



**HAL**  
open science

# Constitution et exploitation d'une base de données pour l'enseignement/apprentissage des phrasèmes NAdj du domaine culinaire français auprès d'apprenants non-natifs

Tong Yang

## ► To cite this version:

Tong Yang. Constitution et exploitation d'une base de données pour l'enseignement/apprentissage des phrasèmes NAdj du domaine culinaire français auprès d'apprenants non-natifs. Linguistique. Université de la Sorbonne nouvelle - Paris III, 2019. Français. NNT : 2019PA030049 . tel-02865525

**HAL Id: tel-02865525**

**<https://theses.hal.science/tel-02865525>**

Submitted on 11 Jun 2020

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

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

UNIVERSITE SORBONNE NOUVELLE – PARIS 3  
ED 268 Langage et langues : description, théorisation, transmission  
Spécialité : didactique des langues et des cultures

**Constitution et exploitation d'une base de données  
pour l'enseignement/apprentissage des phrasèmes  
*NAdj* du domaine culinaire français auprès  
d'apprenants non-natifs**

Tong YANG

Thèse dirigée par Cristelle Cavalla et  
codirigée par Jeanne-Marie Debaisieux

Thèse soutenue publiquement le 4 décembre 2019, devant le jury  
composé de :

**Mme. Cristelle CAVALLA**

Professeur, Université de Sorbonne Nouvelle – Paris 3

**Mme. Jeanne-Marie DEBAISIEUX**

Professeur, Université de Sorbonne Nouvelle – Paris 3

**M. Jean-Marc MANGIANTE**

Professeur, Université d'Artois, Rapporteur

**M. Max SILBERZTEIN**

Professeur, Université de Franche-Comté, Rapporteur

## Résumé

Ce projet de thèse s'inscrit dans le cadre de l'enseignement du FOS (Français sur Objectifs Spécifiques) à des cuisiniers étrangers venus travailler dans des restaurants français ou ayant choisi la restauration comme spécialité. L'objectif de notre recherche est donc d'enseigner les phrasèmes *NAdj* du domaine culinaire auprès d'apprenants étrangers niveau A2. L'enseignement/apprentissage de la phraséologie s'avère nécessaire dans les langues de spécialités et la haute fréquence des phrasèmes *NAdj* a attiré notre attention. Plusieurs questions sont alors abordées : où trouver ce lexique spécifique ? Comment les extraire ? Par quelle approche enseignons-nous les phrasèmes sélectionnés ? Pour répondre à ces questions, nous avons fabriqué notre propre corpus *Cuisitext* – écrit et oral – puis nous avons utilisé *NooJ* pour extraire les phrasèmes *NAdj* du corpus. Enfin, nous avons proposé les trois approches d'utilisation des corpus pour l'enseignement/apprentissage des phrasèmes *NAdj* : approche inductive guidée, approche déductive, approche inductive pure.

## Mot-clé

*Cuisitext*, enseignement/apprentissage de la phraséologie, phrasèmes *NAdj*, français sur objectifs spécifiques, français langue étrangère, extraction automatique de la phraséologie par *NooJ*, approche inductive guidée, approche inductive pure, approche déductive, approches d'utilisation du corpus pour l'enseignement/apprentissage, concordancier, carte mentale.

## Titre

Constitution et exploitation d'une base de données pour l'enseignement/apprentissage des phrasèmes *NAdj* du domaine culinaire français auprès d'apprenants non-natifs

## Abstract

This thesis project aims to study the teaching method of FOS (French on Specific Objectives) catering to foreign cooks who come to work in French restaurants or who have chosen catering as a specialty. The objective of our research is therefore to teach the culinary *NAdj* phrasemes to foreign A2 level learners. The teaching/learning of phraseology is required in specialty languages and the high frequency of *NAdj* phrasemes has caught our attention. Several questions are then addressed: where to find this specific lexicon? How to extract them? By which approach do we teach the selected phrasemes? To answer these questions, we made our own corpus *Cuisitext* - written and oral - and then used *NooJ* to extract the *NAdj* phrasemes from the corpus. Finally, we have proposed the three approaches to the use of corpora for the teaching/learning of *NAdj* phrasemes: guided inductive approach, deductive approach, pure inductive approach.

## Key words

*Cuisitext*, Teaching/Learning of phraseology, *NAdj* Phrasemes, French on Specific Objectives, French as a Foreign Language, Automatic Extraction of the Phraseology Through *NooJ*, Guided Inductive Approach, Pure Inductive Approach, Deductive Approach, Approaches to the Use of the Corpus for Teaching/Learning, Concordance Lines, Mind Mapping.

## Title

Creation and Exploitation of a Database for the Teaching/Learning of *NAdj* Phrasemes of the French Culinary Field with Non-Native Learners

## Remerciements

Avant tout, j'exprime ma profonde reconnaissance envers ma directrice de thèse, CRISTELLE CAVALLA, pour m'avoir formé comme chercheur. Je la remercie pour sa patience, sa tolérance, sa disponibilité et ses encouragements dans mes études. Sans elle, ce travail n'aurait pu être réalisé.

J'adresse ma profonde gratitude à ma codirectrice de thèse, JEANNE-MARIE DEBAISIEUX, pour avoir accepté d'encadrer ma thèse. Avec elle, j'ai pu obtenir la bourse du gouvernement chinois.

Je remercie spécifiquement JEAN-MARC MANGIANTE et MAX SILBERZTEIN pour avoir accepté d'être rapporteurs dans mon jury. Je remercie JEAN-MARC MANGIANTE pour le partage de son corpus du domaine culinaire et MAX SILBERZTEIN pour l'organisation des colloques auxquels j'ai pu participer avec joie.

Je tiens à remercier MARION TELLIER et FABRICE BARTHELEMY pour m'avoir accueilli à bras ouvert comme ATER au sein des départements de FLE (Français Langue Etrangère) de leurs universités AMU (Aix-Marseille Université) et UFC (Université de Franche-Comté).

Je remercie sincèrement mes collègues MARION TELLIER (AMU), NATHALIE THAMIN (UFC), KAOUTHAR BEN ABDALLAH (UFC), ANNE-CLAIRE RAIMOND (UFC), ANNE-SOPHIE CALINON (UFC), ALICE BURROWS (Paris 3) et tant d'autres pour leur aide précieuse à la préparation de mes cours de FLE.

Je remercie chaleureusement HENRY TYNE et MIREILLE BILGER pour leurs cours d'initiation à la linguistique de corpus au département de FLE de l'Université de Perpignan. Avec eux, j'ai pu découvrir les concordanciers pour la première fois.

Je remercie vivement DELPHINE MIGNE pour la relecture minutieuse de mes écrits (mémoire et thèse) durant des années.

Je remercie également les laboratoires DILTEC (EA 2288), LATTICE (UMR 8094), LPL (UMR 6057), ELLIADD (EA 4661) et mon école doctorale (ED 268) auxquels je suis rattaché pour le financement de mes déplacements nationaux (Dijon, Grenoble, Rouen, Reims, Valence...) et internationaux (Canada, Maroc, Italie, Autriche, Tunisie...) dans le but de présenter mes travaux au sein de colloques.

Je remercie le CSC (Chinese Scholarships Council) qui m'a accordé une bourse me permettant de poursuivre mes études.

Un merci tout particulier à mes parents, pour leur soutien constant et leurs encouragements incessants.

# Sommaire

Liste des tableaux.....	9
Liste des figures .....	11
Introduction .....	14
Première partie : théorie.....	22
1. Linguistique de corpus .....	26
1.1 Émergence de la linguistique de corpus .....	26
1.2 Construction et analyse du corpus à l'aide d'outils TAL.....	27
1.3 Élaboration de grands corpus français et anglais .....	29
1.4 Choix des types de textes pour la constitution du corpus.....	31
Bilan .....	37
2. Phraséologie .....	39
2.1 Champs de recherche : la phraséologie .....	39
2.2 Classifications des phrasèmes .....	43
2.3 Critères de la phraséologie.....	44
2.4 Modélisations des données dans des projets phraséologiques.....	47
2.5 Logiciel de la description des langues : approche de <i>NooJ</i> .....	49
Bilan .....	59
3. Utilisation du corpus pour l'E/A de la phraséologie .....	60
3.1 E/A de la phraséologie .....	60
3.2 Utilisation des corpus pour l'E/A des langues .....	78
3.3 Outils didactiques pour l'E/A de la phraséologie .....	84
Bilan .....	92
Deuxième partie : méthodologie et analyse linguistique .....	94
4. Constitution de <i>Cuisitext</i> et réflexion sur son exploitation .....	98
4.1 Constitution de corpus écrit.....	98

4.2 Construction de corpus oraux.....	102
4.3 Réflexion sur l'exploitation de <i>Cuisitext</i> dans notre projet .....	112
Bilan .....	113
<b>5. Repérage des phrasèmes saillants et leur sélection pour l'enseignement.....</b>	<b>115</b>
5.1 Objet d'étude : phrasèmes <i>NAdj</i> .....	115
5.2 Extraction des phrasèmes <i>NAdj</i> .....	117
5.3 Identification et sélection des phrasèmes <i>NAdj</i> pour l'enseignement ..	132
Bilan .....	137
<b>6. Analyse linguistique des phrasèmes <i>NAdj</i> retenus.....</b>	<b>138</b>
6.1 Critères servant à la classification .....	138
6.2 Propriétés transformationnelles du figement des phrasèmes <i>NAdj</i> .....	139
6.3 Classification des phrasèmes <i>NAdj</i> .....	142
6.4 Analyse du figement des phrasèmes <i>NAdj</i> .....	143
Bilan .....	149
<b>Troisième partie : proposition didactique .....</b>	<b>151</b>
<b>7. Propositions des approches d'intégration du corpus et leurs mises en œuvre pour l'E/A des phrasèmes <i>NAdj</i>.....</b>	<b>153</b>
7.1 Trois approches proposées.....	155
7.2 Piste d'apprentissage progressif sur corpus .....	157
7.3 Propositions didactiques.....	159
Bilan .....	180
<b>Conclusion générale.....</b>	<b>182</b>
<b>Perspectives et limites de recherche .....</b>	<b>184</b>
<b>Bibliographie .....</b>	<b>187</b>
<b>Annexe .....</b>	<b>231</b>



<b>Annexe 1 Corpus oraux accessibles/inaccessible en ligne .....</b>	<b>231</b>
<b>Annexe 2 Convention de transcription .....</b>	<b>232</b>

## Liste des tableaux

Tableau 1-1 : Trois corpus britanniques et un français .....	30
Tableau 2-1 : Quelques classifications des phrasèmes .....	43
Tableau 2-2 : Quelques fonctions lexicales .....	48
Tableau 3-1 : Ouvrages complémentaires de FLE concernant l'enseignement de la phraséologie .....	63
Tableau 4-1 : Quelques sites français de recettes de cuisine .....	99
Tableau 4-2 : Trois types de corpus oraux dans notre étude .....	103
Tableau 4-3 : Transcriptions phonologiques et orthographiques .....	108
Tableau 4-4 : Comparaisons succinctes des logiciels de transcription .....	110
Tableau 5-1 : Modélisation de la structure attributive .....	120
Tableau 5-2 : Modélisation avec la variable globale .....	120
Tableau 5-3 : Programmes développés pour l'extraction phraséologique ...	124
Tableau 5-4 : Trois grammaires de la structure attributive .....	128
Tableau 5-5 : Phrasèmes pertinents à une occurrence .....	135
Tableau 6-1 : Transformations de <i>plat préféré</i> .....	144
Tableau 6-2 : transformations de <i>crème fraîche</i> .....	145
Tableau 6-3 : Transformation de <i>filet mignon</i> .....	146
Tableau 6-4 : Transformations de <i>plat facile</i> .....	147
Tableau 6-5 : Transformations d' <i>oignon haché</i> .....	147
Tableau 6-6 : Transformations de <i>poivron rouge</i> .....	148
Tableau 6-7 : Transformations de <i>température ambiante</i> .....	149
Tableau 7-1 : Trois courants d'introduction du corpus aux apprenants .....	154
Tableau 7-2 : Proposition de trois approches d'introduction du corpus .....	155
Tableau 7-3 : Trois catégories de HACHÉ .....	163
Tableau 7-4 : Tableau à remplir .....	168
Tableau 7-5 : Propriétés combinatoires de HACHÉ, MOULU, ÉMIETTÉ, RÂPÉ et BRISÉ .....	172

Tableau 7-6 : Exercice lacunaire..... 178

## Liste des figures

Figure 1-1 : Niveaux de classement de Malrieu & Rastier (2001) .....	33
Figure 1-2 : Discours et genre dans le domaine culinaire .....	34
Figure 2-1 : Évolution de la phraséologie.....	40
Figure 2-2 : Classification des phrasèmes chez Mel'čuk (1998 : 30) .....	43
Figure 2-3 : Quatre types de grammaires génératives .....	49
Figure 2-4 : Graphe à états finis.....	51
Figure 2-5 : Graphe d'un lexique.....	52
Figure 2-6 : Graphe des formes nominales.....	52
Figure 2-7 : Graphe de disjonction du lexique .....	53
Figure 2-8 : Graphe de concaténation lexicale .....	53
Figure 2-9 : Utilisation de l'Opérateur +EXCLUDE .....	54
Figure 2-10 : Graphe de la grammaire Phrase .....	56
Figure 2-11 : Grammaire de groupe nominal .....	56
Figure 2-12 : Accord de la structure N+A.....	57
Figure 2-13 : Accord de la structure N+A avec des variables .....	58
Figure 3-1 : Distinction du synonyme discours/conférence .....	61
Figure 3-2 : Traductions de <i>poser un lapin</i> dans différentes langues.....	62
Figure 3-3 : Trois modèles proposés par Weinreich .....	69
Figure 3-4 : Modèle interactif du lexique mental (Boulton, 1998) .....	70
Figure 3-5 : Conseils d'utilisation de l'ouvrage <i>les expressions imagées</i> .....	77
Figure 3-6 : Trois points de vue didactiques pour l'utilisation du corpus .....	80
Figure 3-7 : Allure des KWOC-KWAC-KWIC-KWUT (Pincemin, 2006) .....	85
Figure 3-8 : Concordances du mot clé HACHÉ dans <i>TXM</i> .....	86
Figure 3-9 : Carte mentale : Mode de cuisson .....	90
Figure 4-1 : Choix d'ajout d'URLs dans <i>Gromoteur</i> .....	100
Figure 4-2 : Paramètre de requête du choix des pages dans <i>Gromoteur</i> ....	100
Figure 4-3 : Extrait de données collectées dans <i>Gromoteur</i> .....	101

Figure 4-4 : Filtrage de corpus par un mot clé « photo » .....	102
Figure 4-5 : Consentement de participation .....	106
Figure 4-6 : Extrait de nos transcriptions de <i>CLAN</i> .....	111
Figure 5-1 : Les dix lexèmes les plus spécifiques de <i>Cuisitext</i> .....	115
Figure 5-2 : Deux approches de définition de la phraséologie .....	116
Figure 5-3 : Modélisation des deux structures en quatre grammaires .....	120
Figure 5-4 : Grammaires du GN.....	121
Figure 5-5 : Modélisation des phrasèmes NAdj .....	121
Figure 5-6 : Définition du GN .....	122
Figure 5-7 : Modélisation et désambiguïsation des phrasèmes NAdj .....	123
Figure 5-8 : Extrait de 2-gram par <i>Antconc</i> .....	126
Figure 5-9 : Trois exigences de la modélisation.....	127
Figure 5-10 : Implémentation de la grammaire nominale.....	127
Figure 5-11 : Désambiguïsation de la grammaire nominale.....	128
Figure 5-12 : Implémentation du syntagme nominal .....	128
Figure 5-13 : Implémentation de la grammaire S + VE + A.....	129
Figure 5-14 : Graphe imbriqué du GN.....	129
Figure 5-15 : Grammaire de la Phrase imbriqué dans la grammaire GN .....	130
Figure 5-16 : Implémentation de la structure attributive .....	130
Figure 5-17 : Graphe total.....	131
Figure 5-18 : Séquences <i>NAdj</i> sous formes de concordances .....	132
Figure 5-19 : Phrasèmes <i>NAdj</i> affichés par fréquence .....	134
Figure 5-20 : Comparaison de <i>Cuisitext</i> écrit avec <i>Cuisitext</i> oral.....	136
Figure 6-1 : Expressions figées, expressions semi-figées et groupes ordinaires .....	143
Figure 7-1 : Piste d'apprentissage progressif sur corpus .....	158
Figure 7-2 : Concordances de HACHÉ de <i>Cuisitext</i> .....	161
Figure 7-3 : Carte mentale de HACHÉ avec des aromates différents.....	165

Figure 7-4 : Carte mentale de HACHÉ avec des viandes différentes .....	165
Figure 7-5 : Carte mentale de HACHÉ avec des ingrédients différents .....	166
Figure 7-6 : Carte mentale rassemblée de HACHÉ .....	166
Figure 7-7 : Concordances de MOULU, RÂPÉ, ÉMIETTÉ et BRISÉ de <i>Cuisitext</i> .....	169
Figure 7-8 : Carte mentale de base .....	170
Figure 7-9 : Carte mentale de MOULU .....	173
Figure 7-10 : Carte mentale de RÂPÉ .....	174
Figure 7-11 : Carte mentale d'ÉMIETTÉ .....	174
Figure 7-12 : Carte mentale de BRISÉ .....	175
Figure 7-13 : Carte mentale rassemblée de MOULU, RÂPÉ, ÉMIETTÉ et BRISÉ .....	176

## Introduction

### 1. Contexte de l'étude

Connue de par le monde, la gastronomie française attire chaque année de nombreux cuisiniers étrangers venus travailler dans des restaurants français ou ayant choisi la restauration comme spécialité. Comme l'affirme Wilson, « La France est réputée être le pays de la gastronomie » (1977 : 201). Les cuisiniers-apprenants non-natifs auront du mal à acquérir des connaissances, non seulement en raison de la nouveauté que représente le milieu de la restauration française, mais aussi de par leur niveau en langue parfois inadéquat à la situation. Ces difficultés d'intégration dans la restauration française concernent surtout les apprenants allophones de niveau A2 du CECRL (Cadre Européen Commun de Référence des Langues), car ce dernier est le niveau minimum pour que les étrangers puissent obtenir un visa pour étudier en France. Il faut surtout retenir que ces cuisiniers allophones du français se retrouvent dans un milieu professionnel où – comme dans de nombreux métiers – la langue est importante à maîtriser afin de bien réaliser le travail : l'apprenant doit rapidement comprendre les consignes professionnelles et doit se faire comprendre aussi brièvement auprès de ses collègues de travail. De ce fait, « en raison de l'expansion de la mobilité professionnelle et étudiante » (Carras, 2007 : 2), les formations en langue française orientées vers le domaine culinaire sont de plus en plus sollicitées. Comme en témoigne Hervé Fleury, directeur de l'institut Paul Bocuse, à Écully (Rhône), dans un reportage<sup>1</sup> (publié le 04/02/2008) : « s'il y a un secteur dans lequel la France n'a pas perdu son attractivité, c'est bien la gastronomie. Les candidatures venues d'autres pays dans notre établissement sont même en

---

<sup>1</sup> Adresse du reportage : [https://www.lexpress.fr/styles/saveurs/restaurant/ces-chefs-etrangers-qui-bouleversent-notre-cuisine\\_473537.html](https://www.lexpress.fr/styles/saveurs/restaurant/ces-chefs-etrangers-qui-bouleversent-notre-cuisine_473537.html) (consulté le 10/10/2019).

constante augmentation ». En outre, Fleury ajoute que sur les 330 étudiants de l'institut, la moitié est composée d'étrangers de 37 nationalités différentes. Conçu pour tout apprenant ayant besoin d'acquérir le langage employé au sein de la profession qu'il exerce, le Français sur Objectifs Spécifiques, une méthodologie issue du FLE, analyse les besoins langagiers des apprenants et y répond par une méthodologie d'enseignement appropriée au monde professionnel. Concernant les besoins des apprenants en FOS, le lexique tient souvent une place centrale. Comme le dit Challe (2000 : 79), « La première compétence chez les spécialistes est lexicale ». Binon et Verlinde soulignent également cette importance du lexique dans le FOS et disent « que le vocabulaire joue un rôle-clé dans l'enseignement/apprentissage (désormais E/A) du français sur objectifs spécifiques, nul ne le contestera. » (2004 : 271). Le vocabulaire de la cuisine, varié et spécialisé, nécessite une bonne connaissance pour une utilisation appropriée en milieu professionnel. La connaissance de ce lexique ne peut être obtenue que par une pratique continue. En outre, ne pouvant que très peu se référer à son propre lexique, l'apprenant doit acquérir un lexique bien spécifique, renvoyant parfois à des réalités qu'il ne connaît pas. Le terme « bleu », par exemple, qui désigne – en français – le mode de cuisson d'une viande, n'a pas d'équivalent dans certaines langues-cultures. Une carence lexicale chez nos apprenants a déjà été détectée lors de l'étape de l'analyse des besoins dans le cadre de notre projet de Master et des activités didactiques ont été proposées dans ce cadre pour assurer la compétence lexicale de nos apprenants (Yang, 2015). Le sens d'un lexique doit être étudié dans ses contextes lexicaux. Comme Chambers (2010) le constate dans son projet, à défaut de précision, le nombre d'occurrences du verbe « connaître » (81 occurrences) dans son corpus Chambers-Le Baron (Chambers & Le Baron, 2007), peut créer une ambiguïté sémantique : connaissance d'une personne ou connaissance d'une ville ? L'ambiguïté sémantique du lexique isolé peut être également attribuée à la



polysémie qui soulève des problèmes aussi dans notre étude (Yang, 2015) : le lexème *neige* dans l'expression *battre les blancs en neige* ne représente pas la même entité que le lexème *neige* dans un discours sur la météorologie en montagne ; le lexème *moule* dans le domaine culinaire peut être un modèle d'outillage pour faire un gâteau ou un aliment (fruit de mer). Afin de lever cette ambiguïté sémantique, Blumenthal (2007) proposait d'établir « le profil combinatoire du mot ». Mel'čuk *et al.* (1999) précisent également que la combinatoire lexicale détermine le « sens d'un vocable ou d'une lexie » (ibid. : 13). Tutin (2007a, 2007b, 2014) et Grossmann et Tutin (2006) confirment que, les séquences polylexicales présentent souvent moins d'ambiguïtés sémantiques qu'un mot isolé. L'analyse d'un lexique porte donc sur ses combinaisons lexicales (séquences polylexicales).

## 2. Intérêt de l'E/A de la phraséologie en FOS

La présence abondante des séquences préfabriquées attire l'attention des linguistes : « an incredibly large number of prefabs<sup>2</sup> » (Bolinger, 1976 : 1) ; « A great proportion of the most familiar concepts and speech acts can be expressed formulaically<sup>3</sup> » (Wood, 2002 : 7) ; le nombre de ces séquences peut s'élever à « des dizaines de milliers » (Mel'čuk, 1993 : 83) ; « omniprésent dans le langage » (Pecman, 2004 : 29). Pawley & Syder (1983) abondent dans le même sens et découvrent que « by far the largest part of the English speaker's lexicon consists of complex lexical items including several hundred thousand lexicalized sentence stems<sup>4</sup> » (ibid. : 215). Comme l'observent Pawley & Syder, la plupart du lexique maîtrisé par un anglophone est constitué de ces séquences préfabriquées. Ce pourcentage a pu être

---

<sup>2</sup> Un nombre incroyablement élevé de formes préfabriqués (notre traduction).

<sup>3</sup> Une grande partie des concepts les plus familiers et des actes de langage peuvent être exprimés en formule (notre traduction).

<sup>4</sup> De loin la plus grande partie du lexique d'un anglophone se compose d'items lexicaux complexes comprenant plusieurs centaines de milliers d'éléments préfabriqués lexicalisés (notre traduction).

calculé par des linguistes : pour Lewis (2000), ces séquences représenteraient plus de la moitié de nos discours (oraux et écrits) ; pour Altenberg (1998) et Mel'čuk (1993), ce pourcentage est approximativement de 40% ; Selon Erman & Warren (2000), cette part est approximativement de 58% dans la langue parlée et s'élève à 52% dans la langue écrite. Notons que différentes appellations ont été utilisées par des linguistes pour désigner les expressions préfabriquées ; on trouve « idioms<sup>5</sup> » chez Sweet (1891: 156) ; « groupes articulés », « formules », « locution » (Bréal, 1897 : 186) ; « groupes usuels » (Bally, 1909 : 88) ; « formulae<sup>6</sup> » (Jespersen, 1924 : 18) ; « déjà-vu » ( Bally, 1963 : 70) ; « phrasèmes » (Mel'čuk , 1993 : 83 ; 1995 : 16), « expressions figées » (G. Gross, 1996 : 57), « locutions toutes faites » (Dubois, 1994 : 289), « expressions préfabriquées » et « expressions préformées » (Gulich & Krafft, 1997 : 243), « prefabs<sup>7</sup> » (Erman & Warren, 2000 : 329) et « formulaic sequence<sup>8</sup> » (Wray, 2002 : 9). D'après Howarth (1996), il n'existe pas d'étiquette formelle pour désigner tout le domaine, mais il a constaté que le terme « phraséologie » est de plus en plus utilisé comme dénomination de cette discipline :

*There is, as yet, no generally-agreed overall term encompassing the whole spectrum [of phraseological units], though there seems to be a growing acceptance of phraseology as a convenient name for the field of study [...].<sup>9</sup>*  
(Howarth, 1996 : 6).

Le terme « phraséologie » étant la dénomination la plus représentative dans les recherches rédigées en anglais et en français, nous emploierons ce terme pour désigner ce domaine dans notre étude. Les didacticiens ont constaté la nécessité de l'enseignement de la phraséologie auprès des apprenants

---

<sup>5</sup> Phrases spéciales

<sup>6</sup> Séquences formulaires

<sup>7</sup> Préfabriqués.

<sup>8</sup> Séquence formulaire ou séquence en formule

<sup>9</sup> Il n'existe pas encore de terme généralement admis qui engloberait tout le spectre [des unités phraséologiques], bien qu'on puisse noter une généralisation progressive dans l'emploi du terme phraséologie pour désigner le champ de recherche.

étrangers. Binon et Verlinde (2004) constate que les unités phraséologiques posent souvent problème aux apprenants étrangers à la fois sur le plan réceptif et sur le plan productif. En outre, ils pensent que le nombre d'unités phraséologiques est plus important en FOS qu'en français général (ex. en médecine : *crise cardiaque, insuffisance respiratoire*). En définitive, il semble bien que l'enseignement de la phraséologie auprès des apprenants étrangers en milieu professionnel soit une exigence fondamentale.

### 3. Intérêt de l'utilisation du corpus en FLE

Dans une certaine mesure, la collecte des données est le « centre de gravité » (Mangiante & Parpette, 2004 : 46) de la méthodologie du FOS. Cette dernière inspirée de l'ingénierie de formation, permet d'un côté, de prendre en compte les besoins des apprenants et d'un autre côté, de mettre en avant l'utilisation de documents authentiques professionnels sur lesquels la préparation des cours se fonde. La plus grande difficulté de la collecte des données par les enseignants, réside dans la recherche et la récolte des ressources nécessaires pour préparer un cours (Mangiante et Parpette, 2004 ; Carras, 2007). Le corpus numérique peut pallier cette difficulté en offrant de multiples exemples authentiques (O'Sullivan & Chambers, 2006 ; Yoon & Hirvela, 2004) dont l'application est promue par les approches communicatives et actionnelles de la didactique du FLE (Laurens, 2013).

Un corpus, en linguistique, est « a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language<sup>10</sup> » (Sinclair, 1996 : 4) et son utilisation peut remonter au XVIII<sup>ème</sup> siècle dans certains domaines, entre autres la lexicographie (dictionnaire de Johnson, 1755). L'apport majeur du corpus est de remplacer l'introspection basée sur l'intuition pure ou quelques exemples

---

<sup>10</sup> Un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon au langage (notre traduction).

isolés du linguiste par des phénomènes authentiques des langues. Sachant que l'intuition est parfois un mauvais guide (Sinclair, 2005), le corpus est un révélateur de la langue (Tognini-Bonelli, 2001) et permet d'accéder aux « intuitions de milliers de locuteurs » (Frankenberg-Garcia, 2005 : 192). L'intégration du corpus en classe de langue suscite, au cours des trente dernières années, un intérêt considérable dans différents champs de l'E/A des langues : erreurs corrigées (O'Sullivan & Chambers, 2006 ; Gaskell & Cobb, 2004), écrit académique (Yoon & Hirvela, 2004 ; Sun, 2007 ; O'Sullivan, 2010 ; Granger & Paquot, 2015), acquisition langagière (Colletta, 2004 ; Tomasello, 1998 ; Hoey, 2005 ; Tyne, 2009), compétence en écriture (Gaskell & Cobb, 2004 ; O'Sullivan & Chambers, 2006). D'après Landure & Boulton (2010), cet intérêt considérable pour l'E/A des langues peut être attribué à plusieurs raisons liées à l'interaction entre l'être humain et la machine qui permet aux apprenants

- d'avoir une plus grande sensibilité aux processus d'acquisition (O'Sullivan, 2007 ; Tyne, 2009),
- d'améliorer la mémorisation (Cobb, 1997a ; 1999),
- de susciter la motivation des apprenants (Bérard, 1991).

En outre, l'apprentissage sur corpus relève d'une activité « naturelle » dans laquelle les apprenants vont essayer de relever les régularités de la langue par rapport à une activité « artificielle » qu'est la simple transmission de connaissances (Gaskell & Cobb 2004 : 304 ; Scott & Tribble 2006 ; Boulton 2009a). Certains voient dans l'usage du corpus une façon, de la part des apprenants, de s'approcher de la langue cible au lieu d'apprendre seulement les règles de la grammaire (Loewen *et al.*, 2009). Ainsi, l'usage du corpus peut offrir un « échafaudage » pour que les apprenants puissent construire leurs propres savoirs sans renier aucune théorie d'apprentissage (Kirschner *et al.*, 2006 ; Mayer, 2004). Enfin, l'apprentissage sur corpus peut « mener à une plus grande autonomisation qui lui [l'apprenant] servira dans sa vie professionnelle ultérieure » (Landure & Boulton, 2010 : 3). Par conséquent,

cela s'avère nécessaire d'utiliser le corpus en classe de FLE, ainsi que pour notre projet en FOS.

Cette thèse a donc pour objectif l'utilisation des corpus en classe de FOS pour l'enseignement de la phraséologie culinaire. Pour ce faire, nous avons d'abord fabriqué un corpus – écrit et oral – puis nous avons défini la façon d'extraire des données phraséologiques pour leur enseignement. Enfin, nous avons réfléchi à une méthodologie d'enseignement qui conduirait à l'autonomie des apprenants à la fois face aux corpus numériques et à l'apprentissage des éléments extraits ou à extraire. Voici résumé le plan de ce travail.

#### 4. Plan de cette thèse

Cette thèse est composée de trois parties :

- La première partie, qui comprend trois chapitres, est destinée à faire l'état de l'art de l'utilisation du corpus pour l'E/A des phrasèmes. Le premier chapitre est consacré à présenter la linguistique de corpus, l'apport des outils TAL pour la construction et l'analyse du corpus, l'élaboration de grands corpus français et anglais ainsi que les choix des types de textes pour la constitution du corpus. Le deuxième chapitre explicite notre objet d'étude : la phraséologie ainsi que ses classifications et certains critères servant à la classification. Le troisième chapitre clarifie l'E/A de la phraséologie et l'utilisation du corpus pour l'E/A des langues.
- La deuxième partie, qui est divisée en trois chapitres, porte sur la méthodologie adoptée et l'analyse linguistique des phrasèmes *NAdj*. Le quatrième chapitre montre la méthodologie adoptée pour la construction de notre corpus écrit et oral. Le repérage et la sélection des phrasèmes *NAdj* pour l'enseignement sont mis en avant dans le cinquième chapitre. Dans le sixième chapitre, les phrasèmes *NAdj*

retenus sont analysés et classés.

- La dernière partie, qui est composée d'un seul chapitre, est consacrée à la proposition des approches d'utilisation du corpus et leurs mises en place pour l'E/A des phrasèmes *NAdj*.

Première partie : théorie

<b>1. Linguistique de corpus .....</b>	<b>26</b>
<b>1.1 Émergence de la linguistique de corpus .....</b>	<b>26</b>
<b>1.2 Construction et analyse du corpus à l'aide d'outils TAL.....</b>	<b>27</b>
<b>1.3 Élaboration de grands corpus français et anglais .....</b>	<b>29</b>
<b>1.4 Choix des types de textes pour la constitution du corpus.....</b>	<b>31</b>
1.4.1 Importance du corpus oral .....	31
1.4.2 Genres de textes .....	33
1.4.3 Documents authentiques pour l'enseignement .....	35
<b>Bilan .....</b>	<b>37</b>
<b>2. Phraséologie .....</b>	<b>39</b>
<b>2.1 Champs de recherche : la phraséologie .....</b>	<b>39</b>
<b>2.2 Classifications des phrasèmes .....</b>	<b>43</b>
<b>2.3 Critères de la phraséologie.....</b>	<b>44</b>
2.3.1 Fréquence .....	45
2.3.2 Figement .....	46
2.3.3 Critère pragmatique .....	46
<b>2.4 Modélisations des données dans des projets phraséologiques.....</b>	<b>47</b>
<b>2.5 Logiciel de la description des langues : approche de <i>NooJ</i> .....</b>	<b>49</b>
2.5.1 Grammaire rationnelle .....	50
2.5.2 Grammaire hors contexte ou algébrique .....	55
2.5.3 Grammaire contextuelle.....	56
<b>Bilan .....</b>	<b>59</b>
<b>3. Utilisation du corpus pour l'E/A de la phraséologie .....</b>	<b>60</b>
<b>3.1 E/A de la phraséologie .....</b>	<b>60</b>
3.1.1 Nécessité de l'E/A de la phraséologie en FLE .....	60
3.1.2 E/A des phrasèmes dans les ouvrages de FLE .....	62
3.1.3 Acquisition des langues en L2 .....	65
3.1.3.1 Organisation lexicale et lexique mental.....	65
3.1.3.2 Lexique mental en L1 .....	67



---

3.1.3.3 Lexique mental en L2 .....	68
3.1.4 E/A du lexique au genre.....	71
3.1.5 Prise en compte du critère de fréquence dans la sélection du lexique ....	73
3.1.6 Entrée didactique pour l'enseignement du lexique .....	75
<b>3.2 Utilisation des corpus pour l'E/A des langues .....</b>	<b>78</b>
3.2.1 Corpus pour l'E/A des langues.....	78
3.2.2 Trois manières d'enseigner l'utilisation des corpus .....	79
3.2.3 Approches d'utilisation du corpus pour l'E/A des langues .....	81
3.2.3.1 Approche <i>data-driven learning</i> (Johns, 1991a ; 1991b) .....	81
3.2.3.2 Apprentissage sur corpus (Boulton & Tyne, 2014) .....	82
3.2.3.3 Consultation directe et indirecte (Chambers, 2010) .....	83
<b>3.3 Outils didactiques pour l'E/A de la phraséologie .....</b>	<b>84</b>
3.3.1 Concordancier .....	84
3.3.1.1 Présentation de concordancier.....	84
3.3.1.2 Outil représentatif du TAL pour l'enseignement (Yang, 2017a).....	87
3.3.2 Carte mentale .....	89
3.3.2.1 Présentation de la carte mentale.....	89
3.3.2.2 Outil d'aide à la mémorisation du lexique (Yang, 2017b).....	91
<b>Bilan .....</b>	<b>92</b>

L'enseignement du français spécialisé a vu le jour dans les années 20 avec l'apparition du manuel *Règlement provisoire du 7 juillet pour l'enseignement du français aux militaires indigènes* (Charles-Lavauzelle & Cie, 1927). Sans intention de poursuivre tout le parcours historique et méthodologique de l'enseignement du français spécialisé (entre autres : français militaire, français scientifique et technique, français fonctionnel, français instrumental...), notre attention s'arrêtera sur les trois courants de cet enseignement existant de nos jours : le Français de Spécialité (désormais FS), le FOS et le Français Langue Professionnelle (désormais FLP) (Carras *et al.*, 2007). D'après Mangiante et Parpette (2004), le FS et le FOS s'inscrivent dans deux logiques distinctes : le FS met l'accent sur l'offre des formations en FLE dans une institution (ex. cours de « Français des affaires », cours de « Français juridique ») et le FOS s'applique à répondre aux besoins spécifiques des apprenants. Par ailleurs, le FS propose des cours à long terme (un semestre minimum) par rapport à celui du FOS (quelques jours ou semaines). Toutefois, le FLP offre toujours les formations afin de satisfaire à deux exigences : en français et dans le domaine professionnel. Effectivement, nos apprenants veulent seulement rattraper les besoins langagiers dans les plus brefs délais et cette approche rapide et efficace dans l'enseignement/apprentissage du français dans le cadre d'un domaine professionnel fait partie intégrante du domaine du FOS. Après avoir vu l'intérêt de l'E/A de la phraséologie en FOS et de l'utilisation du corpus en FLE (cf. introduction), nous avons l'intention d'utiliser, dans ce travail, le corpus pour l'E/A de la phraséologie auprès d'apprenants étrangers. Ainsi, dans la partie théorique, nous allons circonscrire tout d'abord la discipline de linguistique de corpus, notamment la construction et l'analyse du corpus. Ensuite, nous allons tout d'abord circonscrire le domaine de notre étude, la phraséologie dont les classifications, les critères de classification sont étudiés. Pour finir, nous allons nous accrocher à des fins didactiques : enseignement de la phraséologie, utilisation du corpus pour l'enseignement et présentation des outils didactiques (concordancier et carte mentale).

## 1. Linguistique de corpus

### 1.1 Émergence de la linguistique de corpus

Avec l'avènement de l'ordinateur dans les années 50, Sinclair (1966), disciple de Firth (1890-1960), s'est déjà rendu compte, dans le projet OSTI (UK Government Office for Scientific and Technical Information), qu'on se positionnait juste au début de l'étude de la linguistique de corpus, car l'outil informatique peut « offrir un moyen d'aller plus loin » (Williams, 2006 : 154). La discipline TAL (Traitement Automatique du Langage) apparaît à la même époque que la naissance de l'ordinateur avec pour objectif de modéliser et d'automatiser une partie du langage (Condamines, 2005). Depuis son apparition, de nouveaux outils et méthodologies se multiplient pour l'analyse des données textuelles sur support numérique et ces outils commencent à être standardisés selon les ressources « génériques », comme les concordanciers, les conjugueurs, les étiqueteurs, les lemmatiseurs, les transcrip-teurs, les aspirateurs, etc. (Antoniadis *et al.*, 2006). Dans les années 90, la linguistique de corpus devient une discipline autonome (Sinclair, 1991a ; Leech, 1992 ; Habert *et al.*, 1997 ; Biber *et al.*, 1998 ; McEnery & Wilson, 2001 ; Tognini-Bonelli, 2001 ; Malrieu, 2004 ; McEnery & Hardie, 2012) avec ses propres publications (entre autres, *le Journal of Corpus Linguistics*<sup>11</sup>), mais certains spécialistes (entre autres, Kennedy, 1998 ; Condamines, 2005 ; Azzopardi, 2010) pensent que la linguistique de corpus provient d'une perspective TAL, car l'apparition de la linguistique de corpus a été attribuée au TAL sur deux niveaux : la collecte automatique de données et le traitement du corpus. Comme l'affirme Kennedy, « Over the last three decades the compilation and analysis of corpora stored in computerized databases has led

---

<sup>11</sup> Adresse du journal of corpus linguistique : <http://www.jbe-platform.com/content/journals/15699811> (consulté le 10/10/2019).

---

to a new scholarly enterprise known as corpus linguistics<sup>12</sup> » (Kennedy, 1998 : 1). D'après Williams, les deux perspectives face au statut de la linguistique de corpus « ne sont pas totalement contradictoires et peuvent être tout à fait complémentaires » (Williams, 2005 : 13). On peut conclure que le TAL apporte une aide considérable à la linguistique de corpus, notamment à sa construction.

## 1.2 Construction et analyse du corpus à l'aide d'outils TAL

Force est de constater que le développement de la linguistique de corpus s'accompagne de l'essor des technologies (Sinclair, 1991a ; Leech, 1992 ; Tognini-Bonelli, 2001). L'élaboration du corpus écrit peut être effectuée automatiquement à partir du Web par un *aspireur* (*crawler*) qui permet de parcourir et d'indexer automatiquement le web afin d'en établir la cartographie. La plate-forme d'un aspireur peut être un site internet comme *WebBootCat*<sup>13</sup> (Baroni & Bernardini, 2004) ou un logiciel tel que *Gromoteur*<sup>14</sup> (Gerdes, 2014). Le corpus oral est souvent effectué sous la forme d'une enquête sociolinguistique comme c'est le cas des deux premiers corpus oraux utilisés dans l'enseignement du français langue étrangère et maternelle : *Français Fondamental* (Gougenheim *et al.*, 1956 ; 1964) et *les Orléanais ont la parole* (Biggs & Dalwood, 1976). Les techniques d'enquêtes utilisées et le recueil des données évoluent avec les nouvelles technologies : du magnétophone, les chercheurs sont passés à l'enregistrement audio discret voire l'enregistrement audio-visuel. De ce fait, les enregistrements sont souvent de bonne qualité, ce qui permet une réutilisation quasi-directe pour la recherche ou pour

---

<sup>12</sup> Au cours des trois dernières décennies, la compilation et l'analyse des corpus stockés dans des bases de données informatisées ont conduit à une nouvelle entreprise savante appelée linguistique de corpus (notre traduction).

<sup>13</sup> Site en ligne : <https://www.sketchengine.co.uk/quick-start-guide/create-your-corpus-lesson-4/> (consulté le 10/10/2019).

<sup>14</sup> Logiciel à télécharger en ligne gratuitement à l'adresse suivante : <http://gromoteur.ilpga.fr> (consulté le 10/10/2019).

l'enseignement. Afin de transcrire le corpus oral, plusieurs logiciels d'aide à la transcription existent, tels que *PRAAT*<sup>15</sup> (Boersma & Weenink, 2004) intégré dans le projet *PFC*<sup>16</sup> (Durand & Tarrier, 2008). Il existe également *Transcriber* développé et utilisé dans les travaux de *GARS*<sup>17</sup> (Barras *et al.*, 1998 ; 2000), *ELAN* (Eudio Language Annotator)<sup>18</sup> fabriqué au Max Planck Institute de Nimègue (Pays-Bas) et adopté dans les études sur les interactions de Mondada (2008) ainsi que *CLAN* (Computerized Language ANalysis) développé pour le projet *CHILDES*<sup>19</sup> (MacWhinney, 2000). Une fois la construction du corpus effectuée, les outils d'interrogation peuvent mener à bien des analyses. Il existe divers outils apportant une aide appréciable dans la description linguistique des langues, notamment la textométrie avec par exemple *Antconc*<sup>20</sup> (Anthony, 2005), *Hyperbase*<sup>21</sup> (Brunet, 2006 ; 2012), *Lexico 3*<sup>22</sup> (Salem *et al.*, 2003), *TXM*<sup>23</sup> (Heiden *et al.*, 2010), *WordSmith Tools*<sup>24</sup> (Scott, 1998), *Trameur*<sup>25</sup> (Fleury, 2007), *DtmVic*<sup>26</sup> (Lebart, 2015), *Voyant Tools*<sup>27</sup> (Sinclair & Rockwell, 2014) et des outils de description des langues, entre autres *NooJ*<sup>28</sup> (Silberstein, 2002). Le choix de logiciel raisonné doit se faire en évaluant les avantages et les inconvénients des outils à

<sup>15</sup> Idem : <http://www.fon.hum.uva.nl/praat/> (consulté le 10/10/2019).

<sup>16</sup> Idem : <https://www.projet-pfc.net> (consulté le 10/10/2019).

<sup>17</sup> Site du groupe aixois de recherche en syntaxe : <http://www.net1901.org/association/GROUPE-AIXOIS-DE-RECHERCHES-EN-SYNTAXE-GARS.778734.html> (consulté le 10/10/2019).

<sup>18</sup> <https://tla.mpi.nl/tools/tla-tools/elan/> (consulté le 10/10/2019).

<sup>19</sup> Site du projet *CHILDES* : <http://childes.talkbank.org> (consulté le 10/10/2019).

<sup>20</sup> Logiciel à télécharger en ligne à l'adresse suivante : <http://www.laurenceanthony.net/software.html> (consulté le 10/10/2019).

<sup>21</sup> Idem : <http://ancilla.unice.fr> (consulté le 10/10/2019).

<sup>22</sup> Manuel d'utilisation : <http://lexi-co.com/ressources/manuel-3.41.pdf> (consulté le 10/10/2019).

<sup>23</sup> Logiciel à télécharger en ligne à l'adresse suivante : <http://textometrie.ens-lyon.fr/spip.php?article60&lang=fr> (consulté le 10/10/2019).

<sup>24</sup> Idem : <http://www.lexically.net/wordsmith/version6/> (consulté le 10/10/2019).

<sup>25</sup> Idem : <http://www.tal.univ-paris3.fr/trameur/> (consulté le 10/10/2019).

<sup>26</sup> Idem : <http://www.dtmvic.com> (consulté le 10/10/2019).

<sup>27</sup> Site en ligne : <https://voyant-tools.org> (consulté le 10/10/2019).

<sup>28</sup> Logiciel à télécharger en ligne à l'adresse suivante : <http://www.nooj-association.org> (consulté le 10/10/2019).

disposition en fonction de plusieurs facteurs internes et externes aux logiciels notamment selon l'objectif de notre recherche. S'agissant d'un corpus, il n'est pas seulement une collection de données langagières car les données d'un corpus doivent être sélectionnées en fonction de critères linguistiques. Un survol rapide de grands corpus nous est nécessaire pour la sélection de données pour l'élaboration d'un corpus.

### 1.3 Élaboration de grands corpus français et anglais

Dans les années 2000, l'élaboration de grands corpus français s'est développée dans les pays francophones (par exemple, *Scientext*<sup>29</sup> en France ; *Lextutor*<sup>30</sup> au Canada ; *INLATO*<sup>31</sup> en Belgique) ou même dans d'autres pays (*Wortschatz*<sup>32</sup> en Allemagne). Il existe une liste exhaustive des corpus dans l'ouvrage *Des documents authentiques au corpus : démarches pour l'apprentissage des langues* (Boulton & Tyne, 2014 : 290-293). *Frantext*<sup>33</sup> (Febvre, 1952), en tant que premier grand corpus français, joue toujours un rôle important dans de grands projets lexicographiques, entre autres le *DMF*<sup>34</sup> (Dictionnaire du Moyen Français), les outils du *CNRTL*<sup>35</sup> (Centre National de Ressources Textuelles et Lexicales) et le dictionnaire *TLF*<sup>36</sup> (Trésor de la Langue Française Informatisée). Notons que l'Angleterre a été pionnière en ce qui concerne l'élaboration des grands corpus numériques. Afin de saisir l'enjeu de la construction du corpus, nous allons étudier le tableau 1-1 ci-dessous, *Frantext* et les trois plus grands corpus britanniques (Pecman, 2004) : *Brown*

---

<sup>29</sup> Site en ligne : <https://corpora.aiakide.net/scientext20/> (consulté le 10/10/2019).

<sup>30</sup> Idem : <https://lextutor.ca/conc/fr/> (consulté le 10/10/2019).

<sup>31</sup> Idem : <https://ilt.kuleuven.be/inlato/> (consulté le 10/10/2019).

<sup>32</sup> Idem : <http://wortschatz.uni-leipzig.de/de> (consulté le 10/10/2019).

<sup>33</sup> Idem : <http://www.frantext.fr> (consulté le 10/10/2019).

<sup>34</sup> Idem : <http://www.atilf.fr/dmf/EditionsElectroniques/> (consulté le 10/10/2019).

<sup>35</sup> Idem : <http://www.cnrtl.fr/portail/> (consulté le 10/10/2019).

<sup>36</sup> Idem : <http://atilf.atilf.fr/dendien/scripts/tlfiv4/showps.exe?p=combi.htm:java=no> (consulté le 10/10/2019).

*Corpus*<sup>37</sup> (Kučera & Francis, 1967) ; *The Bank of English* (Sinclair, 1991b) ; *British National Corpus*<sup>38</sup> (Coleman John, 2004).

Corpus Critères	Corpus britanniques			Corpus français
Corpus	Brown Corpus	The Bank of English	British National Corpus	Frantext
Taille	1 million de mots (1961)	450 millions de mots (2002)	100 millions de mots (1993)	254 millions de mots (2019)
Textes	Presse, textes religieux, science-fiction, etc.	Textes de journaux, de magazines, de rapports, de lettres, etc.	Textes oraux (10%) et textes écrits	Textes littéraires, politiques, techniques, etc.

Tableau 1-1 : Trois corpus britanniques et un français

Comme l'illustre le tableau 1-1, au niveau de la taille, *Frantext* (254 millions de mots) se trouve entre *The Bank of English* (450 millions de mots) et *British National Corpus* (100 millions de mots). On se demande si *Frantext* est assez volumineux comparativement à *The Bank of English*. En fait, la taille de corpus est toujours liée à la question de la représentativité de l'aspect étudié. D'après Boulton et Tyne (2014 : 128), « Plus on cible une variété précise, plus on peut fonctionner avec un corpus réduit. En revanche, pour couvrir une langue sous tous ses aspects, il faut quelque chose de bien plus conséquent ». *Frantext* est un corpus littéraire au départ, mais *The Bank of English* a l'ambition de couvrir toute la langue anglaise. Il est donc normal que *The Bank of English* soit plus conséquent que *Frantext*. Comme le montre le tableau 1-1 ci-dessus, chaque créateur de corpus choisit différents textes du corpus à sa convenance selon ses propres besoins, comme l'affirme Pearson, « Each corpus views the world differently, with its texts classified according to a system devised by

<sup>37</sup> Idem : <http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/> (consulté le 10/10/2019).

<sup>38</sup> Idem : <http://www.natcorp.ox.ac.uk> (consulté le 10/10/2019).

compilers of the corpus or using existing classification systems (...)»<sup>39</sup> » (1998 : 54). Par exemple, *Frantext*, qui est un corpus littéraire, contient des textes littéraires du XIV<sup>e</sup> siècle à nos jours. La particularité de la sélection de textes du corpus *The British National Corpus* est que l'oral en fait partie (10%). D'après Portillo Serrano, « Nous sommes confrontés sans cesse aux genres oraux et écrits que nous maîtrisons, bien ou mal, selon le milieu dans lequel nous participons à la vie sociale » (2010 : 127). De ce fait, Il nous semble que le choix des types de textes constitue les enjeux de la construction du corpus.

## 1.4 Choix des types de textes pour la constitution du corpus

### 1.4.1 Importance du corpus oral

Contrairement à la normativité du corpus écrit, le corpus oral a un caractère spontané difficile à saisir. En s'inscrivant dans les « genres professionnels » (Clot, 2008 : 102), les dialogues entre les cuisiniers au travail dans un restaurant, comme genre de textes ou « entités langagières » (Beacco, 2004), pourraient être sélectionnés pour la constitution du corpus oral. Le corpus oral n'est pas une version orale de la langue écrite et des différences entre le corpus oral et le corpus écrit sont déjà prises en compte par Bloomfield (1927) cité et traduit par Blanche-Benveniste : « Nous n'avons pas appris à savoir ce que nous faisons quand nous parlons : structure des sons, grammaire réellement utilisée » (1999 : 14). Boulton et Tyne (2014) constatent également que des spécificités linguistiques et pragmatiques d'actes communicatifs peuvent être décrites grâce aux corpus oraux. En France, les corpus oraux ont été initiés par le *GARS* (Groupe Aixois de Recherches en Syntaxe), pour des études en phonétique et en grammaire de l'oral, sous la direction de Blanche-Benveniste dans les années 70. D'ailleurs, le premier grand corpus

---

<sup>39</sup> Chaque corpus dans le monde est différent : les textes sont classés selon un système mis au point par les auteurs du corpus ou selon des systèmes de classification déjà existantes.



oral *CorpAix*<sup>40</sup> a été réalisé par cette équipe dont l'idée de base était que les règles phonétiques et syntaxiques à l'oral devaient être prises en considération autant que celles de l'écrit dans l'enseignement (Debaisieux, 2000 ; 2005 ; 2008). De nos jours, les corpus oraux en français se multiplient (Chambers, 2009). À partir des inventaires réalisés par certains linguistes (Debaisieux, 2005 ; Cappeau & Seijido, 2005 ; voir aussi Cappeau & Gadet, 2007b), nous présentons une liste de corpus oraux en annexe 1 (p. 231), qui est loin d'être exhaustive, mais qui donne un bon aperçu des corpus oraux accessibles/inaccessibles en ligne avec leurs perspectives. Nombre de corpus oraux sont inaccessibles en ligne car ils restent la propriété des personnes liées aux projets qui ont permis leur élaboration (Debaisieux, 2005). Mais heureusement, quelques projets (entre autres, *CFPP2000*<sup>41</sup>, *CO2*<sup>42</sup>, *ANCOR*<sup>43</sup>, *PFC*<sup>44</sup>, *CNRTL*<sup>45</sup>, *OFROM*<sup>46</sup>, *CLAPI*<sup>47</sup> et *FLORALE*<sup>48</sup>) sont accessibles et leurs corpus interrogeables en ligne. Notons que des difficultés persistent pour la constitution des corpus oraux. Mangiante (2016), insiste sur certaines d'entre elles et les classe en trois niveaux :

1. l'obtention du consentement de la part des entreprises (ou institutions ou personnes) pour le recueil de données ;
2. la difficulté connue de la représentativité et de la taille du corpus ;
3. la coûteuse transcription en temps et en technique informatique (Leech, 1991).

---

<sup>40</sup> En 1911, il existe déjà un corpus oral français *Archives de la parole* de la BNF, Mais *CorpAix* est le premier corpus oral ayant plus d'un million de mots.

<sup>41</sup> Site du projet *CFPP2000* : <http://cfpp2000.univ-paris3.fr/> (consulté le 10/10/2019).

<sup>42</sup> Idem : [http://www.info.univ-tours.fr/~antoine/parole\\_publicue/CO2/index.html](http://www.info.univ-tours.fr/~antoine/parole_publicue/CO2/index.html) (consulté le 10/10/2019).

<sup>43</sup> Idem : [http://www.info.univ-tours.fr/~antoine/parole\\_publicue/ANCOR\\_Centre/index.html](http://www.info.univ-tours.fr/~antoine/parole_publicue/ANCOR_Centre/index.html) (consulté le 10/10/2019).

<sup>44</sup> Idem : <https://www.projet-pfc.net> (consulté le 10/10/2019).

<sup>45</sup> Idem : <http://www.cnrtl.fr/corpus/teof/> (consulté le 10/10/2019).

<sup>46</sup> Idem : <http://www11.unine.ch> (consulté le 10/10/2019).

<sup>47</sup> Idem : <http://clapi.ish-lyon.cnrs.fr> (consulté le 10/10/2019).

<sup>48</sup> Idem : <https://florale.unil.ch/> (consulté le 10/10/2019).

En tentant de surmonter ces trois difficultés, l'utilisation du corpus oral, en comparaison avec le corpus écrit, révèle des différences oral/écrit intéressantes dans ce cadre professionnel spécifique. Par exemple, dans le corpus que nous avons élaboré (Yang, 2015), le lexème<sup>49</sup> THERMOSTAT<sup>50</sup> apparaît très fréquemment à l'écrit (124 occurrences), mais il n'est que rarement identifié à l'oral. En ce qui concerne le choix des types de textes pour la constitution du corpus, le genre de textes est également un sujet qui mérite d'être discuté.

#### 1.4.2 Genres de textes

Malrieu & Rastier (2001 : 548) classent les discours en quatre niveaux comme indiqué dans la figure 1-1 ci-dessous : discours (ex. littérature, juridique, politique), champs génériques (ex. théâtre, poésie, récits) ; genres (ex. comédie, tragédie, drame) ; sous-genres (ex. roman policier, roman de formation, roman par lettres).

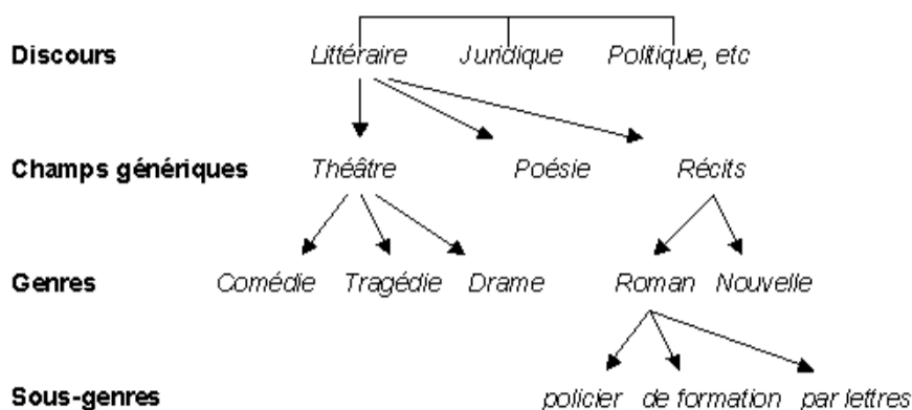


Figure 1-1 : Niveaux de classement de Malrieu & Rastier (2001)

Le discours et le genre sont mis en relief dans l'analyse de la fonction du langage (Biber, 1988). À l'instar de Rastier, « tout texte se rattache à la langue par un discours et à un discours par la médiation d'un genre » (2001 : 230), à savoir que le discours articule tout texte à la langue et qu'un genre est la

<sup>49</sup> D'après Polguère (2002), un lexème regroupe « des mots-formes ne se distinguant que par la flexion » (2002 : 37, par exemple, THERMOSTAT est un lexème qui peut contenir deux mots-formes : thermostat et thermostats).

<sup>50</sup> En lexicologie, les lexèmes sont toujours mentionnés en petite majuscule.

charpente du discours. Malrieu (2004) confirme l'articulation du genre entre la langue et le champ de pratique et qu' « il est impossible de construire un corpus représentatif d'un discours lié à un domaine de pratique. » (ibid. : 74) car chaque domaine comporte des genres hétérogènes et on ne peut considérer certains genres pour représenter un domaine. Par ailleurs, Rastier (1998 : 101) ajoute que les structures textuelles peuvent être déterminées par les genres de textes. De ce fait, « tout corpus doit être constitué de discours représentatifs du genre » (Sinclair, 1996 : 4).

Peu étudié, le genre n'est pas défini clairement. Afin d'identifier le genre dans des pratiques sociales diversifiées, Rastier (2001) propose que les discours correspondent à des types de pratiques (domaine) et que les genres conviennent à des types de situations de communication. Dès lors, dans le domaine culinaire, les discours contiennent tous les textes relatifs à ce domaine et un genre de textes du domaine peut correspondre à une recette de cuisine ou un menu, comme indiqué dans la figure 1-2.

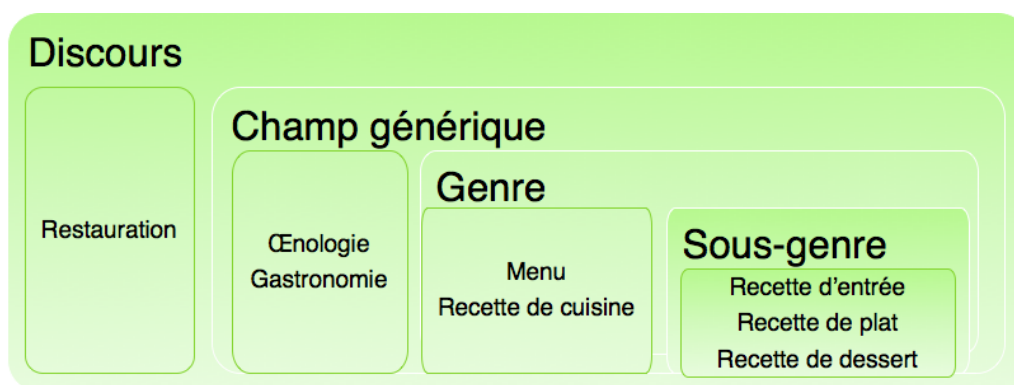


Figure 1-2 : Discours et genre dans le domaine culinaire

En conséquence, la recette de cuisine peut être choisie, dans l'élaboration de notre corpus, comme genre de texte. En outre, elle peut avoir des sous-genres : recette d'entrée, recette de plat et recette de dessert. D'après Malrieu & Rastier (2001), la connaissance d'un genre peut souvent aider à chercher l'information, par exemple, les ingrédients utilisés pour un plat se trouvent souvent au début d'une recette de cuisine. Pour la question du choix

du type de textes, nous ajoutons que les documents authentiques comportent plusieurs avantages pour l'enseignement. Si le corpus est construit à des fins didactiques, l'utilisation des documents authentiques n'est pas à négliger.

#### 1.4.3 Documents authentiques pour l'enseignement

Tout en tenant compte du fait que, dans la méthodologie SGAV (Structuro-Globale Audio-Visuelle), la langue cible doit être enseignée dans sa fonction de communication réelle et que la civilisation française doit être abordée lors de l'apprentissage de la langue, les documents authentiques sont utilisés pour « mettre en contact l'étudiant avec la langue « réelle » » et « concilier l'apprentissage de la langue à celui de la civilisation » (Cuq & Gruca, 2005 : 429). Quant aux avantages apportés par les documents authentiques pour l'enseignement, Abe *et al.* (1979) expliquent que son utilisation permet de motiver les apprenants et d'offrir des possibilités d'autonomiser leurs apprentissages. Prenant en considération le temps limité consacré à l'apprentissage dans une formation de FOS (Mangiante & Parpette, 2004), l'enseignant s'efforce d'autonomiser les apprenants. L'autonomisation des apprentissages avec des documents authentiques répond donc particulièrement aux besoins des apprenants du FOS. D'après Abe *et al.*, les documents authentiques sont définis comme des documents produits « à des fins autres que l'apprentissage d'une langue seconde » (1979 : 2). Sachant que tout document authentique ne peut être utilisé pour l'enseignement, lorsqu'il est question du choix des documents authentiques pour l'enseignement, Abe *et al.* proposent de tenir compte « des besoins, des objectifs et des intérêts des apprenants » (1979 : 10). En fait, l'analyse des besoins des apprenants constitue une étape inévitable dans l'élaboration d'un cours de FOS (Mangiante & Parpette, 2004). En ce qui concerne les documents authentiques pour l'enseignement du français, ils sont constitués des « documents 'bruts', élaborés par des francophones pour des francophones à des fins de communication. Ce sont donc des énoncés

produits dans des situations réelles de communication et non en vue de l'apprentissage d'une seconde langue » (Cuq & Gruca, 2005 : 431-432). En outre, Cuq et Gruca (ibid.) classent les documents authentiques en quatre types : les documents authentiques écrits ; les documents authentiques oraux ; les documents authentiques visuels et télévisuels ; ainsi que les documents authentiques électroniques. Les deux premiers types sont constitués des textes « fonctionnels de la vie quotidienne » (ibid., 433), ex. recette de cuisine et les deuxièmes offrent un contenu linguistique « de la langue parlée » (ibid., 433), par exemple, des discours politiques. Outre les textes fonctionnels et la langue parlée, l'utilisation de l'image est prise en compte dans le troisième type, par exemple, l'image animée et mobile dans une vidéo. L'apport du web n'est pas négligeable dans le quatrième type, par exemple, les journaux sur internet. La question du « web comme corpus » (Kilgarriff & Grefenstette, 2003 ; Baroni & Bernardini, 2006 ; Drago, 2009 ; Hundt *et al.*, 2007) reste encore discutable autour de l'authenticité des données, car les dimensions du web sont inconnues et mises à jour en permanence (Sinclair, 2005), mais d'après Kilgarriff *et al.* (2007) et Shei (2008), le discours sur le web correspond à la langue réelle. En outre, Sinclair pense également que des textes sur le web constituent un bon corpus pour l'enseignement du français à l'école ou l'auto-apprentissage à la maison : « le Web lui-même... constitue une immense source de langue disponible dans la salle de classe ou pour un travail à la maison » (2004 : 297).

Afin d'assurer la compréhension écrite et orale des apprenants dans le cadre de l'enseignement du FOS, les deux premiers types de documents authentiques sont incontournables. Nous avons également besoin du troisième type de documents authentiques, à savoir les documents authentiques visuels et télévisuels, car la compréhension est facilement acquise avec la présence de l'environnement de communication (Cuq & Gruca, 2005). Le multimédia des quatrièmes documents authentiques semble également nécessaire dans un projet de FOS. En effet, l'apport du multimédia

est mis en exergue par Mangiante & Parpette (2004) et Carras *et al.* (2007) dans la collecte des données pour l'élaboration d'un cours de FOS.

Quant à la construction et l'analyse de corpus pour l'enseignement du FOS, l'accent est souvent mis sur les « humanités numériques » (Boulton & Tyne, 2014 ; Mangiante, 2016). D'après Mangiante (2016), dans les « humanités », l'enseignant est en mesure de proposer de nouvelles manières d'exploiter le corpus pour l'enseignement et l'adjectif « numériques » rassemble tous les aspects techniques lors de la recherche et de l'enseignement du FOS. Pour le côté « numériques », les outils TAL ne peuvent être ignorés. La discipline TAL fournit un apport considérable dans la linguistique de corpus, notamment dans la collecte automatique de données et le traitement du corpus (Habert *et al.*, 1997 ; Biber *et al.*, 1998 ; Kennedy, 1998 ; McEnery & Wilson, 2001 ; Malrieu, 2004 ; McEnery & Hardie, 2012).

## Bilan

Le statut de la linguistique de corpus n'est pas stable et sa relation avec le TAL est toujours discutable, mais il s'accroît énormément sous l'impulsion du TAL à deux niveaux : la collecte automatique des données et le traitement de ces données au sein du corpus. Pour la collecte automatique des données écrites, les logiciels d'aspirateur (entre autres, *WebBootCat*, *Gromoteur*) nous servent pour collecter automatiquement des données sur internet. Les logiciels d'aide à la transcription (entre autres, *Praat*, *ELAN*, *CLAN*) nous permettent de synchroniser le signal et la transcription. Afin de procéder à l'analyse pertinente, les logiciels d'interrogation (entre autres, *Antconc*, *Hyperbase*, *Lexico 3*, *TXM*, *WordSmith Tools*, *Trameur*, *DtmVic*, *Voyant Tools*, *NooJ*) restent toujours à notre disposition. Le choix des logiciels appropriés du TAL pour la construction et pour l'analyse du corpus se fait toujours en fonction de l'objectif de l'étude visé. Après avoir fait un survol rapide de l'élaboration de grands corpus français et anglais, nous nous sommes rendu

compte que le choix du type de textes jouait un rôle crucial dans la construction du corpus : les genres de textes peuvent déterminer les structures textuelles ; les règles de l'oral sont différentes de celles de l'écrit. L'utilisation des documents authentiques motive les apprenants en les mettant en contact direct avec la langue réelle. Par ailleurs, elle permet également aux apprenants d'autonomiser leurs apprentissages dans leur vie professionnelle ultérieure. Il en résulte que les genres de textes, le corpus oral, les documents authentiques et l'apport du multimédia doivent être pris en compte pour l'élaboration d'un corpus.

## 2. Phraséologie

### 2.1 Champs de recherche : la phraséologie

Découvert par Paul Hermann (1880), Henry Sweet (1891) et Michel Bréal (1897) à la fin du XIX<sup>ème</sup> siècle, la phraséologie a été, pour la première fois, traitée en tant que discipline par Charles Bally, linguiste suisse, au début du XX<sup>ème</sup> siècle sous l'étiquette « séries phraséologiques » (1909 : 88). Grâce à l'apport au domaine, ce dernier est considéré comme le père de la phraséologie (González-Rey, 2002 ; Čermak, 1998). La prise en compte de son œuvre se fait d'abord chez les linguistes soviétiques dans les années 1940-1960. Grâce à l'allemand et à l'espagnol (notamment des linguistes cubains), la phraséologie a pu revenir en Europe occidentale (González-Rey, 2002). Les linguistes anglo-saxons ont une avance sur la recherche en phraséologie par rapport aux linguistes français (Pecman, 2004). Dans cette évolution de la phraséologie en Europe, certains travaux linguistiques occupent une place importante : Vinogradov (1946 ; 1947) pour le russe ; Häusermann (1977) et Burger (1973 ; 1998) pour l'allemand ; Carneado & Tristá (1985) pour l'espagnol ; Lipka (1974) et Arnold (1973) pour l'anglais ; Thun (1975) pour le français. Une petite chronologie sur la recherche en phraséologie est présentée dans la figure 2-1 ci-dessous.





Figure 2-1 : Évolution de la phraséologie

Dans cette évolution, on peut découvrir le fait que la phraséologie est prise en compte tardivement chez les linguistes français. L'étude sur la phraséologie suscite un intérêt considérable dans le cadre de la recherche appliquée :

- développement de programmes d'extraction (Blanc *et al.*, 2008, 2010) ;
- élaboration de manuels scolaires et de dictionnaires (Granger & Paquot, 2008 ; 2015 ; Tutin, 2010 ; Verlinde, Binon & Selva, 2006) ;
- acquisition (Ellis *et al.*, 2008 ; Forsberg, 2008 ; Conklin & Schmitt, 2008 ; Wood, 2010 ; Edmonds, 2013 ; 2016 )
- et de la recherche fondamentale (Alexander, 1984 ; Gläser, 1988 ; Howarth, 1993 ; Cowie, 1998 ; Moon, 1998 ; Benveniste, 1974 ; Guiraud, 1961 ; Martinet, 1967 ; 1974 ; 1999).

Legallois & Gréa parlent d'un « tournant phraséologique de la linguistique » (2006 : 5), car d'un côté, la phraséologie fait de plus en plus l'objet d'analyses dans des publications et des colloques et de l'autre, elle provoque « un renouvellement de l'analyse syntaxique, sémantique, textuelle et psycholinguistique sur la base d'observations plus empiriques » (ibid.). Malgré l'intérêt considérable soulevé par les recherches du domaine phraséologique,

la phraséologie est un champ flou du langage : « Phraseology is a fuzzy part of language<sup>51</sup> » (Altenberg, 1998 : 100) et ce flou phraséologique est observé sur plusieurs plans. Hormis le plan terminologique mentionné dans l'introduction, ce flou se trouve également au niveau du statut et au niveau définitoire. Pour le flou du statut de cette science, la position de la phraséologie est toujours discutable (Pecman, 2004 ; González-Rey, 2002) : certains spécialistes revendiquent son statut de discipline (Makkai, 1975 ; Čermak, 1988), d'autres pensent que la phraséologie est seulement une sous-discipline de la lexicologie ou de la linguistique (Picoche, 1992 ; Gläser, 1998 ; Pawley, 2001) et pour ces derniers, une unité phraséologique est plutôt considérée comme une unité lexicologique (Heinz, 1997 ; 2003 ; Lamiroy, 2003). Le flou définitoire peut s'incarner dans de multiples définitions proposées comme mentionnées ci-dessous :

*Une séquence, continue ou discontinue, de mots ou de tous autres éléments doués de sens, qui est, ou semble être, préfabriquée : c'est-à-dire, stockée dans la mémoire et extraite de celle-ci dans son ensemble au moment de l'utilisation, plutôt que d'être sujette à la génération ou à l'analyse grammaticale (Wray & Perkins, 2000 : 1).*

*Je considère par combinaisons de mots récurrentes toute chaîne de mots continue apparaissant plus d'une fois sous la même forme (Altenberg, 1998 : 101).*

*Il y a série ou groupement usuel lorsque les éléments du groupe conservent leur autonomie, tout en laissant voir une affinité évidente qui les rapproche, de sorte que l'ensemble présente des contours arrêtés et donne l'impression du < déjà-vu >. (Bally, 1963 : 70).*

*[...] les collocations devraient être définies non seulement comme « des combinaisons récurrentes de mots » mais encore comme des « combinaisons récurrentes et arbitraires de mots » (Benson, 1989 : 3).*

---

<sup>51</sup> La phraséologie est une partie confuse de la langue (notre traduction).

*Un autre type de signification de mot non saillant est exemplifié par voie, qui peut être utilisé dans un sens concret («chemin», «route») ou dans un sens métaphorique étendu («mode», «signifie»). L'étude du corpus montre [...] que ce dernier sens métaphorique est beaucoup plus fréquemment utilisé dans l'anglais actuel, et je l'appelle donc le sens saillant (Hudson, 1998 : 66).*

*Par unité phraséologique nous entendons tout groupement de mots récurrent (expression, locution ou construction au sens large) auquel les locuteurs [...] associent un sens précis, qui n'est pas nécessairement lié au sens de ses composantes prises individuellement (Zinglé, 2003 : iii).*

*Placée à mi-chemin entre le lexique et la syntaxe, la phraséologie s'occupe des signes polylexicaux composés d'au moins deux mots, stables, répétés et souvent figurés. (González-Rey, 2007 : 5).*

*[...] une locution prototypique est caractérisée [...] par sa non-compositionnalité. On a beau comprendre tous les mots qui entrent dans tirer le diable par la queue, cela ne suffit pas pour comprendre ce que cette locution veut dire. (Martin, 1997 : 293).*

La difficulté de la description de la définition tient à deux raisons principales : la première attribue ce flou définitoire à l'obscurcissement des dénominations floues, comme le dit G. Gross, « Le fait linguistique du figement a été obscurci par des dénominations floues et très hétérogènes, de sorte qu'on est en présence de strates définitionnelles très souvent incompatibles » (1996 : 3). La deuxième est due à l'impossibilité de la description de la locution d'un seul point de vue. Comme le constate Heinz, « la complexité du phénomène de la locution et l'impossibilité de la décrire d'un seul point de vue » (1993 : 5). Le flou définitoire de la phraséologie nous fait penser à celui du bilinguisme, car il n'y a pas non plus de définition claire et univoque pour le bilinguisme. Afin de mieux décrire la bilingualité, les linguistes font les typologies du bilinguisme pour avoir l'inventaire de ce phénomène. Nous décidons donc de procéder à la classification des phrasèmes pour mieux cerner la phraséologie.

## 2.2 Classifications des phrasèmes

La question de la typologie de la phraséologie a été traitée par plusieurs chercheurs (entre autres, Gläser, 1988 ; G. Gross, 1996 ; Mel'čuk, 1993, 1998 ; Howarth, 1996, 1998 ; Moon, 1998 ; Cowie, 1978 ; Jernej, 1992 ; González-Rey, 2002 ; Polguère, 2016). Nous allons voir, dans le tableau 2-1 ci-dessous, quelques classifications auxquelles nous pourrions faire référence pour notre étude.

Phraséologues	Critères	
G. Gross (1996)	Figement	Expression semi-figée et expression figée.
Polguère (2016)	Figement	Expression phraséologique et expression semi-phraséologique.
Mel'čuk (1998)	Critère sémantique, critère pragmatique	Phrasèmes complets (pas question, prendre une veste) ; Semi-phrasèmes (café noir, commettre un crime) ; Quasi-phrasèmes (bande dessinée, centre commercial) ; Pragmatèmes (porte ouverte, bonne nuit, à consommer avant...).

Tableau 2-1 : Quelques classifications des phrasèmes

G. Gross (1996) et Polguère (2016) classent, en fonction du figement, les unités phraséologiques en deux grandes catégories : les expressions figées (phraséologiques) et les expressions semi-figées (semi-phraséologiques). Les phrasèmes chez Mel'čuk (1998) sont classés selon quatre catégories en retenant les critères pragmatique et sémantique : *phrasèmes complets* (locution), *semi-phrasèmes* (collocation), *quasi-phrasèmes* et *pragmatèmes*. La figure 2-2 illustre que les trois premières catégories se différencient par le critère sémantique.

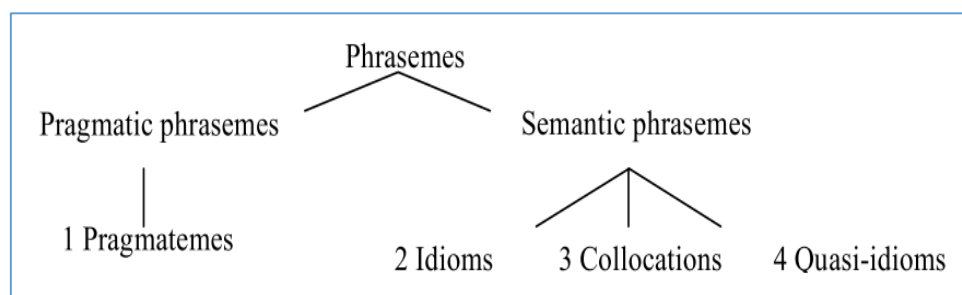


Figure 2-2 : Classification des phrasèmes chez Mel'čuk (1998 : 30)

---

Le sens d'un *phrasème complet* est non-compositionnel, par exemple, *se prendre une veste* (échouer dans une affaire). Un *semi-phrasème* est une collocation dans la vision sémantique : un élément (base) garde son sens habituel ; l'autre (collocatif) prend un sens déviant qui est très souvent imprévisible (Cavalla & Labre, 2009 ; Hill, 1993 ; Nesselhauf, 2005 ; Verlinde, Binon & Selva, 2006). Le sens d'un *quasi-phrasème* (quasi-idioms) prend le sens de l'un de ses composants comme de l'autre (par exemple, *centre commercial*), mais « aucun sens n'est communicativement dominant » (Mel'čuk, 1998 : 25). Les *pragmatèmes* regroupent les expressions associées à un contexte situationnel. Comme nous le voyons, la classification de la phraséologie se fait en fonction de différents critères choisis et par la suite, nous allons étudier les critères servant à la classification de la phraséologie.

## 2.3 Critères de la phraséologie

Cavalla (2009) décrit trois critères principaux dans les définitions de la phraséologie pour son étude dans un cadre d'enseignement des langues : combinatoire, association lexicale et relation sémantique ; mais elle admet en même temps que les éléments mentionnés dans les définitions de la phraséologie sont variés et complexes. En fait, depuis des années, les phraséologues cherchent à définir la phraséologie par les deux grandes approches suivantes (Granger & Paquot, 2008 ; Bolly, 2011 ; Legallois & Tutin, 2013 ; Edmond, 2013) : l'approche distributionnelle (statistique) et l'approche fonctionnelle (ou traditionnelle). La première est basée sur les deux principes suivants : le principe d'idiomaticité (idiom principle) et le principe de libre choix (open choice principle) (Erman & Warren, 2000). Le premier principe met l'accent sur le contexte pour le choix d'un mot, d'après Sinclair, « the choice of one word affects the choice of others in its vicinity<sup>52</sup> » (1991a :

---

<sup>52</sup> Le choix d'un mot affecte le choix des autres dans son voisinage (notre traduction).

173), mais le deuxième principe insiste sur le libre choix pour la combinaison de mots. La fréquence est le critère principal dans la phraséologie statistique (Sinclair, 1991a ; 2000 ; Römer, 2009). Alors que l'approche fonctionnelle fait référence à d'autres critères, tels que les critères syntaxiques, les critères sémantiques et les critères pragmatiques (Edmonds, 2013). La sélection des critères s'effectue selon l'objectif didactique visé : la longueur des unités phraséologiques est retenue dans la typologie de Gläser (1988) ; le figement dans les typologies de G. Gross (1996), de Howarth (1996) et de Polguère (2016) ; le critère pragmatique dans la typologie de Mel'čuk (1993) et de Wray (2002) ; l'étymologie dans la typologie de Jernej (1992). Nous allons détailler les trois critères suivants qui nous semble prioritaires pour notre projet en FOS auprès des apprenants étrangers : fréquence, figement et critère pragmatique.

### 2.3.1 Fréquence

La fréquence de phrasèmes est un révélateur pour identifier les unités phraséologiques, car les phrasèmes ont souvent une fréquence significative. Williams le confirme dans ce sens : « les statistiques ne définissent pas les collocations mais sont un moyen d'évaluer leur seuil de signification. Ainsi, elles peuvent aider à leur extraction » (1999 : 69). Ce critère d'identification a été adopté dans des travaux de phraséologues, tels que Howarth (1996), Moon (1998), Granger (1998), Altenberg (1998), Lewis (2000) et Gledhill (2000). Sachant qu'une faible fréquence peut cacher une haute pertinence pour le domaine. Par exemple, dans sa quête des collocations à partir d'un corpus de textes du domaine des sciences et technologies de l'information, Curado Fuentes (2001 : 112) relève les unités suivantes comme ayant une fréquence très faible : *first and foremost*, *diskette drive*, *ticket booth*, etc. ; Gledhill (2000), à son tour, rencontre également cette faiblesse des études fondées exclusivement sur des analyses statistiques du corpus. Afin de surmonter le problème de négligence, les lexicologues ont l'habitude de réhabiliter les expressions de faible fréquence dans leurs listes d'étude en

---

choisissant d'autres critères nécessaires. Pour notre étude, nous choisissons également le figement et le critère pragmatique comme critères de classification.

### 2.3.2 Figement

Le figement est l'une des propriétés les plus souvent évoquée pour décrire les phrasèmes par rapport aux autres éléments lexicaux de la langue (Pecman, 2004 ; Mejri, 2005). Il est constitué d'« un ensemble de caractéristiques syntaxiques et sémantiques affectant une unité polylexicale » (Neveu, 2004 : 6), par exemple *idée reçue* est un groupe figé, mais *idée stupide* ne l'est pas. Cavalla (2009) constate que le figement pose des problèmes aux apprenants étrangers. D'après Lamiroy & Klein, « le figement est un phénomène extrêmement complexe, polyfactoriel, ce qui explique pourquoi il échappe à toute tentative d'en fournir une définition simple et univoque » (2005 : 135). En effet, il n'existe pas de dichotomie entre « ne pas être figé » et « être figé ». Comme le confirme Neveu, « le figement peut n'être que partiel, ce qui suppose des degrés dans le processus » (2004 : 6). Outre le figement par lui-même, c'est également son degré qui pose problème aux apprenants étrangers : en effet, ces derniers ne sont pas capables de juger si le phrasème *crème fraîche* est plus ou moins figé que le phrasème *plat préféré*.

### 2.3.3 Critère pragmatique

Le troisième critère qui nous semble intéressant en milieu professionnel est le critère pragmatique, parce que certaines expressions libres peuvent apporter une valeur pragmatique. Bossé-Andrieu et Mareschal (1998) prennent le mot « porte » comme exemple : en réalité, on ne peut que « ouvrir » ou « fermer » une porte. En fait, le critère pragmatique est abondamment pris en compte dans la classification de la phraséologie en français (*pragmatèmes* de Mel'čuk, 1998 ; *énoncés liés* de Fónagy, 1997 ; *formules routières* de González-Rey, 2007) et dans celle en anglais (*routines linguistiques* de Coulmas, 1979 ; *unités discursives sociales* de Cowie, 1988 ; *unités situationnelles* de Moon,

1992a, 1992b ; *lexical phrases* de Nattinger & DeCarrico, 1992 ; *unités socio-interactionnelle* d'Aijmer, 1996 ; *unités interactionnelles* de Wray, 2002 ; 2007 et *expressions conventionnelles* de Bardovi-Harlig, 2009). Pour identifier les expressions pragmatiques, mis à part les expériences accumulées en milieu culinaire, le critère pragmatique peut se manifester par la haute fréquence d'utilisation (Bossé-Andrieu & Mareschal, 1998 ; Grossmann, 2011).

Ces trois critères nous semblent pertinents pour l'identification ou la classification des phrasèmes du domaine professionnel pour les apprenants étrangers. En amont de la classification des phrasèmes retenus, l'extraction des phrasèmes doit être mise au premier plan. En fait, la modélisation est souvent présentée comme une méthode efficace pour accéder facilement au texte (Aussenac-Gilles & Condamines, 2001). Tutin souligne également que « le recours à un modèle linguistique plus fin (...) permet à la fois de systématiser les associations syntaxique et sémantique et de proposer un traitement informatique moins *ad hoc*, par classe de phénomènes » (2004 : 210).

## 2.4 Modélisations des données dans des projets phraséologiques

Deux projets sur la phraséologie du français au niveau international nous intéressent : le projet DEC (dictionnaire explicatif et combinatoire du français contemporain) et le projet DELA (dictionnaires électroniques du LADL<sup>53</sup>). Quelles notions remarquables sont utilisées pour la modélisation des données lexicales dans les deux projets ?

Dans le premier projet, Mel'čuk (1981 ; 1984 ; 1995 ; 1999) propose, la fonction lexicale (désormais FL) pour expliciter les relations syntactico-sémantiques. En effet, la FL est un outil conceptuel des

---

<sup>53</sup> LADL : Laboratoire Automatique Documentaire et Linguistique.



descriptions des langues en modélisant et encodant aussi bien les liens paradigmatiques, appelés dérivations sémantiques, que les liens syntagmatiques (collocationnels) (Polguère, 2000). La FL est une notion issue de la théorie Sens-Texte (Mel'čuk & Žolkovskij, 1969 ; 1970). Afin de mieux présenter la FL de manière plus concrète, nous pouvons l'introduire par la formule mathématique suivante :

$$f(x)=y,$$

où x représente l'argument (mot-clé) et y sa valeur. Par exemple, dans la formule  $f(\text{cuire})=\text{cuisson}$ , *cuire* (argument) est la nominalisation (f) de *cuisson* (valeur), c'est-à-dire :

$$\text{Nominalisation}(\text{cuire})=\text{cuisson}.$$

Au total, Mel'čuk *et al.* proposent 80 FL. On peut en citer quelques-unes qui conviennent à notre contexte de cuisine dans le tableau 2-2 ci-dessous :

FL	Liens syntaxiques	Liens sémantiques
$Oper_1$	verbes supports	verbes sémantique vides
$Fact_0$	verbes de réalisation	« réaliser »
<i>Prepar</i>	verbe	« préparer pour »
$A_2$	modificateur adjectival	dérivé sémantique adjectival actanciel

Tableau 2-2 : Quelques fonctions lexicales

La FL [ $Oper_1$ ] est un verbe sémantique vide, elle sert à « verbaliser les noms prédictifs en exprimant le mode et le temps » (Mel'čuk *et al.*, 1995 : 138). Dans cette FL, l'actant est le sujet grammatical et le mot-clé devient le complément d'objet direct.

$$Oper_1(\text{erreur})=\text{faire}.$$

La FL [*Prepar*] est un verbe ayant le sens de « préparer pour » qui a pour mot-clé L comme complément d'objet direct et elle se combine souvent avec le verbe de réalisation [ $Fact_0$ ], par exemple,

$$\text{PreparFact}_0(\text{viande})=\text{hacher}.$$

La FL [A<sub>2</sub>] est un modificateur adjectival et elle peut se combiner avec d'autres FLs, par exemple,

A<sub>2</sub>PreparFact<sub>0</sub>(viande)=haché.

Dans le deuxième projet (M. Gross, 1988 ; G. Gross, 1981 ; 1986a ; 1986b ; 1988 ; 1996 ; Danlos, 1988), une notion a été forgée pour gouverner les rapports des éléments d'un phrasème : *grammaire locale*. Cette notion contient des phrases se référant à un thème (ou à un domaine spécifique). Par exemple, pour le thème « verbes d'état », nous pouvons rassembler une liste de verbes, tels que *être*, *sembler*, *devenir*. Des notions de modélisation nous aident à modéliser les phrasèmes ciblés. Une fois les données lexicales mobilisées, un outil TAL approprié peut réaliser l'extraction automatique de la phraséologie. Pour ce faire, nous avons choisi *Nooj*.

## 2.5 Logiciel de la description des langues : approche de *NooJ*

En tant que logiciel de formalisation des langues, *NooJ* peut décrire, en théorie, toutes les langues naturelles du monde (Silberztein, 2015 ; 2016) et peut également offrir des possibilités de traitement du corpus pour l'enseignement (Silberztein & Tutin, 2005). En se basant sur la hiérarchie de Chomsky-Schützenberger (Chomsky & Schützenberger, 1963), *NooJ* décrit les langues naturelles par le biais de quatre grammaires : grammaire non restreinte, grammaire contextuelle, grammaire algébrique, grammaire rationnelle (dans la figure 2-3 ci-dessous).



Figure 2-3 : Quatre types de grammaires génératives

Les quatre types de grammaires génératives ont une relation hiérarchique : l'ensemble des grammaires rationnelles est compris dans l'ensemble des grammaires algébriques, lui-même inclus dans l'ensemble des grammaires contextuelles, lui-même englobé dans l'ensemble des grammaires non restreintes. Nous verrons par la suite l'utilisation de cet outil par types de grammaires : nous présenterons tout d'abord chaque type de grammaire, pour lequel nous verrons ce que peut faire *NooJ* ou des exemples de l'application. Pouvant décrire toutes les langues naturelles, la grammaire non restreinte est plus puissante que les trois premières. Notre présentation a donc mis l'accent sur les trois premières grammaires : grammaire rationnelle, grammaire algébrique et grammaire contextuelle.

### 2.5.1 Grammaire rationnelle

Ce type de grammaire est composé des deux membres : le membre gauche est constitué d'un symbole auxiliaire et le membre droit peut être soit un symbole terminal (exemple,  $N \rightarrow \text{table}$ ), soit un seul symbole terminal suivi d'un seul symbole auxiliaire (exemple,  $N \rightarrow \text{chat } \textit{ADJECTIF}$ ), soit le mot vide (exemple  $GN \rightarrow \langle E \rangle$ ). L'équivalence entre la grammaire rationnelle et le graphe à états finis a été confirmée dans le théorème de Kleene (1956), donc, les langages rationnels peuvent être également décrits par des graphes à états finis<sup>54</sup>. En effet, plusieurs logiciels peuvent construire une machine à états finis en TAL, mais malheureusement, les grammaires proposées ne s'affichent pas toujours « graphiquement ». *NooJ* propose de vrais graphes pour être lisible structurellement comme indiqué dans la figure 2-4 ci-dessous.

---

<sup>54</sup> La machine à états finis est la machine qui « peut alors analyser des textes automatiquement » (Silberstein, 2015 : 42).

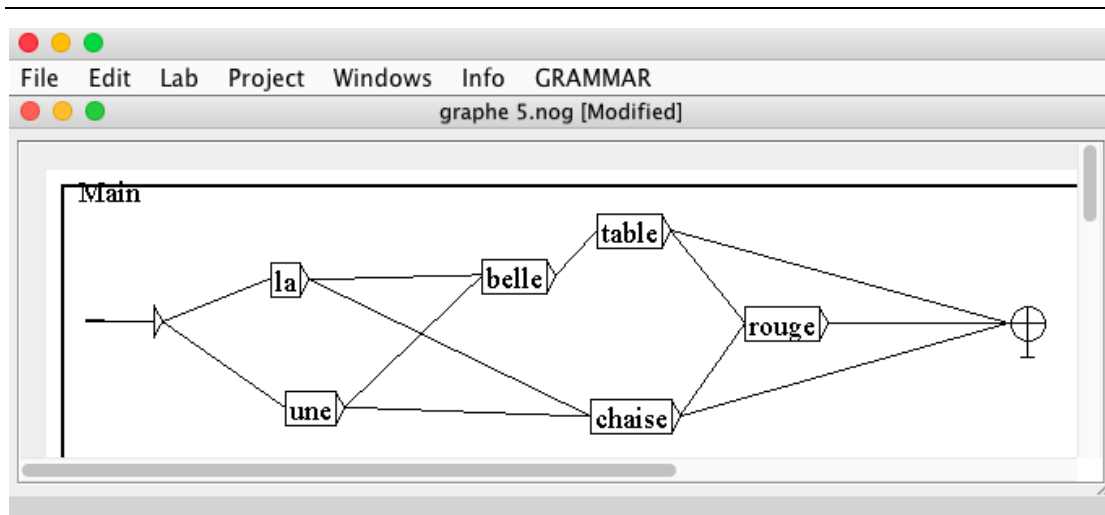


Figure 2-4 : Graphe à états finis

Un graphe à états finis est constitué d'un ensemble de nœuds connectés et orientés entre eux et le premier et le dernier nœud s'appellent respectivement le *nœud initial* et le *nœud terminal*. Chaque nœud représente une unité linguistique atomique (ALU) ou le mot vide. La figure 2-4 ci-dessus peut sortir l'ensemble des huit séquences, comme :

{la belle table ; la belle table rouge ; la chaise ; la chaise rouge ; une belle table ; une belle table rouge ; une chaise ; une chaise rouge }

Il rejette les séquences incorrectes comme :

{la belle chaise; la belle chaise rouge ; une table belle ; ...}

Afin de mieux expliquer la grammaire rationnelle, nous l'illustrons par des exemples d'application dans *NooJ* (3 points : a, b, c)

a. cette grammaire peut exprimer un mot ou une catégorie grammaticale.

Par exemple, le mot table s'exprime en grammaire graphique dans la figure 2-5 ci-dessous :

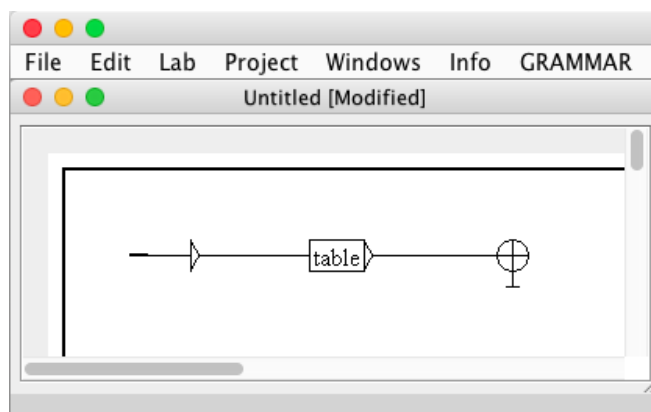


Figure 2-5 : Graphe d'un lexique

et la grammaire graphique dans la figure 2-6 ci-dessous permet de décrire tous les noms (<N><sup>55</sup>) définis dans les dictionnaires de *NooJ*.

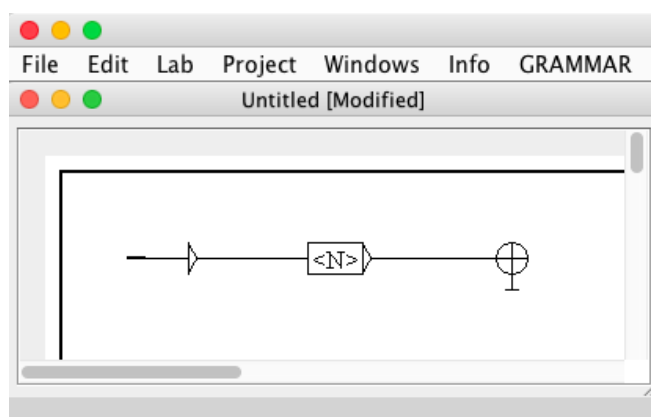


Figure 2-6 : Graphe des formes nominales

Les deux dictionnaires utilisés dans *NooJ* sont le *DEM* (Dictionnaire Electronique des Mots, Dubois & Dubois-Charlier, 2010) et le *LVF* (Les Verbes Français, Dubois & Dubois-Charlier, 1997). D'après Silberztein (2015), ces deux dictionnaires sont en mesure de « décrire le vocabulaire français de façon exhaustive » (ibid. : 149) et « chaque entrée lexicale y est associée à ses propriétés morphologiques, syntaxiques et sémantiques » (ibid.). C'est la raison pour laquelle ces deux dictionnaires sont intégrés dans *NooJ*.

**b.** elle peut exprimer une disjonction.

Nous pouvons construire deux grammaires, comme,  $G_1 = \text{table}$  ;  $G_2 = \text{chaise}$ . Au lieu d'appliquer les deux grammaires précédentes, nous pouvons

<sup>55</sup> Dans *NooJ*, le symbole <N> représente toutes les formes nominales.

également construire une seule grammaire textuelle en utilisant l'opérateur de disjonction « | » (cette opération fonctionne comme l'addition (+) en mathématique) :

$$G_3 = G_1 \mid G_2 = \text{table} \mid \text{chaise}$$

ou en une seule grammaire graphique dans la figure 2-7 :

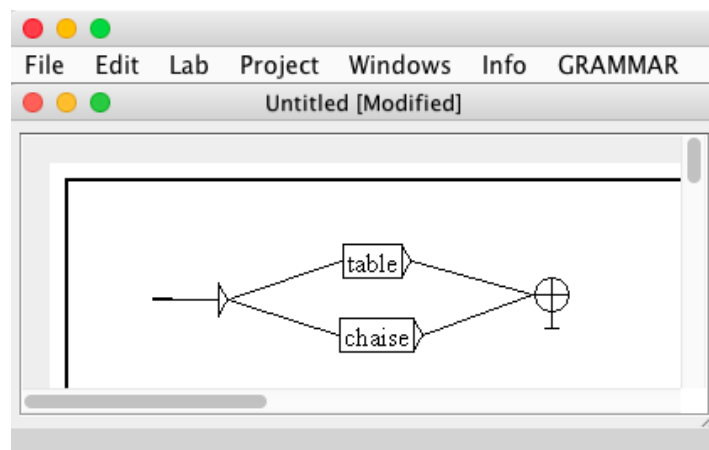


Figure 2-7 : Graphe de disjonction du lexique

c. elle peut exprimer une concaténation.

Afin d'exprimer une concaténation, en premier lieu, nous construisons les deux premières grammaires rationnelles comme :  $G_1 = \text{aim}$  ;  $G_2 = e \mid \text{es} \mid \text{ons} \mid \text{ez} \mid \text{ent}$ . Ensuite, nous pouvons exprimer la concaténation des deux grammaires précédentes en grammaire textuelle :

$$G_3 = G_1 G_2 = \text{aim} (e \mid \text{es} \mid \text{ons} \mid \text{ez} \mid \text{ent})$$

ou en grammaire graphique dans la figure 2-8 ci-dessous :

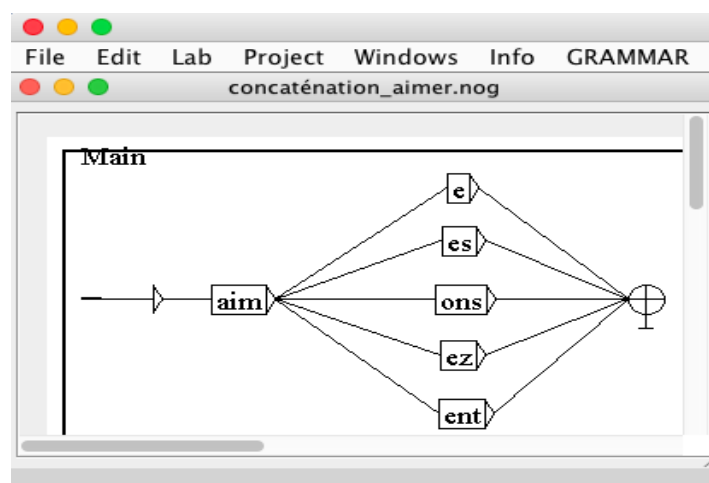


Figure 2-8 : Graphe de concaténation lexicale

Nous pouvons avoir le résultat de cette concaténation comme :

{"aime", "aimes", "aimons", "aimez", "aiment"}

Il est important de noter qu'il ne faut pas oublier les parenthèses dans la troisième grammaire, car sans les parenthèses, le résultat sera différent. En effet, cela donnera les extractions suivantes : {aime, es, ons, ez, ent}. Les parenthèses fonctionnent comme les multiplications (×) en mathématique.

Pour mieux comprendre, nous donnons un exemple mathématique :

$$2 \times (5+6) \neq 2 \times 5 + 6.$$

Avec les parenthèses, le résultat du premier calcul est de 22 et sans les parenthèses, celui du deuxième sera de 16.

Hormis les trois applications (expression d'un mot ou d'une catégorie grammaticale, expression de disjonction du lexique et expression de concaténation) au niveau de la grammaire rationnelle, *NooJ* offre également un opérateur + EXCLUDE pour exclure toutes les séquences indésirables. Dans cette opération, les séquences indésirables seront rejetées en imposant des contraintes, comme par exemple, le graphe à états finis dans la figure 2-9 ci-dessous :

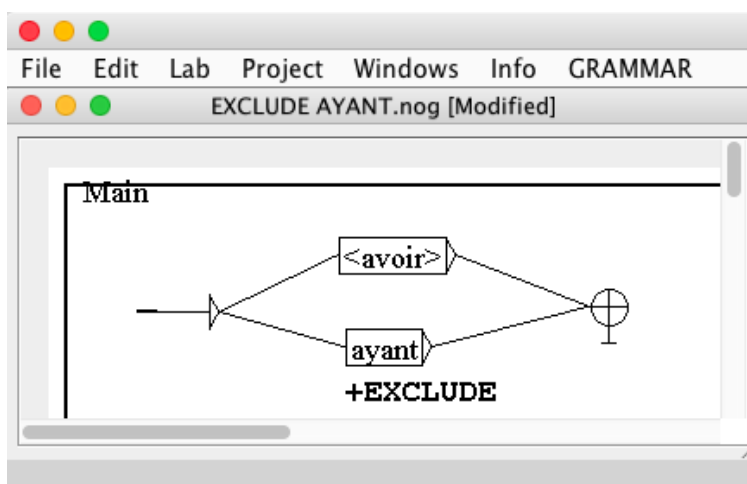


Figure 2-9 : Utilisation de l'Opérateur +EXCLUDE

Il permet de reconnaître toutes les occurrences <avoir><sup>56</sup> et ses variations, sauf la forme ayant. Après avoir vu les applications dans *NooJ* au niveau de la grammaire rationnelle, nous allons voir une grammaire plus puissante : la grammaire algébrique.

### 2.5.2 Grammaire hors contexte ou algébrique

Le membre gauche de ce type de grammaire contient un seul symbole auxiliaire et le membre droit est une combinaison de symboles terminaux et de symboles auxiliaires, par exemple :

PHRASE  $\rightarrow$  GN voit GN

L'atout de ce type de grammaire est d'entraîner la récursivité de règles, c'est-à-dire que les règles sont définies à partir d'elle-même. Par exemple, dans la règle S ci-dessous :

$S = a : \underline{S} b ;$

Dans le membre droit, on trouve la règle « S » (les symboles auxiliaires introduits par le caractère « : » dans *NooJ*), c'est-à-dire qu'il fait apparaître des règles autoréférentielles. Les résultats produits peuvent être, en fonction de la fréquence de l'application de la règle S :

$S \rightarrow ab, aabb, aaabbb, aaaabbbb, \dots$

Nous ne pouvons pas avoir ce langage  $a^n b^n$  dans la règle rationnelle, mais la grammaire algébrique peut entraîner cette récursivité.

Pour l'application de la récursivité de règles, *NooJ* offre les nœuds colorés (préfixés par le caractère « : ») qui permettent d'imbriquer une autre grammaire (ou un autre graphe). Comme l'explique Silberztein, « les nœuds colorés, qui correspondent aux nœuds auxiliaires des règles de réécriture, sont des références à des graphes imbriqués » (2015 : 216). Par exemple, dans la grammaire de phrase dans la figure 2-10 ci-dessous :

---

<sup>56</sup> Dans *NooJ*, les guillemets simples (< >) permettent de retrouver toutes les occurrences de ce terme et de ses variantes.



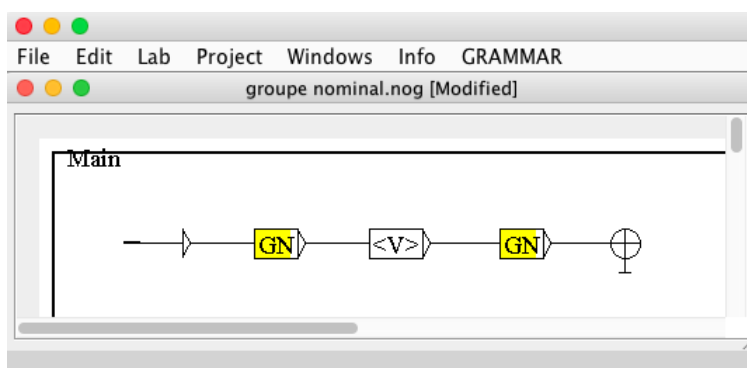


Figure 2-10 : Graphe de la grammaire Phrase

Les nœuds GN sont en couleur, c'est-à-dire que le GN s'explique par un graphe imbriqué (cf. figure 2-11 ci-dessous).

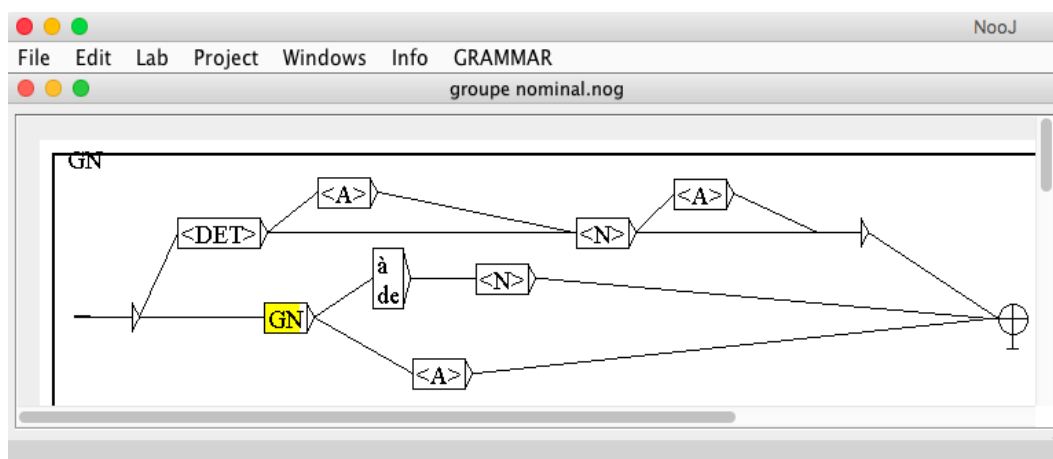


Figure 2-11 : Grammaire de groupe nominal

Dans ce graphe imbriqué, il existe également un GN qui fera référence à un graphe imbriqué (lui-même), donc cette grammaire peut être récursive. Sur le plan de la grammaire algébrique dans *NooJ*, les règles de la grammaire peuvent être récursives. Et qu'apportent les grammaires contextuelles ?

### 2.5.3 Grammaire contextuelle

Un contexte peut s'ajouter aux deux membres de cette grammaire. Les grammaires peuvent s'écrire par un symbole auxiliaire ou terminal dans un quelconque membre de cette grammaire, par exemple :

PLURIEL PHRASE  $\rightarrow$  PLURIEL GN voient GN

Les grammaires contextuelles sont les grammaires génératives limitées par des automates linéairement bornés qui présentent les conditions d'application

des règles en prenant en compte le contexte. Il est impossible de présenter ce que les grammaires contextuelles peuvent faire, car les contextes sont tellement variés que nous ne pouvons pas les présenter au cas par cas. Mais, afin de contextualiser chaque grammaire, l'utilisation des variables semble incontournable. De ce fait, dans cette partie, nous avons tout intérêt à présenter trois variables dans *NooJ*.

Dans la figure 2-12 ci-dessous, nous décrivons l'accord (en genre et en nombre) de la structure Nom + Adjectif avec les grammaires algébriques. Dans un premier chemin, un nom masculin singulier (N+m+s) reconnaît un adjectif masculin singulier (A+m+s) et il pourrait avoir un élément d'insertion (WF), par exemple un adverbe. Afin de décrire cette seule structure de deux éléments, nous sommes obligés de construire au minimum quatre chemins (grammaires algébriques).

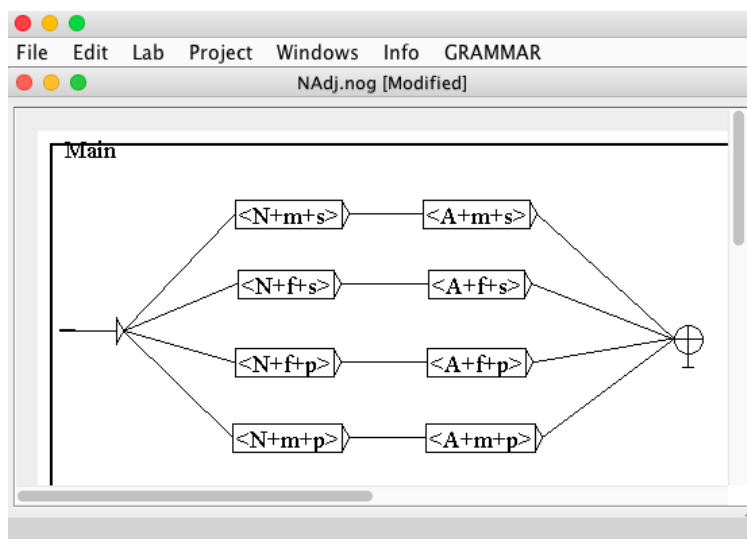


Figure 2-12 : Accord de la structure N+A

Imaginons que nous aurons beaucoup plus de chemins pour une structure plus complexe. C'est l'une des raisons pour lesquelles nous avons recours à d'autres solutions. En effet, il suffit d'un seul chemin dans la figure 2-13 ci-dessous pour décrire la structure N+A en utilisant des variables dans *NooJ*.

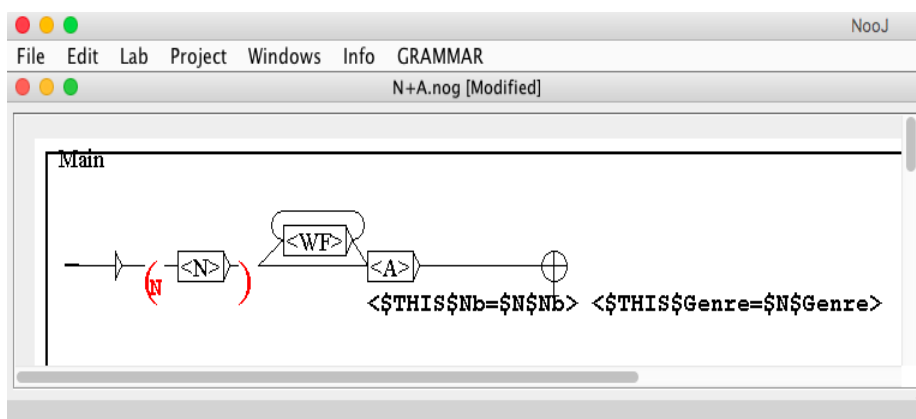


Figure 2-13 : Accord de la structure N+A avec des variables

Dans cette grammaire, deux types de variables sont utilisés : variable `$THIS` et variables locales. Le premier type dans *NooJ* « a toujours pour valeur l'ALU courante » (ibid. : 239), l'ALU courante dans la figure 2-13 ci-dessus désigne les adjectifs. La valeur du deuxième type peut être obtenue dans le même graphe ou un graphe imbriqué. La figure 2-13 affecte à la variable `$N` les noms. Deux contraintes `<$THIS$Nb=$N$Nb>` et `<$THIS$Genre=$N$Genre>` sont employées pour vérifier que les adjectifs s'accordent en genre et en nombre avec les noms : la première contrainte vérifie que l'adjectif (`$THIS`) et le nom (`$N`) ont la même valeur pour la propriété « Nb » (le nombre) et la deuxième contrainte vérifie qu'ils ont la même valeur pour la propriété « Genre ». Si l'on applique cette grammaire à l'expression *viande hachée*, la variable `$THIS` prendra donc la valeur « haché » et la variable `$N` prendra la valeur « viande ».

Hormis les deux premières variables (`$THIS` et variable locale), *NooJ* peut également fournir un troisième type de variables : les variables globales (préfixées par le caractère « @ »). Ces variables peuvent avoir leurs valeurs affectées dans le même graphe ou dans un graphe de niveau supérieur. Chaque variable globale n'a qu'une seule valeur, indépendamment de sa place dans la structure grammaticale et cette valeur peut être transmise au générateur de *NooJ*. Puisque les phrasèmes *NAdj* peuvent subir des transformations, les variables globales peuvent nous aider à implanter les données dans *NooJ* pour en extraire les phrases transformationnelles.

## Bilan

En raison de sa présence abondante et du besoin de désambiguïisation sémantique, la phraséologie *NAdj* constitue un choix judicieux pour notre étude. La phraséologie provoque un intérêt considérable dans la recherche fondamentale et la recherche appliquée. Elle reste pourtant un domaine très flou sur plusieurs plans (entre autres, les critères de reconnaissance et la classification). L'extraction et l'enseignement de la phraséologie, évoqués ci-après, nous imposent de procéder à un traitement linguistique. Les trois critères pertinents (fréquence, figement, pragmatique) sont retenus pour notre classification des phrasèmes culinaires. À la suite des trois critères, les trois types de phrasèmes (expression phraséologique, expression semi-phraséologique, pragmatèmes) font partie de notre étude. La modélisation des données lexicales est l'étape préliminaire pour l'extraction des phrasèmes et des notions de modélisation (lexique grammaire, grammaire locale, fonction lexicale) qui peut se voir dans des projets phraséologiques. Pour extraire exhaustivement les phrasèmes, nous avons fait l'inventaire des programmes développés et des logiciels pour la description des langues (notamment *NooJ*). En observant des exemples de l'application *NooJ*, nous nous rendons compte que *NooJ* est plus flexible que d'autres programmes développés : au niveau de la grammaire rationnelle, nous pouvons décrire un mot, une catégorie grammaticale, une disjonction du lexique et une concaténation lexicale et nous pouvons également faire une désambiguïisation lexicale avec l'opérateur +EXCLUDE ; au niveau de la grammaire algébrique, les règles de la grammaire peuvent se réaliser récursivement ; au niveau de la grammaire contextuelle, les trois types de variables (\$THIS, variables locales et variables globales) peuvent faciliter une certaine description. Par la suite, nous allons étudier la phraséologie dans l'enseignement.

### 3. Utilisation du corpus pour l'E/A de la phraséologie

#### 3.1 E/A de la phraséologie

##### 3.1.1 Nécessité de l'E/A de la phraséologie en FLE

Les linguistes et les didacticiens ont constaté la nécessité de l'utilisation de la phraséologie chez les apprenants étrangers. Le premier constat établi (entre autres, Cavalla, 2009 ; González-Rey, 2010) consiste en la présence abondante d'unités phraséologiques (cf. introduction). À propos de cette idée de récurrence, González-Rey (2007) remarque que les locuteurs natifs utilisent couramment les « prêt à parler » (ibid. : 7), mais ils ne s'en rendent pas compte. Prenons l'exemple de Cavalla (2009) : les locuteurs natifs oublient le sens habituel du verbe *poser* dans la collocation *poser une question*, tandis que les locuteurs non natifs tendent à appliquer le verbe *poser* de façon concrète : *poser la question sur la table*. Comme l'explique Cavalla (ibid.), certaines unités phraséologiques « passent quasiment inaperçues aux yeux des locuteurs natifs qui métaphorisent sans s'en rendre compte », mais un apprenant étranger « ne peut prévoir la suite d'une collocation sans la connaître en raison notamment de l'aspect culturellement marqué de la métaphore utilisée » (2009 : 4) et le manque de connaissances culturelles chez les apprenants étrangers conduit à l'impuissance dans la distinction du synonyme et dans la traduction d'unités phraséologiques :

- Distinction de synonymes : les lexies *discours* et *conférence* sont synonymes dans l'expression *faire un(e) discours/conférence* pour dire que quelqu'un présente une communication dans un congrès ou un colloque, mais Bossé-Andreu et Mareschal (1998 : 166) distinguent ces deux lexies dans les deux expressions suivantes : *prononcer un discours* et *donner une conférence* comme indiqué dans la figure 3-1 ci-dessous.



Figure 3-1 : Distinction du synonyme discours/conférence

Car on ne dit ni \*donner un discours ni \*prononcer une conférence en français de France <sup>57</sup>. En ce qui concerne la distinction des synonymes, nous pouvons reprendre les exemples célèbres de Mel'čuk et de Bally : on dit *plonger dans le désespoir*, mais *mettre en rage* <\*mettre en désespoir, \*plonger dans la rage> et on a de grands blessés <\*grave blessés>, alors qu'une blessure est grave <\*grande blessure> (Mel'čuk, 1997 : 23) ; on peut être *gravement ou grièvement blessé*, mais seulement *gravement malade* <\*grièvement malade> (Bally, 1909 : 70).

- Traduction d'unités phraséologiques : afin de décoder une langue, la culture nous est nécessaire. On dit, par exemple, en français *malin comme un singe*, mais en espagnol et en chinois, le renard est considéré comme l'animal le plus malin et on dit donc *intelligent comme un renard*. Un autre exemple sera illustré dans la figure 3-2 ci-dessous.

<sup>57</sup> Bossé-Andreu et Mareschal (1998) ont trouvé « 164 occurrences de *prononcer une conférence* et 31 de *donner un discours* » dans un corpus canadien (ibid. : 166).

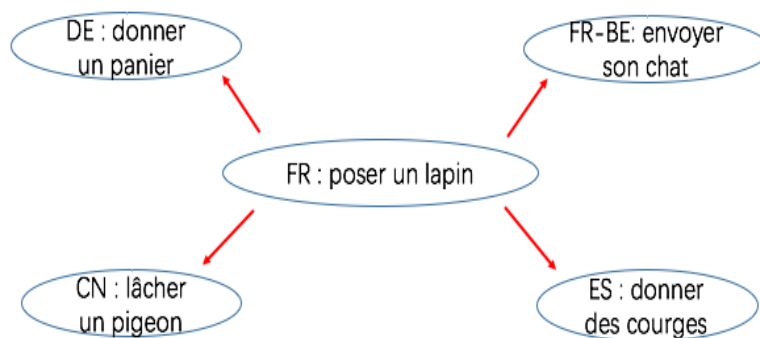


Figure 3-2 : Traductions de *poser un lapin* dans différentes langues

Comme indiqué dans la figure 3-2 ci-dessus, *poser un lapin* en français est exprimé différemment dans d'autres langues : *envoyer son chat* en français de Belgique ; *donner des courges* en espagnol ; *donner un panier* en allemand et *lâcher un pigeon* en Chinois.

Imaginons que si un locuteur étranger ne connaissait pas l'expression *poser un lapin*, cette expression ne pourrait être traduite correctement, car principalement, comme le dit Cavalla (2009), l'absence de connaissances culturelles ne le permet pas. En outre, le sens de cette expression est non compositionnelle, les apprenants ne peuvent pas les traduire mot à mot. Cela nous permet de rappeler deux exemples parlants : *avoir le cafard* de Cavalla (2009 : 7) et *Paris vaut bien une messe* de González-Rey (2010 : 4).

### 3.1.2 E/A des phrasèmes dans les ouvrages de FLE

L'importance de l'enseignement de la phraséologie a été soulignée dans l'introduction de ce travail et le CECRL exige, au niveau de la compétence lexicale, une certaine connaissance et manipulation des structures lexicales de la langue : « Il s'agit de la connaissance et de la capacité à utiliser le vocabulaire d'une langue qui se compose : 1. d'éléments lexicaux et 2. d'éléments grammaticaux ainsi que de la capacité à les utiliser. » (Conseil de l'Europe, 2001 : 87). L'E/A des phrasèmes est donc incontournable. Cet E/A doit se faire à partir du niveau A1, car des phrasèmes peuvent s'employer « tout au long des consignes, des textes, des exercices des manuels »

(Cavalla, 2008 : 6), par exemple, *poser une question et résoudre un problème*. Mais la phraséologie est un domaine peu abordé en FLE (Binon & Verlinde, 2003 ; González-Rey, 2007 ; Cavalla, 2008), comme le soulignent Cavalla et Labre, les phrasèmes « ne font que très rarement l'objet d'un enseignement spécifique » (2009 : 305) et l'absence de cet enseignement se reflète notamment dans les manuels de FLE : la phraséologie « n'apparaît pas encore systématiquement dans les manuels d'enseignement » (Cavalla, 2016 : 68). Nous avons cependant trouvé des soutiens phraséologiques dans les ouvrages complémentaires de FLE organisés dans l'ordre chronologique dans le tableau 3-1 ci-dessous.

Année	Noms des ouvrages complémentaires
1984a	<i>Les expressions imagées</i> (Galison Robert)
2004	<i>Le chemin des mots, pour un apprentissage méthodique du vocabulaire français</i> (Dumarest Danièle & Morsel Marie-Hélène)
2005	<i>Émotions-Sentiments, nouvelle approche lexicale du FLE</i> (Cavalla Cristelle & Crozier Elsa)
2005	<i>Dites-moi un peu, méthode pratique du français oral</i> (Hingue Anne-Marie & Ulm Karine)
2007	<i>La didactique du français idiomatique</i> (González-Rey Isabel, 2007)
2008	<i>Les expressions idiomatiques</i> (Chollet Isabelle & Robert Jean-Michel)
2008	<i>De l'espagnol...au français. Expressions et mots usuels du français</i> (Rolland Jean-Claude & Rolland-Pérez Daria)

Tableau 3-1 : Ouvrages complémentaires de FLE concernant l'enseignement de la phraséologie

*Les expressions imagées* (Galison, 1984a) édité chez Clé International est le premier ouvrage destiné à l'enseignement de la phraséologie aux apprenants étrangers (notamment à un public anglophone). Puisque cet ouvrage est un livret d'auto-apprentissage, Galison propose à l'utilisateur, dans son avant-propos, de faire les exercices en se référant à son *dictionnaire des expressions imagées – Compréhension et Production* (Galison, 1984b).



L'inconvénient de ce livre réside dans le fait que les activités proposées ne sont pas organisées selon une progression. En outre, il manque de la contextualisation pour les activités, ce qui n'aide pas à la mémorisation. À la suite de ce premier ouvrage, trois ouvrages dans le champ de l'enseignement du lexique sont publiés chez PUG (Presse Universitaire Grenoble) : *Le chemin des mots, pour un apprentissage méthodique du vocabulaire français* (Dumarest & Morsel, 2004) ; *Émotions-Sentiments, nouvelle approche lexicale du FLE* (Cavalla & Crozier, 2005) et *Dites-moi un peu, méthode pratique du français oral* (Hingue et Ulm, 2005). Ces trois ouvrages proposent des activités très pertinentes pour l'appropriation du lexique, mais d'après González-Rey, les trois ouvrages « ne sont pas déclarés comme étant proprement phraséologiques » (2010 : 6). Ce point de vue nous semble discutable, car nous pensons que la langue est un tout et qu'on ne peut sortir un élément à part. Ainsi donc, pour nous, la phraséologie doit être enseignée au fil des rencontres, avec le reste, dans les discours spécifiques où on les trouve. Deux ouvrages *Les expressions idiomatiques* (Chollet et Robert, 2008) et *De l'espagnol...au français. Expressions et mots usuels du français* (Rolland, 2008) sont tout particulièrement destinés « aux apprenants de français langue étrangère » (Chollet & Robert, 2008 : 3). Mais le manque d'exercices d'application ne permet pas de réaliser un apprentissage efficace. L'ouvrage *La didactique du français idiomatique* (González-Rey, 2007) constitue un grand apport dans l'histoire de la didactique de la phraséologie. L'auteur de cet ouvrage a pris en considération le genre (contexte) dans l'élaboration de l'activité d'apprentissage des phrasèmes, tels que poésie, théâtre, chanson, conte et roman. Comme le souligne Gonzalez-Rey (2018), « l'enseignement de la phraséologie doit se faire à travers la langue utilisée dans des genres littéraires » (ibid. : 304). Et bien entendu, cet ouvrage comporte également des imprécisions : les expressions phraséologiques et semi-phraséologiques ne sont pas bien distinguées dans certaines parties (ex. thème sur les animaux) ; les *formules routinières* (*les pragmatèmes* de

Mel'čuk, 1998) sont également prises en compte dans un ouvrage de l'enseignement du français idiomatique. En somme, s'inscrivant dans une perspective de proposer des phrasèmes sur le français général ou sur des domaines spécifiques (émotion, littéraire...), ces sept ouvrages peuvent offrir des soutiens phraséologiques aux apprenants de FLE, mais peu d'aide au public de notre projet. En ce qui concerne le domaine culinaire, deux ouvrages sont récemment édités chez Clé International au service de l'E/A de la langue utilisée dans la cuisine française : *En cuisine* A1-A2 (Cholvy, 2014) et *En cuisine et en salle* B1-B2 (Bencini, Cangiolli, Naldini & Paris, 2016) dans lesquels des éléments phraséologiques font l'objet d'enseignement, tels qu'*enlever la queue*, *monter des blancs en neige*. Toutefois, les auteurs du manuel ne consacrent que peu de place à l'E/A de la phraséologie du domaine culinaire. En l'absence de recours aux ouvrages de FLE, nous nous occupons de tous les aspects didactiques des phrasèmes culinaires. Autrement dit, nous devons proposer des phrasèmes culinaires pertinents à enseigner/apprendre, des activités ludiques à réaliser et des séquences à donner. Afin de mieux comprendre le phénomène de mémorisation pour l'apprentissage d'une langue, notamment la mémorisation du lexique, nous verrons, par la suite, comment se servir des connaissances de la discipline psycho-cognitive (via les neurosciences) pour aider à l'apprentissage du lexique.

### 3.1.3 Acquisition des langues en L2

#### 3.1.3.1 Organisation lexicale et lexique mental

Combien de mots de sa langue maternelle une personne possède-t-elle en mémoire ? En neurosciences, la réponse est très variable, mais les neurologues de la fin du XX<sup>ème</sup> siècle estimaient cette quantité à 50 000 mots environ (Aitchison, 1987 ; Miller & Gildea, 1987 ; Nagy & Herman, 1987) situés dans l'hémisphère gauche du cerveau (Lively, Pisoni & Goldinifer, 1994). Aujourd'hui, le stockage n'est plus aussi précisément décrit qu'auparavant car les chercheurs commencent à comprendre que le stockage du lexique se

ferait sous forme de réseaux neuronaux et il devient donc très difficile d'affirmer un chiffre exact quant au nombre de mots connus par un individu. En fait, l'idée de réseau lexical n'est pas nouvelle (introduit par Trier, 1931, cité par Bierwisch, 1970 ; Collins & Loftus, 1975 ; Heiden, 2004 ; Tournier, 2002 ; Williams, 1999 ; 2001 ; Duffau, 2016) et suit les études en neurosciences qui expliquent que le cerveau fonctionne en réseaux de neurones et chaque neurone possède plusieurs milliers de dendrites et d'axones agissant comme connecteurs réagissant au contact d'autres neurones. Afin de mieux comprendre l'organisation lexicale dans le cerveau, certains psycholinguistes utilisent une analogie entre la représentation lexicale et le neurone. Comme l'explique Boulton,

*La représentation mentale d'une unité lexicale correspond à un neurone ; les liens entre ces différentes unités sont représentés par les connexions neurales ; la récupération d'une unité correspond à l'activation d'un neurone ; l'ensemble du lexique forme un réseau d'une complexité et d'une puissance considérable (1998 : 33).*

D'après cette analogie, chaque représentation mentale (entrée lexicale) correspond à un neurone et les relations entre différentes unités lexicales ressemblent aux connexions neurales. De toute évidence, une grande part du lexique – d'une grande ampleur – s'organise d'une manière logique pour que, lors de la production, un élément de ce lexique puisse être tiré en moins d'une fraction de seconde. Cette organisation doit être complexe, mais cela ne fait pas partie de mes études. Toutefois, il faut reconnaître que nous pouvons nous interroger sur l'accès au lexique dans ces conditions, comme l'ont fait les chercheurs autour de Edmonds (2013). Cette organisation entre les éléments du lexique dans le cerveau peut être envisagée comme étant le « lexique mental » tel que rencontré chez Nagy, Herman & Anderson (1985) puis Van der Linden (2006). Au départ, ces études se font notamment dans le domaine de la psychologie et de l'acquisition pour étudier le fonctionnement des mots déjà existants dans la mémoire. La situation a changé dans les années 80 et les linguistes-didacticiens s'y intéressent pour développer des stratégies

d'aide à l'E/A du lexique. Retenons que dans les deux disciplines (psychologie et acquisition des langues), la notion de lexique mental est considérée comme une description de l'organisation lexicale dans le cerveau. En outre, le lexique mental ne décrit pas seulement le stockage lexical, mais aussi les mécanismes du traitement. Comme le dit Boulton, « le stock lexical et les mécanismes qui le traitent constituent ensemble le lexique mental » (1998 : 17). En effet, le lexique mental en L1 (langue maternelle) est différent du lexique mental en L2 (langue seconde/étrangère) et notre étude porte sur ce dernier. Cependant, il nous semble important de commencer par l'acquisition en L1 pour comprendre cette organisation chez les apprenants en L2.

### 3.1.3.2 Lexique mental en L1

Les psychologues et acquisitionnistes essaient de comprendre la manière dont s'organise le lexique en notant les lapsus émis par les locuteurs, car ces phénomènes révèlent certains fonctionnements. Par exemple, un locuteur dit « on a eu une conversion téléphonique » au lieu de dire « on a eu une conversation téléphonique ». Les deux unités lexicales se ressemblent au niveau de la forme phonologique et cela nous permet d'émettre l'hypothèse selon laquelle le lexique serait aussi organisé selon la forme orthographique. Afin d'émettre davantage d'hypothèses et de les confirmer, Arnaud (1993) réalise une analyse quantitative en se basant sur une collection de lapsus des locuteurs. Le résultat révèle qu'une représentation lexicale dans le réseau lexical L1 est composée de trois domaines : orthographique, phonologique et sémantique. Parmi ces trois domaines, il confirme également que les deux premiers se stabilisent rapidement et que le dernier pourrait évoluer en permanence tout au long de l'apprentissage. D'après Prince (1999) et Van der Linden (2006), le lexique mental en réseau serait davantage fondé sur le sens que sur les formes et les sons. Comme le rappelle Zinglé (2003), les phrasèmes expriment toujours « un sens précis » (2003 : iii). Nous nous demandons alors si chaque phrasème est rangé dans le cerveau comme un

mot isolé. Notre hypothèse est confirmée par Edmonds : « Toute unité phraséologique aurait sa place dans le lexique mental comme toute unité lexicale simple (mot) » (2013 : 122), autrement dit que l'individu n'accède pas plus rapidement à un mot isolé qu'à une expression constituée de plusieurs mots (Bally, 1951 ; Cavalla, 2016 ; Edmonds, 2016) et qu'il n'y a pas de relation entre la longueur du mot et le fait de le retenir (Chambers, 2005 ; 2007 ; Lewis ; 2000). Cela nous permet de comprendre pourquoi les phrasèmes sont conçus comme une inscription mémorielle chez les linguistes anglais (entre autres, Pawley & Syder, 1983 ; Skehan, 1998 ; Wood, 2002) ; en outre, cette propriété des phrasèmes est décrite dans la définition dominante anglophone de la phraséologie :

*a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subjected to generation or analysis by the language grammar<sup>58</sup>.* (Wray & Perkins, 2000 : 1)

Cette propriété explique également pourquoi les phrasèmes sont appréhendés et produits facilement par rapport aux séquences non associées chez les locuteurs natifs (Cacciari & Tabossi, 1993 ; Cronk & Schweigert, 1992 ; Nenonen, Niemi & Laine, 2002 ; Qualls *et al.*, 2003 ; Swinney & Cutler, 1979). Ainsi, la thèse de Lewis à propos du stockage de « chunks » est tout à fait soutenable : il est facile pour le cerveau de retenir des éléments figés (ou semi-figés), car leur récupération n'est pas plus complexe que pour un mot simple tandis que leur sens est plus large ou précis selon les cas.

### 3.1.3.3 Lexique mental en L2

Comme en L1, les mots en L2 seraient organisés sous forme de lexique mental selon trois niveaux : orthographique, phonologique et notamment sémantique. Il va de soi que chaque phrasème chez les apprenants étrangers

---

<sup>58</sup> Une séquence, continue ou discontinue, de mots ou de tous autres éléments doués de sens, qui est, ou semble être, préfabriquée : c'est-à-dire, stockée dans la mémoire et extraite de celle-ci dans son ensemble au moment de l'utilisation, plutôt que d'être sujette à la génération ou à l'analyse grammaticale (notre traduction).

correspond également à une entrée lexicale du lexique mental (Conklin & Schmitt, 2008 ; Underwood, Schmitt & Galpin 2004). Pourtant, les apprenants étrangers construisent un réseau au fil de leur apprentissage, leur stock de lexique augmente et les éléments syntaxiques se mettent en place afin de créer un réseau pour la L2. La question des relations entre les deux langues L1 et L2 se pose alors. Il y a quelques décennies déjà, Weinreich (1953) avait soulevé cette question et proposait les trois modèles (figure 3-3 ci-dessous) pour décrire trois relations potentielles entre les deux langues :

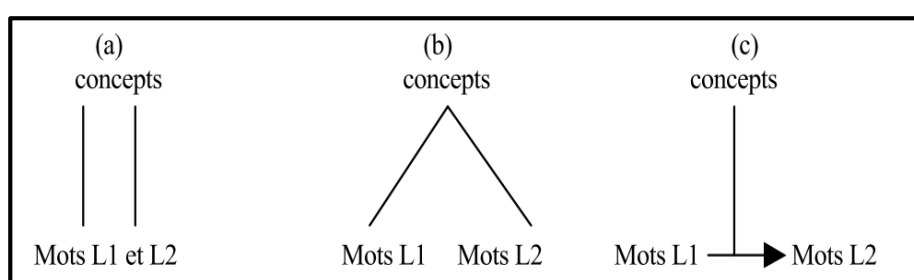


Figure 3-3 : Trois modèles proposés par Weinreich

Dans le modèle (a), les deux langues sont intégrées ; dans le modèle (b), les deux langues sont bien séparées et il n'y a pas de lien direct entre elles ; dans le modèle (c), les deux langues sont également séparées, mais pour accéder à la L2, on passe obligatoirement par la L1. Les deux premiers modèles ont été testés par Kerkman & De Bot (1989). Le résultat montre que les deux langues ne sont ni intégrées, ni complètement séparées. Potter, So, Von Eckardt & Feldman (1984) ont rejeté le troisième modèle, car le locuteur ne commence pas toujours par la L1 pour accéder à la L2 mais parfois c'est l'inverse qui se produit. À partir des années 90, les modèles des relations entre deux langues se sont multipliés (entre autres, Kroll & Stewart, 1994 ; Boulton, 1998), mais le débat est centré autour de deux questions : le degré de séparation des deux langues et l'accès au lexique mental. Afin de répondre aux deux questions, Boulton (1998) propose le modèle interactif dans la figure 3-4 ci-dessous :

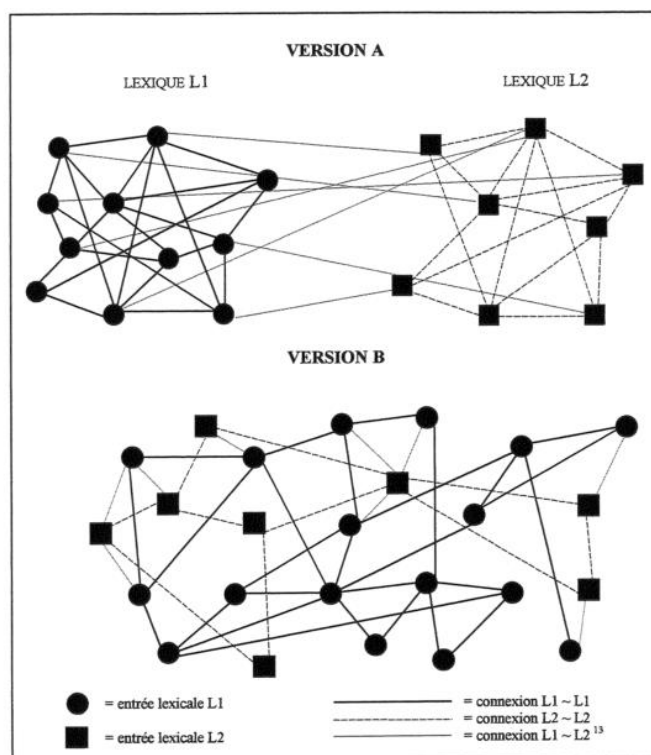


Figure 3-4 : Modèle interactif du lexique mental (Boulton, 1998)

Dans la version (a) de ce modèle, les deux langues sont séparées, mais elles ont encore des connexions interlinguistiques et dans chaque langue, les connexions intralinguistiques apparaissent clairement. Les cercles pour la L1 et les carrés pour la L2 représentent une entrée lexicale (d'un mot ou d'une unité lexicale) et les traits rendent compte de la conscience métalinguistique. À partir de ce modèle, nous constatons que les entrées lexicales existent à la fois en L1 et en L2. Au moment de s'exprimer, les locuteurs cherchent un exemple parmi les langues qu'ils connaissent déjà. C'est pourquoi on cherche de l'aide dans sa langue maternelle quand on n'arrive pas à s'exprimer en langue étrangère. Plus on améliore la L2, plus celle-ci devient autonome sans recourir à une autre langue. La version (b) a le même concept que celui de la version (a) et elle simule la manière dont les multiples entrées lexicales sont stockées dans le cerveau : que ce soit en L1 ou L2, les entrées lexicales se situent côte à côte dans le cerveau.

Étant donné que le lexique en L1 et en L2 est principalement organisé en se basant sur le sens, Krashen (1989) pense que la meilleure façon d'augmenter

le vocabulaire consiste à lire beaucoup en accédant au sens pour le plaisir et non pas pour l'analyse de la syntaxe. Peut-on donc supposer que l'augmentation de la fréquence des rencontres de vocabulaire de manière inconsciente peut aider à l'apprentissage des mots en L2 ? Selon le résultat de Chen (1990) et de De Groot (1995), cette supposition a peu d'effet sur l'apprentissage lexical spécifique, car le traitement des mots inconnus conduit souvent à des malentendus ou à des erreurs sémantiques (Hulstijn, 1992) et l'avancement est lent et incertain (Paribakht & Wesche, 1997). De plus, il est nécessaire de faire une demande au niveau de la compétence des apprenants pour bien comprendre la lecture selon Coady (1997) et Huckin & Coady (1999). Enfin, la consolidation de cet apprentissage se fait par hasard. Il est constaté qu'un espacement optimal et un contexte approprié sont deux critères importants dans cette rencontre pour l'apprentissage du lexique en L2 (Prince, 1999). Quant à l'espacement optimal des rencontres, Gairns & Redman (1986) proposent de s'espacer davantage dans le temps pour consolider les connaissances acquises. À l'égard du contexte des rencontres, il s'avère que l'E/A du lexique dans le contexte L2 est nécessaire (Nation & Newton, 1997 ; Lewis, 1993), mais Prince (1996) constate que le fait de choisir un contexte approprié n'est pas facile.

#### 3.1.4 E/A du lexique au genre

La mise en contexte favorise l'apprentissage (Moulin, 1984 ; Prince, 1996 ; 1999 ; Ellis, 1997 ; Boulton, 2000 ; Hilton, 2002 ; 2009). Les premières recherches sur l'influence du contexte dans l'apprentissage du lexique ont débuté dans les années 80. Schouten-van Parreren (1985) réplise une expérimentation sur l'effet du contexte dans l'apprentissage du lexique. Il divise ses informateurs en deux groupes. Il demande à un groupe d'informateurs de deviner le sens d'un mot dans un contexte naturel et l'autre groupe sans contexte. Les résultats des deux groupes révèlent que l'apprentissage du lexique est plus efficace en contexte que sans contexte. Il



est évident que l'enseignement du lexique doit se faire à l'aide de son voisinage qui l'entourent (Laufer, 1992 ; Hazenberg & Hulstijn, 1992). Comme le dit Cavalla, il faut toujours enseigner le lexique dans le « contexte dans lequel la lexie ou le polylexical s'insère » (2016 : 37). Schouten-van Parreren (1989) souligne également que les unités 'en isolation' sont difficilement apprises. Par exemple, le *dictionnaire d'apprentissage du français des affaires* (Binon *et al.*, 2000) fournit des contextes de mots pour l'auto-apprentissage des collocations. Toutefois, l'utilisation des contextes inappropriés dans l'apprentissage n'aident pas à la mémorisation (Mondria, 1996 ; Edmonds, 2013). Le contexte est ici pris au sens de genre (cf. 1.4 supra). En effet, la recherche portant sur le genre a beaucoup été menée dans la didactique des langues : en langue maternelle (Bronckart, 1996 ; Schneuwly *et al.*, 1997) et en FLE (Beacco, 1991 ; Chiss, 1987 ; Darot *et al.*, 1984 ; Portillo Serrano, 2010). À propos de l'intérêt de l'E/A du lexique dans un genre, Cavalla (2016) l'explique en trois points. Tout d'abord, « le genre va certainement aider à sélectionner le bon sens en fonction des associations lexico-sémantiques possibles dans un genre déterminé » (ibid. : 52). C'est-à-dire que le sens de la lexie peut être précisé dans un genre de textes, ce qui permet de présenter moins d'ambiguïté sémantique. En outre, partant d'une perspective didactique, elle pense que « le fait d'être dans un genre permettrait au cerveau d'ancrer des liens profonds entre certains éléments, ce qui expliquerait que nous accédons rapidement à ces informations dès l'instant où nous connaissons le genre en question » (ibid. : 49). Enfin, les phrasèmes d'un domaine spécialisé sont spécifiques au genre. Autrement dit, le genre nous permet de sélectionner les structures phraséologiques spécifiques. Sitri et Tutin (2016) confirment cette nécessité, « la question du genre (de discours), devient centrale, et accorde une grande importance à ces formules récurrentes qui jouent un rôle de premier plan dans les différents sociolectes » (ibid. : 6). Mais « L'absence de la phraséologie associée aux genres » dans les manuels de FLE est confirmée par Cavalla (2016 : 14). Cavalla (2016) souligne que la

notion de genre n'a jamais été assez expliquée pour les enseignants et le genre est seulement considéré comme un « type de communication ritualisée dont les normes, d'origine culturelle, ont une incidence sur la forme linguistique elle-même » (Portillo Serrano, 2010 : 120).

Notre enseignement doit se focaliser sur l'utilisation du genre : d'une part, les apprenants pourront préciser le sens des phrasèmes dans un genre, d'autre part, comme le font les autres spécialistes du FOS, l'étude du genre nous permet de sélectionner les structures spécifiques de la cuisine comme d'autres l'ont fait dans des disciplines spécifiques :

- Tutin (2002) a sélectionné les phrasèmes du domaine scientifique dans les articles scientifiques, comme *résultats encourageants* et *mener une analyse* ;
- Analysant des discours de mathématiques, Mangiante (2004) a pu relever les phrasèmes des mathématiques dans, par exemple, *si et seulement si*.
- À travers les articles de presse quotidienne, Cavalla (2005) a pu extraire les collocations du domaine de l'affect, telles que *sauter de joie* et *peur bleue* ;
- Afin de dégager les phrasèmes du domaine des affaires de type : *poursuivre les négociations*, Verlinde, Binon & Selva, (2006) ont collecté des textes journalistiques.

Une fois choisi le genre de texte que nous allons étudier, se pose alors pour l'enseignement la manière dont nous allons sélectionner le lexique.

### 3.1.5 Prise en compte du critère de fréquence dans la sélection du lexique

Depuis longtemps, dans la recherche de l'enseignement des langues, une langue spécialisée qui est considérée « comme une sous-langue ou un sous-système linguistique » (Frandsen, 1998 : 15) est différente d'une langue courante. Des enseignants, notamment les lexicalistes, pensent que les lexiques en FOS ne sont constitués que de lexiques spécialisés, comme le constate Lehmann (1993 : 94) « nombreux sont encore les enseignants pour

---

qui les discours spécialisés se réduisent à des spécificités lexicales ; telle est d'ailleurs la représentation spontanée que s'en fait le commun des mortels », à savoir qu'il n'y a que les terminologies qui font partie des lexiques spécialisés. Par exemple, dans le domaine de la cuisine, les *planchas*, les *fours* et les *planches à découper* sont des terminologies. Les fonctionnalistes consentent à l'importance des lexiques spécialisés dans le cadre de l'enseignement du lexique en FOS ; en revanche, ils soutiennent que « les mots les plus étranges pour un étudiant étranger ne sont pas forcément les termes les plus techniques » (Challe, 2002 : 30) et que ce sont les mots courants qui posent le plus de problèmes aux apprenants, parce qu'ils peuvent conduire à une mauvaise compréhension chez les apprenants de faible niveau (un mot courant peut avoir une connotation dans un contexte précis et avoir un sens très différent dans des contextes professionnels). De plus, Challe nous rappelle également que le lexique spécialisé ne comprend pas seulement des terminologies : « le français de spécialité est loin de se réduire à un problème de lexique scientifique et technique » (2002 : 99). D'où le caractère incontournable de l'approche lexicale en FOS pour les novices ; de plus nos apprenants étrangers ont à la fois besoin des mots courants (du côté fonctionnaliste) et des terminologies de la cuisine française (du côté lexicaliste).

Que ce soit en L1 ou en L2, les acquisitionnistes (entre autres, Hilton, 2002 ; Edmonds, 2013) s'accordent sur le fait que la compétence lexicale n'est pas définie par la quantité de lexique connu mais qu'elle serait davantage déterminée par la rapidité et la facilité de récupération d'un mot parmi un grand volume de mots. Grâce à des tests psychométriques, les psychologues arrivent même à mesurer la rapidité et la facilité de l'identification d'un lexique. Par exemple, dans le test de la décision lexicale, on peut mesurer le temps que les informateurs mettent pour identifier une forme donnée avec un mot existant. Le résultat montre que les mots grammaticaux sont récupérés plus rapidement que les mots lexicaux, que les mots concrets ressortent plus

rapidement que les mots abstraits et que les mots fréquents apparaissent plus vite que les mots plus rares (Hilton, 2002 ; Siyanova & Schmitt 2008 ; Siyanova-Chanturia, Conklin & Van Heuven, 2011 ; Cavalla, 2016). La vitesse de l'activation d'un mot dans le cerveau diffère en fonction de la tendance de l'apprentissage : par exemple, au cours d'une expérience, Plane (2004) constate que les mots de faible fréquence sont souvent des mots difficiles pour ses élèves et Tomasello (2003) remarque que les mots les plus fréquents sont souvent appris au premier plan. En fait, l'apprentissage des mots les plus fréquents de la langue représente également un besoin immédiat pour les apprenants : Nation & Newton (1997) découvrent que les 2000 mots les plus fréquents peuvent représenter 87% de la langue écrite et que si les apprenants comprennent environ 90% des mots d'un texte, ils peuvent comprendre ce texte sans problème (Nation & Coady, 1988 ; Nation, 1990). C'est pourquoi les didacticiens tiennent compte du critère de fréquence dans leurs enseignements (entre autres, Sinclair & Renouf, 1988 ; Lewis, 1993 ; 1997 ; Willis, 1990 ; Ellis, 2002). Comme le dit Prince, « le critère de la fréquence des mots dans la langue reste un bon indice pour le choix des mots à enseigner » (1999 : 342). En vue d'asseoir la compétence lexicale, les didacticiens de différentes langues ont fait leurs tentatives en construisant une liste de mots fréquents : « vocabulaire de base » (Grossmann, 2011 : 166), par exemple, *Basic English* (Ogden, 1934) pour l'anglais, *Français Fondamental* (Gougenheim *et al.*, 1956 ; 1964) pour le français et *Vocabolario di base* (De Mauro, 1983) pour l'italien. Dès que nous cernons les phrasèmes pour l'enseignement, nous nous demandons par quel biais seront introduits les phrasèmes en classe de langue.

### 3.1.6 Entrée didactique pour l'enseignement du lexique

Quant à l'enseignement du lexique, Higa (1963) s'oppose aux activités d'associations sémantiques car ces activités peuvent fragiliser le réseau L2 (rappelons que le lexique mental est basé notamment sur le sens), puisque les

représentations mentales (les formes) ne sont pas précises dans les activités d'associations. En outre, lorsque le réseau L2 est instable, l'interférence des formes apparaît souvent, par exemple, *montrer (monter)* et *entrer (rentrer)*. L'apprentissage du lexique est donc de mémoriser le sens et la forme (Ellis, 1997 ; Cavalla, 2016). D'après Ellis (1997), l'apprentissage de la forme qui correspond à une acquisition implicite et inconsciente peut se réaliser par une activité de répétition, mais l'apprentissage du sens fait partie de l'acquisition consciente et explicite, car l'apprenant doit faire des liens entre les connaissances acquises et les savoirs anciens (Ausubel, 1968 ; Ellis, 1996 ; Raynal & Rieunier, 1997 ; Cavalla, 2016). En tenant compte des deux biais du lexique (le sens et la forme), nous nous interrogeons sur la question de l'entrée de l'enseignement du lexique. Autrement dit, nous envisageons l'enseignement allant du sens à la forme ou de la forme au sens. Ces deux entrées de l'enseignement du lexique sont appelées de façon didactique : approche onomasiologique et approche sémasiologique. L'utilisation des deux approches peut s'employer dans toutes les méthodes d'enseignement. D'ailleurs, les deux approches sont tout à fait propices à la mémorisation du lexique (Lieury, 1991). À l'origine, les deux approches sont « entrées méthodologiques développées par des linguistes allemands pour l'analyse lexicologique et sémantique du lexique des langues » (Cavalla, 2016 : 54). L'approche onomasiologique va du sens à la forme et prend pour point de départ une notion (un ensemble de concepts), c'est-à-dire que l'approche onomasiologique regroupe dès le départ « les synonymies ou les expressions sémantiquement proches » (Picoche, 1992 ; Tran, 2014). Les expressions sémantiquement proches ne sont pas forcément constituées d'un même mot clé. Prenons l'exemple de Cavalla (2009), l'expression *avoir le cafard* ne contient pas le mot-clé *triste*, mais elle participe du même concept que la tristesse. Comme les lexies peuvent être connectées en se basant sur le sens dans le lexique mental, l'idée de l'application de l'approche onomasiologique dans l'enseignement est d'aider les apprenants à construire leurs réseaux

sémantiques : l'approche onomasiologique « permet de cerner le lexique, permet de rester dans un réseau lexico-sémantique facilitateur d'apprentissage » (Cavalla, 2016 : 64). Par ailleurs, l'approche onomasiologique est également préconisée par de grands didacticiens comme Bronckart, Verlinde, Binon, Nesselhauf, Hill, Hilton, Firth, Rastier, Lewis.... À l'inverse de l'approche onomasiologique, l'approche sémasiologique prend source dans les grammaires de constructions qui ont, en tant que courant de la linguistique cognitive, « pour but de décrire et d'expliquer les structures et les principes linguistiques mentaux » (González-Rey, 2010 : 9). Cette approche est également recommandée par certains linguistes (entre autres, François, 2008 ; Legallois, 2005) et des didacticiens (entre autres, Galisson, 1984a ; Bárdosi, 1986 ; Nation, 1990 ; 2001 ; González-Rey, 2002 ; 2010). Par exemple, dans l'ouvrage *Les expressions imagées* (cf. 3.2.4 supra), Galisson procède à l'enseignement de la phraséologie par l'approche sémasiologique qui se révèle dans les conseils d'utilisation de cet ouvrage, comme le montre la figure 3-5 ci-dessous.

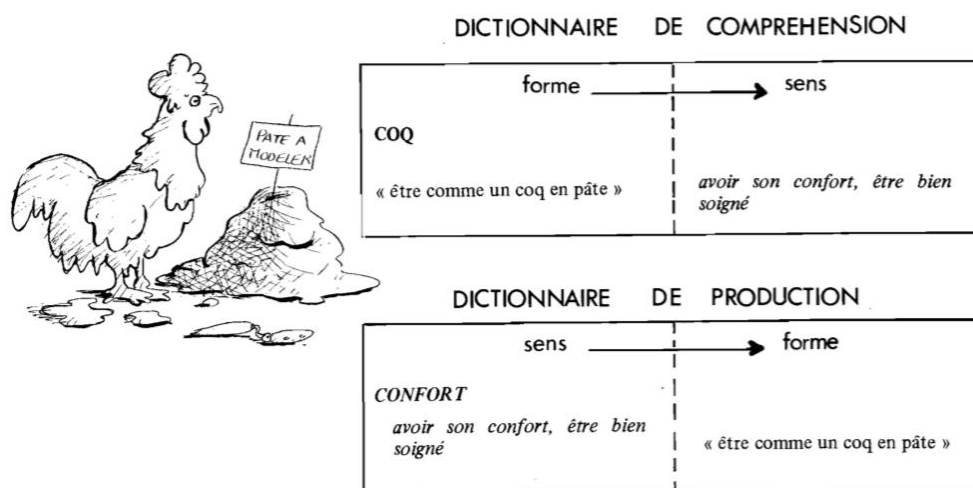


Figure 3-5 : Conseils d'utilisation de l'ouvrage *les expressions imagées*

Les approches sémasiologique et onomasiologique sont opposées et l'approche sémasiologique est peu mobilisée pour l'enseignement du lexique dans les manuels de FLE (Cavalla, 2016). En fait, les deux approches peuvent

être utilisées de façon complémentaire : « les deux entrées sont pertinentes dans l'enseignement du lexique car elles sont complémentaires : des listes de lexies (sémasiologie) liées par le sens dans un champ spécifique (onomasiologie) » (Cavalla, 2009 : 124). De ce fait, les deux entrées seront privilégiées pour notre enseignement : nous aborderons, tout d'abord, un champ spécifique (onomasiologie) ; ensuite, l'approche sémasiologique sera adoptée pour enseigner la liste de mots liés à ce champ.

L'E/A des phrasèmes culinaires s'avère nécessaire pour les apprenants étrangers. Malheureusement, il n'existe pas d'ouvrage consacré spécifiquement à cet E/A. À travers le lexique mental, nous apprenons que l'enseignement au genre aide à stocker les phrasèmes dans le cerveau. Le corpus est une base de données constituée en fonction du choix des genres de textes et peut aider à préciser et mémoriser le sens des phrasèmes. Une telle base de données permet également de sélectionner des phrasèmes *NAdj* à enseigner. De surcroît, l'utilisation des corpus apporte des avantages souhaités pour l'E/A des langues.

## 3.2 Utilisation des corpus pour l'E/A des langues

### 3.2.1 Corpus pour l'E/A des langues

Selon Tognini-Bonelli (1993), il existe trois types de corpus en didactique des langues : corpus-based, data-based et corpus-driven. L'approche corpus-based s'emploie pour confirmer l'hypothèse du linguiste, comme le dit Tognini-Bonelli (1993 : 1) comme un « (...) repository of examples to back pre-defined theories<sup>59</sup> ». L'approche data-based est utilisée par les étudiants pour émettre des hypothèses dans le cadre de l'apprentissage des langues. L'approche corpus-driven est adoptée pour découvrir des phénomènes linguistiques et confirmer des hypothèses au sujet de phénomènes

---

<sup>59</sup> (...) Référentiel d'exemples pour confirmer les théories prédéfinies.

linguistiques, comme l'explique Pearson (1998 : 49) : cette approche « is being increasingly adopted by researchers as a basis for refining linguistic theories<sup>60</sup> ». L'intérêt de l'utilisation du corpus pour l'enseignement a été abordé en introduction. Certains spécialistes pensent que l'intégration du corpus en didactique des langues provoque même un changement révolutionnaire et radical (Johns, 1988 ; 1990 ; McCarthy, 2008 ; Römer, 2011). L'intégration du corpus n'est pourtant pas assez prise en compte dans l'E/A des langues en France, d'après certains didacticiens (entre autres, Debaisieux, 2009 ; Boulton, 2008a ; 2010 ; Boulton & Tyne, 2014 ; Chambers, 2007 ; Römer, 2006). Comme le souligne Ciekanski, « un écart entre le consensus théorique qui se dégage des recherches menées sur l'utilisation effective des corpus et leur intégration dans les pratiques enseignantes en contexte d'apprentissage » (2014 : 112). D'après Chambers, les plus importants obstacles qui retardent cette intégration en France sont « la disponibilité des corpus oraux et le temps qu'il faut y consacrer... » (2009 : 28). Boulton (2008b) pense que la barrière culturelle ne doit pas être sous-estimée dans cette intégration tardive et que le fait que l'apport de l'utilisation du corpus pour l'E/A n'est pas confirmé influence également cette intégration. Kübler (2014) souligne qu'une compétence supplémentaire demandée à l'utilisation du corpus chez les apprenants et les enseignants peut ralentir cette intégration. Certes, l'exposition seule à une base de donnée ne garantit pas l'acquisition d'une langue et l'utilisation du corpus pour l'E/A des langues est censée être connue à la fois par les enseignants et les apprenants.

### 3.2.2 Trois manières d'enseigner l'utilisation des corpus

D'après Fligelstone (1993), l'utilisation des corpus peut être enseignée de l'une des trois manières suivantes : *teaching about*, *teaching to exploit*, *exploiting to teach*, comme reprié dans la figure 3-6 ci-dessous :

---

<sup>60</sup> Cette approche est de plus en plus adoptée par les chercheurs comme base pour les théories linguistiques affinées.



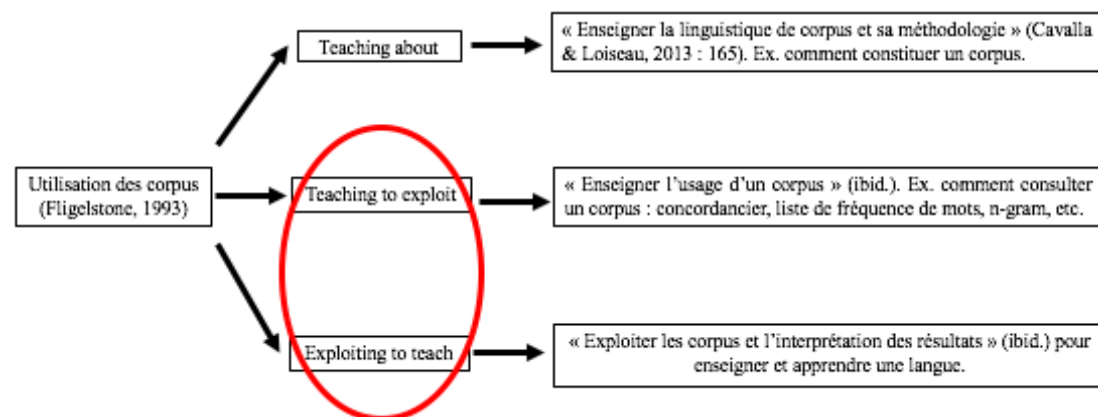


Figure 3-6 : Trois points de vue didactiques pour l'utilisation du corpus

« *Teaching about* » : dans cet usage, le corpus est considéré « comme objet d'enseignement » (Cavalla & Loiseau, 2013 : 165). Les contenus de cet enseignement sont organisés autour de la linguistique de corpus et de sa méthodologie. Comme précédemment évoqué la linguistique de corpus est souvent conçue pour une branche du TAL, cette manière d'enseigner l'utilisation du corpus se voit souvent dans les formations TAL. Les cours concernés peuvent être, par exemple, l'extraction d'information, la programmation algorithmique spécifique, etc.

« *Teaching to exploit* » : l'esprit de cette perspective est que l'apprenant a tout intérêt à « s'approprier l'outil d'exploitation du corpus afin de le manipuler à sa guise » (ibid., p. 167). Cet usage du corpus concerne la consultation et l'interrogation du corpus : quelle requête peut se faire à des fins d'apprentissage des langues ? Par exemple, une requête représentative *liste de fréquence de mot* propose une liste de mots d'un corpus classés par ordre décroissant dans laquelle les mots grammaticaux d'un corpus sont souvent prédominants : l'article défini anglais *the* peut représenter presque un mot sur quinze (Boulton & Tyne, 2014). Grâce à une telle requête, les mots les plus fréquents peuvent être sélectionnés pour l'E/A (Yang & Cavalla, 2019).

« *Exploiting to teach* » : dans cet usage, le corpus est utilisé « comme support d'enseignement » (Cavalla & Loiseau, 2013 : 165). Cet usage consiste à «

exploiter les corpus et l'interprétation des résultats pour enseigner une langue » (ibid. :165).

Comme le font certains didacticiens (entre autres, Cavalla *et al.*, 2014), l'optique dans laquelle notre étude s'inscrit concerne également les deux derniers usages (cerclés en rouge dans la figure 3-6) : en vue d'apprendre une langue, l'apprenant a tout intérêt à passer par l'approche *teaching to exploit* pour consulter et interroger un corpus ; afin d'exploiter le corpus et d'interpréter des résultats pour l'enseignement, l'approche *exploiting to teach* doit être choisie : ex. comme support d'enseignement, par quelle approche le corpus est présenté aux apprenants par l'enseignant ?

### 3.2.3 Approches d'utilisation du corpus pour l'E/A des langues

#### 3.2.3.1 Approche *data-driven learning* (Johns, 1991a ; 1991b)

L'E/A des langues sur corpus s'est principalement propagé sous l'impulsion de l'approche *data-driven learning* (désormais DDL) de Tim Johns (1991a ; 1991b) et c'est pour cette raison qu'il est qualifié de « père fondateur des applications pédagogiques de la linguistique de corpus » (Boulton & Tyne, 2014 : 75). L'esprit de l'approche DDL est que les apprenants peuvent formaliser eux-mêmes leurs savoirs de façon inductive au lieu d'apprendre les règles transmises par l'enseignant (Landure & Boulton, 2010). L'approche inductive « décrit un type de raisonnement qui va du particulier au général, des données particulières à des propositions plus générales » (Tran, 2014). Cette approche prend sa source dans le constructivisme de Vygotsky (1896-1934) dont l'esprit est que les étudiants peuvent créer de nouvelles règles en faisant des exercices ou répétitions. Dans cette approche, l'enseignant n'est plus le seul décideur de l'E/A, surtout de l'enseignement pour l'enseignant. Comme le dit Johns, « la tentative d'éliminer autant que faire se peut l'intermédiaire pour donner à l'apprenant un accès direct aux données » (1991b : 30). L'élimination de certaines médiations (enseignant ou manuel de cours) permet, dans la mesure du possible, aux apprenants de déduire eux-mêmes les régularités

---

sous-jacentes de la langue par une série d'activités : l'observation, la classification et la génération. D'après certains didacticiens (entre autres, Cobb, 1999 ; O'Sullivan, 2007 ; Tyne, 2009, Chambers, 2010), l'apprentissage par l'approche inductive est un processus considéré comme « naturel » pour les apprenants, car il améliore la motivation, la mémorisation et l'autonomisation (Landure & Boulton, 2010 ; Boulton, 2010). En effet, l'approche inductive est privilégiée dans l'utilisation du corpus pour l'enseignement, comme le rappelle Tran : l'approche inductive est l'« un des principes sous-jacents du DDL » (2014 : 218). D'après Cavalla (2009) et Tran (2014), une telle approche se marie parfaitement avec l'approche onomasiologique car cette dernière permet également de relever des concepts (ou conceptualiser les connaissances), comme expliqué dans la sous-partie 3.1.6. À titre d'exemple, Cavalla (2016) recense les expressions à mots multiples concernant le corpus humain (le dos, le bras...) et les expressions les plus fréquentes dans *Lexicoscope*<sup>61</sup> et *Corpus Français de Leipzig* (ex. *tourner le dos à*, *être le dos au mur*, *prendre à bras le corps* et *être le bras droit de*) expriment un sens autre que celui d'un élément du corps humain.

### 3.2.3.2 Apprentissage sur corpus (Boulton & Tyne, 2014)

L'Apprentissage Sur Corpus (désormais ASC) se fait principalement par l'approche inductive, mais l'approche déductive peut également être proposée (entre autres, Boulton & Tyne, 2014 ; Cresswell, 2007). Comme le disent Boulton & Tyne, « le travail sur corpus peut tout aussi bien se faire de façon déductive, ce qui permettrait plus facilement la réutilisation de matériels généralisables » (2014 : 166). Dans une approche déductive, un corpus peut être exploité pour illustrer des règles données ou valider une théorie existante à l'aide d'exemples concrets (Aston, 1997 ; Boulton & Tyne, 2014).

---

<sup>61</sup> Lexicoscope, corpus journalistique français en ligne : <http://phraseotext.univ-grenoble-alpes.fr/lexicoscope/> (consulté le 10/10/2019).

---

Contrairement à l'approche inductive, l'approche déductive va du général au particulier, des règles aux exemples et se révèle être le style approprié d'enseignement humain (Rézeau, 1997 ; 2001 ; Felder & Silverman, 1988). C'est pourquoi ce paradigme est celui qui est le plus proposé dans l'enseignement. Dans une telle approche, l'enseignant explicite les règles des fonctions grammaticales de la langue et les illustre « à l'aide d'un exemple choisi pour son caractère typique » (Colletta, 2004 : 311). Cette approche est censée être la bonne pour présenter des listes de mots (l'entrée sémasiologique, cf. 3.6.1 supra) aux apprenants. Hormis l'application des approches inductive et déductive pour l'E/A sur corpus, nous pouvons également procéder à la consultation directe et indirecte.

### 3.2.3.3 Consultation directe et indirecte (Chambers, 2010)

Deux types de consultations du corpus ont été introduits par Chambers (2010) : la consultation indirecte et la consultation directe. Dans la consultation indirecte, les apprenants consultent les concordances créées par l'enseignant et peuvent être un « utilisateur novice des corpus » (Kübler, 2014). Cet apprentissage peut donc rendre possible un « apprentissage multicontextuel » (Cobb, 1997b : 303) (sans l'obligation de l'ordinateur et des logiciels hypertextuels). Pour la consultation directe, les apprenants se trouvent devant un corpus et ils sont censés être en mesure d'interroger eux-mêmes un corpus. D'après Cavalla et Loiseau (2013), avant de faire la consultation directe, les apprenants ont besoin d'une formation d'initiation dans laquelle il s'agit de respecter trois principes : exercices d'entraînement, corpus de petite taille et prise en main d'un outil d'interrogation du corpus. D'après McEnery & Wilson (1997), ce premier principe encourage un apprentissage constructiviste, car des exercices d'entraînement permettent aux apprenant de tendre à assurer un apprentissage semi-autonome sur corpus. Afin de découvrir l'utilisation du corpus pour l'apprentissage de l'anglais, le site<sup>62</sup> de Tim Johns propose des

---

<sup>62</sup> Adresse du site de Tim Johns : <https://lexically.net/TimJohns/> (consulté le 10/10/2019).

exercices d'entraînement aux apprenants. En ce qui concerne l'apprentissage du français sur corpus, les projets ALFALEX<sup>63</sup> (Selva, Verlinde & Binon, 2004) et FULS<sup>64</sup> (Cavalla & Loiseau, 2013) s'inscrivent également bien dans cette découverte. Pour le deuxième principe, Granger (2001), Chambers (2005) et Boulton (2008a) conseillent déjà aux enseignants de placer, dans un premier temps, les apprenants devant un petit corpus pour ne pas les noyer dans une immense base de données. Le troisième principe, la prise en main d'un outil d'interrogation du corpus, est indispensable pour faire la consultation directe, car avec l'aide d'un outil d'interrogation du corpus, les apprenants peuvent utiliser le corpus à bon escient pour leur apprentissage.

### 3.3 Outils didactiques pour l'E/A de la phraséologie

#### 3.3.1 Concordancier

##### 3.3.1.1 Présentation de concordancier

Les concordanciers sont des logiciels qui peuvent produire des concordances (Boulton & Tyne, 2014). Le terme « concordance » emprunté à l'anglais signifie une liste de mots accentués dans leurs contextes et cette conception de présentation de mot a été créée par Hans Peter Luhn à la fin des années 50. Le concordancier est un outil d'analyse pouvant tirer profit de l'ordinateur. Comme l'explique Levy, « it makes full use of the innate qualities of the computer, its storage capacity, its speed and its accuracy<sup>65</sup> » (1997: 195). Les concordances permettent de faire l'analyse sémantique, lexicale et syntaxique en tenant compte des contextes du mot-clé. L'idée de la prise en compte du contexte dans la détermination du « sens d'un vocable ou d'une lexie »

---

<sup>63</sup> Adresse de la présentation du projet ALFALEX (Actieve Leeromgeving Frans voor Anderstaligen Lexicon, environnement d'apprentissage actif de français pour allophones Lexique) : <https://edutice.archives-ouvertes.fr/edutice-00000735/document> (consulté le 10/10/2019).

<sup>64</sup> Adresse du projet FULS (Formes et Usages des Lexiques spécialisés) : <http://webtek-66.iut2.upmf-grenoble.fr> (consulté le 10/10/2019).

<sup>65</sup> Il utilise pleinement les qualités innées de l'ordinateur, sa capacité de stockage, sa vitesse et son exactitude (notre traduction).

(Mel'čuk, 1999 : 13) est déjà apparue dans l'approche contextualiste anglaise (Williams, 2003) proposée par Firth en 1935. C'est là où naîtra la célèbre phrase : « You should know a word by the company it keeps<sup>66</sup> » (Firth, 1957 : 11). Pincemin (2006) confirme également que, « l'interprétation se nourrit d'informations locales ». D'après elle, les concordances se caractérisent par trois paramètres : pivot (mot-clé), contextes (droit et gauche), tri (d'un critère pour analyse). Dans la communauté scientifique internationale et donc dans les outils disponibles, le seul format des concordances connu et utilisé est le « KWIC » (Key Word In Context), pour désigner toutes sortes de présentations de mots-clés en contexte, que le mot-clé soit centré et aligné ou non. Quand un logiciel propose plusieurs modes d'affichage des mots en contexte, alors il peut utiliser « KWIC » pour l'un et par exemple « context view » pour l'autre. Mais Pincemin (2005) montre que les différentes formes de présentation de mot-clé en contexte ne sont pas équivalentes. Afin de pouvoir énumérer les différentes formes de mots-clés en contexte, elle utilise les quatre désignations pour chacune : KWAC (Key Word And Context), KWIC (Key Word In Context), KWUT (Key Word Up to Text), KWOC (Key Word Out of Context) affichées dans la figure 3-7 ci-dessous.

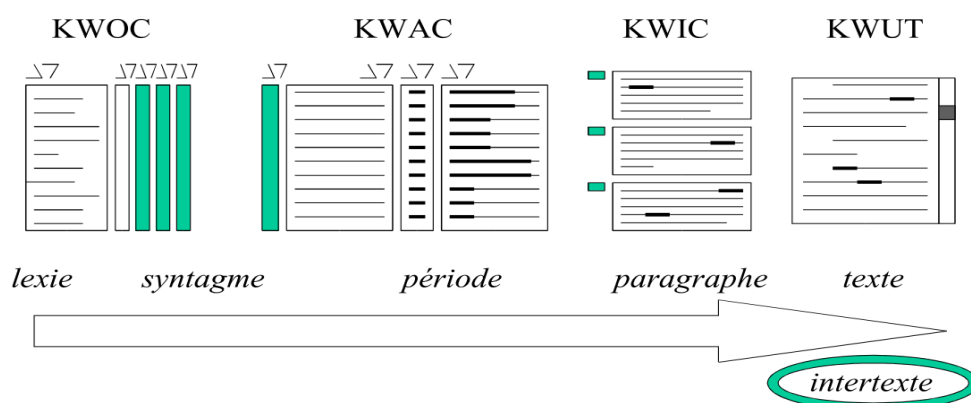


Figure 3-7 : Allure des KWOC-KWAC-KWIC-KWUT (Pincemin, 2006)

<sup>66</sup> « Vous devrez connaître un mot tout en tenant compte de ses compagnies » (notre traduction).

Le KWOC représente seulement une liste de mots (index) sans contexte. Les mots clés dans les trois formes (KWAC, KWIC et KWUT) sont mis en évidence dans les contextes dont la taille est de la plus petite (KWAC) à la plus grande (KWUT) : des syntaxes à périodique pour le KWAC ; des blocs textuels (ex : paragraphe) pour le KWIC ; le texte entier pour le KWUT. L'atout de concordance est la forme KWAC dans laquelle toutes les occurrences du mot-clé sont alignées verticalement en colonne et entourées horizontalement des contextes droits et gauches (Pincemin, 2006). Prenons comme exemple le mot-clé HACHÉ<sup>67</sup> au format KWAC affiché dans TXM sur la figure 3-8 ci-dessous. Le mot-clé HACHÉ, le pivot (1), se trouve dans une colonne au centre et est entouré par les contextes gauches et droits (2). Un tri (3) peut se faire sur le mot-clé ou sur les contextes afin d'arriver à un arrangement plus clair en fonction de l'objet de recherche.

Contexte gauche	Pivot	Contexte droit
aux pistaches, faite de 30g de pistaches coupées en dés et les cerneaux de noix et la levure, puis les noix grossièrement fraîche - 1 cuillère à soupe de persil	hachées	, de 20 g de sucre et de 10 cl de crème (la pâte est alors très épaisse et dure à mélanger)
sope de beurre. Ajouter un petit oignon saupoudrer d'une cuillère à soupe de persil	hachés	. Versez la préparation dans un moule rectangulaire (20 X 28
cuillères à soupe de basilic et de ciboulette , les pommes de terre, l'échalote	haché	- sel - poivre Faire fondre dans une casserole 3 cuillères à et les champignons de Paris coupés en tout petits morceaux. Saler
2 grosses pommes de terre - 2 oignons	haché	. Remuer bien les champignons dans le beurre chaud. Couvrir et
2 oignons hachés - 3 gousses d'ail	hachées	- sel et poivre Laver les courgettes et couper les extrémités.
. Faire revenir les oignons et l'ail	hachés	grossièrement et le thym. Laisser le cresson ramollir et ajouter l'
Y étaler la moutarde. Rajouter le cantal	hachée	- 3 gousses d'ail hachées - 10 cl de crème fraîche
frais - 1 cuillère à soupe de persil	hachés	- 10 cl de crème fraîche - 1 cuillère à café de
olive 20 à 25 mn avec l'ail	hachés	dans un peu de beurre à feu doux. Ajouter les pommes
naturel - 1 cuillère à soupe de basilic	haché	ou coupé en fines tranches, égaliser. Couper les tomates en
350 g chacun) - 500 g porc	haché	- 100 g de crevettes - 20 cl de crème liquide -
et le mélanger à la chair de porc	hachée	, le sel, le poivre et le basilic. D'autre
1 carotte - 600 g viande de bœuf	haché	- 2 cuillères à soupe de crème fraîche - 1 cuillère à
- 2 tomates - 400 g de viande	hachée	(chair à saucisse) - 1 cuillère à soupe farine -
de la recette : Faire revenir les oignons	hachés	. Mélanger. Ajouter l'œuf, la farine, le mélange
les tomates coupées en dés, la viande	hachée	- 800 g purée de tomates - 1 verre d'eau -
pommes de terre - 500 g de bœuf	haché	- farine - sel, poivre - herbes de provence - 1
Dans un fait-tout, faire revenir la viande	hachée	et les gousses d'ail mixées avec un bon morceau de beurre
terre, étaler la moitié de la viande	hachée	, 1 cuillère à soupe de farine, sel, poivre,
. Sur une moitié, étaler le jambon	haché	- 5 tomates (+ / - une demie brique de coulis
le faitout rincé. Ajoutez oignon et poireaux	hachés	au beurre à feu assez fort, saler, poivrer et retirer
		avec la tomate, puis la moitié des aubergines, puis l'
		, puis les champignons, le fromage, et enfin les œufs
		, carottes, échalotes, ail, céleri et bouquet garni.

Figure 3-8 : Concordances du mot clé HACHÉ dans TXM

D'après Pincemin, « cette présentation s'avère un outil heuristique extrêmement efficace pour la lecture des contextes proches et l'observation

<sup>67</sup> En lexicologie, les lexèmes sont toujours mentionnés en petite majuscule.

globale des constructions dans lesquelles s'insère le mot étudié » (2006 : 19). Grâce à cette présentation, les régularités ou les particularités d'emplois du mot-clé au niveau sémantique, lexical et syntaxique peuvent être facilement interprétées en comparant les contextes par superposition et répétition. À titre d'exemple, cette présentation peut être utilisée pour révéler la collocation (Fligestone, 1993 ; Gledhill, 2011 ; Habert *et al.*, 1997 ; Pecman, 2004 ; Polguère, 2003 ; Hoey, 2005). Comme le soulignent Cavalla *et al.*, grâce au concordancier, « l'utilisation des corpus est un atout pour l'enseignement des collocations car l'extraction sous forme de concordances permet à l'enseignant de repérer rapidement les formes figées associées à une lexie » (2014 : 334). Des concordanciers en ligne restent à la disposition de toutes les personnes intéressées, comme *Sketch Engine* (Kilgarriff *et al.*, 2004), *Lextutor* (Gaskell & Cobb, 2004), *Itrameur*<sup>68</sup> (Fleury, 2017), *Hyperbase web*<sup>69</sup> (Vanni & Mittmann, 2016). Les talistes inventent également les concordanciers insérables dans les logiciels textométriques (cf. 1.2 supra). À travers les concordanciers disponibles, les linguistes, les didacticiens et les apprenants peuvent réaliser les interrogations dans leurs propres corpus à des fins d'analyse linguistique, d'enseignement ou d'auto-apprentissage. En outre, le concordancier est un outil facile à maîtriser, comme le soulignent Boulton et Tyne, « avec un concordancier, même le novice peut très rapidement faire des recherches assez sophistiquées qui répondent à des questions réelles » (2014 : 119).

### 3.3.1.2 Outil représentatif du TAL pour l'enseignement (Yang, 2017a)

Comme nous l'avons mentionné plus haut (cf. 3.2.2 supra), l'utilisation du corpus *teaching to exploit* n'est pas assez prise en compte dans l'E/A des langues. L'une des raisons de ce retard est liée à la méfiance de l'outil informatique (Chambers *et al.*, 2004 ; Chambers *et al.*, 2011 ; Salaberry, 2001).

---

<sup>68</sup> Adresse du site Itrameur : <http://www.tal.univ-paris3.fr/trameur/iTrameur/> (consulté le 10/10/2019).

<sup>69</sup> Adresse du site Hyperbase : <http://hyperbase.unice.fr> (consulté le 10/10/2019)..



---

Parmi les outils informatiques du TAL, le concordancier est le plus couramment préconisé par des didacticiens français (Rézeau, 2001 ; 2002 ; 2007 ; Boulton, 2008a ; Chambers, 2011 *et al.* ; Cavalla *et al.*, 2014). À l'instar de Cavalla *et al.*, « le concordancier est probablement l'outil le plus représenté parmi les dispositifs pédagogiques visant à intégrer la notion de corpus en classe de langue » (2014 : 335). En effet, l'utilisation de concordancier dans l'apprentissage des langues devient en vogue dans les pays anglo-saxon depuis les années 90 (Tribble & Jones, 1990 ; Flowerdew, 1993 ; 1996 ; McEnery & Wilson, 1997 ; Sinclair, 2004 ; Gaskell & Cobb, 2004 ; Gavioli, 2005) et Johns (1986 ; 1988a ; 1988b) est le premier promoteur du concordancier dans l'apprentissage des langues (Rézeau, 1997 ; 2001). Même si l'utilisation inappropriée de concordancier peut engendrer un travail « laborieux » pour les apprenants (Yoon & Hirvela, 2004 : 274 ; Chambers, 2005 : 120) et donner des contextes réduits pour l'illustration et l'interprétation des résultats de recherche (Widdowson, 2000), le concordancier peut offrir les concordances « crédibles, utilisables et transférables » (Rézeau, 2001 : 277) pouvant être utilisées comme un type de document semi-didactique entre le corpus et le document didactisé :

*By concentrating and making it easy to compare the contexts within which a particular item occurs, (a concordance) organises data in a way that encourages and facilitates inference and generalization<sup>70</sup> (Johns, 1986 : 159).*

D'après certains didacticiens, l'analyse des concordances peut améliorer la compétence des apprenants en écriture (Gaskell & Cobb, 2004 ; O' Sullivan & Chambers, 2006) et leur donner également confiance pour leur écriture (Yoon & Hirvela, 2004). Après le concordancier, nous allons proposer un autre outil heuristique pour l'E/A du lexique : carte mentale.

---

<sup>70</sup> En concentrant et en facilitant la comparaison des contextes dans lesquels un élément particulier se produit, (une concordance) organise les données de manière à encourager et à faciliter l'inférence et la généralisation (notre traduction).

### 3.3.2 Carte mentale

#### 3.3.2.1 Présentation de la carte mentale

Développée par Tony Busan dans les années 70, la carte mentale est largement utilisée dans l'enseignement en Finlande et depuis 2010, elle est préconisée par l'éducation nationale française pour l'enseignement d'exploration : « l'utilisation d'un logiciel de création de « cartes mentales » facilite la structuration de leurs (les élèves) réflexions et la présentation collective (diaporama, note de synthèse, affiche, compte-rendu de projet, etc.) » (Bulletin officiel de l'éducation nationale (n°4, 29 avril 2010)). Mais malheureusement, elle est beaucoup moins développée dans l'enseignement en France que dans d'autres pays européens (Cavalla, 2016 : 98). La carte mentale est un schéma de type arborescent utilisé pour organiser des connaissances (Cavalla *et al.*, 2014). Rappelons que dans l'approche psycholinguistique, le lexique mental contient le lexique et son organisation, donc dans une certaine mesure, l'idée de la carte mentale simule la connexion et l'association du lexique mental. La construction de la carte mentale lexicale peut se dérouler en deux temps, comme suit :

- D'abord, introduire un mot-clé (il n'y a pas forcément de centre dans une carte mentale),
- Ensuite, ajouter des éléments nouveaux à l'extrémité de branches ou de flèches.

Plusieurs logiciels gratuits aident à réaliser une carte mentale, par exemple, *CmapTools*<sup>71</sup>, *FreeMind*<sup>72</sup>, *XMind*<sup>73</sup>, mais elles peuvent également être réalisées sur papier.

---

<sup>71</sup> Logiciel à télécharger en ligne gratuitement à l'adresse suivante : <http://cmaptools.fr.malavida.com> (consulté le 10/10/2019).

<sup>72</sup> Idem : <https://framasoftware.org/article3854.html> (consulté le 10/10/2019).

<sup>73</sup> Idem : <https://www.xmind.net/fr/> (consulté le 10/10/2019).

En prenant ici pour exemple le mot clé *cuisson*, une carte mentale (construite par le logiciel *CmapTools*) au sujet du mode de cuisson (à l'eau, à la vapeur, grillées, sautés) peut être construit dans la figure 3-9 ci-dessous.

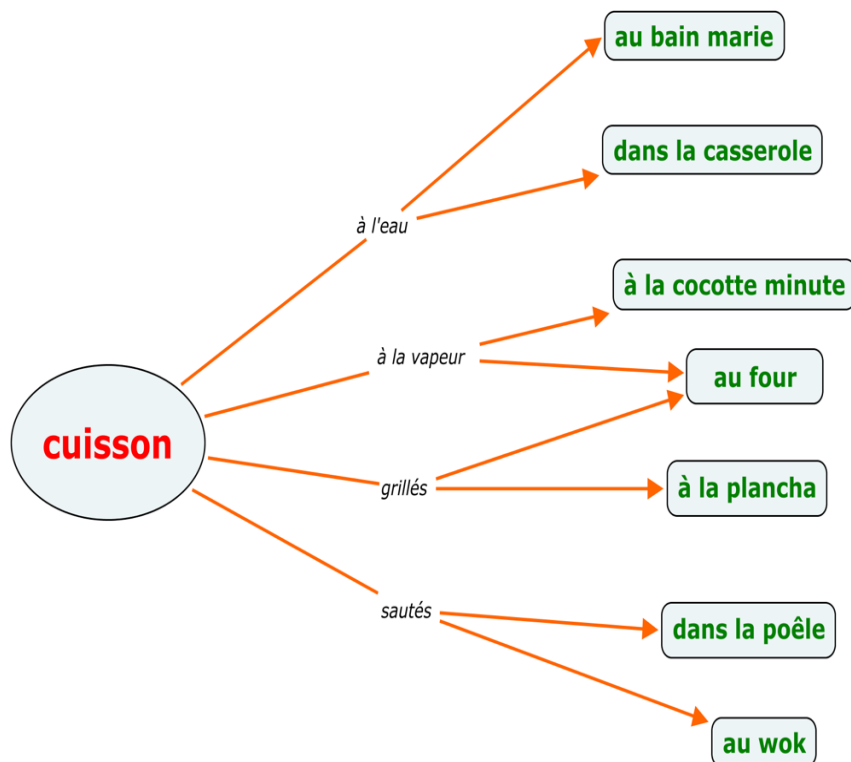


Figure 3-9 : Carte mentale : Mode de cuisson

Nous voyons bien que, dans la figure 3-9 ci-dessus, la carte mentale commence par le mot-clé *cuisson*, que des éléments nouveaux (à l'eau, à la vapeur, grillés, sautés) sont ajoutés à l'extrémité de branches et que le sens des flèches indique également l'ordre syntaxique. En effet, le sujet d'étude détermine la nature des mots à ajouter et en théorie, la carte mentale peut s'étendre à l'infini. La raison de proposer la carte mentale pour l'E/A des langues est qu'elle correspond assez bien au fonctionnement cérébral, comme le précisent *Mongin et al.*, la carte mentale est « en phase avec la manière dont le cerveau fonctionne » (2011 : 31) et la contribution de cette ressemblance n'est pas encore attestée, mais le fait que la fabrication des cartes mentales favorise l'apprentissage, notamment la mémorisation des langues, est déjà prouvé par plusieurs dimensions.

### 3.3.2.2 Outil d'aide à la mémorisation du lexique (Yang, 2017b)

Au moment de la construction de la carte mentale, l'apprenant devra se concentrer sur chaque élément pour les classer les uns par rapport aux autres, et ce fait de lister des éléments va renforcer ou créer des liens dans le lexique mental. En fait, depuis Goody (1979) et Boulton (2000) plus récemment, nous constatons qu'organiser des listes et se concentrer dessus sont deux activités qui aident à la mémorisation. Par ailleurs, le geste de construire la carte mentale aide également à la mémorisation du lexique, comme le soulignent Goody (1979) et Tellier (2006), le geste associé à la langue va aider à mieux mémoriser la langue. Avant que les apprenants commencent leurs constructions, l'enseignant leur donne des consignes à suivre et Cavalla (2016) constate que la précision des consignes pour la construction aide aussi à la mémorisation. De ce fait, le principe de l'utilisation de la carte mentale est l'aide à la mémorisation, dès lors que les apprenants réalisent eux-mêmes leur fabrication. Si la construction de la carte mentale favorise la mémorisation et aide-t-elle également à l'apprentissage ?

La création de cartes se fait toujours dans une situation ou un contexte<sup>74</sup> donné et la mise en contexte favorise sans aucun doute l'apprentissage (cf. 3.2.2 supra). L'apprentissage ne consiste pas seulement dans le fait d'apprendre de nouveaux éléments, mais aussi d'établir des liens entre les éléments nouveaux, de les organiser et de les connecter grâce aux acquis antérieurs (Ausubel, 1968 ; Ellis, 1996 ; Raynal & Rieunier, 1997 ; Cavalla, 2016). La carte mentale est alors utilisée pour établir ces liens en réseau : le savoir « for the learner they [concept maps] help to make evident the key concepts and propositions to be learned and also suggest linkages between the new knowledge and what he or she already knows »<sup>75</sup> (Novak et Gowin,

---

<sup>74</sup> Le terme contexte est à prendre au sens large et dans notre étude, le contexte est à prendre au sens du genre (Rastier, 2001)

<sup>75</sup> « Pour l'apprentissage, elles (les cartes mentales) aident à mettre en évidence les concepts clés et les propositions à apprendre et suggèrent aussi les liens entre ses nouvelles connaissances et ce qu'il ou elle sait déjà » (notre traduction).

1984 : 23). D'après Novak et Gowin, la carte mentale aide à conceptualiser et formaliser des connaissances afin de favoriser l'apprentissage en général. Le fait de construire la carte mentale aide à la mémorisation et « l'aide à la mémorisation reste au cœur de l'apprentissage et de l'enseignement d'une langue étrangère » (Cavalla, 2016 : 100), il est donc logique d'appliquer cet outil didactique dans l'E/A. En effet, l'utilisation de la carte mentale en classe de langue a déjà été prise en compte par certains didacticiens :

- Peters et *al.* (2005) considèrent la carte mentale comme « outil pédagogique puissant et versatile », quel que soit le niveau, le concept, l'âge ;
- Cavalla (2009 ; 2014) recourt à la carte mentale pour l'apprentissage du lexique auprès d'un public FLE ;
- Barna (2009) l'utilise également pour aider à la production écrite/orale ;
- Holme (2010a ; 2010b) s'en sert pour les grammaires de constructions ;
- Nous l'apprécions beaucoup pour l'enseignement du lexique culinaire (Yang & Cavalla, 2019), etc.

Il existe des sites<sup>76</sup> d'enseignants où des exemples et des explications pour la construction d'une carte mentale sont en libre accès.

## Bilan

L'enseignement de la phraséologie s'avère nécessaire en FOS. Toutefois, la phraséologie n'est pas traitée systématiquement ni dans les manuels, ni dans les ouvrages complémentaires du FLE ou du FOS. Ne pouvant pas recourir aux ouvrages de FLE, il y a lieu d'engager une recherche en psycholinguistique afin de mieux aborder la phraséologie. L'étude sur le lexique mental nous permet de tenir compte de deux éléments : d'un côté, le lexique dans le cerveau est principalement organisé sur le sens et l'E/A du

---

<sup>76</sup> Site de présentation de la formation à la carte mentale : <https://ifd-formation.org/formations/creer-et-exploiter-les-cartes-mentales-en-pedagogie/> (consulté le 10/10/2019).

lexique au genre permet de préciser le sens d'une unité lexicale : de ce fait, nous proposons d'enseigner et d'apprendre les phrasèmes au sein d'un genre ; de l'autre, les mots fréquents s'activent plus vite que les mots plus rares et l'apprentissage des mots fréquents est un besoin fondamental pour l'apprenant, ce qui nous facilite la sélection des phrasèmes pour l'enseignement. Quant aux entrées d'enseignement du lexique, les entrées onomasiologique et sémasiologique sont opposées et complémentaires. Dans notre étude, nous comptons adopter les deux entrées afin que l'une et l'autre se complètent. L'utilisation du corpus pour l'enseignement peut se faire par deux approches (inductive et déductive) ou par deux types de consultation (directe et indirecte).

Deuxième partie : méthodologie et analyse linguistique

<b>4. Constitution de <i>Cuisitext</i> et réflexion sur son exploitation .....</b>	<b>98</b>
<b>4.1 Constitution de corpus écrit.....</b>	<b>98</b>
4.1.1 Collecte de données .....	98
4.1.2 Nettoyage et exportation de données .....	101
<b>4.2 Construction de corpus oraux.....</b>	<b>102</b>
4.2.1 Support d'enregistrement.....	102
4.2.2 Cadre juridique pour le corpus .....	104
4.2.2.1 Domaine public ou droit d'auteur.....	104
4.2.2.2 Préparation d'un consentement .....	105
4.2.3 Convention de transcription .....	107
4.2.3.1 Choix de type de transcription.....	107
4.2.3.2 Choix du logiciel .....	109
4.2.4 Exemple d'extrait de transcription .....	111
<b>4.3 Réflexion sur l'exploitation de <i>Cuisitext</i> dans notre projet .....</b>	<b>112</b>
<b>Bilan .....</b>	<b>113</b>
<b>5. Repérage des phrasèmes saillants et leur sélection pour</b>	
<b>l'enseignement.....</b>	<b>115</b>
<b>5.1 Objet d'étude : phrasèmes <i>NAdj</i> .....</b>	<b>115</b>
<b>5.2 Extraction des phrasèmes <i>NAdj</i>.....</b>	<b>117</b>
5.2.1 Modélisation lexicale des données .....	118
5.2.1.1 Modélisation du syntagme nominal .....	119
5.2.1.2 Modélisation de la structure attributive.....	119
5.2.1.3 Exigences de notre modélisation des quatre grammaires .....	121
5.2.2 Choix du logiciel d'extraction .....	123
5.2.3 Implémentation des données lexicales dans <i>NooJ</i> .....	127
5.2.3.1 Implémentation du syntagme nominal.....	127
5.2.3.2 Implémentation de la structure attributive .....	128
<b>5.3 Identification et sélection des phrasèmes <i>NAdj</i> pour l'enseignement ..</b>	<b>132</b>
<b>Bilan .....</b>	<b>137</b>



<b>6. Analyse linguistique des phrasèmes <i>NAdj</i> retenus.....</b>	<b>138</b>
<b>6.1 Critères servant à la classification .....</b>	<b>138</b>
<b>6.2 Propriétés transformationnelles du figement des phrasèmes <i>NAdj</i>.....</b>	<b>139</b>
<b>6.3 Classification des phrasèmes <i>NAdj</i>.....</b>	<b>142</b>
<b>6.4 Analyse du figement des phrasèmes <i>NAdj</i>.....</b>	<b>143</b>
<b>Bilan .....</b>	<b>149</b>

Dans les études sur le domaine culinaire, les corpus explorés sont divers, par exemple,

- dans l'étude éthologique du végétarisme<sup>77</sup> en Suisse (Ossipow, 1997), le corpus contient des menus récoltés dans différents restaurants et des publicités et ouvrages concernant des aliments ;
- Régnier (2004) fait une étude sur l'exotisme culinaire et son corpus est constitué de 9758 recettes de cuisine parues dans des journaux de presse féminine des années 1930 à la fin des années 1990 ; en tant que base de données, dix mille recettes de cuisine provenant du site Marmiton ont été traitées dans le cadre de compétition scientifique DEFT<sup>78</sup> 2013 sur le thème culinaire.

En revanche, il n'existe pas de tel corpus qui plus est serait à la fois écrit et oral pour le domaine culinaire. L'efficacité de l'utilisation des corpus numériques n'étant plus à prouver pour l'enseignement des langues, nous avons eu l'intention, depuis le début de ce projet de thèse, d'élaborer notre propre corpus numérique baptisé *Cuisitext*. Notre ambition est de constituer un corpus qui puisse être utile pour plusieurs types d'études en linguistique et surtout dont on puisse extraire des exemples pour l'E/A du lexique culinaire en FOS. Dans cette partie méthodologique, nous allons essayer de répondre aux questions suivantes : comment construisons-nous un corpus écrit et oral ? Comment ressortons-nous l'élément le plus saillant du corpus ? Comment extrayons-nous exhaustivement toutes les séquences de l'élément étudié ? Et enfin, comment sélectionnons-nous les séquences extraites pour l'enseignement ?

---

<sup>77</sup> Un adresse d'un compte-rendu de ce livre : [https://www.persee.fr/doc/assr\\_0335-5985\\_1999\\_num\\_106\\_1\\_1212\\_t1\\_0085\\_0000\\_3](https://www.persee.fr/doc/assr_0335-5985_1999_num_106_1_1212_t1_0085_0000_3) (consulté le 10/10/2019).

<sup>78</sup> Le site du DEFT (Défi Fouille de Textes) : <http://defi.limsi.fr/2013/> (consulté le 10/10/2019).

## 4. Constitution de *Cuisitext* et réflexion sur son exploitation

### 4.1 Constitution de corpus écrit

La construction d'une base de données permet de se familiariser avec le domaine de spécialité (Yang, 2016). Comme le soulignent Mangiante et Parpette (2004 : 46), la collecte des données permet à l'enseignant de « sortir de son cadre habituel de travail pour entrer en contact avec un milieu qu'il ne connaît pas, a priori, et pour lequel il doit expliquer ses objectifs et le sens de sa démarche pour obtenir les informations dont il a besoin ». Comme expliqué dans la sous-partie 1.4.2, la recette de cuisine peut être choisie comme un genre de texte approprié à l'élaboration d'un corpus culinaire. Nous décidons de prendre des recettes de cuisine sur internet comme textes de référence du corpus écrit *Cuisitext*. Au vu de l'abondance des données, les copier/coller ne suffisent pas pour construire rapidement un tel corpus, donc dans notre étude, nous avons adopté le principe d'un crawler, c'est-à-dire une aspiration des contenus de site. *Gromoteur* est le logiciel de collecte de données écrites, que nous avons choisi car il est gratuit (*WebBootCat* est seulement gratuit pour seulement 30 jours) et ses fonctionnalités multiples permettent de sélectionner des recettes en évitant de tomber dans certains pièges évoqués ci-dessous. En outre, sa prise en main est facile et rapide.

#### 4.1.1 Collecte de données

L'utilisation de *Gromoteur* pour la collecte de données est la suivante : il faut inscrire, dans le formulaire de requête, des adresses web (URL : Uniform Resource Locator) ou des mots-clés et la collecte de données peut se dérouler de façon automatique. En France, il existe quelques sites culinaires connus sur lesquels de nombreuses recettes de cuisine sont proposées, nous en présentons quelques-uns dans le tableau 4-1 ci-dessous.

Noms des sites	URL
Marmiton	<a href="http://www.marmiton.org">http://www.marmiton.org</a>
750g	<a href="http://www.750g.com">http://www.750g.com</a>
Cuisine AZ	<a href="http://www.cuisineaz.com">http://www.cuisineaz.com</a>
Ôdélice	<a href="http://www.odelices.com">http://www.odelices.com</a>
Cuisine actuelle	<a href="http://www.cuisineactuelle.fr">http://www.cuisineactuelle.fr</a>

Tableau 4-1 : Quelques sites français de recettes de cuisine

Le tableau 4-1 ci-dessus rassemble les sites les plus populaires dans le domaine culinaire en France, par exemple, le site « Marmiton », qui existe depuis 2003, recense plus de 4,18 millions de visiteurs. Dans notre étude, nous pouvons collecter davantage de données (recettes de cuisine) en faisant des requêtes par les URL plutôt que par des mots-clés. De nombreuses données nous permettent d'avoir davantage de résultats intéressants en termes de fréquence et d'éléments lexicaux différents. Le principe d'aspiration des ressources d'un crawler est le suivant : la page web associée est téléchargée et les liens sur cette page sont également identifiés. Pour chaque nouveau lien, le crawler cherche encore les liens qu'il contient et ainsi de suite. Ce processus se répète indéfiniment. Les sites sélectionnés ne contiennent pas seulement des recettes (ils contiennent également des publicités, des vidéos, des zones de discussions), donc pour éviter cette accumulation de données inutiles pour notre recherche, nous devons trouver les liens qui peuvent aboutir uniquement aux recettes présentes sur ces sites, par exemple, l'URL (<http://www.marmiton.org>) est l'adresse du site Marmiton, mais l'adresse d'accès direct aux recettes de ce site est <http://www.marmiton.org/recettes/>. Une fois que les accès sont identifiés, nous pouvons alors démarrer l'aspiration des recettes par *Gromoteur*. Cette aspiration, sur différents sites, peut se faire simultanément. Dans la figure 4-1 ci-dessous, nous pouvons décider d'ajouter une ou plusieurs adresses URL en choisissant l'option A ;

nous pouvons également lancer, par l'option B, toutes les aspirations en ajoutant un fichier qui contient toutes les URL.

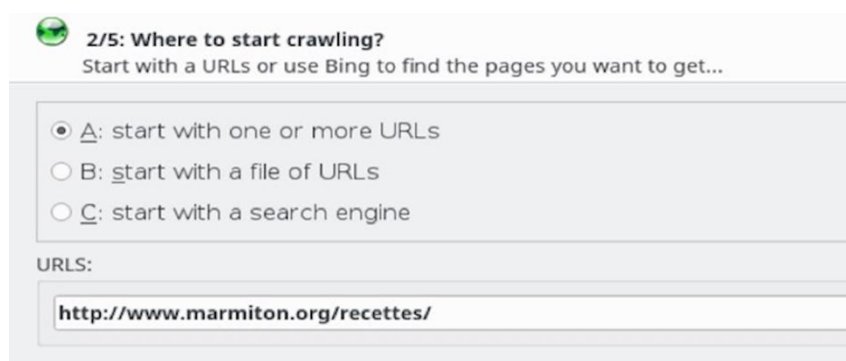


Figure 4-1 : Choix d'ajout d'URLs dans *Gromoteur*

En effet, même si nous avons déjà réduit le champ de requête en choisissant les sites d'aspirations, *Gromoteur* cherche encore tous les liens des sites donnés pour aller plus loin et cette aspiration peut durer plusieurs jours, voire plusieurs semaines. Pour ne pas tomber dans ce piège chronophage, nous devons indiquer à *Gromoteur* où s'arrêter (level from...to...) comme dans la figure 4-2 ci-dessous. Le « level from 0 to 0 » signifie que l'aspiration de données reste sur la même page et le « level from 0 to 1 » permet d'indiquer à l'aspirateur de recueillir des données sur la même page et suivre aussi les liens indiqués à partir de cette page. En choisissant le « level from 0 to 1 » dans notre étude, cette collecte des données se déroule seulement en quelques minutes.

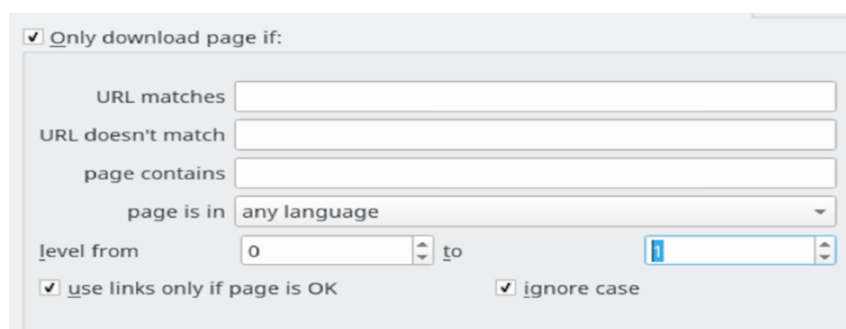


Figure 4-2 : Paramètre de requête du choix des pages dans *Gromoteur*

Dès que l'aspiration est finie, les données sont classées par URL dans un tableau (cf. la figure 4-3 ci-dessous). Le logiciel *Gromoteur* peut donner

l'accès aux données par différents paramètres : l'URL (adresse de chaque page), les titres des textes, la date de création du document, les textes de chaque page, etc.

url	time	linkOrigin_rowid	title_nbCharacters	title_nbSentences	title	text_nbCharacters	text_nbSentences	text
http://...	2017/4/13 18:27	1	118	3	...	9330	146	11 recettes ...
http://...	2017/4/13 18:27	1	96	3	...	7496	152	7 idées à b...
http://...	2017/4/13 18:23	1	79	2	...	4598	61	Accéder au ...
http://...	2017/4/13 18:23	1	79	2	...	4598	61	Accéder au ...
http://...	2017/4/13 18:24	1	72	2	...	6422	135	Bio Bien Bo...
http://...	2017/4/13 18:30	1	44	2	...	10269	260	Cake : nos ...
http://...	2017/4/13 18:24	1	43	2	...	8575	166	Chandeleur...
http://...	2017/4/13 18:24	1	71	2	...	8100	158	Chandeleur...
http://...	2017/4/13 18:31	1	52	2	...	10211	255	Chocolat : n...
http://...	2017/4/13 18:25	1	66	2	...	9455	184	Chocolat : r...
http://...	2017/4/13 18:27	1	111	3	...	7111	142	Comment u...
http://...	2017/4/13 18:21	1	112	2	...	9869	253	Communau...
http://...	2017/4/13 18:30	1	54	2	...	10277	260	Courgette : ...
http://...	2017/4/13 18:30	1	48	2	...	9779	255	Crepes : no...

Figure 4-3 : Extrait de données collectées dans *Gromoteur*

Nous pouvons visualiser chaque case de cette figure 4-3 dans une fenêtre indépendante en cliquant dessus. Après avoir fini cette aspiration de données, nous devons opérer un nettoyage des données pour que celles-ci soient propres donc lisibles et réutilisables.

#### 4.1.2 Nettoyage et exportation de données

Dans les textes recueillis, nous constatons qu'il n'y a pas de problème d'encodage, à savoir, d'éléments non reconnus par la machine. En revanche, il existe des URL non pertinentes qui ne contiennent pas de recettes et dans *Gromoteur*, nous pouvons filtrer (option « current filtrer » dans la figure 4-4 ci-dessous) les données recueillies pour en éliminer les éléments non pertinents. Par exemple, nous avons constaté que le mot-clé « photo » apparaît toujours dans les titres de ces éléments non pertinents, donc afin de supprimer ces éléments non pertinents, il suffit de filtrer les titres.

Database content

standard		
url	title	text
http://instagram.com/marmiton...	Marmiton	Marmiton (@mar...
http://www.marmiton.org/recet...	Photo	Photo de la recet...
http://www.marmiton.org/maga...		Diaporamiam : di...
http://www.marmiton.org/recet...	Photo	Photo de la recet...
http://www.marmiton.org/recet...	Photo	Photo de la recet...
http://www.marmiton.org/recet...	Photo	Photo de la recet...
http://www.marmiton.org/recet...	Photo	Photo de la recet...
http://www.marmiton.org/recet...	Photo	Photo de la recet...
http://www.marmiton.org/recet...	Photo	Photo de la recet...

database information   cell content   current filter   Clean

column	=	condition	remove
1 title	contains the word	photo	<input checked="" type="checkbox"/>
2 1	rowid		<input checked="" type="checkbox"/>

Figure 4-4 : Filtrage de corpus par un mot clé « photo »

Ce filtrage peut se faire au niveau des titres. Dans les deux premières colonnes en bas de la figure 4-4 ci-dessus, nous choisissons respectivement les paramètres « titre » et « contains the word ». Nous écrivons le mot-clé « photo » dans la troisième colonne (« condition »). En validant cette requête, nous pouvons facilement sortir et supprimer toutes les adresses URL non pertinentes. Une fois que toutes les données sont propres, leur exportation s'effectue au format TXT par défaut et notre corpus écrit peut être créé dans un fichier choisi. Nous allons construire notre corpus oral dans ce qui suit et le résultat de la construction de *Cuisitext* est exprimé à la fin de ce chapitre.

## 4.2 Construction de corpus oraux

### 4.2.1 Support d'enregistrement

Nos corpus oraux s'inscrivent dans les genres professionnels et en milieu professionnel, l'interdépendance du « langage au travail et du contexte d'action » (Boutet, 2005 : 26) doit être prise en considération dans la construction. Par conséquent, nos corpus oraux sont sous format vidéo. Nous

avons récupéré trois types de corpus oraux représentatifs du domaine culinaire et les résumons dans le tableau 4-2 ci-dessous.

Corpus oral	Nombre de vidéos	Mins/Vidéo	Année	Source
Clips vidéos culinaires sur internet	Cent	5	Après 2010	Chaines de télévision
Filmages dans une école hôtelière	Vingt	3-5	2016	Mangiante J.M. (Université d'Artois)
Filmages dans deux cuisines	Vingt	60	2015	Nous même

Tableau 4-2 : Trois types de corpus oraux dans notre étude

Pour ces trois types de corpus oraux, nous avons effectué différentes enquêtes. Les premiers corpus indiqués dans ce tableau sont des clips vidéo culinaires que nous avons sélectionnés sur internet en fonction de trois « familles » de plats français (entrées, plats, desserts) et cette classification des « familles » est bien respectée dans tous les restaurants classiques de France, comme expliqué sur le site « [economie.gouv.fr](http://economie.gouv.fr) »<sup>79</sup> :

*Nos repas français sont dictés par la tradition. Pour preuve, le repas gastronomique français tel qu'il a été inscrit au Patrimoine Mondial de l'Unesco se définit ainsi : « Le repas gastronomique doit respecter un schéma bien arrêté : il commence par un apéritif et se termine par un digestif, avec entre les deux au moins quatre plats, à savoir une entrée, du poisson et/ou de la viande avec des légumes, du fromage et un dessert. » Et d'ailleurs, tous les restaurants classiques de France reprennent en cœur ces mêmes familles dans leur carte : entrées, plats, desserts.*

Ces clips sont édités par divers sites et chaînes de télévision, comme *750g*, *Marmiton*, *Cuisine rapide*, *RTL\_TVI* (une chaîne belge), etc. Les deuxièmes vidéos sont des filmages dans une école hôtelière. Mangiante (2016),

<sup>79</sup> Adresse du site : <https://www.economie.gouv.fr> (consulté le 10/10/2019).



spécialiste du FOS, nous a offert vingt petits filmages réalisés dans le cadre de conventions de partenariat avec le lycée hôtelier Marguerite Yourcenar de Beuvry<sup>80</sup>. Nous avons également effectué une enquête, pour le troisième type de vidéos, dans deux cuisines de restaurant français, soit 20 heures d'enregistrement audiovisuel pendant le temps du service dans deux restaurants de Font Romeu fin décembre 2015. D'après Baude *et al.* (2006), nous avons effectué deux types de techniques d'enquêtes : les deux premiers types de vidéos (Clips vidéos sur internet et Filmages dans une école hôtelière) correspondent à une reprise d'enregistrements médiatiques et le troisième type de vidéo est un enregistrement d'activités professionnelles. Pour les exploitations et les diffusions des corpus oraux, les chercheurs ont des obligations ou des droits à respecter dans le cadre juridique selon des techniques d'enquêtes effectuées.

#### 4.2.2 Cadre juridique pour le corpus

##### 4.2.2.1 Domaine public ou droit d'auteur

Quel est le statut juridique de chaque corpus oral constitué ? Deux notions interviennent dans l'éclairage de cette question : le domaine public et le droit d'auteur. Le domaine public est un concept devenu courant avec l'utilisation de données personnelles en ligne via les réseaux sociaux par exemple. Selon Baude *et al.* (2006), étant donné que les informations dans le domaine public manquent de monopole, tout le monde peut y avoir accès, consulter et exploiter ces informations mais celles-ci ne peuvent faire l'objet d'une appropriation. Hormis les informations produites et tombées dans le domaine public, les autres œuvres sont protégées par le droit de la propriété intellectuelle (le droit d'auteur et les brevets). Comme dans l'article 113-1 du CPI (Code de la propriété intellectuelle), en principe, l'auteur est la personne physique sous le nom duquel l'œuvre est divulguée. Afin de déterminer le

---

<sup>80</sup> La commune de Beuvry se situe dans le département Pas-de-Calais en région Nord-Pas-de-Calais-Picardie.

statut d'un corpus, il est nécessaire de distinguer si les données recueillies font partie du domaine public et qui en est l'auteur, car pour exploiter une œuvre protégée par le droit d'auteur, il faut que le corpus soit soumis aux autorisations de ce dernier. Rappelons que dans notre étude, nous effectuons deux types d'enquêtes : la reprise d'enregistrements médiatiques (les filmages empruntés à Mangiante (2016) et les clips vidéos sur internet) et l'enregistrement d'activités professionnelles. Pour les filmages empruntés, il nous faut demander une permission à Mangiante, car il a pris l'initiative de ces œuvres collectives ; pour les clips vidéos sur internet, nous pouvons les consulter sans permissions ; pour les filmages dans deux cuisines, même si nous sommes l'auteur de cette œuvre, il implique la vie professionnelle des personnes concernées, de ce fait, il faut demander l'accord des participants. Afin d'avoir les permissions de Mangiante et des participants, nous devons préparer un consentement à signer.

#### 4.2.2.2 Préparation d'un consentement

Le consentement peut se faire de manière écrite ou orale. Dans le cadre de notre projet, nous avons choisi de recourir à un accord par écrit afin d'en conserver une trace. Par respect pour le patron et les cuisiniers, nous les avons informés à l'avance de la période de nos enregistrements, mais pas pour chaque enregistrement. Pour l'élaboration d'un consentement, toutes les informations sur les enquêtes peuvent être abordées, telles que le processus ou le contexte de cette enquête. Pour que notre consentement soit clair et formel, nous nous sommes appuyés sur les consentements de PFC (Phonologie du Français Contemporain dans l'espace Francophone) et sur ceux utilisés dans les projets d'ICAR<sup>81</sup>. Dans notre consentement de participation (figure 4-5 ci-dessous), nous avons mentionné le projet de

---

<sup>81</sup> Site du laboratoire ICAR (Interaction, Corpus, Apprentissage, Représentation) : <http://icar.univ-lyon2.fr> (consulté le 10/10/2019).

recherche, le format d'enregistrement, le contexte de diffusion, la finalité de l'enquête ainsi que l'anonymisation.

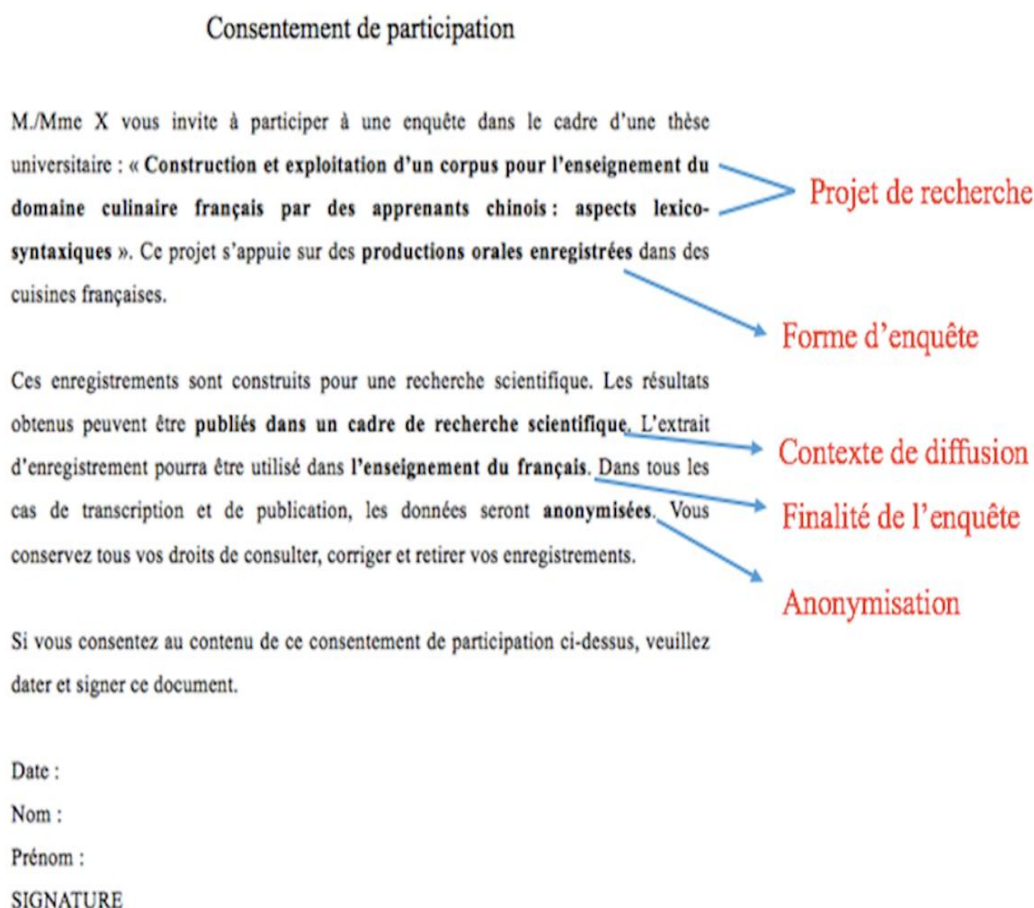


Figure 4-5 : Consentement de participation

Le fait d'avoir fait signer les gens permet d'entamer mes recherches sereinement et de faire les manipulations et exploitations nécessaires pour l'étude du corpus (notamment transcription, analyses et application didactique). Un corpus oral n'est pourtant pas exploitable directement, car « On ne peut pas étudier l'oral par l'oral, en se fiant à la mémoire qu'on en garde. On ne peut pas, sans le secours de la représentation visuelle, parcourir l'oral en tous sens et en comparer des morceaux. » (Blanche-Benveniste, 2010 : 33). L'analyse d'un corpus oral se fonde sur sa transcription. Pour encadrer harmonieusement toutes les transcriptions, en amont de notre engagement, nous avons besoin de préparer une convention de transcription.

#### 4.2.3 Convention de transcription

Les transcrip-teurs ne traduiront pas obligatoirement tous les éléments qui apparaissent dans les supports d'enregistrement et la convention de transcription sera différente selon les finalités de l'étude. De ce fait, le choix des données transcrites sera basé sur les besoins réels de la recherche en cours (Delais-Roussarie, 2003). Par exemple, les évènements paralinguistiques (ex. mimique faciale), les structures segmentales (élision, allongement, amorce de mot, onomatopée), les productions vocales (rires, bruits produits par des machines), la prosodie et les gestes du locuteur ne sont pas traduits pour l'étude lexico-syntaxique. En nous appuyant sur des projets existants (*CHILDES*, *PFC* et projets d'ICAR, etc.), nous avons conservé certains paramètres qui peuvent être utiles, dans notre convention en annexe 2 (p. 232), comme l'identité du participant qui nous permet d'identifier le locuteur et les recommandations sur le tour de parole, sur la pause et sur le chevauchement. Hormis ces indications, en vue d'assurer la fiabilité et la qualité de la transcription, il est fait également appel aux choix du type de transcription ainsi que du logiciel d'aide à la transcription.

##### 4.2.3.1 Choix de type de transcription

La qualité de la transcription est décisive dans la constitution du corpus oral et le choix du type de transcription va assurer la fiabilité et la qualité de la transcription. Il existe des types variés de transcriptions, tels que la transcription orthographique, la transcription phonétique large/étroite, la transcription acoustique-phonétique, etc. Ces types de transcription se rangent dans deux grandes catégories : les transcriptions phonétiques et les transcriptions orthographiques. Dans le choix d'une transcription appropriée/cohérente, plusieurs critères doivent être pris en considération : les objectifs du projet, les critères de taille du corpus ainsi que le degré de détails recherché (Delais-Roussarie, 2003 : 99). Ainsi, deux critères vont guider

prioritairement notre choix : la taille du corpus et les objectifs visés (cf. tableau 4-3 ci-dessous).

	Taille du corpus	Objectifs visés
Transcriptions phonologiques	à petite échelle	phonologique
Transcriptions orthographiques	de grande ampleur	lexicale, syntaxique ou interactionnelle

Tableau 4-3 : Transcriptions phonologiques et orthographiques

Les premières transcriptions du tableau 4-3 sont souvent utilisées pour un corpus de petite taille, parce qu'elles sont chronophages du fait de la traduction juxtalinéaire (traduction mot à mot le plus strict possible). Ces premières transcriptions sont proposées pour l'étude phonologique. Par exemple, Abou-Haidar (2002) adopte ces transcriptions phonologiques pour une étude de la notation de certaines formes d'aphasie comme les jargons. Les transcriptions orthographiques peuvent être recommandées pour un corpus de grande taille, par exemple, le corpus de PFC qui contient près d'un million et demi de mots. Ces transcriptions orthographiques ne peuvent pas être proposées pour refléter le caractère phonémique d'un corpus, mais elles permettent une étude lexicale, syntaxique ou interactionnelle. Pour les transcriptions orthographiques, les transcripteurs sont souvent confrontés à des problèmes phonologiques : par exemple, l'expression 'il y a' peut être transcrite par [ilia, ilja, ija, ja, jæ]. Les différentes prononciations ou les accents pourraient nous empêcher de faire des requêtes orthographiques ultérieures et donc il faut retrouver le même mot écrit de la même façon même s'il est prononcé différemment. Heureusement, dans la transcription orthographique, il existe une approche de « transcription orthographique standard » (TOS) qui ne comprend qu'une prononciation standard, à savoir qu'elle pourrait éviter ce problème phonologique et nous transcrivons toutes ces différentes séquences en [il y a]. Cependant, la question de la fidélité de la représentation de l'oral

pourrait se poser avec l'utilisation de la TOS, car cette transcription pourrait détruire cette représentation. Blanche-Benveniste et Jeanjean (1987 : 115) constatent déjà ce paradoxe et nous confirment qu'une exigence de transcription se manifeste à la lisibilité sur l'écrit :

*transcrire de la langue parlée tient un peu du paradoxe : garder dans une représentation écrite certaines caractéristiques de l'« oralité » ; faire le « rendu » de la chose orale tout en restant dans des habitudes de lecture établies depuis longtemps pour la chose écrite... On va se trouver tiraillé entre deux exigences : la fidélité à la chose parlée et la lisibilité de son rendu par écrit.*

Comme les corpus que nous voulons créer doivent être interrogés sur le plan lexico-syntaxique pour l'extraction et l'enseignement des séquences figées et semi-figées, la lisibilité de la transcription doit être prioritairement prise en compte. D'ailleurs, le projet PFC adopte la TOS qui peut « indexer rapidement toutes les séquences d'une unité lexicale ou de séquences de mots données » (Durand & Tarrier, 2008 : 53). La rapidité de l'indexation est également importante pour nous, donc la TOS est le meilleur choix de transcription dans notre étude.

#### 4.2.3.2 Choix du logiciel

Des logiciels connus existent et aident à la transcription de corpus oraux avec éditeur intégré et alignement sur le signal, tels que *PRAAT* intégré dans le projet PFC. Il existe également *Transcriber* développé et utilisé dans les travaux de Barras *et al.* (1998 ; 2000), *ELAN* fabriqué au Max Planck Institute (Nijmegen, The Netherlands) et adopté dans les études sur les interactions de Mondada (2008) ainsi que *CLAN* développé pour le projet CHILDES (MacWhinney, 2000). Nous les présentons dans le tableau 4-4 ci-dessous pour donner ensuite nos critères de choix de l'un d'entre eux.

	PRAAT	Transcriber	ELAN	CLAN
Projet	PFC	Projet de GRAS (Barras <i>et al.</i> , 2000)	Projet d'ICAR (Mondada, 2008)	CHILDES (MacWhinney, 2000)
Format d'entrée	Audio	Audio	Audio et vidéo	Audio et vidéo
Édition			Fichier excel avec les balises temporelles	Texte seul sans balises temporelles
Avantage			Annotation et codage possibles	Communauté d'utilisateurs

Tableau 4-4 : Comparaisons succinctes des logiciels de transcription

Pour notre étude, le logiciel de transcription sera choisi selon trois critères : le format d'entrée, le format d'édition et la communauté d'utilisateurs. Le format d'entrée (format de traitement) indique si le logiciel est compatible avec le format d'enregistrement. Puisque notre corpus oral se présente sous forme de vidéo (format *MP4* ou *MOV*) et que les deux premiers logiciels (*PRAAT*, *transcriber*) ne peuvent pas traiter ce format de vidéos, il ne nous reste donc plus que *CLAN* et *ELAN*. Comme de multiples annotations peuvent s'ajouter à la transcription, *ELAN* est plutôt utilisé pour des analyses multimodales (entre autres, Mondada, 2008), mais cela n'est pas l'objectif de notre projet. Quant à *CLAN*, il peut produire les textes sans balise temporelle (les balises temporelles installées facilitent la transcription en coupant un enregistrement en petits morceaux) et ce format d'édition permet de donner davantage de possibilités pour le traitement des données récoltées ultérieurement. En outre, *CLAN* est plus largement utilisé qu'*ELAN* par la communauté d'utilisateurs. Il en résulte que *CLAN* est le logiciel de transcription le plus adapté au traitement de corpus que nous voulons développer.

#### 4.2.4 Exemple d'extrait de transcription

Afin de pouvoir avoir une image globale de notre transcription, nous concrétisons une partie de nos transcriptions dans la figure 4-6 ci-dessous.

1 @Begin  
 2 @Languages: fra  
 3 @Participants: CUI2 Male, CUI1 Female  
 4 @ID: fra|change\_corpus\_later|CUI2||||Male||  
 5 @ID: fra|change\_corpus\_later|CUI1||||Female||  
 6 @MEDIA:LASAGNEBOLOGNAISE.MP4  
 7 \*CUI1: une recette copieuse et que tout le monde aime aussi. \*  
 8 \*CUI1: les lasagnes <à la bolognaise>. \*  
 9 \*CUI2: ça, c'est la lasagne préférée de tout le monde, les petits et les  
 10 grands. \*  
 11 \*CUI1: de toute la famille. \*  
 12 \*CUI2: alors, on commence par huiler légèrement un plat à gratin, ça peut  
 13 être un petit peu facultatif, ça. \*  
 14 \*CUI2: on part sur une base de lasagnes sèches. \*  
 15 \*CUI2: et ce qui est important c'est que tout l'ensemble soit suffisamment  
 16 liquide. \*  
 17 \*CUI2: pour que ça puisse absorber le jus de cuisson. \*  
 18 \*CUI1: et avoir les lasagnes très très moelleuses, <donc on commence par une  
 19 belle. \*  
 20 \*CUI1: couche de béchamel>. \*  
 21 \*CUI1: un centimètre à peu près, <ok> ensuite je mets. \*  
 22 \*CUI1: de la sauce bolognaise, <ça se mélange un peu, mais c'est vraiment  
 23 pas grave>. \*

17feb16[E|CHAT] \* 6

En-têtes  
 Lien vidéo  
 Identité du participant  
 Chevauchement




Figure 4-6 : Extrait de nos transcriptions de CLAN

D'après les recommandations de la TEI<sup>82</sup> (cf. Sperberg-McQueen & Burnard, 1999), les renseignements utiles sur les corpus oraux seront intégrés dans les en-têtes de transcription, par exemple, dans notre étude, la date afin de savoir si le document est récent, et le lieu de la chaîne pour connaître notamment le caractère français ou francophone du document puisque nous avons une chaîne belge et que le lexique peut être différent en Belgique. Dans ce fichier vierge de CLAN, les en-têtes peuvent être formées de manière automatique. La ligne des en-têtes est symbolisée par @:. Celles-ci sont suivies du lien vidéo qui est composé de « @MEDIA: » : le nom de la vidéo en majuscule, un point et le format de vidéo (ex. : @MEDIA:LASAGNEBOLOGNAISE.MP4). En

<sup>82</sup> TEI (Text Encoding Initiative) : c'est un format de codage international qui a pour objectif de donner des recommandations pour l'encodage de documents textuels.



amont de chaque énoncé, le locuteur est identifié et un point noir, en aval de chaque énoncé, correspond à une balise temporelle qui permet de synchroniser la transcription et la vidéo. Le chevauchement est marqué dans cet extrait, par exemple, dans l'énoncé 21 de la figure 4-6 ci-dessus, par « ok » qui est inscrit entre deux chevrons. Le format natif (format de sortie) finit par « .cha » ou « .ca ». Pour changer ce format en format *TXT* pour que les transcriptions puissent être utilisées dans des concordanciers (ex. *antConc*), il suffit de copier/coller les textes dans un fichier *TXT*.

Comme prévu par Mangiante (2016), et mentionné plus haut dans cet article, nous avons été confrontés à des problèmes de transcription, lors de la transcription des filmages dans deux cuisines françaises : la cuisine est un lieu bruyant (fonctionnement des machines, ustensiles qui s'entrechoquent lors de la préparation de plats, etc.). Afin d'assurer l'exactitude des transcriptions, celles-ci doivent être validées par des contrôles. Comme l'indique Debaisieux (2005), l'essentiel de la transcription ne peut être automatisée et la transcription doit être vérifiée par plusieurs personnes pour atteindre un bon niveau de fiabilité. Dans nos études, le travail de vérification des transcriptions est effectué par deux correcteurs francophones<sup>83</sup>, ceci dans le but d'atteindre le meilleur niveau de qualité possible.

#### 4.3 Réflexion sur l'exploitation de *Cuisitext* dans notre projet

Une fois *Cuisitext* construit, il est temps de discuter de la question de son exploitation. *Cuisitext* sera exploité, dans notre projet, des quatre manières suivantes :

- Découverte d'éléments saillants : rappelons que, par l'approche *corpus-driven* (cf. 3.2.1 supra), les chercheurs peuvent découvrir des faits de langue saillants et vérifier des hypothèses au sujet de phénomènes

---

<sup>83</sup> Nous remercions Migné Delphine et Élisabeth Cyril pour le travail de vérification.

linguistiques. En conséquence, dans notre projet, *Cuisitext* sera exploité, de prime abord, pour découvrir des éléments lexicaux saillants.

- Description linguistique (modération lexicale) : rappelons que, par l'approche corpus-based (cf. 3.2.1 supra), l'enseignant peut confirmer l'hypothèse émise par le linguiste. En outre, en se basant sur le corpus, l'enseignant peut aussi exemplifier une structure (Aston, 1997 ; Kübler, 2014) et le fait d'exemplifier permet de décrire ou extraire exhaustivement un fait de langue. *Cuisitext* servira ensuite au support d'observation pour parfaire une description linguistique d'un fait de langue.
- Enrichissement de la liste de mots pour l'enseignement : les faits de langue saillants méritent notre attention pour l'enseignement. Dans notre projet, la sélection lexicale pour l'enseignement se fait principalement sur le corpus écrit, mais le corpus oral peut servir à enrichir notre répertoire lexical à enseigner.
- Exploitation pour l'aspect didactique : la richesse des corpus offre des possibilités didactiques. Rappelons que, par l'approche *exploitation to teach* (cf. 3.2.2 supra), le corpus sera utilisé comme support d'enseignement, à savoir que *Cuisitext* sera exploité par l'enseignant ou l'apprenant pour l'E/A des langues.

## Bilan

Par manque de corpus approprié au domaine culinaire, nous sommes obligés de construire notre propre corpus baptisé *Cuisitext* pour remédier aux lacunes lexico-syntaxiques des cuisiniers étrangers en France. Pour la construction de notre corpus écrit, comme une recette de cuisine qui est un genre de textes culinaires et qui peut contenir tous les matériaux lexicaux du domaine culinaire, nous décidons de collecter des recettes de cuisine sur internet de manière

automatique grâce au crawler *Gromoteur*. Nous élaborons également un corpus oral, et nous choisissons *CLAN* comme logiciel de transcription et la TOS comme type de transcription. Comme dans d'autres projets (*CHILDES*, *PFC* et projets d'*ICAR*, etc.), nous avons établi notre consentement de participation et notre convention de transcription. Pour l'instant, notre corpus *Cuisitext* contient 1 million de mots et continue de s'élargir par nos soins. À la fin de cette partie, nous réfléchissons sur la manière dont notre corpus sera exploité dans notre projet. Après notre discussion, quatre approches d'exploitation du corpus sont envisageables : la découverte d'éléments saillants, la description linguistique, l'enrichissement de la liste de mots pour l'enseignement et l'exploitation pour l'aspect didactique. Dans la partie suivante, *Cuisitext* sera exploité pour la découverte des phrasèmes saillants.

## 5. Repérage des phrasèmes saillants et leur sélection pour l'enseignement

### 5.1 Objet d'étude : phrasèmes *NAdj*

Comme nous l'avons vu précédemment, les phrasèmes occupent une place importante dans la langue, mais quels phrasèmes devons-nous enseigner auprès des apprenants étrangers cuisiniers ? En s'appuyant sur des projets (Ahmad, 1994 ; Drouin, 2002 ; Chung, 2003 ; Drouin & Bae, 2005 ; Gillam *et al.*, 2005) dans lesquels le fait de comparer différents corpus permet de faire ressortir les termes spécifiques, nous décidons donc de rechercher les lexèmes spécifiques du domaine culinaire en comparant *Cuisitext* avec d'autres corpus afin de nous aider à cerner les phrasèmes à enseigner. Le CERF<sup>84</sup> a été choisi comme corpus de référence dans notre comparaison, car les textes du CERF sont issus de différents genres de textes tels que conte, discours, poésie, roman, théâtre, correspondance, critique. Grâce au *Trameur*, nous avons pu obtenir les dix lexèmes les plus spécifiques de *Cuisitext* par rapport au CERF (indiqués dans la figure 5-1 ci-dessous).

Item (Fq > 10)	Sp. cuisitext	Sp. correspondance	Sp. critique	Sp. poesie	Sp. roman	Sp. discours	Sp. theatre	Sp. contes
50	50	0	0	-5	-4	-2	-7	-7
couche	50	-4	-3	0	0	-6	-4	-5
persil	48	-3	-3	-3	-4	-4	-3	-3
feuilletée	48	-3	-3	-3	-4	-4	-3	-3
ajoutez	48	-3	-4	-3	-4	-3	-4	-4
poulet	48	-5	-5	-3	-5	-6	-4	2
jaunes	48	-5	-4	0	-3	-6	-5	0
basilic	47	-3	-3	-3	-4	-4	-3	-3
gousse	47	-3	-3	-3	-4	-4	-3	-3
muscade	47	-3	-3	-3	-4	-4	-3	-3

Figure 5-1 : Les dix lexèmes les plus spécifiques de *Cuisitext*

<sup>84</sup> CERF (Corpus Évolatif de Référence du Français) : un corpus constitué par l'équipe Delic (Description linguistique informatisée sur corpus) de l'Université d'Aix-en-Provence sous la direction de Jean Véronis (Debaisieux, 2013).

*Le Trameur* associe chaque lexème de *Cuisitext* à un indice de spécificité et ce dernier est calculé à partir de la comparaison avec le CERF. L'indice se trouve dans la deuxième colonne du tableau et plus il est grand, plus l'item est spécifique. Nous constatons que l'un des lexèmes les plus spécifiques est le chiffre 50 dans la 1<sup>re</sup> colonne (c'est la colonne « Item », c'est-à-dire les mots). Comme les chiffres sont beaucoup employés dans une recette de cuisine, contrairement à un poème ou un roman, nous trouvons cela tout à fait normal. Si l'on observe les dix lexèmes les plus spécifiques, la plupart sont des noms : les six noms *couche*, *persil*, *poulet*, *basilic*, *gousse*, *muscade*. De surcroît, dans notre étude précédente (Yang & Cavalla, 2019), nous avons trouvé, dans *Cuisitext*, un nombre de séquences de type nom + adjectif (désormais *NAdj*), tels que *feu très doux*, *pêche plate*, *thé vert*. Comme mentionné en introduction, l'importance de la phraséologie doit être prise en compte pour l'enseignement. De la sorte, en tant qu'éléments remarquables, les phrasèmes *NAdj* constituent l'objet de notre étude. Rappelons (dans la sous-partie 2.3) que la phraséologie peut être définie par deux grandes approches (statistique et fonctionnelle), comme montré dans la figure 5-2 ci-dessous.

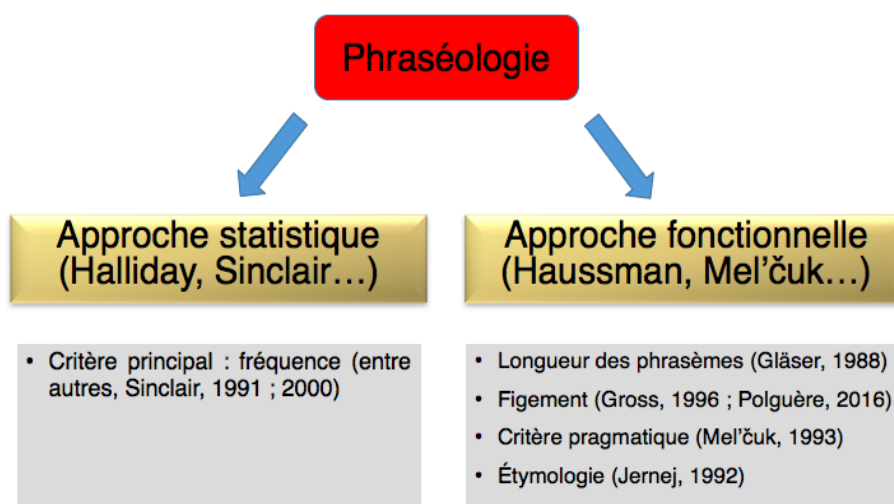


Figure 5-2 : Deux approches de définition de la phraséologie

Afin de définir la phraséologie dans notre projet, nous décidons d'adopter l'approche statistique dans laquelle le critère de fréquence est un révélateur pour identifier les unités phraséologiques. Autrement dit, nous allons identifier les phrasèmes de notre étude en fonction de leurs fréquences. Pour obtenir leurs fréquences, leur extraction du domaine culinaire (Yang, 2018a) reste au premier plan.

## 5.2 Extraction des phrasèmes *NAdj*

Différentes méthodes s'emploient pour l'extraction phraséologique, par exemple, la méthode des questionnaires (cf. dans *BBI Dictionary of English Word Combinations* de Benson *et al.*, 1986) ou celle de l'introspection (cf. dans DEC<sup>85</sup> de Mel'čuk *et al.*, 1995 et dans DCF<sup>86</sup> de Zinglé, 2003). Mais Williams signale que ces méthodes seraient impossibles « à mettre en pratique dans le cadre des langues de spécialités. La solution peut se trouver dans l'utilisation des corpus » (1999 : 156). Kübler ajoute que « le corpus et les outils d'interrogation de corpus permettent des interrogations beaucoup plus spécifiques et structurées, qui tiennent compte des genres textuels et des domaines de spécialité » (2014 : 40). *Cuisitext* peut servir de support d'observation pour extraire les phrasèmes *NAdj*. Dès le début de notre projet (cf. 1.1 supra), nous nous sommes rendu compte que le TAL apportait une aide considérable à l'apparition de la discipline linguistique de corpus, notamment à sa construction et à son traitement. L'utilisation de l'outil TAL peut rendre l'extraction efficace. Les linguistes, les didacticiens, les terminologues et les traducteurs ont recours à des logiciels développés pour l'extraction de séquences polylexicales. En conséquence, un bon logiciel d'extraction peut assurer une extraction exhaustive. En effet, le choix du

---

<sup>85</sup> DEC : Dictionnaire Explicatif et Combinatoire du français contemporain

<sup>86</sup> DCF : Dictionnaire Combinatoire du Français : expression, locution et construction.

logiciel approprié est souvent en lien avec les exigences de la modélisation lexicale de l'objet étudié. La modélisation ne nous offre pas seulement une vue structurée mais nous aide aussi à choisir le logiciel approprié, comme le disaient Aussenac-Gilles & Condamines (2001 : 2) : « La modélisation du contenu d'un document offre une vue structurée qui permet d'imaginer des outils plus variés pour leur consultation ou leur exploitation ». Nous commençons donc notre extraction des phrasèmes par leurs modélisations. Et en fonction des exigences de leurs modélisations, nous choisirons le logiciel approprié de notre projet.

### 5.2.1 Modélisation lexicale des données

La modélisation est souvent présentée comme une méthode efficace pour accéder facilement au texte et il s'agit d'un moule doté de toutes les propriétés sémantiques et syntaxiques associées. Inspiré par Tutin (2004), notre modélisation est construite à partir d'observations relevées dans *Cuisitext* et dans les dictionnaires (*le TLFI* et *le Larousse* en ligne). Comme notre modélisation des phrasèmes *NAdj* touche la collocation (ex. *steak bleu* : *steak* garde son sens habituel et *bleu* prend un sens déviant), notre étude a également recours à la fonction lexicale (FL). Rappelons que, dans la sous-partie 2.4, la FL est un outil conceptuel des descriptions des langues en modélisant et encodant aussi bien les liens paradigmatiques que les liens syntagmatiques. La FL [**Prepar**] est un verbe ayant le sens de « préparer pour » qui a pour mot-clé L comme complément d'objet central ; la FL [**Fact<sub>0</sub>**] est un verbe de réalisation. En prenant le cas de la lexie *ail* comme mot-clé, cette combinaison des deux fonctions lexicales peut être exprimée comme suit :

$$\mathbf{PreparFact_0}(\mathit{ail}) = \mathit{hacher} [\mathbf{ART}^{87} \sim]$$

La FL [**A<sub>2</sub>**] est un modificateur adjectival, donc l'adjectif du verbe [**PreparFact<sub>0</sub>**] (*ail*) peut être exprimé comme suit :

---

<sup>87</sup> Dans *NooJ*, ART signifie les articles définis ou indéfinis.

**A<sub>2</sub>PreparFact<sub>0</sub>(ail) = haché**

Cette FL aide à encoder les liens sémantiques et syntaxiques : au niveau sémantique, la lexie HACHÉ est l'adjectif du verbe « hacher » qui exprime la « façon de préparer » ; au niveau syntagmatique, cette lexie se trouve toujours après l'ingrédient qui va être qualifié. D'après nos observations sur *Cuisitext*, les grammaires des phrasèmes *NAdj* peuvent être exprimées selon deux structures principales : un syntagme nominal (nom + adjectif, ex. *ail haché ; gingembre râpé ; échalote coupée*) et une structure attributive (nom + verbe + adjectif, ex. *ail est haché*). Les modélisations des deux structures seront présentées ci-dessous : celle concernant l'extraction de la structure nominale, suivie de celle permettant l'extraction de la structure attributive.

5.2.1.1 Modélisation du syntagme nominal

En ce qui concerne le syntagme nominal, il peut y avoir l'insertion d'un adverbe entre les composants (nom et adjectif). Une telle insertion a été constatée par Polguère (2016) comme une sorte d'expression « à trous » (ibid., p. 64) et la taille de ce « trou » n'est pas mesurable. Quelle que soit l'insertion (désormais *Ins*), la grammaire peut être décrite comme suit :

$$G = [\text{nom} + \text{Ins ou pas} + \text{adjectif}]$$

5.2.1.2 Modélisation de la structure attributive

En ce qui concerne la structure attributive, elle peut être transformée en d'autres structures (L) et leurs grammaires correspondantes (G) sont très variées et diversifiées, comme indiqué dans le tableau 5-1 ci-dessous.

L=	G=
{“l'ail est écrasé”}	[le + nom + être + adjectif ]
{“ce petit steak est bleu ”}	[ce + petit + nom + être + adjectif ]
{“le caramel, qui ne doit pas devenir noir”}	[le + nom, qui + ne + devoir + pas + verbe + adjectif ]



{“ces aux, ils doivent être écrasés”}	[ces + nom, pronom + devoir + être + adjectif ]
{“ils doivent être écrasés, ces aux”}	[pronom + devoir + être + adjectif, ces + nom]
{“.....”}	.....

Tableau 5-1 : Modélisation de la structure attributive

Il nous semble impossible de créer autant de grammaires que de structures, mais, nous avons fait le constat suivant : toutes les structures présentées dans le tableau 5-2 ci-dessous partagent les mêmes matériaux lexicaux, c'est-à-dire un Sujet (S), un Verbe d'État (VE) et un Attribut (A). À l'aide de ces trois matériaux lexicaux, nous pouvons simplifier les modélisations de la structure attributive reprises dans le tableau 5-2 ci-dessous :

L =	G =
{“l'ail est écrasé”} ; {“ce petit steak est bleu”}.	[S + Ins ou pas + VE + Ins ou pas + A]
{“le caramel, qui ne doit pas devenir noir”} ; {“ces aux, ils doivent être écrasés”}.	[S, pronom + Ins ou pas + VE + Ins ou pas + A]
{“ils doivent être écrasés, ces aux”}.	[pronom + Ins ou pas + VE + Ins ou pas + A, S]

Tableau 5-2 : Modélisation avec la variable globale

Dès lors, la modélisation des deux structures (syntagme nominal et structure attributive) peut être résumée en quatre grammaires principales indiquées dans la figure 5-3 ci-dessous.

Modélisation des phrasèmes NAdj
Nom + Ins ou pas + Adjectif
S + Ins ou pas + VE + Ins ou pas + A
S, Pronom + Ins ou pas + VE + Ins ou pas + A
Pronom + Ins ou pas + VE + Ins ou pas + A, S

Figure 5-3 : Modélisation des deux structures en quatre grammaires

### 5.2.1.3 Exigences de notre modélisation des quatre grammaires

Comme mentionné dans la sous-partie 5.2, le logiciel approprié sera choisi en fonction des exigences de leurs modélisations. Dans cette partie, les exigences de la modélisation des quatre grammaires seront prises en compte pour le choix de notre logiciel approprié. Le Sujet de ces grammaires consiste en un groupe nominal (GN) qui pourrait s'écrire de multiples façons. Comme proposées par des linguistes (entre autres, Dubois, 1973), les grammaires du GN peuvent s'écrire selon trois assertions présentées dans la figure 5-4 ci-dessous.

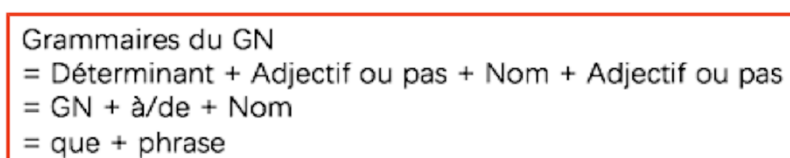


Figure 5-4 : Grammaires du GN

La définition du GN devient notre première exigence pour la modélisation qui nous concerne. Il faudrait définir tous les GN des quatre grammaires principales comme indiqué dans la figure 5-5 ci-dessous.

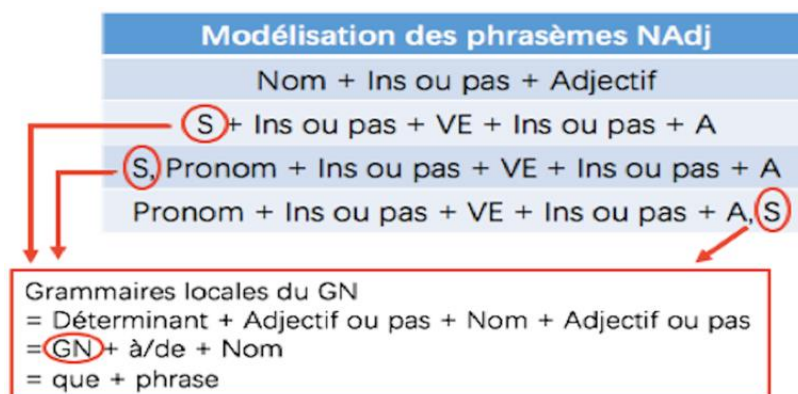


Figure 5-5 : Modélisation des phrasèmes NAdj

Dans les grammaires du GN, nous remarquons qu'il y a un autre GN (entouré en rouge dans la figure 5-5 ci-dessus). Si nous essayons de définir le GN dans les grammaires du GN, nous finirons par avoir un nouveau GN et cette procédure de la définition du GN ne s'arrête jamais comme exposée dans la figure 5-6 ci-dessous ; elle est donc récursive.

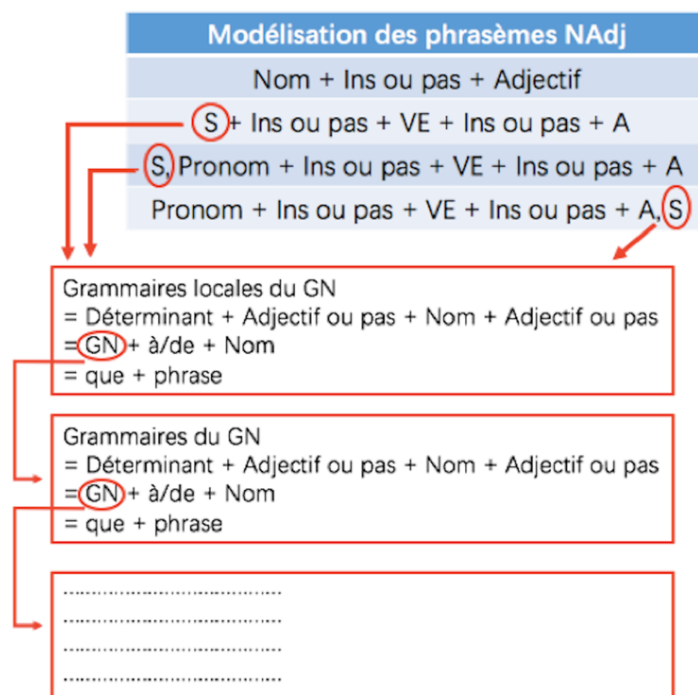


Figure 5-6 : Définition du GN

Dès lors, les grammaires du GN devraient se référer à elle-même. Autrement dit, la récursivité des grammaires est la deuxième exigence de notre modélisation.

Nous nous rendons compte que certaines expressions indésirables connotant au moins un élément polysémique peuvent être incluses dans notre modélisation, par exemple, le lexème *juste* dans l'expression « persil juste » (parsemer de persil juste avant de servir ou dans la poêle juste avant de servir) est utilisé comme un adverbe, mais il peut être reconnu par un logiciel d'extraction comme un adjectif (« persil juste » : Nom + Adjectif). Dans certains exemples (*tout* dans « tout seul » ; *un* dans « un seul » ; *autre* dans « un autre », *reste* dans « il ne reste presque plus de liquide »), les adjectifs et les noms portent un sens très général (les noms : fois, reste, tout, un ; les adjectifs : presque, juste, autre, seul), c'est-à-dire qu'ils ne sont pas spécifiques au domaine culinaire. Afin d'éliminer les expressions non-pertinentes, nous devons procéder à une désambiguïisation lexicale. Comme indiqué dans la figure 5-7 ci-dessous, la désambiguïisation prévue (en

couleur verte) se fera simultanément sur les quatre grammaires principales et sur les grammaires locales du GN.

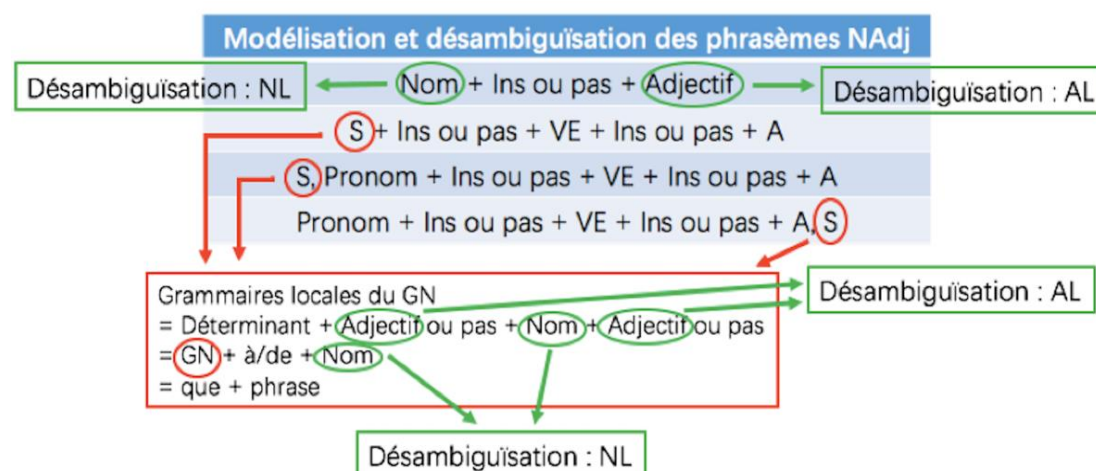


Figure 5-7 : Modélisation et désambiguïsation des phrasèmes NAdj

Nos trois exigences de modélisation (définition de la grammaire locale, récursivité de la grammaire et désambiguïsation lexicale) seront prises en compte dans notre choix du logiciel d'extraction approprié.

### 5.2.2 Choix du logiciel d'extraction

Du fait d'un manque de connaissances informatiques, certains didacticiens, des terminologues et des traducteurs ont recours à des fonctions ou programmes développés pour l'extraction de séquences et nous les présentons brièvement dans le tableau 5-3 ci-dessous :

	Exemples	Avantages dans l'extraction
Logiciels De textométrie	<i>AntConc</i> (Antony, 2005), <i>Lexico 3</i> (Lafon & Salem, 1983 ; Salem, 1987), <i>Trameur</i> (Fleury, 2007), <i>Hyperbase</i> (Brunet, 2006), <i>TXM</i> (Heiden <i>et al.</i> , 2010), <i>DtmVic</i> (Lebart, 2015).	Analyse statistique
Logiciels dédiés à l'extraction polylexicale	<i>Programme de Lin</i> (1998), <i>FipsCo</i> (Goldman <i>et al.</i> , 2001), <i>Programme de Grefenstette et Teufel</i> (1995), <i>sketch engine</i> (Adam Kilgarriff & David Tugwell, 2001).	Analyses linguistique et statistique ( <i>Z-score</i> , <i>Information mutuelle</i> )

Logiciel de formalisation des langues	<i>NooJ</i> (Silberstein, 2002)	Analyses linguistique et statistique flexibilité
---------------------------------------	---------------------------------	---

Tableau 5-3 : Programmes développés pour l'extraction phraséologique

L'Inventaire des segments répétés (ISR) de *Lexico 3* ou « n-gram » d'*Antconc* permet de repérer « les segments dont la fréquence est supérieure ou égale à 2 » (Salem, 1987 : 50) et de constituer « le point de départ d'expériences statistiques intéressantes qui dépassent les descriptions lexicométriques traditionnelles parce qu'elles prennent en charge la dimension syntagmatique » (Lafon & Salem, 1983 : 176). Cette requête pourrait être utilisée pour repérer des séquences plus ou moins figées (Fiala, Habert, Lafon & Pineira, 1987). Par exemple, Lewis (2000) a recours à une telle requête afin d'aider les apprenants à acquérir un répertoire de collocations. Certains spécialistes du TAL se sont rendu compte de la difficulté de l'extraction phraséologique et consacrent leur temps au développement de logiciels destinés à l'extraction des expressions polylexicales (Smadja, 1993 ; Lin, 1998 ; Kilgarriff & Tugwell, 2001 ; Goldman *et al.*, 2001). Précisons que seules des analyses statistiques (par exemple, *Z-score*, *Information mutuelle*) sont intégrées pour cette extraction (Choueka *et al.*, 1983 ; Church & Hanks 1989 ; Church *et al.*, 1991). Il est vrai que cette intégration assure l'exactitude du résultat de l'extraction, mais il manque toujours des analyses linguistiques, auxquelles tout linguiste est confronté régulièrement. Par exemple, nous avons besoin d'un étiquetage morphosyntaxique pour l'extraction. Ensuite, des spécialistes associent le traitement linguistique à l'analyse statistique dans des programmes développés comme indiqué dans le tableau 5-3 ci-dessus : *programme de Lin* (1998), *FipsCo* (Goldman *et al.*, 2001) et *programme de Grefenstette et Teufel* (1995). Le premier *programme de Lin* peut analyser les phrases de plus de 25 mots-occurrences et le deuxième

*FipsCo* peut réaliser de nombreuses transformations grammaticales pour le traitement du sujet direct, tandis que le dernier *programme de Grefenstette et Teufel* ne peut réaliser que certaines désambiguïisations de verbes. Par rapport aux programmes développés, il existe également un logiciel de formalisation des langues qui permet d'extraire les expressions visées par le biais de la description linguistique des langues : *NooJ* (Silberztein, 2002).

D'après Aussenac-Gilles & Condamines (2001), le choix du logiciel approprié dépend des besoins de l'étude. Ainsi, les choix des éléments à extraire pour notre étude se déclinent en trois points : définition du groupe nominal, récursivité de la grammaire, désambiguïisation lexicale. D'abord *L'ISR* ou les *n-gram*, dans des logiciels de textométrie, peuvent faire ressortir les combinaisons lexicales fréquentes dans un espace textuel de corpus. Toutefois, le résultat est souvent frustrant et inexploitable, car d'une part, l'espace textuel détecté est limité dans une fenêtre de 5 mots, et certaines interprétations idiosyncrasiques dépassent cette limite de mots (Sag *et al.*, 2002). Comme l'estiment Goldman *et al.* (2001), 29.26% de combinaisons lexicales sont des occurrences « lointaines » (l'espace textuel est supérieur à la limite de cinq mots). D'autre part, par manque d'analyse linguistique automatique, de faux phrasèmes peuvent être extraits, comme l'a démontré une enquête 2-gram dans la figure 5-8 ci-dessous.

Total No. of N-Gram Types		15785		Total No. of N-Gram Tokens		41008	
Rank	Freq	Range	N-gram				
1	514	1	g de				
2	412	1	de la				
3	282	1	la recette				
4	276	1	préparation de				
5	229	1	à soupe				
6	197	1	dans un				
7	165	1	de sucre				
8	162	1	cuillère à				
9	162	1	cuillères à				
10	161	1	et le				
11	159	1	soupe de				
12	145	1	de beurre				
13	134	1	la pâte				
14	131	1	et les				
15	116	1	dans une				
16	114	1	de crème				
17	114	1	de farine				
18	112	1	à café				
19	105	1	huile d				
20	105	1	un peu				
21	104	1	la crème				
22	104	1	le sucre				
23	103	1	crème fraîche				
24	103	1	d eau				

Figure 5-8 : Extrait de 2-gram par *Antconc*

Ensuite, nous nous sommes tournés vers les logiciels dédiés à l'extraction polylexicale. Ces derniers peuvent désambiguïser des verbes, mais ils ne peuvent pas réaliser les désambiguïsisations sur les noms et sur les adjectifs. En outre, ils ne répondent pas au besoin de définition d'un groupe nominal. Afin de répondre précisément à nos exigences, nous recourons à un logiciel de formalisation des langues : *NooJ* (Silberztein, 2002). En théorie, *NooJ* est capable de décrire toutes les langues de façon exhaustive (Silberztein, 2016). La hiérarchie de Chomsky-Schützenberger (Chomsky, 1963) est utilisée pour décrire les quatre grammaires génératives : grammaire non restreinte, grammaire contextuelle, grammaire algébrique et grammaire rationnelle. Au niveau de la grammaire algébrique, nous pouvons utiliser une grammaire locale (nœud coloré) pour définir le groupe nominal. Par ailleurs, l'atout de la grammaire algébrique permet de réaliser la récursivité de la grammaire. Nos deux premières exigences (définition de la grammaire locale et récursivité de la grammaire) peuvent donc trouver des réponses grâce à la grammaire algébrique. En outre, notre troisième exigence (désambiguïsisation lexicale)

peut être satisfaite en utilisant la grammaire rationnelle comme repris dans la figure 5-9 ci-dessous.

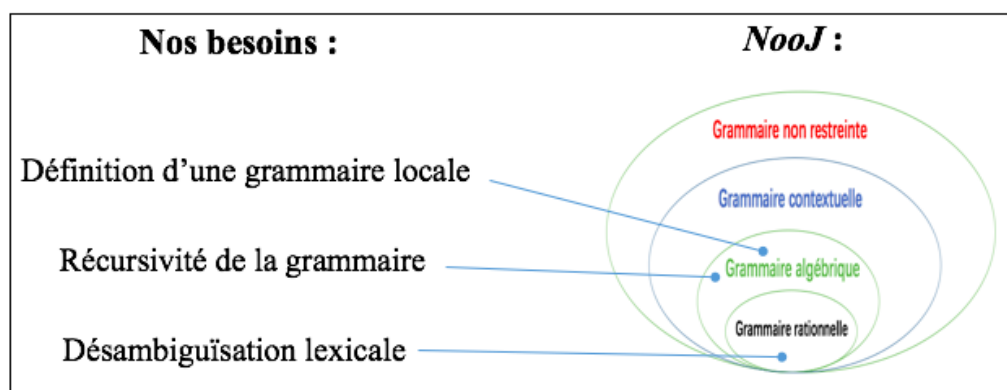


Figure 5-9 : Trois exigences de la modélisation

Étant donné que *NooJ* peut répondre à nos trois exigences, nous le choisissons comme logiciel approprié d'extraction et procédons par la suite à l'implémentation des données lexicales par la suite.

### 5.2.3 Implémentation des données lexicales dans *NooJ*

#### 5.2.3.1 Implémentation du syntagme nominal

La grammaire [Nom + Ins ou pas + Adjectif] peut être implémentée dans *NooJ* comme repris dans la figure 5-10 ci-dessous.

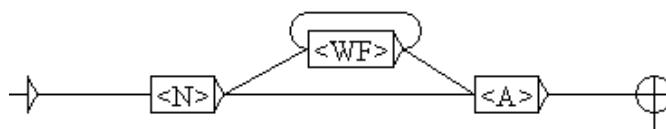


Figure 5-10 : Implémentation de la grammaire nominale

Afin de procéder à la désambiguïstation lexicale à la fois sur les noms et sur les adjectifs, *NooJ* nous offre l'opérateur +EXCLUDE pour rejeter les séquences indésirables. Pour la propreté de notre automate, nous regroupons tous les Noms Polysémiques (NP), sous l'étiquette NP (nœud coloré dans la figure 5-11 ci-dessous), par exemple,  $\langle la \rangle$ <sup>88</sup>,  $\langle être \rangle$ ,  $\langle bien \rangle$ ,  $\langle si \rangle$ ,  $\langle été \rangle$ ,  $\langle petit \rangle$ ,  $\langle y \rangle$ ,  $\langle g \rangle$ ,  $\langle tout \rangle$ ,  $\langle présent \rangle$ ,  $\langle courant \rangle$ ,  $\langle dessus \rangle$ ,  $\langle ce \rangle$ ,  $\langle fois \rangle$ ,  $\langle fait \rangle$ ,

<sup>88</sup> Dans *NooJ*, les guillemets simples ( $\langle \rangle$ ) permettent de retrouver toutes les occurrences de ce terme et de ses variantes.



<reste>, <les>, <le>, <intérieur>, <extérieur>, <morceau> et nous rassemblons tous les Adjectifs Polysémiques (AP) sous l'étiquette AP (nœud coloré dans la figure 5-11 ci-dessous), , comme <moyen>, <seul>, <entier>, <froid>, <inférieur>, <extérieur>, <dur>, <autre>, <fort>, <faible>, <différent>, <juste>, <extra>, <égal>, <nécessaire>, <normal>, <prêt>, <restant>, <suivant>, <régulier>, <allant>, <haut>, <petit>, <grand>, <joli>, <minimum>, <maximum>, <présent>. Ainsi, notre grammaire nominale ressemble au graphe de la figure 5-11 ci-dessous.

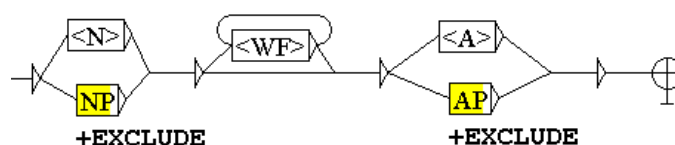


Figure 5-11 : Désambiguïisation de la grammaire nominale

En vue de vérifier que les noms s'accordent avec les adjectifs en genre et en nombre, nous utilisons la variable N en imposant les contraintes <\$THIS\$Nb=\$N\$Nb> et <\$THIS\$Genre=\$N\$Genre> comme repris dans la figure 5-12 ci-dessous.

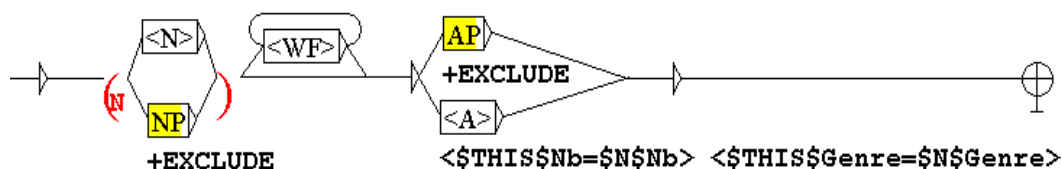


Figure 5-12 : Implémentation du syntagme nominal

### 5.2.3.2 Implémentation de la structure attributive

Rappelons les trois grammaires de la structure attributive dans le tableau 5-4 ci-dessous.

G =
[S + Ins ou pas + VE + Ins ou pas + A]
[S, pronom + Ins ou pas + VE + Ins ou pas + A]
[pronom + Ins ou pas + VE + Ins ou pas + A, S]

Tableau 5-4 : Trois grammaires de la structure attributive

Nous commençons l'implémentation de la première grammaire [S + ins ou pas + VE + ins ou pas + A] dans la figure 5-13 ci-dessous.

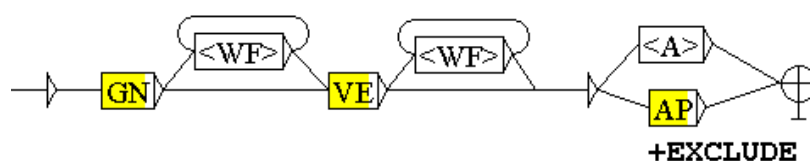


Figure 5-13 : Implémentation de la grammaire S + VE + A

Dans cet automate, les adjectifs sont déjà désambiguïsés en utilisant l'opérateur +EXCLUDE. Nous écrivons tous les verbes d'état sous l'étiquette VE, par exemple, <être>, <rester>, <paraître>, <sembler>, <devenir>, <demeurer>, <redevenir>, <trouver>. Comme nous l'avon vu dans la sous-partie 5.4.1, le Sujet peut être remplacé par le GN, et le graphe imbriqué du GN est indiqué dans la figure 5-14 ci-dessous.

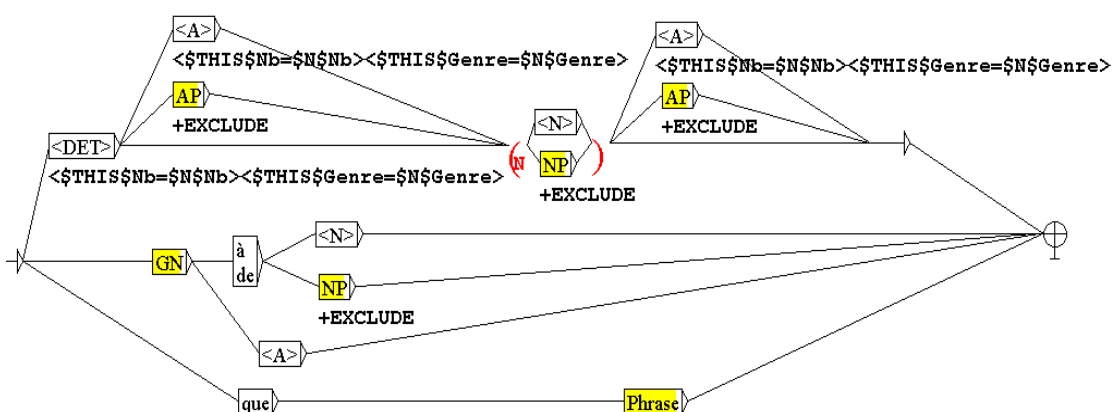


Figure 5-14 : Graphe imbriqué du GN

Dans ce graphe imbriqué du GN, les AP et NP rejetés sont toujours imposés pour la désambiguïsation et la variable N est ainsi définie en nombre et en genre. Rappelons que la grammaire algébrique permet qu'une règle puisse se définir à partir d'elle-même. Le graphe imbriqué du GN se réfère donc à lui-même et cette référence est récursive, car le GN fait partie du graphe. En respectant la même logique, la grammaire locale de la Phrase peut être également définie comme dans la figure 5-15 ci-dessous.

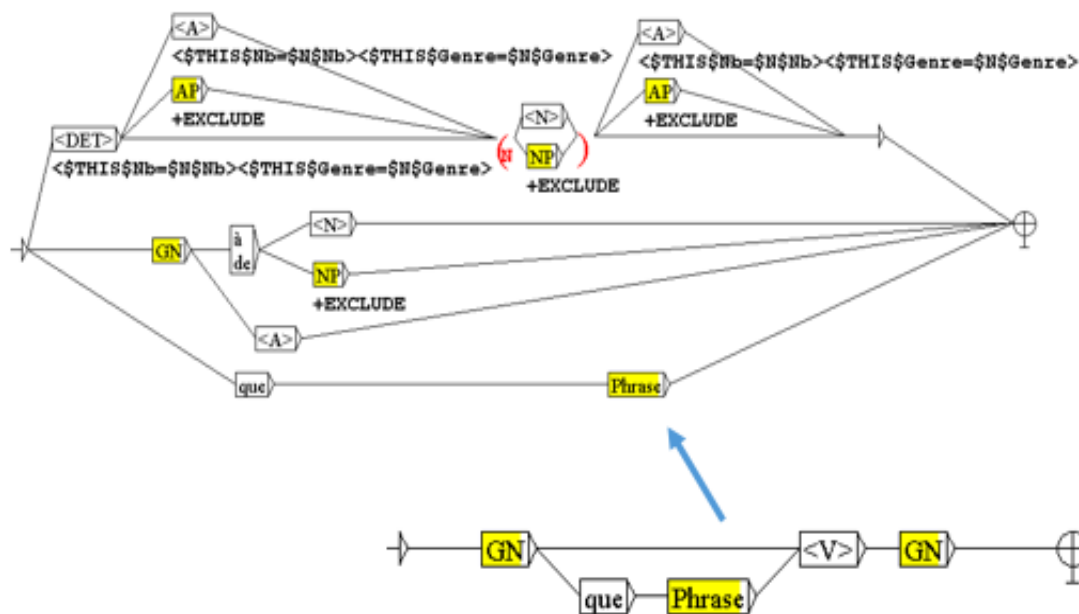


Figure 5-15 : Grammaire de la Phrase imbriqué dans la grammaire GN

En vue de faciliter l'implémentation de la deuxième et de la troisième grammaires de cette structure, nous recourons à la variable globale (préfixée par le caractère « @ »). Rappelons que chaque variable globale a une seule valeur, indépendamment de sa place dans la structure grammaticale et cette valeur est transmise au générateur de *NooJ*. Nous rangeons les GN dans le sujet (@S), les VE dans le verbe (@V) et les adjectifs dans l'attribut (@A). Grâce aux variables globales, l'implémentation de la structure attributive est indiquée dans la figure 5-16 ci-dessous.

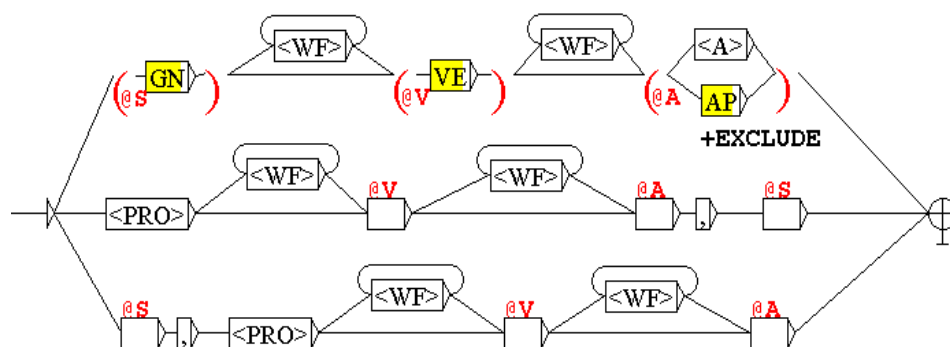


Figure 5-16 : Implémentation de la structure attributive

Chaque structure principale a son propre graphe : le graphe 5-12 pour le syntagme nominal et le graphe 5-16 pour la structure attributive. Pour faciliter

notre interrogation, nous pouvons fusionner les deux graphes en un seul graphe 5-17 (figure 5-17 ci-dessous) qui permet de réaliser notre extraction en un seul lancement.

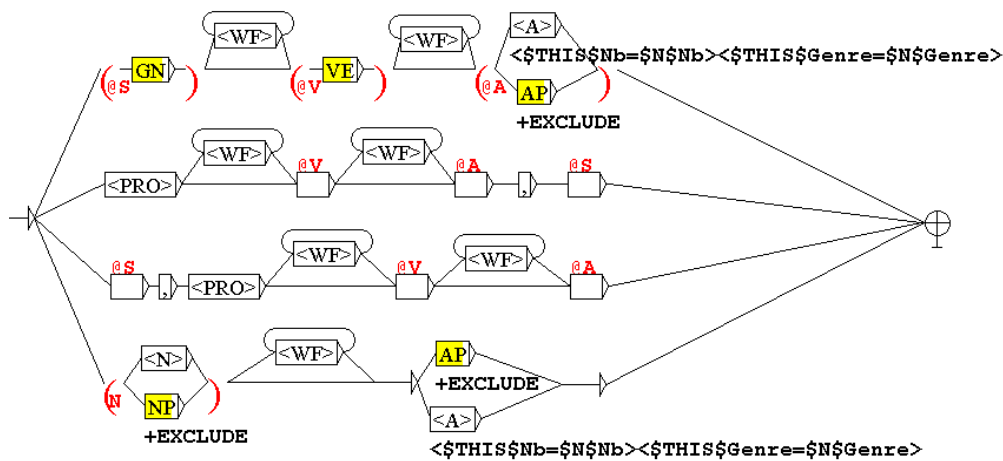


Figure 5-17 : Graphe total

Une fois achevée l'implémentation des données dans *NooJ*, nous pouvons lancer notre requête et le premier résultat est affiché dans la figure 5-18 ci-dessous.

Figure 5-18 : Séquences *NAdj* sous formes de concordances

Comme nous l'avons vu précédemment (cf. 5.1 supra), nous adoptons l'approche statistique pour identifier les phrasèmes *NAdj* et ces derniers seront choisis en fonction de leurs fréquences. Dans la partie suivante, nos phrasèmes *NAdj* seront identifiés et sélectionnés pour l'enseignement.

### 5.3 Identification et sélection des phrasèmes *NAdj* pour l'enseignement

L'importance du critère de fréquence est également prise en compte dans la sélection du lexique pour l'enseignement (cf. 3.1.4 supra). Évidemment, la fréquence des mots est le critère fondamental, mais ce n'est pas le seul critère

dans la sélection : en effet, des didacticiens (entre autres, Gougenheim, Michéa, Rivenc et Sauvageot, 1956 ; Gledhill, 2000 ; Curado Fuentes, 2001) donnent également un poids important aux événements rares, car une faible fréquence peut cacher une haute pertinence pour le domaine spécialisé. Par exemple, les auteurs du *Français Fondamental* découvrent, pendant leur sélection du lexique, le fait que certains mots sont très connus des Français, comme *fourchette*, *coude*, alors qu'ils sont « peu prononcés » (Gougenheim, Michéa, Rivenc et Sauvageot, 1956 : 136-137). Ils pensent qu'il n'est pas judicieux d'ignorer ces mots dans le cadre de l'enseignement du français, « parce que même si nous n'avons pas l'occasion de les prononcer, nous les avons toujours à notre disposition » (Gougenheim, Michéa, Rivenc et Sauvageot, 1956 : 142). Le même cas s'est produit dans notre projet d'enseignement du lexique : « certains mots s'emploient fréquemment quand nous travaillons dans une cuisine, ils n'ont pourtant pas beaucoup d'occurrences dans notre corpus, tels que *orange*, *tabasco*. Nous ne devons pas les ignorer » (Yang, 2015 : 76). À propos de la sélection du lexique en FOS pour l'enseignement, Binon et Verlinde (2004) proposent les deux critères de sélection suivants : un critère objectif et un critère subjectif. Nous conservons leurs critères et nous les croisons avec ceux de Gougenheim, Michea, Rivenc et Sauvageot pour élaborer le *Français Fondamental*, mais ce sont quasiment les mêmes et ils procèdent d'une même logique : la sélection par fréquence des mots est une sélection objective et la recherche des mots disponibles les plus utiles (par exemple, *fourchette*, *coude*) constitue une sélection subjective. Inspiré des travaux mentionnés ci-dessus (Gougenheim, Michea, Rivenc & Sauvageot, 1956 ; Binon & Verlinde, 2004 ; Gledhill, 2000 ; Curado Fuentes, 2001), les phrasèmes de haute fréquence de notre projet doivent être pris en compte, mais nous vérifions également les expressions de fréquence faible pour qu'aucun phrasème pertinent ne nous échappe. Ainsi, en vue d'élaborer notre syllabus lexical, nos phrasèmes *NAdj* du domaine culinaire seront identifiés et sélectionnés en trois étapes comme suit :

- a. La fréquence des phrasèmes dans *Cuisitext*
- b. La recherche des phrasèmes pertinents de fréquence faible
- c. La comparaison avec *Cuisitext* oral (la partie orale de *Cuisitext*)
  - a. Fréquence des phrasèmes dans *Cuisitext*

Grâce à *NooJ*, les phrasèmes extraits peuvent être classés par ordre décroissant des occurrences de mots, comme indiqué dans la figure 5-19 ci-dessous.

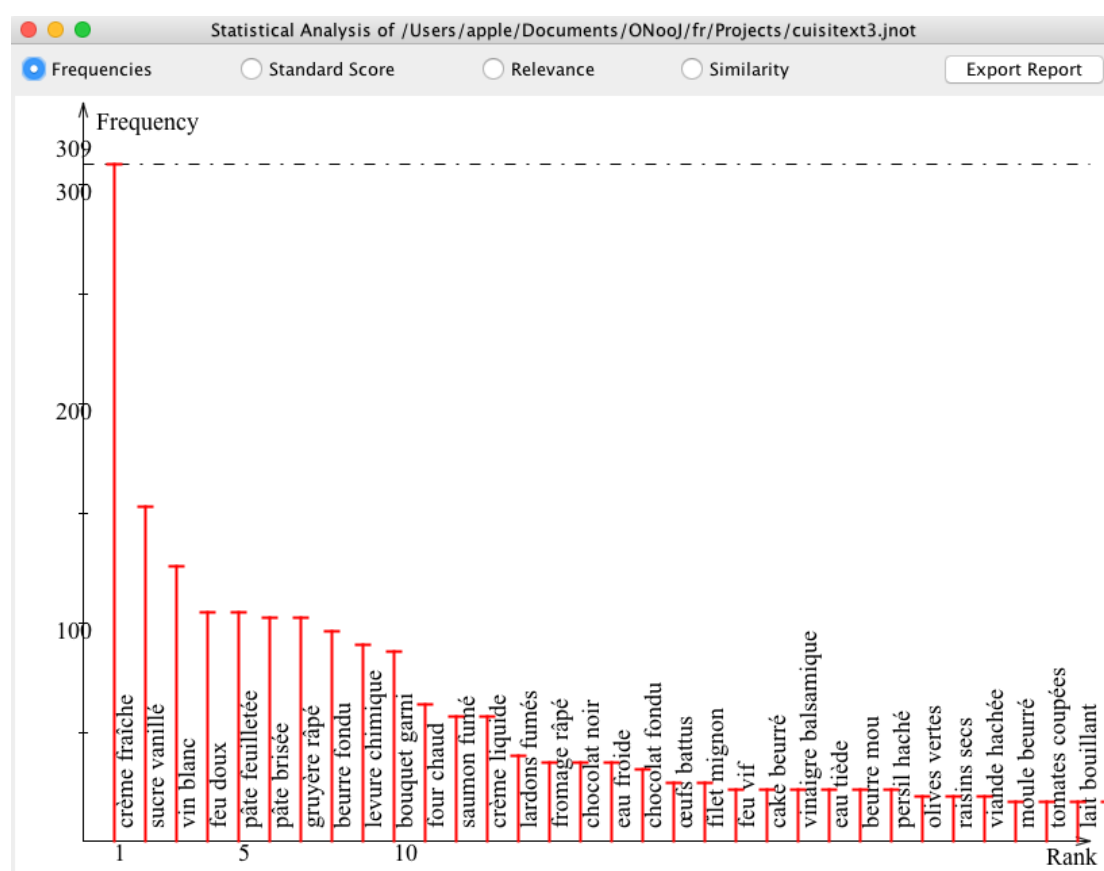


Figure 5-19 : Phrasèmes *NAdj* affichés par fréquence

Le phrasème le plus fréquent dans notre enquête est *crème fraîche* et il se révèle que *crème fraîche* n'est pas considéré comme un terme entier dans les dictionnaires de *NooJ*, *pomme de terre* y est pourtant inscrit comme un terme entier (ex. *pommes de terre paysannes*, *pommes de terre coupées*, *pommes de terre nouvelles*). La fréquence d'emploi d'un phrasème est la totalité des occurrences de ce phrasème et ses « transformations », par exemple, *Feu*

*doux* (25 occurrences, avec majuscule), *feu très doux* (15 occurrences) et *feux doux* (100 occurrences, sans majuscule). Dans notre étude, la limite basse de la fréquence est à trois occurrences, c'est-à-dire que nous retenons tous les phrasèmes qui apparaissent au moins trois fois, car nous constatons que les phrasèmes à faible occurrence ne sont pas tous intéressants pour l'enseignement, par exemple, *recette ultra* (2 occurrences), *pas mauvais* (2 occurrences), *porc pris* (1 occurrence), *général impossible* (1 occurrence), *mal fixé* (1 occurrence).

b. Recherche des phrasèmes pertinents

En se basant sur les connaissances et les expériences provenant du domaine culinaire, nous savons que certains phrasèmes ne sont pas très utilisés dans la cuisine, mais les apprenants sont censés les connaître, par exemple, le phrasème *pâte ferme* n'apparaît qu'une fois dans notre corpus, mais ce phrasème nous apparaît important pour le domaine culinaire. En conséquence, nous ajoutons ce phrasème pertinent dans notre syllabus lexical. À part ce phrasème, d'autres phrasèmes dans le tableau 5-5 ci-dessous nous semblent intéressants.

Peau épaisse	Forme bombée	Poulet mariné	Pommes tranchées
Salade landaise	Taille réduite	Alose farcie	Gaufres moelleuses
Crêpe flambées	Moules marinières	Sucre brun	Biscuit croustillante

Tableau 5-5 : Phrasèmes pertinents à une occurrence

c. Comparaison avec le corpus oral

L'importance du corpus oral a été soulignée dans la sous-partie 1.4.1 et l'utilisation du corpus oral, en comparaison avec le corpus écrit, peut révéler des différences insoupçonnées entre l'oral et l'écrit dans ce cadre professionnel spécifique. Dans notre étude, nous allons comparer *Cuisitext* écrit à *Cuisitext* oral pour compléter notre syllabus lexical. En utilisant *Antconc*, nous pouvons seulement obtenir une liste de mots spécifiques à oral, par



exemple, l'interjection (ex. donc, accord, hein, ça, là, alors). À travers cette liste de mot, nous pouvons également capter des noms ou des adjectifs intéressants, comme l'adjectif *sulfu* encerclé dans la figure 5-20 ci-dessous.

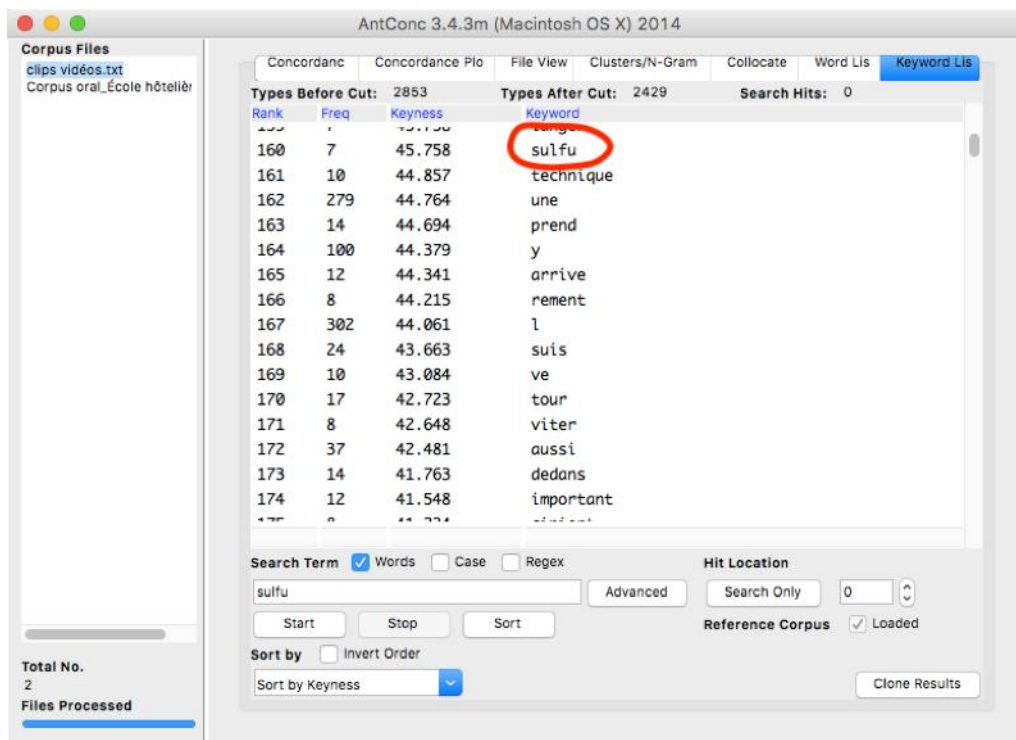


Figure 5-20 : Comparaison de *Cuisitext* écrit avec *Cuisitext* oral

Les phrasèmes *papier sulfu* et *feuille sulfu* s'emploient souvent à l'oral, mais aucune occurrence de *sulfu* ne se trouve dans *Cusitext* écrit, par exemple,

*On va s'assurer de différents éléments. Premièrement, on va mettre une plaque avec du **papier Sulfu**. On a un cercle, propre, et beurré. Beurrer, c'est comme chemiser un moule, ça se fait pas avec le doigt dans le beurre et on le fait comme ça, non. Ça je ne veux pas le voir surtout. (Vidéo 54)*

*Alors une fois qu'on est ici on va faire une petite décoration dessus. Donc une **feuille sulfu** vous aurez besoin de sucre glace, d'une paire de ciseaux. Donc on peut prendre euh un petit morceau de papier comme ceci. (Vidéo 195)*

Bien que ce mot soit absent à l'écrit, les phrasèmes *papier sulfu* et *feuille sulfu* doivent être maîtrisés par nos apprenants. Ces phrasèmes doivent donc être ajoutés dans notre syllabus lexical.

## Bilan

Par l'approche du *corpus-driven*, les phrasèmes les plus saillants du domaine culinaire (phrasèmes *NAdj*) sont mis en évidence. L'extraction des phrasèmes *NAdj* a besoin d'aides informatiques et la modélisation peut aider à choisir un logiciel approprié. Notre modélisation se fait à l'aide d'observations relevées dans les dictionnaires (*le TLFi* et *le Larousse* en ligne) d'extraction de notre corpus *Cuisitext* et d'utilisation des FL. Les trois exigences de nos modélisations lexicales (définition d'une grammaire locale, récursivité de la grammaire et désambiguïsation lexicale) nous conduisent vers un logiciel de formalisation des langues : *NooJ*. En se fondant sur la grammaire algébrique et la grammaire rationnelle, *NooJ*, en tant que système de traitement de corpus, peut répondre parfaitement à nos attentes de modélisation. Après avoir implémenté et lancé les deux structures principales (structure attributive et syntagme nominal) dans *Nooj*, nous pouvons réussir à extraire les phrasèmes *NAdj*. Malgré la désambiguïsation lexicale effectuée dans notre extraction, les phrasèmes *NAdj* extraits ne seront pas tous sélectionnés pour notre étude. Selon l'approche statistique, la fréquence est le critère principal pour l'identification des phrasèmes. Concernant notre étude en FOS, la vérification des phrasèmes de fréquence faible nous permet de ne pas ignorer les phrasèmes pertinents. En nous inspirant de projets précédents, nous avons mis en place trois étapes pour identifier et sélectionner les phrasèmes *NAdj* pour l'enseignement : fréquence des phrasèmes dans *Cuisitext*, recherche des phrasèmes pertinents de fréquence faible, comparaison avec *Cuisitext* oral (la partie orale de *Cuisitext*).

## 6. Analyse linguistique des phrasèmes *NAdj* retenus

Dans la partie précédente, les phrasèmes *NAdj* ont été extraits et sélectionnés pour l'enseignement. Grâce à la réussite de notre extraction, nous avons retenu certaines expressions irrégulières, comme *le caramel qui ne doit pas devenir noir* (Rank 24), *la crème fraîche liquide* (Rank 26), *la pâte soit plus ferme* (Rank 32), *le fromage de chèvre frais* (Rank 49), etc. En résumé, nous avons sélectionné 426 phrasèmes *NAdj* sur 1 million de mots. Suite à la sélection du lexique pour l'enseignement, une classification, notamment linguistique, est souvent demandée pour la préparation des activités didactiques (Tutin, 2014). En outre, Cavalla ajoute que « l'enseignement du lexique est incomplet s'il ne s'appuie pas sur une description préalable de sa structure » (2016 : 28), autrement dit, le manque de classification pourrait prêter à confusion chez les apprenants. Comme expliqué précédemment dans la sous-partie 2.2, les phraséologues établissent leurs propres classifications en fonction des critères retenus. Dans notre projet, nous procéderons de la même manière que les autres phraséologues : chercher nos propres critères nécessaires à la classification de la phraséologie.

### 6.1 Critères servant à la classification

La sélection des critères s'effectue selon l'objectif didactique visé, à savoir l'enseignement du FOS auprès d'apprenants étrangers. Ainsi, au vu du public visé, voici les deux critères que nous retiendrons dans notre étude :

- Figement : d'après Pecman (2004), le figement est la propriété la plus évoquée pour décrire les phrasèmes. En outre, vu que le public de notre étude est constitué d'apprenants étrangers, le premier critère que nous retiendrons est celui du figement. En effet, un apprenant étranger ne peut prévoir la suite d'un phrasème sans la connaître (Cavalla, 2009), par exemple, le phrasème *thon égoutté et émietté* peut être difficilement prévu

par un apprenant étranger et ce phrasème est pourtant maîtrisé inconsciemment par un locuteur natif.

- Critère pragmatique : le critère qui nous semble intéressant en milieu professionnel est le critère pragmatique. Certaines expressions *NAdj*, qui apportent une valeur pragmatique au discours, sont spécifiques au domaine culinaire, par exemple, dans le pragmatème *four chaud*, on ne peut facilement permuter le lexème *chaud* avec d'autres.

G. Gross avait développé cette idée, « le figement n'est pas une valeur absolue mais relève d'une gradation correspondant à des propriétés transformationnelles potentielles réalisées à des degrés différents » (1988 : 63). Non seulement le fait du figement, mais le degré de figement pose aussi problème aux apprenants, par exemple, ces derniers ne sont pas capables de juger si le phrasème *crème fraîche* est plus ou moins figé que le phrasème *plat préféré*. En fait, chaque sorte de phrasème possède ses propres propriétés transformationnelles du figement. Le classement des unités phraséologiques se fait souvent en deux étapes : la première étape consistera d'abord à dresser une liste exhaustive des transformations syntaxiques ; la deuxième étape est de classer les expressions en fonction de leurs degrés de figement (Pecman, 2004). Nous allons d'abord voir les transformations syntaxiques des phrasèmes *NAdj* et ensuite classer les phrasèmes retenus en fonction des degrés de figement.

## 6.2 Propriétés transformationnelles du figement des phrasèmes

### *NAdj*

Pour le au figement des phrasèmes *NAdj*, G. Gross (1988) propose neuf propriétés transformationnelles.

a. Prédicativité

D'après G. Gross (1988), certains adjectifs ne peuvent être prédicatifs, par exemple, *champignon atomique* : \*un champignon est atomique ; *chambre froide* : \*une chambre est froide ; *peur bleue* : \*une peur est bleue. G. Gross souligne que l'impossibilité de la prédicativité est souvent liée à l'emploi métaphorique de l'un des deux éléments.

b. Nominalisation

La nominalisation d'un adjectif est très présente dans le langage ordinaire, par exemple, *les cheveux de Max sont noirs* : *la noirceur des cheveux de Max*. La nominalisation ne s'emploie pas pour tous les adjectifs. Les causes de l'impossibilité de la nominalisation sont multiples et l'une d'entre elles est, d'après G. Gross, la corrélation entre la nominalisation et la prédicativité : « quand l'adjectif n'est pas prédicatif, il n'a pas non plus de nominalisation » (1988 : 65), par exemple, *colère noire* : \* la noirceur de la colère.

c. Restrictions de sélection : rupture paradigmatique

Comme nous le savons, les adjectifs peuvent être utilisés pour qualifier les substantifs, par exemple, *poivron vert*, *poivron rouge*, *poivron jaune* et *voûte est basse*, *voûte est moyenne*, *voûte est élevée*, *voûte est haute*. Dans ces exemples, ces notions de couleur (jaune, blanc et bleu) et de hauteur (basse, moyenne, élevée, haute) sont considérées comme des séries distributionnelles. Il peut également exister la rupture d'une série distributionnelle entre les substantifs et les adjectifs, par exemple, *omelette norvégienne*/\*française/\*écossaise. D'après G. Gross, les causes des contraintes de sélection sont souvent difficiles à établir et la rupture d'une série distributionnelle peut révéler le degré de figement : « plus la série est restreinte par rapport à son extension maximale moins il y a lieu de calculer le sens à partir des éléments » (ibid. : 66), par exemple, *colère noire*, \*colère rouge, \*colère jaune.

d. Variation en nombre

Une expression *NAdj* peut être au pluriel ou au singulier, mais certaines expressions *NAdj* sont obligatoirement au pluriel, par exemple, *eaux usées*, *cordes vocales* et certaines expressions *NAdj* sont obligées de se mettre au singulier, par exemple, *instant présent* et *lexique mental*.

e. Adjonction d'un adverbe

L'emploi de l'adjectif est souvent accompagné de l'adverbe en précisant des « degrés de signification » (superlatif et comparatif) : *une recette très gourmande* ; *un gâteau très moelleux*. Mais l'expression figée n'accepte pas l'adjonction d'un adverbe : ex. *filet mignon*, \**filet très mignon* ; *beurre salé*, \**du beurre très salé* ;

f. Adjonction d'un autre adjectif

Dans une expression libre, un adjectif peut être coordonné avec un autre par une des conjonctions de coordination (*et*, *mais*) : *un livre difficile mais intéressant*, mais cette possibilité n'existe pas dans les expressions figées : *une (étoile filante) superbe* ; *une (chaise longue) délabrée*.

g. Effacement de l'adjectif

Dans une expression libre, l'adjectif n'est pas obligatoire : un livre (difficile), un plat (préféré), mais dans les expressions moins libres comme *eau oxygénée*, *eau sédative*, *eau seconde*, *eau lourde*, l'adjectif n'est pas effaçable, car la suppression d'un adjectif banalise la différence de ces deux produits, par exemple, nous ne pouvons séparer « lourde » de « eau », car l'« eau lourde » est un élément à part entière.

h. Figement du premier terme

Dans une expression *NAdj*, le substantif peut également former un paradigme : (un travail / un devoir / une lecture) difficile. Comme le paradigme formé dans

l'adjectif, la rupture paradigmatique peut intervenir dans le substantif : *beurre noir*, \*margarine noire et \*huile noire.

i. Adjectifs et compléments de nom

Les nominalisations des adjectifs des phrasèmes *NAdj* peuvent également s'appliquer à l'aide des verbes supports *être* ou *avoir* (Meunier, 1981) et le complément de substantif issu de la nominalisation de l'adjectif comprend nécessairement un déterminant, par exemple, *devoir difficile* : \*un devoir de difficulté ; un devoir d'une certaine difficulté. Mais cette transformation n'est pas observable dans les expressions moins libres, par exemple, un professeur de français : \*un professeur a un certain français ; \*un professeur est d'un certain français.

Une fois que nous avons fait l'inventaire de ces neuf propriétés transformationnelles du figement des phrasèmes *NAdj*, nous nous rendons compte que le critère pragmatique retenu dans notre projet correspond à la restriction de sélection, une des neuf propriétés transformationnelles, par exemple, le pragmatème *four chaud* (cf. 6.1 supra) se caractérise par une certaine restriction de sélection lexicale, car le lexème *chaud* est difficilement permutable avec d'autres dans ces contextes. De la sorte, le critère pragmatique est déjà pris en compte dans les propriétés transformationnelles du figement des phrasèmes *NAdj* de G. Gross. En tenant compte du seul critère du figement, notre classification des phrasèmes *NAdj* peut s'établir comme suit.

### 6.3 Classification des phrasèmes *NAdj*

Par le critère de figement, nos phrasèmes sont donc classés en trois catégories : les expressions figées, les expressions semi-figées et les groupes ordinaires. La première est synonyme d'expression idiomatique souvent employée dans la littérature linguistique ; les relations des éléments d'une telle

expression sont figées. Pourtant, dans une expression semi-figée, seul l'un des éléments doit être choisi de façon contrainte en fonction d'autres éléments de cette expression ; dans un groupe ordinaire, tous les éléments sont remplaçables. Les relations des trois catégories de notre projet sont illustrées dans la figure 6-1 ci-dessous.

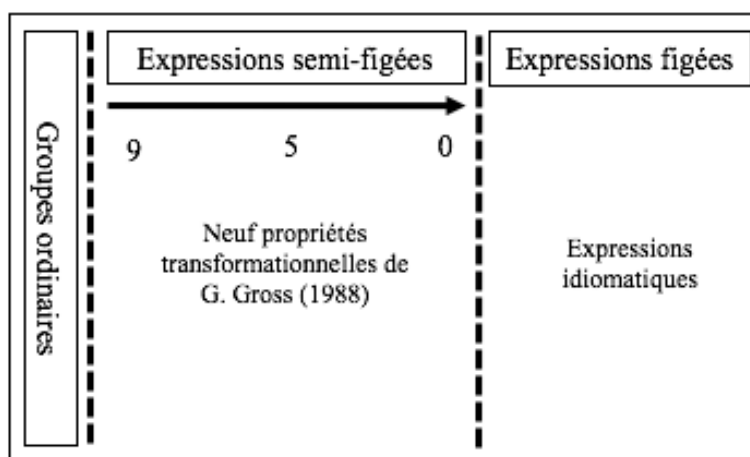


Figure 6-1 : Expressions figées, expressions semi-figées et groupes ordinaires

Afin de préciser le degré de figement, nous nous inspirons du travail de G. Gross (ibid.) dans lequel le degré de figement des phrasèmes *NAdj* est mesuré en fonction de sa déficience transformationnelle sur neuf propriétés transformationnelles proposées. Nos expressions figées ne peuvent subir aucune des neuf transformations et nos expressions semi-figées peuvent accepter une ou plusieurs transformations. D'où le fait que les relations des composantes d'une expression semi-figée soient moins figées que celles d'une expression figée. Les groupes ordinaires sont ceux qui peuvent accepter toutes les transformations de G. Gross et le sens d'un groupe ordinaire est totalement compositionnel.

#### 6.4 Analyse du figement des phrasèmes *NAdj*

En prenant en compte notre classification, les phrasèmes *NAdj* sélectionnés pour l'enseignement seront classés en trois catégories (les expressions figées,



les expressions semi-figées et les groupes ordinaires) en fonction du nombre de critères positifs. À défaut de temps et d'espace, seuls sept phrasèmes seront analysés comme exemple pour illustrer comment les neuf propriétés transformationnelles affectent les phrasèmes *NAdj* : *plat préféré* (tableau 6-1), *crème fraîche* (tableau 6-2), *filet mignon* (tableau 6-3), *plat facile* (tableau 6-4), *oignon haché* (tableau 6-5), *poivron rouge* (tableau 6-6), *température ambiante* (tableau 6-7).

### **Plat préféré**

Prédicativité : plat est préféré	+ <sup>89</sup>
Nominalisation : préférence du plat	+
Restriction de sélection : plat incontournable / favori	+
Variation en nombre : plat préféré / plats préférés	+
Adjonction d'un adverbe : *plat très préféré	- <sup>90</sup>
Adjonction d'un adjectif : *plat préféré et délicieux	-
Effacement d'un adjectif : *plat	-
Figement du premier terme : (assiette + moule + jambon) préféré	+
Adjectif et complément de nom : plat a une certaine préférence plat est d'une certaine préférence	+
Plat préféré	Expression semi-figée

Tableau 6-1 : Transformations de *plat préféré*

Comme indiqué dans le tableau 6-1 ci-dessus, le phrasème *plat préféré* réagit positivement à la plupart des neuf transformations, mais il n'accepte pas l'adjonction d'un adverbe et d'un adjectif : \*plat très préféré et \*plat préféré et délicieux. En conséquence, ce phrasème est une expression semi-figée.

<sup>89</sup> + : réalisable/acceptable

<sup>90</sup> - : irréalisable/inacceptable

**Crème fraîche**

Prédicativité : *crème est fraîche	–
Nominalisation : *fraîcheur de la crème	–
Restriction de sélection : crème fraîche / *chaude / *tiède	–
Variation en nombre : crème fraîche / *crèmes fraîches	–
Adjonction d'un adverbe : *crème très fraîche	–
Adjonction d'un adjectif : *crème fraîche et fermentée	–
Effacement d'un adjectif : *crème	–
Figement du premier terme : (crème+ *huile) fraîche	–
Adjectif et complément de nom : *La crème a une certaine fraîcheur *La crème est d'une certaine fraîcheur	–
Crème fraîche	Expression figée

Tableau 6-2 : transformations de *crème fraîche*

Comme on le voit au tableau 6-2 ci-dessus, le phrasème *crème fraîche* n'accepte pas non plus l'adjonction d'un adjectif ou d'un adverbe : \*crème fraîche et fermentée et \*crème très fraîche. En outre, il n'admet pas l'effacement d'un adjectif, \*crème très fraîche. En effet, la crème fraîche, une sorte de crème, est plus épaisse par rapport à la crème liquide. En raison de sa consistance importante, la crème fraîche est souvent utilisée pour obtenir l'épaisseur souhaitée. C'est pourquoi la crème fraîche ne signifie pas que la crème est fraîche. Puisque la crème fraîche est différente de la crème. La conséquence de l'effacement d'un adjectif est de banaliser la différence de ces deux produits. Au vu du nombre de critères négatifs, ce phrasème est une expression figée.

**Filet mignon**

Prédicativité : *filet est mignon	–
-----------------------------------	---

Nominalisation : mignonnerie du filet	+
Restriction de sélection : *filet canon / *joli / *beau	-
Variation en nombre : filet mignon / filets mignons	+
Adjonction d'un adverbe : *filet très mignon	-
Adjonction d'un adjectif : *crème mignonne et farcie	-
Effacement d'un adjectif : filet	-
Figement du premier terme : (filet + *viande) mignon	-
Adjectif et complément de nom : *le filet a une certaine mignonnerie *le filet est d'une certaine mignonnerie	-
Filet mignon	Expression semi-figée

Tableau 6-3 : Transformation de *filet mignon*

Comme indiqué dans le tableau 6-3 ci-dessus, le phrasème *Filet mignon* n'accepte pas l'adjonction d'un adverbe ou d'un adjectif : \*filet très mignon et \*crème mignon et farci. En outre, le filet mignon dans le TLFi est une pièce de boucherie détaillée à la pointe du filet. Il est donc différent du filet. Il n'est que la pointe. En revanche, ce phrasème accepte les deux transformations suivantes : nominalisation et variation en nombre. Ainsi, le filet mignon est une expression semi-figée.

### ***Plat facile***

Prédicativité : plat est facile	+
Nominalisation : facilité du plat	+
Restriction de sélection : plat (pratique / simple)	+
Variation en nombre : plat facile / plats faciles	+
Adjonction d'un adverbe : plat très facile	+
Adjonction d'un adjectif : plat facile et rapide	+
Effacement d'un adjectif : plat	+

Figement du premier terme : (recette / cuisine / repas) facile	+
Adjectif et complément de nom : plat a une certaine facilité plat est d'une certaine facilité	+
Plat facile	Groupe ordinaire

Tableau 6-4 : Transformations de *plat facile*

Le tableau 6-4 ci-dessus montre que le phrasème *plat facile* se prête à toutes les modifications envisagées. Ce phrasème est donc un groupe ordinaire.

### ***Oignon haché***

Prédicativité : oignon est haché	+
Nominalisation : hachage de l'oignon	+
Restriction de sélection : oignon haché / *émietté / *moulu	-
Variation en nombre : oignon haché / oignons hachés	+
Adjonction d'un adverbe : oignon finement haché	+
Adjonction d'un adjectif : oignon haché et parfumé	+
Effacement d'un adjectif : oignon	+
Figement du premier terme : (oignon + *thym) haché	-
Adjectif et complément de nom Oignon a un certain hachage Oignon est d'un certain hachage	+
Oignon haché	Expression semi-figée

Tableau 6-5 : Transformations d'*oignon haché*

Comme illustré dans le tableau 6-5 ci-dessus, le phrasème *oignon haché* peut accepter la plupart des modifications (ex. prédicativité, nominalisation, variation en nombre, adjonction d'un adjectif et d'un adverbe, complément de nom), mais il subit la restriction de sélection et le figement du premier terme : oignon \*émietté et \*thym haché. Ainsi, ce phrasème est une expression semi-figée.

**Poivron rouge**

Prédicativité : poivron est rouge	+
Nominalisation : rouge du poivron	+
Restriction de sélection : poivron rouge / jaune / vert	+
Variation en nombre : poivron rouge / poivrons rouges	+
Adjonction d'un adverbe : *poivron très rouge	-
Adjonction d'un adjectif : *poivron rouge et émincé	-
Effacement d'un adjectif : poivron	+
Figement du premier terme : (poivron + *patate) rouge	-
Adjectif et complément de nom : Le poivron a un certain rouge Le poivron est d'un certain rouge	+
Poivron rouge	Expression semi-figée

Tableau 6-6 : Transformations de *poivron rouge*

Le tableau 6-6 indique que le phrasème *poivron rouge* supporte la plupart des modifications (prédicativité, nominalisation, restriction de sélection, variation en nombre, effacement d'un adjectif, complément de nom), mais il n'accepte pas l'adjonction d'un adverbe et d'un adjectif : \**poivron très rouge* et \**poivron rouge et émincé*. Ce phrasème est donc une expression semi-figée.

**Température ambiante**

Prédicativité : *température est ambiante	-
Nominalisation : *ambiance de la température	-
Restriction de sélection : *température environnementale	-
Variation en nombre : température ambiante / *températures ambiantes	-
Adjonction d'un adverbe : *température très ambiante	-
Adjonction d'un adjectif : *température ambiante et agréable	-

Effacement d'un adjectif : *température	–
Figement du premier terme : (température + *humidité) ambiante	–
Adjectif et complément de nom : *la température a une certaine ambiance *la température est d'une certaine ambiance	–
Température ambiante	Expression figée

Tableau 6-7 : Transformations de *température ambiante*

Le tableau 6-7 ci-dessus montre que le phrasème *température ambiante* n'accepte aucune transformation et ce phrasème est donc une expression figée.

## Bilan

Une classification des phrasèmes *NAdj* sélectionnés est nécessaire pour procéder à l'enseignement. Selon le public visé (apprenants étrangers), deux critères (figement et critère pragmatique) sont retenus dans notre classification. À propos du figement des phrasèmes *NAdj*, G. Gross propose les neuf propriétés transformationnelles dans lesquelles le critère pragmatique est également pris en compte. Compte tenu de ces propriétés transformationnelles, notre classification se fait en trois catégories : expressions figées, expressions semi-figées et groupes ordinaires. Sept phrasèmes sont analysés comme exemples pour illustrer l'influence des neuf transformations sur les phrasèmes *NAdj* : *plat préféré* (expression semi-figée), *crème fraîche* (expression figée), *filet mignon* (expression semi-figée), *plat facile* (groupe ordinaire), *oignon haché* (expression semi-figée), *poivron rouge* (expression semi-figée), *température ambiante* (expression figée). La petite conclusion que nous pouvons tirer de nos analyses est que les phrasèmes sélectionnés pour cette étude sont des expressions figées, des expressions

semi-figées et des groupes ordinaires. Comme nous l'avons expliqué au début de ce chapitre, notre classification ou notre analyse linguistique sert à l'enseignement des phrasèmes *NAdj*. Les apprenants ont tout intérêt à connaître les transformations réalisables et irréalisables d'un phrasème. Par exemple, lors de l'E/A du phrasème *oignon haché*, les enseignants peuvent faire prendre conscience aux apprenants des transformations irréalisables : \*oignon émietté / \*oignon moulu (restriction de sélection), \*thym haché (figement du premier terme) et des transformations réalisables : oignon est haché (prédictivité), hachage de l'oignon (nominalisation), oignon haché et parfumé (adjonction d'un adjectif), oignons hachés / oignon haché (variation en nombre), oignons finement hachés (adjonction d'un adverbe).

---

Troisième partie : proposition didactique



<b>7. Propositions des approches d'intégration du corpus et leurs mises en œuvre pour l'E/A des phrasèmes <i>NAdj</i> .....</b>	<b>153</b>
<b>7.1 Trois approches proposées.....</b>	<b>155</b>
<b>7.2 Piste d'apprentissage progressif sur corpus .....</b>	<b>157</b>
<b>7.3 Propositions didactiques.....</b>	<b>159</b>
7.3.1 Deux séances de formation .....	160
7.3.1.1 Première séance autour de la lexie HACHÉ .....	160
7.3.1.2 Deuxième séance à partir des champs sémantiques de HACHÉ.....	168
7.3.2 Conseils pour la mise en pratique de l'AIP.....	179
<b>Bilan .....</b>	<b>180</b>

## 7. Propositions des approches d'intégration du corpus et leurs mises en œuvre pour l'E/A des phrasèmes *NAdj*

Comme nous l'avons vu dans la sous-partie 3.2.1, l'intégration tardive du corpus pour l'E/A des langues peut être attribuée à plusieurs facteurs : disponibilité des corpus oraux (Chambers, 2009), investissement en temps (ibid.), aspect culturel (Boulton, 2008b), manque de résultats positifs sur l'efficacité de l'apprentissage sur corpus (ibid.) et manque de compétence supplémentaire pour l'utilisation du corpus pour l'E/A (Kübler, 2014). Ce dernier peut concerner deux types d'usages du corpus pour l'E/A des langues comme mentionné dans la sous-partie 3.2.2 : *teaching to exploit* et *exploiting to teach* (Fligelstone, 1993).

- Dans le premier (*teaching to exploit*), les apprenants sont censés être en mesure d'utiliser l'outil d'exploitation du corpus à leur guise au service de leurs apprentissages. Ajoutons à cela que cette compétence d'utilisation du corpus est également demandée chez les enseignants (Tran, 2014 ; Kübler, 2014). Comme le dit Tran, « pour que les apprenants apprécient l'apprentissage sur corpus, il faudrait avant tout que les enseignants prennent plaisir à la découverte qu'apporte le travail sur corpus, et qu'ils prennent conscience de l'importance de l'enseignement de certains phénomènes linguistiques » (2014 : 216). Cette compétence permet aux enseignants non seulement de prendre en compte les éléments linguistiques saillants dans n'importe quel domaine de spécialité, mais aussi de mieux proposer des activités didactiques. Mais la difficulté de développer cette compétence réside dans le fait que les côtés technique et méthodologique de la maîtrise des outils d'exploitation du corpus font peur à certains apprenants voire aux enseignants (Zanettin, 2002 ; Tyne, 2012).
- Le deuxième usage *exploiting to teach* concerne les façons d'introduire le corpus aux apprenants et les trois grands courants sont recensés dans la

sous-partie 3.2.3 : DDL de Johns (1991a ; 1991b), ASC de Boulton & Tyne (2014) et Consultations Directe/Indirecte de Chambers (2010). Après avoir étudié les trois grands courants d'utilisation du corpus pour l'E/A, nous nous rendons compte que ces types d'introduction du corpus sont hybrides, voire confus. Cette confusion proviendrait du chevauchement des trois grands courants comme montré indiqué le tableau 7-1 ci-dessous.

Courants d'introduction du corpus aux apprenants	Approches	
Approche DDL (Johns, 1991a ; 1991b)	Inductive	
Approche ASC (Boulton & Tyne, 2014)	Déductive	Inductive
Consultation du corpus (Chambers, 2010)	Indirecte	Directe

Tableau 7-1 : Trois courants d'introduction du corpus aux apprenants

Comme indiqué dans le tableau 7-1 ci-dessus, l'approche inductive est préconisée par l'approche DDL de Johns (1991a ; 1991b), mais d'après Boulton et Tyne (2014), le corpus peut être introduit aux apprenants par les approches inductive et déductive. En outre, Chambers (2010) propose deux manières de consultation : la consultation directe et la consultation indirecte. Dans la consultation indirecte, l'enseignant peut faire découvrir les régularités d'un élément langagier à partir de séquences présélectionnées aux apprenants eux-mêmes par l'approche inductive ou bien illustrer des règles données à l'aide d'exemples authentiques du corpus par l'approche déductive. Dans la consultation directe, les apprenants sont capables de conceptualiser et de systématiser les règles du fonctionnement qu'ils ont découvertes. Puisque le choix du type d'introduction du corpus aux apprenants peut déterminer la mise en pratique de l'activité langagière, la confusion ou ambiguïté de ces approches pourrait mystifier l'utilisation du corpus ou freiner l'intégration du corpus pour l'E/A des langues. En vue de démystifier l'utilisation du corpus pour l'E/A des langues, nous décidons de proposer des approches plus précises de consultation du corpus.

## 7.1 Trois approches proposées

Nous proposons trois approches d'introduction des corpus aux apprenants dans le tableau 7-2 ci-dessous : Approche Déductive (AD), Approche Inductive Guidée (AIG) et Approche Inductive Pure (AIP).

Courants d'introduction du corpus aux apprenants	Approches		
Approche DDL (Johns, 1991a; 1991b)	Inductive		
Approche ASC (Boulton & Tyne, 2014)	Déductive	Inductive	
Consultation du corpus (Chambers, 2010)	Indirecte		Directe
Trois approches (Yang, 2018b)	AD	AIG	AIP

Tableau 7-2 : Proposition de trois approches d'introduction du corpus

La proposition des trois approches d'introduction du corpus remet en question les rôles des apprenants et de l'enseignant ainsi que les supports utilisés dans la mise en place des activités langagières.

- Dans l'AD, l'enseignant explicite directement les règles de la langue aux apprenants. Comme dans l'enseignement traditionnel, l'enseignant est transmetteur du savoir, alors que les apprenants jouent le rôle de « consommateur » des « produits » d'apprentissage livrés par l'enseignant « fournisseur » (Holec, 1991 : 2). Dans cette approche, les séquences sont déjà didactisées par l'enseignant pour montrer aux apprenants les règles ou vérifier leurs connaissances acquises par des exercices lacunaires ou exercices d'appariement.
- Dans l'AIG, l'enseignant fait découvrir les régularités sous-jacentes de la langue aux apprenants. Dans ce processus d'apprentissage, l'apprenant a un rôle de découvreur en suivant les consignes de l'enseignant et ce dernier joue un rôle de « facilitateur ou médiateur »

(Kübler, 2014 : 44) en fournissant les séquences présélectionnées. Il faut savoir que, dans cette approche, les séquences fournies ne sont pas trop complexes, sinon elles pourraient compliquer l'interprétation par les apprenants.

- Dans l'AIP, comme dans la consultation directe, les apprenants utilisent les séquences vierges (non traitées) pour déduire des régularités ou/et des exceptions de la langue cible et l'enseignant n'intervient presque pas dans cet apprentissage. Dans une telle approche, le rôle réservé aux apprenants est donc « détective linguistique » (Johns, 1997 : 101) et les apprenants se trouvent « dans une posture réflexive, rompant ainsi avec une attitude passive consistant à recevoir et à (plus ou moins bien) utiliser une forme trop peu questionnée » (Boch & Buson, 2012 : 47), car pour formuler les règles du langage, les apprenants doivent observer, faire des hypothèses et interpréter les résultats obtenus pour voir « what is going on » (Hanks, 2008 : 220). Le rôle dévolu à l'enseignant est « guide et mentor » (Willis, 2003 : 167) et il n'intervient que pour guider les apprenants vers le repérage des propriétés de la langue, si c'est nécessaire.

Si l'on compare cette approche à celle de Boulton & Tyne, on peut en déduire qu'ils estiment que l'apprentissage par l'AIG est quelquefois plus intéressant que celui par l'AIP : « le travail direct sur le corpus électronique est parfois perçu comme étant moins intéressant ou efficace que le travail avec des données présélectionnées par l'enseignant et appuyées par des activités ciblées » (2014 : 169). Toutefois, d'après nous, aucune des trois approches n'est meilleure que les deux autres. En effet, se pose déjà la question de la préférence de l'approche de l'apprentissage sur corpus chez plusieurs chercheurs (entre autres, Rézeau, 1997 ; 2001 ; Colletta, 2004). Hofstede (2006) précise que les apprenants ont l'habitude de consommer les produits fournis par l'enseignant et qu'ils n'apprécieront pas l'approche inductive. Les

chercheurs taiwanais (Lee & Liou, 2003 ; Chan & Liou, 2005 ; Yeh *et al.*, 2007) constatent que le DDL fonctionne mieux pour les apprenants ayant une préférence inductive. En fait, la préférence de l'approche touche à la question de la culture éducative (Harmer, 2001), de l'expérience passée (Yoon, 2008) et de la motivation (Chambers, 2005) des apprenants. Afin de profiter de tous les avantages des approches d'utilisation du corpus pour l'E/A, certains chercheurs (entre autres, Tran, 2014) proposent de combiner les approches inductive et déductive. Rappelons (cf. introduction) que l'une des contraintes des apprenants étrangers en FOS est le temps limité consacré à leur apprentissage (Mangiante et Parpette, 2004 ; Carras *et al.*, 2007). De ce fait, l'enseignant doit s'efforcer d'autonomiser les apprenants dans une formation de FOS pour que ces derniers puissent s'autoformer pour leurs vies professionnelles ultérieures. En effet, l'autonomie de l'apprenant est un concept qui est pris en considération dans le CECRL et ce concept consiste principalement à permettre à l'apprenant de devenir « maître des modalités de réalisation des actes d'apprentissage (il réalise les activités proposées à son heure, à son rythme, dans le lieu de son choix), et, en partie, de la progression de l'apprentissage » (Holec, 1991 : 60). Rappelons (cf. 3.2.3.1 supra) que l'AIP prend sa source dans le constructivisme et l'apprentissage constructiviste tend à développer un apprentissage autonome. Nous proposons donc une piste d'apprentissage progressif sur corpus à suivre pour aider à développer l'autonomie des apprenants, c'est-à-dire que les apprenants peuvent procéder à leur apprentissage par l'AIP.

## 7.2 Piste d'apprentissage progressif sur corpus

Le développement de l'autonomie consiste à la mutation des rôles des intervenants : l'apprenant devient progressivement responsable et indépendant face à son apprentissage et l'enseignant abandonne petit à petit le rôle dominant. Le fait de la disparition du guidage permet à l'apprenant de

devenir de plus en plus autonome. L'objectif de notre proposition est donc d'entraîner les apprenants à auto-diriger leurs apprentissages et de diminuer le guidage de l'enseignant. Notre proposition combine nos trois nouvelles approches et commence par l'AIG. Les activités de cette approche suivent souvent la logique suivante : l'observation, la classification et la généralisation. Dans une telle approche, les apprenants s'entraînent à formuler eux-mêmes les règles de la langue en consultant les concordances présélectionnées par l'enseignant. Ensuite, cette AIG est suivie de l'AD, parce que les règles formulées par les apprenants à l'issue de la première approche peuvent être confirmées et consolidées à travers les séquences didactisées par l'enseignant (ex. exercices lacunaires ou exercices d'appariement). Dans les deux premières approches, l'enseignant offre des séquences travaillées et l'apprenant suit le guidage (les consignes) de l'enseignant. L'utilisation des deux approches (AIG et AD) peut être récursive comme indiqué dans la figure 7-1 ci-dessous.

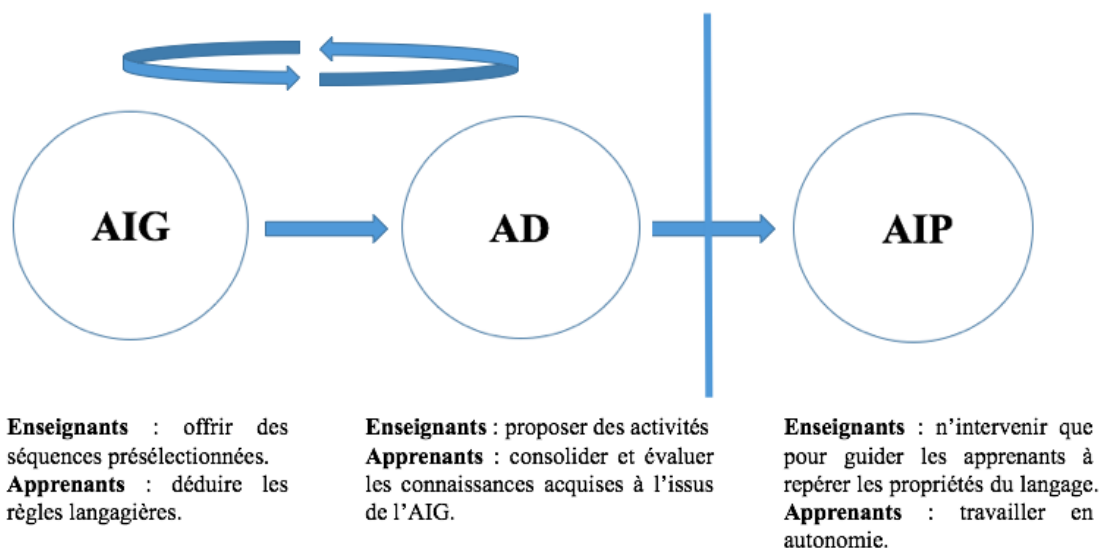


Figure 7-1 : Piste d'apprentissage progressif sur corpus

Cette récursivité, en tant qu'entraînement, va s'appliquer jusqu'au moment où les apprenants peuvent capter un élément langagier saillant, classer des séquences qui leur sont pertinentes et générer eux-mêmes des régularités de la langue. Un entraînement adéquat acquis est l'une des conditions

nécessaires pour procéder à l'AIP (consultation directe). Comme le confirme Chambers, « la consultation indirecte peut servir d'introduction à la consultation directe » (2010 : 15). En procédant à l'AIP, les apprenants peuvent auto-diriger leurs apprentissages, c'est-à-dire qu'ils peuvent être en mesure de gérer les méthodologies et les stratégies de leurs apprentissages. L'apprentissage autonome ne veut pas dire l'apprentissage sans enseignant. L'enseignant intervient pour guider les apprenants à repérer des propriétés linguistiques et il reste l'accompagnateur de l'apprentissage des apprenants. Comme l'affirme Di Vito (2013), selon ses différentes expériences, l'apprenant est capable « de devenir chercheur et artisan de son apprentissage ». Les approches proposées d'utilisation du corpus peuvent être utilisées pour enseigner/apprendre les éléments linguistiques saillants, que ce soit en termes d'éléments lexicaux ou de règles grammaticales. Dans le cadre de cette étude, nous allons proposer des activités en respectant notre cheminement d'apprentissage proposée pour l'E/A des phrasèmes *NAdj*.

### 7.3 Propositions didactiques

Selon la progression didactique que nous proposons pour l'E/A à l'aide de corpus numériques, les activités des deux approches (AIG et AD) peuvent être mobilisées comme des exercices d'entraînement pour pouvoir réaliser l'apprentissage par l'AIP. Afin de guider les apprenants dans l'utilisation du corpus (même pour les enseignants), nous allons proposer deux séances successives dans la sous-partie 7.3.1 et dans chacune des deux, l'activité de l'AIG est suivie par celle de l'AD, comme indiqué dans notre circuit proposé dans la figure 7-1. L'apprentissage par l'AIP consiste principalement en l'apprentissage individualisé : nous ne pouvons rassembler toutes les stratégies d'apprentissage de chaque individu, mais nous donnons également des conseils pour la réalisation de cette approche dans la sous-partie 7.3.2.



### 7.3.1 Deux séances de formation

#### 7.3.1.1 Première séance autour de la lexie HACHÉ

##### A. Cadre de la séquence

- Niveau du public

Apprenants étrangers de niveau A2

- Choix des types d'exercices

Dans l'AIG, les apprenants déduisent eux-mêmes les régularités sous-jacentes de la langue par une série d'activités : l'observation, la classification et la généralisation et le fait de construire les cartes mentales suit également ce principe. Par conséquent, pour la mise en œuvre de l'AIG, les apprenants seront amenés à construire leurs propres cartes mentales. S'agissant de la mise en place de l'AD, l'activité *tableau à remplir* sera mobilisée pour que les apprenants confirment et consolident les connaissances acquises dans la première approche.

- Choix d'un mot-clé

Le mot-clé choisi devrait être l'un des mots représentatifs de la cuisine. Parmi les phrasèmes sélectionnés pour l'enseignement, le lexème HACHÉ est pris comme mot-clé de cette séance, car d'une part, ce lexème n'est pas maîtrisé par nos apprenants de niveau A2 selon le CECRL tant au niveau syntaxique que sémantique, et d'autre part, les ingrédients qualifiés par ce lexème peuvent être facilement classés dans plusieurs catégories et celles-ci permettent de construire aisément différentes cartes mentales.

##### B. Déroulement

###### a. Construction de la carte mentale

- Préparation de l'activité

L'extraction des concordances et le travail d'épuration

Grâce à un concordancier (cf. 3.3.1 supra), l'enseignant peut sortir le lexème HACHÉ avec ses contextes (contextes gauche et droit) dans *Cuisitext*. En intégrant des astérisques dans notre enquête d'extraction (haché\*\*) dans *Antconc*, l'enseignant peut enlever simultanément toutes les séquences qui contiennent ce lexème et ses formes fléchies (*hachée, hachés, hachées*) comme indiqué dans la figure 7-2 ci-dessous.

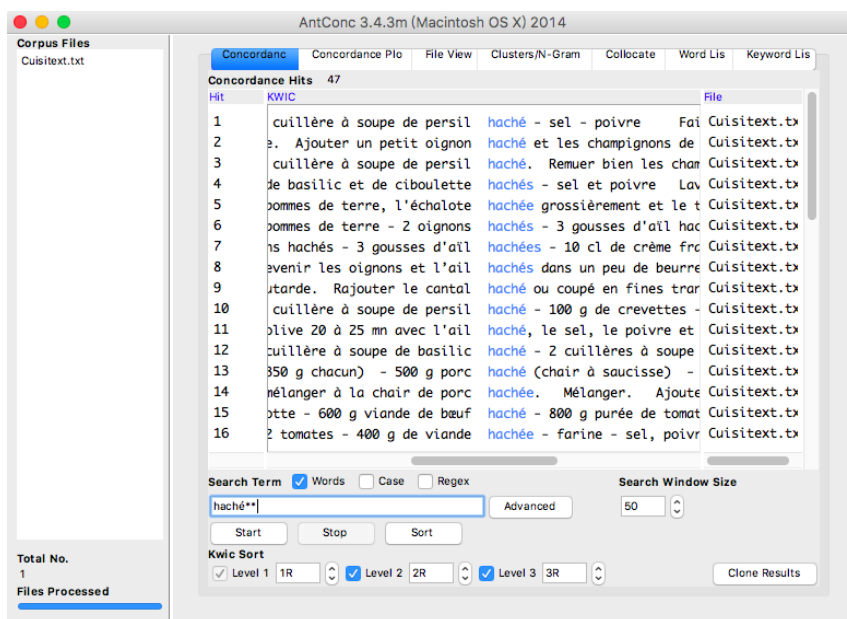


Figure 7-2 : Concordances de HACHÉ de *Cuisitext*

Une fois les séquences corrélatives extraites, l'enseignant a besoin de procéder à un « nettoyage », c'est-à-dire d'éliminer « les occurrences qui risqueraient de compliquer l'interprétation des concordances, voire d'induire en erreur nos étudiants de FLE » (Di Vito, 2013 : 164).

✓ Dispositif pédagogique

- l'enseignant forme des groupes afin que les apprenants puissent confronter et échanger leurs idées à propos du sens et des constructions de ce lexème en contexte.

- chaque groupe reçoit des concordances diverses.

✓ Consignes de l'exercice

- Repérer les mots<sup>91</sup> hachés (hachée, hachées, hachés) dans les phrases ci-dessous.
- Souligner les noms situés après ce terme dans les phrases.
- Observer et analyser les sens des noms que vous soulignez.
- Classer ces éléments soulignés dans différentes listes selon la catégorie d'ingrédients à laquelle ils appartiennent.

Les exemples de phrases ci-dessous sont issus de *Cuisitext* :

1. « Ajouter un petit oignon **haché** et les champignons de Paris coupés en tout petits morceaux. »
2. « Saler, poivrer, saupoudrer d'une cuillère à soupe de persil **haché**. »
3. « Lorsqu'il commence à grésiller, ajouter le cresson, les pommes de terre, l'échalote **hachée** grossièrement. »
4. « Faire revenir les oignons et l'ail **hachés** dans un peu de beurre à feu doux. »
5. « Rajouter le cantal **haché** ou coupé en fines tranches, égaliser. »
6. « Former des galettes avec le poulet **haché** et les déposer sur une feuille de papier d'aluminium huilée. »
7. « Par dessus la couche de pommes de terre, étaler la moitié de la viande **hachée** avec la tomate. »
8. « Sur une moitié, étaler le jambon **haché**, puis les champignons, le fromage, et enfin les œufs. »
9. « Ajouter le boursin, la ciboulette **hachée** et le jus de citron. »
10. « D'autre part, faire revenir le bœuf **haché** dans l'huile d'olive avec l'oignon, l'ail, les herbes. »

---

<sup>91</sup> Nous employons *mot* à la place de *lexème* dans cette consigne pour que les consignes soient compréhensives pour les apprenants non spécialistes de langue.

11. « Les arroser du jus des citrons, les parsemer d'estragon **haché**, saler et poivrer. »
12. « Et pour finir 2 c. à café de feuilles d'aneth **hachées**. »
13. « Ajouter le jambon, le gruyère, les olives et le persil ou les herbes hachées. »
14. « Ajoutez oignon et poireaux **hachés**, carottes, échalotes. »
15. « Réaliser une tarte avec de la viande de veau hachée au lieu du steak haché que je mets habituellement. »
16. « Cette brochette de viande d'agneau **hachée** est un grand classique ! »
17. « Tu verses sur le basilic **haché** un filet d'huile d'olive, suffisamment pour lier le tout, mais pas trop quand même : il ne faut pas que ça baigne. »
18. « L'hiver, du gingembre **haché**, du miel et du citron forment une tisane revigorante. »
19. « Garnissez avec la ciboulette **hachée** et servez immédiatement. »
20. « Ajouter ensuite du boulgour, des feuilles de coriandre et de menthe hachées, et cuire une vingtaine de minutes. »

✓ Mise en commun :

Après avoir observé, analysé et classé, les trois catégories du classement peuvent être présentées dans le tableau 7-3 ci-dessous.

Trois catégories	Exemples
Aromate frais	Persil, oignon, ail, basilic, échalote, ciboulette, gingembre, poireau, menthe, estragon, fenouil, feuilles d'aneth ;
Viande	Porc, bœuf, jambon, veau, agneau, poulet, steak ;
Autres	concombre, chocolat, cantal, saumon.

Tableau 7-3 : Trois catégories de HACHÉ

Dans la culture française, l'ail, le persil, l'oignon et l'échalote sont des aromates, nous les classons donc dans la première catégorie. Les types de

viandes constituent la deuxième catégorie. Outre les viandes et les aromates, HACHÉ peut également qualifier le *concombre* (légume), le *chocolat* (fève de cacao), le *cantal* (fromage), le *saumon* (poisson) qui sont dans des catégories très éloignées des précédentes. C'est la raison pour laquelle nous les regroupons dans une troisième catégorie qui devient générale, contrairement aux autres qui sont très spécifiques à des types d'aliments. Il faut expliquer aux apprenants que les aliments repris dans la troisième catégorie ne sont pas autant utilisés que ceux qui se trouvent dans les deux premières catégories. Il importe de laisser du temps aux apprenants après la proclamation des réponses pour qu'ils puissent compléter et corriger leurs propres listes de catégories. Après cette mise en commun, les apprenants peuvent commencer à établir vraiment leurs propres cartes mentales.

✓ Phase de construction

La construction de la carte mentale permet non seulement de vérifier si les apprenants comprennent bien chaque lexie ainsi que les liens syntaxiques et sémantiques existant entre ces lexies, mais aussi de les aider à les mémoriser, car comme expliqué dans la sous-parite 3.3.2.2, le geste de construire la carte mentale aide à la mémorisation du lexique. Nous pouvons construire trois cartes mentales selon les trois catégories dégagées précédemment autour du lexème HACHÉ. L'enseignant donne des consignes précises et compréhensibles de fabrication d'une carte mentale aux apprenants : les apprenants peuvent commencer par le mot-clé HACHÉ et ajouter des éléments potentiels dans les branches ; le sens des flèches correspond à l'ordre des constructions ; les apprenants peuvent choisir différentes couleurs pour les branches d'une carte mentale afin de distinguer les catégories d'ingrédients, autrement dit, la couleur des branches représente le sens de la catégorie.

Les éléments de la catégorie des aromates frais pourraient être ajoutés autour du mot-clé dans la figure 7-3 ci-dessous réalisé à l'aide du logiciel *CmapTools*.

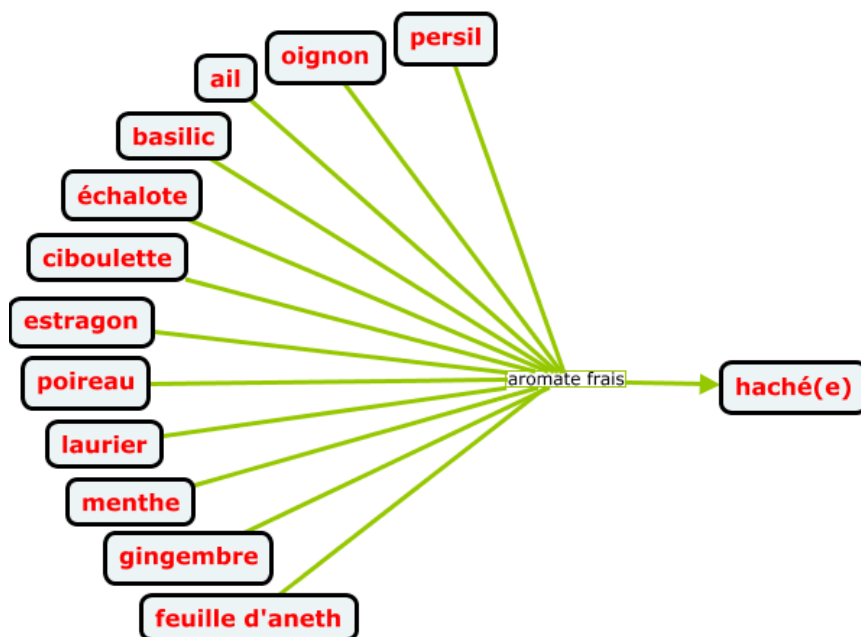


Figure 7-3 : Carte mentale de HACHÉ avec des aromates différents

Le sens des flèches indique l'ordre syntaxique et donc la combinatoire : *persil haché* ; *ail haché*, etc. Nous choisissons le vert pour la couleur des branches de cette catégorie.

Les éléments de la catégorie des viandes pourraient être affichés autour de HACHÉ dans la figure 7-4 ci-dessous :

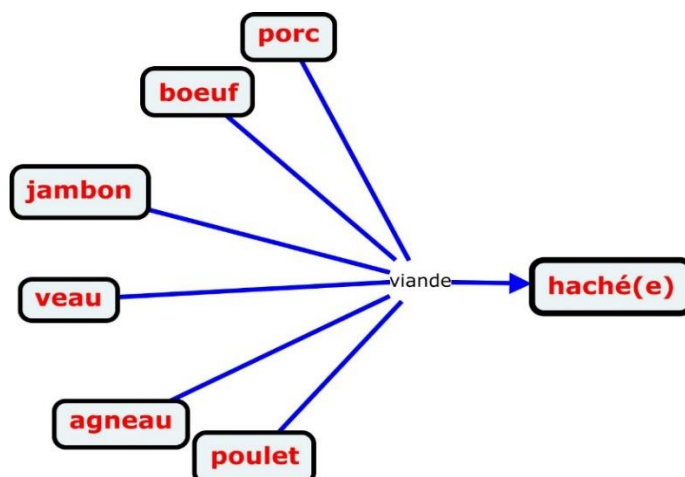


Figure 7-4 : Carte mentale de HACHÉ avec des viandes différentes

Le sens des flèches indique l'ordre syntaxique : *agneau haché*, *boeuf haché*, etc. nous utilisons le bleu pour la couleur des branches de cette liste.

La dernière catégorie peut être représentée de la sorte dans la figure 7-5 ci-dessous :

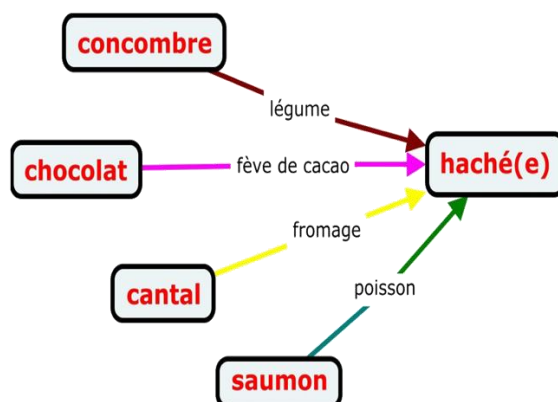


Figure 7-5 : Carte mentale de HACHÉ avec des ingrédients différents

De même, le sens des flèches indique l'ordre syntaxique : *concombre haché*, *chocolat haché*, *cantal haché*, *saumon haché*. Nous pouvons adopter différentes couleurs pour les branches de cette carte mentale afin de la distinguer des autres.

Nous pouvons rassembler, à la fin de la séance, les trois cartes mentales sur le lexème HACHÉ pour que les apprenants aient une vision globale de l'utilisation de cet adjectif, comme dans la figure 7-6 ci-dessous :

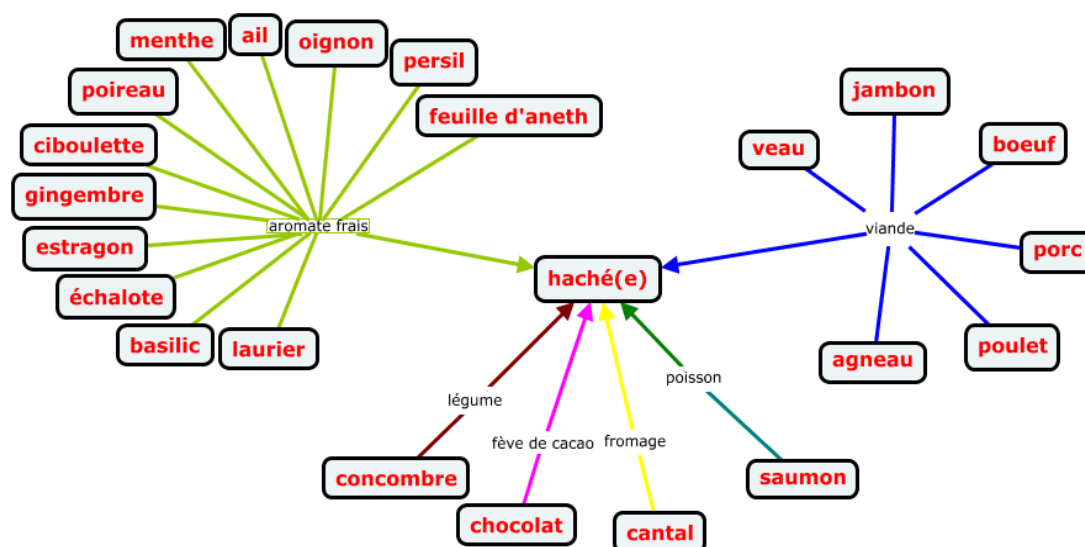


Figure 7-6 : Carte mentale rassemblée de HACHÉ

Dans cette figure 7-6 ci-dessus, nous pouvons voir clairement que l'adjectif HACHÉ qualifie principalement des viandes et des aromates et que ce lexème peut également être employé avec d'autres ingrédients, comme le concombre, le chocolat, le cantal et le saumon mais de façon marginale. Nous ne nous attendons pas à ce que les apprenants dessinent une carte identique à la nôtre, parce qu'ils peuvent choisir d'autres formes, d'autres couleurs, mais nous espérons qu'ils représentent les mêmes logiques.

Cet exercice par l'AIG a pour objectif d'aider les apprenants à découvrir l'emploi du lexème HACHÉ et par la suite, après l'explication donnée par l'enseignant (le lexème HACHÉ qualifie principalement les viandes et les aromates frais), les apprenants sont amenés à consolider et confirmer les règles acquises dans l'activité suivante par l'AD.

b. Tableau à remplir

- ✓ Dispositif :

Un travail individuel

- ✓ Consignes de l'exercice :

Remplissez le tableau 7-4 ci-dessous avec les mots suivants : ail, bœuf, agneau, jambon, échalote, basilic, ciboulette, viande, persil, aromate.

Catégories	Ingrédients	
		HACHÉ
	poulet	




Tableau 7-4 : Tableau à remplir

## ✓ Explication

Un travail individuel sera proposé en vue de vérifier l'acquis de chaque apprenant. Cette activité permet aux apprenants non seulement de ranger les ingrédients en deux catégories (aromate frais : ail, échalote, basilic, ciboulette, persil ; et viande : bœuf, agneau, jambon, etc.), mais aussi de donner une occasion de mémoriser ces phrasèmes (ail haché, viande hachée...).

## 7.3.1.2 Deuxième séance à partir des champs sémantiques de HACHÉ

## A. Cadre de la séquence

- Niveau du public

Apprenants étrangers de niveau A2

- Choix des types d'exercice

L'activité *carte mentale à compléter* sera choisie comme activité de la mise en œuvre de l'AIG et l'enseignant fera travailler les apprenants sur un exercice lacunaire pour la mise en pratique de l'AD.

- Choix d'un mot-clé

Rappelons (cf. 3.1.3 supra) que, que ce soit en L1 ou L2, la mémorisation des mots s'organise principalement autour du sens. Afin d'aider à la mémorisation des phrasèmes sélectionnés, nous décidons d'organiser cette séance autour de l'entrée onomasiologique (cf. 3.1.6 supra) qui va du sens à la forme. Nous prenons comme point de départ les expressions sémantiquement proches. Suite à la première séance, dans laquelle le lexème HACHÉ constitue le mot-clé, cette séance va s'organiser à partir des champs sémantiques de HACHÉ. Parmi les phrasèmes sélectionnés, les adjectifs ayant un sens

proche de HACHÉ seront pris en considération dans notre choix : MOULU, RÂPÉ, ÉMIETTÉ et BRISÉ. En outre, ces lexèmes ne sont que rarement maîtrisés par les apprenants de niveau A2.

## B. Déroulement

### a. Développement de la carte mentale

#### ✓ Préparation de l'activité

L'enseignant peut toujours extraire les concordances de chaque lexème (ÉMIETTÉ, RÂPÉ, MOULU et BRISÉ) en formulant chaque enquête (émietté\*\*, râpé\*\*, moulu\*\* et brisé\*\*) dans *Antconc*. En intégrant des slashes droits dans notre enquête d'extraction (émietté\*\*|râpé\*\*|moulu\*\*|brisé\*\*), l'enseignant peut enlever simultanément toutes les concordances de ces lexèmes et leurs formes fléchies (émietté, émiettée, émiettés, émiettées, râpé, râpée, râpés, râpées, moulu, moulue, moulus, moulues, brisé, brisée, brisés et brisées) comme indiqué dans la figure 7-7 ci-dessous.

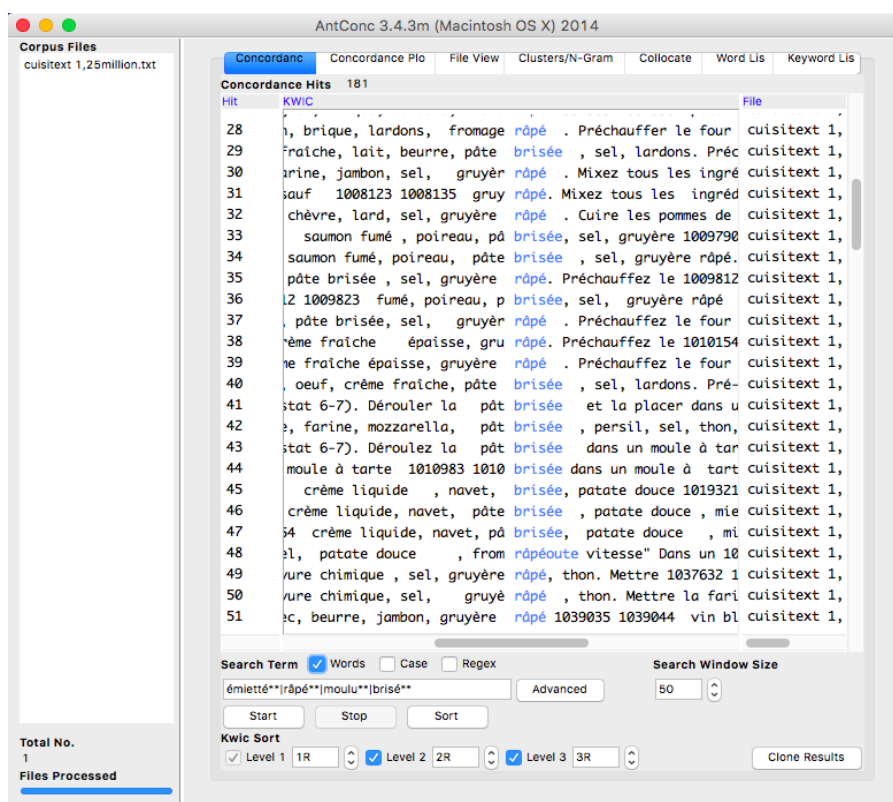


Figure 7-7 : Concordances de MOULU, RÂPÉ, ÉMIETTÉ et BRISÉ de *Cuisitext*

Bien entendu, avant de distribuer les séquences extraites aux apprenants, l'enseignant doit effectuer un « nettoyage » comme nous l'avons fait dans la séance précédente.

✓ Dispositif pédagogique

- l'enseignant forme des groupes
- chaque groupe reçoit des concordances diverses

✓ Consignes de l'exercice :

- Compléter la carte mentale de base suivante (cf. figure 7-8 ci-dessous).

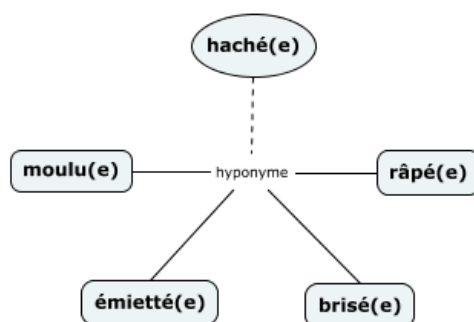


Figure 7-8 : Carte mentale de base

- Repérer les mots émietté, émiettée, émiettés, émiettées, râpé, râpée, râpés, râpées, moulu, moulue, moulus, moulues, brisé, brisée brisés et brisées dans les phrases ci-dessous.
- Souligner les noms situés avant ce terme dans les phrases.
- Observer et analyser les sens des noms que vous soulignez.
- Classer ces éléments soulignés dans différentes listes selon la catégorie d'ingrédients à laquelle ils appartiennent.

Ces exemples de phrases sont issus de *Cuisitext* :

1. « Mélanger le zeste avec le jus de l'orange pressé, la coriandre **moulue** et l'huile. »
2. « Incorporer le thon **émietté** et égoutté, les olives et le fromage. »
3. « Verser la farine, le sucre et la cannelle **moulue**. »

4. « Salez, poivrez et mélangez. Ajoutez la lotte émietée. »
5. « Ajouter la farine, la levure, le gingembre moulu et le bicarbonate de soude. »
6. « Hors du feu, ajouter le parmesan râpé et mélanger rapidement. »
7. « Mélangez les ingrédients secs : la farine, le sucre, le sel, la levure et les amandes moulues. »
8. « Savourez ce Cake au jambon et comté râpé. »
9. « Bien mélanger le beurre, le sucre, la noix moulue, l'œuf, la levure chimique et le sucre. »
10. « Dans une poêle, faire chauffer l'huile d'olive et faire revenir l'aubergine râpée avec l'ail haché. »
11. « La cacahuète moulue sert de base à des sauces, des flans, tartes. »
12. « Tout dépend si vous avez besoin d'utiliser la morue émietée cuite ou crue. »
13. « Faire suer les oignons hachés finement dans un peu de beurre et y ajouter les noix de cajou moulues. »
14. « Recouvrir avec la chair de homard émietée et décorer avec le thym. »
15. « Bien remuer pour obtenir une ganache homogène et y ajouter les pistaches moulues. »
16. « Disposer une couche de lieu noir émieté. »
17. « Mettre la crème à chauffer sur feu moyen dans une casserole avec le gruyère râpé et l'ail finement haché. »
18. « Répartir la crème de concombre râpé dans les verrines, ajouter les fruits de mer marinés et parsemer de feuilles de cerfeuil. »
19. « Mélanger la betterave râpée et le jus d'orange. »
20. « Dérouler la pâte brisée et la placer dans un plat à tarte. »

✓ Mise en commun :

Après avoir observé, analysé et classé, les propriétés combinatoires des lexèmes MOULU, BRISÉ, ÉMIETTÉ et RÂPÉ sont formées dans le tableau 7-5 ci-dessous.

Propriétés combinatoires de HACHÉ, MOULU, ÉMIETTÉ et RÂPÉ	
MOULU	<p><b>Aromates secs</b> : poivre noir, piment d'Espelette, poivre blanc, cannelle, coriandre (grain) , cumin, gingembre, piment de Cayenne, thym, noix de muscade, noix de coco ;</p> <p><b>Fruits secs</b> : amande, noisette, noix, cacahuète, noix de cajou, pistache ;</p>
BRISÉ	Pâte.
ÉMIETTÉ	<p><b>Poissons</b> : poisson, thon, saumon, cabillaud/morue, merlan, sol, limande, maquereau, lieu noir, espadon, anchois, bar/loup, lotte/baudroie, carpe, flétan, grondin, merlu, sardine.</p> <p><b>Fruits de mer</b> : crabe, araignée, crevette, langouste, homard.</p>
RÂPÉ	<p><b>Fromages</b> : gruyère, parmesan, emmental, comté, Tomme de Savoie, mimolette, cantal.</p> <p><b>Légumes</b> : carotte, betterave, pomme de terre, aubergine, céleri, courgette, navet, tomate, fenouil, concombre.</p> <p><b>Fruits</b> : citron, pomme, orange, poire, mangue.</p>

Tableau 7-5 : Propriétés combinatoires de HACHÉ, MOULU, ÉMIETTÉ, RÂPÉ et BRISÉ

Le lexème MOULU s'emploie toujours avec les aromates secs (poivre noir, cumin, etc.) et les fruits secs (amande, noisette, etc.), mais le lexème ÉMIETTÉ est souvent utilisé pour les poissons (thon, saumon, etc.) et les fruits de mer (crabe, crevette, etc.). En outre, les fromages (gruyère, parmesan, etc.), les légumes (carotte, betterave, etc.) et les fruits (citron, pomme, etc.) sont généralement qualifiés par le lexème RÂPÉ. En revanche, nous ne nous servons du lexème BRISÉ que pour la pâte. Pour compléter la carte mentale de base, nous proposons de construire les cartes mentales par lexème.

## ✓ Phase de construction

Les deux catégories (aromates secs et fruits secs) pourraient être ajoutées autour du mot-clé MOULU dans la figure 7-9 ci-dessous créée par *CmapTools*.

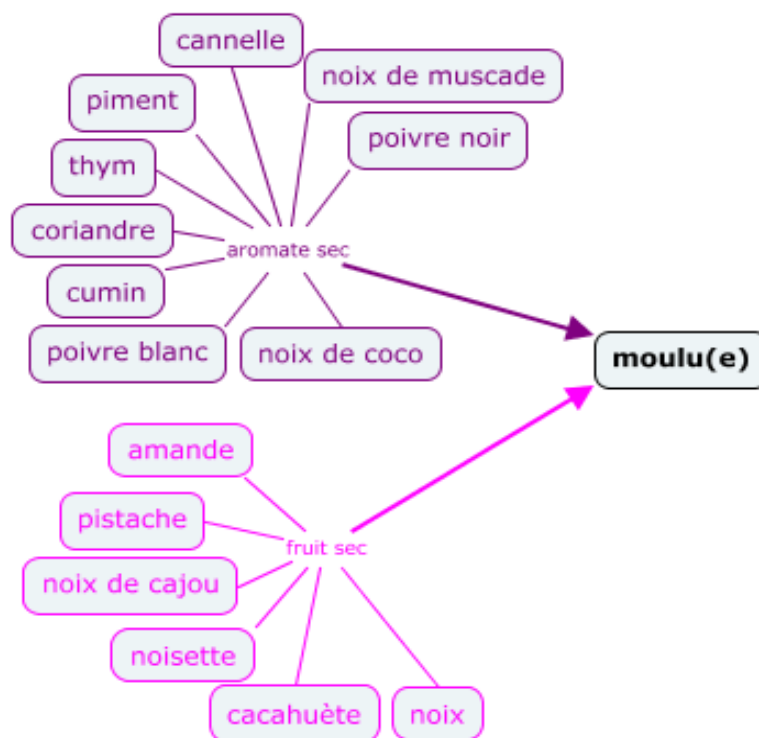


Figure 7-9 : Carte mentale de MOULU

Le sens des flèches indique l'ordre syntaxique et donc les phrasèmes *NAdj* sont tels que : *poivre noir moulu*, *cumin moulu*, *amande moulue*, *noisette moulue*, etc. Nous choisissons le violet pour la couleur de la catégorie aromate sec et le rose pour la couleur de la catégorie fruit sec.

Les trois catégories (légume, fruit, fromage) pourraient être affichées autour du mot-clé RÂPÉ dans la figure 7-10 ci-dessous :

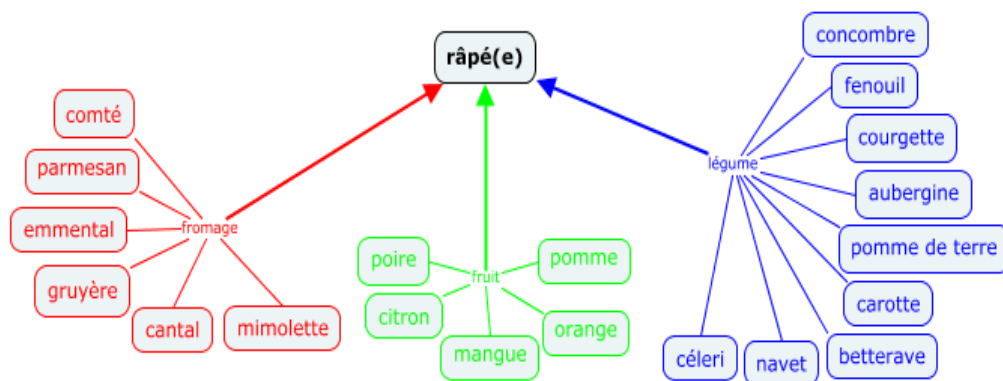


Figure 7-10 : Carte mentale de RÂPÉ

Le sens des flèches indique l'ordre syntaxique et les phrasèmes *NAdj* sont donc : *gruyère râpé*, *comté râpé*, *citron râpé*, *pomme râpée*, *carotte râpée*, *céleri râpé*, etc. Nous choisissons le rouge pour la catégorie fromage, le vert pour la catégorie fruit et le bleu pour la catégorie légume.

Les deux catégories (poisson et fruit de mer) pourraient être présentées autour du mot-clé ÉMIETTÉ dans la figure 7-11 ci-dessous :

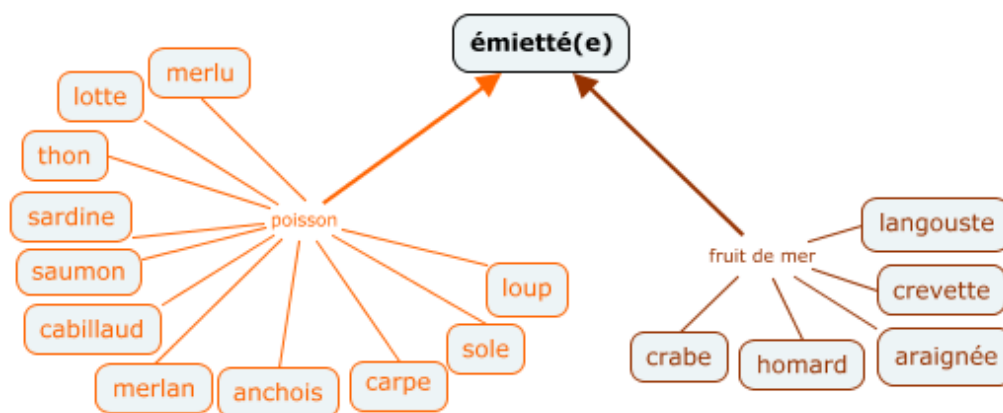


Figure 7-11 : Carte mentale d'ÉMIETTÉ

Le sens des flèches indique toujours l'ordre syntaxique et les phrasèmes *NAdj* sont donc : *thon émietté*, *saumon émietté*, *crevette émiettée*, *crabe émietté*, etc. Nous choisissons l'orange pour la catégorie poisson et le marron pour la catégorie fruit de mer. Enfin, la couleur cyan est utilisée pour le lexème BRISÉ dans la figure 7-12 ci-dessous.



Figure 7-12 : Carte mentale de BRISÉ

Une fois les quatre cartes mentales élaborées, nous pouvons envisager de compléter la carte mentale de base. Autrement dit, nous pouvons ajouter toutes les cartes mentales construites (figure 7-9, figure 7-10, figure 7-11, figure 7-12) sur la carte mentale de base comme indiqué dans la figure 7-13 ci-dessous.



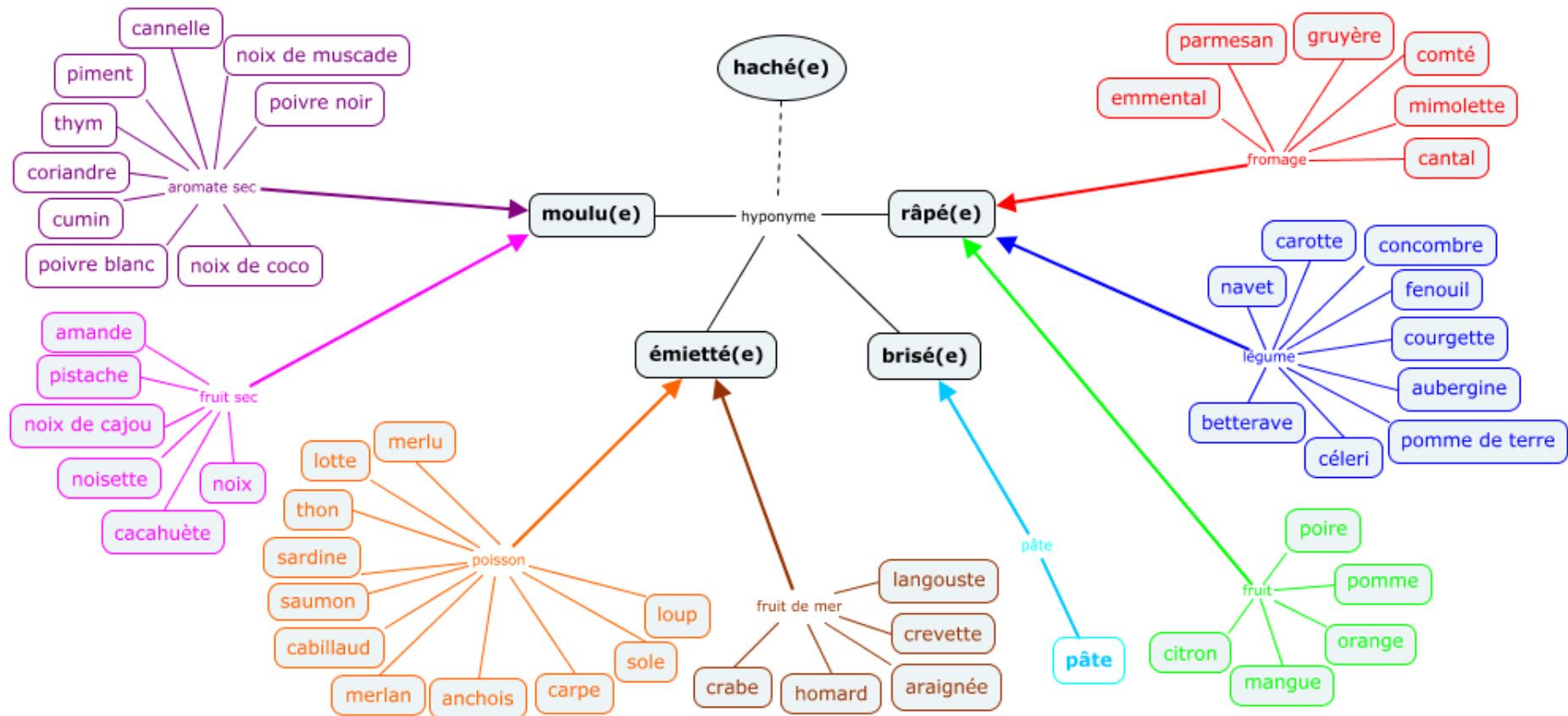


Figure 7-13 : Carte mentale rassemblée de MOULU, RÂPÉ, ÉMIETTÉ et BRISÉ

Nous répétons que nous ne nous attendons pas à ce que les apprenants dessinent une carte identique à la nôtre car ils peuvent choisir d'autres formes, d'autres couleurs, mais nous espérons qu'ils suivent les mêmes logiques pour l'élaboration de la carte mentale. Après avoir explicité et éclairci les propriétés combinatoires de ces lexèmes, les enseignants font travailler les apprenants sur un exercice lacunaire pour consolider et vérifier leurs connaissances.

b. Exercice lacunaire

✓ Dispositif :

Un travail individuel

✓ Consignes de l'exercice :

Remplissez les phrases suivantes dans le tableau 7-6 ci-dessous avec les mots appropriés (haché, hachée, hachés, hachées, émietté, émiettée, émiettés, émiettées, râpé, râpée, râpés, râpées, moulu, moulue, moulus, moulues, brisé, brisée, brisés et brisées) et justifiez votre réponse.

Phrases issues de <i>Cuisitext</i>		Explications
1	Ajouter un petit oignon ..... et les champignons de Paris coupés en tout petits morceaux.	
2	Pendant ce temps, dérouler la pâte ..... sur un moule à tarte, la piquer avec une fourchette.	
3	Ajouter le cantal ..... ou coupé en fines tranches, égaliser.	
4	Le saupoudrer largement de poivre noir.....	
5	Par dessus la couche de pommes de terre, étaler la moitié de la viande ..... avec la tomate, puis la moitié des aubergines.	
6	Incorporer le bœuf ..... et laisser cuire doucement 10 min.	
7	Recoupez cette chair en petits morceaux et mettez-la dans votre mixeur avec la crème bien froide, le blanc d'œuf, du sel et du	

	poivre blanc .....	
8	Faites fondre le beurre dans une poêle et ajoutez-y le fenouil .....	
9	Ajoutez dans le saladier le thon ..... et les tomates séchées égouttées.	
10	Lorsqu'ils sont tendres, ajoutez les tomates, l'ail ....., le thym et le laurier.	
11	Une fois terminé, vous pouvez saupoudrer de persil ..... et garnir de petits piments espagnols.	
12	Dans un saladier, mélangez la viande d'agneau ..... avec du sel, du poivre et le cumin .....	
13	Ajouter du fromage ..... et de la menthe ....., saler et poivrer.	
14	Ajoutez le fromage ....., l'ail ..... et le crabe .....	
15	Au centre d'un plat à gratin, mettez la morue ....., les pommes de terre, et couvrez avec les oignons salés et poivrés.	
16	Ajoutez dans le saladier le thon ..... et les tomates séchées égouttées.	
17	Faites fondre le beurre dans une poêle et ajoutez-y le fenouil .....	
18	Incorporer le maquereau .....	
19	Disposer une couche de purée puis une couche de poisson ..... et ainsi de suite jusqu'à épuisement des ingrédients.	
20	Etaler la pâte ..... et la piquer avec une fourchette.	

Tableau 7-6 : Exercice lacunaire

## ✓ Explication

Ce travail individuel peut être utilisé pour vérifier l'assimilation des cinq lexèmes (HACHÉ, MOULU, ÉMIETTÉ, BRISÉ, RÂPÉ) des apprenants. Nous prenons la première phrase comme exemple : en dépit du fait que l'oignon est un aromate frais, les apprenants doivent choisir le lexème HACHÉ pour qualifier l'oignon. Nous pouvons résumer les corrections apportées à cet

exercice comme suit : 1. haché (aromate frais) 2. brisée (pâte) 3. râpé (fromage) 4. Moulu (aromate sec) 5. haché (viande) 6. haché (viande) 7. Moulu (aromate sec) 8. râpé (légume) 9. émietté (poisson) 10. haché (aromate frais) 11. haché (aromate frais) 12. haché (viande) 13. râpé (fromage) et hachée (aromate frais) 14. râpé (fromage), hachée (aromate frais), émietté (fruit de mer) 15. émietté (poisson) 16. émietté (poisson) 17. râpé (légume) 18. émietté (poisson) 19. émietté (poisson) 20. brisée (pâte).

Les deux séances proposées ne sont pas des exercices en soi. Du fait du repérage et de la réflexion autour du fonctionnement de la langue effectués lors des activités précédentes, nous pensons que ce serait idéal que les deux séances soient utilisées comme exercices de repérage et de conceptualisation dans une séquence didactique ayant un lien avec les champs sémantiques de HACHÉ. Ces activités peuvent être suivies d'activités de production écrites ou orales. Si on revient sur l'apprentissage sur corpus, les deux séances proposées peuvent constituer des exercices d'entraînement pour procéder à l'AIP.

### 7.3.2 Conseils pour la mise en pratique de l'AIP

Rappelons (cf. 3.2.3.3 supra) que trois principes doivent être satisfaits pour pouvoir procéder à l'apprentissage par l'AIP : 1/exercices d'entraînement, 2/corpus de petite taille et 3/prise en main d'un outil d'interrogation du corpus. Pour le premier principe, notre proposition des deux séances ci-dessus peut servir comme exercices d'entraînement, mais il en faut davantage, car un entraînement adéquat est l'une des conditions nécessaires pour procéder à l'AIP (consultation directe) ; notons qu'un entraînement insuffisant risque de décourager les apprenants (Di Vito, 2013 ; Giuliani & Hannachi, 2010). Pour le deuxième principe, nous proposons, dans un premier temps, de placer les apprenants devant une petite partie de *Cuisitext* pour ne pas les perdre dans

de nombreuses données. Pour le troisième principe, parmi les outils d'interrogation du corpus, nous proposons *Antconc*, car ce dernier est « extrêmement simple d'utilisation et satisfera à la plupart des usages pédagogiques » (Boulton & Tyne, 2014 : 241). Même s'il n'a pas autant de fonctionnalités que *WordSmith Tools*, sa prise en main est plus aisée, d'autant plus qu'il est gratuit pour tous. Une fois que les trois principes sont satisfaits, les apprenants (cuisiniers) peuvent s'autoformer pour leurs vies professionnelles ultérieures.

## Bilan

L'une des raisons de retarder l'intégration du corpus pour l'E/A des langues provient des modes d'utilisation du corpus dont les deux types suivants sont impliqués : *teaching to exploit* et *exploiting to teach*. Pour le premier mode d'utilisation, les côtés techniques et méthodologiques empêchaient cette intégration. Pour le deuxième, le chevauchement des manières d'introduire le corpus aux apprenants peut obscurcir l'intégration du corpus pour l'E/A des langues. Afin de banaliser cette utilisation, trois approches d'utilisation du corpus ont été proposées dans cette étude : AIG, AD et AIP. Une piste d'apprentissage progressive sur corpus par combinaison des trois approches a été suggérée afin de développer l'individualisation et l'autonomisation pour l'apprentissage. Pour que les enseignants ne se perdent pas dans le processus de la préparation des cours sur corpus, nous avons proposé deux séances de formation sur les champs sémantiques de HACHÉ dans lesquelles l'accent est mis sur les activités par les deux approches (AIG et AD). Sans vouloir unifier différents comportements d'apprentissage, des conseils sont également donnés pour réaliser l'AIP à la fin de cette partie. Comme dans toute nouvelle approche proposée, nos trois nouvelles approches d'utilisation du corpus pour l'E/A comportent également des avantages et des limites.

L'avantage le plus important, de notre point de vue, est que nous avons affiné les approches existantes (approche DDL, approche ASC et consultation du corpus) en précisant les rôles des intervenants (enseignants et apprenants) et les supports utilisés pour la mise en œuvre des activités langagières. Les questions sous-jacentes à l'origine de l'E/A sur corpus (mutation des rôles des intervenants, temps consacré à la préparation, etc.) existent toujours dans les approches que nous proposons, mais il nous semble que la plus grande limite est peut-être le fait qu'il n'existe pas de frontières claires et exactes entre ces trois approches. Par exemple, à partir de quel moment exact les apprenants peuvent-ils être en mesure de procéder à l'AIP ? Nous pensons que cette limite serait réduite avec davantage de critères de référence provenant de la mise en œuvre des trois approches sur un terrain pédagogique.

## Conclusion générale

Notre étude s'inscrit dans le cadre de l'enseignement du FOS à des apprenants A2 venus travailler dans des restaurants français ou ayant choisi la restauration comme spécialité. Puisque l'enseignement du lexique est la « clé de voûte du FOS » (Binon & Verlinde, 2004 : 272) et la phraséologie constitue un des principe-clé de l'E/A du lexique (ibid., 273), notre travail s'est donc centré sur l'E/A de la phraséologie du domaine culinaire. Le livre « Le Français sur Objectif Spécifique : de l'analyse des besoins à l'élaboration d'un cours » de Mangiante et Parpette en 2004, éclaire les difficultés des apprenants en FOS (entre autres : manque de ressource nécessaire, temps limité). Dix ans plus tard (2014), la parution de l'ouvrage « Des documents authentiques aux corpus : Démarches pour l'apprentissage des langue » de Boulton et Tyne nous semble fournir toutes les solutions novatrices pour surmonter les difficultés repérées en FOS. Ces deux ouvrages nous ont permis de cerner notre champ d'étude et nos objets d'analyse. Ils ont contribué à entrer de façon plus approfondie à la fois dans la vaste question que se posent de nombreux enseignants de FOS : quel lexique enseigner en FOS ? Notre réponse étant « la phraséologie spécifique à des métiers » car elle est fréquente et peu enseignée. Où trouver ce lexique spécifique étant la deuxième grande question que l'enseignant se pose. Je propose donc d'élaborer des corpus par corps de métier afin de pouvoir extraire les éléments linguistiques qui intéressent l'enseignant. Enfin la dernière question que nous avons traitée s'attache à la méthodologie de l'enseignement de ces formes spécifiques. Nous y répondons par le développement d'une approche qui associe plusieurs approches existantes et qui nous semble appropriée à un public de professionnels car elle conduit rapidement à l'autonomie. Pour finir, nous proposons l'utilisation pédagogique des cartes mentales car cet outil permet à chacun de créer une représentation de son apprentissage et de le

construire à sa guise, voire à son rythme. Ceci semble important en milieu professionnel également. En guise de conclusion, nous allons rappeler les résultats essentiels obtenus dans ce travail en suivant le fil de conducteur de notre projet.

Du fait du manque de corpus approprié, dès le début de notre travail, nous avons pour ambition de constituer un corpus représentatif du domaine culinaire qui peut être utilisé dans plusieurs domaines de recherche, que ce soit en linguistique ou en didactique. Pour l'élaboration de notre corpus nommé *Cuisitext*, les recettes de cuisine sur l'internet (entre autres : *Marmiton*, *750g*) sont prises en compte dans le choix du genre de textes. Compte tenu de l'importance de l'oral, notre corpus est constitué de trois sortes de corpus oraux : filmages dans deux cuisines, clips vidéos culinaires et filmages dans une école hôtelière. Les outils TAL apportent une aide précieuse à la construction et au traitement du corpus. Après comparaison avec d'autres outils TAL, l'outil *Gromoteur* a été choisi pour collecter automatiquement les recettes de cuisine sur internet et l'outil *CLAN* a été utilisé pour nous aider à la transcription. Une fois construit notre corpus *Cuisitext*, par l'approche *corpus-driven*, les phrasèmes *NAdj* attirent notre attention et notre curiosité comme éléments saillants du domaine culinaire. Le repérage des phrasèmes *NAdj* s'est déroulé en deux étapes : l'extraction des séquences *NAdj* et l'identification des phrasèmes. Afin de rendre notre extraction phraséologique efficace, l'outil TAL *NooJ* a été choisi en fonction des trois exigences (définition d'une grammaire locale, récursivité de la grammaire et désambiguïsation lexicale) mises en évidence dans notre modélisation. Une fois extraites les séquences *NAdj* par *NooJ*, l'identification et la sélection des phrasèmes *NAdj* pour l'enseignement se sont faites en trois étapes : nous avons d'abord retenu les séquences apparaissant au moins 3 fois dans notre corpus et ensuite, afin de ne pas laisser passer les séquences pertinentes selon notre expérience, nous avons également vérifié les séquences de faible



fréquence (1 ou 2 occurrences) comme *sucre brun* (2 occurrences), *gaufres moelleuses* (1 occurrence). Puisque notre extraction s'est basée sur *Cuisitext* écrit, à la fin de notre sélection, nous avons enrichi notre syllabus lexical en comparant *Cuisitext* écrit à *Cuisitext* oral. Notre syllabus lexical qui est constitué de 426 phrasèmes *NAdj* devait être classé et analysé pour l'enseignement. Selon les besoins de notre public, les deux critères de classification sont sélectionnés : le critère pragmatique et le critère du figement. En adoptant les neuf propriétés transformationnelles de G. Gross qui contiennent déjà le critère pragmatique, nos phrasèmes retenus sont classés en trois catégories : expressions figées, expressions semi-figées et groupes ordinaires. Afin de préciser les rôles des intervenants (enseignants et apprenants), et les supports utilisés dans les approches d'utilisation du corpus pour l'enseignement par l'approche *exploiting to teach*, nous avons proposé les trois approches suivantes : AIG, AD et AIP. A partir de ces trois approches, nous avons développé deux séances d'activités didactiques dans lesquelles la carte mentale et le concordancier sont associés. Nous avons également proposé une piste d'apprentissage sur corpus pour que les apprenants puissent s'autoformer pour leurs vies professionnelles ultérieures, ce qui est très important pour les apprenants en FOS.

## Perspectives et limites de recherche

La phraséologie fait l'objet de notre étude et par l'approche *corpus-driven*, les phrasèmes *NAdj* ont été mis au jour comme étant les éléments les plus saillants du domaine culinaire. En même temps, bon nombre des phrasèmes verbaux (Verbe + Nom, ex. *hacher l'oignon*, *prendre couleur*, *monter les blancs en neige*) et des paternes verbaux (Verbe + qch + préposition + qch, ex. *parsemer de noisettes de beurre*; *séparer les blancs des jaunes*) ont également été relevés lors de ces extractions. En effet, le projet de l'enseignement des phrasèmes verbaux et des paternes verbaux auprès des

apprenants étrangers a déjà déclenché (Yang, 2019a ; 2019b) et à l'aide de *NooJ*, nous avons déjà extrait ces éléments (Yang, 2019c ; 2019d). Après ces projets sur les phrasèmes verbaux, nous poursuivons les études d'autres éléments lexicaux du domaine culinaire dans le but d'élaborer un « Français Culinaire Fondamental » (désormais FCF) qui permet de faciliter l'intégration langagière des apprenants étrangers dans le domaine. Sans vouloir mener un projet lexicographique du français de spécialité, comme le *dictionnaire d'apprentissage du français des affaires* (Binon *et al.*, 2000) pour le domaine du français des affaires, le FCF a pour objectif d'établir une liste de mots plus usuelle du domaine culinaire en suivant la même méthodologie de la sélection lexicale que celle utilisée lors de cette thèse (fréquence, pertinence et comparaison avec le corpus oral). Inspiré du *Français Fondamental* dans lequel le premier degré est limité à moins de 1500 mots et le deuxième degré comporte environ 1700 mots, le nombre de mots du FCF sera également limité à environ 3000. À la différence du Français Fondamental, le FCF sur internet est bien associée aux ressources multimédia, autrement dit, les utilisateurs peuvent être amenés vers la ressource (texte ou vidéo) dans laquelle le mot de leur choix est utilisé (sous forme de concordance). Des séquences d'activités langagières par les approches AIP et AD sont également proposées dans le FCF pour illustrer son exploitation à des fins d'apprentissage et d'enseignement. Afin de mieux alimenter notre projet, notre corpus a besoin d'être encore amélioré. En effet, *Cuisitext* suffisant à répondre à l'objectif de notre travail portant sur l'enseignement des phrasèmes *NAdj* : la sélection des phrasèmes *NAdj* basé sur *Cuisitext* écrit et *Cuisitext* oral utilisé pour compléter notre syllabus lexical. Mais comme le dit Rastier, le corpus doit être construit à des fins spécifiques : « un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications »

(2002 : 2). Afin de recueillir tout type de lexique du domaine culinaire, il nous faut développer davantage notre corpus à la fois au niveau de la taille et au niveau de l'annotation pour répondre à tous les besoins lexicaux attendus.

---

## Bibliographie

### A

- Abe Danielle, Carton Francis, Cembalo Sam Michel & Regent Odile, (1979). Didactique et authentique : du document à la pédagogie, *Mélanges pédagogiques*, p. 1-14.
- Abou-Haidar Laura, (2002). Transcription de la parole normale et pathologique, *Revue Parole*, n°22-24, p. 1373-1955.
- Ahmad Khurshide, Davies Andrea, Fulford Heather & Rogers Margaret, (1994). What's in a Term? The semi-automatic Extraction of Terms from Text, dans *Translation Studies. An Inter-discipline*, John Benjamins (Amsterdam/Philadelphie), p. 267- 278.
- Aijmer Karin, (1996). Conversational routines in English: Convention and Creativity, Longman (New York).
- Aitchison Jean, (1987). *Words in the Mind: An introduction to the mental lexicon*, Blackwell (Oxford).
- Alexander Richard, (1984). Fixed expressions in English: reference books and the teacher. *ELT Journal*, vol. 38(2), p. 127-134.
- Altenberg Bengt, (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations, dans Cowie Anthony-Paul (Éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press (Oxford), p. 101-122.
- Anthony Laurence, (2005). AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom, *Professional Communication Conference Proceedings*, p. 729-737. En ligne sur <http://ieeexplore.ieee.org/document/1494244/?tp=&arnumber=1494244&url=http:%2F%2Fieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D1494244> (consulté le 10/10/2019).
- Antoniadis Georges, Fairon Cédric, Granger Sylvianne, Medori Julia & Zampa Virginie, (2006). Quelles machines pour enseigner la langue ?, *The 13th Conference on Natural Language Processing* (Leuven), p. 795-805.

En ligne sur

---

<https://telearn.archives-ouvertes.fr/hal-00197393/document> (consulté le 10/10/2019).

Aussenac-Gilles Nathalie & Condamines Anne, (2001). Entre textes et ontologies formelles : les Bases de Connaissances Terminologiques, Ingénierie et capitalisation des connaissances, dans Zacklad Manuel & Grundstein Michel (Éd.), *Ingénierie et capitalisation des connaissances*, Hermes (Paris), p. 153-176.

Ausubel David Paul, (1968). *Educational psychology a cognitive view*, Holt, Rinehart & Winston (New York).

Arnold Irina Vladimirovna, (1973). *The English Word*, Vysšaja Škola (Moscow).

Arnaud Pierre, (1993). *Mille lapsus*. Manuscrit non-publié, CRTT, Université Lumière-Lyon 2.

Aston Guy, (1997). Involving learners in developing learning methods: exploiting text corpora in self-access, dans Benson Phil & Voller Peter (Éd.) *Autonomy and independence in language learning*, Longman (London), p. 204-263.

Azzopardi Sophie, (2010). Présentation : La linguistique « de » corpus au-delà des champs disciplinaires : questions et enjeux transversaux, *Cahiers de praxématique*, n°54, p. 11-24.

## B

Bally Charles, (1951). *Traité de stylistique française*, 3e éd. (1909 pour la 1e éd.), vol. I & II, Klincksieck (Paris).

Bally Charles, (1963), *Traité de Stylistique française*, Librairie de l'Université Georg & Cie (Genève).

Bárdosi Vilmos. (1986). *De fil en aiguille : les locutions françaises ; recueil thématique et livre d'exercices; kalandozás a francia szólások világában*, Tankönyvkiadó (Budapest).

- 
- Bardovi-Harlig Kathleen, (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics, *Language Learning*, n°59, p. 755-795.
- Barna Judith, (2009). L'article de presse à l'épreuve, *Les langues modernes*, n°2, p. 21-27.
- Baroni Marco & Bernardini Silvia (2004). BootCaT: Bootstrapping corpora and terms from the web, *Proceedings of LREC 2004*, Universidade Nova de Lisboa (Lisbonne).
- Baroni Marco & Bernardini Silvia (2006). *WaCky! Working Papers on the Web as Corpus*, Gedit Edizioni (Bologne). En ligne sur <http://wackybook.sslmit.unibo.it> (consulté le 10/10/2019).
- Barras Claude, Geoffrois Edouard, Wu Zhibiao & Liberman Mark (1998). Transcriber: A Free Tool for Segmenting, Labelling and Transcribing Speech, *Proceedings of the First International Conference on Language Ressources and Evaluation (LREC)*, Grenade (Espagne), p. 1373-1376.
- Barras Claude, Geoffrois Edouard, Bird Steven & Wu Zhibiao (2000). Transcribing with Annotation Graph, *Proceedings of the First International Conference on Language Ressources and Evaluation (LREC)*, Grenade (Espagne), p. 517-1521.
- Baude Olivier, Blanche-Benveniste Claire, Calas Marie-France, Cappeau Paul, Cordereix Pascal [...] Mondada Lorenza, (2006). *Guide des bonnes pratiques*, CNRS Editions, Presses Universitaires Orléans. En ligne sur [https://impec.sciencesconf.org/conference/impec/pages/Corpus\\_Oraux\\_guide\\_des\\_bonnes\\_pratiques\\_2006.pdf](https://impec.sciencesconf.org/conference/impec/pages/Corpus_Oraux_guide_des_bonnes_pratiques_2006.pdf) (consulté le 10/10/2019).
- Beacco Jean-Claude, (1991). Types ou genres ? Catégorisations des textes et didactique de la compréhension et de la production écrites, *Études de linguistique appliquée*, n°83, p. 19-28.
- Beacco Jean-Claude, (2004). Trois perspectives linguistiques sur la notion de genre discursif, *Langages*, vol. 38(153), p. 109-119.
- Bencini Vera, Paola Cangiolli Monique, Naldini Francesca & Paris Aurélie (2016). *En cuisine et en salle B1-B2*, Clé international.

- 
- Benson Morton, Benson Evelyn & Ilsen Robert, (1986). *The BBI combinatory dictionary of English: A guide to word combination*, John Benjamins (Amsterdam/Philadelphia).
- Benson Morton, (1989). The Structure of the Collocational Dictionary, *International Journal of Lexicography*, vol. 2(1), p. 1-14.
- Benveniste Emile, (1974). Fondements syntaxiques de la composition nominale, *Problèmes de linguistique générale*, Gallimard (Paris), p. 145-162.
- Bérard Evelyne, (1991). *L'approche communicative. Théorie et pratiques*. CLE International (Paris).
- Biber Douglas, (1988). *Variety across speech and writing*, Cambridge University Press.
- Biber Douglas, Conrad Susan & Randi Reppen, (1998). *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge University Press.
- Bierwisch Manfred, (1970). Semantics, *New Horizons in Linguistics*, Penguin Books (Harmondsworth), p. 166-184.
- Biggs Patricia & Dalwood Mary, (1976). *Les orléanais ont la parole*, Longman (London).
- Binon Jean, Verlinde Serge, Van Dyck Jan & Bertels Ann, (2000). *Dictionnaire d'apprentissage du français des affaires*. Paris, Didier.
- Binon Jean & Verlinde Serge, (2003). Les collocations : clef de voûte de l'enseignement et de l'apprentissage du vocabulaire d'une langue étrangère ou seconde, *La lettre de l'AIRDF*, n°33, p. 31-36.
- Binon Jean & Verlinde Serge, (2004). L'enseignement/apprentissage du vocabulaire et la lexicographie pédagogique du français sur objectifs spécifique (FOS) : Le domaine du français des affaires, *Études de linguistique appliquée*, n°135, p. 271–283.
- Blanc Olivier, Constant Mattieu, Dister Anne & Watrin Patrick, (2008). Corpus oraux et chunking, 27èmes Journées d'Études sur la Parole (Aix-en-Provence).

- Blanc Olivier, Constant Mattieu, Dister Anne & Watrin Patrick, (2010). Partial parsing of spontaneous spoken French, 7th International Conference on Language Resources and Evaluation (Miyazaki), May 2010.
- Blanche-Benveniste Claire & JeanJean Colette, (1987). *Français parlé : transcription et édition*, Didier Erudition (Paris).
- Blanche-Benveniste Claire, (1999). *Approches de la langue parlée en français*, Ophrys (Paris).
- Blanche-Benveniste Claire, (2010), *Approches de la langue parlée en français* (nouvelle édition), Ophrys (Paris).
- Bloomfield Leonard, (1927). Literate and illiterate speech, *American Speech*, vol. 2(10), p. 432-439.
- Blumenthal Peter, (2007). A usage-based dictionary of French collocations, dans Kawaguchi Yuji, Takagaki Toshihiro, Tomimori Nobuo & Tsutuga Yoichiro (Éd.), *Corpus-Based Perspectives in Linguistics*, John Benjamins (Amsterdam/Philadelphia), p. 287-300.
- Boch Françoise & Buson Laurence, (2012). Orthographe & grammaire à l'université, quels besoins ? quelles démarches pédagogiques ? *Scripta*, vol. 16(30), p. 31–51.
- Boersma Paul & Weenink David, (2004). *Praat: Doing Phonetics by Computer* (Version 4.3.01). En ligne sur <http://www.fon.hum.uva.nl/praat/> (consulté le 10/10/2019).
- Bolinger Dwight, (1976). Meaning and Memory, *Forum Linguisticum*, n°1, p. 1-14.
- Bolly Catherine, (2011). *Phraséologie et collocations*, Peter Lang (Berne).
- Boulton Alex, (1998). *Le Lexique Mental L2 : Connexions et Associations*. Humanities and Social Sciences, Université Nancy II.
- Boulton Alex, (2000). Aspects lexicaux de l'acquisition < naturelle > et de l'apprentissage < artificiel > en L2, *Mélanges CRAPEL*, n°26, p. 63-90. En ligne sur <https://hal.archives-ouvertes.fr/hal-00114345/document> (consulté le 10/10/2019).



- Boulton Alex, (2008a). Esprit de corpus : promouvoir l'exploitation de corpus en apprentissage des langues, *Texte et Corpus*, n°3, p. 37-46. En ligne sur <https://hal.archives-ouvertes.fr/hal-00326988/document> (consulté le 10/10/2019).
- Boulton Alex, (2008b). DDL is in the details... and in the big themes, dans Davies Matthew, Rayson Paul, Hunston Susan & Danielsson Pernilla (Éd.), *Proceedings of the 4th Corpus Linguistics Conference* (Birmingham), 27-30 juillet. En ligne sur <https://hal.archives-ouvertes.fr/hal-00326994> (consulté le 10/10/2019).
- Boulton Alex, (2009). Testing the limits of data-driven learning: language proficiency and training, *ReCALL*, vol. 21(11), p. 37-51.
- Boulton Alex, (2010). Consultation de corpus et styles d'apprentissage, *Cahiers de l'APLIUT*, vol. 29(1). En ligne sur <https://journals.openedition.org/apliut/967> (consulté le 10/10/2019).
- Boulton Alex & Tyne Henri (2014). *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*, Didier (Paris).
- Boutet Josiane, (2005). Genres de discours en situation de travail, dans Filliettaz Laurent & Bronckart Jean-Paul (Éd.), *L'analyse des actions et des discours en situations de travail. Concepts, méthodes et applications*, Peeters (Louvain), BCILL, p. 19-35.
- Bossé-Andrieu Jacqueline & Mareschal Geneviève, (1998). Trois aspects de la combinatoire collocationnelle, *TTR : traduction, terminologie, rédaction*, vol. 11(1), p. 157-171.
- Bréal Michel (1897). *Essai de sémantique* (Science des significations), Paris, Hachette.
- Bronckart Jean-Paul, (1996). Genres de textes, types de discours et opérations psycholinguistiques, *Voies livres*, n°78, p. 1-20. En ligne sur [http://www.revue-texto.net/docannexe/file/86/bronckart\\_rastier.pdf](http://www.revue-texto.net/docannexe/file/86/bronckart_rastier.pdf) (consulté le 10/10/2019)
- Brunet Étienne, (2006). *Hyperbase, Logiciel documentaire et statistique pour la création et l'exploitation de bases hypertextuelles, Manuel de référence*, Université de Nice.

---

Brunet Étienne, (2012). Nouveau traitement des cooccurrences dans Hyperbase, *Corpus*, n° 11, p. 219-246. En ligne sur <http://journals.openedition.org/corpus/2275> (consulté le 10/10/2019).

Burger Harald, (1973). *Idiomatik des Deutschen*, Max Niemeyer (Tübingen).

Burger Harald, (1998). *Phraseologie: Eine Einführung am Beispiel des Deutschen*, Erich Schmidt (Berlin).

## C

Cacciari Cristina & Tabossi Patrizia, (1993). *Idioms: Processing, Structure and Interpretation*, Lawrence Erlbaum Associates (Hillsdale).

Cappeau Paul & Sejjido Magali, (2005). *Les corpus oraux en français* (inventaire 2005 v. 1.0), Direction générale à la langue française et aux langues de France. En ligne sur [http://www.dgflf.culture.gouv.fr/recherche/corpus\\_parole/Presentation\\_Inventaire.pdf](http://www.dgflf.culture.gouv.fr/recherche/corpus_parole/Presentation_Inventaire.pdf) (consulté le 10/10/2019).

Cappeau Paul & Gadet Françoise, (2007). Où en sont les corpus sur les français parlés ?, *Revue française de linguistique appliquée*, vol. 12(1), p. 129-133.

Carneado Moré Zoila & Tristá Pérez Antonia, (1985). *Estudios de Fraseología*, Academia de Ciencias de Cuba (La Habana).

Ciekanski Maud, (2014). Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ?, *Les Cahiers de l'Acedle : Recherches en Didactique des Langues et Cultures*, vol. 11(1), p. 111- 135.

Carras Cathrine, Tolas Jacqueline, Kohler Patricia & Sjlilagi Elisabeth, (2007). *Le français sur objectifs spécifiques et la classe de langue*, CLE International (Paris).

Cavalla Cristelle & Elsa Crozier, (2005). *Émotions-Sentiments, nouvelle approche lexicale du FLE*, presses universitaires de Grenoble.

Cavalla Cristelle, (2008). Propositions didactiques pour l'enseignement d'éléments phraséologiques en FLEM, dans Raemdonck Dan Van & Damar Marie-Eve (Éd.), *Enseigner les structures langagières en FLE*. En

- 
- ligne sur <https://hal.archives-ouvertes.fr/hal-00397807/document> (consulté le 10/10/2019).
- Cavalla Cristelle, (2009). La phraséologie en classe de FLE. *Les Langues Modernes*, n°1. En ligne sur <https://hal.archives-ouvertes.fr/hal-00699916/document> (consulté le 10/10/2019).
- Cavalla Cristelle & Labre Virginie, (2009). L'enseignement en FLE de la phraséologie du lexique des affects, dans Agnès Tutin et Iva Novakova (Éd.), *Le lexique des émotions et sa combinatoire lexicale et syntaxique*, Ellug, p. 297-316.
- Cavalla Cristelle & Loiseau Mathieu, (2013). Scientext comme corpus pour l'enseignement, dans Tutin Agnès & Grossmann Francis (Éd.), *L'écrit scientifique : du lexique au discours. Autour de Scientext*, Presses Universitaires de Rennes, p. 163-182.
- Cavalla Cristelle, Loiseau Mathieu, Lascombe Valérie & Socha Joanna, (2014). Corpus, base de données, cartes mentales pour l'enseignement, dans Blumenthal Peter, Iva Novakova & Siepmann Dirk (Éd.), *les émotions dans le discours*, Peter Lang (Francfort), p. 327–341.
- Cavalla Cristelle, (2016). *Les apprentissages lexicaux : des unités linguistiques à l'enseignement du FLE*. En ligne sur <https://hal-univ-paris3.archives-ouvertes.fr/tel-01468588/document> (consulté le 10/10/2019).
- Čermak František, (1988). On the Substance of Idioms, *Folia linguistica*, vol. 22(3-4), p. 413-438.
- Čermak František, (1998). La identificación de las expresiones idiomáticas, dans Juan de Dios Luque Duran & Antonio Pamies Bertrán (Éd.), *Léxico y fraseología*, Método Ediciones (Granada), p. 1-18.
- Condamines Anne, (2005). Linguistique de corpus et terminologie, *Langages*, n°157, *La terminologie : nature et enjeux*, p. 36-47. En ligne sur [https://www.persee.fr/doc/lgge\\_0458-726x\\_2005\\_num\\_39\\_157\\_973](https://www.persee.fr/doc/lgge_0458-726x_2005_num_39_157_973) (consulté le 10/10/2019).
- Challe Odile, (2000). *Le français de spécialités*. CLE International (Paris).
- Challe Odile, (2002). *Enseigner le français de spécialité* Economica (Paris).

- Chambers Angela & O'sullivan Ide, (2004). Corpus consultation and advanced learners' writing skills in french, *ReCALL*, vol. 16(1), p. 158-172. En ligne sur <https://www.cambridge.org/core/journals/recall/article/corpus-consultation-and-advanced-learners-writing-skills-in-french/1BD75F72F01A7F265167D7DF3AE636CF> (consulté le 10/10/2019).
- Chambers Angela, (2005). Integrating corpus consultation in language studies, *Language Learning & Technology*, vol. 9(2), p. 111-125.
- Chambers Angela & Le Baron Florence, (2007). *Le Corpus Chambers-Le Baron d'articles de recherche en français / The Chambers-Le Baron Corpus of Research Articles in French*, Oxford Text Archive (Oxford). En ligne sur <http://ota.ahds.ac.uk/headers/2527.xml> (consulté le 10/10/2019).
- Chambers Angela, (2009). Les corpus oraux en français langue étrangère : authenticité et pédagogie, *Mélanges CRAPEL*, n° 31, p. 15-33. En ligne sur <http://www.atilf.fr/spip.php?article3724> (consulté le 10/10/2019).
- Chambers Angela, (2010). L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé, *française de linguistique appliquée*, vol. 15(2), p. 9-20.
- Chambers Angela, Farr Fiona & O'Riordan Stéphanie, (2011). Language teachers with corpora in mind: from starting steps to walking tall, *The Language Learning Journal*, vol. 39(1), p. 85-104.
- Chan Tunpei & Liou Hsien-Chin, (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations, *Computer Assisted Language Learning*, vol. 18(3), p. 231-251.
- Charles-Lavauzelle Henri & Cie, (1927). *Règlement provisoire du 7 juillet pour l'enseignement du français aux militaires indigènes*, Editeurs militaires, Paris, Limoges, Nancy.
- Chen Hsuan-Chih, (1990). Lexical processing in a non-native language: effets of language proficiency and learning strategy, *Memory & Cognition*, vol. 18(3), p. 279-288. En ligne sur <https://link.springer.com/content/pdf/10.3758/BF03213881.pdf> (consulté le 10/10/2019).
- Chiss Jean-Louis, (1987). Malaise dans la classification, *Langue Française*, n°74, p. 10-28.

- 
- Chollet Isabelle & Robert Jean-Michel, (2008). *Les expressions idiomatiques*, Clé International.
- Cholvy Jérôme, (2014). *En cuisine !* Clé international.
- Chomsky Noam & Schützenberger Marcel-Paul, (1963). The algebraic theory of context-free languages, dans Braffort Paul & Hirschberg Davide (Éd.), *Computer programming and formal systems*, North-Holland (Amsterdam), p. 118–161. En ligne sur <http://www-igm.univ-mlv.fr/~berstel/Mps/Travaux/A/1963-7ChomskyAlgebraic.pdf> (consulté le 10/10/2019).
- Choueka Yaacov, Klein Shmuel Toni & Neuwitz Erich, (1983). Automatic Retrieval of Frequent Idiomatic and Collocational Expressions in a Large Corpus, *Journal of the Association for Literary and Linguistic Computing*, n°4, p. 34-38.
- Chung Teresa Mihwa (2003). A Corpus Comparison Approach for Terminology Extraction, *Terminology*, vol. 9(2), p. 221.246.
- Church Kenneth Ward & Hanks Patrick, (1989). Word Association Norms, Mutual Information, and Lexicography, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver (Canada), p. 76-83.
- Church Kenneth Ward, Hanks Patrick & Hindle Donald, (1991). Using Statistics in Lexical Analysis, dans Zernik Uri (Éd.), *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, Lawrence Erlbaum (Hillsdale, New Jersey), p. 115-164.
- Clot Yves, (2008). *Travail et pouvoir à agir*, Presses Universitaires de France (Paris).
- Coady James, (1997). L2 vocabulary acquisition through extensive reading, dans Coady James & Thomas Huckin (Éd.), *Second Language Vocabulary Acquisition*, Cambridge University Press, p. 225-237.
- Cobb Thomas, (1997a) *From concord to lexicon: Development and test of a corpus-based lexical tutor*. Thèse de doctorat. Concordia University (Montréal). En ligne sur <https://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp04/nq25913.pdf> (consulté le 10/10/2019).

- 
- Cobb Thomas, (1997b). Is there any measurable learning from hands-on concordancing? *System*, vol. 25(3), p. 301-315.
- Cobb Thomas, (1999a) Breadth and depth of lexical acquisition with hands-on concordancing. *CALL*, vol. 12(4), p. 345-360.
- Coleman John (2014). *Transcription textGrids for the audio edition of the British National Corpus*, UK Data Archive (Colchester).
- Colletta Jean-Marc, (2004). *Le développement de la parole chez l'enfant âgé de 6 et 11 ans : corps, langage et cognition*, Editions Mardaga (Sprimont).
- Collins Allan & Loftus Elizabeth, (1975). A spreading activation theory of semantic processing, *Psychological Review*, n°82, p. 407-428.
- Conklin Kathy & Schmitt Norbert, (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?, *Applied Linguistics*, vol. 29(1), p. 72-89.
- Conseil de l'Europe, (2001). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Didier (Paris).
- Coulmas Florian, (1979). On the sociolinguistic relevance of routine formulae, *Journal of Pragmatics* 3, p. 239-266.
- Cowie Anthony-Paul, (1978). The place of illustrative material and collocations in the design of a learner's dictionary, dans Strevens Peper (Éd.), *In Honour of A.S. Hornby*, Oxford University Press, p. 127-139.
- Cowie Anthony-Paul, (1988). Stable and Creative Aspects of Vocabulary Use, dans Carter Ronald & McCarthy Michael (Éd.), *Vocabulary and language teaching*, Longman (London), p. 126-139.
- Cowie Anthony-Paul, (1998). Introduction, dans Cowie Anthony-Paul (Éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, p. 1-20.
- Curado Fuentes-Alejandro, (2001). Lexical Behaviour in Academic and Technical Corpora: implications for ESP development, *Language Learning & Technology*, vol. 5(3), p. 106-129.

---

Cuq Jean-Pierre & Gruca Isabelle, (2005). *Cours de didactique du français langue étrangère et seconde*, Grenoble, PUG

Cresswell Andy, (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course, dans Hidalgo Encarnación, Quereda Luis & Santana Juan (Éd.), *Corpora in the foreign language classroom*, Rodopi (Amsterdam), p. 267-287.

Cronk Brian & Schweigert Wendy, (1992). The comprehension of idioms: The effects of familiarity, literalness, and usage, *Applied Psycholinguistics*, n°13, p. 131-146.

## D

Danlos Laurence, (1988). Les phrases à verbe support être Prép, *Langages*, n°90, p. 23-37.

Darot Mireille & Beacco Jean-Claude, (1984). *Analyses de discours, lecture et expression*, Hachette/Larousse (Paris).

Debaisieux Jeanne-Marie, (2000). Réalités du discours oral et didactique du français langue étrangère, *Une didactique des langues pour demain (Le français dans le monde : Recherches et applications)*, Clé International, p. 143-156.

Debaisieux Jeanne-Marie, (2005). Les corpus oraux : Situation, exploitation linguistique, bilan et perspectives, *De la linguistique de corpus à la relation "partie/tout"*, *Scolia*, Publications de l'Université Marc Bloch (Strasbourg), n°19, p. 9-40.

Debaisieux Jeanne-Marie, (2008). Corpus oraux et didactique des langues : un rendez-vous à ne pas manquer, *Le français dans le Monde*, numéro spécial : *Quel oral enseigner, cinquante ans après le Français Fondamental ?*, p. 102-114.

Debaisieux Jeanne-Marie, (2009). Des documents authentiques oraux aux corpus : un défi pour la didactique du FLE, *Mélanges CRAPEL*, n°31, p. 35-56. En ligne sur <http://www.atilf.fr/spip.php?article3725> (consulté le 10/10/2019).

- Debaisieux Jeanne-Marie, (2013). Introduction, dans Debaisieux Jeanne-Marie (Éd.), *Analyses linguistiques sur corpus*, Lavoisier (Paris), p. 17-35.
- De Groot Annette, (1995). Determinants of bilingual lexicosemantic organisation, *Computer Assisted Language Learning*, n°8, p. 151-180.
- Delais-Roussarie Elisabeth, (2003). Constitution et annotation de corpus : méthodes et recommandations, dans Delais-Roussarie Elisabeth & Durand Jacques (Éd.), *Corpus et variation en phonologie du français : Méthodes et analyses*, Presses Universitaires du Mirail (Toulouse), p. 89-125.
- De Mauro Tullio, (1983). Il vocabolario di base, *Prometeo*, n°3, p. 127-31.
- Di Vito Sonia, (2013). L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE, *Linx*, n°68-69, p. 159-176. En ligne sur <https://journals.openedition.org/linx/1519> (consulté le 10/10/2019).
- Drago Lucia, (2009). Le Web comme corpus : documents authentiques et exploitation en FLE, *Mélanges CRAPEL*, n°31, p. 237-257.
- Drouin Patrick (2002). *Acquisition automatique des termes : l'utilisation des pivots lexicaux spécialisés*, thèse de doctorat, Université de Montréal.
- Drouin Patrick & BAE Hee Sook (2005). Korean Term Extraction in the Medical Domain by Corpus Comparison, *Terminology and Knowledge Engineering (TKE-2005)*, Copenhague, Copenhagen Business School, p. 349-361.
- Dubois Jean, (1973). *La nouvelle grammaire du français*, Larousse (Paris).
- Dubois Jean, (1994). *Dictionnaire de Linguistique et des Sciences du Langage*. Paris : Larousse.
- Dubois Jean & Dubois-Charlier Françoise, (1997). *Les verbes français*, Larousse (Paris).
- Dubois Jean & Dubois-Charlier Françoise, (2010). La combinatoire lexico-syntaxique dans le Dictionnaire électronique des mots, *Langages*, n°3, p. 31-56. En ligne sur <https://www.cairn.info/revue-langages-2010-3-page-31.htm> (consulté le 10/10/2019).



---

Duffau Hugues, (2016). *L'erreur de Broca - Exploration d'un cerveau éveillé*, Michel Lafon (Paris).

Dumarest Danièle & Morsel Marie-Hélène, (2004). *Le chemin des mots, pour un apprentissage méthodique du vocabulaire français*, Presses Universitaires de Grenoble.

Durand Jacques & Tarrier Jean-Michel, (2008). Deux thèses sur la transcription et le codage d'un grand corpus oral : le cas de PFC, dans Bilger Mireille (Éd.), *Données orales : les enjeux de la transcription*, Presses Universitaires de Perpignan, n°37, p. 48-77.

## E

Edmonds Amanda, (2013). Une approche psycholinguistique des phénomènes phraséologiques : le cas des expressions conventionnelles, *Langages*, n°189, p. 121-138. En ligne sur <https://www.cairn.info/revue-langages-2013-1-page-121.htm> (consulté le 10/10/2019).

Edmonds Amanda, (2016). Le rôle du contexte dans le traitement des expressions conventionnelles, *Cahier de lexicologie*, n°108, p. 107-123.

Ellis Nick, (1996). Sequencing in SLA: Phonological Memory, Chunking, and Points of Order, *Studies in Second Language Acquisition*, n°18, p. 91-126.

Ellis Nick, (1997). Vocabulary acquisition: word structure, collocation, word-class, and meaning, dans Schmitt Norbert & McCarthy Michael (Éd.), *Vocabulary description, acquisition and pedagogy*, Cambridge University Press, p. 122-139.

Ellis Nick, (2002). Frequency effects in language processing, *Studies in Second Language Acquisition*, n°24, p. 143–188.

Ellis Nick, Simpson-Vlach Rita & Maynard Carson, (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL, *Tesol Quarterly*, n°42, p. 375-396.

Erman Britt & Warren Beatrice (2000), The Idiom Principle and the Open Choice Principle, *Text*, vol. 20(1), p. 29-62.

## F

- 
- Febvre Lucien, (1952). *Combats pour l'histoire*, Armand Colin (Paris).
- Felder Richard & Silverman Linda, (1988). Learning and teaching styles in engineering education. *Engineering Education*, n°78(7), p. 674-681. En ligne sur <https://www.engr.ncsu.edu/wp-content/uploads/drive/1QP6kBI1iQmpQbTXL-08HSI0PwJ5BYnZW/1988-LS-plus-note.pdf> (consulté le 10/10/2019).
- Fiala Pierre, Habert Benoit, Lafon Pierre & Pineira Carmen. (1987). Des mots aux syntagmes, figements et variations dans la Résolution générale du congrès de la CGT de 1978, *Mots*, n°14, 45-87.
- Firth John, (1935). The Semantic of Linguistic Science, dans Firth John (Éd.), *Papers in Linguistics 1934-1951*, Oxford University Press.
- Firth John, (1957). Modes of Meaning. *Papers in Linguistics 1934-1951*, p. 190-215.
- Fleury Serge, (2007). *Le Métier Textométrique : Le Trameur, Manuel d'utilisation*. En ligne sur <http://tal.univ-paris3.fr/trameur/> (consulté le 10/10/2019).
- Fleury Serge, (2017). iTrameur : Application utilisable dans votre navigateur habituel et reproduisant des calculs disponibles dans le trameur. En ligne sur <http://www.tal.univ-paris3.fr/trameur/iTrameur/> (consulté le 10/10/2019).
- Fligelstone Steve, (1993). Some reflections on the question of teaching, from a corpus linguistics perspective, *ICAME journal*, n°17, p. 97-109.
- Flowerdew John, (1993). Concordancing in language learning, *Perspectives*, n°5, p. 87-101.
- Flowerdew John, (1996). Concordancing in language learning, dans Pennington Martha (Éd.), *The Power of CALL*, Athelstan (Houston), p. 87-101.
- Fónagy Ivan, (1997). Figement et changement sémantique, dans Martins-Baltar Michel (Éd.), *La locution entre langue et usages*, ENS Éditions (Fontenay/Saint-Cloud), p. 131-164.

---

Forsberg Fanny, (2008). Formes, fonctions et fréquences en français parlé L2 et L1, *Le langage préfabriqué*, Peter Lang (Bern-Berlin).

François Jacques, (2008c). Une approche diachronique quantitative de la polysémie verbale, *Cahiers du CRISCO*, CRISCO (Caen), n°24. En ligne sur <https://hal.archives-ouvertes.fr/hal-01870490/document> (consulté le 10/10/2019).

Frandsen Finn, (1998). Langue générale et langue de spécialité – une distinction asymétrique ?, dans Gambier Yves (Éd.), *Discours professionnels en français*, Peter Lang (Frankfurt), p. 15-34.

Frankenberg-Garcia Ana, (2005). A peek into what today's language learners as re- searchers actually do, *International Journal of Lexicography*, vol. 18(3), p. 335-355.

## G

Gairns Ruth & Redman Stuart, (1986). *Working with Words: A guide to teaching and learning vocabulary*. Cambridge University Press.

Galisson Robert, (1984a). *Les expressions imagées, pratique des langues étrangères, les mots : mode d'emploi*, Clé International.

Galisson Robert, (1984b). *Dictionnaire de compréhension et de production des expressions imagées*, Clé International.

Gasparov Boris, (2004). *Speech, Memory, and Meaning: Intertextuality in Everyday Language*, Columbia University Press (New York).

Gavioli Laura, (2005). *Exploring corpora for ESP Learning*, John Benjamins (Amsterdam).

Gaskell Delian & Cobb Thomas, (2004). Can learners use concordance feedback for writing errors? *System*, vol. 32(3), p. 301-319.

Gerdes Kim, (2014). Corpus collection and analyse for the linguistic layman: The Gromoteur, actes de 12<sup>ème</sup> Journées internationales d'Analyse statistique des Données Textuelles (Paris). En ligne sur <https://gerdes.fr/papiers/2014/Gromoteur-jadt2014.pdf> (consulté le 10/10/2019).

- 
- Gillam Lee, Tariq Mariam & Ahmad Khurshid (2005). Terminology and the Construction of Ontology, *Terminology*, vol. 11(1), p. 55-81.
- Giuliani Delphine & Hannachi Radia, (2010). Linguistique de corpus et didactique du F.L.E, Une exploitation du corpus IntUne, *Cahiers de praxématique*, n°54- 55, p. 145- 160.
- Gläser Rosemarie, (1988) The grading of idiomaticity as a presupposition for a taxonomy of idioms, dans Hüllen Werner & Schulze Rainer (Éd.), *Understanding the Lexicon*, Max Niemeyer (Tübingen), p. 264-279.
- Gläser Rosemarie, (1998). The stylistic potential of phraseological units in the light of genre analysis, dans Cowie Anthony-Paul (Éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, p. 125-143.
- Gledhill Christophe, (2000). *Collocations in science writing*, Gunter Narr Verlag Tübingen (Tübingen).
- Gledhill Christophe, (2011). The 'lexicogrammar' approach to analysing phraseology and collocation in ESP texts, *ASp*, n°59, p. 5-23. En ligne sur <https://journals.openedition.org/asp/2169> (consulté le 10/10/2019).
- Goldman Jean-Philippe, Nerima Luka & Wehrli Eric, (2001). Collocation extraction using a syntactic parser, *Proceedings of the ACL '01 Workshop on Collocation* (Toulouse), p. 61-66.
- González-Rey Isabel, (2002). *La phraséologie du français*, Presses Universitaires du Mirail (Toulouse).
- González-Rey Isabel, (2007). *La didactique du français idiomatique*, E.M.E, Fernelmont (Belgique).
- González-Rey Isabel, (2010). La phraséodidactique en action : les expressions figées comme objet d'enseignement, *La Clé des Langues*, ENS de Lyon/DGESCO. En ligne sur <http://cle.ens-lyon.fr/espagnol/langue/traduction/la-phraseodidactique-en-action-les-expressions-figees-comme-objet-d-enseignement> (consulté le 10/10/2019).
- González-Rey Isabel, (2018). Phraséotext – Le français idiomatique : une méthode d'enseignement-apprentissage en phraséodidactique du FLE, dans Soutet Olivier, Mejri Salah & Sfar Inès (Éd.), *la phraséologie : théories et applications*, Honoré Champion (Paris), p. 301-318.

- 
- Goody Jack, (1979). *La raison graphique – la domestication de la pensée sauvage*, Minit (Paris).
- Gougenheim Georges, Michea René, Rivenc Paul & Sauvageot Aurélien Aurélien (1956). *L'élaboration du français élémentaire : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Didier (Paris).
- Gougenheim Georges, Michea René, Rivenc Paul & Sauvageot Aurélien (1964). *L'élaboration du français fondamental : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Didier (Paris).
- Granger Sylviane, (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae, dans Cowie Anthony-Paul (Éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, p. 145-160.
- Granger Sylviane, (2001). Didactique des langues étrangères, linguistique de corpus et traitement automatique des langues, dans Marquillo Larruy (Éd.), *Questions d'épistémologie en didactique du français (langue maternelle, langue seconde, langue étrangère)*, Cahier FORELL, p. 105-109.
- Granger Sylviane & Paquot Magali, (2008), Disentangling the phraseological web, dans Granger Sylviane & Meunier Fanny (Éd.), *Phraseology: An Interdisciplinary Perspective*, John Benjamins (Amsterdam/Philadelphia).
- Granger Sylviane & Paquot Magali, (2015). Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid. *Lexicographica - International Annual for Lexicography/Internationales Jahrbuch für Lexikographie*, vol. 31(1), p. 118- 141.
- Grefenstette Gregory & Teufel Simone, (1995). Corpus-based Method for Automatic Identification of Support Verbs for Nominalizations, *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics* (Dublin), p. 98-103.
- Gross Gaston, (1981). Lexicographie et grammaire, *Cahiers de Lexicologie*, n°39, p. 36-46.
- Gross Gaston, (1986a). Étude syntaxique du déterminant possessif, *Recherches Linguistiques*, Klincksieck (Paris), n°11, p. 87-111.

- 
- Gross Gaston, (1986b). *Typologie des noms composés*, Rapport A.T.P. du CNRS.
- Gross Gaston, (1988). Degré de figement de noms composés, *Langages*, n°90, p. 57- 72, Larousse (Paris).
- Gross Gaston, (1996). *Les expressions figées en français : noms composés et autres locutions*, Ophrys (Paris/Gap).
- Gross Maurice, (1988). Les limites de la phrase figée. *Langages*, n°90, p. 7-22, Larousse (Paris).
- Grossmann Francis & Tutin Agnès, (2007). Motivation of lexical associations in collocations: the case of intensifiers for “nouns of joy”, dans Wanner Leo (Éd.), *Selected Lexical and Grammatical Issues in the Meaning-Text Theory In honor of Igor Mel’čuk*, John Benjamins (Amsterdam), p. 139-165.
- Grossmann Francis, (2011). Didactique du lexique : état des lieux et nouvelles orientations, *Pratiques*, n°149-150, p. 163-183.
- Guiraud Pierre, (1961). *Les locutions françaises*, Presses Universitaires de France.
- Gülich Elisabeth & Krafft Ulrich, (1997). Le rôle du préfabriqué dans les processus de production discursive, dans Martins-Baltar Michel (Éd.), *La locution entre langue et usages*, ENS Éditions (Fontenay/Saint-Cloud), p. 241-276.
- Günthner Susanne, (2006). Von Konstruktionen zu kommunikativen Gattungen : Die Relevanz sedimentierter Muster für die Ausführung kommunikativer Aufgaben, *Deutsche Sprache*, n°34, p. 173-190.
- H
- Habert Benoît, Nazarenko Adeline & Salem André, (1997). *Les linguistiques de corpus*, Armand Colin (Paris).
- Hanks Patrick, (2008). The lexicographical legacy of John Sinclair, *International Journal of Lexicography*, vol. 21(3), p. 219-229

- 
- Harmer Jeremy, (2001). *The Practice of English Language Teaching*, Longman (Harlow).
- Häusermann Jürg, (1977). *Phraseologie: Hauptprobleme der deutschen Phraseologie auf der Basis sowjetischer Forschungsergebnisse*, Niemeyer (Tübingen).
- Hazenberg Suzanne & Hulstijn Jan, (1992). Defining a minimal receptive second language vocabulary for non-native university students: an empirical investigation, *Applied linguistics*, n°17, p.145-163. En ligne sur [https://pure.uva.nl/ws/files/1979914/165030\\_Hazenberg\\_Hulstijn\\_Applied\\_Linguistics\\_1996.pdf](https://pure.uva.nl/ws/files/1979914/165030_Hazenberg_Hulstijn_Applied_Linguistics_1996.pdf) (consulté le 10/10/2019).
- Heiden Serge, (2004). Interface hypertextuelle à un espace de cooccurrences : implémentation dans Weblex, *Proceedings of 10th International Conference on the Statistical Analysis of Textual Data JADT2004*, Le poids des mots, Presses universitaires de Louvain, p. 577-588.
- Heiden Serge, Magué Jean-Philippe & Pincemin Bénédicte, (2004). TXM : une plateforme logicielle open-source pour la textométrie – conception et développement, *JADT2010*, Roma, Edizioni Universitarie di Lettere Economia, n°2, p. 1021-1032.
- Heinz Michaela, (1993). Les locutions figurées dans le Petit Robert : description critique de leur traitement et propositions de normalisation, *Lexicographica*, Max Niemeyer (Tübingen).
- Heinz Michaela, (1997). L'à-peu-près dans les locutions et son traitement lexicographique, dans Fiala Pierre Lafon & Piquet Marie-France (Éd.), *La locution : entre lexicque, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage*, Klincksieck (Paris), p. 213-229.
- Heinz Michaela, (2003). Les locutions figurées dans le Nouveau Petit Robert : évolution de quelques traitements entre 1993 et 2003, dans Cormier Monique, Boulanger Jean-Claude & Francœurn Aline (Éd.), *Les dictionnaires Le Robert : genèse et évolution*, Les Presses Universitaires de Montréal, p. 227-245.
- Hermann Paul, (1880). *Prinzipien der Sprachgeschichte*, Niemeyer (Tübingen).

- Higa Masanori, (1963). Interference effects of intralist word relationships in verbal learning, *Journal of Verbal Learning and Verbal Behavior*, n°2, p. 170-175.
- Hill Jimmie, (1993). Revising priorities: from grammatical failure to collocational success, dans Lewis Michael (Éd.), *The lexical approach, The State of ELT and the Way Forward*, Language Teaching Publications (Londres), p. 47-67.
- Hilton Heather, (2002). Modèles de l'acquisition lexicale en L2 : où en sommes-nous ? *ASp*, n°35-36, p. 201-217. En ligne sur <https://journals.openedition.org/asp/1668> (consulté le 10/10/2019).
- Hilton Heather, (2009). *Systèmes émergents : acquisition, traitement et didactique des langues*, Humanities and Social Sciences, Université Lumière - Lyon II.
- Hingue Anne-Marie & Ulm Karine, (2005). *Dites-moi un peu. Méthode pratique de français oral, Niveau Intermédiaire et avancé (B2-C1)*, Presses Universitaires de Grenoble.
- Hoey Michael, (2005). *Lexical Priming*, Routledge (London).
- Hofstede Geert, (2006). Dimensionalizing cultures: The Hofstede model in context, dans Lonner Walter, Dinnel Dale, Hayes Susanna & Sattler Davide (Éd.), *Psychology and Culture, Unit 2: Conceptual, Methodological and Ethical Issues in Psychology and Culture*, Western Washington University.
- Holec Henri, (1991). Autonomie de l'apprenant : de l'enseignement à l'apprentissage, *Education permanente*, n°107, p. 59-66. En ligne sur [http://epc.univ-lorraine.fr/EPCT\\_F/pdf/Autonomie.pdf](http://epc.univ-lorraine.fr/EPCT_F/pdf/Autonomie.pdf) (consulté le 10/10/2019).
- Holme Randal, (2010a). A construction grammar for the classroom. *International Review of Applied Linguistics and Language Teaching*, n° 48, p. 355-377.
- Holme Randal, (2010b). Construction grammars: Towards a pedagogical model, *AILA Review*, n° 23, p. 115-133.



- Howarth Peter, (1993). A phraseological approach to academic writing, dans Blue George (Éd.), *Language, Learning and Success: Studying through English*, Macmillan (London), p. 58-69.
- Howarth Peter, (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*, Niemeyer (Tübingen).
- Howarth Peter, (1998). The Phraseology of Learners' Academic Writing, dans Cowie Anthony-Paul (Éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, p. 161-186.
- Huckin Thomas & Coady James, (1999). Incidental vocabulary acquisition in a second language, *Studies in Second Language Acquisition*, n°21, p. 181–193.
- Hudson Jean, (1998). *Perspectives on Fixedness: Applied and Theoretical* (Lund Studies in English, 94), Lund University Press.
- Hulstijn Jan, (1992). Retention of inferred and given meanings: Experiments in incidental vocabulary learning, dans Arnaud Pierre & Béjoint Henry (Éd.), *Vocabulary and Applied Linguistics*, Macmillan (Londres).
- Hundt Marianne, Nesselhauf Nadja & Biewer Carolin, (2007). *Corpus Linguistics and the Web*, Rodopi (Amsterdam).
- J
- Jernej Josip, (1992). O klasifikaciji frazema, *Fililogija*, Hrvatska akademija znanosti i umjetnosti (Zagreb), n°20-21, p. 191-197.
- Jespersen Otto, (1992). *La philosophie de la grammaire*, traduction de l'anglais par Anne-Marie Léonard, première édition en 1924, George Allen & Unwin Ltd (London).
- Johns Tim, (1986). Micro-Concord: A language learner's research tool, *System*, vol. 14(2), p. 151-162.
- Johns Tim, (1988a). Whence and Whither Classroom Concordancing, dans Bongaerts Theo, Haan Pieter de, Lobbe Sylvia & Wekker Herman (Éd.), *Computer applications in language learning*, Foris (Dordrecht), p. 9-27.

- Johns Tim, (1988b). Implications et applications des logiciels de concordance dans la salle de classe, *Les langues modernes*, vol. 82(5), p. 29-45 (traduction française de Johns, 1988a).
- Johns Tim, (1991a). Should you be persuaded: two examples of data-driven learning, dans Johns Tim & King Philip (Éd.), *Classroom Concordancing. English Language Research Journal*, n°4, p. 1-16.
- Johns Tim, (1991b). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning, dans Johns Tim & King Philip (Éd.), *Classroom Concordancing. English Language Research Journal*, n°4, p. 27-45.
- Johns Tim, (1997). Contexts: The background, development and trialling of a concordance-based CALL program, dans Wichmann Anne, Fligelstone Steve, Mcenery Tony & Knowles Gerry (Éd.), *Teaching and Language Corpora*, Addison Wesley Longman (Harlow), p. 100-115.
- K
- Kahn Gisèle, (1995). Différentes approches pour l'enseignement du français sur objectifs spécifiques, *Le Français dans le Monde, Recherches et Applications « Méthodes et méthodologies »*, Clé international, p.144-152.
- Kennedy Graeme, (1998). *An introduction to Corpus Linguistics*, Longman (London & New York).
- Kerkman Hans & De Bot Kees, (1989). De organisatie van het tweetalige lexicon. *Toegepaste Taalwetenschap in artikelen*, n°34, p. 115-121.
- Kilgarriff Adam & Tugwell David, (2001). WORD SKETCH: Extraction, Combination and Display of Significant Collocations for Lexicography, *Proceedings of the Workshop on Collocations: Computational Extraction, Analysis and Exploitation*, ACL-EACL 2001 (Toulouse), p. 32-38.
- Kilgarriff Adam & Grefenstette Gregory (2003). Web as corpus, *Computational Linguistics*, vol. 29(3), p. 333-347.
- Kilgarriff Adam, Rychly Pavel, Smrz Pavel & Tugwell David, (2004). The Sketch Engine, *Proceedings EURALEX* (Lorient), p. 105-116.

- Kilgarriff Adam, Atkins Sue & Rundell Michael, (2007). BNC design model past its sell-by, dans Davies Matthew, Rayson Paul, Hunston Susan & Danielsson Pernilla (Éd.), *Corpus Linguistics Conference CL2007* (Birmingham). En ligne sur <http://www.kilgarriff.co.uk/Publications/2007-KilgAtkinsRundell-CL-Sellby.pdf> (consulté le 10/10/2019).
- Kirschnerl Paul, Sweller John & Clark Richard, (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching, *Educational Psychologist*, vol. 41(2), p. 75-86.
- Kleene Stephen Cole, (1956). Representation of events in nerve nets and finite automata, *Automata Studies, Annals of Mathematics Studies*, n°34, p. 3-41.
- Krashen Stephen, (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis, *Modern Language Journal*, n°73, p. 440-464.
- Kroll Judith & Stewart Erika, (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations, *Journal of Memory and Language*, n°33, p. 149-174.
- Kübler Natalie, (2014). Mettre en œuvre la linguistique de corpus à l'université - Vers une compétence utile pour l'enseignement/apprentissage des langues ? *Les Cahiers de l'Acedle*, vol. 11(1), p. 37-77.
- Kučera Henry & Francis Winthrop Nelson, (1967), *Computational Analysis of Present-Day American English*, Brown University Press (Providence).
- L
- Lafon Pierre & Salem André, (1983). L'inventaire des segments répétés d'un texte, *Mots*, vol. 6(1), p. 161-177. En ligne sur [https://www.persee.fr/doc/mots\\_0243-6450\\_1983\\_num\\_6\\_1\\_1101](https://www.persee.fr/doc/mots_0243-6450_1983_num_6_1_1101) (consulté le 10/10/2019).
- Lamiroy Béatrice, (2003). Les notions linguistiques de figement et de contrainte, *Lingvisticae Investigationes*, n°26, p. 1-14. En ligne sur <http://wwwling.arts.kuleuven.be/franitalco/papers/Lamiroy2003.pdf> (consulté le 10/10/2019).

- Lamiroy Béatrice & Klein Jean René, (2005). Le problème central du figement est le semi-figement, *Linx*, n°53, p. 135-154. En ligne sur <https://journals.openedition.org/linx/271> (consulté le 10/10/2019).
- Lamy Marie-Noëlle & Hans Jørgen Klarskov Mortensen, (2000). Using concordance programs in the modern foreign languages classroom, *Information and Communications Technology for Language Teachers*. En ligne sur [http://www.ict4lt.org/en/en\\_mod2-4.htm](http://www.ict4lt.org/en/en_mod2-4.htm) (consulté le 10/10/2019).
- Landure Corinne & Boulton Alex, (2010). Corpus et autocorrection pour l'apprentissage des langues, *ASp*, n°57. En ligne sur <http://asp.revues.org/931> (consulté le 10/10/2019).
- Laufer Batia, (1992). How much lexis is necessary for reading comprehension?, dans Henry Béjoint & Arnaud Pierre (Éd.), *Vocabulary and Applied Linguistics*, Macmillan (London), p.126-132. En ligne sur <http://nflrc.hawaii.edu/rfl/April2010/articles/laufer.pdf> (consulté le 10/10/2019).
- Laurens Véronique, (2013). *Formation à la méthodologie de l'enseignement du FLE et développement de l'agir enseignant*. Education. Université Sorbonne Nouvelle – Paris 3. En ligne sur <https://halshs.archives-ouvertes.fr/tel-01419749/document> (consulté le 10/10/2019).
- Lebart Ludovic, (2015). DtmVic: Data and Text Mining - Visualization, Inference, Classification. Exploratory statistical processing of complex data sets comprising both numerical and textual data. En ligne sur <http://www.dtmvic.com/> (consulté le 10/10/2019).
- Lee Chuen-Yi & Liou Hsien-Chin, (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context, *English Teaching and Learning*, vol. 27(3), p. 35-56.
- Leech Geoffrey, (1991). The state of the art in corpus linguistics, dans Aijmer, Karin & Altenberg Bengt (Éd.), *English corpus linguistics: Studies in honour of Jan Svartvik*, Longman (London), p. 8-29.
- Leech Geoffrey, (1992). Corpus and theories of linguistic performance, dans Svartvik Jan (Éd.), *Corpus Linguistics, Proceedings of Nobel Symposium*, Mouton de Gruyter (Berlin/New York), p. 105-122.

- 
- Legallois Dominique, (2005). Du bon usage des expressions idiomatiques dans l'argumentation de deux modèles anglo-saxons : la Grammaire de Construction et la Grammaire des Patterns, *Cahiers de l'institut de linguistique de Louvain*, n°31, p. 109-127.
- Legallois Dominique & Gréa Philippe, (2006). La Grammaire de Construction, *Cahier du CRISCO*, n°21, p. 5-27.
- Legallois Dominique & Tutin Agnès, (2013). Présentation : Vers une extension du domaine de la phraséologie, *Langages*, n°189, p. 3-25.
- Lehmann Denis, (1993). *Obejectifs spécifiques en langue étrangère*. Hachette.
- Levy Michael, (1997). *Computer-assisted language learning: Context and conceptualization*, Oxford University Press.
- Lewis Michael, (1993). *The Lexical Approach*, Language Teaching Publications (Hove).
- Lewis Michael, (1997). *Implementing the lexical approach: Putting theory into practice*, Language Teaching Publications (Hove).
- Lewis Michael, (2000). There is nothing as practical as a good theory, dans Michael Lewis (Éd.), *Teaching collocation: Further developments in the lexical approach*, Language Teaching Publications (Hove), p. 10-27.
- Lieury Alain, (1991). *Mémoire et réussite scolaire*, Dunod (Paris).
- Lin Dekang, (1998). Extracting collocations from text corpora, *First Workshop on Computational Terminology*, Montréal, p. 57-63.
- Lipka Leonhard, (1974). Probleme der Analyse englischer Idioms aus struktureller und generativer Sicht, *Linguistik und Didaktik*, n°20, p. 274-285.
- Lively Scott, Pisoni David & Goldinger Stephen, (1994). Spoken word recognition: Research and theory, dans Gernsbacher Morton Ann (Éd.), *Handbook of Psycholinguistics*, Academic Press (San Diego), p. 265-302.
- Loewen Shawn, Li Shaofeng, FeiFei, Thompson Amy, Nakatsukasa Kimi, Ahn Deongmee & Chen Xiaoqing, (2009). Second language learners' beliefs

---

about grammar instruction and error-correction, *Modern Language Journal*, vol. 93(1), p. 91-104.

## M

Macwhinney Brian, (2000). *The CHILDES Project: Tools for Analyzing Talk*, troisième édition, Mahwah, Lawrence Erlbaum Associates (New Jersey).

Makkai Adam, (1975). The Cognitive Organization of Idiomaticity: Rhyme ou Reason?, *Georgetown university working papers on language and linguistics*, n°11, p. 10-29.

Malrieu Denise & Rastier François, (2001). Genres et variations morphosyntaxiques. *Traitement automatique des langues*, vol. 42(2), p. 548-577.

Malrieu Denise, (2004). Linguistique de corpus, genres textuels, temps et personnes, *Langages*, n° 153, p. 73-85.

Mangiante Jean-Marc & Parpette Chantal, (2004). Le français sur objectif spécifique : de l'analyse de besoins à l'élaboration d'un cours, Hachette.

Mangiante Jean-Marc, (2016). Analyse de données en FOS : caractéristiques méthodologique et outils numériques, *Points Communs - Recherche en didactique des langues sur objectif(s) spécifique(s)*, CCIP, p. 25-43.

Martin Robert, (1997). Sur les facteurs du figement lexical, dans Martins-Balter (Éd.), *la locution entre langue et usages*, Fontenay Saint-Cloud : ENS Éditions, p. 291-305.

Martinet André, (1967). Syntagme et syntème, *La linguistique*, vol. 3(2), p. 1-14.

Martinet André, (1974). *Éléments de linguistique générale*, Armand Colin (Paris).

Martinet André, (1999). Le syntème, *La linguistique*, vol. 35(2), p. 11-16.

Mayer Richard, (2004). Should there be a three-strike rule against pure discovery learning? The case for guided methods of instruction, *American Psychologist*, vol. 59(1), p. 14-19.

- 
- McCarthy Michael, (2008). Accessing and interpreting corpus information in the teacher education context, *Language Teaching*, vol. 41(4), p. 563-574.
- McEney Tony & Wilson Andrew, (1997). Teaching and language corpora, *ReCALL*, vol. 9(1), p. 5-14.
- McEney Tony & Wilson Andrew, (2001). *Corpus Linguistics: An Introduction*, Edinburgh University Press (Edimbourg).
- McEney Tony & Hardie Andrew, (2012). *Corpus Linguistics: Method, theory and practice*, Cambridge University Press.
- Mel'čuk Igor & Žolkovskij Aleksandr, (1969). K postroeniju dejstvujuščej modeli jazyka "Smysl ⇔ Tekst" [Vers un modèle fonctionnel linguistique du type "Sens ⇔ Texte"], *Mašinnyj perevod i prikladnaja lingvistika*, n°11, p. 5–35.
- Mel'čuk Igor & Žolkovskij Aleksandr, (1970). Towards a Functioning Meaning-Text Model of Language, *Linguistics*, n°57, p. 10–47.
- Mel'čuk Igor, Iordanskaja Lidija & Arbatchewsky-Jumarie Nadia, (1981). Un nouveau type de dictionnaire : le dictionnaire explicatif et combinatoire du français contemporain, *Cahiers de lexicologie*, n°39, p. 3-34.
- Mel'čuk Igor, Arbatchewsky-Jumarie Nadia, Iordanskaja Lidija, Mantha Suzanne & Polguère Alain, (1984). *Dictionnaire explicatif et combinatoire du français contemporain : recherches lexico-sémantiques I*, Les Presses de l'Université de Montréal.
- Mel'čuk Igor, (1993). La phraséologie et son rôle dans l'enseignement-apprentissage d'une langue étrangère, *Études de linguistique appliquée*, n°92, p. 82-113.
- Mel'čuk Igor, Clas André & Polguère Alain, (1995). *Introduction à la lexicologie explicative et combinatoire*, Éditions Duculot/AUPELF-UREF (Louvain-la-Neuve/Paris).
- Mel'čuk Igor, (1998). Collocations and lexical functions, dans Cowie Anthony-Paul (Éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, p. 23-53.

- Mel'čuk Igor, Arbatchewsky-Jumarie Nadia, Iordanskaja Lidija, Mantha Suzanne & Polguère Alain (1999). *Dictionnaire explicatif et combinatoire du français contemporain : recherches lexico-sémantiques IV*, Les Presses de l'Université de Montréal.
- Meunier Annie, (1981). Nominalisations d'adjectifs par verbes supports, thèse, Université Paris 7.
- Mondria Jan-Arjen, (1996). *Vocabulaireverwerving in het vreemde-talenonderwijs*. Thèse de Doctorat, Université de Groningue.
- Moon Rosamund, (1992a). Textual Aspects of Fixed Expressions in Learners Dictionaries, dans Arnaud Pierre & Béjoint Henri (Éd.), *Vocabulary and applied linguistics*, Macmillan Academic and Professional, Ltd (U.K.), p. 13-27.
- Moon Rosamund, (1992b). There is Reason in the Roasting of Eggs: A Consideration of Fixed Expressions in Native-Speaker Dictionaries, dans Tommola Hannu, Varantola Krista, Salmi-Tolonen Tarja & Schopp Jürgen (Éd.), *Euralex 92 Proceedings*, Series translologica A, vol. II, University of Tampere (Tampere), p. 493-502.
- Moon Rosamund, (1998). Frequencies and forms of phrasal lexemes in English, dans Cowie Anthony-Paul (Éd.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, p. 79-100.
- Miller George & Gildea Patricia, (1987). How children learn words, *Scientific American*, vol. 257(3), p. 86-91.
- Mondada Lorenza, (2008). Documenter l'articulation des ressources multimodales dans le temps : la transcription d'enregistrements vidéos d'interactions, dans Bilger Mireille (Éd.), *Données orales : les enjeux de la transcription*, Presses Universitaires de Perpignan, n°37, p. 127-156.
- Mongin Pierre, Delengaigne Xavier & Garcia Luis, (2011). *Organisez vos notes avec le Mind Mapping – Dessinez vos idées !*, Dunod (Paris).
- Moulin André, (1984). The problem of vocabulary teaching: An answer to Robbins Burling, *System*, vol. 12(2), p. 147-149.



---

Murison-Bowie Simon, (1993). *MicroConcord Manual, an Introduction to the Practices and Principles of Concordancing in Language Teaching*, Oxford University Press.

## N

Nagy William, Herman Patricia & Anderson Richard, (1985). Learning words from context, *Reading Research Quarterly*, n°20, p. 223-253.

Nagy William & Herman Patricia, (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction, dans McKeown Margaret & Curtis Mary (Éd.), *The Nature of Vocabulary Acquisition*, p. 19-35, Lawrence Erlbaum (Hillsdale, New Jersey).

Nation Paul & Coady James, (1988). Vocabulary and reading, dans Carter Ronald & McCarthy Michael (Éd.), *Vocabulary and Language Teaching*, Longman (Londres), p. 97-110.

Nation Paul, (1990). *Teaching and Learning Vocabulary*, Newbury House (New York).

Nation Paul & Jonathan Newton, (1997). Teaching vocabulary, dans Coady James & Thomas Huckin (Éd.), *Second Language Vocabulary Acquisition*, Cambridge University Press, p. 238-254.

Nation Paul, (2001). *Learning vocabulary in another language*, Cambridge University Press.

Nattinger James & DeCarrico Jeanette, (1992). *Lexical Phrases and Language Teaching*, Oxford University Press.

Nenonen Marja, Niemi jussi & Laine Matti, (2002). Representation and processing of idioms: Evidence from aphasia, *Journal of Neurolinguistics*, n°15, p. 43-58. [https://www.academia.edu/684569/Representation\\_and\\_processing\\_of\\_idioms\\_evidence\\_from\\_aphasia](https://www.academia.edu/684569/Representation_and_processing_of_idioms_evidence_from_aphasia) (consulté le 10/10/2019).

Nerima Luka, Seretan Violeta & Wehrli Eric, (2006). Le problème de collocation en TAL, *Nouveaux cahiers de linguistiques Française*, p. 95-115.

Nesselhauf Nadja. (2005). *Collocation in a Learner Corpus*. John Benjamins Publishing Company (Amsterdam / Philadelphia).

---

Neveu Franck, (2004). *Dictionnaire des sciences du langage*, Armand Colin (Paris).

Novak Joseph & Gowin Bob, (1984). *Learning how to learn*, Cambridge University Press.

O

Ogden Charles-Kay (1934). *The system of Basic English*, Harcourt Brace and Co (New York).

Ossipow Laurence, (1997). La cuisine du corps et de l'âme : approche ethnologique du végétarisme, du crudivorisme et de la macrobiotique en Suisse, Maison des sciences de l'homme (Paris).

O'sullivan Ide & Chambers Angela, (2006). Learners' writing skills in French: Corpus consultation and learner evaluation, *Journal of Second Language Writing*, vol. 15(1), p. 49-68. En ligne sur <https://www.sciencedirect.com/science/article/abs/pii/S1060374306000038> (consulté le 10/10/2019).

O'sullivan Ide, (2007). Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy, *ReCALL*, vol. 19(3), p. 269-286. En ligne sur <https://www.cambridge.org/core/journals/recall/article/enhancing-a-process-oriented-approach-to-literacy-and-language-learning-the-role-of-corpus-consultation-literacy/A0A8096A78A86D6B3DBE01F01BDA226D> (consulté le 10/10/2019).

O'sullivan Ide, (2010). Using corpora to enhance learners' academic writing skills in French, *Revue française de linguistique appliquée*, n°15, p. 21- 35. En ligne sur <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2010-2-page-21.htm> (consulté le 10/10/2019).

P

Paribakht Sima & Wesche Marjorie, (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition, dans Coady James & Thomas Huckin (Éd.), *Second Language Vocabulary Acquisition*, Cambridge University Press, p.174-200.

- Pawley Andrew & Syder Frances Hodgetts, (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency, dans Richards Jack & Schmidt Richard (Éd.), *Language and communication*, Longman (London), p. 191-225.
- Pawley Andrew, (2001). Phraseology, Linguistics and the Dictionary, *International Journal of Lexicography*, vol. 14(2), p. 122-134. En ligne sur <https://academic.oup.com/ijl/article/14/2/122/957811> (consulté le 10/10/2019).
- Pearson Jennifer, (1998). *Terms in context*, John Benjamins (Amsterdam/Philadelphia).
- Pecman Mojca, (2004). Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique. Thèse de doctorat, Université de Nice Sophia Antipolis.
- Peters Martine, Chevrier Jacques, Leblanc Raymond, Forlin Gilles & Malette Judith, (2005). Compétence réflexive, carte conceptuelle et webfolio à la formation des maîtres, *Canadian Journal of Learning and Technology*, vol. 31(3). En ligne sur <https://www.cjlt.ca/index.php/cjlt/article/view/26491/19673> (consulté le 10/10/2019).
- Picoche Jacqueline, (1992). *Précis de lexicologie française : L'étude et l'enseignement du vocabulaire*, Nathan (Paris).
- Pincemin Bénédicte, (1999). Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents. Thèse de Doctorat en Linguistique, Université Paris IV Sorbonne.
- Pincemin Bénédicte, (2006). Concordances et concordanciers -De l'art du bon KWAC, Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation, Actes du XVIIe Colloque d'Albi *Langages et Signification* (Albi), p. 33-42.
- Planck Max, (2008). ELAN Version 3.6.0. En ligne sur <https://tla.mpi.nl/tools/tla-tools/elan/> (consulté le 10/10/2019).
- Plane Sylvie, (2004). Un consensus à interroger : la question des "mots difficiles". Quelques éléments d'une enquête sur le repérage des problèmes lexicaux, Actes du 9e colloque de l'AIRDF (Association

- internationale pour la recherche en didactique du français, Québec), 26, 27 et 28 août. En ligne sur [http://www.colloqueairdf.fse.ulaval.ca/fichier/Symposium\\_Ronveaux-Dolz/sylvie-plane.pdf](http://www.colloqueairdf.fse.ulaval.ca/fichier/Symposium_Ronveaux-Dolz/sylvie-plane.pdf) (consulté le 10/10/2019).
- Polguère Alain, (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. *Proceedings of EURALEX'2000* (Stuttgart), p.517-527.
- Polguère Alain, (2003). Collocations et fonctions lexicales : pour un modèle d'apprentissage, *les collocations, analyse et traitement*, De Werelt (Amsterdam), p. 23-34.
- Polguère Alain, (2016). *Lexicologie et sémantique lexicale. Notions fondamentales*. Troisième édition (première édition en 2003), les presses de l'Université de Montréal.
- Portillo Serrano, (2010). Problématique des genres dans les productions écrites universitaires : cas du résumé scolaire chez les étudiants français et mexicains. Thèse de Doctorat, Université de Franche-Comté.
- Potter Mary, So Kwok-Fai, Von Eckardt Barbara & Feldman Laurie, (1984). Lexical and conceptual Representation in Beginning and Proficient Bilinguals, *Journal of Verbal Learning & Verbal Behavior*, n°23, p. 23-28.
- Prince Peter, (1996). Second language vocabulary learning: The role of context versus translation as a function of proficiency, *Moderne Language Journal* n°80, p. 478–393.
- Prince Peter, (1999). L'apprentissage lexical en deuxième langue : des réseaux évolutifs, *Academic and Scientific Publishers (ASp)*, n°23-26, p. 335–347. En ligne sur <https://journals.openedition.org/asp/2596> (consulté le 10/10/2019).
- Qualls Constance Dean, Treaster Beth, Blood Gordon & Hammer Carol Scheffner, (2003). Lexicalization of idioms in urban fifth graders: A reaction time study, *Journal of Communication Disorders*, n°36, p. 245-261.

R

- Rastier François, (1998). Le Problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage, *Langages*, n°129, p. 97-111.
- Rastier François, (2001). Eléments de théorie des genres, *Texto !* En ligne sur [http://www.revue-texto.net/1996-2007/Inedits/Rastier/Rastier\\_Elements.html](http://www.revue-texto.net/1996-2007/Inedits/Rastier/Rastier_Elements.html) (consulté le 10/10/2019).
- Rastier François, (2002). Enjeux épistémologiques de la linguistique de corpus, dans Geoffrey Williams (Éd.), *Deuxièmes Journées de Linguistique de Corpus*, Lorient. En ligne sur [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html) (consulté le 10/10/2019).
- Raynal Françoise & Rieunier Alain, (1997). *Pédagogie, dictionnaire des concepts clés*, Editions Sociales Française (Paris).
- Régnier Faustine, (2004), *L'exotisme culinaire. Essai sur les saveurs de l'autre*, Le lien social, Maison des Sciences de l'Homme, Presses Universitaires de France.
- Rézeau Joseph, (1997). Applications des concordanciers à l'enseignement de la grammaire anglaise en DEUG, *ASp*, n°15-18, p. 117-132. En ligne sur <https://journals.openedition.org/asp/2997> (consulté le 10/10/2019).
- Rézeau Joseph, (2001). Médiatisation et médiation pédagogique dans un environnement multimédia. Le cas de l'apprentissage de l'anglais en Histoire de l'art à l'université. Thèse de doctorat en didactique de la langue, Université Victor Segalen Bordeaux 2. En ligne sur <https://tel.archives-ouvertes.fr/edutice-00000222/document> (consulté le 10/10/2019).
- Rézeau Joseph, (2002). Médiation, médiatisation et instruments d'enseignement : du triangle au « carré pédagogique », *ASp*, n°35-36, p. 183-200. En ligne sur <http://asp.revues.org/1656> (consulté le 10/10/2019).
- Rézeau Joseph, (2007). L'apport du concordancier à l'analyse et à la remédiation des erreurs des apprenants dans les forums de discussion en ligne. *Apprentissage des Langues et Systèmes d'Information et de Communication (Alsic)*, vol. 10(2), p. 27-43. En ligne sur <https://journals.openedition.org/alsic/561> (consulté le 10/10/2019).

---

Rolland Jean-Claude & Rolland-Pérez Daria, (2008). *De l'espagnol...au français, Expressions et mots usuels du français*, DeBoeck (Belgique).

Römer Ute, (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments, *Zeitschrift für Anglistik und Amerikanistik*, vol. 54(2), pp. 121-134.

Römer Ute, (2011). Corpus research applications in second language teaching, *Annual Review of Applied Linguistics*, n°31, p. 205-225.

Römer Ute, (2009). Corpus research and practice: what help do teachers need and what can we offer?, dans Aijmer Karin (Éd.), *Corpora and Language Teaching*, John Benjamins (Amsterdam/Philadelphia), p. 83-98.

## S

Sag Ivan, Baldwin Timothy, Bond Francis, Copestake Ann & Flickinger Dan, (2002). Multiword expressions: A pain in the neck for NLP, *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics* (Mexico City), p. 1-15.

Salaberry Rafael, (2001). The use of technology for second language learning and teaching: a retrospective, *Modern Language Journal*, vol. 85(1), p. 39-56.

Salem André, (1987). *Pratique des segments répétés : Essai de statistique textuelle*, Klincksieck (Paris).

Salem André, Lamalle Cédric, Fleury Serge Martinez & William Martinez, (2003). *Lexico 3 : Outils de statistique textuelle*. En ligne sur <http://lexi-co.com/ressources/manuel-3.41.pdf> (consulté le 10/10/2019).

Schneuwly Bernard & Dolz Joachim, (1997). Les genres scolaires : Des pratiques langagières aux objets d'enseignement, *Repères*, n°15, p. 27-40.

Schouten-Van Parreren, (1985). *Woorden leren in het vreemde-talenonderwijs*, Van Walraven (Apeldoorn).

Schouten-Van Parreren, (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts?, dans Nation

- 
- Paul & Carter Ron (Éd.), *AILA Review 6: Vocabulary Acquisition*, Free University Press (Amsterdam), p. 75-85.
- Scott Mike, (1998). *WordSmith Tools Version 3*, Oxford University Press.
- Scott Mike & Christopher Tribble (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. John Benjamins (Amsterdam/Philadelphia).
- Selva Thierry, Verlinde Serge & Binon Jean, (2004). ALFALEX, un environnement d'aide à l'apprentissage lexical du français langue étrangère. *Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et de l'Industrie*, Compiègne, Université de Technologie de Compiègne, p. 515-522.
- Shei Chris, (2008). Discovering the hidden treasure on the Internet: Using Google to uncover the veil of phraseology, *Computer Assisted language Learning*, vol. 21(1), p. 67-85.
- Silberztein Max, (2002). NooJ : an oriented object approach, dans Royauté Jean & Silberztein Max (Éd.), *INTEX pour la Linguistique et le Traitement Automatique des Langues*, Presses universitaires de Franche-Comté (Besançon).
- Silberztein Max & Tutin Agnès, (2005). NooJ, un outil TAL pour l'enseignement des langues : application pour l'étude de la morphologie lexicale en FLE, *Alsic*, vol. 8(2), p. 123-134. En ligne sur <https://edutice.archives-ouvertes.fr/edutice-00109619/document> (consulté le 10/10/2019).
- Silberztein Max, (2015). *La formalisation des langues : l'approche de NooJ*, International Society for Technology in Education (ISTE).
- Silberztein Max, (2016). *Formalizing Natural Languages: The NooJ Approach*, Wiley.
- Sinclair John McHardy, (1966). Beginning the Study of Lexis, dans Bazell Charles, Catford John, Halliday Michael & Robins Robert (Éd.), *In Memory of J.R. Firth*, Longmans (London/Harlow), p. 410-430.
- Sinclair John McHardy & Renouf Antoinette, (1988). A lexical syllabus for language learning, *Vocabulary and Language Teaching*, Longman (Harlow).

- Sinclair John McHardy, (1991a). *Corpus Concordance Collocation*, Oxford University Press.
- Sinclair John McHardy, (1991b). *Collins COBUILD English Grammar*. Collins (London/Glasgow).
- Sinclair John McHardy, (1996). Preliminary Recommendations on Corpus Typology, Rapport technique, *Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale*. En ligne sur <http://www.ilc.cnr.it/EAGLES96/corpus/corpus.html> (consulté le 10/10/2019).
- Sinclair John McHardy, (2004). *How to use corpora in language teaching*, John Benjamins (Amsterdam/Philadelphia).
- Sinclair John McHardy, (2005). Corpus and text-basic principles, dans Wynne Martin (Éd.), *Developing Linguistic Corpora: A Guide to Good Practice*, Oxbow Books (Oxford), p. 1-16.
- Sinclair Stéfán & Rockwell Geoffrey, (2014). Les potentialités du texte numérique, dans Sinatra Michael, Vitali-Rosati Marcello (Éd.), *Pratiques de l'édition numérique*, collection Parcours Numériques, Les Presses de l'Université de Montréal, p. 191-204.
- Sitri Frédérique & Tutin Agnès, (2016). Phraséologie et genre de discours : Présentation, *Lidil*, n°53, p. 5-18. En ligne sur <http://journals.openedition.org/lidil/3914> (consulté le 10/10/2019).
- Siyanova Anna & Schmitt Nobert, (2008). L2 learner production and processing of collocation: A multi-study perspective, *The Canadian Modern Language Review*, n°64, p. 429-458. En ligne sur <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.722.789&rep=rep1&type=pdf> (consulté le 10/10/2019).
- Siyanova-Chanturia Anna, Conklin Kathy & Van Heuven Walter, (2011). Seeing a phrase « time and again » matters: the role of phrasal frequency in the processing of multiword sequences, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 37(3), p. 776-784. En ligne sur <https://pdfs.semanticscholar.org/f4ae/78f8bdec94bfb3338a9a9ccb187e0996a508.pdf> (consulté le 10/10/2019).
- Skehan Peter, (1998). *A Cognitive Approach to Language Learning*. Oxford University Press.



---

Smadja Frank, (1993). Retrieving collocations form text: Xtract, *Computational Linguistics*, vol. 19(1), p. 143-177

Sperberg-McQueen Michael & Burnard Lou, (1999). *Guidelines for Electronic Text Encoding and Interchange: TEI P3 Text Encoding Initiative*, Revised reprint (Oxford).

Sun Yu-Chih, (2007). Learner Perceptions of a Concordancing Tool for Academic Writing, *Computer Assisted Language Learning*, vol. 20(4), p. 323-343.

Sweet Henry, (1891). *A new English Grammar, logical and historical*, Oxford University Press.

Swinney David & Cutler Anne, (1979). The access and processing of idiomatic expressions, *Journal of Verbal Learning & Verbal Behavior*, vol. 18(5), p. 523-534.

## T

Tellier Marion, (2006). L'impact du geste pédagogique sur l'enseignement/apprentissage des langues étrangères : Etude sur des enfants de 5 ans. Thèse de Doctorat en linguistique, Université Paris Diderot - Paris 7.

Thun Harald, (1975). Quelques relations systématiques entre groupements de mots figés, *Cahiers de lexicologie*, n°27, p. 52-71.

Tognini-Bonelli Elena, (1993). *Rationale and Aims of Corpus Linguistics, Internal memorandum*, Corpus Linguistics Group (Birmingham).

Tognini-Bonelli Elena, (2001). *Corpus linguistics at work*, John Benjamins (Amsterdam/Philadelphia).

Tomasello Michael, (1998). *The new psychology of language. Cognitive and Functional Approaches to Language Structure*, Lawrence Erlbaum Associates (Mahwah, New Jersey).

Tomasello Michael, (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press.

- 
- Tournier Maurice, (2002). *Des sources du sens - Propos d'étymologie sociale*, vol. 3, Collection Langages École Normale Supérieure de Lyon.
- Tran Thi Thu Hoai, (2014). *Description de la phraséologie transdisciplinaire des écrits scientifiques et réflexions didactiques pour l'enseignement à des étudiants non-natifs. Application aux marqueurs discursifs*. Thèse de Doctorat, Université Grenoble Alpes. En ligne sur <https://tel.archives-ouvertes.fr/tel-01330952/> (consulté le 10/10/2019).
- Tribble Christopher & Jones Glyn, (1990). *Concordances in the classroom: A resource book for teachers*. Longman (London).
- Trier Jost, (1931), *Der deutsche Wortschatz im Sinnbezirk des Verstandes. Dies Geschichte eines sprachlichen Feldes*, Winter (Heidelberg).
- Tutin Agnès, (2004). Pour une modélisation dynamique des collocations dans les textes, *Actes d'Euralex* (Lorient), p. 207-221.
- Tutin Agnès, (2007a). Autour du lexique et de la phraséologie des écrits scientifiques, *Revue française de linguistique appliquée* 2007/2, n°12, p. 5-14.
- Tutin Agnès, (2007b). Modélisation linguistique et annotation des collocations : application au lexique transdisciplinaire des écrits scientifiques, dans Koeva Svetta, Maurel Denis & Silberztein Max (Éd.), *Formaliser les langues avec l'ordinateur*, Presses Universitaires de Franche-Comté (Besançon).
- Tutin Agnès, (2010). Le traitement des collocations dans les dictionnaires monolingues de collocations du français et de l'anglais, dans Neveu Franck, Durand Jacques, Klingler Thomaks, Mondada Lorenza, Prévost Sophie & Toke Muni (Éd.), *2ème Congrès Mondial de Linguistique Française* (Paris), p. 1075-1090. En ligne sur [http://www.linguistiquefrancaise.org/articles/cmlf/pdf/2010/01/cmlf2010\\_000141.pdf](http://www.linguistiquefrancaise.org/articles/cmlf/pdf/2010/01/cmlf2010_000141.pdf) (consulté le 10/10/2019).
- Tutin Agnès, (2014). La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétoriques, dans Tutin Agnès & Grossmann Francis (Éd.), *L'écrit scientifique : du lexique au discours. Autour de scientext*, Presses Universitaires de Rennes, p. 27-44.

---

Tyne Henri, (2009). Corpus oraux par et pour l'apprenant, *Mélanges CRAPEL*, n°31, p. 91-111. En ligne sur <http://www.atilf.fr/spip.php?article3728> (consulté le 10/10/2019).

Tyne Henri, (2012). Corpus work with ordinary teachers: data-driven learning activities, dans Thomas James & Boulton Alex (Éd.), *Input, process and product: developments in teaching and language corpora*, Masaryk University Press (Brno), p. 114-129.

## U

Underwood Geoffrey, Schmitt Norbert & Galpin Adam, (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences, dans Schmitt Norbert (Éd.), *Formulaic sequences: acquisition, processing and use*, John Benjamins (Amsterdam/Philadelphia), p. 153-172.

## V

Van der Linden Elisabeth, (2006). Lexique mental et apprentissage des mots, *Revue française de linguistique appliquée*, vol. 11(1), p. 33-44.

Van Lancker-Sidtis Diana & Rallon Gail, (2004). Tracking the incidence of formulaic expressions in everyday speech: methods for classification and verification, *Language & Communication*, vol. 24(3), p. 207-240.

Vanni Laurent & Mittmann Adiel, (2016). Cooccurrences spécifiques et représentations graphiques, le nouveau 'thème' d'Hyperbase, actes de *13<sup>ème</sup> Journées internationales d'Analyse statistique des Données Textuelles* (Nice). En ligne sur <https://hal.archives-ouvertes.fr/hal-01359413/document> (consulté le 10/10/2019).

Verlinde Serge, Binon Jean & Selva Thierry, (2006). Corpus, collocations et dictionnaires d'apprentissage, *Langue française*, n°150, p. 84-98.

Vinogradov Viktor, (1946). *Osnovnij ponjatija russkoj fraseologii, kak lingvisticeskoj discipliny*, Trudy jubilejnoj naucnoj sessi, Leningrad.

Vinogradov Viktor, (1947). *Ob osnovnij Tipaj fraseologicheskij iedinits v russkom jazike*, Trudy komissii po istorii Akademii Nauk SSSR (Moscou).

## W

- 
- Weinreich Uriel, (1953). *Languages in Contact*, Publications of the Linguistic Circle of New York.
- Widdowson Henry, (2000). On the limitations of linguistics applied, *Applied Linguistics*, vol. 21(1), p. 3-25.
- Williams Geoffrey, (1999). Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique. Thèse de doctorat, Université de Nantes.
- Williams Geoffrey, (2001). Sur les caractéristiques de la collocation, *Actes de TALN*, Tours, p. 9-16.
- Williams Geoffrey, (2003). Les collocations et l'école contextualiste britannique, dans Grossman Francis & Tutin Agnès (Éd.), *les collocations, analyse et traitement*, De Werelt (Amsterdam), p. 33-44.
- Williams Geoffrey, (2005). Introduction, dans Williams Geoffrey (Éd.), *la linguistique de corpus*, Presses Universitaires de Rennes, p. 13-18.
- Williams Geoffrey, (2006). La linguistique de corpus : une affaire prépositionnelle, dans Rastier François & Ballabriga Michet (Éd.), *Corpus en lettres et sciences sociales : des documents numériques à l'interprétation*, Actes du colloque international d'Albi, Texto (Paris), p. 151-158.
- Willis Dave, (1990). *The Lexical Syllabus. A New Approach to Language Teaching*, Harper Collins (London).
- Willis Dave, (2003). *Rules, Patterns and Words*, Cambridge University Press.
- Wilson Anne, (1977). *Les Tables du Monde : Recettes et Ingrédients de A à Z*. Könnemann (Cologne).
- Wood David, (2002). Formulaic language in acquisition and production: Implications for teaching, *TESL Canada Journal*, n°20, p. 1-15. En ligne sur <https://files.eric.ed.gov/fulltext/EJ659409.pdf>.
- Wood David, (2010). *Perspectives on formulaic language: Acquisition and communication*, Continuum (London/New York).

---

Wray Alison & Perkins Michael, (2000). The functions of formulaic language: an integrated model, *Language & Communication*, n°20, p. 1-28. En ligne sur <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.558.6248&rep=rep1&type=pdf> (consulté le 10/10/2019).

Wray Alison, (2002). *Formulaic language and the lexicon*, Cambridge University Press.

Wray Alison, (2007). Needs only' analysis in linguistic ontogeny and phylogeny, dans Lyon Caroline, Nehaniv Chrystopher & Cangelosi angelo (Éd.), *Emergence of communication and language* (Part 1), Springer Verlag (London), p. 53-70.

Y

Yang Tong, (2015). *L'enseignement du lexique culinaire auprès d'apprenants étrangers de niveau A2*, Sous la direction de Cavalla Cristelle, mémoire du Master 2 en didactique du Français Langue Étrangère, Université Sorbonne-Nouvelle Paris 3.

Yang Tong, (2016). Cuisitext : un corpus écrit et oral pour l'enseignement, communication présentée au *Langues sur objectifs spécifiques : perspectives croisées entre linguistique et didactique - LOSP*, Novembre 2016, Grenoble (France). En ligne sur <http://losp2016.u-grenoble3.fr/> (consulté le 10/10/2019).

Yang Tong, (2017a, novembre). L'apport des concordanciers dans l'apprentissage/enseignement du lexique, communication présentée au *Colloque Jeune Chercheur* (CJC) organisé par le laboratoire DyLis (Dynamique du langage in Situ) de l'Université de Rouen les 23 et 24 novembre 2017. En ligne sur <https://cjcrouen2017.sciencesconf.org> (consulté le 10/10/2019).

Yang Tong, (2017b). La carte mentale : un outil didactique d'aide à la mémorisation, pour l'enseignement/apprentissage du lexique, dans l'acte du colloque *Rencontre des Jeunes Chercheurs* (Juin 2017) organisé par l'ED 268 de l'Université Sorbonne Nouvelle. En ligne sur <https://hal.archives-ouvertes.fr/hal-02013296/document> (consulté le 10/10/2019)

Yang Tong, (2018a). Automatic Extraction of the Phraseology Through NooJ, dans Mbarki Samir, Mourchid Mohammed & Silberztein Max (Éd.)

---

*Formalizing Natural Languages with NooJ and Its Natural Language Processing Applications*. NooJ 2017, *Communications in Computer and Information Science*, n°811, p.167-177.

Yang Tong, (2018b). Proposition de trois approches d'utilisation du corpus pour l'enseignement/apprentissage des langues présentée au colloque international du REMADDIF sur le thème Enjeux, réflexions et pratique en didactique des langues et des cultures, Kénitra (Maroc), les 25, 26 et 27 octobre. En ligne sur <https://remaddif2018.sciencesconf.org/resource/page/id/13> (consulté le 10/10/2019).

Yang Tong & Cavalla Cristelle, (2019). L'enseignement de la terminologie de la cuisine auprès des apprenants de niveau A2, *Revue ReDILLeT (Revista Digital Internacional de Lexicología, Lexicografía y Terminología)*, n°2, Cordoba (Argentine).

Yang Tong, (2019a). Enseignement/apprentissage des patternes grammaticaux du domaine culinaire auprès d'apprenants étrangers, la quatrième édition du colloque les terminologies professionnelles de la gastronomie et de l'œnologie à Innsbruck (Autriche), les 19 et 20 septembre. En ligne sur <https://www.uibk.ac.at/congress/gastro2019/index.html.fr> (consulté le 10/10/2019).

Yang Tong, (2019b). Enseignement/apprentissage des phrasèmes verbaux du domaine culinaire auprès d'apprenants étrangers au 87<sup>ème</sup> Congrès de l'ACFAS (association francophone pour le savoir) à Gatineau (Canada), les 29, 30 et 31 mai. En ligne sur <https://www.acfas.ca/evenements/congres/programme/87/500/503/d> (consulté le 10/10/2019).

Yang Tong, (2019c). Automatic Extraction of Verbal Phrasemes in the Culinary Field with NooJ, dans Mirto Ignazio Mauro, Monteleone Mario, Silberztein Max (Éd.) *Formalizing Natural Languages with NooJ and Its Natural Language Processing Applications*, NooJ 2018 *Communications in Computer and Information Science*, n°987, p.83-94.

Yang Tong, (2019d). Automatic extraction of grammatical patterns of verbs in the culinary field with NooJ au colloque international NooJ 2019, Hammamet (Tunisie), les 7, 8 et 9 juin. En ligne sur [http://www.nooj-association.org/nooj2019/index.php?option=com\\_content&view=article&id=4&Itemid=101](http://www.nooj-association.org/nooj2019/index.php?option=com_content&view=article&id=4&Itemid=101) (consulté le 10/10/2019).

Yeh Yuyi, Liou Hsien-Chin & Li Yi-Hsin, (2007). Online synonym materials and concordancing for EFL college writing, *Computer Assisted Language Learning*, vol. 20(2), p. 131-152.

Yoon Hyunsook & Hirvela Alan, (2004). ESL student attitudes towards corpus use in L2 writing, *Journal of Second Language Writing*, n°13, p. 257-283.

Yoon Hyunsook, (2008). More than a linguistic reference: the influence of corpus technology on L2 academic writing, *Language Learning & Technology*, vol. 12(2), p. 31-49.

## Z

Zanettin Federico, (2002). Corpora for translation practice, dans Yuste-Rodrigo Elia (Éd.), *Language resources for translation work and research, LREC 2002 Workshop Proceedings*, Las Palmas de Gran Canaria. p. 10-14.

Zinglé Henri, (2003). Constitution et exploitation d'une base de données phraséologiques en français, *Congrès international des linguistes* (Prague), 24-29 juillet.

## Annexe

## Annexe 1 Corpus oraux accessibles/inaccessible en ligne

Corpus inaccessible en ligne		Corpus accessible en ligne	
Nom du corpus oral	Perspectives	Nom du corpus oral	Perspectives
Corpus de S. Mellet (Université de Nice)	Variations régionales	CFPP2000 (Corpus de Français Parlé Parisien, Université de Paris 3)	Variations parisiennes
Corpus de Barbéris, Boyer (Université de Montpellier)	Analyse de discours	CO2 (Université de Tours et d'Orléans)	Analyse de discours
Corpus de Françoise Gadet (Université de Nanterre)	Sociolinguistique, français « populaire »	ANCOR (Université de Tours, d'Orléans et de Paris 3)	Analyse de discours
Corpus de Durand et Laks (Université de Toulouse, Université de Paris X et Université d'Oslo )	Phonologie	PFC (Université de Toulouse 2)	Variations francophones
Corpus du groupe de Mary-Annick Morel, Université de Paris III	Prosodie		
Corpus de Pierre Bange (Lyon II)	Interaction		
Corpus du Groupe de Recherches sur les Interactions Communicatives, GRIC	Interaction		



---

## Annexe 2 Convention de transcription

### 1. Identité du participant

Chaque locuteur est désigné par les trois caractères du métier, comme cuisinier est identifié par CUI1 ou CUI2, CUI3 (on ne fait pas des distinctions entre cuisiniers, sauciers et pâtisseries dans notre étude), s'il y a plusieurs cuisiniers dans une séquence. Comme la ligne commence par un astérisque (\*) et les trois caractères sont suivi de deux points (:) dans CLAN, chaque participant se présente dans une combinaison des trois éléments (un astérisque, trois caractères et deux points) avant les énoncés, par exemple, une combinaison de \*CUI représente l'énoncé d'un cuisinier.

### 2. Tour de parole

Les participants sont identifiés en amont de leurs propres énoncés. Pour que la transcription soit exacte, l'énoncé pourrait être coupé de façon régulière.

### 3. Pause

Une pause brève est assignée par une virgule et une pause longue est indiquée par un point.

### 4. Chevauchement

Lorsqu'un locuteur L1 s'exprime et qu'un autre locuteur L2 l'interrompt, les interventions de L2 seront transcrites entre chevrons (< >) à l'intérieur des énoncés de L1, si le dernier reprend la parole.

### 5. Ponctuation

À part la virgule (pause courte) et le point (pause longue), le point interrogatif à la fin d'un énoncé forme une question. Le point exclamatif s'emploie à la fin d'une exclamation ! Un énoncé en suspens (incomplet) est suivi de trois points.

### 6. Mots composés

Les noms propres seront transcrits avec un tiret en bas (\_), par exemple, marie\_ange.

Les noms propres ne portent pas de majuscule sur l'initiale.

---

## Table des matières

Liste des tableaux.....	9
Liste des figures .....	11
Introduction.....	14
Première partie : théorie.....	22
<b>1. Linguistique de corpus .....</b>	<b>26</b>
1.1 Émergence de la linguistique de corpus .....	26
1.2 Construction et analyse du corpus à l'aide d'outils TAL.....	27
1.3 Élaboration de grands corpus français et anglais .....	29
1.4 Choix des types de textes pour la constitution du corpus.....	31
1.4.1 Importance du corpus oral .....	31
1.4.2 Genres de textes .....	33
1.4.3 Documents authentiques pour l'enseignement .....	35
<b>Bilan .....</b>	<b>37</b>
<b>2. Phraséologie .....</b>	<b>39</b>
2.1 Champs de recherche : la phraséologie .....	39
2.2 Classifications des phrasèmes .....	43
2.3 Critères de la phraséologie.....	44
2.3.1 Fréquence .....	45
2.3.2 Figement .....	46
2.3.3 Critère pragmatique .....	46
2.4 Modélisations des données dans des projets phraséologiques.....	47
2.5 Logiciel de la description des langues : approche de <i>NooJ</i> .....	49
2.5.1 Grammaire rationnelle .....	50
2.5.2 Grammaire hors contexte ou algébrique .....	55
2.5.3 Grammaire contextuelle.....	56

---

<b>Bilan .....</b>	<b>59</b>
<b>3. Utilisation du corpus pour l'E/A de la phraséologie .....</b>	<b>60</b>
<b>3.1 E/A de la phraséologie .....</b>	<b>60</b>
3.1.1 Nécessité de l'E/A de la phraséologie en FLE .....	60
3.1.2 E/A des phrasèmes dans les ouvrages de FLE .....	62
3.1.3 Acquisition des langues en L2 .....	65
3.1.3.1 Organisation lexicale et lexique mental.....	65
3.1.3.2 Lexique mental en L1 .....	67
3.1.3.3 Lexique mental en L2 .....	68
3.1.4 E/A du lexique au genre.....	71
3.1.5 Prise en compte du critère de fréquence dans la sélection du lexique ....	73
3.1.6 Entrée didactique pour l'enseignement du lexique .....	75
<b>3.2 Utilisation des corpus pour l'E/A des langues .....</b>	<b>78</b>
3.2.1 Corpus pour l'E/A des langues.....	78
3.2.2 Trois manières d'enseigner l'utilisation des corpus .....	79
3.2.3 Approches d'utilisation du corpus pour l'E/A des langues .....	81
3.2.3.1 Approche <i>data-driven learning</i> (Johns, 1991a ; 1991b) .....	81
3.2.3.2 Apprentissage sur corpus (Boulton & Tyne, 2014) .....	82
3.2.3.3 Consultation directe et indirecte (Chambers, 2010) .....	83
<b>3.3 Outils didactiques pour l'E/A de la phraséologie .....</b>	<b>84</b>
3.3.1 Concordancier .....	84
3.3.1.1 Présentation de concordancier.....	84
3.3.1.2 Outil représentatif du TAL pour l'enseignement (Yang, 2017a) .....	87
3.3.2 Carte mentale .....	89
3.3.2.1 Présentation de la carte mentale.....	89
3.3.2.2 Outil d'aide à la mémorisation du lexique (Yang, 2017b).....	91
<b>Bilan .....</b>	<b>92</b>
<b>Deuxième partie : méthodologie et analyse linguistique .....</b>	<b>94</b>

---

<b>4. Constitution de <i>Cuisitext</i> et réflexion sur son exploitation .....</b>	<b>98</b>
<b>4.1 Constitution de corpus écrit.....</b>	<b>98</b>
4.1.1 Collecte de données .....	98
4.1.2 Nettoyage et exportation de données .....	101
<b>4.2 Construction de corpus oraux.....</b>	<b>102</b>
4.2.1 Support d'enregistrement.....	102
4.2.2 Cadre juridique pour le corpus .....	104
4.2.2.1 Domaine public ou droit d'auteur.....	104
4.2.2.2 Préparation d'un consentement .....	105
4.2.3 Convention de transcription .....	107
4.2.3.1 Choix de type de transcription.....	107
4.2.3.2 Choix du logiciel .....	109
4.2.4 Exemple d'extrait de transcription .....	111
<b>4.3 Réflexion sur l'exploitation de <i>Cuisitext</i> dans notre projet .....</b>	<b>112</b>
<b>Bilan .....</b>	<b>113</b>
<b>5. Repérage des phrasèmes saillants et leur sélection pour</b>	
<b>l'enseignement.....</b>	<b>115</b>
<b>5.1 Objet d'étude : phrasèmes <i>NAdj</i> .....</b>	<b>115</b>
<b>5.2 Extraction des phrasèmes <i>NAdj</i>.....</b>	<b>117</b>
5.2.1 Modélisation lexicale des données .....	118
5.2.1.1 Modélisation du syntagme nominal .....	119
5.2.1.2 Modélisation de la structure attributive.....	119
5.2.1.3 Exigences de notre modélisation des quatre grammaires .....	121
5.2.2 Choix du logiciel d'extraction .....	123
5.2.3 Implémentation des données lexicales dans <i>NooJ</i> .....	127
5.2.3.1 Implémentation du syntagme nominal.....	127
5.2.3.2 Implémentation de la structure attributive .....	128
<b>5.3 Identification et sélection des phrasèmes <i>NAdj</i> pour l'enseignement ..</b>	<b>132</b>

---

<b>Bilan .....</b>	<b>137</b>
<b>6. Analyse linguistique des phrasèmes <i>NAdj</i> retenus.....</b>	<b>138</b>
6.1 Critères servant à la classification .....	138
6.2 Propriétés transformationnelles du figement des phrasèmes <i>NAdj</i> .....	139
6.3 Classification des phrasèmes <i>NAdj</i> .....	142
6.4 Analyse du figement des phrasèmes <i>NAdj</i> .....	143
<b>Bilan .....</b>	<b>149</b>
<b>Troisième partie : proposition didactique .....</b>	<b>151</b>
<b>7. Propositions des approches d'intégration du corpus et leurs mises en œuvre pour l'E/A des phrasèmes <i>NAdj</i> .....</b>	<b>153</b>
7.1 Trois approches proposées.....	155
7.2 Piste d'apprentissage progressif sur corpus .....	157
7.3 Propositions didactiques.....	159
7.3.1 Deux séances de formation .....	160
7.3.1.1 Première séance autour de la lexie HACHÉ .....	160
7.3.1.2 Deuxième séance à partir des champs sémantiques de HACHÉ.....	168
7.3.2 Conseils pour la mise en pratique de l'AIP.....	179
<b>Bilan .....</b>	<b>180</b>
<b>Conclusion générale.....</b>	<b>182</b>
<b>Perspectives et limites de recherche .....</b>	<b>184</b>
<b>Bibliographie .....</b>	<b>187</b>
<b>Annexe .....</b>	<b>231</b>
<b>Annexe 1 Corpus oraux accessibles/inaccessible en ligne .....</b>	<b>231</b>
<b>Annexe 2 Convention de transcription .....</b>	<b>232</b>

## Résumé

Ce projet de thèse s'inscrit dans le cadre de l'enseignement du FOS (Français sur Objectifs Spécifiques) à des cuisiniers étrangers venus travailler dans des restaurants français ou ayant choisi la restauration comme spécialité. L'objectif de notre recherche est donc d'enseigner les phrasèmes *NAdj* du domaine culinaire auprès d'apprenants étrangers niveau A2. L'enseignement/apprentissage de la phraséologie s'avère nécessaire dans les langues de spécialités et la haute fréquence des phrasèmes *NAdj* a attiré notre attention. Plusieurs questions sont alors abordées : où trouver ce lexique spécifique ? Comment les extraire ? Par quelle approche enseignons-nous les phrasèmes sélectionnés ? Pour répondre à ces questions, nous avons fabriqué notre propre corpus *Cuisitext* – écrit et oral – puis nous avons utilisé *NooJ* pour extraire les phrasèmes *NAdj* du corpus. Enfin, nous avons proposé les trois approches d'utilisation des corpus pour l'enseignement/apprentissage des phrasèmes *NAdj* : approche inductive guidée, approche déductive, approche inductive pure.

## Mot-clé

*Cuisitext*, enseignement/apprentissage de la phraséologie, phrasèmes *NAdj*, français sur objectifs spécifiques, français langue étrangère, extraction automatique de la phraséologie par *NooJ*, approche inductive guidée, approche inductive pure, approche déductive, approches d'utilisation du corpus pour l'enseignement/apprentissage, concordancier, carte mentale.

## Titre

Constitution et exploitation d'une base de données pour l'enseignement/apprentissage des phrasèmes *NAdj* du domaine culinaire français auprès d'apprenants non-natifs

## Abstract

This thesis project aims to study the teaching method of FOS (French on Specific Objectives) catering to foreign cooks who come to work in French restaurants or who have chosen catering as a specialty. The objective of our research is therefore to teach the culinary *NAdj* phrasemes to foreign A2 level learners. The teaching/learning of phraseology is required in specialty languages and the high frequency of *NAdj* phrasemes has caught our attention. Several questions are then addressed: where to find this specific lexicon? How to extract them? By which approach do we teach the selected phrasemes? To answer these questions, we made our own corpus *Cuisitext* - written and oral - and then used *NooJ* to extract the *NAdj* phrasemes from the corpus. Finally, we have proposed the three approaches to the use of corpora for the teaching/learning of *NAdj* phrasemes: guided inductive approach, deductive approach, pure inductive approach.

## Key words

*Cuisitext*, Teaching/Learning of phraseology, *NAdj* Phrasemes, French on Specific Objectives, French as a Foreign Language, Automatic Extraction of the Phraseology Through *NooJ*, Guided Inductive Approach, Pure Inductive Approach, Deductive Approach, Approaches to the Use of the Corpus for Teaching/Learning, Concordance Lines, Mind Mapping.

## Title

Creation and Exploitation of a Database for the Teaching/Learning of *NAdj* Phrasemes of the French Culinary Field with Non-Native Learners

École doctorale 268  
4, rue des irlandais, 75005, PARIS