



HAL
open science

Lexique transdisciplinaire et enseignement aux étudiants allophones

Cristelle Cavalla

► **To cite this version:**

Cristelle Cavalla. Lexique transdisciplinaire et enseignement aux étudiants allophones. Agnès Tutin; Marie-Paule Jacques. Lexique transversal et formules discursives des sciences humaines, ISTE Editions, pp.191-214, 2018. hal-01917082

HAL Id: hal-01917082

<https://hal.science/hal-01917082>

Submitted on 16 Nov 2018

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.

Chapitre 9

Lexique transdisciplinaire et enseignement aux étudiants allophones

Cristelle Cavalla (Diltec EA2288, Université Sorbonne Nouvelle Paris 3)

Résumé

Dans ce chapitre nous aborderons l'enseignement du Lexique Scientifique Transdisciplinaire auprès d'étudiants et d'enseignants allophones. Les étudiants ont pour objectif la rédaction d'un écrit long universitaire en français tandis que les enseignants découvrent ce lexique pour son enseignement ensuite auprès d'étudiants. Pour cet enseignement, nous utiliserons les interfaces des corpus numériques développés dans le cadre de TermITH et Scientext et nous guiderons les apprenants – grâce à ces outils – vers la découverte de la phraséologie transdisciplinaire. Nous nous interrogerons alors sur le type d'éléments à enseigner, sur l'aide à leur mémorisation et enfin, nous présenterons notamment une séquence didactique mise en place auprès d'étudiants de Master.

Abstract

In this chapter we deal with the teaching of the Scientific interdisciplinary Lexicon with allophone students and teachers. The student's objective is to write a university long paper in French whereas the teachers discover this lexicon for its teaching then to students. For teaching these elements, we used the interfaces of the electronic corpus developed in TermITH and Scientext's project and we guided the learners - thanks to these tools - towards the discovery of the interdisciplinary phraseology. Then, we wonder about the type of elements to be taught, on the help to their memorization and finally, we present a sequence of teaching, organized with students of Master's degree.

9.1. Introduction

Dans ce chapitre nous allons aborder le Lexique Scientifique Transdisciplinaire (LST) et son enseignement/apprentissage en français langue étrangère (FLE). Le cadre est celui de l'université, le public est double : des étudiants allophones venus suivre un cursus de Master ou de Doctorat en France, des enseignants de français académique désireux d'aider ces apprenants.

L'objectif de ce travail est de développer une aide à l'écriture universitaire à partir de choix linguistiques susceptibles de participer à tout écrit scientifique dans une discipline des SHS. Rédiger met en jeu non seulement le lexique, mais tous les éléments qui contribuent à sa bonne insertion discursive. Nous allons rassembler les études qui ont été la source de développements de séquences didactiques autour des projets Scientext¹ puis TermITH² en passant par ScienQuest³. Les corpus numériques développés dans le cadre de ces projets sont des outils pour les linguistes, et c'est comme tels qu'ils ont d'abord été utilisés. Ensuite, nous souhaiterions faire bénéficier tant les enseignants que les apprenants de ces outils souvent peu connus dans le cadre de la classe de langue.

Sans nous attarder outre mesure sur ce qu'est le LST – cf. les auteurs qui ont contribué à son développement de Phal [PHA 71] à Hatier [HAT 16a] et dans cet ouvrage en passant par Granger [GRA 98], Pecman [PEC 04],

¹ URL du corpus en ligne : <http://scientext.msh-alpes.fr/scientext-site/spip.php?article9>

² URL : <http://www.atilf.fr/ressources/termith/>

³ URL : <https://corpora.aiakide.net/scientext19/>

Drouin [DRO 07] et Tutin [TUT 07] – nous aborderons la question des choix lexicaux à opérer pour son enseignement en FLE et les facteurs d'aide à l'apprentissage. Enfin, nous présenterons des exemples de séquences didactiques. Les deux parties de ce chapitre seront complémentaires, mais chronologiquement consécutives : le recensement des descriptions linguistiques, la réflexion didactique pour l'enseignement et l'apprentissage de ce lexique.

9.2. Études pour l'enseignement autour du LST

Quand, en 2007, Tutin [TUT 07] introduit le LST, à la suite des auteurs précédemment cités, il était déjà question d'enseignement. En effet, dans le projet Scientext, les données constituées devaient aboutir à plusieurs types d'applications et notamment l'enseignement et l'apprentissage en langue étrangère. L'objectif, toujours d'actualité, est de proposer des outils d'aide à l'écriture scientifique aux étudiants étrangers. Ces éléments linguistiques identifiés – la phraséologie scientifique – l'équipe de chercheurs s'est alors tournée vers la description et la réflexion didactique pour l'enseignement et l'apprentissage de ces éléments.

9.2.1. Éléments linguistiques singuliers

Pour nombre d'enseignants de langue, le lexique est difficile à enseigner et peu de méthodes de langue l'abordent de façon à les satisfaire. Pour les apprenants, plusieurs difficultés apparaissent, nous en retiendrons trois : la mémorisation, l'usage sémantique de la lexie, l'usage combinatoire de la lexie. En découvrant la fréquence d'utilisation des formes figées et semi-figées (les phrasèmes de Mel'čuk [MEL 95]) par les natifs, on mesure l'importance de ces formes dans la langue. Nous utiliserons la terminologie de la théorie Sens-Texte dans laquelle un phrasème est un syntagme non libre, c'est-à-dire contraint selon des règles syntagmatiques et paradigmatisées [MEL 11, p. 43]. Par exemple, le locuteur choisira *pull* et *bleu* pour dire *j'ai un pull bleu* (association non contrainte), le même locuteur ne choisira pas *bleue* dans *j'ai une peur bleue*. C'est la langue qui lui impose cet élément. Donc *peur bleue* est un phrasème et plus spécifiquement un phrasème sémantique : une collocation.

La langue française utilise nombre de phrasèmes très peu enseignés en FLE. Le discours universitaire fait usage de phrasèmes qui lui sont spécifiques et les étudiants allophones ne peuvent deviner leur existence, si ce n'est en leur prêtant une attention particulière dans leurs lectures universitaires. Un enseignement structuré serait utile car leur utilisation reste difficile pour les apprenants. Partant de ce constat souvent repris de Hoey [HOE 05], il est communément admis qu'un locuteur allophone est évalué positivement par les natifs (de façon inconsciente) quand il utilise la phraséologie.

« Un des traits qui différencient le langage d'un apprenant avancé de celui d'un locuteur natif, c'est que les énoncés prononcés sont souvent corrects d'un point de vue grammatical, mais inappropriés d'un point de vue collocationnel. » [EDM 13, p. 131]

Les apprenants tentent souvent de traduire les phrasèmes mais cela est le plus souvent impossible. Ainsi, dans notre enseignement, nous avons pensé à une aide à la découverte du phénomène, une prise de conscience de leur existence, puis un repérage avec une reconnaissance des éléments. Ces étapes seront reprises plus tard dans la partie didactique, mais nous pouvons dès lors indiquer qu'elles ont été testées auprès d'étudiants à l'Université d'Arras, à l'université du Luxembourg pendant un semestre en 2016, auprès de doctorants à l'Université de Grenoble, mais aussi dans le cadre de la formation continue d'enseignants à l'Université des Philippines et à celle de Kenitra au Maroc.

9.2.2. Analyse des éléments à enseigner

Tous les phrasèmes rencontrés représentent une manne fort intéressante pour l'enseignement, mais il faut les trier selon plusieurs critères énoncés ci-dessous :

- D'abord le contexte d'énonciation, en d'autres termes le discours et le genre [MAL 01]. Au sein du discours universitaire, on trouve par exemple, le genre du mémoire universitaire et celui de l'article scientifique.
- Puis, le lieu d'insertion de l'élément phraséologique scientifique : son insertion dans la structure syntagmatique du texte en fonction de son sens.
- Enfin, les synonymes et les variations morphosyntaxiques possibles.

Pour prendre en compte ces critères, nous avons utilisé trois corpus numériques : KIAP, Scientext et TermITH. Le corpus KIAP [FLØ 06] rassemble des articles scientifiques. La question du genre discursif se pose car nous souhaitons enseigner l'écrit académique à des étudiants de Master et Doctorat pour rédiger des mémoires et des thèses ; or, nous travaillons à partir d'articles scientifiques. Un corpus d'articles scientifiques rédigés par des locuteurs natifs sert à relever les régularités de fonctionnement linguistique des phrasèmes recherchés dans ce genre particulier. Des règles régissent la forme et le contenu des articles notamment en termes de problématique à résoudre et d'argumentation à développer. Ainsi, le choix des articles comme exemples pour l'écriture académique est un point de départ pour le développement ultérieur des divers éléments à aborder dans le cadre de la rédaction d'un mémoire de Master ou d'une thèse de Doctorat. L'article de recherche comme genre est un point d'entrée pour l'utilisation d'éléments phraséologiques fréquents dans ce genre et présents dans les mémoires et les thèses.

Les projets ANR Scientext puis TermITH ont donné aux chercheurs l'occasion de développer ScienQuest⁴ [FAL 13a] et DiCorpus⁵ [FAL 13b, TRA 16]. Dans ces corpus, toutes les disciplines scientifiques ne sont pas représentées, les utilisateurs ont accès à des articles majoritairement de SHS. L'outil ScienQuest sert à l'interrogation guidée du corpus Scientext tandis que DiCorpus est une aide à la rédaction scientifique (voir aussi le chapitre 7). Afin de voir les lieux d'insertion des phrasèmes transdisciplinaires, l'équipe de Scientext s'est intéressée aux fonctions rhétoriques véhiculant le positionnement et le raisonnement scientifique de l'auteur [BOC 10, GRO 10, RIN 13, TUT 10]. Une typologie des classes des noms les plus fréquents dans le cadre des objets et des activités scientifiques rencontrées [TUT 08] a été élaborée et l'équipe a mis au jour des variations significatives entre les traditions rédactionnelles de différentes disciplines. Par exemple, l'évaluation des résultats de recherche est plus prégnante en économie qu'en linguistique et surreprésentée dans les introductions de ces deux disciplines [CAV 08, TUT 10]. L'entrée par la question de l'évaluation a permis de réfléchir à deux approches pour l'enseignement : soit l'enseignant s'arrête sur ce qu'est une introduction/conclusion et son contenu, soit sur ce qu'est l'évaluation et sa place attendue dans l'écrit scientifique. Dans les deux cas, l'entrée onomasiologique est privilégiée et va dans le sens de l'aide à la mémorisation grâce à la structuration des connaissances dans un cadre contraint.

De façon plus précise, dans le cadre du positionnement, Chavez [CHA 08] a extrait des constructions [Verbe + Nom] autour de la démarcation de l'auteur. De telles listes ont contribué à une réflexion sur le positionnement scientifique des scripteurs débutants en SHS [MRO 14] auprès desquels il est important d'insister sur l'utilisation de ces constructions au sein de l'argumentation scientifique. Par exemple, Tutin [TUT 08] donne cinq verbes – *décrire, détailler, exposer, présenter, retracer* – pour la description d'un « objet construit par l'activité scientifique ». Ces verbes peuvent être quasi-synonymes en contexte et se combinent avec les noms *démarche, méthode, technique*. Ces associations lexicales sont possibles à partir de deux types de listes à combiner, mais finies : une liste de verbes et une liste de noms. Plus tard, en 2016, l'équipe de l'ANR TermITH (cf. chap. 1) a extrait le LST et a constitué les listes [HAT 16b] qui ont servi notamment aux thèses de Hatier [HAT 16a] et Yan YAN [YAN 17] (cf. chapitres 2 et 6) ; ces listes rassemblent les éléments suivants :

⁴ En ligne : <https://corpora.aiakide.net/scientext19/>

⁵ En ligne : https://corpora.aiakide.net/scientext18/services/dicorpus/?login=hoai&base=MD_base_mai2016&type=

[Traduction de l'auteur] « À la fin du processus, notre liste de mots CSL contenait 1,311 articles : 274 adjectifs, 202 adverbes, 493 noms et 342 verbes » [HAT 16b, p. 359]

À partir de ces listes, a été établie une typologie du lexique scientifique transdisciplinaire, organisée en deux niveaux de classes et de sous-classes sémantiques élaborées à partir de propriétés lexico-syntaxiques [HAT 16b, p. 360]. Dans sa thèse, Yan [YAN 17] (cf. chap. 6) a décrit et affiné le classement sémantique des verbes du LST et Tran en 2014 des Marqueurs Discursifs (TRA 14. Ces précisions permettent de cerner les usages de ce lexique en termes d'association lexicale et d'entrevoir des éléments dont la fréquence renforce l'idée de leur rôle structurant dans le genre. Par exemple, dans la sous-classe sémantique « source » (de la classe « relation »), Yan trouve les verbes *découler*, *dépendre*, *hériter*, *procéder_2*⁶, *provenir*, *ressortir_1*, *résulter*, qui entrent dans la construction syntaxique N[sujet] + V + de N [COI]. Un exemple donne un aperçu des phrases alors rencontrées :

« Deuxièmement, la **variation morphosyntaxique peut découler d'une autre source** que l'économie des moyens productifs » (Scientext, HDR – Linguistique – Développement)

Un tel matériau est important pour l'aide à la mémorisation dès l'instant où les éléments sont présentés en association lexicale (comme : *décrire* + dét. + *méthode*_{objet scientifique}) et où l'apprenant prend conscience du paradigme restreint de lexies possibles entre les éléments associés.

L'aspect morphosyntaxique est également très pertinent pour l'enseignement. Il ressort d'une étude [TUT 07] que les variations syntaxiques (cf. chap. 4) diffèrent : les collocations très figées ne subissent que très peu de variations tandis que certaines ont des préférences syntaxiques – la construction directe est majoritaire à côté de la diathèse par exemple. Ce dernier constat participe à privilégier une structure unique, dans un premier temps, auprès des apprenants. Ceci rejoint la remarque précédente à propos des paradigmes limités de substantifs à associer aux verbes du LST. Les apprenants sont souvent rassurés de comprendre qu'un nombre limité d'associations lexicales est à retenir.

Enfin, le caractère souvent peu ambigu des phrasèmes est intéressant au plan sémantique. Tutin [TUT 14] précise que les formes figées présentent souvent moins d'ambiguïté sémantique qu'un mot isolé. L'exemple de *poser une question* met au jour la monosémie de la forme semi-figée. Tutin [TUT 14] indique que les phrasèmes sont plus fréquents que les lexies isolées qui les composent et que leur sens paraît plus accessible que celui d'une lexie isolée, car il reste stable.

À la lumière de ces listes, de ces descriptions linguistiques et de l'outil qu'est le corpus numérique, nous avons conçu des séquences didactiques pour accompagner l'apprenant simultanément dans la découverte de l'écriture scientifique et de l'outil numérique et le rendre autonome. Avant de présenter une séquence, nous avons voulu comprendre l'organisation du lexique mental pour aborder la mémorisation.

9.3. Comment aider l'apprenant à retenir des phrasèmes ?

Dans l'enseignement du FLE, la présentation des formes figées est souvent subordonnée à des thématiques lexicales hors contexte [GAL 84]. Les auteurs des manuels classent souvent ces éléments selon le critère de la forme et non de l'usage. Il existe des listes d'expressions contenant du lexique des fleurs ou des animaux, ce qui n'éclaire pas l'utilisation discursive de ces éléments, mais induit un apprentissage par cœur des formes. Ainsi, serait-il pertinent de proposer un ouvrage contenant un classement onomasiologique des expressions transdisciplinaires en contextes autour du LST ; cela est désormais possible. Pourquoi l'entrée onomasiologique et contextuelle, présente depuis une trentaine d'années dans les manuels de FLE, paraît-elle si intéressante pour les apprenants ? Le contexte aide à la mémorisation comme l'expliquent Van der Linden [VAN 06] et les chercheurs en sciences cognitives. Ainsi, la forme primerait moins que le sens pour retenir des éléments.

⁶ Ce chiffre correspond pour Yan au niveau de l'entrée définitoire de la lexie. Ici il s'agit de la définition n°2 du verbe « procéder » dans le sens de la source comme « Relation mettant en lien l'origine, la raison de qqch et le résultat » [Yan 17, p. 107].

9.3.1. Le sens et la forme

L'enseignement des phrasèmes ou des mots qualifiés de « longs » (par le nombre de caractères) est souvent remis à plus tard dans la progression au fil des niveaux en langue, car considéré comme difficile dans un cadre d'apprentissage. La représentation d'une mémorisation plus difficile d'éléments longs, comme les phrasèmes, persiste alors même que l'on n'a pas de critère de reconnaissance de ce qu'est un « mot long » : à partir de quelle taille peut-on en effet dire qu'un mot est long ? Nous ne pouvons répondre à cette question. En revanche, plusieurs facteurs entrent en ligne de compte pour l'aide à leur mémorisation : la fréquence et la motivation [DÖR 94, GAR 03] sont les deux premiers, puis viennent la représentation que l'on se fait des éléments et enfin, l'utilité qu'on leur attribue. Des exemples complexes – quant aux représentations – sont présentés par Laufer [LAU 94], qui s'attache à expliquer que la difficulté associée à l'apprentissage d'un mot dépend notamment de la langue première de l'apprenant et des autres langues apprises. Ainsi, pas de mots longs ou de mots courts qui seraient plus faciles à retenir, mais des éléments de ressemblance ou de rappel d'éléments connus, ou encore de sonorité à l'oral ou d'orthographe à l'écrit (pour les langues à écriture alphabétique) : autant de facteurs qui vont faciliter ou entraver la mémorisation des éléments.

Enfin, Lewis [LEW 00] et Wray [WRA 02] ont mis en exergue la préférence qu'aurait l'apprenant à retenir des blocs de mots ou « chunks ». Mémoriser des blocs éviterait de devoir assembler des éléments isolés. Le travail d'assemblage est complexe, car, selon le niveau en langue de l'apprenant, le paradigme est soit trop restreint, soit trop grand, et donc le choix est difficile. En outre, le sens des éléments d'une forme figée n'est pas toujours prédictible. De fait, comme Hausmann [HAU 89], nous comprenons que l'apprenant ne peut pas toujours « deviner » l'association lexicale à construire, car elle est rarement prédictible, mais cet obstacle n'entrave pas forcément la compréhension :

« [La] transparence n'empêche nullement la collocation d'être imprédictible. L'apprenant étranger, tout en la comprenant (s'il comprend les mots combinés), ne saurait automatiquement la reproduire. » [HAU 89, p. 1010]

Associer deux lexies pour en former une seule ne correspond que peu à l'addition des sens de ces lexies, mais à la création d'un nouveau sens à côté de celui déjà existant. Par exemple, *gros fumeur* a deux sens, l'un compositionnel (une personne obèse qui fume) et l'autre pas (quelqu'un qui fume beaucoup sans considération quant à son poids ; dans ce cas *gros* signifie « beaucoup »). Ce sens s'élaborera en fonction de divers facteurs de l'environnement syntactico-sémantique. Binon et Verlinde [BIN 04, BIN 03] ont pris le parti du contexte et leur dictionnaire pour l'enseignement des langues étrangères⁷ tient compte des collocations et de leur entour lexicosyntactico-sémantique. Cette entrée est développée par des auteurs [CHA 07, BOU 98, MEA 96] avec des corpus introduits en classe de langue. L'idée pédagogique est l'aide à la recherche d'éléments dans leur contexte. L'aide à la mémorisation de ces blocs ne serait que bénéfique pour la production de l'apprenant qui aurait des accès sémantiques et lexicaux rapides grâce aux corpus.

Enfin, puisqu'un élément phraséologique est simultanément une association lexicale, une combinatoire syntaxique spécifique et un sens particulier, alors il faut tenir compte des trois entrées dans l'enseignement. Le sens serait prioritaire sur la forme pour l'aide à l'apprentissage, car retenir une forme relèverait de l'acquisition implicite, voire inconsciente, donc ne demande qu'un effort minime selon Ellis et Sinclair [ELL 96]. Un tel exercice paraîtrait facile sur le plan cognitif, car la répétition, et donc, la fréquence de la forme jouent un rôle essentiel dans la mémorisation. En revanche, retenir un sens relève de l'acquisition explicite et consciente, car faire des liens sémantiques entre les éléments serait une activité cérébrale complexe. Cette complexité viendrait des liens d'inférence à mettre en place entre les éléments et de la construction ou re-construction du réseau neuronal.

Ellis et Sinclair confirment des intuitions issues de la pratique des enseignants en classe de langue. Si on résume, outre la répétition et la reformulation [MAR 15], l'association de mots est à privilégier, notamment en raison de leur fréquence élevée dans la langue et du fait que le cerveau retiendrait mieux des blocs que des

⁷ ILT : Interactive Language Toolbox, <https://ilt.kuleuven.be/inlato/>

isolats. Ces éléments sont à présenter dans des contextes qui souvent sont reconnus par les locuteurs non-natifs. La difficulté se trouve alors dans le décodage et la compréhension des normes spécifiques à chaque culture qui régissent ces contextes dans les situations quotidiennes, décodage qui permettra l'accès au sens des lexies.

Comme il est davantage préconisé d'entrer par le sens que par la forme, sans renier une approche sémasiologique, l'entrée onomasiologique semble plus adaptée à l'apprentissage et les structures syntaxiques, lexicales et sémantiques apparaissent au fil des situations de communications abordées.

Les applications didactiques, développées autour du LST, associent sens et formes (présentées ci-dessous). Reste une interrogation à aborder : existe-t-il un classement mémoriel du lexique ? Si oui, comment l'apprenant y accède-t-il ?

9.3.2. Du lexique mental vers l'accès au lexique

9.3.2.1. Le lexique mental

Grossmann [GRO 11, p. 174] approuve l'idée que le sens prime sur la forme donc que le lexique s'organise dans le cerveau via les réseaux de neurones qui interagissent. Il est tout à fait probable que les connexions peuvent être en lien avec la forme. Comme l'explique Damasio [DAM 95], le cerveau n'est pas organisé en tiroirs dans lesquels il rangerait les mots, mais en un réseau qui aide à relier les éléments nécessaires pour s'exprimer avec une langue dans un contexte particulier. Ainsi le lexique mental, qui est une hypothèse qui s'avère formulable en termes neuroscientifiques, serait un réseau de liens organisés en fonction de situations de communication spécifiques et dépendrait aussi de la compétence pragmatique des locuteurs.

De telles recherches vont dans le sens de la mise en avant du savoir antérieur des apprenants – le « déjà-là » de Meirieu [MEI 12] et évoqué par tant d'autres –, mais pas seulement. Il faut se méfier des savoirs antérieurs qui pourraient faire émerger des éléments non pertinents pour les nouveaux apprentissages. Dans l'enseignement d'une langue étrangère, l'enseignant peut apporter des éléments de traduction, mais en lien avec des éléments parfois non traduisibles, tels que les phrasèmes, afin de faire prendre conscience des différences entre les langues sources et cibles. Il faudrait ajouter à la comparaison mise en place dans les programmes tels que « l'éveil aux Langues », l'aide à la mise en place d'un nouveau chemin neuronal [DAM 95]. Ceci prend du temps pour les adultes dont les circuits neuronaux sont déjà bien tracés et qui peinent à façonner de nouvelles voies dérivatives.

9.3.2.2. L'accès au lexique

Les procédures d'accès au lexique restent floues. Gaonac'h [GAO 91] recourt à la notion d'« associations possibles », car le sens sélectionnerait les lexies et pas l'inverse. Le genre va certainement aider à sélectionner le bon sens en fonction des associations lexico-sémantiques possibles dans un genre déterminé. C'est – comme le précise la Théorie Sens Texte [MEL 95] – le sens qui va permettre alors le choix lexical et pas l'inverse. L'enseignant peut alors prévoir l'introduction de nouvelles lexies en fonction du sens attendu et du genre dans lequel il fera travailler ses apprenants.

La difficulté à changer de point de vue sur des éléments et sur les concepts nouveaux en langue étrangère, impose, au plan didactique, de savoir si les allophones ont suivi ou pas des enseignements systématisés dans lesquels les phrasèmes apparaissaient. Nous pourrions peut-être en déduire qu'un enseignement de ces éléments semi-figés influencerait la vitesse à laquelle nous y accédons ensuite. La question du type d'enseignement à développer pour fixer ces éléments et donc – peut-être – aider à y accéder plus rapidement lors d'interactions langagières, se pose alors. La vitesse d'accès au lexique en langue étrangère serait-elle influencée par une certaine systématisation dans l'enseignement ou dans le repérage ? Si oui, quelle systématisation ou quel type de repérage ?

Ainsi, retenir un mot et sa structure syntagmatique, c'est mémoriser une nouvelle forme lexicale par appariement de cette forme à un sens [MIL 87]. Dans un contexte d'apprentissage, ceci va s'effectuer soit de façon implicite par inférence en contexte, soit de façon explicite. Van der Linden [VAN 06] a donné à voir les facteurs favorisant la mémorisation sémantique, lexicale et structurelle :

- Le contexte favorise l'apprentissage du lexique à condition que le sens soit facilement décodable ; un apprenant retient un mot qu'il comprend facilement (peut-être grâce aux réseaux qu'il a déjà en tête). Une lexie dont le sens reste flou pour un apprenant ne sera pas mieux retenue parce qu'elle est contextualisée. Ainsi, faut-il que le contexte donne des indices lexico-sémantiques pour permettre de décoder le sens d'une lexie et ce n'est pas toujours le cas.
- Le son et l'image favorisent la mémorisation. Ceci rappelle les théories des apprenants « visuels » ou « auditifs », théorie battue en brèche notamment par Lieury [LIE 97] qui explique la multitude de facteurs en lien avec les organes sensoriels. Ces stratégies de mémorisation que nous mettrions chacun en place seraient variables d'un individu à l'autre et varieraient chez le même individu en fonction des apprentissages (nous ferions appel à plusieurs fonctions sensori-motrices lors d'apprentissages et chaque apprentissage ferait appel à des fonctions différentes, car nous mettrions en place des mécanismes variés selon que nous apprenons à conduire ou à lire). Cependant, le son et l'image vont amplifier la variété d'entrées possibles pour les réseaux neuronaux et donc multiplier les chances de créer un chemin, relier des réseaux et donc construire un sens.

Pour conclure, il faut enseigner les phrasèmes en contexte. Il paraît important de situer le cadre élargi des éléments à enseigner : d'abord le discours (scientifique), puis le genre (article scientifique), enfin la fonction rhétorique (par exemple, le positionnement) et pour finir la phrase. En présentant de la sorte les éléments, les apprenants intégreront les codes du genre et s'interrogeront quant aux éléments à y insérer et de quelle façon. L'entrée dans le genre se ferait alors grâce à des exemples authentiques puisés dans les corpus. Ainsi, il ne s'agit pas seulement de donner des formes de la langue, mais de donner des outils que les apprenants peuvent manipuler rapidement pour entrer dans la langue. Les corpus numériques sont alors intégrés dans la classe pour guider l'apprenant vers l'autonomie dans son apprentissage de la langue. Nous allons maintenant présenter les corpus de textes numériques.

9.3.3. Un outil : le corpus numérique

Depuis Firth [FIR 57], la linguistique de corpus est liée à l'enseignement et à l'apprentissage des langues (en Anglais Langue Étrangère) et se réalise pleinement au sein du projet COBUILD⁸ développé ensuite par Sinclair [SIN 87]. Les corpus deviennent alors des objets de recherche et d'enseignement et deux grandes approches pédagogiques émergent et donnent à découvrir la langue :

« alors que l'option « corpus-based » du courant Leech tient le corpus comme un simple réservoir d'exemples destiné à tester ou vérifier des positions théoriques existantes, l'option « corpus-driven », qu'ils [Firth, Sinclair] préconisent, postule qu'aucune position théorique *a priori* ne préside aux observations sur corpus, la théorie étant induite du corpus. » [LEO 08, p. 12]

Léon explique que soit le linguiste sait ce qu'il cherche et utilise le corpus comme un moyen de vérification (« corpus-based »), soit il cherche ce que le corpus contient (« corpus-driven » ou, dans le contexte de l'enseignement, « data driven learning » (DDL) [JOH 02, TOG 01]). Dans l'approche « corpus-driven », le chercheur effectue par exemple des statistiques en s'attachant aux éléments très fréquents ou à d'autres caractéristiques qui émergent. Selon Johns, l'approche DDL rend les apprenants autonomes dans leur apprentissage et leur permet de développer des stratégies de découverte dans l'esprit de « apprendre à apprendre » (de Holec [HOL 90] ; ou « learn to learn » de Johns [JOH 91]). Toutefois, l'enseignant et l'apprenant s'aperçoivent rapidement que les deux approches sont le plus souvent complémentaires et que l'utilisation du corpus numérique nécessite de former, dans un premier temps, les deux publics à l'outil. Dans le cadre de la formation à ces outils, il serait intéressant de tenir compte des connaissances du public en matière de numérique et d'envisager les étapes suivantes pour d'abord introduire de tels outils puis, viser l'autonomie d'utilisation, telle que la préconise des chercheurs [MAJ 05, p. 99] pour les TICE⁹ :

⁸ Collins Birmingham University International Language Database.

⁹ Technologies d'Information et de Communication pour l'Enseignement.

- La première étape est l'émergence de la prise de conscience par l'utilisateur, à la fois de l'existence des corpus numériques et de sa capacité à utiliser l'outil numérique dans son ensemble (bureautique, Internet...). Cette entrée par le savoir-faire numérique de l'apprenant a été récemment reprise par Boulton et Tyne [BOU 14], précisément pour l'introduction des corpus en classe de langue.
- Dans l'étape suivante, celle de l'application, il est question de mesurer les différentes utilisations des outils numériques, selon les besoins des apprenants et les objectifs des enseignants. Cette étape tend à l'autonomie de l'apprenant, notamment en termes de prise en main des interfaces rencontrées. Cette étape entraîne donc à l'utilisation et à la compréhension de la variété de l'existant après la compréhension du fonctionnement de ces outils dans la première étape.
- La troisième étape donne la possibilité à l'apprenant d'identifier la manière d'utiliser l'outil et donne à l'enseignant celle de l'intégrer à son cours. Cette étape est l'occasion pour l'utilisateur d'explorer des chemins nouveaux pour faire évoluer ses pratiques d'enseignement ou d'apprentissage. L'objectif étant, pour l'apprenant, de comprendre l'utilité des corpus en se les appropriant ou en les combinant à d'autres outils en vue d'un apprentissage personnalisé.
- Enfin, lors de l'étape de transformation ou d'instrumentalisation [RAB 95], l'utilisateur se spécialise dans la manipulation, voire le détournement des corpus en les intégrant comme outil d'aide à son enseignement ou à son apprentissage. Cette étape est celle de la créativité, de l'appropriation et du détournement si nécessaire pour des objectifs personnels d'enseignement et/ou d'apprentissage.

L'autonomie peut alors se concrétiser tandis que l'enseignant devient guide dans le parcours d'apprentissage de l'apprenant. Une telle approche socioconstructiviste, aboutissant à l'appropriation d'un outil d'aide à la rédaction universitaire, est celle que nous avons choisi d'adopter pour les séquences développées pour l'enseignement du LST présentées ci-dessous.

9.4. Séquences didactiques pour l'enseignement du LST

L'approche didactique qui s'apparente le plus à celle décrite pour les TICE est celle développée dans le cadre du Français sur Objectif Universitaire (FOU) [MAN 11] en lien étroit avec l'ingénierie de formation [LEC 03]. Au plan didactique, nous avons pris en compte les étapes suivantes pour l'élaboration des séquences en lien avec nos objectifs :

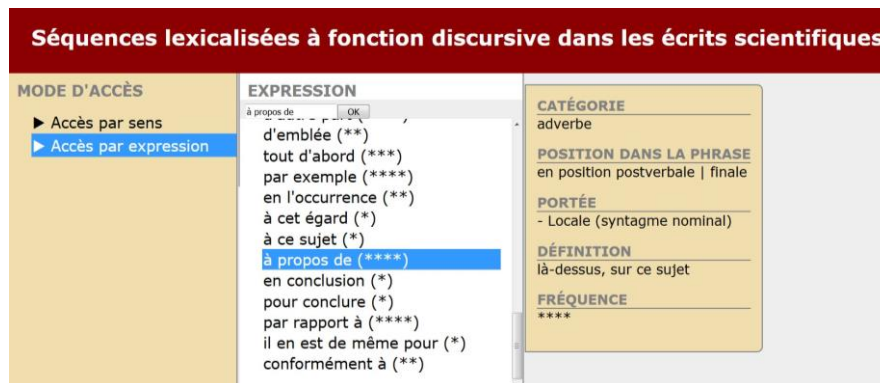
- L'analyse des besoins¹⁰ : enquête auprès de doctorants allophones du domaine universitaire de Grenoble par Bard [BAR 07].
- La collecte de données authentiques : élaboration de corpus numériques d'écrits scientifiques de locuteurs natifs (ANR Scientext puis TermITH).
- Le traitement des données : analyses linguistiques sur la démarcation et le positionnement de l'auteur dans le cadre de plusieurs mémoires de Master FLE à Grenoble, éléments qui ont notamment servi aux thèses de Mroué [MRO 14], Tran [TRA 14], Hatier [HAT 16a] et Yan [YAN 17].
- L'élaboration d'une séquence : d'abord en ligne avec le projet FULS¹¹ [CAV 14, CAV 15], puis en présence à l'aide des corpus numériques dans le cadre des thèses de doctorat de Tran [TRA 14] et Yan [YAN 17].
- L'élaboration des activités didactiques : les activités développées sont en lien avec le corpus et prennent en compte la dimension « numérique » de l'outil.

Les extraits d'activités choisis ci-dessous, intègrent les deux aspects évoqués : les corpus numériques et les phrasèmes. Le corpus numérique utilisé en ligne est l'association de Scientext et TermITH. Deux interfaces ont

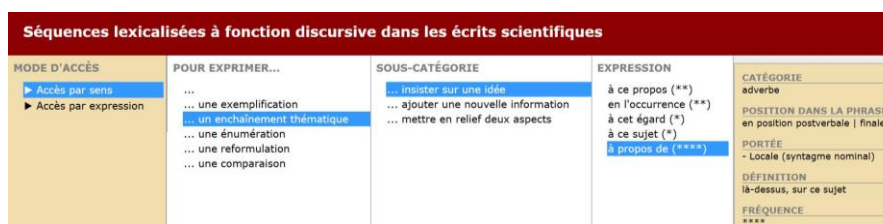
¹⁰ Ces travaux ont été réalisés en partie par des étudiants de Master FLE de Grenoble. Ces derniers ont participé au développement des projets en cours : Scientext et FULS.

¹¹ Formes et Usages des Lexiques de Spécialités : http://webtek-66.iut2.upmf-grenoble.fr/index.php?dossier_nav=689

été développées : ScienQuest et DiCorpus. ScienQuest permet une interrogation libre ou guidée (et avancée pour les plus aguerris) qui donne les résultats sous forme de concordanciers. L'utilisateur peut soit les exporter pour les trier et créer des exercices (côté enseignant), soit les utiliser en les parcourant ou les comparant pour s'aider à l'écriture (côté apprenant). DiCorpus [TRA 16] est orienté vers l'apprenant et l'aide à la rédaction scientifique. L'interface propose de choisir une fonction discursive présente dans les écrits scientifiques (par exemple « l'enchaînement thématique »), puis une sous-catégorie de cette fonction (« insister sur une idée ») et des expressions appropriées (des phrasèmes) apparaissent (dans ce cas : *à ce propos*, *en l'occurrence*, *à propos de...*).



Copie d'écran 1 : Entrée onomasiologique de *à propos de* dans DiCorpus



Copie d'écran 2 : Entrée sémasiologique de *à propos de* dans DiCorpus

Sur la copie d'écran 2 on distingue les fréquences des éléments dans le corpus ce qui donne un aperçu de leur utilisation dans le champ scientifique (ici on voit que *à propos de* est fréquent avec 4*).

La copie d'écran 1 donne à voir la position du phrasème dans la phrase (plutôt postverbale ou finale pour *à propos de*), sa définition sommaire, sa catégorie grammaticale et sa portée sémantique qui contribue à une insertion appropriée dans la phrase. Il manque des exemples, mais les développeurs sont en train d'y réfléchir soit en ajoutant un lien direct vers ScienQuest, soit en indiquant quelques phrases.

Ci-dessous est présentée une séquence qui tient compte des éléments énoncés précédemment : l'aide à la mémorisation en s'appuyant sur les acquis antérieurs des apprenants, l'évitement de la traduction sauf si deux phrasèmes sont identiques en L1 et L2, l'appropriation simultanée de l'outil et des éléments linguistiques. Nous entrons par l'utilisation de l'outil (le corpus numérique), et intégrons ensuite l'élément linguistique. L'approche est celle du FOU puisque l'analyse des besoins a été effectuée en amont comme précédemment indiqué, la collecte des données est recensée dans les corpus et leur traitement a été effectué au fil des recherches de l'équipe du Lidilem. Nous proposons les exercices prenant en compte l'appropriation du corpus par les apprenants et la découverte des phrasèmes spécifiques à l'écriture académique.

La séquence se déroule en trois temps pédagogiques :

- la découverte de l'outil corpus,
- son appropriation d'abord libre, puis avec des contraintes linguistiques (recherche de phrasèmes).
- Enfin, la réutilisation pour la présentation d'un sujet de mémoire avec un diaporama.

Cette séquence a été mise en place auprès d'étudiants de différents Masters en Sciences Humaines (langues, psychologie, droit) avec des L1 variées. Leur niveau en français variait du B2 au C1 et ils devaient rédiger un mémoire en français. Le cours s'est déroulé sur 20h en un semestre.

La séquence débute par un remue-méninges sur les connaissances et les représentations de l'écrit scientifique par les apprenants. Cet échange collectif permet d'ajuster des représentations erronées (« tous les chercheurs savent écrire dès le début de leur carrière ») et de conforter celles qui apparaissent plus appropriées (il faut s'entraîner à rédiger). Les objectifs sont présentés au premier cours dans l'ordre suivant :

- Découvrir et s'appropriier l'utilisation des corpus numériques : trouver les éléments linguistiques qui seront utiles pour la rédaction finale ;
- Préciser les connaissances linguistiques sur les écrits académiques : la structure de la langue, le vocabulaire spécifique ;
- Réfléchir à une méthodologie de rédaction : penser à utiliser l'outil pour vérifier quels sont les éléments fréquents, pour lister/s'approprier les éléments linguistiques utiles à la rédaction...

9.4.1. Découverte et appropriation de l'outil

L'entrée est donc outillée. Cela a pour effet de rassurer les étudiants parfois stressés face à des éléments linguistiques inconnus. La démarche est dite « directe » [CHA 04], car nous plaçons les étudiants devant les ordinateurs et nous ne leur fournissons pas d'éléments présélectionnés pour deux raisons : leur niveau en langue leur permet d'opérer des tris et les choix de mots dépendent trop de leurs écrits personnels ; il nous serait alors difficile de deviner les besoins individuels des étudiants-apprenants.

Le cours commence alors par ce qu'ils connaissent (le « déjà-là ») sous forme de vérification de leur « intuition linguistique ». La question posée (« Dit-on plus souvent *Je prends un risque* ou *J'attrape un risque* ? », « Dites-vous *Je mets la table* ou *Je mets le couvert* ? ») les conduit sur *Ngram Viewer*¹², interface qui permet de vérifier la fréquence de lexies isolées ou de phrasèmes dans un corpus écrit. Souvent les étudiants connaissent cet outil fort apprécié pour la vérification de son intuition en langue étrangère. Il leur est alors demandé de trouver une traduction de l'élément fréquent, appropriée à un contexte particulier de leur choix, grâce à *Linguee* (site connu des apprenants). Enfin, le dernier exercice de découverte permet de vérifier deux choses avec eux et pour eux : 1/ leur méthode de recherche en ligne, 2/ ce que le web ne nous donne pas à voir pour l'aide à l'écrit académique. La demande paraît simple a priori, et reste pourtant sans réponse pour certains étudiants : « Dans quel type de phrase trouve-t-on l'expression *en particulier* ? Comment s'en servir dans les écrits académiques ? ». Dans chacune des trois formations dispensées à des apprenants (Master et Doctorants : environ 100 personnes), la moitié tape *en particulier* dans la case de recherche en ligne, le plus souvent sans guillemet, et l'autre moitié adopte des stratégies variées, de type : recopier la question dans la case (ce qui est fort pertinent, mais non approprié à la recherche en ligne telle qu'elle est proposée actuellement), indiquer une partie de la question, souvent la fin (« expression en particulier »). Peu d'étudiants mettent des guillemets et aucun n'utilise la recherche avancée. Comme le précisent Boulton et Tyne [BOU 14], les étudiants ont besoin de formation à la recherche en ligne, formation qui contribue, nous semble-t-il, à l'appropriation de la méthodologie de la recherche universitaire.

L'appropriation de l'outil est donc progressive, allant d'outils que les apprenants connaissent bien ou partiellement, vers les corpus numériques souvent inconnus des étudiants. Ainsi, la 2^e partie de la consigne du dernier exercice invite l'enseignant à guider les apprenants vers ces corpus. En effet, la recherche sur le web permet difficilement d'extraire le type d'éléments que les corpus numériques nous offrent, sous forme de concordancier. Le premier corpus auquel ils ont été confrontés est *Le corpus français de Leipzig*¹³ parce qu'il propose des chiffres de fréquence d'utilisation immédiate. Une aide à la lecture des données est proposée sous forme d'échange collectif, puis, chacun cherche un mot de son choix et doit expliquer les données obtenues d'abord à son voisin,

¹² URL : <https://books.google.com/ngrams>

¹³ URL : http://corpora.uni-leipzig.de/fr?corpusId=fra_mixed_2012

puis au groupe [KUB 14]. Ensuite, pour présenter un concordancier, nous commençons par *Lextutor*¹⁴ qui est un petit corpus avec lequel l'étudiant peut lire et trier les extractions. Enfin, nous leur présentons le *Lexicoscope*¹⁵ dans lequel les extractions sont nombreuses et l'interface conviviale. Ils doivent alors chercher leurs premiers phrasèmes et faire leurs premiers choix pour un tri spécifique, selon un objectif linguistique lié au choix du phrasème, en accord avec l'enseignant.

Une fois que les apprenants sont à l'aise avec les concordanciers de *Lextutor*, de *Lexicoscope* et les listes fréquentielles du *Corpus Français de Leipzig*, ils sont invités à découvrir ScienQuest puis le DiCorpus.

L'entrée sur Scientext suit la chronologie de la démarche d'interrogation proposée par le site : d'abord le choix du corpus, puis celui du type de recherche et enfin la requête [TUT 13]. Les étudiants sont invités à choisir le corpus en français puis la recherche sémantique dans un premier temps. Cette entrée ne laisse pas de choix à l'apprenant, mais lui permet de visualiser rapidement ce que le site propose. Il est proposé alors de choisir une entrée sémantique (*l'opinion* par exemple) puis le type d'élément linguistique (les verbes), et un concordancier apparaît à ce propos avec des éléments pré-triés. Les étudiants sont alors invités à réfléchir aux choix des entrées sémantiques (notamment sur la place de ces éléments dans leurs écrits) pour en proposer d'autres. Ensuite, pour les guider vers la recherche libre, la réflexion s'oriente vers le détail des extractions : les variations au plan lexical (choix des verbes et des noms associés) et sémantique (divers sens apparaissent) sont à trier pour faire un choix personnel et utiliser ensuite l'élément qui paraît correspondre le mieux aux attentes du scripteur. Avant de passer à la recherche libre, DiCorpus est présenté. L'outil fournit des éléments précis et donc oblige les apprenants à préciser leur questionnement linguistique pour leur écrit, notamment préciser ce qu'il est important de mettre en avant dans leur rédaction. Ils doivent alors rédiger les questions qu'ils se posent et ensuite tenter de trouver les éléments appropriés à la rédaction scientifique pour énoncer leur pensée. Croiser ainsi les données des classes sémantiques proposées dans ScienQuest et celles autour des marqueurs discursifs de DiCorpus fait émerger à la fois des questions, mais aussi des prises de conscience de la part des apprenants. On constate leur soulagement quand ils s'aperçoivent que *par exemple* est très présent dans les écrits scientifiques et qu'ils peuvent l'utiliser sans complexe¹⁶.

Le deuxième temps du cours est l'aide au choix des lemmes à chercher et donc l'aide à affiner les recherches sur les corpus. Ceci conduit à utiliser la recherche libre. C'est à partir de ce moment que les apprenants sont entraînés vers la démarche DDL (Data Driven Learning) puisqu'ils peuvent faire une requête sur n'importe quel lemme afin de découvrir simultanément le contenu du corpus et l'utilisation d'un lemme. Nous remarquons qu'ils ne sont pas complètement dans une démarche DDL, car le temps imparti pour un cours (parfois seulement 45mn selon les pays), ne permettent pas l'extraction et le tri de n-grammes pour une découverte exhaustive des trésors cachés du corpus.

Nous constatons que les étudiants-apprenants s'approprient rapidement l'interface. Ils associent de façon fructueuse les outils présentés en parallèle surtout le CNRTL¹⁷ qui donne accès à de multiples ressources autour de la langue. L'intérêt des étudiants pour les corpus croît et ils s'interrogent alors parfois sur le contenu d'autres interfaces qu'ils connaissent. Comme le préconisent Boulton et Tyne [BOU 14], il est fréquent que les étudiants découvrent les corpus, mais il est tout aussi fréquent qu'ils les maîtrisent rapidement voire qu'ils en détournent l'utilisation pour l'adapter à leurs besoins pour leurs rédactions scientifiques. Qu'en est-il de leur appropriation des phrasèmes du LST ?

¹⁴ URL : <https://lextutor.ca/conc/fr/>

¹⁵ URL : <http://dip01.u-grenoble3.fr/~kraifo/lexicoscope/lexicoscope.php>

¹⁶ Des représentations de l'écrit et de l'oral seraient à étudier dans ces contextes universitaires où les étudiants ont souvent de bon niveau en langue.

¹⁷ Centre National de ressources Textuelles et Lexicales : <http://www.cnrtl.fr/>

9.4.2. Découverte et appropriation des phrasèmes

L'introduction des corpus dans la classe est rapidement suivie par l'introduction des phrasèmes. Sans prononcer le mot « phrasème », les étudiants prennent conscience après quelques exemples, de la fréquence de ces formes dans la langue. L'entrée dans ces éléments linguistiques passe par la traduction qu'ils ne trouvent pas littéralement¹⁸ et les oblige à envisager de nouvelles stratégies de mémorisation. Le chemin que les liens neuronaux du cerveau doivent parcourir, est alors à envisager en relation directe avec le sens du phrasème et non avec sa forme. C'est en ayant pour objectif l'aide à cette mémorisation voire à la mise en place de nouvelles stratégies d'apprentissage que nous avons développé les activités suivantes.

D'abord nous avons mis en relation la forme et le sens grâce aux entrées sémantiques de ScienQuest et DiCorpus. Le sens est premier et la forme suit immédiatement dès l'instant où l'enseignant laisse les apprenants chercher seuls dans l'interface de requêtes libres de Scientext. Ainsi, les apprenants sont conduits à envisager les mots qu'ils cherchent pour énoncer leurs réflexions scientifiques, non pas à partir d'autres mots, mais à partir de concepts dans lesquels s'insèrent ces mots : *avancer une raison* contient un verbe dit « d'opinion » et appartient, dans le classement de Scientext, à la classe sémantique « Évaluation et Opinion ». L'expérience montre que nombre d'étudiants n'a pas l'idée *a priori* d'utiliser le verbe *avancer* dans un tel cadre sémantique. Nous encourageons donc l'apprenant et l'enseignant à adopter une démarche onomasiologique, c'est-à-dire, se situer d'abord dans un concept, puis consulter les entrées lexicales possibles et s'attarder sur l'une d'entre elles si nécessaire. L'expérience montre que l'apprenant réagit d'abord de façon inverse (entrée sémasiologique) : il choisit un mot qu'il tente de traduire et de placer au bon endroit. Les outils présentés permettent l'inverse et surtout donnent à voir les mots fréquents, c'est-à-dire ceux qui ne correspondent pas systématiquement à ceux envisagés par l'apprenant. On constate alors que les apprenants écartent la traduction après quelques exemples d'impossibilité. Ainsi, entrent-ils, de fait, dans la démarche onomasiologique qui simultanément les entraîne davantage encore dans une réflexion quant à leur rédaction. L'évolution des réflexions a permis d'aboutir à la séquence suivante : une fois l'utilisation des corpus maîtrisée ainsi que la compréhension du lien sémantique entre des classes sémantiques et les mots présents dans ces classes, alors les apprenants entrent facilement dans des exercices de repérage tels que présentés ci-dessous. L'exercice s'effectue à partir de l'extrait de l'introduction d'un article pris dans ScienQuest et des questions de découpage sémantique de l'extrait sont alors proposées aux apprenants, pour une première approche, qui doivent travailler en petits groupes. Voici l'extrait et les consignes afférentes :

L'analyse linguistique que nous proposons s'inscrit en réponse à une demande de la SNCF relative à la perception du confort global en train. Les objectifs du projet sont d'identifier les propriétés sémantiques du confort en train et leurs relations de dépendance, à partir d'analyses linguistiques et cognitives. Cet article porte sur la manière dont les formes linguistiques en contexte, utilisant les ressources de la langue mises en discours, renseignent sur les structures cognitives construites à partir des perceptions sensorielles. On tente ici d'identifier, à partir de l'analyse des formes adjectivales, les représentations individuelles et partagées qui se construisent dans les discours des voyageurs, lorsqu'on les interroge sur leur expérience sensible du confort. <...> La première question que l'on se pose alors est la suivante : le confort en train est-il une catégorie cognitive de ce type ? (Scientext – article – Linguistique – Introduction)

1. D'après vous, cet extrait vient de quelle partie de l'article ? Pourquoi ?

A. résumé // B. introduction // C. théorie // D. méthodologie // E. conclusion

2. Découpez l'extrait en paragraphes et justifiez votre choix en précisant les mots et les thématiques de votre découpage : contexte, sujet, objectif, question.

¹⁸ L'exemple de « poser une question » est clair pour tous car ils ont l'anglais comme L1, L2 ou L3 et connaissent « to ask a question ».

3. Trouvez des expressions synonymes pour dire « on tente de... » à l'aide de l'outil CNTRL en ligne.

La réponse attendue pour la distinction des parties est la suivante :

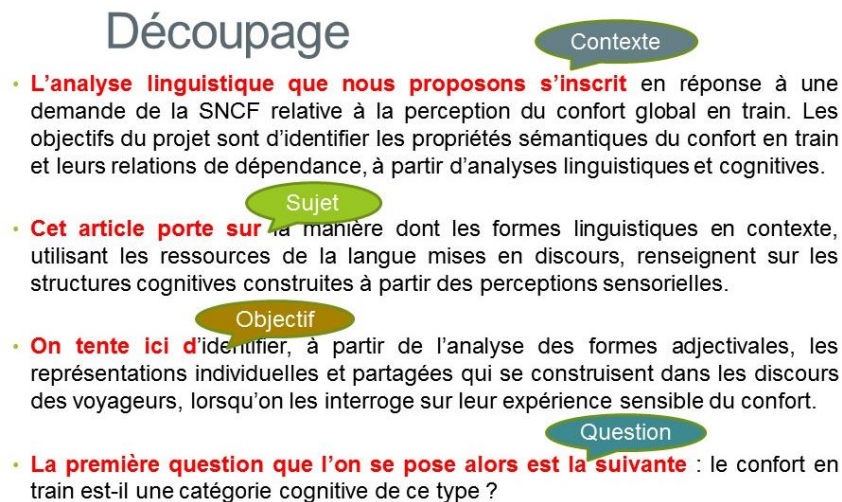


Figure 1.1 : Découpage d'un extrait de ScienQuest

Ce premier exercice permet d'introduire certaines routines verbales du LST. Les suivants sont alors plus précis, car ils sont développés pour s'attarder sur l'une des thématiques telles que « l'introduction du sujet » ou « les questions de recherche ». Après chaque repérage, les apprenants doivent recenser les nouvelles routines rencontrées et ainsi continuer au fil du temps à alimenter leur LST. Par exemple, au fil des extraits et des repérages collectifs et individuels sur corpus, les apprenants peuvent retenir que pour introduire des questions de recherche, on peut utiliser : *il semble légitime de s'interroger sur, il est possible de formuler cette interrogation comme suit, Nous nous sommes demandé si...* De même pour présenter son sujet de recherche, on peut retenir : *Cet article porte sur, Dans cet article nous nous intéresserons à, Nous montrerons que...* Remarquons qu'il leur est demandé d'être attentifs à l'insertion syntagmatique de tels éléments et notamment ici en début de phrase (la majuscule l'atteste) ou non.

Une fois ces éléments rassemblés avec des informations morphosyntaxiques (l'insertion appropriée des éléments), lexicales (le choix des mots associés) et sémantiques (la place rhétorique de l'élément), nous demandons aux étudiants de rédiger une introduction (ou une conclusion ou un développement selon l'entrée choisie) à l'aide de tous les outils auxquels ils ont désormais accès.

Nous constatons un effet bénéfique de l'utilisation des corpus pour l'apprentissage des phrasèmes. En effet, au fil des enseignements dispensés, nous constatons que les étudiants connaissent certains des phrasèmes fréquents mais présentent des difficultés à les utiliser à bon escient. En outre, leur intuition linguistique ne les aide pas à savoir pourquoi insérer certains plutôt que d'autres. L'indication de fréquence trouvée dans les corpus les rassure quant au choix sémantique tandis que les exemples les aident pour l'insertion syntagmatique. Ainsi, pouvons-nous donner quelques résultats à partir des données recueillies auprès des étudiants de Master du niveau B2-C1.

9.4.3. Utilisation pour l'aide à la rédaction personnelle

L'appropriation a été vérifiée, pour les étudiants de Master de niveau B2-C1, à l'aide d'une présentation orale et écrite de leur sujet de recherche. Les phrasèmes devaient à la fois figurer sur le diaporama¹⁹ de présentation et dans le discours oral l'accompagnant. Nous avons alors retrouvé des éléments

¹⁹ Une courte séance a eu lieu pour l'aide à l'élaboration d'un diaporama.

- du LST : *Cette recherche s'adresse à, Cet article porte sur, Nous montrerons que, il semble légitime de s'interroger sur, Nous pourrions donc avancer l'hypothèse que, nous nous sommes demandé si...*
- de DiCorpus : *En d'autres termes, en particulier, Dans un premier temps* suivi de *Dans un second temps*.

Ce qui est encourageant, c'est l'appropriation créative. Certains étudiants ont fait des associations qui sont tout à fait acceptables dans l'écrit scientifique, mais que nous n'avions pas vues en cours : *Notre première supposition est, Notre seconde supposition est*. Nous savons que certains les ont trouvées dans le corpus tandis que pour d'autres, nous n'avons pas d'indice quant à la stratégie développée. Si nous comparons avec des enseignants en formation continue, nous avons constaté qu'ils retrouvaient les extraits dans le corpus, qu'ils comparaient alors avec d'autres extraits contenant un phrasème proche sémantiquement et associaient cela à leur imagination. De fait, nous avons eu des résultats plus riches en termes linguistiques, mais il s'agissait d'enseignants de FLE.

Au fil des exercices, les apprenants étaient fortement encouragés à toujours chercher d'autres éléments dans les corpus. La pédagogie adoptée est celle de la collaboration. Tous les exercices étaient faits en groupe, parfois de petits défis intergroupes étaient proposés et chaque exercice devait finir par un récapitulatif personnel soit sous forme de liste des phrasèmes, de notes à propos d'une procédure pour trouver telle information dans tel corpus, etc. L'aide à la rédaction s'est donc sans cesse déroulée sur deux plans : un premier plan outillé avec les corpus et les sites en ligne, et un deuxième plan linguistique avec les phrasèmes transdisciplinaires. Nombre d'étudiants étaient d'abord impressionnés par l'interface de ScienQuest et se servaient plus aisément de *Lextutor* (dont l'interface n'est pourtant pas si aisée) et du *Lexicoscope*. Cependant, la majorité d'entre eux s'est finalement tournée vers ScienQuest en raison des requêtes infructueuses dans les corpus non spécialisés. Se dirigeront-ils davantage vers *Lexicoscope* dans lequel apparaissent désormais des textes scientifiques ? Il reste une étude à faire à ce propos.

9.5. Conclusion et discussion

Les séquences développées ont une entrée outillée accompagnée d'une démarche onomasiologique pour l'introduction des phrasèmes. Ceci s'est avéré fructueux dans les productions écrites et orales des étudiants.

Pour la séquence nous avons retenu une approche directe avec une entrée outillée et l'entrée onomasiologique. Dans le cadre de la thèse de Tran [TRA 14], les exercices, à partir d'extraits longs de ScienQuest, permettaient de faire des distinctions sémantiques fines entre des marqueurs discursifs. L'exemple de la distinction sémantique entre *à savoir* et *c'est-à-dire* dans des articles de recherche montre une complexité difficile à surmonter pour des apprenants non spécialistes de la langue. Ces exercices ont permis des distinctions sémantiques et la prise en compte de la ponctuation. Ce travail a été repris dans DiCorpus qui est apprécié des étudiants. Enfin, dans la thèse de Yan [YAN 17], les facteurs d'aide à la mémorisation sont réunis. Les classes sémantiques permettent une entrée didactique plus abordable pour des apprenants non-linguistes, car elles représentent des notions qu'ils reconnaissent. Le fait de devoir comparer, décrire ou examiner un élément apparaît comme une action accessible tandis que la « démarcation » reste un concept et donc un exercice difficile pour un étudiant débutant en écriture scientifique.

Ces différents essais ont abouti à la séquence présentée en partie 9.4 et ont permis d'ouvrir aux éléments des classes décrites dans TermITH. L'entrée outillée guide les apprenants qui voient alors la langue comme un réel élément à manipuler. Ensuite, leur objectif rédactionnel les oblige à voir la langue dans toute sa complexité sémantique. Pour l'enseignant, l'utilisation des corpus devient l'objectif premier des séquences mais aussi un tremplin vers l'introduction d'éléments complexes de la langue pour ensuite entraîner les apprenants à s'approprier la méthode d'interrogation du corpus et les phrasèmes du LST. L'apprenant comprend rapidement l'interface d'interrogation des corpus pour effectuer des requêtes. L'autonomie d'interrogation du corpus ScienQuest est facilitée par les entrées de la recherche sémantique immédiatement compréhensible par les apprenants.

Pour reprendre les questions soulevées au fil de l'article, nous avons constaté que les apprenants n'avaient plus systématiquement recours à la traduction dès l'instant où l'outil les aidait à voir les éléments et qu'ils avaient compris que la traduction littérale n'était que rarement possible. Notons que pour l'enseignement du lexique, même

si les neurosciences commencent à nous donner des pistes sur le fonctionnement de la mémoire, la question de « l'aide à la compréhension des discours » reste floue (qu'est-ce que comprendre, au fond ?). La question de « comment apprendre à apprendre » trouve une réponse ici en abordant le contenu à acquérir par l'outillage. Nous espérons guider l'apprenant vers l'autonomie de l'utilisation des corpus et des éléments linguistiques à utiliser pour rédiger. En d'autres termes, une fois que l'apprenant sait où chercher l'information utile à sa rédaction personnelle, une fois qu'il a pris conscience de l'existence du LST, alors son autonomie d'apprentissage et de production scientifique ne peut être qu'en bonne voie.

9.6. Bibliographie

- [BAR 07] BARD A., Constitution d'un corpus d'écrits universitaires d'étudiants étrangers, Mémoire de Master, Université Stendhal-Grenoble3, 2007.
- [BIN 04] BINON J., VERLINDE S., "L'enseignement / apprentissage du vocabulaire et la lexicographie pédagogique du français sur objectifs spécifiques (FOS) : le domaine du français des affaires", *Ela Études de linguistique appliquée*, n° 135, p. 271-83, 2004.
- [BIN 03] BINON J., VERLINDE S., "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. 31-6, 2003.
- [BOC 10] BOCH F., GROSSMANN F., RINCK F., "Le cadrage théorique dans l'article scientifique: un lieu propice à la circulation des discours", dans L. ROSIER, M.-A. PAVEAU (dir.), *Actes du colloque international Cit-dit*, Nota Bene, Laval, Québec, 2010.
- [BOU 98] BOULTON A., "L'acquisition du lexique en langue étrangère", (dir.), *Actes du 26ème Congrès de l'UPLEGESS*, Ecole des Mines, St Etienne, France, 1998.
- [BOU 14] BOULTON A., TYNES H., *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*, Didier, Paris, 2014.
- [CAV 15] CAVALLA C., "Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE", *Linx*, n° 72, p. 95-110, 2015.
- [CAV 14] CAVALLA C., "Collocations transdisciplinaires : réflexion pour l'enseignement", dans M. I. GONZALEZ-REY (dir.), *Outils et méthode d'apprentissage en phraséodidactique*, E.M.E., Fernelmont, Belgique, 2014.
- [CAV 08] CAVALLA C., TUTIN A., "Etude des collocations évaluatives dans les écrits scientifiques", *Europhras 2008*, Helsinki, Finlande, 2008.
- [CHA 07] CHAMBERS A., "Popularising corpus consultation by language learners and teachers", dans E. HIDALGO TENORIO, L. QUEREDA RODRÍGUEZ-NAVARRO, J. SANTANA (dir.), *Corpora in the Foreign Language Classroom*, Rodopi, Amsterdam, New York, 2007.
- [CHA 04] CHAMBERS A., O'SULLIVAN I., "Corpus consultation and advanced learners' writing skills in French", *ReCALL*, n° 16-1, p. 158-72, 2004.
- [CHA 08] CHAVEZ I., *La démarcation dans les écrits scientifiques - Les collocations transdisciplinaires comme aide à l'écrit universitaire auprès des étudiants étrangers*, Université Stendhal-Grenoble3, Grenoble, 2008.
- [DAM 95] DAMASIO A., *L'erreur de Descartes - La raison des émotions*, Odile Jacob, Paris, ed. 2012, 1995.
- [DÖR 94] DÖRNYEI Z. N., "Motivation and Motivating in the Foreign Language Classroom", *The Modern Language Journal*, n° 78, p. 273-84, 1994.
- [DRO 07] DROUIN P., TUTIN A., REID-TRIANTAFYLLOS S. Identification automatique des collocations transdisciplinaires des écrits scientifiques. In: European Association for Terminology, Gatineau (Québec), 2007.
- [EDM 13] EDMONDS A., ""Parfaitement sensé" - La préférence sémantique et l'enseignement des collocations en L2", dans C. GARCIA-DEBANC, C. MASSERON, C. RONVEAUX (dir.), *Enseigner le lexique*, Presses Universitaires de Namur, Namur, 2013.
- [ELL 96] ELLIS N. C., SINCLAIR S. G., "Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order", *The quarterly journal of experimental psychology*, n° 49A, p. 234-50, 1996.
- [FAL 13a] FALAISE A., "Adaptation de la plateforme corporale ScienQuest pour l'aide à la rédaction en langue seconde", *Actes de TALN 2013*, Sables d'Olonne, 2013.
- [FAL 13b] FALAISE A., "Exploitation linguistique de corpus arborés d'écrits scientifiques à l'aide du logiciel ScienQuest", dans A. TUTIN, F. GROSSMANN (dir.), *L'écrit scientifique, du lexique au discours Autour de Scientext*, PUR, Rennes, 2013.
- [FIR 57] FIRTH J. R., "A Synopsis of Linguistic Theory 1930-1955", in J. R. FIRTH (dir.), *Studies in Linguistic Analysis*, Basil Blackwell, Oxford, 1957.
- [FLØ 06] FLØTTUM K., DAHL T., KIN T., *Academic voices. Across languages and disciplines*, Philadelphie: John Benjamins, Amsterdam, 2006.
- [GAL 84] GALISSON R., *Les expressions imagées*, Clé International, Paris, 1984.
- [GAO 91] GAONACH D., *Théories d'apprentissage et acquisition d'une langue étrangère*, CREDIF - Hatier, Paris, 1991.
- [GAR 03] GARDNER R. C., MASGORET A.-M., "Attitudes, Motivation, and Second Language Learning: A Meta-Analysis of Studies Conducted by Gardner and Associates", *Language Learning*, n° 53, p. 123-63, 2003.
- [GRA 98] GRANGER S., "Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae", dans A. P. COWIE (dir.), *Phraseology - Theory, Analysis and Applications*, Oxford University Press, New York, 1998.

- [GRO 11] GROSSMANN F., "Didactique du lexique : état des lieux et nouvelles orientations", *Pratiques*, n° 149-150, p. 163-83, <http://pratiques.revues.org/1732>, 2011.
- [GRO 10] GROSSMANN F., TUTIN A., "Evidential Markers in French Scientific Writing: the Case of the French Verb 'voir'", dans E. SMIRNOVA, G. DIEWALD (dir.), *Evidentiality in European Languages Empirical Approaches to Language Typology (EALT)*, Mouton de Gruyter, Berlin, New York, 2010.
- [HAT 16a] HATIER S., Identification et analyse linguistique du lexique scientifique transdisciplinaire. Approche outillée sur un corpus d'articles de recherche en SHS, Thèse de doctorat, Université Grenoble Alpes, 2016.
- [HAT 16b] HATIER S., AUGUSTYN M., TRAN T. T. H., YAN R., TUTIN A., JACQUES M.-P., "French Cross-disciplinary Scientific Lexicon: Extraction and Linguistic Analysis", *Proceedings of the XVII EURALEX International Congress*, n°, p. 355-66, 2016.
- [HAU 89] HAUSMANN F. J., "Le dictionnaire de collocations", dans F. J. HAUSMANN, O. REICHMANN, H. WIEGAND, L. ZGUSTA (dir.), *Wörterbücher : ein internationales Handbuch zur Lexicographie Dictionaries Dictionnaires*, De Gruyter, Berlin/New-York, 1989.
- [HOE 05] HOEY M., *Lexical Priming: A New Theory of Words and Language*, Routledge, London/New York, 2005.
- [HOL 90] HOLEC H., "Qu'est-ce qu'apprendre à apprendre", *Mélanges Crapel*, p. 75-87, 1990.
- [JOH 02] JOHNS T., "Data-driven Learning: The Perpetual Challenge", dans B. KETTEMANN, G. MARKO (dir.), *Teaching and learning by doing corpus analysis*, The Edwin Mellen Press, New York, 2002.
- [JOH 91] JOHNS T., "Should you be persuaded : Two examples of data-driven learning", *Classroom concordancing English language research journal*, n° 4, p. 1-16, 1991.
- [KUB 14] KUBLER N., "Mettre en oeuvre la linguistique de corpus à l'université - Vers une compétence utile pour l'enseignement/apprentissage des langues ?", *Les Cahiers de l'Acadelle*, n° 11, p. 37-77, 2014.
- [LAU 94] LAUFER B., "Appropriation du vocabulaire : mots faciles, mots difficiles, mots impossibles", *AILE*, n° 3, p. 97-113, 1994.
- [LEC 03] LECLERCQ G., "Quelques usages de l'activité d'ingénierie de formation", *Savoirs*, n° 2003, p. 71-104, 2003.
- [LEO 08] LÉON J., "Aux sources de la 'Corpus Linguistics': Firth et la London School", *Langages*, n° 171, p. 12-33, URL: <http://www.cairn.info/revue-langages-2008-3-page-12.htm>, 2008.
- [LEW 00] LEWIS M., *Teaching collocation : Further developments in the lexical approach*, Language teaching publications LTP, Hove, 2000.
- [LIE 97] LIEURY A., *Mémoire et réussite scolaire*, Dunod, Paris, 1997.
- [MAJ 05] MAJUMDAR S., Regional Guidelines on Teacher Development for Pedagogy-technology Integration. Thailand: UNESCO Asia and Pacific Regional Bureau for Education, URL: <http://unesdoc.unesco.org/images/0014/001405/140577e.pdf>, 2005.
- [MAL 01] MALRIEU D., RASTIER F., "Genres et variations morphosyntaxiques", *Traitement Automatique des langues*, n° 2 vol.42, p. 548-77, URL: http://www.revue-texto.net/Inedits/Malrieu_Rastier/Malrieu-Rastier_Genres.html, 2001.
- [MAN 11] MANGIANTE J.-M., PARPETTE C., *Le français sur objectif universitaire*, Presses universitaires de Grenoble, Grenoble, 2011.
- [MAR 15] MARTINOT C., "La reformulation : de la construction du sens à la construction des apprentissages en langue et sur la langue", *Corela*, n° HS-18, URL: <http://corela.revues.org/4034>, 2015.
- [MEA 96] MEARAP., "The classical research in L2 vocabulary acquisition", dans G. M. ANDERMAN, M. A. ROGERS (dir.), *Words Words Words: The Translator and the Language Learner*, Multilingual Matters, Cleveland, 1996.
- [MEI 12] MEIRIEU P., *Apprendre... oui, mais comment*, ESF (1987, 1re éd.), Issy les Moulineaux, 2012.
- [MEL 95] MEL'ČUK I., CLAS A., POLGUÈRE A., *Introduction à la lexicologie explicative et combinatoire*, Duculot, Louvain la Neuve, 1995.
- [MEL 11] MEL'ČUK I., "Phrasèmes dans le dictionnaire", dans J.-C. ANSCOMBRE, S. MEJRI (dir.), *Le figement linguistique : la parole entravée*, Honoré Champion, Paris, 2011.
- [MIL 87] MILLER G. A., GILDEA P. M., "How Children Learn Words", *Scientific American*, n° 46, p. 94-9, 1987.
- [MRO 14] MROUÉ M., Éléments pour une didactique du lexique transdisciplinaire dans les écrits de recherche en français langue étrangère, Thèse de doctorat, Grenoble Alpes, 2014.
- [PEC 04] PECMAN M., Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique, Nice-Sophia Antipolis, 2004.
- [PHA 71] PHAL A., *Vocabulaire général d'orientation scientifique (V.G.O.S.) - Part du lexique commun dans l'expression scientifique*, Didier, Créteil, Paris, 1971.
- [RAB 95] RABARDEL P., *Les hommes et les technologies, approche cognitive des instruments contemporains*, Armand Colin, Paris, 1995.
- [RIN 13] RINCK F., "Comment gérer les sources dans un article de recherche? Un scénario didactique pour les apprentis-chercheurs", dans C. DESOUTTER, C. MELLET (dir.), *Le discours rapporté : approches linguistiques et perspectives didactiques*, Peter Lang, Berne, 2013.
- [SIN 87] SINCLAIR J. M., *The nature of the evidence. Looking Up: an account of the COBUILD Project in lexical computing*, Collins, London and Glasgow, 1987.
- [TOG 01] TOGNINI-BONELLI E., *Corpus Linguistics at Work*, Benjamins, Amsterdam, 2001.

- [TRA 14] TRAN T. T. H., 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, Grenoble Alpes, 2014.
- [TRA 16] TRAN T. T. H., FALAISE A., "Quelles stratégies pédagogiques pour une introduction des corpus en classe de langues ?", *Colloque Langues sur objectifs spécifiques : perspectives croisées entre linguistique et didactique*, Université Grenoble Alpes, 2016.
- [TUT 10] TUTIN A., "Evaluative adjectives in academic writing in the humanities and social sciences", dans R. LORES-SANZ, P. MUR-DUENAS, E. LAFUENTE-MILLAN (dir.), *Constructing Interpersonality: Multiple Perspectives on Written Academic Genres*, Cambridge Scholars Publishing, Cambridge, 2010.
- [TUT 08] TUTIN A., "Sémantique lexicale et corpus : l'étude du lexique transdisciplinaire des écrits scientifiques", *Lublin Studies in Modern Languages and Literature*, n° 32, p. 242-60, URL: www.ceeol.com), 2008.
- [TUT 07] TUTIN A., "Modélisation linguistique et annotation des collocations : une application au lexique transdisciplinaire des écrits scientifiques", dans S. KOEVA, D. MAUREL, M. SILBERZTEIN (dir.), *Formaliser les langues avec l'ordinateur : de Intex à NooJ*, Presses Universitaires de Franche Comté, Besançon, 2007.
- [TUT 14] TUTIN A., FALAISE A., "Expressions polylexicales dans le discours scientifique: une base de données lexicales basée sur corpus", *Europhras2014*, URL: <https://pro.aiakide.net/publis/2014Europhras-tutin-falaise-slides.pdf>, 2014.
- [TUT 13] TUTIN A., GROSSMANN F., *L'écrit scientifique : du lexique au discours. Autour de Scientext*, Presses Universitaires de Rennes, Rennes, 2013.
- [VAN 06] VAN DER LINDEN E., "Lexique mental et apprentissage des mots", *Revue française de linguistique appliquée*, n° XI(1), p. 33-44, 2006.
- [WRA 02] WRAY A., *Formulaic sequences and the lexicon*, Cambridge University Press, Cambridge, 2002.
- [YAN 17] YAN R., Étude des constructions verbales scientifiques dans une perspective didactique : utilisation des corpus dans le diagnostic des besoins langagiers du FLE à l'aide des techniques de TAL, Thèse de doctorat, Université Grenoble Alpes, 2017.