

Former à l'élaboration de corpus textuels scientifiques et techniques: quelles limites? quelles solutions? quelle méthodologie?

Training to create scientific and technical textual corpora: What limits? What solutions? What methodology?

FRANÇOISE OLMO CAZEVIEILLE
Universitat Politècnica de València
folmo@idm.upv.es

Abstract

Training terminology students to create scientific and technical text corpora requires: firstly, that the teacher has a theoretical knowledge of terminographical processes; and, secondly, the ability to design a didactic unit that facilitates the process. In this study, we will highlight the underlying limitations or difficulties involved. On the one hand, those that the teacher has to overcome when developing the course and on the other hand, those that learners face when developing their textual corpus. The contribution of this research lies in the explicit solutions it provides, bridging gaps detected and offering specific paths for those interested in developing terminological products, educators, or terminology students.

Key words

Terminography, scientific and technical textual corpus, search and selection of texts, digital literacy, project approach.

Resumen

Formar a los estudiantes de terminología en la elaboración de corpus textuales científicos y técnicos implica para el docente, por una parte, un conocimiento teórico del proceso terminográfico y por otra, la capacidad de concebir actividades que les faciliten el proceso. En este estudio, pondremos de manifiesto los límites o dificultades subyacentes a este objetivo. Por un lado, las que tendrá que solventar el docente en la preparación de sus clases y por otro, las que deberán superar los discentes para conseguir confeccionar sus corpus textuales. La contribución principal de esta investigación reside en las soluciones aportadas para dar respuestas a las dificultades expuestas. Estas ofrecen pistas metodológicas concretas a las personas que se interesan por la elaboración de obras terminográficas, profesores, mediadores o futuros terminólogos.

Palabras clave

Terminografía, corpus textual científico y técnico, búsqueda y selección de textos, alfabetización digital, aprendizaje basado en proyectos.

1. Introduction

La numérisation des corpus, pour notre part, en linguistique, a grandement facilité la recherche d'informations et la collecte des données. En effet, le développement de la linguistique de corpus par le biais des Technologies de l'Information et de la Communication (désormais les TIC) permet actuellement d'effectuer des analyses approfondies et exhaustives sur de vastes corpus qui ne pourraient pas être menées de façon manuelle. Néanmoins, très peu d'études abordent la recherche et le choix des textes du corpus. En effet, la littérature actuelle sur le sujet tend plutôt vers le traitement ou l'exploration (Poudat & Landragin, 2017) de ceux-ci. Sélectionner de façon pertinente les textes authentiques qui constitueront le corpus à dépouiller est une étape fondamentale en terminographie car d'elle dépend la qualité du produit final qu'il s'agisse d'un recueil de termes d'un domaine spécialisé ou d'un sujet plus précis.

L'étude que nous développons ici se situe en didactique de la terminologie. Elle propose une méthodologie d'élaboration de corpus textuels scientifiques et techniques afin de former les apprenants, futurs traducteurs ou terminologues, à cette pratique. Elle est issue des cours de terminologie que nous dispensons dans le Master en *Langues et Technologies* de l'Université Polytechnique de Valencia. Sous le terme corpus, de nombreuses conceptions se regroupent selon les domaines d'études et les approches spécifiques d'un chacun. Après avoir repris quelques définitions recensées dans des dictionnaires généraux et terminologiques en ligne ainsi que dans la littérature relative au sujet, nous nous proposons, dans notre partie théorique, d'explicitier tout d'abord, le type de corpus sur lesquels nous construisons nos cours ainsi que l'approche descriptive sur laquelle nous nous appuyons. Puis, nous exposerons les limites méthodologiques qu'enseignants et apprenants peuvent rencontrer respectivement: les uns, dans la confection des cours, les autres, dans l'élaboration de corpus textuels adaptés à leur projet. Ensuite, nous proposerons des solutions tant théoriques que méthodologiques à ces limites et une série d'activités réfléchies afin de les former dans la constitution des corpus scientifiques et techniques et de contribuer, par là-même, à leur alphabétisation numérique, les *e-skills*¹, et à l'éducation aux médias, le tout, dans une approche actionnelle (Conseil de l'Europe, 2000).

2. Définitions, typologie des corpus et caractéristiques des textes spécialisés

Afin de mieux appréhender le terme corpus, regardons de plus près quelques définitions de celui-ci. Selon le Trésor de la Langue Française informatisé² (TLFi), le mot "corpus"

1 Le terme générique *e-skills* englobe à la fois les compétences TIC des professionnels et des usagers. Pour une définition plus précise nous renvoyons à *E-Skills for Europe: Towards 2010 and Beyond*. Disponible sur <http://eskills.cedefop.europa.eu/download/synthesis_report-ESF-2004.pdf>

2 <<http://atilf.atilf.fr/dendien/scripts/tlfiv5/visusel.exe?11;s=367719180;r=1;nat=;sol=0>>

apparaît à la fin du 19^e siècle dans le domaine du droit, dont l’acception correspond à un “Recueil. Collection du droit romain”. Dans le domaine de la linguistique, il date du 20^e siècle et il est défini par le *Larousse encyclopédique* en 1961 comme un “ensemble d’énoncés servant de base à l’analyse linguistique”. Le TLFi en donne trois acceptions chacune reliée à un domaine spécialisé: les sciences humaines, le droit et l’informatique.

1. Recueil réunissant ou se proposant de réunir, en vue de leur étude scientifique, la totalité des documents disponibles d’un genre donné, par exemple épigraphiques, littéraires, etc.

P. anal. Répertoire scientifique.

LING. Ensemble de textes établi selon un principe de documentation exhaustive, un critère thématique ou exemplaire en vue de leur étude linguistique. Le corpus des textes parus d’un journal, d’une revue; un corpus littéraire; le corpus du vocabulaire français.

2. DR. Recueil, collection du droit romain. Le corpus juris, p. abrég., le corpus.

3. ÉLECTRON., INFORM. Ensemble de données exploitables dans une expérience d’analyse ou de recherche automatique d’informations. Perforation de corpus.

Il convient d’observer des différences importantes selon le domaine mentionné. En effet, en droit, la définition de corpus ressemble davantage à celle d’une base de données ou encore à des archives puisqu’elle équivaut à un recueil ou une collection. En revanche, en linguistique, le corpus semble être plus organisé car il est basé “sur un critère” et avoir un but, des objectifs “à des fins d’étude linguistique”.

Le Grand Dictionnaire Terminologique³ (GDT) de l’Office québécois de la langue française propose trois définitions selon trois domaines: la linguistique, la philosophie et le droit. Le droit étant, comme pour le TLFi, le premier domaine dans lequel il a été défini. En linguistique, il le situe dans le domaine de la terminologie et le décrit en ces termes (Office québécois de la langue française, 1985): “Ensemble des sources orales et écrites relatives au domaine étudié et qui sont utilisées dans un travail terminologique.” En philosophie, c’est la définition de De Landsheere (1979) qui est reprise: “Ensemble limité des éléments (énoncés) sur lesquels se base l’étude d’un phénomène” et pour le droit, il le décrit dans le droit civil avec la définition de Guillien (1974): “Le corpus (corps) constitue l’élément matériel de la possession, en désignant le pouvoir de fait exercé sur une chose”. Force est de constater que même si ce dictionnaire est en ligne et fréquemment actualisé, il ne fait aucune référence au corpus numérique. Les premiers travaux sur corpus datent, selon Cori, David et Léon, des années 1940-50 et “la mise à disposition de grands corpus et la possibilité de traitements de très gros volume de données langagières” (2008: 5) des années 1990. Ces mêmes auteurs mentionnent quant à la dénomination de “linguistique de corpus”:

3 <<http://www.granddictionnaire.com/>>.

Elle a été empruntée au courant britannique *Corpus Linguistics*, l'un des plus anciens et des plus structurés, fondé théoriquement sur la tradition firthienne de la London School. Grand producteur de manuels, il est le premier à s'être doté d'une revue: *International of Corpus Linguistics* (qui date de 1997) et d'une collection: *Studies in Corpus Linguistics* (chez Benjamins) (Cori, David & Léon, 2008: 5).

De nombreux linguistes offrent des définitions de corpus. Les plus prisées sont celles de Sinclair que nous adoptons. La première définit le corpus en ces termes: "Un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage" (dans Habert, Nazarenko et Salem, 1997: 144); la seconde, se réfère aux corpus numériques:

A computer corpus is a corpus which is encoded in a standardised and homogenous way for open-ended retrieval tasks. Its constituent pieces of language are documented as to their origins and provenance.
(Un corpus électronique est un corpus qui est encodé de façon standardisée et homogène, ceci afin de permettre de faire des requêtes ouvertes. On y mentionne l'origine et la provenance des textes qui le constituent) (Sinclair, 1996: 4).

La première définition met en exergue le regroupement raisonné du matériau langagier relevé qu'il soit textes, parties de texte ou unités de la langue; la seconde, la spécificité des corpus numériques qui manipulent des données encodées sur lesquelles sont effectuées les interrogations illimitées et l'importance de la validité et de la transparence des sources.

Maints chercheurs (Dubois (1969), Habert & al. (1997), Aston (2001), Blanche-Benveniste (2005), Burnard & McEnery (2000), etc.) proposent des définitions de corpus en reprenant dans d'autres termes et, plus ou moins détaillées et en fonction de leur champ d'étude, les idées exposées *supra*.

Comme Patrick Charaudeau, nous pensons que "le corpus n'existe pas en soi mais dépend – comme en témoigne l'article 'corpus' du *Dictionnaire d'analyse du discours* – du positionnement théorique à partir duquel on l'envisage" (2009: 37). De ce fait, il existe plusieurs types de corpus et leur classification dépend en effet du point de vue du théoricien. Ainsi, par exemple, Habert, Nazarenko et Salem différencient:

- Les corpus de textes (complets) aux corpus d'échantillons
- Les corpus de référence aux corpus spécialisés: "un corpus de référence est conçu pour fournir une information en profondeur sur une langue. Il vise à être suffisamment étendu pour représenter toutes les variétés pertinentes du langage et son vocabulaire caractéristique, de manière à pouvoir servir de base à des grammaires, des dictionnaires et d'autres usuels fiables"; "les corpus spécialisés sont limités à une situation de communication, ou à un domaine. Parmi ces corpus, on trouve les ensembles relevant de sous-langages que l'on trouve dans les domaines scientifiques et techniques".

- Les corpus comparables “constituent des sélections de textes similaires dans plus d’un langage ou dans plusieurs variétés d’un langage.”
- Les corpus parallèles “constitués d’un ou plusieurs document(s) traduits dans une ou plusieurs langues”.
- Les corpus de suivi: “donne la possibilité de ‘capter’ en continu des données dans certains secteurs. [...] Par définition, un tel corpus ne cesse de croître. Il devient alors possible d’étudier l’évolution de certains phénomènes langagiers” (Habert, Nazarenko & Salem, 1997: 144-145).

Le type de corpus sur lequel nous allons travailler ici se situe dans la catégorie des corpus spécialisés dont l’appellation mérite réflexion. L’un des critères qui font qu’un texte soit “spécialisé”, et sur lequel il convient de s’arrêter, ce sont les producteurs de ce type de textes. Selon nous et à l’instar de Cabré, les seuls émetteurs capables de fabriquer les textes spécialisés sont les spécialistes ou experts. À ce propos, cette autrice manifeste:

Defendemos que la etiqueta de ‘especialista’ debe reservarse a los que son capaces de introducir innovación en un campo de conocimiento, en una tecnología o en un área de servicios o, si no lo hacen porque no se encuentran en situación de creación sino en disposición de divulgación del conocimiento consolidado, pueden controlar el estado de conocimiento del campo en cuestión. (Nous soutenons que l’étiquette de ‘spécialiste’ doit être réservée à ceux qui sont en mesure d’introduire l’innovation dans un champ de connaissance, dans une technologie ou une aire de services ou, s’ils ne le font pas car ils ne sont pas dans une situation de création mais plutôt dans une situation de divulgation des connaissances, qui peuvent contrôler l’état des connaissances du domaine en question) (Cabré, 2002: 10).

Elle insiste sur le fait que les spécialistes sont les seuls aptes à introduire l’innovation, à créer et à contrôler les unités terminologiques d’un champ de connaissance concret.

Les textes spécialisés se différencient de ceux qui ne le sont pas car ils combinent deux caractéristiques essentielles “un critère thématique corrélé à une situation énonciative spécifique [...] qui est susceptible de favoriser le recours à une organisation discursive particulière, à des constructions syntagmatiques originales aux yeux des novices, à un vocabulaire compris par les seuls initiés” (Charaudeau & Maingueneau, 2002: 540). Effectivement, seul le critère thématique ne suffit pas pour affirmer qu’un texte est spécialisé; celui-ci doit, comme nous venons de le mentionner ci-avant, être produit dans “une situation énonciative spécifique” donc par un spécialiste, c’est-à-dire quelqu’un qui a été formé et qui travaille dans le domaine, qui, dans la vie professionnelle, se trouve dans une situation communicative spécialisée.

Le niveau de spécialisation peut varier. Il dépend généralement de l’auteur du texte et du public auquel il est adressé. Cabré (2002: 12) propose une typologie des textes spécialisés fondée sur deux critères: la thématique et le niveau de spécialisation. Le premier correspond à la variation horizontale; le second, à la variation verticale. Le résultat concernant la variation verticale est observable dans le tableau suivant que nous avons élaboré à cet effet:

Textes hautement spécialisés	Transmission/échanges de connaissances de spécialistes à spécialistes/experts
Textes moyennement spécialisés	Transmission des connaissances de spécialistes à futurs spécialistes/experts
Textes peu spécialisés	Transmission des connaissances de spécialistes à un public non spécialiste/profane

Tableau 1: Typologie des textes spécialisés selon le critère de variation verticale.

Un projet terminographique peut concerner tous les niveaux de spécialisation, dans le cas de l'élaboration d'un ouvrage lexicographique, ou être focalisé uniquement sur un seul, si le sujet est très précis, *les fromages à pâte molle* par exemple, ou si l'on travaille sur un vocabulaire de pointe.

De concert avec Cabré, nous appelons "texte spécialisé la totalité de productions discursives de caractère spécialisé" (Cabré, 2000: 14). En effet, les activités terminologiques ne se limitent pas aux domaines scientifiques et techniques, la terminologie est présente dans toutes les situations de communication spécialisée. Elle varie d'une spécialisation à l'autre. Les linguistes, chercheurs, enseignants ou apprenants, voulant observer et décrire les unités terminologiques devront "les chercher dans les productions, orales et écrites, des spécialistes" (*ibid.*) car un corpus de textes bien sélectionnés, c'est-à-dire produits par des spécialistes du domaine étudié, offre des exemples authentiques, valables et qui servent, notamment en terminologie, de productions attestées. Dans ce même ordre d'idées, Kocourek précise:

La langue de spécialité, comme la langue tout entière au reste, ce sont d'abord les textes parlés et écrits. C'est ce que les spécialistes disent et écrivent, entendent et lisent pour réaliser la communication de spécialité: entretiens, discours, conférences, débats, réunions; et lettres, rapports, comptes rendus, analyses, procès-verbaux, documents de travail, articles, manuels, livres, publications techniques et scientifiques diverses (1991: 25).

Par conséquent, en accord avec ces auteurs et dans le but d'aider les apprenants à se former en terminologie, le type de corpus sur lequel nous allons baser nos cours sera spécialisé pour des raisons relatives à la fiabilité des données comme nous venons de le voir mais aussi parce qu'ils correspondent au matériau "*in vivo*", "à l'environnement naturel des termes" (Cabré & Estopà, 2002: 6), dans lesquels notre approche descriptive de la terminologie s'applique. Mais, qu'entendons-nous par approche descriptive?

L'approche descriptive provient de la Théorie Communicative de la Terminologie (TCT) de Cabré (2009) qui qualifie, en asymétrie à la théorie classique ou Théorie Générale de la terminologie⁴ (TGT) de Wüster (1979), les termes comme étant "polyédriques". Cet

4 Wüster, considéré comme l'un des pères fondateurs de la terminologie, préconisait la normalisation conceptuelle des termes. Ils devaient être univoques de façon à faciliter la communication internationale. La démarche était

adjectif lui sert à expliquer qu'ils présentent divers aspects (cognitifs, linguistiques et communicatifs) et qu'ils peuvent être abordés sous plusieurs prismes –ou disciplines– l'un d'eux étant la linguistique. Pour ce faire, elle développe l'approche descriptive des termes ou "théorie des portes" (2000) qui consiste à les étudier dans leur contexte réel, les textes spécialisés, et à s'intéresser à tous leurs types de variations, formelles et conceptuelles. Elle les traite pour analyser leur forme comme des unités lexicales de la langue générale mais les différencie parce qu'ils présentent un sens spécialisé.

Apprendre à constituer un corpus de textes spécialisés fiable nous permettra donc en aval, d'observer et d'analyser les unités terminologiques et de connaissance spécialisée et de relever les contextes d'usage, les définitions, les variantes terminologiques, les collocations, etc. afin de systématiser la terminologie, dans le sens de relever de façon rigoureuse le vocabulaire rattaché à un domaine d'activités, à une discipline, à une technique, à un projet, etc. soit dans une langue, soit dans deux ou plusieurs langues. Néanmoins, quelle méthodologie dispenser pour former à l'élaboration de corpus spécialisé numérisé? Avant de répondre à cette question, il convient de nous interroger sur les limites auxquelles enseignants, d'abord, et apprenants, ensuite, vont se confronter.

3. Limites méthodologiques

Le défi d'un enseignant de terminologie au niveau de master réside dans l'élaboration d'une méthodologie qui au-delà de l'initiation à la terminologie et à la terminographie contribue à la formation en littératie numérique et à répondre aux pédagogies actuelles, c'est-à-dire, suive une approche communicative et actionnelle et si possible, les fasse travailler en collaboration ou en coopération afin de développer, par ce biais, les compétences transversales essentielles pour leur avenir professionnel. Pour ce faire, le didacticien devra surmonter quelques difficultés d'ordre organisationnel et méthodologique. Certaines générales puisque récurrentes dans toutes les sphères de l'enseignement (primaire, secondaire, supérieur et professionnel), d'autres plus ciblées au type de cours dispensés. Les premières, de type général, concernent:

–le facteur temporel qui constitue, pour tous les enseignants, la première limite à surmonter puisque la programmation des cours doit forcément s'y adapter et constitue, dans certains cas, une prouesse comme, par exemple, en FOS (Français sur Objectif Spécifique) ou en FOU (Français sur Objectif Universitaire) (Mangiante & Parpette, 2004 et 2011 respectivement);

–l'adaptation des stratégies mises en œuvre pour répondre aux exigences institutionnelles et sociétales en ce qui concerne d'une part, le choix des supports pour dispenser les cours (quels manuels, quels documents authentiques utiliser); les moyens

onomasiologique c'est-à-dire, elle consistait "à postuler d'abord l'existence du concept et à découvrir, par la suite, les formes linguistiques servant à l'étiqueter" (L'Homme, 2005: 1115).

employés, l'intégration des TIC et dans notre cas spécifique, d'un logiciel pour la constitution du corpus et le traitement des données langagières;

– et d'autre part, les dernières tendances en didactique recommandées par le Conseil de l'Europe, notamment, de se situer dans:

Une perspective actionnelle en considérant avant tout l'utilisateur et l'apprenant d'une langue comme des acteurs sociaux ayant à accomplir des tâches (qui ne sont pas seulement langagières) dans des circonstances et un environnement donné, à l'intérieur d'un domaine d'action particulier" (Conseil de l'Europe, 2000: 15).

Ceci se traduit, dans nos cours, par l'adoption d'une approche par tâches ou par projet de façon à faire acquérir aux étudiants plus d'indépendance dans la réflexion et dans l'action mais aussi à développer les compétences transversales, en particulier, la littératie numérique⁵ pour les préparer à leur futur professionnel et à un apprentissage permanent tout au long de leur vie;

– une programmation basée non seulement sur les connaissances à acquérir, mais aussi tenant compte des besoins et des intérêts de la classe et de chaque individu.

Les secondes, spécifiques à la matière dispensée, ici la terminologie, se réfèrent:

–Au manque de manuel de terminologie comme support des cours: bien qu'il en existe quelques-uns comme ceux d'Aguilar ou de Monterde datant de 2001 et de 2002 respectivement ou le précis de terminologie de Pavel et Nolet (2001), élaboré par le bureau de la traduction du gouvernement du Canada, en ligne et existant en sept langues dont l'anglais, le français⁶ et l'espagnol. Ce dernier, véritable chef d'œuvre de la terminologie, offre dans sa version en ligne de 2005 une partie théorique très complète avec des hyperliens qui rendent le contenu hyperactif et des exercices avec corrigés facilement intégrables aux cours de terminologie. Néanmoins, tous ces manuels, sur papier ou numérique, n'intègrent ni pratiques basées sur le numérique qui impliquent recherche d'informations et sélection des données, ni exercices terminologiques intégrant une approche actionnelle. Seul le didacticiel fait référence à la possibilité d'utiliser des moteurs de recherche dont Google pour sélectionner les informations pertinentes relatives aux concepts et à ses dénominations tout en précisant que ces informations peuvent être utilisées pour attester les termes.

–Au profil hétérogène des étudiants du master. Les étudiants que nous accueillons

5 Bien que "littératie" soit un anglicisme puisqu'il est la traduction de "*digital literacy*" nous l'utilisons car il implique une vision plus englobante des compétences numériques; selon l'OCDE (2000), "elle consiste en l'aptitude à comprendre et à utiliser le numérique dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses compétences et capacités". En outre, ce terme a été adopté par le Conseil national du numérique dans son rapport *Citoyens d'une société numérique. Accès, littératie, médiations, pouvoir d'agir: pour une nouvelle politique d'inclusion*. La Documentation française, Collection des rapports officiels, Paris: novembre 2013.

6 <http://www.infotheque.info/cache/8146/www.termiumplus.gc.ca/didacticiel_tutorial/francais/lecon1/indexe_f.html>.

dans notre master proviennent de filières diverses: ils ont suivi des formations pour devenir professeurs d'espagnol, traducteurs ou informaticiens. Les nationalités présentes sont aussi très diversifiées, les représentations les plus importantes étant les Asiatiques en premier lieu et en moindre mesure, les Latinoaméricains.

–Au manque de revues scientifiques pour les langues autres que l'anglais et l'espagnol. Nous reprendrons ce point *infra* quand nous aborderons les solutions proposées dans la méthodologie adoptée pour le français.

–Au type et à la taille du corpus de textes réalisable et exploitable dans un cours de terminologie limité par les heures de cours.

Pour les étudiants, les limites vont de pair avec les difficultés à surmonter lors des différentes étapes du processus terminologique. Dans notre cas, pour cette étude, elles sont sous-jacentes à l'élaboration du corpus textuel et elles correspondent:

–Au choix du domaine et au nombre de langues: un domaine? plusieurs? une langue? plusieurs?

–À la délimitation du domaine ou du sous-domaine et à l'établissement de l'arbre du domaine ou du sous-domaine choisi. L'établissement de l'arbre du domaine se réalise "dans le but d'obtenir une vue d'ensemble du réseau notionnel à traiter et afin de pouvoir situer plus exactement dans son environnement notionnel l'aspect particulier sous étude" (Rondeau, 1981: 72). Ceci implique pour l'apprenant la consultation de la documentation existante (classement, ouvrages, etc.), souvent abondante, afin d'acquérir des connaissances sur les contenus et la structuration du domaine.

–À la collecte de la documentation terminologique qui doit être rapide et fiable pour rassembler rapidement un corpus de textes du domaine valables pour repérer les unités terminologiques du domaine étudié et élaborer un premier fichier de termes.

–Au choix des textes à retenir: quelle taille doit avoir un corpus? comment évaluer la qualité des textes? Concernant la qualité des sources, Rondeau affirme que:

Il n'existe aucun ensemble structuré de critères permettant d'évaluer la qualité linguistique de la documentation scientifique et technique, de sorte que traducteurs, terminologues, rédacteurs techniques et professeurs de LSP doivent établir individuellement et de façon empirique les éléments sur lesquels ils fondent leurs jugements (Rondeau, 1981: 52).

Comment s'y prendre pour surmonter toutes ces limites? C'est ce que nous allons développer dans la partie suivante, les solutions méthodologiques, que nous avons divisée en deux parties: l'une, plus théorique liée au contexte institutionnel dans lequel nous nous situons et l'autre, pratique, consacrée à l'explicitation de notre méthodologie et à son développement au travers d'activités réalisables en cours.

4. Solutions méthodologiques liées au contexte institutionnel

L'enseignement de la terminologie, comme nous l'avons mentionné dans l'introduction, fait partie du cursus du Master *Langues et Technologie* de l'Université Polytechnique de Valencia (UPV). Cette unité d'enseignement (UE) de 4 crédits européens (ECTS) est dispensée chaque deux ans dans le parcours de spécialisation *Terminologie et Traduction assistée par Ordinateur* (TAO). L'UPV, comme son nom l'indique, regroupe plusieurs écoles supérieures d'ingénierie, exactement, dix. Par conséquent, les unités d'enseignement des différents cursus se situent dans les domaines scientifiques et techniques (ST). Afin d'ouvrir notre UE au public de notre université, nous focalisons l'application de la terminologie aux champs des sciences et des technologies d'où son titre *La terminologie et son application au texte technique* ainsi que le genre des textes sur lesquels nous travaillons.

En accord avec Bédard (1986), quand nous faisons référence aux domaines de spécialité nous englobons tous les champs de connaissance dans lesquels la traduction doit faire appel à des connaissances liées à une situation spécialisée, c'est-à-dire des connaissances qui généralement n'apparaissent pas de façon quotidienne dans la langue générale. Cet auteur, adopte une vision simplificatrice des domaines de spécialités pour le traducteur technique afin qu'il puisse situer les connaissances acquises ou à acquérir. Il différencie à l'intérieur de ces domaines, les réalités humaines (le droit, la finance, l'administration, les sciences humaines, les arts, les sports, etc.) des non humaines (sciences exactes et techniques). Au regard de cette proposition, nous pouvons situer notre UE dans le champ des réalités non humaines où sont rassemblées les sciences exactes et les techniques.

Comme nous l'avons vu *supra*, parmi les limites que doivent surmonter les apprenants de terminologie se trouve l'établissement de l'arbre du domaine. Cette étape est fondamentale car elle leur permettra d'une part, d'acquérir des connaissances du domaine et des sous-domaines qui le composent et d'autre part, elle leur servira à situer les termes dans leur aire spécifique d'utilisation. Pour ce faire, ils doivent s'initier au domaine choisi pour leur recherche ce qui implique la consultation d'ouvrages de tout genre (manuels d'initiation au domaine, encyclopédies thématiques, etc.) et quelquefois la consultation d'experts, chercheurs ou enseignants de la discipline en question. Dans un cours de terminologie, le temps nous est compté et les recherches ne peuvent s'effectuer comme il devrait, c'est-à-dire tel que le fait le terminologue professionnel. Pour aider les apprenants à arriver à leur fin, c'est-à-dire, à connaître le champ à étudier, sans avoir besoin d'entreprendre une recherche approfondie sur des documents, nous avons recours aux classifications internationales existantes pour les domaines de la science et de la technologie, en particulier, à la classification proposée par l'Unesco⁷. Dans cette classification, les catégories sont divisées en trois niveaux hiérarchiques: les champs, codifiés par deux chiffres, se rapportent aux sections plus générales composées

7 <<http://skos.um.es/unesco6/?l=fr>>.

de plusieurs disciplines; les disciplines, codées par quatre chiffres, supposent une description générale de groupes de spécialités; les sous-disciplines, de six chiffres, sont les entrées les plus spécifiques et représentent les activités réalisées au sein d'une discipline. Cette classification donne une vision rapide et complète des domaines scientifiques et techniques et nous sert, dans le choix des textes, à mieux sélectionner l'échantillon représentatif du domaine étudié, et dans la saisie des termes, à confirmer leur domaine d'appartenance.

En ce qui concerne la collecte de la documentation terminologique qui doit être rapide et fiable pour rassembler rapidement un corpus de textes du domaine valables pour repérer les unités terminologiques du domaine étudié et élaborer un premier fichier de termes, nous avons recours à des bases de données. En effet, les bases de données stockent des informations, dans notre cas des revues scientifiques et techniques, organisées par disciplines. Il y est aisé de consulter les derniers numéros de certaines d'entre elles afin de voir quels sujets occupent les travaux des chercheurs. Et à partir de cette littérature spécialisée, choisir comme nous le verrons dans la partie dédiée à la méthodologie, les termes-pivots qui serviront pour effectuer les recherches d'autres textes complémentaires sur Internet. D'autre part, les revues scientifiques constituent une source valide car les textes sont évalués avant leur publication par des pairs ce qui garantit la qualité des connaissances et des termes. Ce faisant, le problème de l'évaluation des sources, lié à la fiabilité de celles-ci et de l'auteur ainsi qu'à la qualité et à la pertinence, disparaît.

Quant à la taille du corpus, elle dépendra du projet d'un chacun, de la recherche thématique prévue. Pour concrétiser le projet, après le choix du sujet, chaque apprenant devra s'interroger afin de définir le public cible et de concrétiser le type de vocabulaire qu'il veut établir. À ce propos, Dubuc distingue deux types de vocabulaire: "Un vocabulaire de base comprend généralement 500 à 600 notions. Un vocabulaire de pointe peut n'en comprendre qu'une centaine. Si l'on veut couvrir l'ensemble d'un domaine ou d'un sous-domaine important, il faut songer à traiter 1500 notions et plus" (2002: 50). Pour sa part, L'Homme souligne:

La taille des corpus réunis à des fins terminologiques est souvent plus réduite que celle des corpus utilisés en lexicographie [...] Comme le texte spécialisé porte sur un sujet bien ciblé, il fait appel à un nombre limité de termes. Le terminographe n'a donc pas à parcourir des millions de mots pour retrouver plusieurs occurrences d'un même terme (L'Homme, 2004: 128).

Dans cette voie, nous sommes convaincus qu'effectivement, lorsqu'un sujet à traiter est très précis, ses délimitations sont plus faciles à percevoir et les termes à relever beaucoup moins nombreux.

Le travail pourra s'exécuter dans une langue, dans ce cas et dans notre contexte, l'espagnol; être bilingue, l'espagnol et la langue maternelle de l'étudiant ou l'anglais, ou même multilingue. Pour les recherches de plus d'une langue, le même travail de compilation des textes doit être mené dans chacune d'elles.

Il convient de souligner que le nombre de revues scientifiques publiant en langues autres que l'anglais décroît à grande vitesse. Pour pallier ce gros inconvénient, nous nous voyons obligés d'avoir recours, pour certains sujets et dans certaines langues, aux thèses et aux mémoires publiés dans les domaines concernés. Passons maintenant à la proposition méthodologique et aux activités de classe qui l'accompagnent.

5. Méthodologie suivie et activités didactiques proposées

Pour dispenser notre UE, nous disposons de quarante heures de cours. Les premières vingt heures sont consacrées à la partie théorique de la discipline et les restantes à la pratique terminographique. L'élaboration du corpus textuel constitue la première étape du processus terminographique. Celui-ci est suivi, plus tard, par le repérage des candidats-terms, la collecte des données qui consiste à relever toutes les informations relatives au terme (contexte, occurrences, éléments définitoires, etc.), l'analyse des données qui consiste à remplir les fiches terminologiques avec les renseignements relevés, à leur organisation selon différents aspects (alphabétique, thématique) et finalement, à leur gestion qui consiste à corriger, ajouter ou supprimer des renseignements sur les termes. Dans cette dernière étape, la consultation des experts est souvent nécessaire afin de faire valider les propositions de certaines équivalences notamment pour les "trous lexicaux" (Janssen & Van Campenhoudt, 2005: 1).

Les limites temporelles ainsi que de grandeur d'un corpus représentatif pour l'élaboration d'un vocabulaire relié à un sujet précis, nous ont poussés à proposer aux apprenants de travailler sur une thématique commune ce qui présente, du point de vue didactique, un grand avantage, celui de pouvoir travailler la pédagogie du projet ou plutôt, à l'instar de Perrenoud (1999), une démarche de projet⁸, qui, à notre avis, combine approche actionnelle (Conseil de l'Europe, 2000) centrée sur l'apprenant et approche communicative⁹ (Olmo, 2011) et qui plus est, forme en compétences transversales.

Effectivement, une démarche de projet:

- est une entreprise collective gérée par le groupe-classe (l'enseignant(e) anime, mais ne décide pas de tout);
- s'oriente vers une production concrète (au sens large: texte, journal, spectacle, exposition, maquette, carte, expérience scientifique, danse, chanson, bricolage, création artistique ou artisanale, fête, enquête, sortie, manifestation sportive, rallye, concours, jeu, etc.);
- induit un ensemble de tâches dans lesquelles tous les élèves peuvent

8 "Parler de *démarches de projet* présente l'avantage de ne pas écarter d'emblée tous ceux pour lesquels travailler par projets n'est pas une orientation globale, mais une façon *parmi d'autres* de mettre les élèves au travail" (Perrenoud, 1999: 2).

9 Enseigner la langue d'un domaine concret, et en particulier sa terminologie, pour communiquer implique l'intégration des aptitudes linguistiques (comprendre le discours oral et écrit, s'exprimer à l'oral et à l'écrit) et les capacités de communication.

- s'impliquer et jouer un rôle actif, qui peut varier en fonction de leurs moyens et intérêts;
- suscite l'apprentissage de savoirs et de savoir-faire de gestion de projet (décider, planifier, coordonner, etc.);
- favorise en même temps des apprentissages identifiables (au moins après-coup) figurant au programme d'une ou plusieurs disciplines (français, musique, éducation physique, géographie, etc.) (Perrenoud, 1999: 2).

En outre, cette approche appliquée à une langue cible offre grâce à la recherche et à l'utilisation de textes authentiques de sources fiables (chercheurs, organismes officiels) via Internet, une exposition directe à la langue étudiée.

Le projet consiste à construire un vocabulaire terminologique en coopération puisque chaque apprenant devra investiguer une parcelle du sujet choisi. Plus exactement, les apprenants choisissent un champ scientifique et technique à étudier de commun accord. Ils travaillent pour ce faire, tout d'abord, en collaboration et puis en coopération, puisque chacun d'eux va, à l'intérieur de cette thématique, choisir un sujet précis bien délimité sur lequel travailler de façon individuelle. Avec, au final, une mise en commun des recherches et un produit: le vocabulaire.

Par ailleurs, cette approche facilite l'acquisition, entre autres, de la compétence numérique, définie par la Commission européenne en ces termes:

La compétence numérique implique l'usage sûr et critique des technologies de la société de l'information au travail, dans les loisirs et dans la communication. La condition préalable est la maîtrise des TIC: l'utilisation de l'ordinateur pour obtenir, évaluer, stocker, produire, présenter et échanger des informations, et pour communiquer et participer via l'internet à des réseaux de collaboration (Commission européenne, 2006: 16).

La première tâche du projet à réaliser consiste à élaborer le corpus de textes scientifiques et techniques, objet de notre étude. Voici, sous forme de schéma, les différentes phases de notre méthodologie:

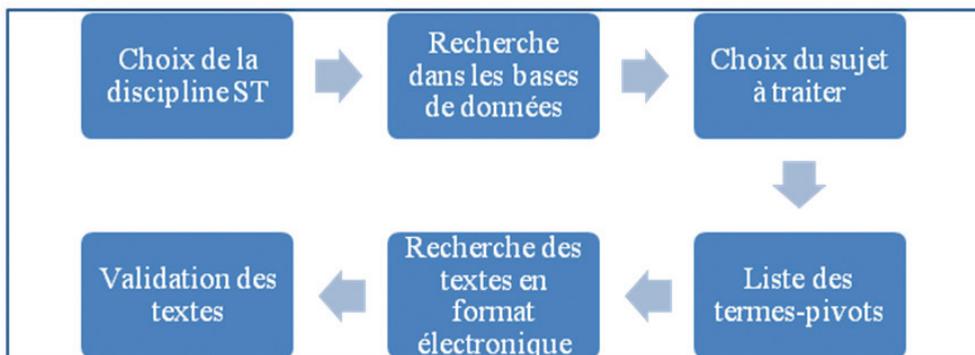


Tableau 2: Méthodologie proposée pour la mise en forme du corpus textuel.

Cette phase de conception du corpus textuel est essentielle dans le travail terminographique car elle conditionne la qualité de celui-ci. Premièrement, elle consiste, comme nous pouvons l'apprécier sur le schéma, à choisir le sujet scientifique et technique à partir de la classification de l'Unesco comme nous l'avons vu. Puis, à la recherche de revues spécialisées à partir de bases de données relatives au domaine choisi. Troisièmement, elle porte sur le choix du sujet à traiter de façon autonome. Celui-ci doit être bien dimensionné pour limiter les ressources à consulter et pour que le relevé des termes soit représentatif. En outre, avant de s'y lancer, une investigation doit être menée dans le but de savoir si le sujet n'a pas déjà été l'objet d'une recherche terminographique. Quatrièmement, on relève dans l'article de cette revue les termes-pivots et cinquièmement, on entreprend à l'aide du moteur de recherche Google la recherche des articles en format électronique qui formeront le corpus à dépouiller. Et finalement, on valide les textes d'abord, à partir du *snippet* qui correspond au petit extrait ou fragment de texte apparaissant sous l'adresse et de l'adresse électronique (URL) puis en appliquant les critères de validation reconnus par les terminologues. Voyons maintenant plus en détail les activités proposées à nos apprenants durant l'année scolaire 2014-2015. Nous avons adapté certaines d'entre elles au contexte français pour en faciliter leur compréhension.

Le projet consistait à créer individuellement, un corpus de textes spécialisés de 100.000 mots approximativement, c'est-à-dire une douzaine de textes, disponibles sur Internet. Pour ce faire, les apprenants devaient choisir tout d'abord, un domaine technique commun car le but comme nous venons de le voir, était, au final, d'assembler les recherches des pairs pour obtenir une grande partie des termes du domaine étudié. La discipline choisie fut celle relative à la Technologie des Aliments, correspondant au code 3309 de la classification de l'Unesco. Nous reprenons ci-après quelques sous-disciplines qui y sont incluses:

– Boissons alcoolisées	– Meunerie
– Aliments pour animaux	– Additifs alimentaires
– Agents antioxydants alimentaires	– Conservation des aliments
– Boulangerie	– Transformation des aliments
– Brasserie	– Hygiène alimentaire
– Conserverie	– Cryodessiccation
– Produits à base de céréales	– Lyophilisation
– Produits laitiers	– Pasteurisation
– Arôme, sapidité	– Microbiologie des aliments, etc.

Tableau 3: Exemples de sous-disciplines.

Comme nous l'apprécions sur ce tableau, le domaine de la Technologie des Aliments est très vaste, les activités variées et les projets terminographiques qui en découlent nombreux:

ceux-ci peuvent se développer autour d'un produit comme la viande, les produits laitiers, les huiles, etc. ou bien être en rapport avec leur conservation, transformation, conditionnement, distribution, etc.

La deuxième activité relative à la recherche de revue consacrée au domaine choisi afin de sélectionner un sujet actuel et de lister les termes-pivots pour qu'ils servent de base à la collecte des textes a été menée à partir de la base de données¹⁰ Dialnet¹¹ qui propose un total de 9806 revues internationales dans les domaines suivants:

<ul style="list-style-type: none"> • <i>Ciencias básicas y experimentales</i> • (543 Revistas) • <i>Geociencias. Medio ambiente</i> • (444 Revistas) • <i>Ciencias biológicas</i> • (267 Revistas) • <i>Ciencias de la salud</i> • (717 Revistas) • <i>Agricultura y alimentación</i> • (382 Revistas) • <i>Tecnologías</i> • (730 Revistas) • <i>Economía y empresa</i> • (902 Revistas) 	<ul style="list-style-type: none"> • <i>Ciencias jurídicas</i> • (1143 Revistas) • <i>Ciencias sociales</i> • (1394 Revistas) • <i>Psicología y educación</i> • (934 Revistas) • <i>Humanidades</i> • (2093 Revistas) • <i>Arte</i> • (499 Revistas) • <i>Filologías</i> • (1147 Revistas)
---	--

Tableau 4: Les revues spécialisées sur Dialnet.

Dans cette base, la Technologie des Aliments est classée avec l'Agriculture:

<p><i>Agricultura y alimentación (382 Revistas)</i></p> <p><i>Jardinería. Forestas (35)</i></p> <p><i>Agricultura y alimentación. Generalidades (69)</i></p> <p><i>Vitivinicultura (51)</i></p> <p><i>Agronomía (99)</i></p> <p><i>Producción animal. Veterinaria (92)</i></p> <p><i>Tecnología alimentaria (55)</i></p>
--

Tableau 5: Revues de la rubrique Agriculture et Alimentation (Dialnet).

10 En France, on pourrait avoir recours à des portails de revues scientifiques comme le Portail d'information scientifique des unités CNRS en Sciences et Technologies de l'Information et de l'Ingénierie sur: <<http://bibliost2i.inist.fr/>> mais pour en bénéficier, il faut que le laboratoire du chercheur y soit rattaché.

11 <https://dialnet.unirioja.es/>.

Sur les 55 revues proposées dans la rubrique Technologie alimentaire, seulement 8 revues de langue espagnole proposent les articles en libre accès, ce sont les suivantes:

- ANS. Alimentación, nutrición y salud*: <<http://digital.grupoaran.com/books/EBKANS2201/#>>
- AquaTIC: revista electrónica de acuicultura*: <<http://www.revistaaquatic.com/>>
- Ciencia y tecnología alimentaria* (Mexique) (jusqu'en 2008 uniquement): <<http://www.redalyc.org/revista.oa?id=724>>
- Cooperación agraria*: <<http://ibercede.ibercaja.es/documenta/documento/revista-cooperacion-agraria.aspx>>
- Grasas y aceites*: <<http://grasasyaceites.revistas.csic.es/index.php/grasasyaceites>>
- Invenio* (Argentine): <<https://dialnet.unirioja.es/servlet/revista?codigo=13932>>
- Tecnología agroalimentaria*: <<http://www.serida.org/publicacionesdetalle.php?id=5562>>
- Tecnología en marcha* (Costa Rica): <http://revistas.tec.ac.cr/index.php/tec_marcha/article/view/2342/2131>

À partir de ces revues, nous avons mené les recherches de thèmes et nous avons relevé les termes-pivots ou mots clés qui nous ont servis pour effectuer la collecte des textes sur Internet.

Les recherches en langue française afin d'adapter pour cette étude les exemples au contexte français ont spécialement été difficiles à mener car les revues existantes sont pour la plupart d'accès restreint. Afin d'illustrer nos propos, nous citons, pour le secteur alimentaire, quelques-unes des revues françaises que nous avons consultées et qui pour la plupart, soit ne sont plus publiées, soit sont réservées aux abonnés:

- Science des aliments* (parution suspendue depuis 2011): <<http://sda.revuesonline.com/accueil.jsp>>
- Revue process alimentaire* (réservée aux abonnés): <<http://www.processalimentaire.com/>>
- La revue des industries agro alimentaires* (réservée aux abonnés): <<http://www.revue-iaa.fr/>>
- RIA- La revue de l'industrie agroalimentaire* (réservée aux abonnés): <<http://www.ria.fr/>>
- Revue Actualité alimentaire* (réservée aux abonnés): <<http://www.actualitealimentaire.com/>>
- Revue L'alimentation* (pour un public général): <http://l-alimentation.com/archives/2016/01_janvfev16.pdf>

Parmi les revues examinées, seule la revue *Sciences des aliments et technologie alimentaire* du CIRAD (L'organisme français de recherche agronomique et de coopération internationale pour le développement durable des régions tropicales et méditerranéennes) offre le texte intégral des articles lorsqu'il est disponible: <<http://publications.cirad.fr/theme>>

php?code=Q01>. C'est donc cette revue que nous avons utilisée pour développer les activités qui suivent.



Figure 1: Extrait des publications du Cirad en Sciences des aliments et technologie alimentaire.

Le Cirad propose une sélection de documents, articles de revues, ouvrages et chapitres d'ouvrages, actes de congrès, thèses et mémoires HDR, issus des publications des agents du Cirad parues depuis 2008. Dans la branche des sciences des aliments et technologie alimentaire qui nous occupe nous trouvons 22 textes dont 10 en anglais et 12 en français. Parmi ceux écrits en français nous avons retenu la thèse de doctorat suivante: *Analyse expérimentale et modélisation de l'hétérogénéité de la qualité et de la maturité des mangues* de Nordey (2014). L'accès au résumé de celle-ci nous a permis de repérer rapidement non seulement les mots-clés mais aussi les thématiques ou sous-disciplines qui s'y entrecroisent (cf. tableau 6).

Mots-clés: mangue; mangifera indica; fruits; qualité; maturation; croissance; xylème; couleur; acidité; composition chimique; modèle de simulation; phénologie; sève; floraison; transpiration; calibre du fruit; charge.

Thématiques: Composition des produits alimentaires; Sciences des aliments et technologie alimentaire; Méthodes mathématiques et statistiques; Culture des plantes; Physiologie végétale: croissance et développement.

Tableau 6: Mots-clés et thématiques de la thèse de Nordey (2014).

À partir de ces données, l'apprenant créera son projet individuel. Celui-ci pourra tourner sur le même sujet, *la mangue*, ou bien porter sur un autre fruit puisque *la qualité, la maturation, la croissance*, etc. sont des termes applicables à tous les produits de l'horticulture. C'est la recherche avancée de *Google* qui nous permettra de localiser et de rassembler sans perte de temps d'autres textes reliés à la thématique choisie. Par conséquent, pour mettre à exécution la recherche et l'orienter rapidement à la sélection des textes complémentaires, nous accédons à *Google* <<https://www.google.fr/>>, nous ouvrons le lien "paramètres" se trouvant en bas et à droite et nous sélectionnons la "recherche avancée."

Recherche avancée

<p>Trouvez des pages avec...</p> <p>tous les mots suivants :</p> <input type="text"/>	<p>Pour effectuer cette opération dans le champ de recherche</p> <p>Saisissez les mots importants : <i>terrier trioloze</i></p>
<p>ce mot ou cette expression exact(e) :</p> <input type="text"/>	<p>Ajoutez des guillemets autour des mots exacts : <i>"terrier"</i></p>
<p>l'un des mots suivants :</p> <input type="text"/>	<p>Saisissez OR entre tous les mots à inclure : <i>miniature OR standard</i></p>
<p>aucun des mots suivants :</p> <input type="text"/>	<p>Placez un signe - (moins) devant les mots à exclure : <i>-rongeur, -"Jack Russell"</i></p>
<p>nombres compris entre :</p> <input type="text"/> et <input type="text"/>	<p>Placez deux points entre les nombres, et ajoutez une unité de mesure : <i>10..35 kilos, 300..500 USD, 2010..2011</i></p>

Affinez ensuite la recherche par...

<p>langue :</p>	<p>toutes les langues</p>	<p>Rechercher des pages dans la langue sélectionnée</p>
<p>région :</p>	<p>tous les pays/territoires</p>	<p>Rechercher des pages publiées dans une région précise</p>

Figure 2: Fragment de la recherche avancée de Google.

Le menu offre plusieurs champs à compléter qui vont nous permettre de lancer différents types de recherche. Le premier intitulé "*tous les mots suivants*" permet de trouver les pages contenant un terme monolexématique ou plusieurs. Il suffit de les séparer par un espace, par exemple: *qualité maturation croissance xylème couleur acidité*. Le second intitulé "*ce mot ou cette expression exact(e)*" permet la recherche d'unités polylexématiques. Celles-ci mises entre guillemets doivent être séparées d'un espace comme dans l'exemple suivant: "*composition chimique*" "*modèle de simulation*" "*calibre du fruit*".

En outre, les recherches peuvent être filtrées par:

- la langue;
- le lieu où doivent apparaître les termes (dans le titre, l'URL, le texte, etc.). En sélectionnant le titre, on s'assure que le document sera spécialisé et pertinent;
- le type de fichiers à récupérer; les formats word (.doc) ou pdf sont privilégiés. Cependant, si l'on retient des fichiers au format pdf il faut s'assurer qu'ils ne soient pas protégés;
- le site ou le domaine;
- les droits d'usage ou type de licence.

Voici le résultat d'une recherche dans laquelle nous avons forcé la langue française, les fichiers au format .doc et introduit les termes monolexématiques suivants: *qualité*, *maturation*, *croissance* et *xylème*.



Figure 3: Fragment du retour de la recherche avancée.

Comme nous pouvons l'apprécier le résultat de la recherche n'affiche que des fichiers .doc facilement téléchargeables et directement transférables dans le logiciel Terminus¹² que nous utilisons pour l'exploitation des données. Aucune publicité ni autre document "hors sujet" n'y est infiltré. Tous les résultats sont pertinents.

À partir de ce retour, pour guider les apprenants dans le choix des textes, certaines pistes à suivre peuvent être données afin d'éviter d'ouvrir les documents négligeables. Celles-ci ont rapport à l'adresse électronique ou URL (*Uniform Resource Locator*) du lien et au *snippet*. La première nous indique qui est l'auteur du document (organismes, chercheurs, etc.) et le type de document trouvé (article, notes de cours, actes, programmes, etc.); la seconde, le sujet traité. D'autres critères peuvent être retenus pour évaluer la qualité des textes numériques saisis. Nous les avons déjà référés dans la partie théorique. Il s'agit du domaine de spécialité, du niveau de spécialisation, de la langue de rédaction et de la date de publication. En ce qui concerne les deux derniers, il convient de préciser qu'en terminographie, les textes, en principe, doivent être en version originale et en ce qui concerne la date, les textes récents

12 *Terminus* comprend toute la chaîne du travail terminologique: recherche, constitution et exploration de corpus textuels, extraction de termes, gestion de glossaires et de projets, création et maintenance de base de données et édition de dictionnaires et publication de l'ouvrage en ligne. Il appartient à l'Institut Universitari de Lingüística Aplicada de l'Universitat Pompeu Fabra de Barcelone. Son accès est conditionné à la souscription d'une licence d'utilisation.

doivent être favorisés. Austermühl propose des “petits trucs” pour évaluer un document web que nous retranscrivons ici:

Le site est-il signalé (le nom et les coordonnées de son concepteur apparaissent-ils en clair)?

Le site est-il d'une autorité confirmée ou d'un organisme public (université, gouvernement, périodique largement diffusé)?

Le site mentionne-t-il la date de la dernière mise à jour?

Y a-t-il une bibliographie?

Le site comporte-t-il des liens vers d'autres sites gérés par des autorités renommées?

D'autres sites pointent-ils vers ce site? (2001: 130)

Pour clore cette partie méthodologique, en parfaite harmonie avec Dubuc pour ce qui est du choix des textes à dépouiller, nous dirons que “la valeur d'une recherche terminologique est directement fonction de la qualité de la documentation qui la fonde. La qualité est nettement plus importante que la quantité” (Dubuc, 2002: 51).

6. Conclusion

Même si nous avons affaire à des apprenants nés dans l'ère digitale, c'est-à-dire, connaisseurs des TIC depuis l'enfance et fervents utilisateurs des réseaux sociaux (Navarro & Olmo, 2016), ce ne sont pas des spécialistes en terminologie. Pour qu'ils le deviennent, il nous faut les guider dans la recherche des textes, les orienter dans la requête d'information linguistique valable afin de dynamiser le processus de compilation et de faciliter le travail terminologique.

À la base de tout travail terminographique, il doit exister un besoin, une nécessité qui souvent émane d'un manque de relevé de termes sur un sujet concret car, comme nous le savons bien, les ouvrages ne peuvent pas aller au rythme des découvertes et des investigations menées dans toutes les aires spécialisées. Une fois le sujet précisé, le domaine ou sous-domaine peut être délimité, le public à qui il va s'adresser ciblé, le niveau de vocabulaire fixé, la taille du corpus et l'ampleur de la nomenclature stipulés. Tous ces paramètres pris en compte contribueront à définir un projet rigoureux.

La constitution d'un corpus de textes va donc dépendre du but de l'étude ou du projet mais aussi et surtout, des données disponibles dans le domaine considéré et la langue étudiée. La sélection des données doit répondre à des critères de rassemblement réfléchis car les textes dépouillés, en aval, sont les sources des termes qui constitueront par la suite les ouvrages terminographiques.

Baser l'arbre du domaine sur la classification internationale de l'Unesco nous a permis d'avoir rapidement une idée du champ à étudier et, forcer les recherches sur Google, nous a donné la possibilité de repérer facilement et rapidement des textes spécialisés. Les

apprenants se sont formés en terminographie dans une démarche de projet et en développant des compétences transversales, en particulier, la littératie numérique.

La méthodologie exposée ici pour un cours de terminologie est tout à fait transposable aux cours de traduction ou de langue, en particulier, ceux dans lesquels les corpus textuels dans la langue cible servent à la recherche de termes, de collocations, etc. Nous pensons ici non seulement à certains cours de FOS mais aussi aux cours de FOU pour un public scientifique et technique. En effet, qu'ils soient futurs chercheurs, ingénieurs, technologues ou techniciens, ils auront tous, tôt ou tard, à manipuler les connaissances de leur domaine soit pour les échanger entre pairs ou dans leur milieu professionnel, soit pour se former, soit pour mener des recherches.

Références bibliographiques

- AGUILAR, Lourdes. 2001. *Lexicología y terminología aplicadas a la traducción. Curso práctico de introducción*. Bellaterra, Universitat Autònoma de Barcelona, Servei de Publicacions.
- ASTON, Guy (ed.). 2001. *Learning With Corpora*. Houston, Athelstan.
- AUSTERMÜHL, Frank. 2001. *Electronic Tools for Translators*. Manchester, St. Jerome Publishing.
- BÉDARD, Christian. 1986. *La traduction technique. Principes et pratique*. Montréal (Québec), Linguatex.
- BLANCHE-BENVENISTE, Claire. 2005. "Le Corpus de français parlé du GARS, Groupe Aixois de Recherches en Syntaxe" in Burr, Elisabeth (ed.), *Tradizione & Innovazione. Il parlato: teoria - corpora - linguistica dei corpora*. Atti del VI Convegno Internazionale della SILFI, Firenze, Cesati, 57-75.
- BURNARD, Lou & Tony MCENERY (eds.). 2000. *Rethinking Language Pedagogy from a Corpus Perspective*. New York, Peter Lang.
- CABRÉ, María-Teresa. 2000. "Terminologie et linguistique: la théorie des portes" in *Terminologies nouvelles. Terminologie et diversité culturelle*, n° 21, 10-15.
- CABRÉ, María-Teresa. 2002. "Textos especializados y unidades de conocimiento: metodología y tipologización" in GARCÍA PALACIOS, Jaime & María-Teresa FUENTES (eds.). *Texto, terminología y traducción*. Salamanca, Ediciones Almar, 15-36.
- CABRÉ, María-Teresa. 2009. "La Teoría Comunicativa de la Terminología, una aproximación lingüística a los términos" in *Revue française de linguistique appliquée*, 2009/2, vol. XIV, 9-15.
- CABRÉ, María-Teresa, & Rosa ESTOPÀ. 2002. "El conocimiento especializado y sus unidades de representación: diversidad cognitiva" in *Sendebarr*, n° 13, 141-153.
- CHARAUDEAU, Patrick. 2009. "Dis-moi quel est ton corpus, je te dirai quelle est ta problématique" in *Corpus*, n°8, 37-66.
- CHARAUDEAU, Patrick & Dominique MAINGUENEAU (sous la direction). 2002. *Dictionnaire d'analyse du discours*. Paris, Seuil.
- COMMISSION EUROPÉENNE. 2006. "Recommandation du Parlement européen et du Conseil du 18 décembre 2006 sur les compétences clés pour l'éducation et la formation tout au long de la vie" in JOL 394 du 30.12.2006, 10-18.
- CONSEIL DE L'EUROPE. 2000. *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Division des langues vivantes, Strasbourg, Didier.
- CORI, Marcel, Sophie DAVID, & Jacqueline LÉON. 2008. "Présentation: éléments de réflexion sur la place des corpus en linguistique" in *Langages* 2008/3, n°171, 5-11.

- DUBOIS, Jean. 1969. "Lexicologie et analyse d'énoncé" in *Cahiers de lexicologie*, n° 2, 115-126.
- DUBUC, Robert. 2002. *Manuel pratique de terminologie*. 4^e édition. Montréal (Québec), Linguatex.
- GUILLIEN, Raymond. 1974. *Lexique de termes juridiques*. 3^e édition. Paris, Dalloz.
- HABERT, Benoît, Adeline NAZARENKO & André SALEM. 1997. *Les linguistiques de corpus*. Paris, Armand Colin - Masson.
- JANSSEN, Maarten & Marc VAN CAMPENHOUDT. 2005. "Terminologie traductive et représentation des connaissances: l'usage des relations hyponymiques" in *Langages*, n°157, 63-79.
- KOCOURECK, Rostislav. 1991. *La langue française de la technique et de la science*, 2.^a ed. Wiesbaden, Oscar Brandstetter Verlag.
- LANDSHEERE (DE), Gilbert. 1979. "Dictionnaire de l'évaluation et de la recherche en éducation" in *Bulletin des bibliothèques de France (BBF)*, n°11, 559-560.
- L'HOMME, Marie-Claude. 2004. *La terminologie: principes et techniques*. Québec, les Presses de l'Université de Montréal (coll. Paramètres).
- L'HOMME, Marie-Claude. 2005. "Sur la notion de 'terme'" in *Meta*, vol. 50, n°4, 1112-1132.
- MANGIANTE, Jean-Marc & Chantal PARPETTE. 2004. *Le français sur objectif spécifique: de l'analyse des besoins à l'élaboration d'un cours*. Paris, Hachette FLE.
- MANGIANTE, Jean-Marc & Chantal PARPETTE. 2011. *Le français sur objectif universitaire*. Grenoble, PUG.
- MONTERDE REY, Ana María. 2002. *Ejercicios de introducción a la terminología para traductores e intérpretes*. Universidad de las Palmas de Gran Canaria, Servicio de Publicaciones y Producción Documental.
- NAVARRO, José Manuel & Françoise OLMO. 2016. "Autoeficacia, actividades con TIC y rendimiento académico" in BOTTI NAVARRO, Vicent & Miguel-Ángel FERNÁNDEZ PRADA (eds). *Actas del Congreso Nacional de Innovación Educativa y Docencia en Red*. València, Universitat Politècnica de València [consulté le 21/08/2016] <<http://ocs.editorial.upv.es/index.php/INRED/INRED2016/paper/viewFile/4430/2000>>
- NORDEY, Thibault. 2014. *Analyse expérimentale et modélisation de l'hétérogénéité de la qualité et de la maturité des mangues*. Thèse de doctorat. Avignon, Université d'Avignonnet des pays du Vaucluse.
- OLMO, Françoise. 2011. "Hacia una competencia comunicativa en francés científico" in MARUENDA-BATALLER, Sergio & Begoña CLAVEL-ARROITIA (eds). *Multiple Voices in Academic and Professional Discourse. Current Issues in Specialised language Research, teaching and New Technologies*. UK, Cambridge Scholars Publishing, 167-176.
- PAVEL, Silvia & Diane NOLET. 2001. *Précis de terminologie*. Canada, Ministre des Travaux publics et Services gouvernementaux.
- PERRENOUD, Philippe. 1999. "Apprendre à l'école à travers des projets: pourquoi? comment?". Faculté de psychologie et des sciences de l'éducation, Université de Genève [consulté le 23/08/2016] <https://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_1999/1999_17.html>.
- POUDAT, Céline & Frédéric LANDRAGIN. 2017. *Explorer un corpus textuel: Méthodes, pratiques, outils*. Louvain-la-Neuve, De Boeck, Champs linguistiques.
- RONDEAU, Guy. 1981. *Introduction à la terminologie*. Chicoutimi, Gaëtan Morin.
- SINCLAIR, John. 1996. *Preliminary recommendations on corpus typology*. EAGLES [consulté le 23/08/2016] <<http://www.ilc.cnr.it/EAGLES96/corpus typ/node5.html#SECTION00041000000000000000>>.
- WÜSTER, Eugene. 1979. *Einführung in die Allgemeine Terminologielehre und Terminologische Lexicographie*. Viena, Springer.