

L'Extraction Terminologique Automatique : Une Approche Centrée sur l'Apprenant

Samah Nasr

Faculté de Jeunes Filles, Université Ain Shams

samah.hassan@women.asu.edu.eg

Résumé : Cette étude présente une expérience d'extraction de terminologie à partir de corpus compilés par des apprenantes universitaires en vue d'explorer les ressources documentaires englobant des textes de spécialité dans les deux langues arabe et française et de montrer l'impact de l'intégration des outils d'extraction automatique dans le cursus universitaire d'après une expérience dans le cadre d'un module portant sur les rapports entre informatique et terminologie. Nous décrivons les résultats d'extraction de la terminologie et la méthode de représentation des données lexicales. Les difficultés liées à la sélection des candidats-termes, la modélisation de la structure actancielle de certains termes, le choix des termes associés et des collocations sont parmi les points abordés ainsi que quelques solutions envisagées pour modéliser les résultats obtenus.

Mots-clés : extraction terminologique – autonomie de l'apprenant – pédagogie par projets - terminotique

1 INTRODUCTION

Depuis le début des années 1990, des études se sont largement développées dans la formation universitaire, tout comme dans la formation scolaire en ce qui concerne les applications numériques et le changement des pratiques des enseignants et des apprenants. Que ce soit au niveau de la documentation et de la recherche d'information dans les établissements ou encore dans les médiathèques, les données digitales sont davantage utilisées ; mis à la disposition des apprenants, elles contribuent à changer la posture de l'enseignant et à individualiser l'apprentissage.

Les modalités de travail se transforment ; le cours magistral est bouleversé. Les contenus évoluent et modifient les paramètres traditionnels de la salle de cours dans son unité de lieu (le cours est donné en ligne), de temps (deux heures vs accès asynchrone aux ressources du Moodle) et d'action (la formation vs dépouillement et activités personnelles).

Ces changements remettent en question les interventions pédagogiques et exigent des réflexions sur l'agencement des contenus qui favorisent les compétences des apprenants. Notre étude présente une expérience d'extraction de terminologie à partir de corpus compilés par des apprenantes universitaires en vue d'explorer les ressources documentaires englobant des textes de spécialité dans les deux langues arabe et française et de montrer l'impact de l'intégration des outils d'extraction automatique dans le cursus universitaire d'après les projets individuels dans le cadre d'un module portant sur les rapports entre *informatique et terminologie*. Nous décrivons les résultats d'extraction de la terminologie et la méthode de représentation des données lexicales. Les difficultés liées à la pratique centrée sur les apprenantes.

2 RECHERCHE TERMINOLOGIQUE : PRINCIPALES ORIENTATIONS

Le traitement automatique des langues naturelles a largement influencé la recherche documentaire et conserve toute sa pertinence dans le domaine des études terminologiques, notamment dans les travaux de L'Homme [1] qui met en relief, dans son approche méthodologique, l'exploitation des corpus. Également, le développement des applications basées sur l'intelligence artificielle ne cesse de s'étoffer. Les recherches dans ce domaine portent sur des thématiques variées telles que la constitution des données terminologiques [2], la modélisation des relations terminologiques [3], la gestion des termes [4], l'étude des dénominations [5], la place de la documentation dans la recherche terminologique [6], l'évaluation des outils technologiques dans la formation des traducteurs [7] et enfin « l'explicitation des apports de la terminologie à la rédaction technique, à la traduction et à l'aménagement linguistique » [8].

Au niveau théorique, la terminologie est née dans un cadre cognitif sous l'influence de Wüster [9], puis Cabré [10] analyse le terme selon une approche communicative et Gaudin [11] aborde l'approche sociolinguistique.

Parmi les étapes de la recherche terminologique, nous pouvons citer la délimitation d'un domaine spécifique, la collecte de la documentation, le repérage des candidats-termes qui seront recensés dans un fichier terminologique formé de rubriques standardisées et définies selon une structure précise. L'approche lexico-sémantique occupe une place importante.

Les logiciels d'extraction ou concordanciers se sont introduits dans cette opération de repérage des données terminologiques ; il s'en suit une transformation des pratiques de dépouillement due aux exploitations de corpus numériques issus de recherches automatisées ou semi-automatisées, ce qui rend les opérations de détection plus facile. Dans cette étude, la modélisation des relations sémantiques entre termes se base sur la Théorie Sens-Texte amorcée dans le cadre d'un projet de fichier informatisé.

3 APPROCHE DIDACTIQUE DE LA TERMINOLOGIE : METHODOLOGIE ET PROGRESSION

Avec la contribution croissante des études sur corpus et la mutualisation des ressources ouvertes de données linguistiques, la structuration des données terminologiques offre de larges possibilités au traducteur-terminologue ; l'exploration des corpus est en fait une des principales étapes de dépouillement terminologique et il s'en suit un intérêt grandissant en vue de l'introduction - dans les cursus de formation en traduction- de quelques éléments de terminologie. Les apprentis-traducteurs approchent l'extraction terminologique et sont ainsi mieux préparés aux outils disponibles dans leur futur environnement professionnel.

Dans notre expérience d'enseignement, le groupe d'apprenantes s'est composé de seize étudiantes universitaires inscrites en deuxième année de traduction dans la Filière des langues appliquées de l'Université d'Alexandrie. La démarche adoptée pendant ce cours entièrement présenté en ligne exploite les outils numériques de dépouillement semi-automatique. Cette approche à distance a été conçue selon une approche centrée sur l'apprenant, d'où la nécessité de prendre en compte certains critères que nous développerons en matière du temps disponible, du référentiel de compétences et des besoins des apprenantes.

Tout d'abord, il semblait essentiel de concevoir ce cours sur quelques éléments théoriques relatifs aux bases de tout travail terminologique : la précision de l'essence de la recherche terminologie, ses approches et son objet d'étude. La collecte de données et la conception du fichier terminologique se font en fonction d'un domaine de spécialité, de la nature des notions et de la finalité de la ressource ou public cible. La compilation des données ne peut se faire qu'après une délimitation du domaine afin d'en dresser la liste de termes, leurs définitions et leur usage. La liste définitive est élaborée après consultation d'experts en cas de néologie ou de variations dénominales.

Identifier et regrouper les unités terminologiques dans les corpus spécialisés est une des compétences terminologiques qui doivent se consolider tout au long de la progression du cours. Cela implique de se questionner au préalable sur l'ensemble des situations pertinentes à retenir en vue d'une évolution progressive à partir d'une série de tâches à réaliser pour résoudre les difficultés terminologiques en simulation d'une liste de missions dans une situation professionnalisante. En premier lieu, il faut répondre aux questions suivantes : quelles sont les compétences requises et comment les intégrer de façon pratique ? Quel est le rôle de l'apprenant et sa responsabilité vis-à-vis de la préparation de son avenir en vue de son insertion rapide dans le monde du travail ? Pour répondre, commençons par une définition des compétences en matière de terminologie, ce qui exige par conséquent une précision des étapes de recherche terminologique et des nouvelles pratiques qui en résultent. La terminologie servirait à l'acquisition de nouvelles compétences numériques renvoyant d'une part à des *ressources cognitives* (maîtrise des méthodes de dépouillement, de fouille documentaire, de sélection de candidats-termes qui doivent se faire suivre, selon Perrenoud [12] à une pratique réussie de *schèmes opératoires* qui permettent, en temps réel, la mobilisation efficace des ressources cognitives. On passe du *savoir* au *savoir-faire*.

A. Référentiel des compétences terminologiques et déroulement des séances

En ce qui concerne la sélection des compétences visées par la formation, nous avons opté pour la démarche lexicosémantique décrite par L'Homme [1] et nous en avons sélectionné les points suivants tels que mentionnés dans le document européen, EMT Competence Framework [13] à savoir :

- S'initier aux principes de base dans toute recherche terminologique ;
- Découvrir les choix terminologiques et les modes de formation des termes dans de spécialité ;
- Préciser ses besoins en information et documentation ;
- Prendre en main des techniques de recherche documentaire et terminologique ;
- Appliquer des critères d'évaluation de la fiabilité des sources ;
- Maîtriser l'emploi des moteurs de recherche ;
- Enregistrer les métadonnées de ses propres documents ;
- Et enfin gérer son environnement informatique et son projet de gestion terminologique.

La réalisation de telles compétences se nourrissent au fil des activités et visent à outiller les apprenantes des différents types de requêtes et de ressources numériques disponibles ainsi que la découverte des principes qui président à la structuration et l'organisation des données. Nous avons commencé par l'examen des fichiers interrogeables, ensuite par les critères de sélection des termes-candidats, d'élaboration d'une fiche (champs, sélection et types de contextes, marques d'usage, etc.) et enfin de systématisation des descripteurs définitionnels. La définition actancielle est privilégiée vu la cible formée de traducteurs et rédacteurs spécialisés.

1) Interrogation des ressources ou bases de données en ligne

Un premier cours dresse les objectifs du module intitulé *terminologie et terminotique* en définissant le terme, la terminologie, la terminographie et la terminotique. En examinant un article de dictionnaire spécialisé, une entrée d'une encyclopédie en ligne et une définition dans une base de données terminologiques, Techno-Science.net, nous entrons dans le vif du sujet : la découverte des ressources ouvertes, gratuites et interrogeables.

La typologie des recherches, déjà évoquée à partir de remue-ménages, concernant les démarches suivies lors de la recherche d'équivalents en arabe de termes français est mise en œuvre lors de la consultation des dites-ressources. L'approche par problèmes portera sur l'absence de dictionnaire spécialisé dans un champ néologique tel que la cybersécurité. Comment trouver les équivalents et comment vérifier l'emploi d'un terme ? Les notions de recherche documentaire et de glossaires thématiques sont revisitées à partir de l'examen du Grand Dictionnaire terminologique et de recherche par terme ou par mot-clé. Dans les fiches de ce dictionnaire, nous nous trouvons face à la découverte des champs de la fiche terminologique ; des icônes informatiques étoffent la fiche : il s'agit de *marques d'usage* ainsi désignées par Corbeil [14].

Nous remarquons dans la fiche extraite du GDT [15] l'emploi de : *terme privilégié, termes utilisés dans certains contextes, ancienne appellation, notes* ; l'usage peut concerner des indications telles que *familier, désuet ou vieilli*. Le terme est suivi des éléments suivants : l'indication du domaine, du sous-domaine, de la note sur les composantes, les termes associés, le registre, la date, l'équivalent, l'illustration. Il est également possible de se déplacer vers la fiche de l'équivalent anglais. La fiche est également imprimable. D'autant plus, le travail collaboratif entre l'Office québécois de la langue française et d'autres partenaires est mis en avant en soulignant l'utilisation d'un « code de couleurs et à des pictogrammes pour renseigner l'usager sur l'acceptabilité des termes français consignés sur la fiche :

- vert : termes privilégiés
- jaune : termes utilisés dans certains contextes
- rouge : termes déconseillés »

La figure 1 présente la fiche du terme *ordinateur* et marque l'évolution de la dénomination de *calculateur à programme enregistré* annoté de l'étiquette *désuet* qui est une remarque sur l'arrêt de l'usage de ce terme. Le terme d'usage familier est indiqué également sur la fiche « *ordi* ».

ordinateur
Anglais (EN)

Domaine informatique > matériel informatique

Auteur Office québécois de la langue française, 2006

Définition
Machine programmable de traitement de l'information, commandée par des programmes stockés en mémoire, qui accepte des données structurées, les traite selon des règles définies et produit automatiquement un résultat en sortie.

Notes
Un ordinateur est constitué d'une unité centrale (pour exécuter les programmes), d'une mémoire centrale (pour stocker les données et les logiciels) et de périphériques, comme le clavier et la souris (pour... [±])

Terme privilégié

ordinateur n. m.

Termes utilisés dans certains contextes

<p>ordi n. m. familier</p> <p>calculateur à programme enregistré n. m. désuet</p>	<p>L'ancienne appellation française de l'ordinateur était <i>calculateur à programme enregistré</i>. On employait ce terme par opposition à programme câblé et à programme par cartes. Tous les ordinateurs modernes étant à programme enregistré, cette distinction n'est plus nécessaire.</p>
---	---

Fig. 1 Prise d'écran du Grand Dictionnaire Terminologique

Il est à noter que l'adaptation systématique des fiches terminologiques repose sur des schémas de standardisation et relèvent du souci de précision afin d'assurer la diffusion de données ; toutefois, ce principe existait avant l'informatisation ou la terminotique et figurait dans les travaux d'Eugen Wüster [9]. La terminologie de l'École de Vienne décrivait le terme d'une manière régulière et systématique. Depuis les années 1960, la gestion terminologique utilisait le support informatique afin de traiter « la masse de données à traiter dans certains organismes, mais également par la nature des termes qui semblent bien se prêter à un encodage rigoureux et systématique. » L'Homme [1].

Apprendre à sélectionner un format unique de description terminologique constitue un premier contact des normes à suivre dans un projet de recherche. La compétence visée porte sur la sélection des champs de la fiche et de la visualisation des données terminologiques, ce qui implique la standardisation de la catégorie grammaticale, de la définition, des critères de sélection des termes associés et du système de renvoi, dans ce cas modélisé sous forme de possibilités de liens hypertextuels pour un plus grand degré de navigation dans le glossaire grâce à son format électronique.

Puis, nous abordons la différence entre recherche thématique et ponctuelle, l'exploration par thèmes est examinée à travers le parcours des listes d'abréviations dans les dictionnaires de langue générale dans le cours suivant. Les apprenantes dressent la liste des abréviations repérées dans les dictionnaires de langue générale : le Petit Robert, Larousse, Hachette et donnent des termes relatifs au domaine choisi. Dès le début, les activités préparent, à l'avance, le contenu du cours et l'apprenante cherche et présente les informations collectées.

Un deuxième modèle de fiche terminologique, illustré par la figure 2, introduit les notions de variantes orthographiques, synonymiques dans le cadre des travaux du Réseau des traducteurs et traductrices en éducation [16]. Des *justifications* concernent des précisions sur l'usage du terme par les services concernés dans le champ disciplinaire spécifique ainsi que des notes sur un autre terme *déconseillé*, et sur la prononciation du terme. Le contexte explicatif montre les conditions d'application du concept et contient des termes associés (ex. *faits allégués, mention, crédits*) et des collocations (ex. *accordée*), etc.

La précision et la révision reste des valeurs essentielles dans tout travail linguistique ; les apprenantes ont pu relever une erreur orthographique dans la fiche objet d'étude : *examens spéicaux* au lieu de *spéciaux*. En effet, sur le plan professionnel, le terminologue se trouve dans l'obligation de respecter les normes exigées pour la présentation typographique des ressources terminologiques.

TERME ÉTUDIÉ : cote aegrotat
<p>VARIANTES ORTHOGRAPHIQUES :</p> <p>SYNONYMES :</p> <p>TERMES DÉCONSEILLÉS : certificat d'indisposition</p> <p>DÉFINITION : Mention correspondant à une note de passage et à une valeur en crédits accordée à une étudiante ou à un étudiant dans des circonstances spéciales, comme la maladie, un accident ou des problèmes familiaux, qui l'empêchent de terminer les travaux d'un cours ou de se présenter à un examen; les raisons invoquées doivent être attestées par des pièces justificatives</p> <p>CONTEXTE : « On accorde rarement la cote aegrotat pour la totalité des cours. Selon les faits allégués et les preuves à l'appui, on permettra plutôt à l'étudiant ou à l'étudiante de se présenter à des examens spéciaux ou différés, sans frais supplémentaires. Lorsque la cote aegrotat est accordée, le cours ne comprend aucune note. On n'accordera pas la cote aegrotat à plus d'un cours dans la matière de concentration ou de spécialisation de l'étudiant ou de l'étudiante. »</p> <p>JUSTIFICATION : Le Réseau écarte le terme «certificat d'indisposition», car il ne couvre pas la notion étudiée; il se peut que l'étudiante ou l'étudiant à qui on accorde une cote aegrotat reçoive aussi un certificat d'indisposition, mais les deux ne sont pas interchangeables.</p> <p>Le Réseau retient le terme «cote aegrotat», car plusieurs établissements postsecondaires l'utilisent déjà dans leurs annuaires et aucun autre terme ne lui fait concurrence. À noter que le terme d'origine latine «aegrotat» s'utilise de préférence en italique. De plus, le «t» final ne se prononce pas et les lettres «a» et «e» sont soudées, respectant ainsi la prononciation et la graphie d'origine.</p> <p>DATE DE RÉDACTION : novembre 1997</p>

Fig. 2 Fiche terminologique du Réseau des traducteurs et traductrices en éducation

Ces activités pratiques se font suivre de lectures d'articles de référence sur les éléments théoriques mentionnées, et nous faisons référence aux travaux de L'Homme [1]; dans cet ouvrage, les structures terminologiques et la confrontation des deux approches, conceptuelle et lexico-sémantique, explicitent l'essence de chacune des démarches avec des exemples. La première s'intéresse à la présentation des relations ontologiques (se plaçant du côté cognitif) alors que la seconde prône l'examen du terme dans le texte spécialisé en tant qu'unité lexicale, d'où l'inclusion des verbes, des adverbes en plus des noms privilégiés dans l'approche onomasiologique.

2) Introduction des réseaux notionnels basés sur l'arbre du domaine

Les relations entre les termes ont longtemps dominé les représentations terminologiques et l'approche cognitive est basée sur le classement des concepts dans un champ selon une organisation précise. Les relations dites logiques classent les termes dans un champ conceptuel selon une certaine hiérarchie : superordonné, subordonné et coordonné. L'arbre du domaine illustre bien ces hiérarchies et organise les relations conceptuelles d'un domaine donné.

L'exemple abordé dans ce cours est la pollution, un domaine qui gagne plus d'intérêt actuellement et dans lequel les concepts sont classés selon diverses catégories ; il renvoie entre autres aux actions (types de pollution), aux causes (*anoxie*) aux résultats (*déforestation*), aux substances (*polluants*), aux processus (*déséquilibre*, *dépollution*). Les apprenantes ont été guidées pour réaliser les cartes conceptuelles de ce domaine et chacune a présenté son travail qu'elle a présenté oralement dans le cours puis déposé sur la plateforme Moodle. Nous présentons deux modèles préparés par les apprenantes qui illustrent la disposition des concepts dans le domaine de la pollution. Dans ce cadre, la carte conceptuelle est un support permettant d'identifier les interrelations conceptuelles.

Cet exercice de conception de carte mentale pour ordonner les concepts relève des compétences cognitives et répond aux besoins de classification et de tri des données terminologiques. Un apprenant universitaire est confronté à soumettre une conception hiérarchisée des concepts dans un domaine, c'est-à-dire que l'apprentissage est couplé d'une série de stratégies de gestion et de tri des données terminologiques pour illustrer la classification en catégories (superordonné, subordonné et coordonné). Il est alors légitime de se demander quels sont les critères selon lesquels on pourrait classer un terme comme *terme associé* au terme vedette. Vu sous cet angle, la recherche terminologique - orientée vers le projet terminologique individuel- suscite certainement des difficultés et des réflexions sur les prises

de décisions vis-à-vis des champs conceptuels. La sélection de la liste des termes-clés et le choix des descripteurs nécessitent un esprit de déduction des relations générique/spécifique, partie/tout, etc.

La pratique envisagée est de proposer des domaines de recherche et de donner à chacun la liberté de choisir son domaine et de préparer son fichier. La mission est totalement soumise à des préférences personnelles que l’enseignant guide. Ce dernier suit le développement du projet et accompagne l’apprenant dans ses décisions : choix du corpus, choix des termes, choix du modèle de fiche terminologique et sélection des données terminologiques pertinentes.

La gestion du projet inclut des compétences méta-cognitives classées en trois catégories : « personne », « tâche » et « stratégie ». Ces trois catégories expliquées par Frenkel [17] impliquent des « connaissances relatives à notre propre comportement d’apprentissage ». Les motivations sont au cœur de l’apprentissage par projets. L’enseignant fournit une stratégie de gestion du temps en fonction des objectifs annoncés, et accompagne la planification, le processus de priorisation, d’autorégulation, recherche autonome de l’information. Les apprenantes cherchent à s’informer sur la meilleure façon de représenter les données recueillies dans un cadre formel uniforme à tous les termes. Cette uniformité repose sur les principes ou normes de la *représentation des connaissances*.

Développer des compétences en ce qui concerne le traitement des informations : extraction des données du corpus ; sélection des éléments terminologiques pertinents ; et maîtrise des techniques informatiques (fonctionnalités des logiciels, les expressions régulières de recherche d’informations).

La présentation des relations décrites et leur nombre repose sur une « *représentation conceptuelle* » résultant de critères variés mais définis à l’avance par les spécialistes du domaine. L’Homme [1] donne l’exemple de classification des animaux en biologie où les espèces animales sont classées en fonction des caractéristiques morphologiques. Un critère est, selon Raynal et Rieunier [18], un «élément auquel on se réfère pour porter une appréciation, un jugement». Cependant, les parentés entre ces espèces ne prennent pas en compte ni l’habitat ni le mode de déplacement.

Par ailleurs, dans le domaine de la pollution, certaines apprenantes ont pu chercher sur le Web et ont alors classé les termes, d’une part, en sous-domaines selon le type de pollution, cela est indiquée dans la carte mentale de la figure 3.

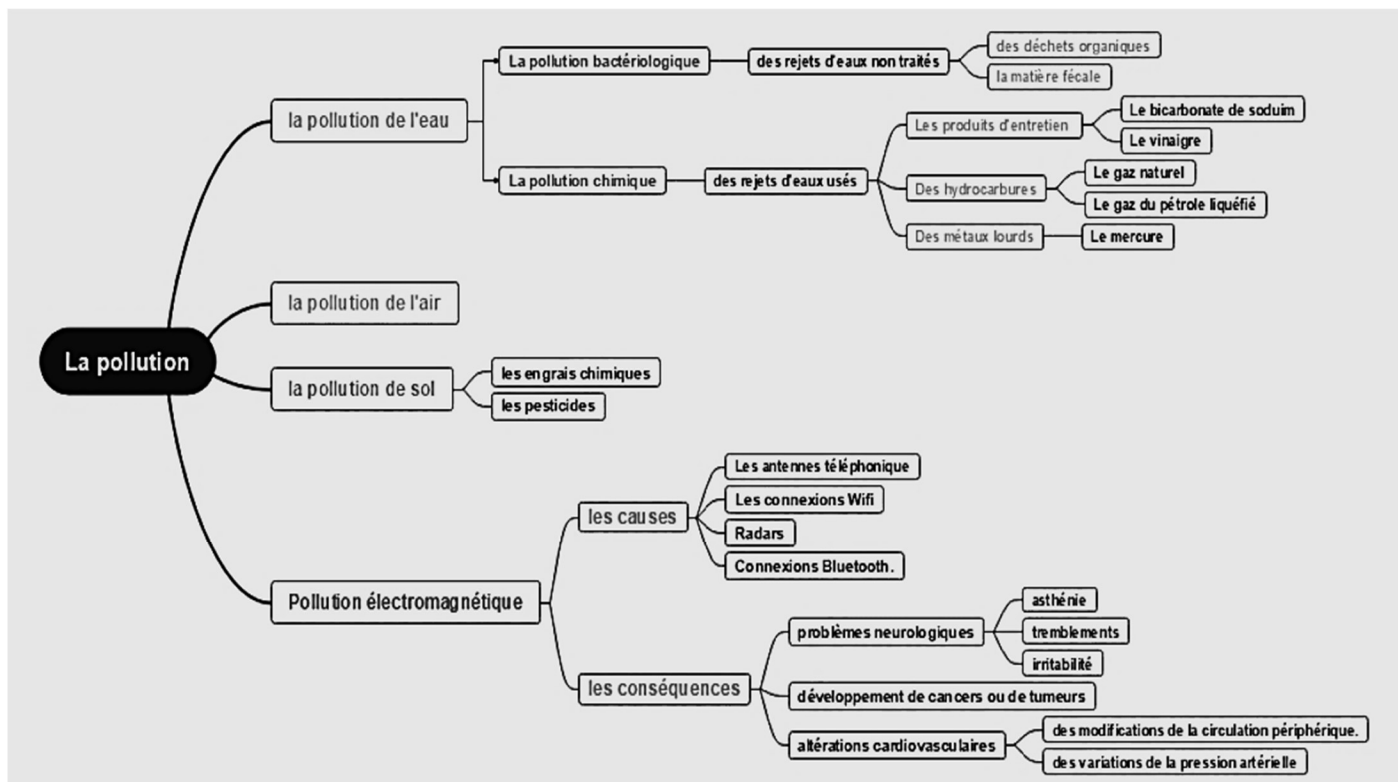


Fig. 3 Carte mentale préparée par une apprenante

D'autre part, une autre apprenante a choisi la forme suivante de la figure 4 portant sur les causes de la pollution de l'eau. Dans cette figure, nous pouvons observer des erreurs de frappe. Lors de la séance de mise en commun des travaux des apprenantes, les corrections visent les modifications visant à respecter les règles orthographiques. Il faut écrire polluants domestiques au lieu de domestiauxes, et corriger l'absence de l'accent aigu dans polluants bacterologiques).

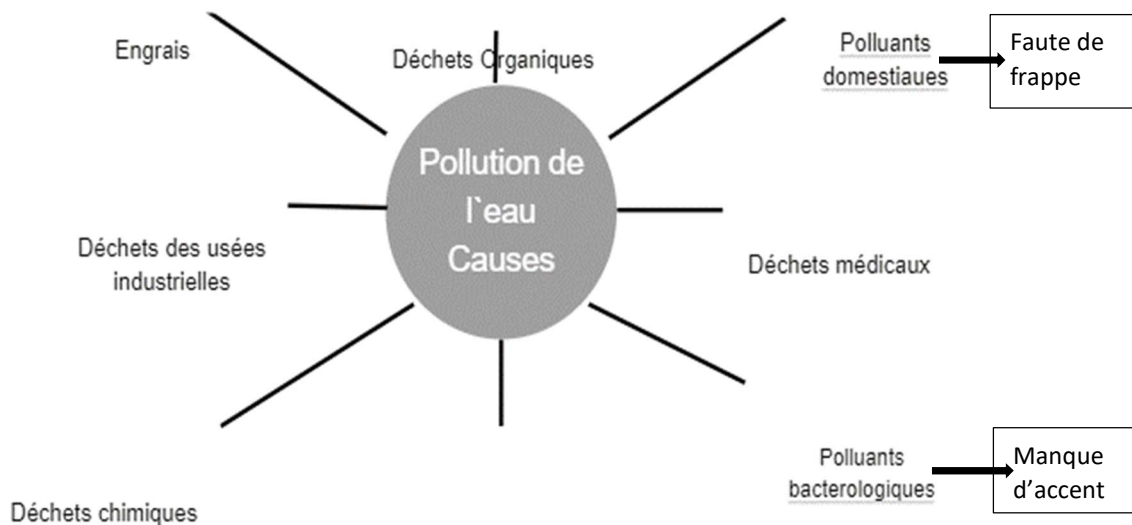


Fig. 4 Carte mentale sur la pollution de l'eau préparée par une apprenante

Bommier-Pincemin [19] précise que la définition d'un corpus orienté vers le traitement doit remplir des conditions de: signifiante, acceptabilité, et exploitabilité. Le corpus de textes se plie aux instructions élémentaires pour le recueil des textes et la constitution de corpus lexico-métriques : normes de saisie, nettoyage des données, opérations nécessaires à tout traitement automatisé. Il faut ajouter la pertinence des documents comme critère de choix : « à quelles conditions ce qui a été rassemblé pour servir un objectif peut être recyclé pour en servir un autre ? ». Les documents doivent apporter des éléments de réponse aux problématiques soulevées et aux objectifs de l'analyse.

Les métadonnées ou les données bibliographiques des documents textuels numériques sont citées d'une façon précise qui sert à délimiter la source et la date. Des instructions sont fournies pour le recueil et la constitution de corpus. Le choix d'un corpus suppose une homogénéité et une représentativité en adaptation à l'objet d'étude ; c'est-à-dire que l'ensemble pourrait servir de réponses aux questions soulevées et les pistes possibles dans le cadre de la gestion des références.

Nous avons proposé aux apprenantes des domaines dans lesquels, se présentaient des lacunes comme : un manque de terminologie équivalente, un manque de documentation en arabe, notre objectif étant l'application d'une approche de l'apprentissage par problèmes (APP) (« problem-based learning »), selon laquelle « les apprenants, regroupés par équipes, travaillent ensemble à résoudre un problème généralement proposé par l'enseignant, problème pour lequel ils n'ont reçu aucune formation particulière, de façon à faire des apprentissages de contenu et à développer des compétences de résolution de problèmes. » [20]

Selon les normes de croissance professionnelle des TIC de l'ISTE [21], l'enseignant accompagne ses apprenants, en participant à la prise de décisions partagée et en amenant à l'acquisition des compétences technologiques, dans une démarche professionnalisante dont l'objectif est d'évaluer les nouvelles pratiques professionnelles. Utiliser efficacement les ressources numériques disponibles et optimiser la vitalité et le renouvellement des outils accorde à l'enseignant un rôle de facilitateur.

Pour ce module de terminotique, le logiciel AntConc, conçu par Anthony [22], est exploité pour des fins d'extraction à partir de corpus compilés par les apprenantes.

C. Utilisation du logiciel Antconc

Extraire des données informatiques signifie avoir la possibilité de formuler des requêtes et de poursuivre l'analyse des résultats affichés, la prise en main des fonctionnalités du logiciel facilite davantage le temps des opérations relatives à l'objectif *s'informer pour extraire les données*. Cette extraction résulte d'une interaction Homme-machine sur laquelle se construit l'utilité du logiciel. Quels outils informatiques à destination des professionnels en terminologie et adéquats aux besoins des apprenantes ?

Ce processus d'exploration de corpus numérique offre des possibilités immenses, mais il englobe toutefois des contraintes relatives à la mise en forme des documents du corpus en format texte ; les problèmes posés par la difficulté de transformer des documents photocopiés par des procédés de reconnaissance optique de caractères, l'interrogation des concordanciers, l'étude des corpus comparables, etc.

AntConc est un logiciel d'analyse textuelle gratuit et téléchargeable pour une exploitation dans les systèmes Windows et Mac. Il sert à extraire les termes simples et complexes dans un corpus de textes en arabe ou en français. Des précisions sont données sur le format des fichiers que supporte le logiciel pour commencer l'application. Il est nécessaire de transformer les données en fichiers texte. Les titres de fichiers sont écrits en français même en cas de fichiers de textes arabes. Une fois le corpus préparé, les requêtes lancées font l'objet d'examen pour la prise de décisions. Il s'agit de compiler un corpus en format numérique et le transformer en fichiers texte.

Interroger un corpus au niveau morphologique est désormais possible, par exemple, dans la figure 5 on peut chercher tous les noms qui se terminent par le morphème *tion et l'on obtient la liste de concordances suivantes relevées par le logiciel AntConc . Nous obtenons une liste de termes à savoir : population, féminisation, éducation, salarisation, organisation, occupation. Ensuite, l'apprenante peut se déplacer de la liste des concordances aux textes en cliquant sur le terme surligné en couleur bleue. Il est également possible de repérer tous les syntagmes dans le corpus à partir de la fonction N-Gram et l'on obtient les listes de syntagmes fréquents : celles-ci contiennent des collocations et des termes complexes.

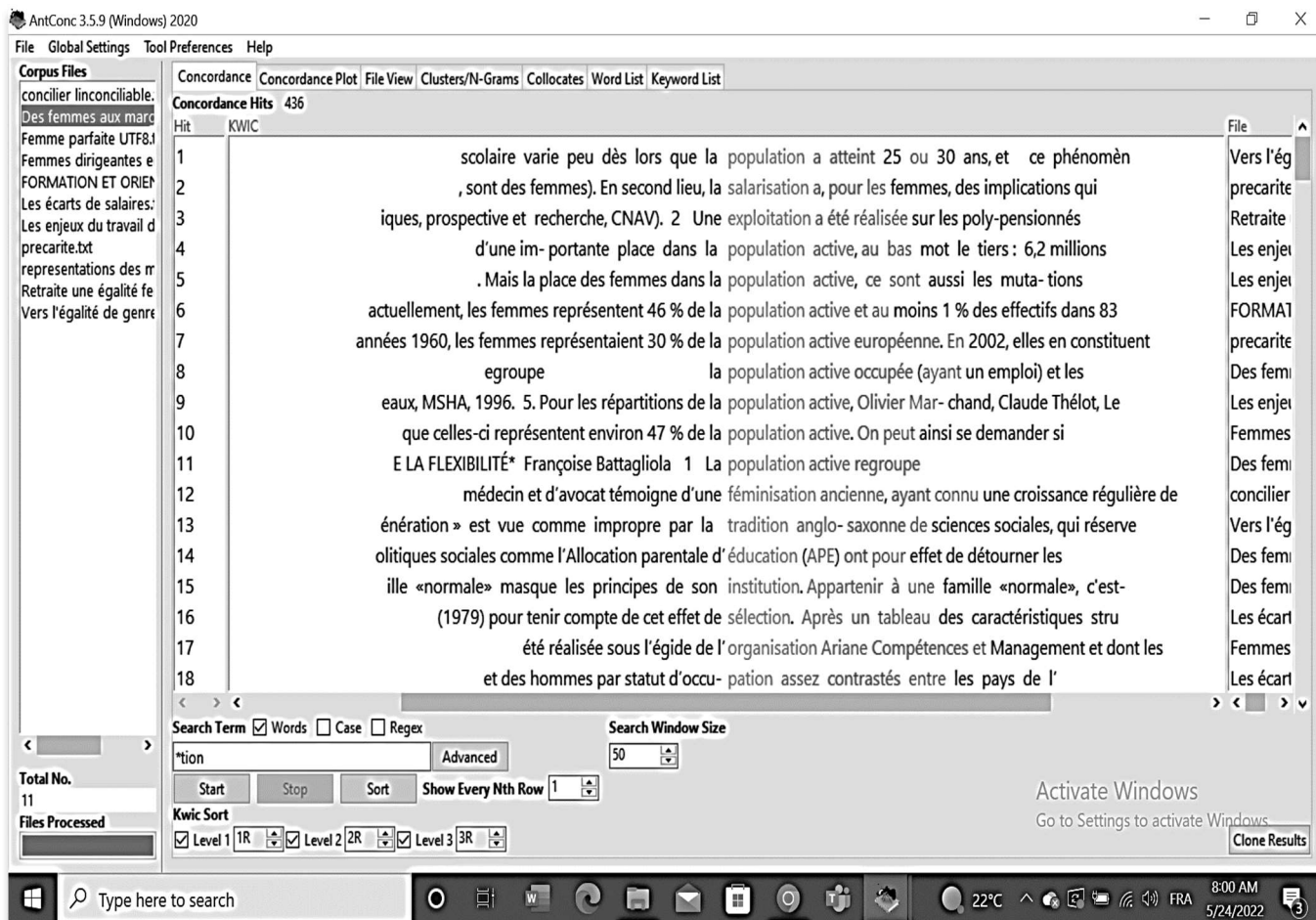


Fig. 5 Capture d'écran du logiciel AntConc : toutes les occurrences des mots se terminant par le morphème (*tion)

Nous avons remarqué que l'affichage des données sous la présentation KWIC (Key Word in Context) permet la consultation des contextes droit et gauche du terme-pivot. Un clic sur le terme permet l'accès vers l'article complet. Les listes de fréquence et ce parcours entre les résultats des recherches (mini-contextes) et le contexte élargi est une opportunité pour trancher sur le degré de lexicalité des faits observés dans les occurrences des termes et des collocations. En cas de doute sur le statut du terme, ou son appartenance au domaine terminologique, il est fortement recommandé d'avoir recours à un corpus de référence.

4 COMPÉTENCES ACQUISES : EXPÉRIMENTER POUR CHANGER LES PRATIQUES D'APPRENTISSAGE

Autonomiser l'exploitation des corpus et le traitement des analyses qui en résultent en vue d'une structuration de la description terminologique. Le repérage des termes et la recherche d'équivalents sont deux objectifs caractéristiques des traducteurs-terminologues. Initialement présents dans les textes spécialisés, ces termes sont recensés ensuite dans des dictionnaires à travers un processus de codification qui repose dans son organisation, selon Cabré [23], sur des étapes qui se rattachent aux démarches de la réalité professionnelle de la recherche documentaire indispensable à la collecte terminologique.

A la fin de ce module, les apprenantes peuvent choisir les candidats-termes et leurs équivalents dans deux corpus comparables. Elles arrivent à classer les termes selon leur mode de formation (formée d'un ou de plusieurs signes linguistiques pour désigner un concept). La méthodologie de recherche assistée par ordinateur déploie des compétences centrées sur les capacités de sélectionner, éliminer et formaliser la description linguistique des données

terminologiques. Elles maximisent les recherches sur des corpus, c'est pourquoi elles assument leurs choix. Deux éléments ont été traités lors des analyses des N-Gram extraits du logiciel : quels groupes syntagmatiques inclure dans les fiches ? La collecte des associations combinatoires se fait par l'analyse des fréquences des collocations.

L'intelligence artificielle se base sur l'utilisation de l'informatique conçue comme outil « véhicule de connaissances » favorisant l'interaction homme-machine [24]. Appliquée en terminotique, les recherches en IA abordent les moyens de dépouillement, de visualisation des données et de modélisation des liens notionnels.

Les extracteurs mettent en œuvre l'analyse de texte assistée par ordinateur qui est considérée comme un complément à la compétence documentaire : par exemple, elle peut fournir une confirmation quantitative des modèles que vous observez dans un texte, vous permettre de repérer rapidement des mots ou des phrases intéressants dans un grand corpus, de les localiser dans le contexte élargi et aider à contextualiser les préférences d'emploi des termes. Par conséquent, l'utilisation d'un outil comme AntConc dans le cadre de pédagogie par projet peut également aider à :

- développer des stratégies d'enseignement centré sur l'apprenant ;
- créer un contenu adapté aux besoins des apprenantes ;
- définir les critères de constitution des corpus ;
- informer sur les fonctionnalités des outils informatiques ;
- orienter les activités pratiques vers la recherche sur corpus.

L'évaluation des compétences liées à notre étude devrait être définie ; à quelle mesure ces compétences sont mesurables ? Quelles compétences mentales sont associées à ces manipulations de recherche et de modélisation des données terminologiques ? Comment agencer ces compétences au niveau des facultés dites de haut niveau (résolution de problème, raisonnement, activités de systématisation des représentations langagières).

Le volet cognitif constitue un des fondements de notre démarche puisque la planification du projet individuel comprend des prises de décision et des argumentations concernant la justification du choix du domaine de spécialité, la définition des objectifs de dépouillement, les critères de sélection de la documentation ou corpus d'étude, la construction d'une part de la fiche terminologique et, et d'autre part de l'organisation des liens hypertextuels pour une prochaine représentation basée sur la navigation hypertextuelle.

Dans la démarche suivie, nous nous limiterons aux compétences liées à l'emploi des outils informatiques ou extracteurs ; en effet, l'Intelligence artificielle se base sur la représentation de données (reproduction des processus mentaux) soit comme véhicule de *connaissances*. Dans le cadre de cette étude, les apprenantes développent des pratiques centrées sur la recherche de données. Les notions d'*ergonomie informatique*, d'utilité et d'utilisabilité sont introduites implicitement lors des séances d'expérimentation d'emploi du site de DicoInfo [25], du GDT (Grand dictionnaire terminologique) [15] et enfin du logiciel AntConc [21].

La période de prise en main de ce dernier dépend de critères de motivation, d'expertise et de littéracie numérique. L'*ergonomie informatique* désigne selon les travaux de l'Institut Grasset [26] « la capacité à identifier les freins qui compliquent l'usage d'un site web, d'un logiciel ou d'une application web ».

Le volet pratique de notre module porte essentiellement sur une formation en techniques informatiques : collecte de données, préparation des formats txt, expérimentation du logiciel, analyse des résultats et modélisation sous forme de fichier terminologique, éventuellement informatisé. La première étape consiste alors à faire le lien entre les méthodes conventionnelles de documentation exigée pour la réalisation d'une traduction correcte (le choix de la bonne terminologie, la consultation des thésaurus, la structuration classique des arbres de connaissances, etc.) et le développement des approches informatiques de traitement des données (visualisation des interfaces) afin d'assurer leur adéquation aux capacités perceptives et cognitives des utilisateurs.

Pendant les cours, les apprenantes sont amenées à se familiariser avec les fonctionnalités du logiciel et puis à mesurer à quel point les résultats sont pertinents par rapport aux besoins initiaux de leur recherche. Comment lancer une recherche sur la fréquence des combinaisons ? Comment préciser la taille des N-Gram ? De telles interrogations dépendent du taux d'implication et de pratique autonome à propos de termes précis. Ce système permet à chaque apprenant de développer, selon Seus-Walker [27], l'esprit d'initiative et d'indépendance. Face à son ordinateur,

l'apprenante s'investit et vérifie elle-même ses compétences en suivant des requêtes similaires à celles présentées en cours en progressant sans besoin d'interventions systématiques de l'enseignant, et donc elle gagne du temps.

Par ailleurs, la communication sur le forum du Moodle offre à tout moment la possibilité de solliciter l'avis de l'enseignant en cas de besoin de vérification ou de manque d'information sur un cas précis. Ainsi, il est plus raisonnable d'opter pour une stratégie de semi-autonomie qui exige une préparation des difficultés possibles et de les surmonter afin de pouvoir répondre aux besoins particuliers de chacun. Les demandes sont tantôt formulées en public dans le cadre du groupe classe, tantôt en privé où l'enseignement se fait « sur mesure » sur les détails dans la constitution des données ou les difficultés de l'extraction. Discuter les remarques collectives s'alterne avec l'accompagnement des difficultés personnelles réglées au cas par cas.

A. Sélection des candidats-termes

La méthodologie a été bien décrite de sorte à pouvoir détailler les étapes et situer la progression du projet. Savoir nommer les opérations de compilation du corpus, de constitution des données, de fouille et de modélisation, tel est l'objectif de la démarche terminologique. La mesure de la qualité des fichiers dépend de la systématisme des champs ou rubriques, le choix de contextes adéquats au fonctionnement conceptuel, la description des données terminologiques repérées. L'exemple dans la figure 6 détaille les opérations de collecte faite par une apprenante.

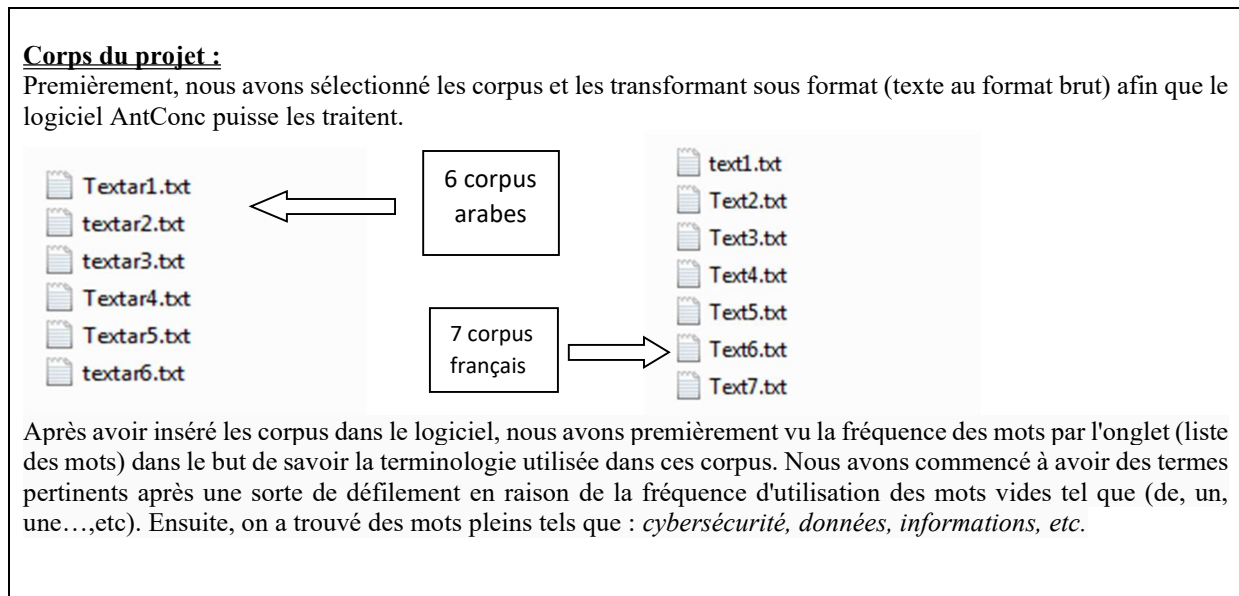


Fig.6 Étapes de constitution du corpus d'études présentées par une apprenante en traduction

Une première remarque porte sur les listes de fréquence, les mots les plus fréquents sont les mots vides ; puis chaque apprenante détermine les étapes de recherche. La figure no. 7 constitue la liste de mots (*Word List*) résultant de l'extraction d'un corpus de textes dans le domaine de la cybersécurité collectés dans des sites Internet. Les termes simples ou mots pleins sont formés de noms (البيانات- التهديدات- الأمن- الأمان) ou d'adjectifs (السيبراني-الالكتروني). Les mots vides sont essentiellement des prépositions, des pronoms (أن- إلى - في - أو- من- التي).

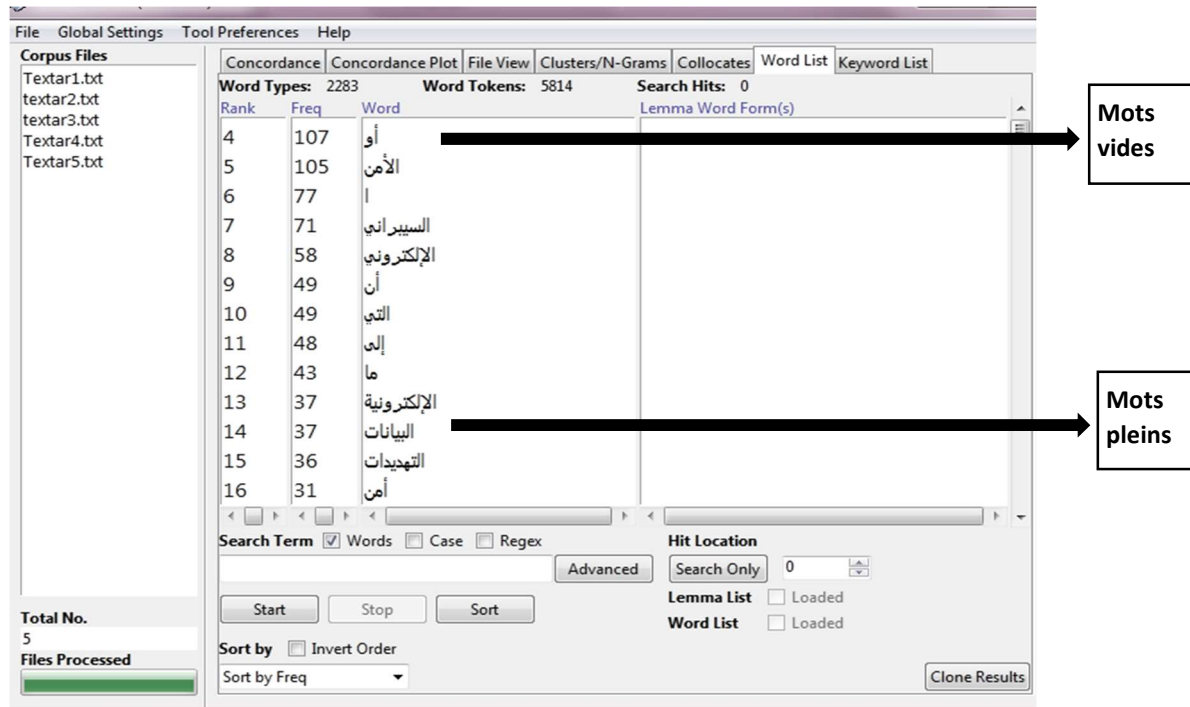


Fig. 7 Classification par une apprenante des listes de termes en mots pleins et mots vides

La méthodologie de travail, exposée dans les projets individuels, suit ce qui a été fait en cours pour explorer un corpus déjà compilé à l'avance. Notons les commentaires des apprenantes pour décrire leur méthodologie.

« Premièrement, nous allons choisir des textes arabes et des textes français sur le même domaine puis nous les transformerons en format txt d'encodage (Utf-8) pour préparer les corpus desquels les données vont être extraites. Deuxièmement, nous allons obtenir la liste des mots en pressant sur l'onglet (Word List). Troisièmement, nous chercherons les différents contextes dans les concordances à travers l'onglet (concordanciers), ce qui affiche les concordances et le nombre de résultats obtenus (Concordance Hits). Finalement, nous allons obtenir les N-grams et les collocations pour relever les expressions les plus fréquentes pour les patrons syntaxiques.» (extrait du projet d'une apprenante N°5)

Concevoir et organiser une méthodologie claire, tel est le premier principe à respecter. Lors du dépouillement, l'apprenante recense, dans la figure no.8, la liste des termes détectés par le logiciel AntConc et indique le critère de classement par ordre décroissant des mots pleins ou candidats-termes tels que : *utilisateur ; entreprises, informatique ; systèmes ; attaque ; logiciels ; malveillants ; phishing ; système ; réseaux ; etc.* Ces termes proviennent d'un corpus de cybersécurité.

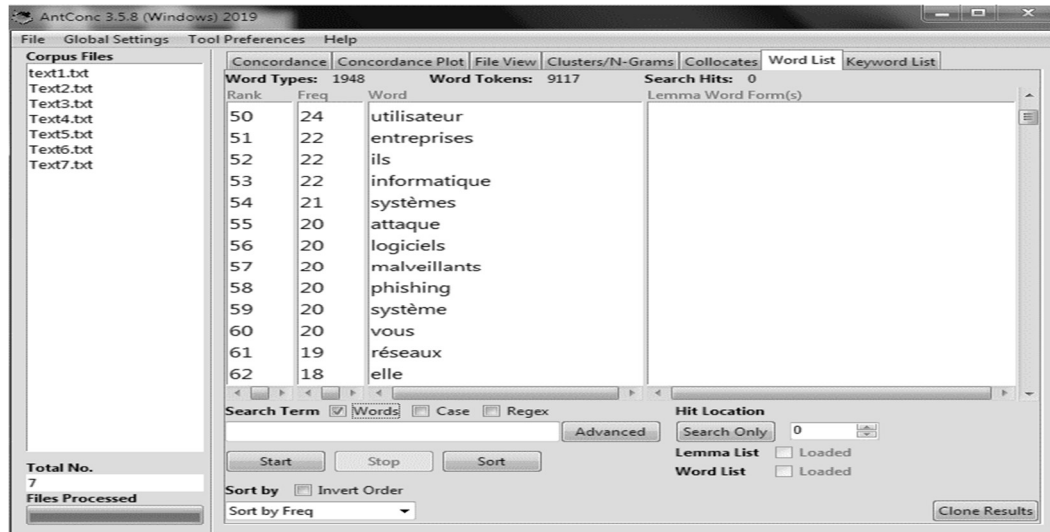


Fig. 8 Capture d'écran de la suite des résultats de l'onglet (liste des mots)

- Remarques sur les observables (liste des mots) :

L'apprenante arrive dans sa recherche à préciser le classement décroissant des termes d'après la fréquence d'apparition dans le concordancier.

« Lors des captures d'écran, on a trouvé que le terme le plus utilisé était la cybersécurité avec une occurrence de 88 fois. C'est logique puisque c'est le domaine de la cybersécurité. Ensuite, le terme (données) a été utilisé fréquemment avec une occurrence de (81) fois. Ensuite, il y avait d'autres termes tel que : (sécurité, informations, systèmes, utilisateurs, logiciels, phishing...etc) » (extrait de la recherche de l'apprenante N°6)

- Analyse des concordances du terme *logiciel*

Les concordances dans la figure no. 9 font apparaître les différents syntagmes dans lesquels le terme apparaît. Ces syntagmes sont classés en termes complexes :

- Termes complexe Nom+de+Nom (*logiciel de sécurité*)
- Termes complexe Nom+adjectif (*logiciel malveillant ; logiciel authentique*)

Dans le cours, nous avons déjà abordé les modes de formation des termes, la partie théorique sert d'introduction aux applications pratiques en rapport avec le repérage des termes et à la description de leurs formes après l'extraction.

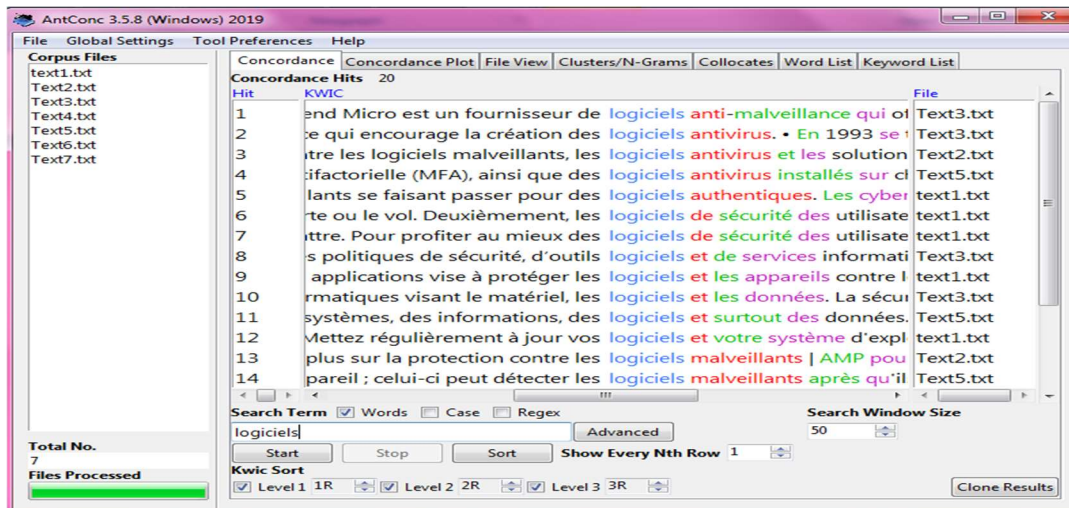


Fig. 9 Capture d'écran du terme 'logiciel' dans l'onglet 'concordances' (projet apprenante N°7)

Les concordances du terme sont une source riche de contextes d'apparition du terme-pivot dans le corpus, elles s'affichent et décrivent les réalisations du terme dans le discours, ce qui révèle son usage indirectement par le recueil d'un ensemble de variantes synonymiques, paronymiques, de rapports générique/spécifique, relations d'hyponymes/hyperonymes, etc.

La sélection et l'organisation de données terminologiques résultent de l'exploration des concordances dans le but d'y trouver des modèles de classification des désignations ou des représentations pertinentes et correctes. Par ailleurs, les relations entre les termes semblent un des fondements de la terminologie qui figurent parmi les objectifs du module et que nous exposons dans la partie suivante de notre recherche.

B. Élaboration des relations entre les termes dans l'arbre du domaine

Le repérage des termes se fait généralement suivre d'une analyse des différentes relations sémantiques dans lesquelles entrent les termes. Cette analyse est une compétence indispensable pour déduire les structures qui expliquent ces relations et les représentent selon des modes de classement variés ; le premier ensemble est formé de classements thématiques, qui s'intéresse à l'organisation des notions selon des index thématiques. Le deuxième est l'arbre de connaissances ou arbre de domaine dans lequel apparaissent les relations conceptuelles, appréhende les termes englobant les connaissances dans un domaine de spécialité et les catégorise selon des liens hiérarchiques divisés en: génériques/spécifiques, partie/tout, cause/effet, etc.

Les liens entre les termes sont indispensables dans une représentation onomasiologique. Polguère [28] assure que « le lexique est un gigantesque réseau où tout se tient. Il suffit de « tirer » sur une lexie de ce réseau pour que vienne avec elle toute une série d'autres lexies auxquelles elle semble attachée par des liens parfois très subtils. » Ce réseau de termes dépend de la classification variée selon les domaines de spécialité. Le tableau no. I présente une contribution d'une apprenante dans le domaine de la pollution pour décrire les termes associés au domaine de la pollution lors d'une première recherche personnelle.

TABLEAU I
TERMES ASSOCIÉS REGROUPÉS PAR UNE APPRENANTE (No 5)

Termes associés à la pollution	
Terme	Mode de formation
1) Antipollution (antonyme de pollution)	préfixation
2) Dépollution (des concepts en opposition par rapport à pollution).	préfixation
3) Hypothèque environnementale : désigne <i>Charge polluante libérée</i> et abandonnée dans la nature de toute région de la planète à la prochaine génération. (Définition du GDT)	Composition (Nom+adjectif)
4) Ignifuge	Terme simple en français Terme complexe en arabe (مضاف ومضاف اليه) (مانع الاحتراق)

La collecte des termes associés à la pollution de l'air/ atmosphérique a donné la liste suivante : contamination, écosphère, atmosphère, couche d'ozone, composé organique volatil, fréon, fumée, gaz à effet de serre, smog, particules fines).

L'extraction dans le corpus comparable dans ce même domaine sert à trouver les équivalents et à déterminer les moyens de formation des termes comme dans le cas de :

Halocarbure = الهيدروكربونات المهلجنة (Un terme simple en français est traduit par un terme complexe en arabe) ; *smog* الضباب الدخاني (Un terme simple en français rendu par un autre complexe en arabe).

C. Choix des contextes en fonction de leur typologie

Le contexte occupe une place importante en terminologie. Toutefois, Hmida [29] souligne qu'« avoir accès à un terme ou à sa traduction ne suffit pas, encore faut-il être capable de l'employer correctement et d'en appréhender le sens exact.» Au niveau du discours, le contexte fournit des informations concernant l'usage des termes ; sur le plan terminologique, il informe sur les relations sémantiques entre les termes et fait partie de la classe des contextes riches en connaissances conceptuelles.

L'Homme [1] affirme que les contextes renferment souvent des éléments pour la définition ; des indices sur les relations entre les concepts (rapports de synonymie, de co-hyponymie ou d'antonymie) ; des variantes terminologiques et parfois des cooccurrents.

D. Rédaction des structures actancielles

La rédaction des définitions ressort des principes de modélisation en fonction de la lexico-sémantique ou actancielle. Cette approche relève de la combinatoire lexicale, nous cherchons à regrouper, pour un terme donné, ce que L'Homme et Meynard [30] appellent *combinaisons lexicales spécialisées* ou CLS. Elle donne l'exemple de la possibilité de retrouver tous les cooccurrents à partir de *programme* dans le domaine informatique. Les contextes relevés explicitent les unités lexicales avec lesquelles le terme se combine dans les textes spécialisés : *écrire, exécuter, exécution, actif, etc.* Le terminologue doit insérer ces combinaisons dans une liste subordonnée au terme-vedette *programme*. L'usage des CLS est essentiel dans la construction des savoirs spécialisés, leur présence est formalisée en fonction des structures actancielles telles qu'elles ont été définies par Mel'čuk et al. [31] : chaque verbe est accompagné des *actants* (*arguments*), c'est-à-dire les *éléments participant à l'action qu'il dénote*. Nous commençons par l'examen de la présentation des actants dans le DicoInfo (site du dictionnaire fondamental de l'informatique et de l'Internet) comme indiqués dans les exemples suivants où le terme est suivi d'une précision de sa structure actancielle (ex. X enregistre Y sur Z).

Nous étudions plus particulièrement le processus de collecte de la combinatoire dans les textes spécialisés tirés des projets des apprenantes selon les domaines variés. Dans le domaine de la sécurité informatique, la figure no. 10 présente une fiche élaborée par une apprenante et illustrant la structure actancielle du terme *phishing*.

L'entrée
Phishing
Définition
Technique de fraude basée sur l'usurpation d'identité, qui consiste à envoyer massivement un message en se faisant passer pour une institution financière ou une entreprise commerciale afin d'induire les destinataires en erreur et de les inciter à révéler des informations sensibles à leur insu.
Structure actancielle
Phishing : ~ fait par un <u>fraudeur</u> pour un <u>organisme</u>.
Catégorie grammaticale
Nom masculin
Le contexte
Les messages du phishing ciblé sont généralement conçus pour ressembler à des messages envoyés par une source fiable.
Source du contexte
Kaspersky. "Conseils Sur La Façon De Vous Protéger Contre La Cybercriminalité." Www.kaspersky.fr , 9 Mar. 2022, https://www.kaspersky.fr/resource-center/threats/what-is-cybercrime .
Synonymes
Hameçonnage
Termes associés
Fraudeur-internet- escroquerie
Equivalent en arabe
التصيد الاحتيالي

Fig. 10 Fiche finale réalisée par une apprenante du terme Phishing

La fiche est conçue de sorte de faciliter l'accès aux actants par hyperliens. La même modalisation se fait pour les termes arabes.

Vishing : ~ commis un fraudeur par un appel téléphonique (l'outil), contre une victime (destination).

En arabe, c'était compliqué au début, mais petit-à-petit, on réussissait à préciser les actants et à systématiser les couleurs correspondantes ; vert pour le patient, rouge pour l'agent et marron pour le moyen.

الاحتيال: ~ يقوم به شخص يسرق المعلومات الشخصية عن طريق رسائل خادعة

Cette modélisation peut servir à une mémorisation plus facile des champs conceptuels formés de termes sémantiquement liés selon une structure actancielle ; précisant les arguments ou actants (agent, patient).

E. Repérage des collocations

L'extraction des collocations dans le corpus se réalise par un examen des listes de N-Gram recensées par AntConc. Le profil combinatoire de chaque terme est basé sur une approche inductive ; les données du corpus selon les occurrences servent de méthode de collecte de l'environnement contextuel et des fréquences des segments répétés ou N-gram. Les caractères sont appelés uni-grammes ; les Bi-grammes désignent les combinaisons de deux caractères alors que les Tri-grammes sont formés de combinaisons de trois caractères.

L'apprenante cherche à analyser les formes collectées et puis à collectionner les collocations, en interpréter les acceptations et déduire les relations de dépendance syntaxique entre les termes et les collocations.

Pour traduire les collocations localisées dans le corpus français, les apprenantes ont recours à l'examen des N-Gram repérés dans le corpus comparable en langue arabe et insèrent les équivalents dans le fichier terminologique. Par exemple dans un corpus de textes commerciaux sur les conditions de vente, les concordances du terme شروط

(conditions), que nous présentons dans la figure no. 11, regroupent plusieurs collocations : conditions du site, conditions des commandes, conditions de vente, conditions des services, conditions d'utilisation, conditions de retour, conditions de paiement, conditions supplémentaires, conditions de garanties, autres conditions.

Ce type de recherche débouche sur une méthode d'encodage incluse dans le logiciel prévoyant l'accès à partir du mot-clé aux cooccurents qui, dans figureront dans la fiche du terme lors de la création du fichier comme collocations ; sur ce point, L'Homme et Meynard [30] réclament de consacrer une entrée pour le cooccurrent, plutôt que sa seule présence dans l'entrée descriptive du terme.

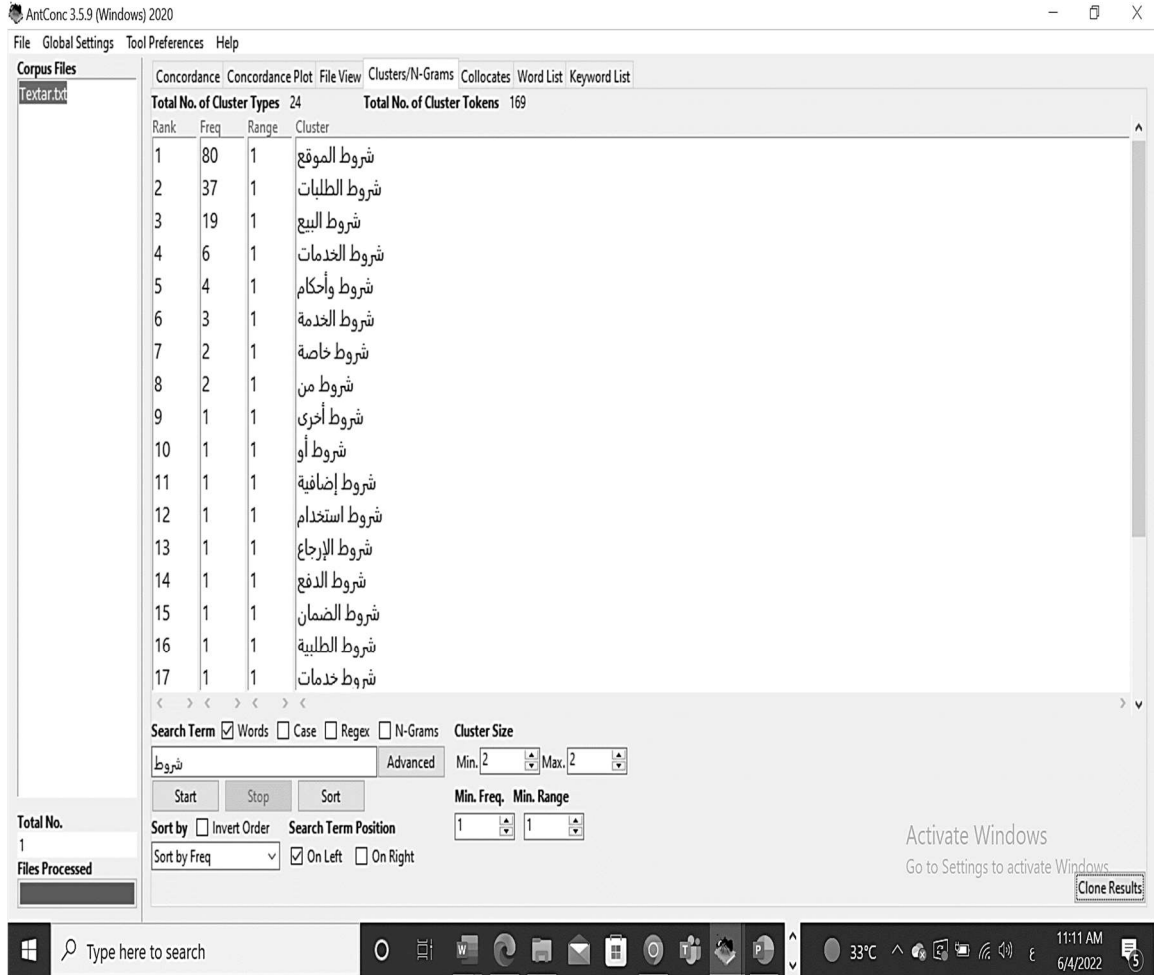


Fig. 11 Capture d'écran N-Gram du terme «شروط»

Au niveau des productions, l'essentiel à cette étape c'est de repérer les collocations et de les consigner dans les fiches comme précisé dans le tableau II.

TABLEAU II
FICHE TERMINOLOGIQUE DU TERME « VISHING »

Terme	Vishing
Catégorie grammaticale	Nom masculin
Définition	C'est une méthode d'escroquerie vocale par laquelle les criminels volent les données personnelles de victimes peu méfiantes.
Source de définition	C'est un néologisme qui ne figure pas dans un dictionnaire

Structure actancielle	Vishing : ~ commis un fraudeur par un appel téléphonique (l'outil), contre une victime (destination).
Contexte	Le vishing se sert d'arnaques vocales pour conduire les gens à faire des choses qu'ils pensent être dans leur intérêt.
Source du contexte	Kaspersky. (2021, January 13). <i>Qu'est-ce que le vishing ?</i> www.kaspersky.fr , https://www.kaspersky.fr/resource-center/definitions/vishing , consulté le 28 juin 2022.
Note	-
Source de la note	-
Collocation	se protéger contre le smishing.
Termes associés	Fraudeur, phishing, smishing, cyberattaques, cybercrimes, cybersécurité
Equivalent arabe	التصيد الصوتي

L'onglet N-Gram a affiché dans quelques travaux les résultats suivants dans la sécurité informatique :

- **Termes complexes tels que** →
(*Protection des données, mots de passe, sécurité informatique, programmes de sécurité, logiciel malveillant...etc*)
- **De collocations telles que** →
(**Cliquer** sur un lien, **voler** des données, serveur **contrôlé par**, **détecter** les nouvelles menaces...)

F. Sélection des termes associés

Lorsqu'un apprenant est formé à la recherche terminologique, il doit chercher à formuler ses requêtes selon les options de recherche présentes dans l'interface du logiciel qui, selon Manes-Gallo et Paganelli [24], permet « à l'utilisateur de traduire son besoin d'information dans le langage du système et de le transformer ainsi en requête. Celle-ci constitue une représentation du besoin d'information qui se fait au travers de l'utilisation du vocabulaire d'indexation. » Dans la figure no. 12, la recherche porte sur les concordances du terme *travail* dans un corpus d'études sociologiques. L'examen de cette liste montre deux groupes de données lexicales concernant les syntagmes nominaux formés autour du mot-pivot :

- relation type de travail (travail domestique, travail éducatif, travail des femmes) ;
- termes complexes (marché du travail, heures de travail) ;
- collocations nominales (enjeux du travail, exigences du travail, monde du travail) ;
- expressions (être sans travail, trouver du travail) ;
- termes dérivés (travailler, travailleur) ;
- termes en relation (employabilité).

L'apport de l'informatique est la variété des moyens de recherche et surtout à partir d'un mot-clé ; nous pouvons accéder aux différents termes d'un champ terminologique (synonymes, dérivés, abréviations, hyponymes, hyperonymes).

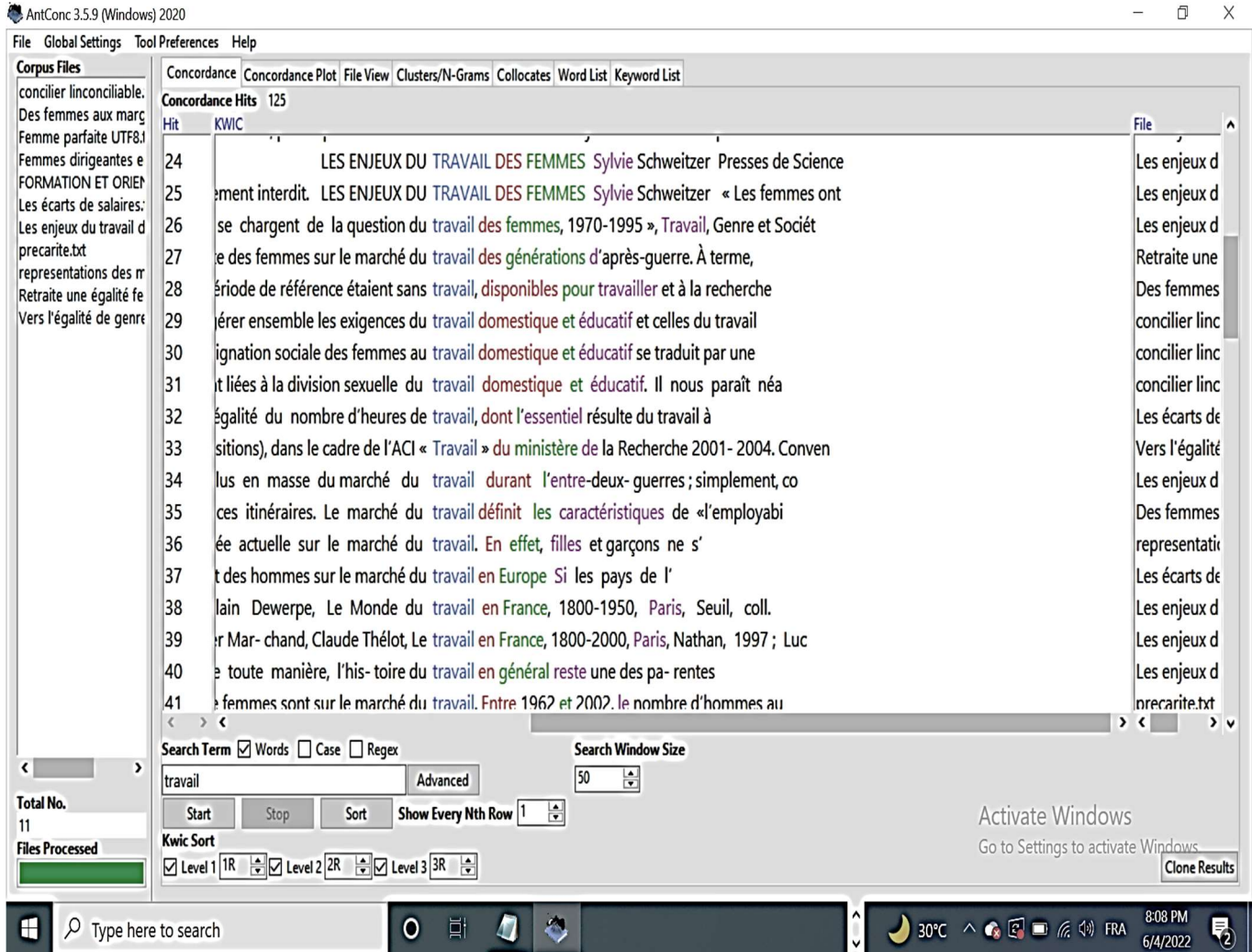


Fig. 12 Capture d'écran des concordances de travail

G. Modélisation sous forme d'hyperliens

L'hyperlien lie formellement deux éléments textuels appartenant soit à deux documents ou dans des parties différentes d'un même document. Les liens de ce genre sont utilisés dans la représentation des données terminologiques. L'Homme [1] souligne l'importance des hyperliens qui assurent une navigation rapide aux rapports préalablement établis par le terminographe. La diffusion finale des descriptions terminographiques gagne en efficacité et aide à se déplacer d'une fiche à une autre. L'exemple du terme « vishing » représente la structure actancielle conçue pour montrer comment les termes contenus dans des définitions sont reliés, au moyen d'hyperliens, aux articles où ces termes sont définis.

A partir d'un mot-clé, on peut avoir accès à tous les termes associés, à toutes les collocations, aux actants, aux équivalents et aux variantes synonymiques.

Vishing : ~ commis par **fraudeur** par **un appel téléphonique** (l'outil), contre **une victime** (destination).

La navigation hypertextuelle reste la caractéristique de la culture numérique, passer d'une page web à une autre et pouvoir revenir en un clic, cela offre un plus haut degré de déplacement partout dans les ressources ouvertes. C'est le

point positif de changement de la lecture linéaire des écrits traditionnels. La possibilité de visualiser ces déplacements est également une compétence à acquérir pendant la formation terminologique.

Du terme au texte et vice versa, les apprenantes gagnent en rapidité et en jugement de la pertinence des données observées. Les hésitations du départ s'évanouissent et se trouvent remplacées par une tendance à systématiser les régularités recueillies pour une meilleure gestion terminologique basée sur l'interactivité des logiciels. Ce parcours dans le fichier implique une bonne maîtrise des liens et de l'outil. Tout au long des phases du projet, l'apprenante se sert de ses connaissances, de références en ligne, de documentation du domaine et sollicite l'enseignant pour se rassurer à plusieurs reprises, tant au niveau de l'acquisition de techniques informatiques qu'au niveau du tri des données linguistiques. Alors, à quel point cette démarche sert aux objectifs initiaux d'autonomie et de réussite de la description linguistique ?

5 STRATÉGIES D'AUTONOMISATION

A. Accompagner pour guider

Seus-Walker [27] avance qu'une approche de l'autonomisation des apprentissages exige deux points complémentaires à savoir : la *responsabilisation* de l'apprenant et l'*investissement* de l'enseignant ; cette approche exige une réflexion sur les exploitations possibles et imaginables des outils technologiques et leur intégration selon une démarche selon des stratégies efficaces et convenables aux objectifs. En d'autres termes, les activités doivent se baser sur les acquis visés ou les résultats attendus de l'apprentissage.

Un autre facteur est celui de la gestion du temps, comment gérer le temps dont nous disposons ? L'informatique ne nous dispense en aucun cas d'une conception planifiée selon le volume horaire et les finalités. Comment peut-on déterminer si l'outil d'apprentissage est utile et efficace ? Les résultats des recherches sont-ils acceptables ? Quel est le taux de bruit ? Quelle approche didactique peut-on suivre concernant l'appréhension des outils ? Comment gérer l'angoisse lors des premières tentatives ?

Dans un enseignement à distance, l'enseignant occupe la place de « *facilitateur de savoir* » et ses pratiques, centrées sur l'interactivité et l'harmonisation entre partage et suivi, œuvrent en faveur d'un apprivoisement doux mais progressif des techniques de fouille désormais indispensable à des tâches actuelles telles que la recherche d'information. Cependant, il est à noter que certains obstacles résultent des opérations d'extraction que tout apprenant ne doit pas affronter seul. L'usage de telles techniques expose l'apprenant à des obstacles qui méritent plus de réflexion. Il est question d'un processus continu de partage et de communication de réponses aux points de doute et d'insécurité tel que l'interrogation portant sur la confusion liée à la distinction entre les termes complexes et les collocations. (Exemples : stratégie efficace et stratégie politique ; guerre souterraine et guerre totale ; différends internationaux, différends internes ; diplomatie préventive ; diplomatie parlementaire ; etc.)

Lors des séances, l'enseignant explore des sites de référence, donne des exemples de projets et met à la disposition des apprenants des ressources ouvertes et des stratégies de collaboration expliquées par Nissen [32]. Par exemple, lorsqu'une apprenante présente une partie de son projet aux autres apprenantes (les objectifs, le domaine et les étapes de sa démarche), nous traitons des questions résultant de l'étape de dépouillement et des hésitations vis-à-vis de certains choix des candidats-termes. Ensuite, chacune assume sa charge du travail dans le cadre des délais et des missions du projet.

B. Susciter la motivation

L'usage des extracteurs peut, dans certaines tâches, constituer des difficultés à trouver une information dans des fichiers ; plutôt qu'à négocier le sens des termes en contexte, l'information est modélisée à travers les résultats des outils informatiques. Face aux difficultés de la prise en main de l'outil, il importe de « gagner la confiance des étudiants en les aidant à retrouver la leur ». [27] Les dites-difficultés sont surmontables au moyen de communication par les réseaux sociaux ; les scénarios pédagogiques réalisables évoluent en ligne et les possibilités de collaboration des étudiants augmentent. Les relations enseignant/apprenant sont rendues plus conviviales, ce qui permet aux enseignants de garder la maîtrise du processus : leur rôle de facilitateur de connaissances et de médiateur des

fonctionnalités des logiciels. L'enseignant explique, renvoie au tutoriel, procède aux extractions afin de démontrer les options et puis guide les opérations personnelles pour intervenir en cas de besoin. Les démarches supposent une connaissance des exigences du travail personnel et des obstacles de l'individualisation des tâches de découverte [33]. Encourager et suivre de près nécessitent des capacités d'écoute et de réactions positives de l'enseignant durant la formation.

C. Vérifier les progressions et mesurer les qualités des résultats

Le suivi des apprenantes consiste à s'assurer que les résultats du dépouillement automatique sont valables et existent dans un corpus de référence constitue une étape importante dans le contrôle de qualité des données terminologiques recueillies ? Ce contrôle se fera à quel rythme ? Les candidats-termes seront sélectionnés selon quelles règles ? La limite minimale de fréquence requise pour inclure ou écarter le terme reste un des points réguliers qui reviennent dans les cours lors de la phase de dépouillement dans les projets des apprenantes.

A cet égard, nous présentons des exemples de situations dans lesquelles l'apprenante pose des questions précises pendant la phase de dépouillement et demande de l'aide pour continuer. Une apprenante me contacte pour l'aider à trouver la raison pour la non-reconnaissance des lettres en arabe et m'envoie une prise d'écran de son ordinateur puis je propose de changer les noms des fichiers et de les écrire en lettres latines. Enfin, elle m'envoie la prise d'écran des listes de termes formés à partir du terme-vedette الطاقة (Energie) pour s'assurer de son cheminement. Elle est arrivée à repérer la liste des termes complexes exposés dans la figure no. 13 et leurs contextes. Un besoin de vérification de la progression se fait en répartissant les tâches sur l'axe du temps, par semaine, par objets délivrables ou par résultats définis.

Rank	Freq	Range	Cluster
1	33	1	الطاقة المتجددة
2	16	1	الطاقة الشمسية
3	10	1	الطاقة البديلة
4	7	1	الطاقة في
5	5	1	الطاقة الكهربائية
6	3	1	الطاقة التقليدية
7	3	1	الطاقة المحلية
8	3	1	الطاقة من
9	2	1	الطاقة إلى
10	2	1	الطاقة الأولية
11	2	1	الطاقة المائية
12	2	1	الطاقة الهوائية
13	2	1	الطاقة لأغراض

Fig. 13 Capture d'écran des termes complexes affichés lors de la recherche à partir du terme simple الطاقة

Les expressions suivantes sont formulées par les apprenantes pour indiquer les difficultés rencontrées ; nous les avons reçues soit sur le groupe de what's App consacré au cours, soit en message privé :

« Bonjour Madame, j'ai du mal à trouver la structure actancielle trouver. La structure actancielle dans le domaine des énergies renouvelables et du développement durable. »

« Est-ce que dans le domaine des énergies renouvelables, je peux utiliser le verbe produire comme entrée ? Est-ce que cela, c'est la structure actantielle, n'est-ce pas, Madame ? » »

« Quelques-unes d'entre nous confrontent un problème technique avec le logiciel. »

« Voici ma fiche divisée sous forme de Tableau. Est-ce que c'est lisible ? C'est le format que vous avez demandé ? Ça va ? »

« C'est difficile pour moi d'identifier les collocations, est-ce qu'il y a une méthode facile de les trouver ? Est-ce que vous pouvez vérifier ces collocations s'ils sont corrects ou non ?

Défendre : défendre activement

Drogue : trafic de drogues

Victimes : victimes tués »

« Bonjour, j'espère que vs allez bien. J'ai une question. [Elle m'envoie une liste des collocations (lutte informatique, secours informatique, gestion informatique) et de termes associés (informatique sécurisée, informatique en nuage)]

Est-ce que c'est juste ou bien je suis confuse ? »

Ces expressions indiquent des besoins de confirmation du bon cheminement, de sécurisation face à l'interface et une demande d'aide en ce qui concerne les notions relatives aux choix entre collocations et termes associés, à la modélisation nouvellement acquise en structures actanciennes, précision des requêtes à partir de dépouillement par les N-Gram.

Une fois le dépouillement fait en français, nous passons aux corpus arabes et les problèmes liés à la reconnaissance des lettres arabes surgissent. Mais enfin, les apprenantes traduisent les rubriques en négociant les équivalents possibles durant le cours, ensuite les structures actanciennes sont appliquées pour les termes arabes.

6 ANALYSE DES CONCLUSIONS DES APPRENANTES ET LEUR SENS CRITIQUE

Comprendre la valeur ajoutée des applications informatiques en TAL - en rapport avec les données massives- ne passe pas sans la reconnaissance des difficultés rencontrées par les apprenantes dont le manque de ressources numériques arabes. Cela se situe au début de la recherche, toutefois lors des sélections des champs disciplinaires pour les projets individuels, nous avons présenté les plateformes qui supportent la langue arabe, qui sont parfois des bases textuelles accessibles et interrogeables en ligne ou des ressources sur des sites de vulgarisation scientifique. La mutualisation des ressources ouvertes semble d'une grande utilité.

Toutefois, les apprenantes ont explicité les obstacles rencontrés au début des projets, faute de pratique, ce qui s'améliore petit à petit. Sinon, au moins les notions de données quantitatives, fréquence, concordanciers ont été acquises à partir de recherche sur corpus bilingues comparables.

« Pour conclure, nous avons donc exploiter les avantage d AntConc en faisant notre projet. Au début, nous avons souffert de quelques difficultés en l'utilisant spécifiquement dans le cas des corpus arabes mais après cela, tout était bien. De même, nous avons appris des néologismes dans le domaine de la cybercriminalité, des collocations, des équivalents dans l'autre langue. AntConc était un logiciel très efficace dans l'extraction des donnees, les N-grams et les concordanciers. » (Apprenante 6)

L'intérêt de la compilation et de l'extraction quantitative c'est sa comparaison avec la vérification des données recueillies dans un corpus de référence afin de gérer la qualité des candidats-termes retenus. Les recherches des clusters/N-Grams permettent l'accès aux groupes syntagmatiques de 2 mots - au moins - les plus fréquents contenant le mot de la requête. La taille des n-grammes varie (Min. Size & Max. Size) ainsi que la fréquence minimale (Min. Cluster Frequency) et la position du mot de la requête (Search Term Position).

La pratique et les tâches sont conditionnées par les réalités des applications informatiques, en d'autres termes il s'agit de réorganiser les contenus des programmes en fonction des nouvelles tendances de l'approche communicative de la terminologie et de se baser sur les corpus vs dictionnaires. Autrefois centré sur la normalisation, le travail actuel s'inscrit dans une optique de terminologie textuelle rendue efficace à la suite de l'automatisation des modèles de fouille. Le corrélat immédiat de ces nouvelles approches selon laquelle le terme est considéré comme « une unité de

discours », selon Cabré [34]. La dimension communicationnelle et discursive justifiait une approche basée sur corpus (Data Driven learning).

L'acquisition des notions relatives aux recherches terminologiques repose sur l'application d'une méthodologie visant la collecte de la documentation, le recueil des données qualitatives (la qualité et fiabilité des ressources, la représentation du domaine, le niveau de spécialité, etc.), le repérage ou dépouillement, vérification, formalisation des définitions actanciennes puis choix des contextes et des équivalents.

La perception des apprenantes vis-à-vis de l'emploi et de l'évaluation de l'outil et de leurs travaux constituent des éléments fort intéressants : à notre avis, leurs progrès est l'indice principal de l'efficacité de la formation. L'évaluation de l'outil se base sur l'examen des résultats fournis aux requêtes soumises au logiciel ; au début, le groupe teste les recherches à partir d'une liste de mots-clés, puis postérieurement aura lieu l'observation de toutes les occurrences du terme ainsi que l'extraction des collocations.

7 CONCLUSION

Les apprenantes ont saisi l'apport des analyses quantitatives et la valeur des fréquences des mots dans un corpus donné, dans ce cas le corpus personnel compilé de façon individuelle. Au niveau de la linguistique informatique, la modélisation de la connaissance linguistique porte sur les analyses des mots afin d'en trouver des régularités et décrire certains phénomènes linguistiques (collocations). Le module a permis de détecter à partir des N-grammes les segments répétés. L'approche réflexive compare les résultats du système aux données du corpus de référence et vérifie la liste des candidats-termes.

Sur le plan de l'analyse morphosyntaxique, les apprenantes peuvent classer les termes recueillis selon leur formation. A travers les occurrences des termes, les apprenantes se sont entraînées à rédiger les définitions selon des structures actanciennes, mais la mise en forme des structures actanciennes n'était pas facile ; cela nécessite plus de temps et de travail afin de se familiariser et d'extraire plus facilement les données du corpus.

Le recensement des erreurs commises par les apprenantes et leurs commentaires font apparaître la difficulté liée à l'analyse automatique des textes : la recherche de documents ou du corpus d'étude, la phase de préparation, la sélection des candidat-termes, les critères de classification des données et puis leur représentation au moyen des fiches. Une grille d'analyse contrôlée des listes de termes et des collocations pour recenser la catégorie grammaticale des termes complexes, le fonctionnement des termes dans le discours, la sélection du contexte. Dans une approche terminologique, le choix des termes associés a posé certains problèmes.

Des interrogations ont été formulées à propos de l'élaboration de critères linguistiques de classification des contextes afin de choisir ceux à introduire dans le fichier terminologique. Tout au long de la finalisation des projets individuels, il faut accompagner les apprenants pour répondre à leurs hésitations et trancher au niveau des prises de décisions, ce qui mène à un plus haut degré d'autonomie et une simulation des situations de travail réelles et des défis de traitement des données textuelles.

La répartition du travail sur l'axe du temps et par catégorie de tâches selon le domaine et les différences de classement thématique liées à la nature de chaque domaine. Les compétences de littéracie informatique varient d'une personne à une autre. La différence significative relève du domaine choisi : la présence ou non de documents en langue arabe, la fiabilité des ressources, la qualité de la documentation, le caractère néologique de certains termes nourrissent les soucis et les doutes parfois justifiés et récurrents durant la période de dépouillement.

Autonomiser implique une démarche de nature motivante, patiente et attentive de la part de l'enseignant-formateur, et un travail sérieux, minutieux et progressif de la part des apprenantes. A la fin du module, les projets terminologiques ont été élaborés et évalués selon une grille déterminée en collaboration de tous les groupes. D'autres données théoriques ont été fournies par l'enseignante comme pistes de consolidation des opérations de repérage sur corpus. Au moment de l'évaluation des fiches des ressources ou dictionnaires terminologiques disponibles en ligne, des remarques spontanément formulées par les apprenants méritaient des précisions enrichies d'exemples d'autres ressources, d'où l'aller-retour entre corpus et références précédentes ; la mise à jour et la comparaison des options de navigation électronique exige des études plus approfondies : quand enrichir les fiches, comment inclure les notes explicatives ?

Quelle est la place accordée aux régionalismes dénommatifs dans certains domaines ? Tant de questions qui dépassaient l'horizon de la présente étude et qui peuvent susciter des réflexions en rapport avec la réalité du travail terminologique.

RÉFÉRENCES

- [1] L'Homme M.- C., *La terminologie : Principes et Techniques*, Montréal, Presses Universitaires de Montréal, 2004.
- [2] D. Gouadec, *Terminologie : constitution des données*, Paris, AFNOR, 1990.
- [3] G. Otman, "*Les représentations sémantiques en terminologie : la modélisation des unités terminologiques sous la forme de réseaux sémantico-terminologiques*", Thèse de doctorat, Université Paris IV, 1995.
- [4] F. Vezzani, *Terminologie numérique : conception, représentation et gestion*, Bern, Berlin, Bruxelles, New York, Oxford, Peter Lang, série Linguistic Insights, vol. 290, 2022. <https://library.oapen.org/handle/20.500.12657/56656> (consulté le 15 octobre 2022).
- [5] C. Boisson, et P. Thoiron, *Autour de la dénomination*, Lyon, Presses universitaires de Lyon, 1997.
- [6] L. Depecker, "Contribution de la terminologie à la linguistique", *Langages*, no.157, pp.6-13, 2005.
- [7] P. Gardy, "*L'évaluation en didactique de la traduction et l'intégration des outils technopédagogiques : étude qualitative et expérimentation*", thèse de doctorat de l'Université Laval, 2015.
- [8] J. Humbley, "Présentation. Terminologie : orientations actuelles", *Revue française de linguistique appliquée*, XIV, pp. 5-8, 2009, <https://doi.org/10.3917/rfla.142.0005>, (consulté le 23 octobre 2022).
- [9] Wüster, E. *Dictionnaire multilingue de la machine-outil : notions fondamentales, définies et illustrées*, Technical Press, Londres, 1968.
- [10] M.T. Cabré, *La terminologie. Théorie, méthode et applications*, Paris, Armand Colin, 1998.
- [11] F. Gaudin, *Pour une socioterminologie : des problèmes sémantiques aux pratiques institutionnelles*, Rouen, Publications de l'Université de Rouen, 1993.
- [12] P. Perrenoud, "*Construire un référentiel de compétences pour guider une formation professionnelle*", Université de Genève, Faculté de psychologie et des sciences de l'éducation, 2001, http://www.unige.ch/fapse/SSE/teachers/perrenoud/php_main/php_2001/2001_33.html, (consulté le 15 septembre 2022).
- [13] EMT competence framework disponible sur le site https://ec.europa.eu/info/sites/default/files/emt_competence_fwk_2017_en_web.pdf, (consulté le 15 avril 2021)
- [14] J.-C. Corbeil, "Les marques d'usage comme technique de description des aspects connotatifs du lexique", contribution à la table ronde sur les marques lexicographiques, Montréal, 3 et 4 novembre 1994 in *Les marques lexicographiques en contexte québécois*, Office de la langue française, coll. « Études, Recherches et Documentation », Gouvernement du Québec, pp. 29-50, 1998, disponible sur <http://corbeil.recherche.usherbrooke.ca/document-corbeil-1994-a>, (consulté le 03 mars 2021).
- [15] GDT (Grand dictionnaire terminologique) disponible sur <https://gdt.oqlf.gouv.qc.ca/>, (consulté le 10 mai 2022)
- [16] Les recommandations terminologiques produites par le Réseau de Traducteurs et Traductrices en Education (RTE Canada) 2010, https://termino.rte-nte.ca/fiche_pdf.php?fld=242, (consulté le 03 avril 2021).
- [17] S. Frenkel, "Composantes métacognitives ; définitions et outils d'évaluation", *Enfance*, no. 4, pp. 427-457, 2014. <https://doi.org/10.3917/enf1.144.0427>, (consulté le 19 mai 2022)
- [18] F. Raynal, et A. Rieunier, *Pédagogie, dictionnaire des concepts clés. Apprentissage, formation, psychologie cognitive*, Issy-les-Moulineaux, ESF éditeur, 2012.
- [19] B. Bommier-Pincemin, "Caractérisation d'un texte dans un corpus : du quantitatif vers le qualitatif. Définir un corpus", in "*Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*", Thèse de Doctorat en Linguistique, Université Paris IV Sorbonne, 6 avril 1999, chapitre VII, pp. 415-427, 1999.
- [20] PISTES : projets interdisciplinaires, site réalisé par le Centre de services et de ressources en technopédagogie, une collaboration de l'Université Laval, Faculté des sciences de l'éducation, *L'approche par problèmes*, https://www.pistes.fse.ulaval.ca/sae/?onglet=contenu&no_version=1952, (consulté le 01 septembre 2021)
- [21] ISTE, International Society for Technology in Education Standard (2008), disponible sur <https://www.iste.org/standards/iste-standards-for-teachers>, (consulté le 03 mai 2021).
- [22] L. Anthony (2007), *AntConc 3.4.1w*. Concordancier téléchargeable sur http://www.antlab.sci.waseda.ac.jp/antconc_index.html (consulté le 11 mars 2022).
- [23] M.T. Cabré, "Terminologie et dictionnaires", *Meta*, vol. 39, no. 4, 1994, pp. 589-597.

- [24] M. Manes-Gallo et C. Paganelli, " La recherche d'information assistée par ordinateur : quelle représentation des connaissances ?", *Les Enjeux de l'information et de la communication*, vol. 2003 (1), pp. 36-51, GDT, 2003. <https://doi.org/10.3917/enic.003.0036>, (consulté le 23 octobre 2022).
- [25] DicoInfo (Dictionnaire de l'Informatique) disponible sur http://olst.ling.umontreal.ca/?page_id=356 , (consulté le 10 mai 2022)
- [26] Institut Grasset, *L'ergonomie informatique c'est quoi exactement ?*, disponible sur <https://www.institut-grasset.qc.ca/batiment/ergonomie-informatique-cest-quoi-exactement/>, (consulté le 10 octobre 2022)
- [27] Seus-Walker, K. " Pour développer l'autonomie des apprenants", *Cahiers de l'APLIUT*, Vol. XXII N° 2 | 2003, 43-58, consulté le 15/11/2022. <http://journals.openedition.org/apliut/3715>(consulté le 01 septembre 2022).
- [28] A. Polguère, *Lexicologie et sémantique lexicale. Notions fondamentales*, Montréal, Presses de l'Université de Montréal, 2003.
- [29] F. Hmida, *Identification et exploitation de contextes riches en connaissances pour l'aide à la traduction terminologique*, Université de Nantes, 2017, <https://tel.archives-ouvertes.fr/tel-01725324>, (consulté le 01 septembre 2022).
- [30] L'Homme, M.-C. et Meynard, I., "Le point d'accès aux combinaisons lexicales spécialisées : présentation de deux modèles informatiques ", *TTR*, vol. 1, 1no.1, pp.199–227, <https://doi.org/10.7202/037322ar>, (consulté le 11 novembre 2021).
- [31] I. Mel'čuk, A. Clas et A. Polguère, *Introduction à la lexicologie explicative et combinatoire*, Bruxelles, Duculot, 1995.
- [32] E. Nissen, "Quelles aides les formations hybrides en langues proposent-elles à l'apprenant pour favoriser son autonomie ? ", *Apprentissage des Langues et Systèmes d'Information et de Communication (Alsic)*, vol. 10 no.1, pp.129-144, 2007. <http://alsic.revues.org/617>, (consulté le 23 octobre 2022).
- [33] Zhong, Y. (2008). Teaching translators through self-directed learning, *The Interpreter and Translator Trainer*, vol. 2, no.2, pp. 203-220, 2008. <https://doi.org/10.1080/1750399x.2008.10798774> , (consulté le 15 octobre 2022).
- [34] M.T. Cabré, "Technologie et terminologie : changements méthodologiques et épistémologiques", *Cahiers de lexicologie*, no. 94, 2009-1, pp. 31-56, 2009.

BIOGRAPHIE



Samah Nasr est professeur- adjoint en linguistique, Département de langue et de littérature françaises, Faculté de Jeunes Filles pour les arts, les sciences et l'éducation, Université Ain Shams. Titulaire du diplôme de Doctorat en 2009 : « Étude d'un dictionnaire spécialisé dans le domaine de la mercatique ». Titulaire du diplôme de Magistère en langue française en 2005 : « La terminologie de l'informatique ». Titulaire du Master 2 en Lexicologie, Terminologie Multilingues et Traduction de l'Université Lyon II et le diplôme de Master 1 de FLE de l'Université Stendhal-Grenoble III en 1999. Elle a publié plusieurs articles sur les ressources électroniques et les besoins des traducteurs, sur les dictionnaires informatisés et sur la post-édition dans la formation des traducteurs.

ENGLISH ABSTRACT

Automatic Terminology Extraction: a Student-Centered Approach

Samah Nasr

Department of French, Faculty of Women for Arts, Sciences and Education, Ain Shams University, Egypt

samah.hassan@women.asu.edu.eg

Abstract: *This study presents an experience of extracting terminology from corpus compiled by university learners in order to explore resource materials that include specialized texts in both Arabic and French and to show the impact of the integration of automatic retrieval tools into the university curriculum based on a teaching experience in a module on terminotics. We describe the terminology extraction results and the method of representation of lexical data. We also discussed the difficulties related to the selection of candidate terms, the modelling of the combinatory structure of terms and the choice of associated terms and collocations. We presented some solutions proposed to model the results obtained.*

Keywords: *Terminology, automatic extraction, student's autonomy- project based learning.*

ARABIC ABSTRACT

الاستخراج الآلي للمصطلحات: منهجية التركيز على الطالب

سامح نصر

قسم اللغة الفرنسية وآدابها، كلية البنات للآداب والعلوم والتربية، جامعة عين شمس، مصر

samah.hassan@women.asu.edu.eg

ملخص

تقدم هذه الدراسة تجربة في استخراج المصطلحات بصورة آلية من مجموعة الذخائر اللغوية التي يجمعها الطلاب الجامعيون من أجل تكوين مسارد للمصطلحات المستخرجة من نصوص متخصصة باللغتين العربية والفرنسية، وتهدف إلى توضيح أثر إدماج أدوات الاستخراج الحاسوبي في المناهج الدراسية الجامعية استناداً إلى خبرة التدريس في مقرر جامعي عن المصطلحات لبيان نتائج استخراج المصطلحات وطريقة تمثيل البيانات المعجمية. ناقش البحث أيضاً الصعوبات المتعلقة باختيار المصطلحات المقترحة من الآلة، ووصف البنية التركيبية للمصطلحات واختيار المصطلحات ذات الصلة. كما قدمنا بعض الحلول المقترحة لنمذجة النتائج التي تم الحصول عليها.

الكلمات المفتاحية

علم المصطلح الحاسوبي، نموذج حاسوبي، استقلالية المتعلم، التعلم القائم على المشاريع.