 我 替 她 担着 心。
 moi pour elle porter-PROG cœur
 'Je m'inquiète pour elle.'
- c. 她们 害了 怕。
 elles produire-PERF peur
 'Elles ont eu peur.'
- d. 他 依然 发着 火。
 lui encore lancer-PROG feu
 'Il est encore fâché.'
- e. 他 现在 正 生着 气。
 lui maintenant justement générer-PROG air
 'Il est en colère en ce moment.'
- (15) a. 他 吃 你的 醋。
 lui manger toi DE₁ vinaigre
 'Il est jaloux de toi.'
- b. 他 发了 很 大 的 火。
 lui lancer-PERF très grand DE₁ feu.
 'Il a été très fâché.'

Maintenant, pour les décrire dans le dictionnaire, il faut déterminer si ce sont des locutions ou des collocations, en nous appuyant sur la compositionnalité des signes et leur figement (cf. §2.3, p. 16). S'il s'agit d'une locution, elle va avoir sa propre entrée, cf. (13)a et (13)b. Si c'est une collocation, nous allons les encoder avec la fonction lexicale qui lui convient, par exemple en (13)c, le verbe 害^{hài} est clairement un verbe support du nom 怕^{pà}, ainsi, nous l'encodons avec la fonction lexicale **Oper₁** sous l'entrée du lexème 怕^{pà}.

Cependant, il y a des cas moins évidents, tels (13)d et (13)e. Au premier coup d'œil, ces deux expressions ressemblent à des locutions parce que leur signifié ne semble pas compositionnel : 'lancer' ⊕ 'feu' ≠ 'être fâché' et 'générer' ⊕ 'air' ≠ 'être fâché'. En effet, dans la thèse de Nguyen (2006, p. 320), (13)e est traité comme une locution verbale qualificative. Néanmoins, nous les considérons plutôt comme des collocations, parce que les deux noms 火^{huǒ} et 气^{qì} possèdent en fait un sens figuré pour exprimer un sentiment de colère. Imaginons une scène où une personne est tellement en colère que c'est comme s'il y avait du feu ou de l'air qui en sortait. Ils peuvent ainsi s'utiliser seuls comme des mots-formes pour exprimer le sens de colère dans les phrases comme (16). Ils peuvent également être combinés avec d'autres verbes supports

comme en (17) pour exprimer le sens ‘être fâché’, et ils peuvent accepter leurs propres modificateurs dans ces syntagmes, cf. (15)b. En plus, ils peuvent s’utiliser comme verbes pour exprimer le même sens, cf. (18). Donc, nous avons affaire à deux copolysèmes 火_{n1} ‘feu’ et 火_{n2} ‘colère’, avec un lien métaphorique, et ils appartiennent au même vocable nominal 火_n, ainsi que deux copolysèmes 气_{n1} ‘air’ et 气_{n2} ‘colère’, qui font partie du vocable nominal 气_n. Les syntagmes collocationnels en (13)d et (13)e vont respectivement être décrits sous les entrées 火_v et 气_v. Nous avons ainsi deux autres vocables verbaux, 火_v et 气_v pour encoder les deux verbes en (18).

- (16) a. 我 现在 火 很 大。
 moi maintenant colère très grand
 ‘Je suis très fâché maintenant.’
 b. 他 的 气 消 了。
 lui DE₁ colère se.dissiper-PERF
 ‘Sa colère s’est dissipée.’
- (17) a. 挂 火
 accrocher feu
 ‘être fâché’
 b. 动 火
 toucher feu
 ‘être fâché’
 c. 动 气
 toucher air
 ‘être fâché’
- (18) a. 他 火 了。
 lui se.fâcher-PERF
 ‘Il est fâché.’
 b. 他 好 气。
 lui très se.fâcher
 ‘Il est très fâché.’

À part ces syntagmes locutionnels et collocationnels que nous avons discutés ci-dessus, il existe aussi des expressions phraséologisées de la forme « V?N » qui sont des mots-formes complexes, cf. (19), parce qu’elles répondent aux critères des mots-formes (cf. Nguyen, 2006). Par exemple, les constituants de ces signes ne peuvent être séparés par d’autres signes, même pas par un suffixe, cf. (20), et les constituants ne sont pas transmutables. Ainsi, ce sont bien des mots-formes complexes. Mais ces expressions peuvent s’utiliser soit comme un nom soit comme un verbe, donc nous les décrivons séparément dans les vocables nominaux 开心_n, 好奇_n, 遗憾_n et les vocables verbaux 开心_v, 好奇_v, 遗憾_v.

- (19) a. 开心
 ouvrir-cœur
 ‘joie’/‘être joyeux’
 b. 好奇
 aimer-nouveauté
 ‘curiosité’/‘être curieux’
 c. 遗憾
 laisser-regret
 ‘regret’/ ‘regretter’
- (20) a. *他 开 着 心
 lui ouvrir-PROG-cœur
 b. *他 好 着 奇
 lui aimer-PROG-nouveauté
 c. *他 遗 着 憾
 lui laisser-PROG-regret

Finalement, il existe certains cas très particuliers où une expression phraséologisée se comporte tantôt comme un mot-forme, tantôt comme un syntagme. Ici, nous reprenons les exemples de Nguyen (2006, p. 148). En (21)a, ^{dān xīn}担心 ressemble beaucoup à un mot-forme parce qu'il est inséparable quand il est accompagné d'un COD : si nous voulons ajouter un suffixe aspectuel -^{zhe}着, il faut l'ajouter après ^{dān xīn}担心, pas au milieu. Par contre, en (21)b, nous avons clairement un syntagme locutionnel ^{dān xīn}担心 qui peut être séparé par d'autres signes. En même temps, ces deux signes peuvent être utilisés dans le même contexte, comme en (21)c. Ainsi, Nguyen (2006, p.138) a postulé l'existence de deux signes différents « qui devront être inclus, l'un dans un lexème, l'autre dans une locution, et ces deux lexies devront figurer dans le dictionnaire en entrées séparées ».

- (21) a. ^{wǒ}我 ^{dān xīn zhe}担心(着) ^{zhè jiàn shì}这件事 ^{shì}事。
 moi porter-cœur(-PROG) ce CL affaire
 'Je m'inquiète pour cette affaire.'
- b. ^{wǒ}我 ^{dān le}担了 ^{sān nián de}三年的 ^{xīn}心。
 moi porter-PERF trois année DE₁ cœur
 'Je me suis inquiété pendant trois ans.'
- c. ^{wǒ}我 ^{dān xīn}担心 ^{zhè jiàn shì}这件事, ^{dān le}担了 ^{sān nián de}三年的 ^{xīn}心 ^{le}了。
 moi porter-cœur ce CL affaire porter-PERF trois année DE₁ cœur LE₂
 'Je m'inquiète pour cette affaire, et m'en inquiète depuis trois ans.'

Néanmoins, nous ne pensons pas que c'est une bonne solution parce que cela est trop lourd pour un dictionnaire d'avoir deux entrées pour « deux » signes qui ont exactement les mêmes signifié et signifiant, ainsi que des collocations presque pareilles, sauf quelques différences dans leur combinatoire : l'un accepte un COD, l'autre un COI, mais ces compléments expriment le même actant sémantique, cf. (24)a et (26)³⁷. Alors, il faut se demander si ce sont vraiment deux signes différents. Sans aucun doute, il existe bien un syntagme ^{dān xīn}担心, car il satisfait le critère de séparabilité, cf. (21)b et (14)b. Concernant l'hypothèse d'un mot-forme complexe ^{dān xīn}担心, Nguyen (2006) a postulé son existence pour trois

³⁷ Nous discuterons un peu plus loin dans cette section de leur comportement syntaxique différent selon la séparabilité des deux caractères en question.

raisons, mais nous avons trouvé des contre-exemples pour prouver qu’il y a seulement un syntagme ^{dān xīn}担心.

D’abord, Nguyen (2006) suppose que la construction verbale résultative de la forme «V₁ N V₂» est impossible en mandarin (cf. §4.4.3, p. 64), donc pour expliquer que la construction en (22) est possible, il faut admettre que ^{xīn}心 fait partie du mot-forme ^{dān xīn}担心. Cependant, nous avons trouvé des phrases dans le corpus BCC où le V₁ est clairement une locution, cf.(13)a, ou une collocation, cf. (13)e. En (23)a, le mot-forme nominal ^{cù}醋 fait partie du syntagme locutionnel, ^{chī cù}吃醋 mais il est bien situé entre les deux verbes. De la même façon, en (23)b, le mot-forme nominal ^{qì}气 fait partie du syntagme collocationnel ^{shēng qì}生气 et il est aussi situé entre son verbe support ^{shēng}生 et V₂. Par conséquent, la construction « V₁ N V₂ » est bel et bien possible, ce n’est donc pas une bonne raison pour postuler l’existence d’un mot-forme ^{dān xīn}担心.

(22) ^{wǒ}我 ^{dān xīn}担?心 ^{sǐ}死³⁸ ^{nǐ}你 ^{le}了。
 moi porter?cœur mourir toi LE₂
 ‘Je suis mort d’inquiétude pour toi.’

(23) a. ^{wǒ}我 ^{chī}吃 ^{cù}醋 ^{sǐ}死 ^{le}了。
 moi manger vinaigre mourir LE₂
 ‘Je suis mort de jalousie.’

b. ^{wǒ}我 ^{shēng}生 ^{qì}气 ^{qǐ}起 ^{lái}来³⁹。
 moi générer air se.lever venir
 ‘Je commence à me fâcher.’

La deuxième raison de Nguyen (2006) pour conclure que ^{dān xīn}担心 en (24)a est un mot-forme est que normalement un suffixe aspectuel ne peut pas se placer après un nom, cf. (24)b-c, ainsi « pour expliquer la présence du suffixe aspectuel ^{zhe}-着 après N, il faut admettre que N est une partie de mot-forme et que V-N est un mot-forme composé ». Néanmoins, nous avons

³⁸ En fait, dans la section §4.4.4, p. 73, nous montrerons que le mot-forme ^{sǐ}死 fonctionne comme un adverbe au lieu d’un verbe.

³⁹ Nous allons voir plus loin que ^{qǐ lái}起来 est une locution qui fonctionne comme V₂ dans les CVSR et qui peut s’intercaler avec les noms dans V₁. Voir la section suivante (§4.4.3) pour plus de détails sur son comportement.

encore une fois trouvé des contre-exemples dans le corpus qui montrent que -^{zhe}着 peut bien suivre un nom qui fait partie d'une locution ou une collocation, cf. (25).

(24) a. 我 担?心着 你。
 wǒ dān xīn zhe nǐ
 moi porter?cœur-PROG toi
 'Je m'inquiète pour toi.'

b. 他 爱着 你。
 tā ài zhe nǐ
 lui aimer-PROG toi
 'Il est amoureux de toi'

c. *他 爱 你着。
 tā ài nǐ zhe
 lui aimer toi-PROG
 'Il est amoureux de toi.'

(25) a. 他 吃 醋着。
 tā chī cù zhe
 lui manger vinaigre-PROG
 'Il est jaloux.'

b. 他 生 气着。
 tā shēng qì zhe
 lui générer air-PROG
 'Il est fâché.'

Finally, quand ^{dān}担 et ^{xīn}心 sont séparés, l'ajout d'un COD est impossible, cf. (26)a-b. Quand ils ne sont pas séparés, l'ajout d'un objet devient possible, cf. (24)a. Selon Nguyen (2006), le fait que ^{dān}担? ^{xīn}心 possède ces comportements syntaxiques différents renforce son hypothèse qu'il existe deux signes complexes ^{dān}担? ^{xīn}心 dont le sens est identique : un mot-forme ^{dān}担 ^{xīn}心, cf. (24)a et un syntagme ^{dān}担 ^{xīn}心, cf. (26)a-b. Cependant, nous pensons que ces faits ne sont pas suffisants pour postuler l'existence d'un mot-forme ^{dān}担 ^{xīn}心. Nous pouvons simplement supposer qu'il n'existe qu'un syntagme ^{dān}担 ^{xīn}心, mais qu'il est sujet à certaines contraintes quand il est séparé : nous ne pouvons pas ajouter son complément d'objet directement après ce syntagme en tant que COD, mais nous pouvons bien l'ajouter en avant en tant que COI, comme en (26)c.

(26) a. *我 担了 心 你。
 wǒ dān le xīn nǐ
 moi porter-PERF cœur toi
 'Je me suis inquiété pour toi.'

b. *我 担着 心 你
 wǒ dān zhe xīn nǐ
 moi porter-PROG cœur toi
 'Je m'inquiète pour toi.'

c. 我 为 你 担着 心。
 wǒ wéi nǐ dān zhe xīn
 moi pour toi porter-PROG cœur
 'Je m'inquiète pour toi.'

Ainsi, nous utilisons le terme *locution verbale* pour décrire la partie du discours de la lexie ^{dān}担 ^{xīn}心 's'inquiéter' dans notre dictionnaire. Il est à noter qu'il existe également une lexie

nominale 担心^{dān xīn} ‘inquiétude’ qui est une nominalisation de ce verbe : $S_0(\text{担心}_n) = \text{担心}_v$ ⁴⁰. Néanmoins, il n’existe pas de moyen en mandarin pour tester la séparabilité de cette sorte de signe nominal comme nous l’avons fait pour les verbes en insérant les marqueurs aspectuels. En conséquence, nous ne pouvons pas distinguer s’il s’agit d’un mot-forme ou d’un syntagme. Puisque la composante verbale 担^{dān} ‘porter’ perd totalement sa fonction verbale dans la lexie nominale 担心^{dān xīn} ‘inquiétude’, cette lexie en question peut simplement être considérée comme un lexème au lieu d’une locution nominale. Pour les autres lexies nominales converties à partir d’une locution verbale à la forme « V N », nous les traitons également comme des lexèmes nominaux simples.

4.4.3 Constructions verbales sérielles résultatives (= CVSR)

Pour certaines lexies dans notre dictionnaire, nous avons rencontré des constructions particulières ressemblant à une collocation, mais où la base et le collocatif sont tous les deux des verbes (ce qui est relativement rare en français), comme les expressions 爱上^{ài shàng} ‘tomber amoureux’ et 感到^{gǎn dào} ‘ressentir’ en (27) ci-dessous.

- (27) a. 我 爱 上 了 他。 b. 我 感 到 悲 伤。
- moi aimer commencer-PERF lui moi ressentir atteindre tristesse
‘Je suis tombé(e) amoureux(euse) de lui’ ‘Je ressens la tristesse.’

Nguyen (2006) appelle ce genre de construction constituée de deux verbes consécutifs non coordonnés les **constructions verbales sérielles résultatives** (= CVSR)⁴¹, car le deuxième verbe exprime souvent un résultat du premier verbe, cf. (27)b et(28).

⁴⁰ Dans notre dictionnaire, nous utilisons les notations « _v » , « _n » pour différencier les vocables n’ayant pas la même partie du discours.

⁴¹ Il est à noter que cette construction est une sous-catégorie de la construction de « série verbale » ou « prédicat complexe » dans la littérature (Aikhenvald et Dixon, 2005; Delplanque, 1998; Müller, 2002), mais elle est plus

- (28) a. 他 吃 饱了。 b. 他 考 上了 大学。
 lui manger être.rassasié-PERF lui passer.examen réussir-PERF université
 ‘Il est rassasié.’ ‘Il est admis à une université.’

Il existe aussi des constructions spécifiques où V₂ indique la direction de l’action exprimée par V₁, cf. (29). Dans la littérature de la grammaire chinoise traditionnelle, ce genre de construction est appelé 动趋式 ‘construction verbale directionnelle’ et est souvent traité séparément de la construction 动结式 ‘construction verbale résultative’, et le V₂ dans les constructions directionnelles est ainsi appelé verbe directionnel (Huang et Liao, 2002; Lü, 1999; Wang et al., 1993). Toutefois, nous pouvons bien considérer cette direction de déplacement comme un résultat de l’action V₁, ainsi Nguyen (2006) les classe comme CVSR de déplacement.

- (29) a. 他 走 进 教室。 b. 风筝 飞 上 天空。
 lui marcher entrer classe cerf-volant voler monter ciel
 ‘Il entre (en marchant) dans la classe.’ ‘Le cerf-volant s’envole dans le ciel.’

En outre, il y a certaines constructions où V₂ exprime un changement d’état de V₁, si V₁ est un verbe statique, cf. (27) a et (30). L’interprétation du sens de V₂ peut être différente selon les linguistes chinois : Chen (1994) pense que V₂ 上 ‘commencer’ dans la CVSR 爱上 ‘tomber amoureux’ exprime un sens de commencement et de continuation, tandis que Liu (1998) le considère comme un sens de résultat. Alors, selon Dong (2017), le changement d’état peut être considéré soit comme un commencement, soit comme un accomplissement, donc en (30)b, 喜欢上 欢上 peut s’interpréter soit comme le commencement et la continuation de l’état d’aimer, soit un accomplissement du changement d’état de ‘ne pas aimer’ à ‘aimer’.

contrainte que la série verbale en général, parce que dans cette dernière construction en mandarin, le premier verbe peut être suivi par son complément. Par exemple, dans la phrase ci-dessous, nous trouvons le complément du premier verbe 去 ‘aller’ entre les deux verbes. Dans la CVSR, par contre, les deux verbes sont plus liés et nous ne pouvons pas insérer d’autres lexies au sens plein.

- 我 去 学校 写 论文。
 moi aller école écrire mémoire.
 ‘je vais à l’école pour écrire mon mémoire.’

- (30) a. 他 ^{tā} 想 ^{xiǎng} 起 ^{qǐ} 了 ^{le} 往 ^{wǎng} 事 ^{shì}。
 lui se.rappeler commencer-PERF passé.chose
 ‘Il commence à se rappeler du passé.’
- b. 他 ^{tā} 喜 ^{xǐ} 欢 ^{huan} 上 ^{shàng} 了 ^{le} 摇 ^{yáo} 滚 ^{gǔn} 乐 ^{yuè}。
 lui aimer commencer-PERF rock.musique
 ‘Il commence à aimer la musique rock.’

Dans cette perspective, le V₂ exprime le résultat d’accomplissement de V₁, donc ce type de construction peut être classé dans les CVSR. De plus, Chung (2006) utilise le terme *inchoative aspect* (aspect inchoatif) pour désigner tout *initiator aspect* (aspect initiateur) ou *realization aspect* (aspect de réalisation) qui exprime le commencement d’une action, et selon elle, ces deux derniers contrastent en mandarin : dans les verbes composés, l’aspect initiateur se trouve à la position initiale d’une composante de verbe aspectuel suivi par un complément qui décrit l’action spécifique, par exemple, 起 ^{qǐ} 舞 ^{wǔ} ‘commencer-danser’; alors que l’aspect d’accomplissement est exprimé dans les constructions résultatives et directionnelles, cf. 想 ^{xiǎng} 起 ^{qǐ} ‘se rappeler’ en (30)a. En outre, les V₂ de ce type ayant un sens inchoatif sont des lexies grammaticalisées de verbes directionnels du 2^e type de CVSR, tels que 上 ^{shàng} ‘monter’, 下 ^{xià} ‘descendre’, 起 ^{qǐ} ‘lever’, 起 ^{qǐ} 来 ^{lái} ‘lever-venir’. Ainsi, nous utilisons le terme CVSR pour englober ces trois sortes de constructions dans ce mémoire.

Au premier coup d’œil, les CVSR peuvent ressembler aux mots-formes composés. Mais selon l’analyse de Nguyen (2006, p. 168-171), les CVSR sont bien des syntagmes pour principalement deux raisons. Premièrement, selon le critère de séparabilité du signe, puisque les constituants du CVSR sont séparables par un clitique. Par exemple, en (31), nous pouvons insérer respectivement les clitics atones 得 ^{de} ‘pouvoir’ et 不 ^{bu} ‘ne pas pouvoir’ entre les deux verbes 爱 ^{ài} ‘aimer’ et 上 ^{shàng} ‘commencer’. Deuxièmement, le premier verbe peut garder des modificateurs, donc en (32), 爱 ^{ài} ‘aimer’ peut être modifié par son **Magn** 深 ^{shēn} 深 ^{shēn} 地 ^{de} ‘profondément’.

- (31) a. 他 ^{tā} 爱 ^{ài} 得 ^{de} 上 ^{shàng} 你 ^{nǐ}。
 lui aimer DE₄ commencer toi
 ‘Il peut tomber amoureux de toi’
- b. 他 ^{tā} 爱 ^{ài} 不 ^{bu} 上 ^{shàng} 你 ^{nǐ}。
 lui aimer BU₂ commencer toi
 ‘Il ne peut pas tomber amoureux de toi.’

- (32) 我 深深 地 爱 上了 他。
 moi être.profond-être.profond DE₂ aimer commencer-PERF lui
 ‘Je suis profondément tombé(e) amoureux(euse) de lui.’

Zhu (1982) indique que les CVSR fonctionnent comme des verbes, et peuvent accepter les suffixes d’aspect tels que 了^{le} et 过^{guò}, donc elles ressemblent beaucoup aux verbes composés de type prédicat-complément, tels que 改良^{gǎi liáng} ‘améliorer (litt. changer-être.bon)’ et 扩大^{kuò dà} ‘élargir, étendre (litt. étendre-être.grand)’, la seule différence est que DE₄ ‘pouvoir’ et BU₂ ‘ne pas pouvoir’ peuvent s’insérer dans les CVSR, tandis que les verbes composés ne le permettent pas, cf. *改得良^{gǎi de liáng} ‘pouvoir améliorer’, *扩不大^{kuò bu dà} ‘ne pas pouvoir élargir’. Chung (2006) considère également les CVSR comme des syntagmes en raison de la séparabilité de leurs composantes et ainsi les exclut des verbes composés. Toutefois, il est à noter que la séparabilité des CVSR est extrêmement limitée : sauf dans les CVSR de déplacement dont V₂ est constitué d’une autre CVSR⁴² où le suffixe d’accomplissement 了^{le} peut être inséré entre V₁ et V₂, cf. (33), il n’existe que les deux clitiques mentionnés ci-haut qui peuvent s’insérer dans les CVSR. De plus, Lü (1999) a mentionné que les deux clitiques ne peuvent pas toujours être insérés dans les CVSR qui contiennent un V₂ 上^{shàng} qui exprime un sens inchoatif du V₁, cf. (34).

- (33) a. 他 走了 进 来。 b. 他 跑了 出 去。
 lui marcher-PERF entrer venir lui courir-PERF sortir aller
 ‘Il est entré en marchant ‘Il est sorti en courant
 (en s’approchant du locuteur).’ (en s’éloignant du locuteur).’
- (34) a. 外边 飘 (*得) 上 雪花 了。
 dehors voltiger DE₄ commencer neige-fleur LE₂
 ‘Les flocons de neige commencent à voltiger dehors.’
- b. 最近 又 忙 (*不) 上 了。
 récemment encore être.occupé BU₂ commencer LE₂
 ‘Récemment, je commence à être occupé de nouveau.’

En outre, les V₂ directionnels sont souvent atones (B. Huang et Liao, 2002; Zhu, 1982), y compris ceux qui sont grammaticalisés, cf. (33), (34), alors que les prononciations atones se produisent généralement dans les clitiques et les suffixes, les morphes redupliqués ou les

⁴² Nous discuterons plus loin ce type de CVSR.

derniers morphes de certains mots-composés. Ces faits montrent que les CVSR ont une structure interne cohésive ressemblant aux mots-formes, mais en considérant la séparabilité, même limitée, nous les traitons comme des syntagmes.

Selon la compositionnalité et la nature des contraintes des composantes, il existe des CVSR libres et phraséologisées (cf. Nguyen, 2006, p. 171-175). Par exemple, les CVSR en (35) sont des syntagmes libres, car elles sont à la fois compositionnelles sur le plan sémantique et libres au niveau syntaxique; et celles en (36) sont des syntagmes locutionnels parce qu'elles sont non compositionnelles dans leur signifié et figées sur le plan syntaxique.

- | | | | | | |
|---------|----------------------|-------------------|----|-------------------------|-------------------|
| (35) a. | ^{dǎ} 打 | ^{sù} 碎 | b. | ^{dǎ} 打 | ^{pò} 破 |
| | frapper | être.brisé | | frapper | être.cassé |
| | 'briser en frappant' | | | 'casser en frappant' | |
| | | | | | |
| (36) a. | ^{dǎ} 打 | ^{dòng} 动 | b. | ^{gǎn} 感 | ^{dòng} 动 |
| | frapper | bouger | | sentir | bouger |
| | 'émouvoir' | | | 'être ému' / 'émouvoir' | |

À part ces deux types que nous pouvons distinguer facilement selon leur compositionnalité, il existe un autre type de CVSR dont le statut n'est pas évident, cf. (27), (30) et (34), et c'est surtout ce dernier type qui nous intéresse dans ce mémoire. Comme nous l'avons mentionné au début, les CVSR de ce type ressemblent beaucoup à des syntagmes collocationnels, car elles sont compositionnelles, mais les V₂ sont sélectionnés en fonction des premiers (donc elles ne sont pas libres). Il est aussi à noter que les V₂ de ce type sont souvent des verbes de déplacement grammaticalisés exprimant un sens d'accomplissement ou de commencement. Cependant, ces CVSR ne peuvent pas accepter toutes les mêmes collocations que la base. Reprenons l'expression ^{ài shàng} 爱上 'tomber amoureux' : si la base ^{ài} 爱 'aimer' peut être modifiée par un **Magn** ^{shēnshēn de} 深深地 'profondément' cf. (32), elle ne peut pas accepter tous ses collocatifs habituels quand elle est suivie d'un V₂ ^{shàng} 上 'commencement'. Par exemple, en (37)b, ^{ài} 爱 'aimer' ne peut plus être modifié par ses **Magn** ^{hěn} 很 'très' ou ^{fēi cháng} 非常 'extrêmement' comme en (37)a. Cela indique que la CVSR ^{ài shàng} 爱上 'tomber amoureux' ressemble plus à une locution qu'à une collocation.

- (37) a. ^{tā hěn fēi cháng} 他 很/非常 ^{ài nǐ} 爱 你。 b. * ^{tā hěn fēi cháng} 他 很/非常 ^{ài shàng} 爱 上 ^{nǐ} 你。
 lui très/extrêmement aimer toi lui très/extrêmement aimer commencer toi
 ‘Il t’aime beaucoup.’ ‘Il tombe amoureux fou de toi.’

Toutefois, s’il s’agit d’une locution (faible), il y aura un problème concernant le pivot sémantique (§2.3), parce que, par définition, le pivot sémantique d’une locution n’est pas contenu dans le sens de ses composantes, alors que la base sert de pivot sémantique pour une collocation. Dans le cas de ^{ài shàng} 爱上 ‘tomber amoureux’, nous voyons bien que le pivot sémantique est le sens de ^{ài} 爱 ‘aimer’, qui est un argument de ^{shàng} 上 ‘commencer’. Si nous considérons cette expression comme une locution, cela va contredire la théorie. En conséquence, c’est un type de syntagme très particulier, à mi-chemin entre les locutions faibles et les collocations. Pour l’instant, nous nous contentons de supposer que ce type de CVSR correspond à une collocation dont V₁ est la base et V₂ est le collocatif.

Maintenant, pour encoder ces CVSR collocationnelles dans notre dictionnaire, il faut choisir les bonnes fonctions lexicales. Prenons le cas de ^{ài shàng} 爱上 ‘tomber amoureux’, où V₂ est un collocatif de V₁ qui exprime le début du sentiment amoureux (ou l’accomplissement de l’état d’aimer dans un autre point de vue tel que discuté). La FL phasique **Incep** permet de modéliser des collocations françaises avec un sens inchoatif⁴³, par exemple **Incep**(aimer) = commencer [à ~] ou la FL complexe **IncepOper**₁(amoureux) = tomber [~]. Cependant, nous ne pouvons pas utiliser directement cette fonction pour les CVSR en mandarin, parce que le collocatif retourné par la FL **Incep** est toujours un verbe phasique qui est le gouverneur syntaxique de la base. Or, si nous retournons à notre CVSR collocationnelle en mandarin ^{ài shàng} 爱上, V₁ ^{ài} 爱 ‘aimer’ est bien un actant du V₂ ^{shàng} 上 ‘commencer’ sur le plan sémantique comme en français, mais dans la structure syntaxique, il y a une inversion du gouverneur et V₂ devient un dépendant modificateur de V₁

⁴³ Il existe deux autres FL phasiques, **Fin** ‘cesser’ et **Cont** ‘continuer’, mais (Mel’čuk et al., 1995, p. 142) indiquent que la combinatoire de ces trois FL phasiques avec les verbes français n’est pas très intéressante du point de vue lexicographique, puisque « auprès d’un verbe, ces fonctions sont presque toujours exprimées de façon régulière : **Incep**(chanter) = commencer [à ~], **Fin**(lire) = cesser [de ~], etc. ». Cependant, en combinant avec d’autres FL actanciennes, la combinatoire des FL phasiques avec les noms prédicatifs est très riche et variée en français.

(Nguyen, 2006). Nguyen (2006, p. 154-156) a comparé les CVSR aux verbes en série où le premier verbe est clairement la tête syntaxique de la phrase, comme en (38), et il a constaté que le premier verbe dans une CVSR se comporte comme le premier verbe d'une construction en série :

- 1) Nous pouvons construire une phrase interrogative de la forme «V₁ MÉI₂ V₁V₂», mais jamais *«V₁ V₂ MÉI₂ V₂» même si l'interrogation porte sémantiquement sur V₂, cf. (39);
- 2) Le marqueur de négation au perfectif MÉI₂ se place toujours avant V₁, mais jamais avant V₂, cf. (40);
- 3) V₁ ne peut pas admettre le suffixe d'accomplissement, 了^{le} mais V₂ le peut, cf. (41).

(38) 我 kǎi shǐ xiě lùn wén。
 moi commencer écrire mémoire
 'je commence à écrire le mémoire.'

(39) a. 你 kǎi shǐ méi kǎi shǐ xiě lùn wén ?
 toi commencer NEG_{Perf} commencer écrire mémoire
 'Est-ce que tu as commencé à écrire le mémoire ?'

b. 你 ài méi ài shàng tā ?
 toi aimer NEG_{Perf} aimer commencer lui
 'Est-ce que tu es tombé amoureux de lui ?'

(40) a. 我 méi kǎi shǐ xiě lùn wén 。 B. 我 méi ài shàng
 tā 。
 Moi NEG_{Perf} commencer écrire mémoire moi NEG_{Perf} aimer commencer lui
 'Je n'ai pas commencé à écrire le mémoire.' 'Je ne suis pas tombé amoureux de lui.'

(41) a. 我 kǎi shǐ xiě le jǐ háng 。 b. 我 ài shàng le tā 。
 Moi commencer écrire-PERF quelques lignes moi aimer commencer-PERF lui
 'J'ai commencé à écrire quelques lignes.' 'Je suis tombé amoureux de lui.'

Ainsi, par analogie avec les verbes en série, nous admettons que V₁ dans une CVSR est le gouverneur syntaxique de V₂. De plus, le complément d'objet direct (=COD) de V₁ est toujours placé après V₂ au lieu de V₁ comme en (42)a, alors que la phrase (42)b est agrammaticale. Si V₂ est le gouverneur du V₁, V₁ et son COD vont être séparés par V₂ pour aucune raison, il y aura donc un problème pour la projectivité et la linéarisation de la phrase.

(42) a. 我 爱 上 他。
 moi aimer commencer lui
 'Je tombe amoureux de lui'

B. *我 爱 他 上。
 moi aimer lui commencer
 'Je tombe amoureux de lui'

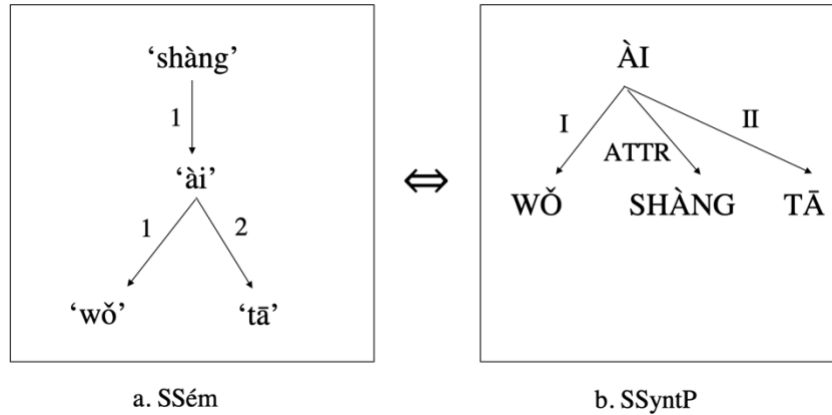


Figure 5. SSém et SSyntP de la collocation 爱上^{ài shàng}

En outre, comme nous l'avons mentionné plus haut, V₂ 上^{shàng} au sens inchoatif a subi une neutralisation sémantique en perdant son sens propre 'monter' et en portant le sens inchoatif, et il doit toujours s'utiliser après un autre verbe pour exprimer ce sens. Néanmoins, il n'est pas totalement grammaticalisé au point de devenir une particule aspectuelle, car son usage comme marque de l'aspect inchoatif n'est pas généralisé : il n'est sélectionné que par certains verbes. Par conséquent, son statut grammatical est à mi-chemin entre un verbe et un clitique (X. Dong, 2017). Il est ainsi plausible que V₂ fonctionne comme un dépendant modificateur de V₁ au lieu de son gouverneur.

En conséquence, pour indiquer cette dépendance syntaxique, Nguyen (2006) a proposé d'encoder la collocation 爱上^{ài shàng} à l'aide de la FL complexe **Adv₁Incep** : **Adv₁Incep**(爱)^{ài} = 上^{shàng}. La fonction **Adv₁** indique que le collocatif V₂ fonctionne comme un adverbe par rapport à la base V₁. En effet, au niveau sémantique, le sémantème de V₁ est un actant de V₂, et au niveau syntaxique profond, V₂ devient un modificateur de V₁ comme un adverbe. Ainsi, une paraphrase française de **Adv₁Incep**(爱)^{ài} = 上^{shàng} serait : *aimer en commençant* (Nguyen, 2006, p. 274). Nous admettons cet encodage dans notre dictionnaire, et nous utiliserons ainsi cette fonction pour les autres cas similaires.

Il est à noter que dans les CVSR en mandarin, V₂ peut être soit un lexème verbal comme 上^{shàng} ‘commencer’ dans 爱上^{ài shàng} ‘tomber amoureux’ ou 起^{qǐ} ‘commencer’ dans 想起^{xiǎng qǐ} ‘se rappeler’, cf. (30)a, soit une autre CVSR phraséologisée de la forme « V₂'-V₂'' »⁴⁴, comme 起来^{qǐ lái} ‘commencer’ dans les deux syntagmes en (43). Ce dernier type est un peu particulier du fait qu’il peut s’intercaler avec le complément d’objet direct du V₁ ou les composantes d’un V₁ phraséologisé, ce qui rend ce type de CVSR plus complexe. Donc, nous verrons ci-dessous son comportement syntaxique par rapport à V₁ afin de bien classer ce type de collocation dans notre dictionnaire.

- (43) a. 恨^{hèn} 起来^{qǐ lái} b. 担心^{dān xīn} 起来^{qǐ lái}
 haïr se.lever-venir porter-cœur se.lever-venir
 ‘commencer à haïr’ ‘commencer à s’inquiéter’

En fait, le sens littéral de 起^{qǐ} est ‘se lever’ et 来^{lái} ‘venir’, et ces deux verbes forment une CVSR collocationnelle dont le sens est ‘se lever’ comme dans la phrase (44). Mais dans les deux syntagmes en (43), 起来^{qǐ lái} forme plutôt une CVSR locutionnelle au sens de ‘commencer’⁴⁵, car elle n’est pas compositionnelle au niveau du signifié, et le V₂ 来^{lái} devient atone⁴⁶.

- (44) 他^{tā} 起^{qǐ} 来^{lái} 了^{le}。
 lui se.lever venir LE₂
 ‘Il s’est levé.’

En outre, 起来^{qǐ lái} est une locution séparable. Si V₁ possède un COD, ce dernier doit être inséré entre V₂' et V₂'' comme en (45)a et(46)a au lieu d’être placé après un mot-forme V₂ comme en (42)a.

⁴⁴ Quant au statut de V₁, Nguyen (2006, p. 162) pense que V₁ est toujours un mot-forme verbal, mais en fait, il peut aussi être un syntagme collocationnel ou locutionnel, dont la partie nominale va soit se placer après V₂, soit s’intercaler avec le « V₂'- V₂'' ». Nous verrons plus loin ces structures complexes.

⁴⁵ Cette CVSR peut aussi s’utiliser après certains verbes pour exprimer un sens de l’état complet ou une impression/jugement du locuteur, par exemple, 想起来^{xiǎng qǐ lái} ‘se rappeler’ (litt. ‘penser se.lever venir’) et 看起来^{kàn qǐ lái} ‘avoir l’air’ (litt. ‘regarder se.lever venir’); dans ces cas-là, il s’agit de deux autres locutions différentes.

⁴⁶ Nous pouvons aussi garder le ton de 来^{lái}, mais cela semble moins naturel à l’oral.

(45) a. 我 恨 起 他 来
 moi haïr se.lever lui venir
 ‘Je commence à le haïr’

b. *我 恨 起来 他
 moi haïr se.lever lui

(46) a. 我 担心 起 他 来
 moi porter-cœur se.lever lui venir
 ‘Je commence à m’inquiéter pour lui’

b. *我 担心 起来 他
 moi porter-cœur se.lever lui
 ‘Je commence à m’inquiéter pour lui’

Néanmoins, si V₁ est une locution comme 担心 ‘s’inquiéter’ en (47)a ou une collocation comme 害怕 ‘avoir peur’ en (47)b de la forme «V-N» et il ne possède pas d’autres compléments d’objet qui le suivent, la composante nominale peut avoir un comportement syntaxique similaire au COD de V₁ en (45)a et (46)a, donc les deux noms 心 ‘cœur’ et le 怕 ‘peur’ en (47) peuvent tous être insérés entre 起 et 来.

(47) a. 我 担 起 心 来
 moi porter se.lever cœur venir
 ‘Je commence à m’inquiéter’

b. 我 害 起 怕 来
 moi produire se.lever peur venir
 ‘Je commence à avoir peur’

Nous pouvons utiliser la même FL complexe **Adv₁Incep** comme 上 ‘commencer’ dans 爱上 ‘tomber amoureux’ pour encoder cette CVSR locutionnelle en ajoutant une contrainte pour la séparabilité. Toutefois, nous ne l’avons pas décrit dans notre dictionnaire, étant donné que la locution 起来 peut se combiner avec la plupart des verbes d’action ou des verbes qualificatifs pour exprimer un commencement d’action ou d’état.

Il existe d’autres CVSR locutionnelles phasiques en mandarin, telles que 下来 ‘commencer’/‘continuer’ (litt. ‘descendre venir’) et 下去 ‘continuer’ (litt. ‘descendre aller’), dont les comportements syntaxiques sont plus ou moins similaires à ceux de 起来 ‘commencer’. Nous pouvons utiliser les FL complexes comme **Adv₁Incep** et **Adv₁Cont** pour les encoder tout en précisant les différentes contraintes.

4.4.4 Adv₁Magn vs Magn postposé : le cas de 死

En mandarin, les adverbes se placent toujours avant les verbes qu’ils modifient sauf rares exceptions : c’est le cas de 极 ‘extrêmement’, qui peut se placer soit devant soit derrière (mais quand il se trouve après le verbe, il faut ajouter un clitique exclamatif 了), cf. (48). Ici, la lexie

极^{jí} ‘extrêmement’ est un **Magn** parce qu’elle exprime un sens d’intensification pour la base 担^{dān}心^{xīn} ‘s’inquiéter’. Par défaut, les **Magn** pour les lexies verbales se placent en avant, et nous ajoutons des contraintes pour ceux qui peuvent se placer après la base.

- (48) a. 我^{wǒ} 极^{jí} 担^{dān}心^{xīn}。
 moi extrêmement porter-cœur
 ‘Je m’inquiète extrêmement.’
- b. 我^{wǒ} 担^{dān}心^{xīn} 极^{jí} 了^{le}。
 moi porter-cœur extrêmement LE₂
 ‘Je m’inquiète extrêmement.’

Il y a d’autres expressions semblables à (48)b où la lexie portant le sens d’intensification se trouve également après la base, cf. (49).

- (49) a. 我^{wǒ} 担^{dān}心^{xīn} 死^{sǐ} 了^{le}。
 moi porter-cœur mourir LE₂
 ‘Je m’inquiète à mourir.’
- b. 我^{wǒ} 担^{dān}心^{xīn} 坏^{huài} 了^{le}。
 moi porter-cœur s’abîmer LE₂
 ‘Je m’inquiète terriblement.’

Au premier regard, ces structures ressemblent beaucoup aux CVSR (§4.4.3) : 死^{sǐ} ‘mourir’ et 坏^{huài} ‘s’abîmer’ fonctionnent comme les autres V₂ de CVSR, exprimant un résultat de V₁ et étant un dépendant syntaxique de ce dernier en SSyntP. Ainsi, Nguyen (2006, p. 312) traite 死^{sǐ} ‘mourir’ comme un **Adv₁Magn** pour le mot-clé 担^{dān}心^{xīn}.

Néanmoins, si nous examinons plus précisément le sens de 死^{sǐ} dans cette construction, nous pouvons constater un figement : 死^{sǐ} perd le sens propre de ‘mourir’ parce que le sujet ne meurt pas vraiment, c’est plutôt une exagération pour exprimer un sens d’intensification du verbe qu’il modifie. Dans les deux phrases en (50), nous pouvons voir clairement que 死^{sǐ} modifie seulement les verbes 气^{qì}₂ ‘se fâcher’ et 气^{qì}₂ ‘fâcher’ pour exprimer une intensification de ceux-ci, et ces lexies sont le seul actant sémantique du prédicat ‘intense’. La lexie 死^{sǐ} peut également modifier un verbe qualificatif qui exprime une sensation ou un sentiment du locuteur, cf. (51).

- (50) a. 我^{wǒ} 气^{qì} 死^{sǐ} 了^{le}。
 moi se.fâcher mourir LE₂
 ‘Je me fâche à mourir.’
- b. 你^{nǐ} 气^{qì} 死^{sǐ} 我^{wǒ} 了^{le}。
 toi fâcher mourir moi LE₂
 ‘Tu me fâches à mourir.’
- (51) a. 这^{zhè}道^{dào}菜^{cài} 辣^{là} 死^{sǐ} 了^{le}。
 Ce CL plat être.piquant mourir LE₂
 ‘Ce plat est piquant à mourir.’
- b. 风^{fēng}景^{jǐng} 美^{měi} 死^{sǐ} 了^{le}。
 paysage être.beau mourir LE₂
 ‘Le paysage est beau à mourir.’

Ainsi, nous pouvons constater que, quand la lexie 死^{sǐ} exprime un sens d'intensification, elle a un seul actant sémantique qui est le sens qu'elle modifie, mais elle ne possède pas un autre actant qui doit être généralement un être animé comme son sens propre 'mourir', parce que ni le sujet 我^{wǒ} 'moi' de la phrase en (50)a ni le sujet 你^{nǐ} 'toi' de (50)b ne meurent, ni existe-t-il un tel actant en (51). En fait, cela ressemble à la locution adverbiale française à mourir qui exprime également un sens d'intensification comme dans *plate à mourir, s'inquiéter à mourir*. Cependant, il y a une différence entre l'usage de à mourir en français et de 死^{sǐ} 'à mourir' en mandarin : ce dernier peut faire ressortir en syntaxe un actant caché des verbes qu'il modifie. Par exemple, pour les deux phrases en (51), un pronom personnel 我^{wǒ} 'moi' peut être ajouté après 死^{sǐ}, cf. (52), pour exprimer explicitement le 2^e actant des verbes 辣^{là} 'x est piquant pour y' et 美^{měi} 'x est beau pour y', alors que sans la présence de 死^{sǐ}, ces phrases sont agrammaticales, cf. (53). Cet usage n'est compatible qu'avec les lexies qui dénotent des expériences sensorielles ou subjectives comme 辣^{là} 'être piquant' et 美^{měi} 'être beau' (l'actant exprimé après 死^{sǐ} désigne souvent le locuteur lui-même, c'est-à-dire, 我^{wǒ} 'moi', mais pas nécessairement), mais il est incompatible avec les lexies qui décrivent une réalité objective comme 快^{kuài} 'être rapide' ou 高^{gāo} 'être élevé, cf. (54).

- (52) a. zhè dào cài là sǐ wǒ le. b. fēngjǐng měi sǐ wǒ le.
 ce CL plat être.piquant mourir moi LE₂ paysage être.beau mourir moi LE₂
 'Ce plat est piquant à mourir pour moi.' 'Le paysage est beau à mourir pour moi.'
- (53) a. *zhè dào cài là wǒ le. b. *fēngjǐng měi wǒ le.
 ce CL plat être.piquant moi LE₂ paysage être.beau moi LE₂
 'Ce plat est piquant pour moi.' 'Le paysage est beau pour moi.'
- (54) a. rè liàng gāo sǐ le. b. *rè liàng gāo sǐ wǒ le.
 calorie être.élevé mourir LE₂ calorie être.élevé mourir moi LE₂
 'C'est riche à mourir en calories.' 'C'est riche à mourir pour moi en calories.'

En outre, 死^{sǐ} peut s'utiliser avant certains verbes qualificatifs pour exprimer une intensification de ces derniers, cf. (55)a, où il se comporte comme un adverbe; ou il peut encore s'utiliser comme un verbe intensificateur pour modifier un autre verbe, cf. (55)b-d. Toutefois, dans les deux premiers cas, la lexie 死^{sǐ} porte plutôt un sens statique 'être mort' au lieu d'un sens télique 'mourir', il se peut qu'il s'agisse d'une lexie différente de celle en (49)a, mais les tests

du zeugme et du contraste⁴⁷ pour distinguer l’ambiguïté des deux mots-formes sont difficiles à appliquer à cause de la syntaxe chinoise, donc nous n’allons pas plus loin sur ce point.

- (55) a. 死^{sǐ} 沉^{chén} 的^{de} 书^{shū} 包^{bāo} b. 他^{tā} 睡^{shuì} 得^{de} 很^{hěn} 死^{sǐ}。
 mourir lourd DE₁ livre.sac lui dormir DE₃ très mourir
 ‘un cartable extrêmement lourd’ ‘Il dort très profondément.’
- c. 他^{tā} 死^{sǐ} 要^{yào} 面^{miàn} 子^{zi}。 d. 她^{tā} 死^{sǐ} 不^{bù} 要^{yào} 脸^{liǎn}。
 lui mourir désirer face elle mourir NEG désirer visage
 ‘Il tient extrêmement à sa réputation.’ ‘Elle est extrêmement éhontée.’

En considérant les faits ci-dessus, nous supposons qu’il existe une lexie adverbiale 死^{sǐ}² ‘extrêmement’ dérivée de la lexie verbale 死^{sǐ}¹ ‘mourir’ par métaphore et conversion. De même pour la lexie adverbiale 坏^{huài}², dont l’usage ressemble beaucoup à celui de 死^{sǐ}², et qui provient de la lexie verbale 坏^{huài}¹ ‘s’abîmer’. Ainsi, nous pouvons facilement les encoder avec la fonction **Magn** dans le dictionnaire, tout en ajoutant une contrainte qu’il faut les placer après leur base.

4.4.5 有^{yǒu} comme Oper₁ ou Manif

Le verbe 有^{yǒu} a deux sens principaux: ‘avoir’ et ‘exister’, qui correspondent à deux lexèmes différents, 有₁^{yǒu} ‘avoir’ et 有₂^{yǒu} ‘exister’, cf. (56).

- (56) a. 他^{tā} 有^{yǒu} 一^{yī} 只^{zhī} 猫^{māo}。 b. 床^{chuáng} 上^{shàng} 有^{yǒu} 一^{yī} 只^{zhī} 猫^{māo}。
 lui avoir un CL chat lit sur exister un CL chat
 ‘Il a un chat.’ ‘Il y a un chat sur le lit.’

⁴⁷ Ces tests sont utilisés pour distinguer si un mot est vague ou ambigu. Le test du zeugme (jeu de mot) demande d’utiliser une occurrence du mot en activant ses deux sens perçus; si on a un effet bizarre, cela indique que le mot est ambigu. Par exemple, la phrase « #La scène est triste et moi aussi » est bizarre, donc *triste* au sens ‘qui provoque un sentiment négatif’ et *triste* ‘qui éprouve un sentiment négatif’ sont deux lexies différentes. Le test du contraste demande d’utiliser deux occurrences du mot avec chacune un sens différent; s’il y a un contraste, il s’agit d’un mot ambigu. Par exemple, dans « mon amour m’a avoué son amour », on a un contraste entre *amour* ‘personne aimée’ et *amour* ‘sentiment d’affection’, donc, il s’agit de deux lexies différentes (Cf. Cruse, 1986; Murphy, 2010).

Le lexème 有₁ ‘avoir’ peut servir de verbe support pour plusieurs lexies dans le champ des émotions, cf. (57). Dans ces deux phrases, le verbe 有₁ ‘avoir’ a deux actants : le sujet est le 1^{er} actant de la base, et le COD est la base. Ainsi, c’est un verbe support **Oper₁**.

- (57) a. 他 有 一 丝 悲 伤。
 lui avoir un fil tristesse
 ‘Il a un peu de tristesse.’
- b. 我 总 是 有 很 多 担 心。
 moi toujours avoir beaucoup porter-cœur
 ‘J’ai toujours beaucoup d’inquiétude.’

Le lexie 有₂ ‘exister’ peut également servir de verbe support, cf. (58). Cependant, ce n’est pas un **Oper₁** comme en (57), car la base est le premier actant, et le deuxième actant est une partie du corps de la personne qui ressent la tristesse. Cela fait penser à la fonction **Manif** ‘se manifester dans’, qui est utilisée pour décrire un collocatif verbal dont le sujet est la base et le complément d’objet est le nom de l’entité dans laquelle la chose dénotée par la base se manifeste (Mel’čuk et al., 1995, p. 145). Ainsi, nous pouvons décrire la collocation en (58)a avec cette fonction: **Manif** (悲伤 ‘tristesse’) = [x (的) 眼 里/中 ‘dans les yeux de x’] 有 ‘exister’ ~. Les structures sémantique et syntaxique profonde et de surface de cette phrase sont données à la Figure 6.

- (58) a. 他 (的) 眼 里/中 有 一 丝 悲 伤。
 lui DE₁ œil dans exister un fil tristesse
 ‘Il y a un peu de tristesse dans ses yeux.’
- b. 他 (的) 心 里/中 有 一 丝 悲 伤。
 lui DE₁ cœur dans exister un fil tristesse
 ‘Il y a un peu d’inquiétude dans son cœur.’

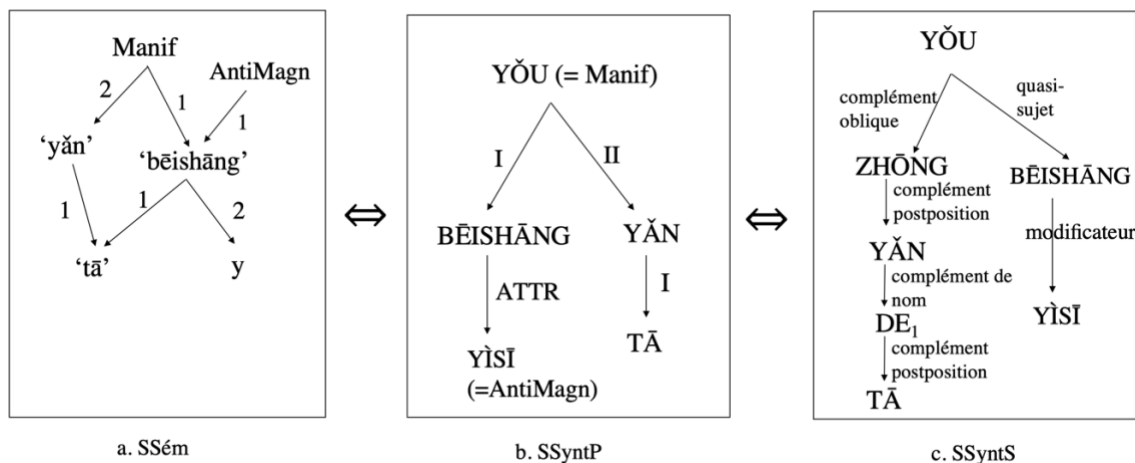


Figure 6. SSém, SSyntP et SSyntS de la phrase (58a)

Toutefois, pour la phrase en (58)b, le deuxième actant 心 ‘cœur’ est une chose invisible, contrairement à 眼 ‘œil’ en (58)a. dans ce cas-là, la FL **Manif** n’est pas très pertinente puisqu’il n’y a pas de « manifestation » observable. En outre, nous pouvons changer le verbe support 有 par d’autres verbes, cf. (59). En fait, 感到 ‘ressentir’ en (59)a et 充满 ‘se remplir’ en (59)b sont respectivement le **Oper**₁ et le **Magn+Oper**₁ de 悲伤 ‘tristesse’, cf. (60). De plus, 心里 et 心中 ‘cœur-dedans’ peuvent interchanger en (59) avec une lexie nominale 内心⁴⁸ ‘cœur intérieur’. Ces faits suggèrent que 心里 et 心中 ‘cœur-dedans’ fonctionnent comme un syntagme nominal plutôt que prépositionnel.

- (59) a. 他(的) 心 里/中 感到 一 丝 悲伤。
 lui DE₁ œil dans ressentir un fil tristesse
 ‘Il ressent un peu de tristesse dans son cœur.’
- b. 他(的) 心 里/中 充满了 悲伤。
 lui DE₁ cœur dans se.remplir-LE₁ tristesse
 ‘Son cœur se remplit de tristesse.’
- (60) a. 他 感到 一 丝 悲伤。
 lui ressentir un fil tristesse
 ‘Il ressent un peu de tristesse.’
- b. 他 充满了 悲伤。
 lui se.remplir-LE₁ tristesse
 ‘Il se remplit de tristesse.’

En fait, 心 ‘cœur’⁴⁹ en mandarin est considéré comme un organe de la pensée, dans un sens figuré. Aujourd’hui, il existe beaucoup de lexies qui contiennent ce sinogramme dans le

⁴⁸ Dans 《汉语规范词典》(*Dictionnaire standard de mandarin moderne*), on utilise « 心中, 心里 » pour définir le nom 内心 ‘cœur intérieur’. Nous y trouvons également les entrées 心里 ‘cœur-dedans’, 心间 ‘cœur-dedans’, 心上 ‘cœur-dessus’, qui ressemblent à des syntagmes prépositionnels, mais leur partie du discours est marquée comme nom.

⁴⁹ D’ailleurs, c’est aussi une des clés les plus importantes des sinogrammes. Dans 《说文解字》(*Explication des pictogrammes et des caractères composés*), il y a 262 sinogrammes sous la catégorie de clé de 心 ‘cœur’, tous ces mots sont liés aux pensées, sentiments, ou activités mentales. Même aujourd’hui, nous pouvons trouver beaucoup de sinogrammes qui possèdent une partie de 心, tels que 悲 ‘chagrin’, 怒 ‘colère’, 忘 ‘oublier’, 忍 ‘endurer’ et ainsi de suite. Les sens de ces mots sont liés à la fonction du cœur en tant qu’organe de la pensée, et aussi aux sens figurés de 心 (X. Zhao, 2017).

champ des pensées et des émotions, par exemple, 心想 ^{xīn xiǎng} ‘penser’ (litt. ‘cœur-penser’), 心算 ^{xīn suàn} ‘calculer mentalement’ (litt. ‘cœur-calculer’), 心动 ^{xīn dòng} ‘être touché’ (litt. ‘cœur-battre’), 心怀 ^{xīn huái} ‘porter au cœur’, 心存 ^{xīn cún} ‘garder au cœur’ (ces deux dernières lexies sont des mots-valises de 心中怀有 ^{zhōnghuái yǒu} et 心中存有 ^{xīn zhōng cún yǒu}). Par un mécanisme de métonymie, l’endroit où se trouve 心 ^{xīn} ‘cœur’ peut représenter son propriétaire pour exprimer un sentiment. En conséquence, nous pouvons considérer 心里 ^{xīn lǐ} et 心中 ^{xīn zhōng} ‘cœur-dedans’ comme des locutions nominales qui peuvent remplacer l’individu qui éprouve un certain sentiment et servir de sujet dans les phrases telles que (58)b et (59). Ainsi, le verbe support en (58)b serait 有₁ ^{yǒu} ‘avoir’ qui fonctionne comme un **Oper₁**, au lieu de 有₂ ^{yǒu} ‘exister’ qui est un **Manif** en (58)a.

4.4.6 高兴 ^{gāo xìng} : une lexie ou deux ?

Le vocable 高兴 ^{gāo xìng} ‘être content; aimer’ possède souvent deux acceptions dans les dictionnaires chinois : un adjectif et un verbe. Par exemple, dans 《现代汉语词典》 (*Dictionnaire du mandarin moderne*), les deux acceptions sont respectivement ‘joyeux et excité’ et ‘faire quelque chose avec une humeur joyeuse; aimer’; et dans un dictionnaire sino-français en ligne⁵⁰, une acception est ‘content, heureux’, l’autre est ‘être content de faire quelque chose’. Ainsi, en (61), 高兴 ^{gāo xìng} est considéré comme un adjectif dans la première acception, alors qu’il est considéré comme un verbe dans la deuxième acception en (62) selon les dictionnaires chinois.

(61) 我 很 高兴 见 到 你。
 moi très content voir arriver toi
 ‘Je suis content de te voir’

(62) 你 高兴 去 哪儿 就 去 哪儿。
 toi content aller où alors aller où
 ‘Va où tu veux’

Cependant, à part les parties du discours, les deux acceptions sont en fait très semblables. Dans ces deux phrases ci-haut, les deux lexies en question ont une forme propositionnelle

⁵⁰ Voir <http://www.frdic.com/dicts/fr/高兴> [consulté le 8 oct. 2018]

similaire ‘x ^{gāo xìng}高兴 y’ dont les deux actants ont les mêmes étiquettes sémantiques (l’actant x est un individu et y est un fait), et il est difficile de distinguer les différences dans leur sens. Étant donné que nous considérons les adjectifs en mandarin comme des verbes qualificatifs (§3.1), nous avons seulement un vocable verbal pour ^{gāo xìng}高兴 dans notre dictionnaire, alors, il faut se demander s’il existe vraiment deux lexies séparées pour ce même vocable. Nous supposons qu’il s’agit bien de deux lexies différentes ^{gāo xìng}高兴₁ et ^{gāo xìng}高兴₂ en (61) et (62) pour les deux raisons suivantes.

Premièrement, même si les sens de ^{gāo xìng}高兴 en (61) et (62) se ressemblent beaucoup, il existe quand même une nuance par rapport à leur deuxième actant y. En (61), l’actant y est une cause qui fait que x est content, ainsi, la forme propositionnelle pour ce sens serait ‘x ^{gāo xìng}高兴 en raison de y’. Dans ce sens, y peut être une proposition dont le sujet est différent de ^{gāo xìng}高兴, cf. (63). En outre, y peut se déplacer au début de la phrase, ainsi les deux phrases en (61) et (63) peuvent être reformulées en (64). Dans certains contextes, l’actant y peut s’exprimer implicitement sans apparaître. En revanche, en (62), la forme propositionnelle pour cette lexie serait plutôt ‘x ^{gāo xìng}高兴 pour faire y’, où l’actant x de ^{gāo xìng}高兴 doit également être le 1^{er} actant du fait y. Dans ce sens-là, y ne peut plus se déplacer en avant, cf. (65). L’usage de cette 2^e lexie est plus restreint que la première.

(63) ^{wǒ hěn gāo xìng nǐ néng lái}
 moi très content toi pouvoir venir
 ‘je suis content que tu peux venir’

(64) a. ^{jiàn dào nǐ wǒ hěn gāo xìng} 见到你, 我很高兴。
 voir arriver toi moi très content
 ‘je suis content de te voir’
 b. ^{nǐ néng lái wǒ hěn gāo xìng} 你能来, 我很高兴。
 toi pouvoir venir moi très content
 ‘je suis content que tu peux venir’

(65) *^{qù nǎr nǐ gāo xìng jiù qù nǎr} 去哪儿, 你高兴就去哪儿。
 aller où toi content alors aller où
 ‘Va où tu veux’

Deuxièmement, les deux lexies ont des fonctions lexicales différentes. Par exemple, les quasi-synonymes de ^{gāo xìng}高兴₁ sont ^{kāi xīn}开心 ‘être.joyeux’, ^{kuài lè}快乐 ‘être.joyeux’, et nous pouvons les

interchanger avec 高兴₁ dans les phrases en (64)⁵¹ pour former des paraphrases; alors que ceux de la deuxième lexie sont 爱 ‘aimer’, 喜欢 ‘aimer’, et nous pouvons les interchanger avec 高兴₂ en (62). De plus, la première lexie possède plusieurs modificateurs d’intensification (**Magn**) comme beaucoup d’autres verbes qualificatifs, tels que 很 ‘très’, 好 ‘bien’, 特别 ‘extraordinairement’, tandis que 高兴₂ n’accepte pas de **Magn** en général, cf. (66), et elle s’utilise souvent avec un clitique de négation 不 ‘NEG’, cf. (67).

(66) 他 (*很) 高兴 看 电影, 对 看 戏 不 感 兴趣。
 lui très content regarder film, pour regarder théâtre NEG sentir intérêt
 ‘Il aime regarder les films, et il ne s’intéresse pas au théâtre.’

(67) 你 不 高兴 去 就 算 了。
 toi NEG content aller alors laisser-passer LE₂
 ‘Si tu n’aimes pas y aller, alors laisse tomber.’

Par conséquent, 高兴₁ et 高兴₂ sont bel et bien deux lexies différentes : 高兴₁ est un verbe qualificatif qui peut être transitif pour accepter des propositions comme 2^e actant, mais nous pouvons aussi déplacer cet actant au début de la phrase ou ne pas l’exprimer; 高兴₂ est un verbe transitif qui doit toujours être accompagné d’une proposition qui partage le même 1^{er} actant que la lexie.

En outre, dans les expressions en (68)a, nous avons l’impression que nous avons affaire à la lexie 高兴₁, mais est-ce que la personne qui parle est vraiment contente de servir quelqu’un ? En fait, c’est plutôt une expression pragmatique pour exprimer une certaine politesse. Et cette expression est assez figée parce que nous ne pouvons pas échanger 高兴 avec les quasi-synonymes de la lexie 高兴₁, cf. (68)b. Selon la typologie des phrasèmes, le sens de cette expression est compositionnel et il est contraint d’une manière sémantico-lexémique, il s’agit donc d’un cliché (§2.3). De même, dans les phrases en (69) qui contiennent la lexie 不安

⁵¹ Ici, nous ne pouvons pas les échanger avec leurs quasi-synonymes dans les phrases en (61), (62) et (63) parce que ces deux verbes qualificatifs sont intransitifs et ne peuvent pas accepter un COD.

‘troublé’ sont des formules de politesse pour exprimer des excuses, le locuteur n’est pas vraiment troublé. Ce sont aussi des clichés.

Idéalement, il faut créer une entrée pour chaque cliché, mais nous nous concentrons sur les phrasèmes lexémiques (i.e. les locutions et collocations) dans ce mémoire, donc nous n’avons pas inclus les clichés dans notre dictionnaire pour l’instant.

- (68) a. (我) ^{wǒ} 很 ^{hěn} 高兴 ^{gāo xìng} 为 ^{wéi} 您 ^{nín} 服务 ^{fú wù}。 b. *很 ^{hěn} 快乐/开心 ^{kuài lè/kāi xīn} 为 ^{wéi} 您 ^{nín} 服务 ^{fú wù}。
moi très content pour vous servir très joyeux/ouvrir.cœur pour vous servir
‘Je suis content de te servir.’ ‘Je suis joyeux de te servir.’
- (69) a. ^{ràng} 让 ^{nín} 您 ^{shòu} 受 ^{lèi} 累 ^{shēn}, 深 ^{gǎn} 感 ^{bù} 不 ^{ān} 安。
laisser vous subir fatigue profondément sentir troublé
‘Désolé de vous déranger.’
- b. ^{zhè yàng} 这样 ^{má fan} 麻烦 ^{nín} 您 ^{zhēn shì}, 真是 ^{bù ān} 不安。
tellement déranger vous vraiment troublé
‘Vraiment désolé de vous déranger.’

Chapitre 5 Conclusion

Les collocations sont un phénomène récurrent à travers les langues. La lexicologie explicative et combinatoire [=LEC] (Mel'čuk et al., 1995) utilise un outil formel, les fonctions lexicales, qui a été appliqué à la description de collocations dans plusieurs langues, comme le français, l'anglais, l'espagnol, le russe et le coréen. Cependant, il n'existe pas encore un dictionnaire de ce type en mandarin qui traite systématiquement les collocations, donc nous avons comme objectif d'établir un modèle de dictionnaire des collocations dans cette langue dans le cadre de la LEC.

Dans le premier chapitre, nous avons montré les lacunes des dictionnaires du mandarin existants ainsi que la nécessité d'établir un modèle de dictionnaire pour décrire les collocations.

Dans le deuxième chapitre, nous avons présenté le cadre théorique de notre recherche : la lexicologie explicative et combinatoire. Après avoir présenté les notions générales de lexicologie, nous avons surtout clarifié les notions de morphe, mot-forme et syntagme en mandarin dans ce cadre. Étant donné que le système d'écriture tente de représenter le signifié plutôt que le signifiant et qu'il n'y a pas d'espaces entre les sinogrammes, il est difficile de déterminer le statut morphologique d'un signe complexe. Nous avons ainsi présenté les critères de Nguyen (2006), que nous avons utilisés pour déterminer le statut des signes complexes dans notre nomenclature. Nous avons également parlé de la typologie des phrasèmes pour bien définir la collocation, objet principal de ce travail, et la distinguer des autres types de phrasèmes. Ensuite, nous avons abordé le métalangage formel des fonctions lexicales que nous avons utilisé pour décrire les collocations et dérivations sémantiques. Nous avons présenté les différents types des fonctions lexicales selon trois axes : l'axe standard/non standard, l'axe paradigmatique/syntagmatique et l'axe simple/complexe/configuration (Jousse, 2010; Mel'čuk et al., 1995). Puis, nous avons présenté le modèle de dictionnaire explicatif et combinatoire [=DEC] et une ressource informatisée basée sur ce modèle, le DiCoLiLex.

Dans le troisième chapitre, nous avons présenté la méthodologie de ce travail. À partir des entrées choisies en français dans le DiCoLiLex, nous avons établi une nomenclature

correspondante en mandarin dans le champ des émotions. Nous avons discuté le classement des parties du discours en mandarin pour justifier la séparation des vocables nominaux et verbaux ayant le même signifiant dans notre nomenclature. Puis, nous avons choisi un corpus d'environ cinq millions de messages Weibo, que nous avons nettoyé et segmenté pour en extraire les cooccurrents des lexies à l'étude. Afin de trouver une mesure optimisée pour trouver les collocations, nous avons comparé plusieurs méthodes d'extraction, telles que le score z, le score t, le PMI et ses variantes (NPMI, PMI², PMI³), et avons constaté que le PMI³ donnait de meilleurs résultats.

Dans le quatrième chapitre, nous avons présenté la rédaction de notre dictionnaire, un des principaux objectifs de ce travail. Après avoir présenté la structure XML du DiCoLiLex, nous l'avons adaptée en fonction des caractéristiques du mandarin afin de faciliter la description lexicographique des lexies dans la nomenclature chinoise. Nous avons traité des étiquettes sémantiques des lexies et des actants utilisés dans notre dictionnaire. Ensuite, nous avons discuté les cas difficiles dans la description lexicographique, des phénomènes qui ne cadrent pas bien dans la théorie, et nous avons proposé des pistes de solutions. En particulier, nous avons traité des « mots élastiques » (§4.4.1) et nous avons montré que la forme courte et la forme longue ayant le même signifié appartiennent en fait à deux lexies différentes, car ils n'ont pas le même signifiant ni la même combinatoire. Nous avons discuté des critères pour déterminer le statut morphologique des signes de la forme « V?N » (§4.4.2) et trouvé que ce sont soit des locutions verbales, soit des collocations avec verbe support. Ces phrasèmes peuvent également se convertir en nom, mais il manque d'indices linguistiques pour distinguer s'il s'agit d'une locution ou d'un lexème composé. Nous avons également traité la construction verbale sérielle résultative [=CVSR] (§4.4.3), et les différents types selon la compositionnalité et la nature des contraintes de ses composantes. Ainsi, nous avons remarqué un type de syntagme très particulier, à mi-chemin entre les locutions faibles et collocations, représenté par l'expression ^{ài shàng}爱上 'tomber amoureux'. Ce type de CVSR ressemble beaucoup à une collocation parce que le sens est compositionnel et la contrainte est de nature lexémique. Toutefois, sa séparabilité est très restreinte, et elle n'est pas compatible avec toutes les autres collocations, ce qui est plutôt

caractéristique des locutions. Puisque le pivot sémantique est le premier verbe, qui ressemble à la base d'une collocation, nous l'avons traité comme une collocation. Cependant, il manque de critères dans la théorie pour nous aider à distinguer le vrai statut de ce type de CVSR; il faudrait faire des études plus approfondies pour combler cette lacune. Une fois que nous avons décidé que cette expression est une collocation, nous avons étudié sa structure syntaxique, et montré que le premier verbe est la tête syntaxique, et le deuxième verbe est un dépendant. Ainsi, nous avons utilisé la fonction complexe **Adv₁Incep** pour encoder cette collocation dans le dictionnaire. Nous avons ensuite montré une autre construction qui ressemble aux CVSR (§4.4.4), par exemple, ^{ài} ^{sǐ} 爱死 'aimer à mourir', mais nous avons montré que ^{sǐ} 死 'extrêmement' est en fait un adverbe postposé, et il peut faire ressortir en syntaxe un actant caché du verbe modifié dénotant des expériences sensorielles ou subjectives comme ^{là} 辣 'être piquant' et ^{méi} 美 'être beau'.

En conclusion, nous avons atteint les trois objectifs de ce mémoire : nous avons décrit les caractéristiques des collocations en mandarin; nous avons adapté la microstructure du DiCoLiLex aux caractéristiques du mandarin; et nous avons rédigé un petit dictionnaire des collocations du champ sémantique des émotions en mandarin constituant de 43 fichiers XML au total, qui contiennent 78 vocables et 126 lexies différentes. Ces fiches décrivent 2 630 liens lexicaux à l'aide de 73 fonctions lexicales uniques. En nous inspirant de l'interface du DiCoLiLex, nous avons créé, avec l'aide de Benoît Robichaud, une interface en ligne⁵² avec une nomenclature de tous les vocables que nous avons décrits afin que le lecteur puisse plus facilement consulter le dictionnaire.

Ce mémoire apporte deux contributions. D'abord, nous avons proposé un modèle de dictionnaire des liens lexicaux en mandarin, en particulier une modélisation des collocations chinoises au moyen des fonctions lexicales. Ce modèle de DEC en mandarin nous permet de

⁵² C'est Benoît Robichaud qui s'est chargé des aspects techniques. Voir : <http://olst.ling.umontreal.ca/Xiaoyu/>. Voir aussi l'annexe 3 pour quelques exemples de vocables provenant de cette interface.

comblent certaines lacunes des dictionnaires monolingues chinois traditionnels en décrivant les collocations de manière systématique et formelle. Puis, nous avons discuté certaines difficultés dans les analyses des collocations en mandarin, parce que la théorie actuelle ne permet pas de bien décrire certaines caractéristiques des constructions en mandarin. Nous avons ainsi proposé des solutions possibles, ce qui apporte un éclairage nouveau à la théorie.

Dans les futures recherches, notre dictionnaire pourrait servir de modèle pour un dictionnaire chinois pédagogique des collocations à l'intention des apprenants. Pour ce faire, il faut vulgariser le dictionnaire, surtout le formalisme, en ajoutant une glose pour chaque fonction lexicale. Nous pouvons aussi nous inspirer du *Lexique Actif Français* (Meřčuk et Polguère, 2007), qui est généré à partir du DiCo. C'est un dictionnaire grand public de dérivations sémantiques et de collocations du français, dans lequel les fonctions lexicales sont remplacées par des paraphrases en français. D'ailleurs, Diachenko (2006) a proposé un système d'apprentissage du lexique, surtout des phrasèmes des langues étrangères, en utilisant les fonctions lexicales. Le système contient des jeux de fonctions lexicales, des jeux de définitions analytiques, des jeux de traduction, et les jeux de lexies. Nous pourrions construire la matière d'apprentissage à partir de notre dictionnaire pour intégrer le mandarin dans ce système.

De plus, notre dictionnaire pourrait être utile dans le domaine du traitement automatique du langage. Par exemple, nous pourrions intégrer ce dictionnaire au système de génération multilingue GenDR, un réalisateur profond générique pour la génération de texte (Lambrey et Lareau, 2015; Lareau et al., 2018). Le système contient trois dictionnaires pour réaliser la lexicalisation : un dictionnaire sémantique (*semanticon*), un dictionnaire lexique (*lexicon*) et un dictionnaire des fonctions lexicales. Nous pouvons transférer les contenus de notre dictionnaire dans le *semanticon* et *lexicon* tout en ajustant les règles d'arborisation en fonction des caractéristiques chinoises, pour ainsi faire de la génération automatique de texte en mandarin.

Bibliographie

- Aikhenvald, A. Y. et Dixon, R. M. W. (2005). *Serial verb constructions: A cross-linguistic typology*. Oxford : Oxford University Press.
- Alonso-Ramos, M. (2004). Elaboración del Diccionario de colocaciones del español [Élaboration du Dictionnaire des collocations de l'espagnol]. Dans P. Battaner et J. DeCesaris (dir.), *De lexicografía: Actes del I Symposium internacional de lexicografía* (vol. 15, p. 149-162). Barcelona : Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Alonso-Ramos, M. et Tutin, A. (1996). Classification and description of lexical functions for the analysis of their combination. Dans L. Wanner (dir.), *Lexical Functions in Lexicography and Natural Language Processing* (p. 147–168). Amsterdam/Philadelphia : John Benjamins.
- ATILF. (2017). Réseau lexical du français (RL-fr). *ORTOLANG (Open Resources and TOols for LANGuage)*, <https://hdl.handle.net/11403/lexical-system-fr/v1>.
- Barnbrook, G., Mason, O. et Krishnamurthy, R. (2013). *Collocation: applications and implications*. Palgrave Macmillan, London.
- Bloomfield, L. (1933). *Language*. London : Allen and Unwin.
- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. Dans *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL)* (p. 31–40). Potsdam, Germany.
- Chao, Y. R. (1948). *Mandarin primer, an intensive course in spoken Chinese*. Cambridge : Harvard University Press.
- Chao, Y. R. (1965). *A grammar of spoken Chinese*. Berkeley : University of California Press.
- Chen, C. (1994). 论动后趋向动词的性质——兼谈趋向动词研究的方法 [Sur la nature des verbes de déplacement postérieurs des verbes - ainsi que sur la méthode d'étude des verbes de déplacement]. *鲁东大学学报: 哲学社会科学版 [Journal de l'Université de Ludong: Philosophie et sciences sociales]*, 4, 64–70.

- Chen, Yaju. (2006). *现代汉语词语搭配自动抽取方法 [Méthode d'extraction automatique des collocations du mandarin moderne]* (Mémoire de maîtrise). Université Normale de l'Est de Chine à Shanghai, Shanghai.
- Chen, Ying, Li, Y. et Hung, J. (2009). 非监督的汉语感情语料库的构建及分析 [Construction of Chinese Emotion Corpus with an unsupervised approach]. Dans M. Sun et Q. Chen (dir.), *中国计算机语言学研究前沿进展 (2007-2009)[Progress à la frontière de la recherche en linguistique informatique en Chine 2007-2009]* (p. 325-331). Beijing : 清华大学出版社.
- Chung, K. S. (2006). *Mandarin compound verbs*. Taipei : Crane Publishing CO. Ltd.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge/New York : Cambridge University Press.
- Daille, B. (1994, 1 janvier). *Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques* (Thèse de doctorat). Paris 7.
- DeFrancis, J. (1986). *The Chinese language: Fact and fantasy*. Honolulu : University of Hawaii Press.
- Delplanque, A. (1998). Le mythe des « séries verbales ». *Faits de langues*, 6(11), 231–250.
- Diachenko, P. (2006). Lexical functions in learning the lexicon. *Current Developments in Technology-Assisted Education*, 1, 538–542.
- Dong, X. (2002). *词汇化: 汉语双音词的衍生和发展 [Lexicalisation: la dérivation et le développement des mots dissyllabiques en mandarin]* (1ère ed.). Chengdu : 四川民族出版社.
- Dong, X. (2017). 动词后虚化完结成分的使用特点及性质 [On post-verbal completive elements]. *中国语文 [Studies of the Chinese Language]*, 3, 290-298.
- Dong, Y. (2015). *The prosody and morphology of elastic words in Chinese: annotations and analyses* (Thèse de doctorat). University of Michigan.
- Duanmu, S. (1997). Wordhood in Chinese. Dans J. L. Packard (dir.), *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese* (p. 135–196). Berlin/New York : Mouton de Gruyter.
- Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). Oxford/New York : Oxford University Press.

- Duanmu, S. (2012). Word-length preferences in Chinese: a corpus study. *Journal of East Asian Linguistics*, 21(1), 89–114.
- Duanmu, S. (2013). How many Chinese words have elastic length. *Eastward flows the Great river: Festschrift in honor of Prof. William S.-Y. Wang on his 80th birthday*, 1–14.
- Duanmu, S. et Dong, Y. (2016). Elastic words in Chinese. Dans S.-W. Chan (dir.), *Routledge Encyclopedia of the Chinese Language* (p. 452–468). London/New York : Taylor & Francis Ltd.
- Gader, N., Ollinger, S. et Polguère, A. (2014). One lexicon, two structures: So what gives? Dans *Seventh Global Wordnet Conference (GWC2014)* (p. 163–171). Tartu : Global WordNet Association.
- Gao, M. (1953). 关于汉语的词类分别 [Distinction des parties du discours en mandarin]. *中国语文 [Studies of the Chinese Language]*, 10, 13–16.
- Guo, R. (2002). *现代汉语词类研究 [Étude des parties du discours en mandarin moderne]*. Beijing : 商务印书馆.
- Guo, S. (1938). 中国语词之弹性作用 [The function of elastic word length in Chinese]. *燕京学报 [Yenching Journal of Chinese Studies]*, 24, 1-34.
- Huang, B. et Liao, X. (2002). *现代汉语 [Mandarin moderne]*. Beijing : 高等教育出版社.
- Huang, C.-T. J. (1984). Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association*, 19(2), 53–78.
- Jousse, A.-L. (2010). *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales* (Thèse de doctorat). Université de Montréal/Université Paris 7.
- Jousse, A.-L. et Polguère, A. (2005). *Le DiCo et sa version Dicouèbe: Document descriptif et manuel d'utilisation*. Montréal : OLST, Université de Montréal.
- Kahane, S. et Polguère, A. (2001). Formal foundation of lexical functions. Dans *Proceedings of the ACL Workshop on Collocation: Computational Extraction, Analysis and Exploitation* (p. 8–15). Toulouse.
- Karlgren, B. (1923). *Sound & Symbol in Chinese*. Oxford : Oxford University Press.
- Krylosova, S. (2017). *О проекте создания лексической сети русского языка (RL-RU) [À propos du projet de création d'un réseau de russe (RL-RU)]*. Paris : Centre de recherches Europes-Eurasie, Institut national des langues et civilisations orientales (CREE, INaLCO).

- Krylosova, S., Polguère, A., Chepiga, V. et Baïandina-Soujaeff, N. (2014). Réseau lexical du français (RLF) et Réseau lexical du russe (RL-ru). Dans *Dictionnaires et corpus à l'INaLCO (Institut national des langues et civilisations orientales)*. Paris. Repéré à <https://hal.archives-ouvertes.fr/hal-01463410/>
- Lambrey, F. et Lareau, F. (2015). Le traitement des collocations en génération de texte multilingue. Dans *Actes de TALN 2015* (p. 579–585). Caen.
- Lareau, F., Lambrey, F., Dubinskaite, I., Galarreta-Piquette, D. et Nejat, M. (2018). GenDR: A Generic Deep Realizer with Complex Lexicalization. Dans *Proceedings of 11th Edition of the Language Resources and Evaluation Conference (LREC)*. Miyazaki.
- L'Homme, M.-C. (2008). Le DiCoInfo–Méthodologie pour une nouvelle génération de dictionnaires spécialisés. *Traduire. Revue française de la traduction*, (217), 78–103.
- L'Homme, M.-C. (2009). *DiCoInfo, Dictionnaire fondamental de l'informatique et de l'Internet*. Montréal : OLST, Université de Montréal. Repéré à <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>
- L'Homme, M.-C. (2017). Définition des statuts dans le DiCoLilex. Document interne non publié.
- L'Homme, M.-C. et Laneville, M.-E. (2009). *DiCoEnviro. Le dictionnaire fondamental de l'environnement*. Montréal : OLST, Université de Montréal. Repéré à <http://olst.ling.umontreal.ca/dicoenviro/manuel-DiCoEnviro.pdf>
- Li, C. N. et Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley : University of California Press.
- Liu, Y. (1998). *趋向补语通释 [Interprétation générale des compléments directionnels]*. Beijing : 北京语言文化大学出版社.
- Lü, S. (1955). *汉语的词类问题 [Questions sur les parties du discours en mandarin]*. Beijing : 中华书局.
- Lü, S. (dir.). (1999). *现代汉语八百词 (增订本) [Huit cents mots du mandarin moderne]*. 北京 : 商务印书馆.
- Lü, S. et Zhu, D. (1952). *语法修辞讲话 [Introduction sur la grammaire et la rhétorique]*. Beijing : 中国青年出版社.

- Lu, Z. (1953). 汉语的词的分类 [Classification des mots en mandarin]. *语文学学习 [Études du chinois]*, 12, 13–14.
- Lux-Pogodalla, V. et Polguère, A. (2011). Construction of a French Lexical Network: Methodological Issues. Dans *First International Workshop on Lexical Resources, WoLeR 2011* (p. 54-61). Ljubljana, Slovenia.
- Ma, J. (1898). *马氏文通 [Principes de base pour écrire clairement et de manière cohérente de Maître Ma]*. Beijing : 商务印书馆.
- Marengo, S. et Robichaud, B. (2017). Une nouvelle ressource pour la description des collocations : le DicoLiLex, dictionnaire de liens lexicaux. Dans *Représentations du sens linguistique VII (RSL VII)*. Université de Sherbrooke, Sherbrook.
- McCawley, J. D. (1992). Justifying part-of-speech assignments in Mandarin Chinese. *Journal of Chinese linguistics*, 20(2), 211–246.
- Mei, J. (1999). *现代汉语搭配词典 [Dictionnaire de collocation du mandarin moderne]*. Shanghai : 汉语大词典出版社.
- Mel'čuk, I. (1993). *Cours de morphologie générale (théorique et descriptive): Vol. 1. Introduction et première partie: Le mot*. Montreal: Les presses de l'Université de Montréal.
- Mel'čuk, I. (1996). Lexical functions: a tool for the description of lexical relations in a lexicon. *Lexical functions in lexicography and natural language processing*, 31, 37–102.
- Mel'čuk, I. (1998). Collocations and lexical functions. Dans A. P. Cowie (dir.), *Phraseology. Theory, analysis, and applications* (p. 23–53). Oxford : Clarendon Press.
- Mel'čuk, I. (2001). *Communicative organization in natural language*. Amsterdam : John Benjamins.
- Mel'čuk, I. (2003). Collocations dans le dictionnaire. Dans T. Szende (dir.), *Les écarts culturels dans les dictionnaires bilingues* (p. 19–64). Paris : Honoré Champion.
- Mel'čuk, I. (2008). Phraséologie dans la langue et dans le dictionnaire. Dans *Repères & Applications VI, Actes des XXIVe Journées Pédagogiques sur l'Enseignement du Français en Espagne* (p. 187–200). Barcelone.
- Mel'čuk, I. (2013). Tout ce que nous voulions savoir sur les phrasèmes, mais.... *Cahiers de lexicologie*, 102(1), 129–149.

- Mel'čuk, I. (2015). Clichés, an understudied subclass of phrasemes. *Yearbook of Phraseology*, 6(1), 55–86.
- Mel'čuk, I., Arbatchewsky-Jumarie, N. et Iordanskaja, L. (1982). *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques* (vol. I-IV). Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk, I., Clas, A. et Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Mel'čuk, I. et Milićević, J. (2014a). *Introduction à la linguistique* (vol. 1). Paris : Hermann.
- Mel'čuk, I. et Milićević, J. (2014b). *Introduction à la linguistique* (vol. 3). Paris : Hermann.
- Milićević, J. (1997). *Étiquettes sémantiques dans un dictionnaire formalisé du type Dictionnaire Explicatif et Combinatoire* (Thèse de doctorat). Université de Montréal, Montréal.
- Müller, S. (2002). *Complex predicates: Verbal complexes, resultative constructions, and particle verbs in German*. Stanford : CSLI Publications.
- Murphy, M. L. (2010). *Lexical meaning*. Cambridge/New York : Cambridge University Press.
- National Institute of Korean Language. (2006). *세종전자사전 결과보고서 [A Report on Sejong Electric Dictionary]*. Seoul : Hanul Publishing Company.
- Nguyen, É. V. T. (2006). *Unité lexicale et morphologie en chinois mandarin : vers l'élaboration d'un dictionnaire explicatif et combinatoire du chinois* (Thèse de doctorat). Université de Montréal.
- Norman, J. (1988). *Chinese*. Cambridge : Cambridge University Press.
- Packard, J. L. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge : Cambridge University Press.
- Polguère, A. (2007). Lexical function standardness. Dans *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Mel'čuk* (p. 43–95). Amsterdam/Philadelphia : John Benjamins Publishing.
- Polguère, A. et Mel'čuk, I. (2006). Dérivations sémantiques et collocations dans le DiCo/LAF. *Langue française*, 150(2), 66-83.
- Role, F. et Nadif, M. (2011). Handling the Impact of Low Frequency Events on Co-occurrence based Measures of Word Similarity - A Case Study of Pointwise Mutual Information. Dans

- Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (p. 226–231). SciTePress.
- Shen, J. (2016). *名词和动词 [Noms et verbes]*. Beijing : 商务印书馆.
- Sproat, R. et Shih, C. (1988). Prenominal adjectival ordering in English and Mandarin. Dans *Proceedings of NELS 18, vol.2* (p. 465–489). Amherst : GLSA.
- Wang, L., Lu, J., Fu, H., Ma, Z. et Su, P. (1993). *现代汉语 [Mandarin moderne]* (1ère ed.). Beijing : 商务印书馆.
- Zhang, S. et Lin, X. (dir.). (1992). *现代汉语实词搭配词典 [Dictionnaire de collocation des mots lexicaux du mandarin moderne]* (1ère éd.). Beijing : 商务印书馆.
- Zhao, K. (1994). *古代汉语词汇学 [Lexicologie du chinois classique]* (1ère éd.). Beijing : 商务印书馆.
- Zhao, X. (2017). *Études comparatives sur la polysémie des termes des organes internes du corps humain en français et en chinois dans la perspective cognitive* (Mémoire de maîtrise). Université des études internationales de Shanghai.
- Zhou, J. (1999). 双字组合与词典收条 [Double character combinations and dictionary entries]. *中国语文 [Studies of the Chinese Language]*, (04), 304-309.
- Zhu, D. (1982). *语法讲义 [Cours de grammaire]*. Beijing : 商务印书馆.
- Žolkovskij, A. K. et Mel'čuk, I. (1967). O semantičeskom sinteze [A propos de la synthèse sémantique]. *Problemy kibernetiki*, 19, 177-238.

Annexe 1. Structure XML du DiCoLiLex-FR

fichier «_flexidico.rnc» :

```
# Créé par Benoit Robichaud, le 1 mars 2016.
# Édité par François Lareau, le 3 mai 2016.
# Schema validant la structure XML du FlexiDiCo

default namespace = ""
namespace rng = "http://relaxng.org/ns/structure/1.0"
namespace xlink = "http://www.w3.org/1999/xlink"

start =
  element mot { element vocable { vocable }+ }

vocable =
  attribute identificateur { text },
  attribute numéro { text }?,
  attribute dico { text }?,
  element caractéristiques { caractéristiques }?,
  element remarque { text }*,
  element lexie { lexie }+

lexie =
  attribute identificateur { xsd:integer },
  attribute numéro { text }?,
  attribute source { text }?,
  attribute date { xsd:date },
  attribute rédacteur { text }?,
  attribute statut { xsd:integer },
  element caractéristiques { caractéristiques }?,
  element étiquette-sémantique { text },
  element forme-propositionnelle { forme-propositionnelle },
  element remarque { text }*,
  element fonction-lexicale { fonction-lexicale }+,
  element exemple { exemple }*

caractéristiques =
  element pdd { "adj." | "loc. adj." | "loc. adj./adv." | "loc. adv." | "loc.
nom." | "loc. prép." | "loc. verb." | "n." | "v." }?,
  element genre { "f." | "m." | "m. ou f." }?,
  element définitude { "déf." | "déf. ou avec A-poss" | "indéf." | "ne s'emploie
qu'avec l'article partitif" | "non comptable" | "pas d'ART-indéf ni d'A-poss" |
"seulement avec l'article défini" | "surtout déf." | "surtout indéf." }?,
  element nombre { "pl." | "sing." | "surtout pl." | "surtout sing." }?,
  element marque-usage { "affectueux" | "angl." | "environnement" | "fam." |
"formel" | "informatique" | "litt." | "péjor." | "pop." | "rare" | "spéc." |
"vieilli" | "vieillissant" | "vulg." }*,
  element autres { text }*

forme-propositionnelle =
  attribute rôle-sémantique { text }*,
  mixed { (element lexie-ref { text } | element actant { actant }* | element
étiquette-sémantique { text })* },
  text
```

```

actant =
  attribute numéro { "1" | "2" | "3" | "4" | "5" | "1+2" | "1+3" | "2+3" | "2+4" |
"3+4" }?,
  attribute rôle { "Agent" | "Destination" | "Matériau" | "Patient" | "Cause" |
"Bénéficiaire" | "Source" | "" }?,
  text

fonction-lexicale =
  element fonction { fonction },
  element glose { glose }?,
  element remarque { text }*,
  element valeur { valeur }+,
  element contrainte { contrainte }*

fonction =
  attribute type { "standard" | "non standard" },
  mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ | element étiquette-sémantique { text }+ | element marque-usage
{ text }* }*,
  text

glose =
  mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ }*,
  text

valeur =
  attribute fusion { xsd:boolean }?,
  attribute gradation { xsd:integer }?,
  attribute statut { xsd:integer },
  element élément { élément }?,
  element marque-usage { text }*,
  element régime { régime }?,
  element contrainte { contrainte }*,
  element exemple { exemple }*,
  element remarque { text }*

élément =
  mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ }*,
  text

contrainte =
  mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ | element étiquette-sémantique { text }+ | element nombre { text }? |
element définitude { text } }*,
  text

régime =
  mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ | element étiquette-sémantique { text }+ | element nombre { text } }*,
  text

exemple =
  attribute source { text },
  mixed { element lexie-ref { text } | element valeur { text }+ | element actant
{ actant }* }*,
  element remarque { text }*

```


Annexe 2. Structure XML du DiCoLiLex-ZH

fichier «_flexidico.rnc»⁵³ :

```
# Créé par Benoit Robichaud, le 1 mars 2016.
# Édité par François Lareau, le 3 mai 2016.
# Édité par François Lareau et Xiaoyu Zhao, le 5 juin 2018.
# Édité par Xiaoyu Zhao, le 29 juin 2018.
# Schema validant la structure XML du DiCoLiLex-ZH

default namespace = ""
namespace xlink = "http://www.w3.org/1999/xlink"
namespace rng = "http://relaxng.org/ns/structure/1.0"
datatypes xsd = "http://www.w3.org/2001/XMLSchema-datatypes"

start =
  element mot { element vocable { vocable }+ }

vocable =
  attribute identificateur { text },
  attribute numéro { text }?,
  attribute dico { text }?,
  attribute pinyin { text },
  element caractéristiques { caractéristiques }?,
  element remarque { text }*,
  element lexie { lexie }+

lexie =
  attribute identificateur { xsd:integer },
  attribute numéro { text }?,
  attribute source { text }?,
  attribute date { xsd:date },
  attribute rédacteur { text }?,
  attribute statut { xsd:integer },
  element glose { text },
  element caractéristiques { caractéristiques }?,
  element étiquette-sémantique { étiquette-sémantique },
  element forme-propositionnelle { forme-propositionnelle },
  element régime { régime }*,
  element remarque { text }*,
  element fonction-lexicale { fonction-lexicale }+,
  element exemple { exemple }*

caractéristiques =
  element pdd { "adj." | "loc. adj." | "loc. adj./adv." | "loc. adv." | "loc.
nom." | "loc. prép." | "loc. verb." | "n." | "v." }?,
  element genre { "f." | "m." | "m. ou f." }?,
  element définitude { "déf." | "déf. ou avec A-poss" | "indéf." | "ne s'emploie
qu'avec l'article partitif" | "non comptable" | "pas d'ART indéf ni d'A-poss" |
"seulement avec l'article défini" | "surtout déf." | "surtout indéf." }?,
  element nombre { "pl." | "sing." | "surtout pl." | "surtout sing." }?,
```

⁵³ Les parties modifiées sont en gras.

```

    element marque-usage { "affectueux" | "angl." | "environnement" | "fam." |
"formel" | "informatique" | "litt." | "péjor." | "pop." | "rare" | "spéc." |
"vieilli" | "vieillissant" | "vulg." }*,
    element autres { text }*

étiquette-sémantique =
    attribute pinyin { text }?,
    attribute glose { text }?,
    text

forme-propositionnelle =
    attribute rôle-sémantique { text }*,
    mixed { (element lexie-ref { text } | element actant { actant }* | element
étiquette-sémantique { étiquette-sémantique })* },
    text

actant =
    attribute numéro { "1" | "2" | "3" | "4" | "5" | "1+2" | "1+3" | "2+3" | "2+4" |
"3+4" }?,
    attribute rôle { "Agent" | "Destination" | "Matériau" | "Patient" | "Cause" |
"Bénéficiaire" | "Source" | "" }?,
    text

fonction-lexicale =
    element fonction { fonction },
    element glose { glose }?,
    element remarque { text }*,
    element valeur { valeur }+,
    element contrainte { contrainte }*

fonction =
    attribute type { "standard" | "non standard" },
    mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ | element étiquette-sémantique { text }+ | element marque-usage
{ text }* }*,
    text

glose =
    mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ }*,
    text

valeur =
    attribute fusion { xsd:boolean }?,
    attribute gradation { xsd:integer }?,
    attribute statut { xsd:integer },
    element régime { régime }?,
    element élément { élément }?,
    element marque-usage { text }*,
    element régime { régime }?,
    element contrainte { contrainte }*,
    element exemple { exemple }*,
    element remarque { text }*

élément =
    attribute pinyin { text }?,
    attribute glose { text }?,
    mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ }*,
    text

```

```
contrainte =
  mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ | element étiquette-sémantique { text }+ }*,
  text

régime =
  mixed { element lexie-ref { text } | element pdd { text } | element actant
{ actant }+ | element étiquette-sémantique { text }+ }*,
  text

exemple =
  attribute source { text },
  mixed { element lexie-ref { text } | element valeur { text }+ | element actant
{ actant }* }*,
  element glose { text },
  element remarque { text }*
```

Annexe 3. Exemples de vocables du DiCoLiLex-ZH

Voir <http://olst.ling.umontreal.ca/Xiaoyu/> pour la nomenclature complète.

Note : **ÉS** = Étiquette sémantique ; **FP** = Forme propositionnelle ; **FL** = Fonction lexicale

爱 (ài) n.

1. ‘affection’

ÉS	正面情感	zhèngmiàn qíngǎn		'sentiment positif'
FP :	个人 X 对 Y 的 ~			
Régime :	N = X 对 N = Y 的 ~			
FL :	Magn	浓浓 [的]	nóngnóng	'dense'
		强烈 [的]	qiángliè	'fort'
		深深 [的]	shēnshēn	'profond'
	AntiMagn	淡淡 [的]	dàndàn	'léger'
		一丝	yīsī	'un fil'
		一点	yīdiǎn	'un peu'
	Ver	真诚 [的]	zhēnchéng	'sincère'
		真挚 [的]	zhēnzhi	'sincère'
	Propt	出于 [~]	chūyú	'de'
	V0	爱_v#1	ài	'aimer'
	QSyn	感情#2	gǎnqíng	'sentiment'
		(litt.) 青睐_n	qīnglài	'faveur'
		喜爱_n	xǐ'ài	'affection'
		喜欢_n	xǐhuān	'affection'
		欣赏_n	xīnshǎng	'admiration'
	QAnti	鄙视_n	bǐshì	'mépris'
		反感_n	fǎngǎn	'répulsion'
		讨厌_n	tǎoyàn	'dégoût'

2. ‘amour’

ÉS	正面情感	zhèngmiàn qíngǎn		'sentiment positif'
FP :	个人 X 对个人 Y 的 ~			
Régime :	N = X 对 N = Y 的 ~			
FL :	Magn	浓烈 [的]	nóngliè	'fort'
		浓浓 [的]	nóngnóng	'dense'
		强烈 [的]	qiángli	'fort'

		深深 [的]	shēnshēn	'profond'
		一生一世 [的]	yìshēngyíshí	'une vie un siècle'
AntiMagn		淡淡 [的]	dàndàn	'léger'
		一丝	yīsī	'un fil'
Ver		真诚 [的]	zhēnchéng	'sincère'
		真挚 [的]	zhēnzhì	'sincère'
Propt		出于 [~]	chūyú	'de'
Figur		// 爱河	àihé	'rivière d'amour'
Gener		// 爱情	àiqíng	'amour-sentiment'
		// 爱意	àiyì	'amour-sentiment'
V0		爱_v	ài	'aimer'
QS1/QS2		爱人	àirén	'amour'
		对象	duìxiàng	'objet'
		恋人	liànrén	'amoureux'
		朋友	péngyǒu	'ami'
		情人	qíngrén	'amant/maîtresse'
QSyn		感情#3	gǎnqíng	'sentiment'
		恋爱_n	liàn'ài	'amour/relation amoureuse'
		喜爱_n	xǐ'ài	'affection'
		喜欢_n#2	xǐhuān	'affection'
QAnti		鄙视_n	bǐshì	'mépris'
		反感_n	fǎngǎn	'répulsion'
		讨厌_n	tǎoyàn	'dégoût'
		厌恶_n	yànwù	'aversion'

3. 'passion'

ÉS	正面情感	zhèngmiàn qínggǎn		'sentiment positif'
FP :	个人 X 对活动 Y 的 ~			
Régime :	N = X 对 Y 的 ~			
FL :	V0	爱_v#3	ài	'aimer'
	QSyn	爱好_n	àihào	'goût'
		喜爱_n	xǐ'ài	'affection'
		喜欢_n#1	xǐhuān	'affection'
QAnti		鄙视_n	bǐshì	'mépris'
		反感_n	fǎngǎn	'répulsion'
		讨厌_n	tǎoyàn	'dégoût'
		厌恶_n	yànwù	'aversion'

爱 (ài) v.

1. 'aimer'

ÉS	感到一种正面情感	gǎndào yī zhǒng zhèngmiàn qínggǎn		'éprouver un sentiment positif'
FP :	个人 X ~ Y			
Régime :	N = X ~ N = Y			
FL :	Magn	(fam.) 超	chāo	'super'
		(fam.) 超级	chāojí	'super'
		好	hǎo	'bien'
		很	hěn	'très'
		非常	fēicháng	'extrêmement'
		热	rè	'chaudement'
		十分	shífēn	'très'
		特别	tèbié	'extraordinairement'
		真	zhēn	'vraiment'
		极	jí	'extrêmement' peut être postposé
		死	sǐ	'à mourrir' postposé
	AntiMagn	有点儿	yǒudiǎnr	'un peu'
	S0	爱_n#1	ài	'aimer' postposé
	Qual2	// 可爱 [的]	kě'ài	'adorable'
		(fam.) // 有爱 [的]	yǒu'ài	'adorable'
	QSyn	喜爱_v	xǐ'ài	'aimer'
		喜欢_v	xǐhuān	'aimer'
		欣赏_v#1	xīnshǎng	'admirer'
	QAnti	鄙视_v	bǐshì	'mépriser'
		反感_v	fǎngǎn	'répugner'
		讨厌_v#1	tǎoyàn	'détester'
		厌恶_v	yànwù	'détester'

2. 'aimer'

ÉS	感到一种正面情感	gǎndào yī zhǒng zhèngmiàn qínggǎn		'éprouver un sentiment positif'
FP :	个人 X ~ 个人 Y			
Régime :	N = X ~ N = Y			
FL :	Magn	(fam.) 超	chāo	'super'
		(fam.) 超级	chāojí	'super'
		好	hǎo	'bien'
		很	hěn	'très'

	非常	fēicháng	'extrêmement'	
	深	shēn	'profondément'	
	深深 [地]	shēnshēn	'profondément'	
	十分	shífēn	'très'	
	特别	tèbié	'extraordinairement'	
	真	zhēn	'vraiment'	
	一生一世	yìshēngyíshí	'une vie un siècle'	postposé
	极	jí	'extrêmement'	peut être postposé
	死	sǐ	'à mourir'	postposé
AntiMagn	有点儿	yǒudiǎnr	'un peu'	
Adv1 Incep	上	shàng	'commencer'	postposé
S0	爱_n#2	ài	'amour'	
QS1	情种	qíngzhǒng	'espèce sentimentale'	
QS1/QS2	爱人	àirén	'amour'	
	对象	duìxiàng	'objet'	
	恋人	liànrén	'amoureux'	
	朋友	péngyǒu	'ami'	
	情人	qíng rén	'amant/maîtresse'	
QSyn	恋爱_v	liàn'ài	'être en relation amoureuse'	
	喜欢_v	xǐhuān	'aimer'	
QAnti	鄙视_v	bǐshì	'mépriser'	
	反感_v	fǎngǎn	'répugner'	
	讨厌_v#1	tǎoyàn	'détester'	
	厌恶_v	yànwù	'détester'	

3. 'aimer'

ÉS 感到一种正面情感 gǎndào yī zhǒng zhèngmiàn qínggǎn 'éprouver un sentiment positif'

FP : 个人 X ~ 活动 Y

Régime : N = X ~ Y

FL :	Magn	(fam.) 超	chāo	'super'
		(fam.) 超级	chāoji	'super'
		好	hǎo	'bien'
		很	hěn	'très'
		非常	fēicháng	'extrêmement'
		热	rè	'chaudement'
		十分	shífēn	'très'
		特别	tèbié	'extraordinairement'
		十分	shífēn	'très'
		特别	tèbié	'extraordinairement'

		极	jí	'extrêmement'	peut être postposé
		死	sǐ	'à mourir'	postposé
AntiMagn		有点儿	yǒudiǎnr	'un peu'	
Adv1Incep		上	shàng	'commencer'	postposé
S0		爱_n#3	ài	'amour'	
QSyn		爱好_v	àihào	'aimer'	
		喜爱_v	xǐ'ài	'aimer'	
		喜欢_v#1	xǐhuān	'aimer'	
QAnti		鄙视_v	bǐshì	'mépriser'	
		反感_v	fǎngǎn	'répugner'	
		讨厌_v#1	tǎoyàn	'détester'	
		厌恶_v	yànwù	'détester'	

悲伤 (bēishāng) n.

1. 'tristesse'

ÉS	负面情感	fùmiàn qíngǎn		'sentiment négatif'
FP :	个人 X 因为 Y 的 ~			
Régime :	N = X 的 ~			
FL :	Magn	极大 [(的)]	jídà	'extrême'
		巨大 [的]	jùdà	'énorme'
		浓浓 [(的)]	nóngnóng	'dense'
		强烈 [(的)]	qiángliè	'fort'
		深深 [(的)]	shēnshēn	'profond'
		特别 [的]	tèbié	'spécial'
	AntiMagn	淡淡 [(的)]	dàndàn	'léger'
		一丝	yīsī	'un fil'
		一点	yīdiǎn	'un peu'
	Oper1	感到 [~]	gǎndào	'ressentir'
		有 [~]	yǒu	'avoir'
	Magn+Oper1	充满 [~]	chōngmǎn	'se remplir de'
		深感 [~]	shēngǎn	'ressentir profondément'
		满怀 [~]	mǎnhuái	'avoir le cœur plein de'
	IncepOper1	陷入 [~]	xiànrù	's'enfoncer'
	ContOper1	怀着 [~]	huáizhe	'porter'
		心怀 [~]	xīnhuái	'porter au cœur'
	FinOper1	走出 [~]	zǒuchū	'sortir de'
	FinFunc0	释怀	shìhuái	'éliminer au cœur'
	Liqu1Func0	释怀 [~]	shìhuái	'éliminer au cœur'

LiquFunc0	消除 [~]	xiāochú	'éliminer'
Reall	沉浸 [(于/在)~(中)]	chénjìn	'se prolonger'
	沉溺 [(于/在)~(中)]	chénì	's'abîmer'
AntiReall	承受 [~]	chéngshòu	'endurer'
Perm1Manif	表达 [~]	biāodá	'exprimer'
	流露出 [~]	liúlù chū	'révéler'
	露出 [~]	lùchū	'montrer'
NonPerm1Manif	掩饰 [~]	yǎnshì	'dissimuler'
	隐藏 [~]	yǐncáng	'cacher'
Gener	// 悲情_n	bēiqíng	'triste-sentiment'
	[~] 感	gǎn	'sentiment'
	[~] 情绪	qíngxù	'émotion'
	[~] 心情	xīnqíng	'humeur'
V0	悲伤_v#1	bēishāng	'être triste'
QSyn	难过_n	nánguò	'tristesse'
	伤悲_n	shāngbēi	'tristesse'
	伤心_n	shāngxīn	'tristesse (litt. 'cœur blessé)'
	痛苦_n	tòngkǔ	'douleur'
	忧伤_n	yōushāng	'chagrin'
QAnti	高兴_n	gāoxìng	'contentement'
	开心_n	kāixīn	'joie'
	快乐_n	kuàilè	'joie'
	幸福_n	xìngfú	'bonheur'

悲伤 (bēishāng) v.

1. 'être triste'

ÉS 感到一种负面情感 gǎndào yī zhǒng fùmiàn qínggǎn 'éprouver un sentiment négatif'

FP : 个人 X 因为 Y 而 ~

Régime : N = X ~

FL :	Magn	(fam.) 超	chāo	'super'
		(fam.) 超级	chāojí	'super'
		好	hǎo	'bien'
		很	hěn	'très'
		非常	fēicháng	'extrêmement'
		十分	shífēn	'très'
		特别	tèbié	'extraordinairement'
		真	zhēn	'vraiment'

	极	jí	'extrêmement'	peut être postposé
	死	sǐ	'à mourir'	postposé
	(litt.) // 肝肠寸断	gānchángcùnduàn	'Le foie et les intestins sont découpés en pouces'	
AntiMagn	有点儿	yǒudiǎnr	'un peu'	
Qual1	// 多愁善感 [的]	duōchóushàngǎn	'mélancolique et sentimental'	
Qual2	// 可悲 [的]	kěbēi	'déplorable'	
	// 悲情_v [的]	bēiqíng	'triste'	
	// 悲伤_v#3 [的]	bēishāng	'triste'	
S0	悲伤_n	bēishāng	'tristesse'	
S1	悲伤者	bēishāngzhě	'personne triste'	
QSyn	难过_v	nánguò	'être triste'	
	伤悲_v	shāngbēi	'être triste'	
	伤心_v	shāngxīn	'être triste'	
	痛苦_v#1	tòngkǔ	'être douloureux'	
QAnti	高兴_v#1	gāoxìng	'être content'	
	开心_v#1	kāixīn	'être joyeux'	
	快乐_v#1	kuàilè	'être joyeux'	
	幸福_v	xìngfú	'être heureux'	

2. 'être triste'

ÉS	表现出一种负面情感	biǎoxiàn chū yī zhǒng fùmiàn qínggǎn	'exprimer un sentiment négatif'
FP :	表情 X ~		
Régime :	N = X ~ ~ 的 N = X		
FL :	Magn	(fam.) 超	chāo 'super'
		(fam.) 超级	chāojí 'super'
		好	hǎo 'bien'
		很	hěn 'très'
		非常	fēicháng 'extrêmement'
		十分	shífēn 'très'
		特别	tèbié 'extraordinairement'
AntiMagn	有点儿	yǒudiǎnr	'un peu'
QSyn	忧伤_v	yōushāng	'être chagrin'
	难过_v	nánguò	'être triste'
	伤心_v	shāngxīn	'être triste'
	痛苦_v#2	tòngkǔ	'être douloureux'
QAnti	开心_v#2	kāixīn	'être joyeux'
	快乐_v#2	kuàilè	'être joyeux'
	幸福_v	xìngfú	'être heureux'

3. ‘être triste’

ÉS	引起一种负面情感	yǐnqǐ yī zhǒng fùmiàn qínggǎn		'provoquer un sentiment négatif'
FP :	事件 X ~			
Régime :	N = X ~ ~ 的 N = X			
FL :	Magn	(fam.) 超	chāo	'super'
		(fam.) 超级	chāojí	'super'
		好	hǎo	'bien'
		很	hěn	'très'
		非常	fēicháng	'extrêmement'
		十分	shífēn	'très'
		特别	tèbié	'extraordinairement'
	AntiMagn	有点儿	yǒudiǎnr	'un peu'
	QSyn	可悲	kěbēi	'déplorable'
		难过_v	nánguò	'être triste'
		伤心_v	shāngxīn	'être triste'
		痛苦_v#3	tòngkǔ	'être douloureux'
	QAnti	高兴_v#1	gāoxìng	'être content'
		欢喜	huānxǐ	'être joyeux'
		开心_v#3	kāixīn	'être joyeux'
		快乐_v#3	kuàilè	'être joyeux'
		幸福_v	xìngfú	'être heureux'
		愉快	yúkuài	'être gai'

吃醋 (chīcù) loc.verb.

1. ‘être jaloux (litt. ‘manger vinaigre’) (souvent dans une relation amoureuse)’

ÉS	感到一种负面情感	gǎndào yī zhǒng fùmiàn qínggǎn		'éprouver un sentiment négatif'
FP :	个人 X 吃 个人 Y 的醋			
Régime :	N = X 吃 N = Y 的醋 N = X ~			
FL :	Magn	(fam.)超	chāo	'super'
		(fam.)超级	chāojí	'super'
		好	hǎo	'bien'
		很	hěn	'très'
		非常	fēicháng	'extrêmement'
		十分	shífēn	'très'
		特别	tèbié	'extraordinairement'
		大	dà	'grand'

modifie 醋

	酸	suān	'acide'	modifie 醋
AntiMagn	有点儿	yǒudiǎnr	'un peu'	
	小	xiǎo	'petit'	modifie 醋
Ver	真	zhēn	'vraiment'	
A0NonManif	闷	mèn	'étouffé'	modifie 醋
S0	醋劲儿	cùjìn	'jalousie (litt. 'attitude de vinaigre)'	
	醋心	cùxīn	'jalousie (litt. 'cœur de vinaigre)'	
	醋意	cùyì	'jalousie (litt. 'sentiment de vinaigre)'	
S1--usual	醋坛子 [~]	cùtánzi	'personne jalouse (litt. 'pot de vinaigre)'	
	醋罐子 [~]	cùguànzǐ	'personne jalouse (litt. 'cruche de vinaigre)'	
QSyn	嫉妒_v	jídù	'être jaloux'	
	妒忌_v	dùjì	'être jaloux'	
	眼红	yǎnhóng	'être jaloux (litt. 'yeux rougir)'	

担心 (dānxīn) n.

1. 'inquiétude (litt. 'porter coeur)'

ÉS	负面情感	fùmiàn qínggǎn		'sentiment négatif'	
FP :	个人 X 对 Y 的 ~				
Régime :	N = X 对 N = Y 的 ~				
FL :	Magn	浓浓 [的]	nóngnóng	'dense'	
		强烈 [(的)]	qiángliè	'fort'	
	AntiMagn	一丝	yīsī	'un fil'	
	Oper1	感到 [~]	gǎndào	'ressentir'	pas d'actant Y sur 担心
		有 [~]	yǒu	'avoir'	pas d'actant Y sur 担心
	Oper12	[对/为 N = Y] 感到 [~]	gǎndào	'ressentir'	
		[对 N = Y] 有 [~]	yǒu	'avoir'	
	Magn+Oper1	充满 [~]	chōngmǎn	'se remplir de'	
		深感 [~]	shēngǎn	'ressentir profondément'	
	Perm1Manif	流露出 [~]	liúlù chū	'révéler'	
		露出 [~]	lùchū	'montrer'	
	NonPerm1Manif	掩饰 [~]	yǎnshì	'dissimuler'	
	V0	担心_v#1	dānxīn	's'inquiéter'	
	QSyn	不安_n	bù'ān	'trouble'	
		担忧_n	dānyōu	'inquiétude'	
		烦恼_n#1	fǎnnǎo	'tracas'	
		害怕_n	hàipà	'peur'	
		焦虑_n	jiāolǜ	'anxiété'	

恐惧_n	kǒngjù	'peur'
忐忑_n	tǎntè	'inquiétude'
忧虑_n	yōulǜ	'souci'

担心 (dānxīn) loc. verb.

1. 's'inquiéter'

ÉS	感到一种负面情感	gǎndào yī zhǒng fùmiàn qínggǎn	'éprouver un sentiment négatif'		
FP :	个人 X ~ Y				
Régime :	N = X ~ Y N = X 为 N = Y ~				
FL :	Magn	(fam.) 超 (fam.) 超级 好 很 非常 十分 真 特别 坏 极 死 (litt.) // 提心吊胆	chāo chāojí hǎo hěn fēicháng shífēn zhēn tèbié huài jí sǐ tíxīndiàodǎn	'super' 'super' 'bien' 'très' 'extrêmement' 'très' 'vraiment' 'extraordinairement' 'terriblement' 'extrêmement' 'à mourir' 'porter le cœur et pendre la bile'	
	AntiMagn	有点儿	yǒudiǎnr	'un peu'	
	S0	担心_n	dānxīn	'inquiétude'	
	QSyn	不安_v#1 担忧_v 烦恼_v#1 焦虑_v 恐惧_v#1 忐忑_v 忧虑_v	bù'ān dānyōu fánnǎo jiāolù kǒngjù tǎntè yōulǜ	'être troublé' 's'inquiéter' 'se tracasser' 'être anxieux' 's'effrayer' 's'inquiéter' 'être soucieux'	
	QAnti	安宁_v 安心 放心	ānníng ānxīn fàngxīn	'calme' 'se rassurer (litt. 'calmer le coeur)' 'se rassurer (litt. 'déposer le coeur)'	

2. 'être inquiet'

ÉS	表现出一种负面情感	biǎoxiàn chū yī zhǒng fùmiàn qínggǎn	'exprimer un sentiment négatif'
FP :	表情 X ~		

Régime :	~ 的 N = X				
FL :	Magn	很	hěn	'très'	
		非常	fēicháng	'extrêmement'	
		十分	shífēn	'très'	
		特别	tèbié	'extraordinairement'	
	QSyn	不安_v#2	bù'ān	'être troublé'	
		担忧_v	dānyōu	's'inquiéter'	
		烦恼_v#2	fǎnnǎo	'se tracasser'	
		焦躁_v	jiāozào	'être nerveux'	
		焦虑_v	jiāolù	'être anxieux'	
		焦灼_v	jiāozhuó	'se tourmenter'	
		紧张_v	jǐnzhāng	'être stressé'	
		忧愁_v#2	yōuchóu	'être triste et soucieux'	
		忧虑_v#2	yōulù	'être soucieux'	
	QAnti	安宁_v	ānníng	'être calme'	
		安心	ānxīn	'se rassurer (litt.'calmer le cœur)'	
		放心	fàngxīn	'se rassurer (litt.'déposer le cœur)'	

恨 (hèn) n.

1. 'haine'

ÉS	负面情感	fùmiàn qíngǎn	'sentiment négatif'		
FP :	个人 X 对 Y 的 ~				
Régime :	N = X 对 N = Y 的 ~				
FL :	Magn	大	dà	'grand'	
	Magn+Oper1	充满 [~]	chōngmǎn	'se remplir de'	
		满怀 [~]	mǎnhuái	'avoir le cœur plein de'	
	Oper1	抱 [~]	bào	'porter'	pas d'actant Y sur 恨
		含 [~]	hán	'garder'	pas d'actant Y sur 恨
		怀 [~]	huái	'porter'	pas d'actant Y sur 恨
		饮 [~]	yǐn	'boire'	pas d'actant Y sur 恨
		有 [~]	yǒu	'avoir'	pas d'actant Y sur 恨
	Oper12	[对 N = Y] 怀 [~]	huái	'porter'	
		[对 N = Y] 有 [~]	yǒu	'avoir'	
	IncepOper1	产生 [~]	chǎnshēng	'produire'	pas d'actant Y sur 恨
		心生 [~]	xīnshēng	'produire au cœur'	pas d'actant Y sur 恨
	IncepOper12	[对 N = Y] 产生 [~]	chǎnshēng	'produire'	
		[对 N = Y] 心生 [~]	xīnshēng	'produire au cœur'	
	Liqu1Func0	雪 [~]	xuě	'laver'	pas d'actant Y sur 恨

Propt	出于 [~]	chūyú	'de'
Gener	// 恨意	hènyì	'haine-sentiment'
V0	恨_v	hèn	'haïr'
QSyn	仇恨_n	chóuhèn	'haine'
	愤恨_n	fènhèn	'indignation'
	怨恨	yuànhèn	'rancune'
	怨气	yuànqì	'ressentiment'
	憎恨_n	zēnghèn	'haine'

恨 (hèn) loc.verb.

1. 'haïr'

ÉS	感到一种负面情感	gǎndào yī zhǒng fùmiàn qínggǎn	'éprouver un sentiment négatif'
FP :	个人 X ~ Y		
Régime :	N = X ~ N = Y		
FL :	Magn	(fam.) 超	chāo 'super'
		(fam.) 超级	chāojí 'super'
		好	hǎo 'bien'
		很	hěn 'très'
		非常	fēicháng 'extrêmement'
		十分	shífēn 'très'
		特别	tèbié 'extraordinairement'
		痛	tòng 'douloureusement'
		真	zhēn 'vraiment'
		极	jí 'extrêmement'
		死	sǐ 'à mourir'
		(litt.) // 不共戴天	bùgòngdàitiān 'ne pouvoir coexister sous le ciel'
		(litt.) // 恨之入骨	hènzhīrùgǔ 'haïr qqn. jusqu'aux os'
		(litt.) // 恨入骨髓	hènrùgǔsuǐ 'haïr qqn. jusqu'aux os'
	AntiMagn	有点儿	yǒudiǎnr 'un peu'
	Qual2	// 可恨 [的]	kěhèn 'haïssable'
		// 可憎 [的]	kězèng 'odieux'
	S0	恨_n	hèn 'haine'
	QS2	仇家	chóujiā 'ennemi'
		仇人	chóurén 'ennemi'
		冤家	yuānjia 'ennemi'
	QSyn	仇恨_v	chóuhèn 'haïr'
		愤恨_v	fènhèn 's'indigner'
		怨恨_v	yuànhèn 'en vouloir à'

憎恨_v zēnghèn 'hair'

高兴 (gāoxìng) n.

1. 'contentement'

ÉS	正面情感	zhèngmiàn qínggǎn		'sentiment positif'
FP :	个人 X 因为 Y 的 ~			
Régime :	N = X 的 ~			
FL :	Magn	特别 [的]	tèbié	'spécial'
		最大 [的]	zuìdà	'le plus grand'
	AntiMagn	一丝	yīsī	'un fil'
		一点	yīdiǎn	'un peu'
	Ver	由衷 [的]	yóuzhōng	'sincère'
	Oper1	感到 [-]	gǎndào	'ressentir'
	Magn+Oper1	深感 [-]	shēngǎn	'ressentir profondément'
	NonPerm1Manif	掩饰 [-]	yǎnshì	'dissimuler'
		// 暗喜	àn xǐ	'cacher'
	V0	高兴_v#1	gāoxìng	'être content'
	QSyn	开心_n	kāixīn	'joie'
		快乐_n	kuàilè	'joie'
		幸福_n	xìngfú	'bonheur'
	QAnti	悲伤_n	bēishāng	'tristesse'
		难过_n	nánguò	'tristesse'
		伤心_n	shāngxīn	'tristesse (litt. 'cœur blessé)'
		痛苦_n	tòngkǔ	'douleur'
		忧伤_n	yōushāng	'chagrin'

高兴 (gāoxìng) loc. verb.

1. 'être content, heureux'

ÉS	感到一种正面情感	gǎndào yī zhǒng zhèngmiàn qínggǎn		'éprouver un sentiment positif'
FP :	个人 X 因为 Y 而 ~			
Régime :	N = X ~ PROP= Y N = X ~			
FL :	Magn	(fam.) 超	chāo	'super'
		(fam.) 超级	chāojí	'super'
		大为	dàwéi	'grandement'
		好	hǎo	'bien'
		很	hěn	'très'
		非常	fēicháng	'extrêmement'

	分外	fēnwài	'particulièrement'	
	十分	shífēn	'très'	
	特别	tèbié	'extraordinairement'	
	真	zhēn	'vraiment'	
	坏	huài	'terriblement'	postposé
	极	jí	'extrêmement'	peut être postposé
	死	sǐ	'à mourir'	postposé
	(litt.) // 兴高采烈	xìnggāocǎiliè	'être pétillant de verve'	
AntiMagn	有点儿	yǒudiǎnr	'un peu'	
Magn+Adv1	// 兴冲冲 [地]	xìngchōngchōng	'gaiement'	
S0	高兴_n	gāoxìng	'contentement'	
QSyn	欢喜	huānxǐ	'être joyeux'	
	开心_v#1	kāixīn	'être joyeux'	
	快乐_v#1	kuàilè	'être joyeux'	
	幸福_v	xìngfú	'être heureux'	
	愉快	yúkuài	'être gai'	
QAnti	悲伤_v#1	bēishāng	'être triste'	
	难过_v	nánguò	'être triste'	
	伤心_v	shāngxīn	'être triste'	
	痛苦_v#1	tòngkǔ	'être douloureux'	
	忧伤_v	yōushāng	'être chagrin'	

2. 'être content'

ÉS	表现出一种正面情感	biǎoxiàn chū yī zhǒng zhèngmiàn qínggǎn	'exprimer un sentiment positif'	
FP :	表情 X ~			
Régime :	N = X ~ ; ~ 的 N = X			
FL :	Magn	很	hěn	'très'
		非常	fēicháng	'extrêmement'
		十分	shífēn	'très'
		特别	tèbié	'extraordinairement'

3. 'aimer'

ÉS	感到一种正面情感	gǎndào yī zhǒng zhèngmiàn qínggǎn	'éprouver un sentiment positif'	
FP :	个人 X ~ 活动 Y			
Régime :	N = X ~ PROP = Y			
FL :	QSyn	爱_v#3	ài	'aimer'
		喜欢_v#2	xǐhuān	'aimer'
	QAnti	鄙视_v	bǐshì	'mépriser'
		反感_v	fǎngǎn	'répugner'

讨厌_v#1	tǎoyàn	'détester'
厌恶_v	yànwù	'détester'

气 (qì) n.

1. 'colère'

ÉS	负面情感	fùmiàn qínggǎn	'sentiment négatif'	
FP :	个人 X 对 Y 的 ~			
Régime :	N = X 对 Y 的 ~			
FL :	Magn	大	dà	'grand'
		很大 [的]	hěndà	'très grand'
	AntiMagn	小	xiǎo	'petit'
	Oper1	生 [~]	shēng	'générer'
		动 [~]	dòng	'bouger'
		有 [~]	yǒu	'avoir'
	IncepOper1	来 [~]	lái	'venir'
	FinOper1	消 [~]	xiāo	'éliminer'
	FinFunc0	平	píng	'apaiser'
		消	xiāo	'éliminer'
	Cul+A1	[~] 头上	tóu shang	'sur la tête'
	V0	气_v#2	qì	'se fâcher'
	QSyn	愤怒_n	fènnù	'fureur'
		火_n	huǒ	'colère'
		火气	huǒqì	'colère'
		怒_n	nù	'colère'
		怒火	nùhuǒ	'colère'
		怒气	nùqì	'colère'
		脾气_n	píqì	'colère'
		气愤_n	qìfèn	'indignation'

气 (qì) v.

1. 'se fâcher'

ÉS	感到一种负面情感	gǎndào yī zhǒng fùmiàn qínggǎn	'éprouver un sentiment négatif'	
FP :	个人 X ~ 因为 Y			
Régime :	N = X ~			
FL :	Magn	(fam.) 超	chāo	'super'
		(fam.) 超级	chāojí	'super'
		好	hǎo	'bien'
		很	hěn	'très'

	非常	fēicháng	'extrêmement'	
	十分	shífēn	'très'	
	特别	tèbié	'extraordinairement'	
	真	zhēn	'vraiment'	
	坏	huài	'terriblement'	postposé
	极	jí	'extrêmement'	peut être postposé
	死	sǐ	'à mourir'	postposé
	(litt.) // 七窍生烟	qīqiàoshēngyān	'très fâché (litt.'sept orifices fument!')	
AntiMagn	有点儿	yǒudiǎnr	'un peu'	
Adv1Magn	(fam.) 炸	zhà	'à exploser'	postposé
Magn+A1	// 气冲冲 [的]	qìchōngchōng	'très fâché'	
	// 气鼓鼓 [的]	qìgǔgǔ	'très fâché'	
	// 气哼哼 [的]	qìhēnghēng	'très fâché'	
	// 气呼呼 [的]	qìhūhū	'très fâché'	
Magn+Adv1	// 气冲冲 [地]	qìchōngchōng	'très fâché'	
	// 气鼓鼓 [地]	qìgǔgǔ	'très fâché'	
	// 气哼哼 [地]	qìhēnghēng	'très fâché'	
	// 气呼呼 [地]	qìhūhū	'très fâché'	
Qual2	// 可气 [的]	kěqì	'irritant'	
S0	气_n	qì	'colère'	
QSyn	愤怒_v	fènnù	'être furieux'	
	火_v	huǒ	'se fâcher'	
	怒_v	nù	'se fâcher'	
	气愤_v	qìfèn	's'indigner'	

2. 'fâcher'

ÉS	引起一种负面情绪		yǐnqǐyī zhǒng fùmiàn qínggǎn	'éprouver un sentiment négatif'
FP :	Y ~ 个人 X			
Régime :	N = Y ~ X			
FL :	Magn	真	zhēn	'vraiment'
		死	sǐ	'à mourir'
				postposé