



Extraction des termes sur la COVID-19 et leurs emplois sémantico-syntaxiques à partir d'un corpus spécialisé

Ma'moun Alshtaiwi

Université du Yarmouk, Jordanie

m.alshtaiwi@yu.edu.jo

Reçu le 20-05-2020 / Évalué le 04-10-2020 / Accepté le 30-10-2020

Résumé

Dans cette étude, un corpus électronique de documentations issues de l'Organisation mondiale de la santé (OMS) traitant de la situation épidémique de COVID-19 est constitué afin d'en extraire la terminologie existante (formation morpho-syntaxique) et l'emploi des termes. Cette analyse intègre particulièrement le verbe terminologique ainsi que ses arguments sémantico-syntaxiques pour pouvoir définir son sens spécialisé. Le corpus sert à analyser non seulement les unités terminologiques existantes et ses occurrences, mais aussi l'environnement contextuel de l'unité, c'est-à-dire l'emploi des termes en contexte ayant pour but d'enrichir les dictionnaires électroniques et les bases de données ciblant l'utilisation et l'emploi de termes pour une thématique d'une importance majeure telle que la pandémie de COVID-19.

Mots-clés : terminologie, extraction semi-automatique, emploi, COVID-19

COVID-19 terimlerinin bulunması ve özel bir derlemede anlamsal-sözdizimsel kullanımları

Özet

Bu çalışma, mevcut biçimsel-sözdizimsel oluşum terminolojisini ve terimlerin kullanımını ortaya çıkarmak için Dünya Sağlık Örgütü (DSÖ)'den COVID-19 salgınıyla ilgili elektronik bir dokümantasyon derlemesinden oluşmaktadır. Bu analiz, özellikle terminolojik fiili ve onun anlamsal-sözdizimsel argümanlarının özel anlamını tanımlayabilmeyi içermektedir. Bu derleme, COVID-19 gibi büyük bir öneme sahip olan konuyla ilgili terimlerin kullanımını hedefleyen elektronik sözlükleri ve veri tabanlarını zenginleştirmek amacıyla sadece terminolojik birimleri ve bunların oluşumlarını değil aynı zamanda da birimin bağlamsal durumunu, yani bağlam içinde kullanımını analiz etmeye yardımcı olacaktır.

Anahtar sözcükler: terminoloji, yarı otomatik çıkarma, kullanım, COVID-19

Extraction of COVID-19 terms and their uses semantico-syntactic from a specialized corpus

Abstract

In this study, an electronic corpus of documentation from WHO dealing with the COVID-19 epidemic situation will be assembled in order to extract the existing terminology (morpho-syntactic formation) and the use of terms. This analysis particularly integrates the terminological verb as well as its semantic-syntactic arguments in order to be able to define its specialized meaning. The corpus is used to analyze not only the existing terminological units and their occurrences, but also the contextual environment of the unit, i.e. the use of terms in context with the aim of enriching electronic dictionaries and databases targeting the use and use of terms for a topic of major importance such as the COVID-19 pandemic.

Keywords: terminology, semi-automatic extraction, use, COVID-19

Introduction

Le terme COVID-19¹ (dont l'équivalent anglais est coronavirus disease 2019) est la dénomination officielle retenue en février 2020 par l'Organisation mondiale de la santé (OMS). Lors de l'apparition de cette maladie, avant même que le virus ne soit identifié, les médias, notamment, utilisaient plusieurs unités linguistiques pour exprimer le concept comme le sigle «2019 nCov» ou «la pneumonie de Wuhan» qui était la traduction du terme qui était alors employé en Chine. Ce dernier, jugé inadéquat, a graduellement été remplacé dans l'usage par les appellations officielles de l'OMS, à savoir, «COVID-19» et «maladie à coronavirus 2019» (*le Grand dictionnaire terminologique, GDT*). Le contenu de ces termes récents constitue les concepts qui forment une structure, et dans laquelle ils occupent une place déterminée dans le domaine de l'épidémiologie² : ces concepts gardent des relations avec tous les autres concepts de la structure. En ce sens, l'objectif de notre travail est de décrire formellement et fonctionnellement les termes relevés sur le corpus, d'expliquer leur activation dans le langage et leurs rapports avec les autres éléments du discours, au sein d'un même système, en vue de faire avancer la communication spécialisée et les unités étudiées. Aujourd'hui, la compilation et l'examen des unités terminologiques y compris les verbes terminologiques ont diverses applications : dictionnaires électroniques, glossaires, bases de données terminologiques, systèmes de veilles, traduction, etc.

1. Caractéristiques et niveau de spécialisation du corpus

L'analyse de l'environnement linguistique de la terminologie dans le cadre d'une langue de spécialité nécessite une méthodologie claire et efficace pour l'extraction des termes y compris le verbe terminologique (L'Homme, 2012) et l'identification du contexte. Afin d'analyser l'environnement textuel des termes simples et complexes de la situation épidémiologique de COVID-19, nous avons constitué un corpus grâce aux documentations de l'OMS. Les textes choisis sont destinés principalement au grand public car ils ont pour but de l'informer de l'évolution de la COVID-19. Le matériau textuel peut être arrangé différemment selon les niveaux de spécialisation. Or, la spécialisation est souvent définie en fonction de l'auteur du texte et des destinataires. Les textes portant sur la COVID-19 sont en effet homogènes à caractères normatifs, mais organisés en plusieurs niveaux, en fonction des critères de classifications qui sont pris en considération. Pearson en identifie quatre :

« d'expert à expert (par exemple dans un article tiré d'une revue scientifique) ; d'expert à expert d'un domaine connexe (exemple : un médecin à des infirmiers, un technicien à un ingénieur) ; didactique (texte s'adressant à des spécialistes en devenir) ; de vulgarisation [...] » (Pearson, 1998 : 39).

Aussi, les textes sur la COVID-19 peuvent être divisés en trois catégories.

Textes très spécialisés (ces textes à des fins scientifiques sont : ensemble ou partie des ouvrages et des articles médicale portant sur le sujet).

Textes spécialisés (exemple, textes à caractères normatifs) : comme ceux qui appartiennent à l'OMS. Ils sont homogènes en matière textuelle et ils ne comportent pas de textes de genre narratif ni argumentatif.

Textes de vulgarisation ou de semi-vulgarisation : ensemble ou partie des articles de presse généraliste, magazines et brochures portant sur le sujet traité, etc. Ce type de discours relève d'articles tirés de journaux généraux et de sites web non spécialisés.

Les textes traités appartiennent principalement à la catégorie de textes spécialisés et de vulgarisation, c'est-à-dire des textes à caractères normatifs écrits par un expert ou non expert qui s'adresse à une personne ne possédant pas *à priori* les connaissances abordées dans ceux-ci.

Tous les textes rassemblés sont principalement homogènes, ils sont réunis en corpus et ont la forme de rapports d'activité. Cette étude se repose donc sur du seul contexte définitoire portant sur des documentations de l'OMS. Dans la pratique de ce travail, l'exhaustivité du corpus n'est pas possible à atteindre (Biber, 1993).

Parmi les textes sur lesquels nous travaillons, tous ne présentent pas le même degré de spécialisation et de hiérarchisation. Il arrive que d'autres discours interviennent dans un texte. Le corpus n'est pas exhaustif, mais réunit une part importante des événements relatifs à la COVID-19 du 1^{er} mars au 30 octobre 2020.

Selon nous, dans le traitement du corpus, il faudrait en outre prendre en compte d'autres critères de sélections.

1. *La thématique abordée, son historique et son développement* : la pandémie de COVID-19 est très récente et renvoie à une réalité nouvelle et les médias en traitent massivement tous les jours. Dans la situation de communication présente de la COVID-19, les auteurs s'échangent des informations sur une réalité nouvelle selon différents degrés de scientificité et de technicité. Dans le corpus, nous pouvons distinguer des textes traitant de la COVID-19 et d'autres qui abordent des thèmes connexes, c'est pourquoi nous avons sélectionné certaines parties pertinentes et écartons d'autres.
2. *Les sources scientifiques* : les décisions de l'OMS jouent un rôle essentiel dans le choix de la terminologie.
3. *Le rapport entre les langues* : la situation traitée renvoie à une pandémie mondiale et devrait être traitée dans beaucoup de langues. La majorité de nos textes français sont traduits de l'anglais.

2. Constitution et traitement du corpus sur la COVID-19

Nous avons essayé de constituer un corpus équilibré qui est un ensemble de documents homogènes, cohérents, similaires. Notre corpus ouvert peut être développé et mis à jour particulièrement pendant la pandémie de COVID-19 qui progresse rapidement. Le corpus est utilisé pour extraire les termes ainsi que leurs contextes et se rapporte directement à l'étude terminologique. Il renvoie exclusivement à un ensemble de documents qui concernent une thématique particulière (la COVID-19). Généralement, en ce qui concerne le corpus en terminologie, Cabré note que les textes spécialisés sont définis comme :

« les situations professionnelles par les interlocuteurs qui interagissent, par le sujet évoqué qui relève du domaine ou des domaines concernés par la profession, et par la finalité essentielle de rechercher l'information auprès du récepteur, bien que pour ce faire on utilise des stratégies discursives différentes ». (Cabré, 2008 : 38).

Notre corpus français est traduit de l'anglais, c'est pourquoi nous avons décidé de constituer un corpus parallèle comprenant des textes avec leurs traductions.

Afin d'aligner les deux termes parallèlement, nous avons utilisé l'outil *YouAlign*² en ligne permettant d'organiser deux textes traduits au même niveau.

Le corpus de textes spécialisés représente ainsi une collection de données langagières qui sont sélectionnées pour l'extraction des termes et l'analyse linguistique de l'environnement contextuel du terme (Condamines, 1993, 2003, 2005). Ces données du corpus sont organisées selon des critères linguistiques (morpho-syntaxique, sémantique et discursif) et extralinguistiques explicites pour servir d'échantillons d'emploi d'une langue.

Nous avons utilisé l'outil *TermoStat Web* et sa méthodologie (Drouin, 2003) pour extraire les termes simples (nom, verbe, adjectif, adverbe) et complexes. Cet outil permet d'identifier les contextes dans lesquels ils sont utilisés et nous donne des informations utiles sur la distribution syntaxique de chaque terme qu'il soit un nom, un verbe, un adjectif ou un adverbe. L'outil *TermoStat Web* permet de situer le contexte dans lequel le terme est utilisé : il est en effet possible de prendre connaissance des fréquences, des occurrences, ou encore des phraséologismes. De plus, il permet d'en reconnaître les schémas textuels de chacun des termes. Mais cette phase nécessite aussi la révision et l'analyse humaines afin que soient assurées la validité et la précision. Cette intervention humaine révèle les limites de l'automatisation, qui constitue néanmoins une avancée considérable dans la rapidité de traitement. Au cours du traitement, à chaque lecture d'un document, nous avons recensé sur une fiche les termes spécifiques au domaine. Ainsi, une partie de nos textes du corpus sont utilisés sous forme brute dans le cadre de logiciels de traitement de l'information. Il paraît logique de tenir compte des ressources et des outils de traitement semi-automatique de langues. Notons que *TermoStat Web* implique le marquage structural et linguistique, ce logiciel ayant un dictionnaire initial de traitement de la terminologie, il possède un système d'analyse morphologique, un système de lemmatisation et un système de structuration syntaxique.

L'étude des documents qui sont issus d'une organisation mondiale montre qu'il existe une terminologie relativement homogène, malgré le maintien de certaines spécificités au sein des institutions productrices de ces documents. Dans ce corpus, les termes se ressemblent au niveau des structures conceptuelles observées. Cette méthodologie pourrait être un argument en faveur de l'utilisation d'une terminologie cohérente.

3. Structuration conceptuelle

Le point de départ de notre étude était une structuration conceptuelle, fondée sur une documentation rédigée par les spécialistes et conçue selon le découpage scientifique qui est le leur. Cette structuration conceptuelle nous permet de nous

consacrer exclusivement à la terminologie du domaine (Dahlberg, 1981 ; Cabré, 1998), car il est inutile d'alourdir la liste terminologique avec des termes superflus. Le texte spécialisé se constitue de deux dimensions, à savoir la dimension conceptuelle d'une part et la dimension linguistique d'autre part. Pour représenter et extraire les concepts et leurs classes conceptuelles relatifs à la pandémie de COVID-19, nous avons analysé la littérature médicale relative à la COVID-19 dans l'encyclopédie médicale sur le site Internet suivant : *Vocabulaire-medicale.fr*.

Il est important de définir les trois classes « endémie », « épidémie » et « pandémie » appartenant au vocabulaire de « l'épidémiologie ». Ils ont leur équivalent dans le domaine des maladies animales, parfois transmissibles à l'homme. Ces quatre concepts étaient initialement dédiés aux maladies infectieuses. Le terme « maladie infectieuse » exprime « toute maladie causée par des agents microbiologiques qui peut être transmise d'une personne à une autre » (GDT) qu'il s'agisse d'une bactérie, d'un virus, d'un parasite ou encore d'un protozoaire. L'épidémiologie renvoie à une des disciplines sur lesquelles s'appuient la santé publique et la médecine préventive. Donc, l'épidémiologiste « effectue notamment des recherches visant à mieux comprendre et à maîtriser les facteurs qui influencent la fréquence et la distribution de différents problèmes de santé dans une population donnée » (GDT). Notons qu'il existe deux termes attachés à l'épidémiologie : « épidémiologie descriptive » et « épidémiologie évaluative ». Le terme « endémie » se définit par la présence habituelle d'une maladie, en général infectieuse, dans une population déterminée ou une région précise, avec une incidence stable. Une épidémie se définit par la croissance rapide de l'incidence d'une maladie dans une région donnée et pendant une période.

Initialement, le terme d'« épidémie » ne concernait que les maladies infectieuses, comme la variole autrefois et la grippe tous les ans depuis fort longtemps. Depuis le début du XXI^e siècle on assiste à l'émergence d'épidémies dues à un nouveau type de virus, les « coronavirus ». L'adjectif épidémique a deux contraires : « endémique », que nous venons de voir, et « sporadique », qui signifie occasionnel, ou au cas par cas.

D'autres concepts sont relatifs à la pandémie de COVID-19. Nous analysons les termes suivants : « contamination », « contagion », « contagiosité » et « période d'incubation ». Nous avons consulté le GDT pour chercher les définitions de ces quatre termes qui sont relatifs aux champs étudiés.

La contamination : « envahissement d'un être vivant ou d'un objet par des germes pathogènes ou des contagés » (GDT).

La contagion : « transmission d'une maladie infectieuse d'un sujet malade à un sujet sain, par un contact direct ou indirect » (GDT). Quand la transmission est indirecte, par exemple par le biais d'un moustique, comme dans le cas du paludisme, il ne s'agit pas d'une maladie contagieuse.

La contagiosité : « le terme contagiosité est utilisé surtout dans des contextes de maladies infectieuses (virales, microbiennes) animales et humaines. C'est le caractère d'un agent infectieux pathogène qui lui permet de se transmettre, d'un individu à un autre, par la contagion » (GDT). Le patient contaminé peut en contaminer un nombre plus ou moins grand selon le type de l'agent pathogène. Par exemple il est admis qu'un patient contaminé par le virus de la grippe va, en moyenne, contaminer une seule personne, ce qui est une contagiosité faible, par opposition au virus de la rougeole, à forte contagiosité, un individu porteur du virus pouvant contaminer jusqu'à neuf personnes. (*V-Médical*)

La période d'incubation : « Intervalle qui sépare l'infection de l'apparition du premier signe ou du premier symptôme de la maladie » (GDT). On notera qu'il n'est pas toujours facile de connaître avec précision la date de contamination. Par défaut, on fera débiter la période d'incubation du jour où le test diagnostique revient positif. Deux exemples : la période d'incubation de la grippe est de 1 à 3 jours ; celle de la Covid-19 de 2 à 14 jours. Cette période d'incubation indique la durée de la période de confinement quand elle est imposée : 14 jours pour la Covid-19. (*V-médical*).

Dans le cadre de ce classement conceptuel, nous procéderons à la gestion de la terminologie. Notons que le recours à l'analyse de différents textes permet de dégager, modéliser et étudier les connaissances afin de mieux situer les concepts et élaborer des arborescences conceptuelles qui désignent le domaine concerné. Dans cette perspective pratique, les catégories conceptuelles ci-dessus facilitent et favorisent le dépouillement et le traitement de la terminologie pour des dictionnaires, glossaires ou thésaurus. Cet objectif méthodologique s'intéresse au traitement de la terminologie qui s'appuie sur la représentation du système conceptuel d'après les documentations constituées dans le cadre du domaine.

Le domaine de l'épidémiologie comprend ses propres termes ainsi que des termes venant de plusieurs autres domaines connexes. Les termes étudiés empruntent principalement à ceux d'« infectiologie, de « pneumologie » et de « virologie », etc. Dans cette perspective, le corpus n'est pas seulement une référence sur la constitution et les transformations sémantiques qui s'opèrent au sein d'une langue, mais également un témoin des événements historiques et des variations terminologiques qui sont les principaux instigateurs de ce changement. Les textes analysés

sont souvent liés aux événements scientifiques qui font date dans l'histoire du domaine. La situation de la COVID-19 s'est développée principalement à partir du 1er décembre 2019 : le premier cas de Covid-19 a été détecté à Wuhan.

4. Critères de sélection des candidats-termes

Plusieurs critères peuvent être pris en considération lors du choix du terme existant dans le corpus. Les critères de sélection peuvent être classés sur deux catégories : critères généraux et critères linguistiques.

- **Critères généraux** : celui-ci porte sur l'analyse du corpus et sur la fréquence du terme dans le corpus. Le premier critère important est la fréquence des occurrences. Il faut souligner qu'une fréquence non négligeable dans le texte de spécialité permet à un terme d'entrer dans la catégorie des candidats-termes. Cependant ce critère n'est pas une condition *sine qua non* à l'inclusion d'un terme. Il s'agit de connaître les unités terminologiques les plus utilisées en parcourant le corpus. « L'unité lexicale a un sens qui est lié à un domaine de spécialité ; ce dernier est délimité au préalable pour un projet terminographique donné. » (L'Homme, 2004 : 64).
- **Critères linguistiques** : il implique principalement le sémantisme et la structure syntaxique du terme. Nous avons utilisé deux méthodes afin de repérer les unités terminologiques et leurs variantes dans le corpus. Premièrement, certains termes sont a priori transparents, ils figurent néanmoins comme vedettes en raison de leur appartenance à l'épidémiologie. L'ensemble des candidats-termes suivants « personne infectée », « transmission par aérosols » et « personne porteuse du virus » tirés de notre corpus, présentent des unités terminologiques complexes dans le discours. Nous voyons dans ces exemples que la lemmatisation des mots contenus dans les textes consiste à ramener les formes fléchies des mots variables à une forme canonique. Les termes sont ramenés au singulier, les adjectifs au masculin et les verbes à l'infinitif. Les unités terminologiques sont généralement présentées dans leur forme pleine avec la citation de la forme réduite quand elle existe. La forme pleine se trouve forcément en contexte. Deuxièmement, il nous semble difficile de définir les critères de différenciation entre une unité terminologique et un phraséologisme. La phraséologie d'une langue est l'ensemble constitué par ses lexies prépositionnelles et ses lexies complexes (Tournier, Tournier, 2009 : 278).

5. Formation de la terminologie de la COVID-19

Le corpus a aussi pour but de fournir une liste de termes qui ont trait à la pandémie concernée. Nous relevons deux types d'unités terminologiques nominales : unité terminologique simple et unité terminologique complexe. Les matrices syntaxiques des unités terminologiques complexes nominales dans le corpus électronique se présentent comme suit.

- 42 % pour le groupe : noms – adjectifs « affection préexistante », « maladie respiratoire », « voie respiratoire » ;
- 22 % pour le groupe : nom – préposition – nom « hygiène des mains », « maladie à coronavirus » ;
- 8 % pour le groupe : nom – participe passé « personne contaminée », « cas confirmé » ;
- 8 % pour le groupe : nom – préposition – nom – adjectif « équipements de protection individuelle » ;
- 5 % pour le groupe : nom – adjectif – adjectif « Atteinte pulmonaire aiguë » ;
- 3 % pour le groupe : nom – adjectif – préposition – nom « Unité néonatale de soins intensifs ».
- 3 % pour le groupe : nom – adjectif – adjectif – adjectif « syndrome respiratoire aigu sévère », « infections respiratoires aiguës sévères », « maladie respiratoire aiguë grave ».

La fréquence pour les autres groupes apparaît relativement faible dans le discours traité. Les suites formées d'un nom – préposition – nom – adjectif - préposition - nom – adjectif représentent également 0,2 % « urgence de santé publique de portée internationale ». La suite nom–nom « hôte réservoir » apparaît très faible dans le discours.

Enfin, les groupes de chaînes restants sont très peu nombreux dans la totalité du corpus. Ces groupes sont d'ailleurs constitués de deux termes issus de groupes précédents. Ces deux termes se trouvent reliés à des éléments tels que « en matière de » dans « information en matière de santé ».

Unité terminologique simple	Unité terminologique complexe
quarantaine	masque chirurgical
distance	période d'incubation
confinement	taux de létalité
masque	infections respiratoires aiguës sévères
rumeur	saturation en oxygène
protection	prévention et contrôle des infections
transmission	système cardiovasculaire
traitement	équipements de protection individuelle

Plusieurs sigles ont été relevés. Ces sigles français sont souvent calqués sur les sigles anglais.

Terme français — Siglaison	Équivalent anglais — siglaison
Covid-19	COVID-19
SDRA	ARDS
IRAS	SARI
USI	ICU
PEP	PEEP
SpO2	SpO2
EVS	SBT
CDC	CDC
PAPR	PAPR
EPI	PPE
SRAS	SARS
SARS-Cov2	SARS-CoV-2
VM	MV
UNSI	NICU
PAVM	VAP

La recherche terminologique consiste à ne pas rechercher seulement des unités terminologiques nominales, mais aussi des unités verbales.

Verbe spécialisé	Équivalent anglais
éternuer	Sneeze
se propager	Spread
protéger	Protect
prévenir	Prevent
désinfecter	Disinfect
protéger	Shielding
confiner	lockdown
deconfiner	lift the lockdown
écouvillonner	Swab
reconfiner	reinstate the lockdown
s'isoler	self-isolate
mettre en quarantaine	quarantine

Le verbe « confiner » fait l'objet d'un calque en anglais sur le terme français.

Verbe spécialisé anglais	Équivalent français
reconfine	Reconfiner
deconfinement	Deconfinement
deconfine	Déconfiner
confinement	Confinement
confined	Confiné
deconfinement plan	plan de déconfinement

D'autres termes sont calqués sur l'anglais.

Unité terminologique française	Équivalent anglais	Unité terminologique française
tracer les contacts	trace contacts	rechercher les contacts
application de traçage des contacts	contact tracing application	application de recherche des contacts
traçage des contacts	contact tracing	recherche des contacts

Nous avons relevé également des adjectifs relatifs à la COVID-19.

Terme français — adjectif	Équivalent anglais
Présymptomatique	presymptomatic
Symptomatique	symptomatic
Respiratoire	respiratory
Contagieux	contagious
Asymptomatique	asymptomatic
Immunodéprimé	immunodepressed
Immunosupprimé	immunosuppressed
Confiné	locked down

Une liste de termes communs à l'ensemble du corpus s'est constituée progressivement, qui n'était cependant pas définitive. L'étape de dépouillement du corpus a bien pour objectif d'établir une liste de termes susceptibles de figurer comme vedettes. Durant cette recherche, nous avons réduit l'ensemble des termes à une terminologie qui appartient exclusivement à la thématique traitée. Nous avons finalement choisi de réviser la correspondance entre les termes relevés et les classes conceptuelles en fonction de leur pertinence.

6. L'emploi du verbe terminologique

Un enjeu qu'il faut prendre en considération vise l'analyse de l'environnement contextuel du verbe spécialisé afin d'évaluer et d'alimenter les ressources terminologiques (L'Homme, 2004, 2008). Un corpus est une référence qui, par l'organisation, les définitions et les informations qu'il fournit, pourrait refléter une partie de l'organisation conceptuelle du domaine. Ainsi la présence et l'absence du verbe dans le corpus nous en disent long sur son utilisation et son importance chez les spécialistes. Une analyse des verbes donnés dans le corpus nous permet de mieux situer le contexte tel qu'exprimé par les locuteurs, en lien avec la réalité et les concepts contigus. Nous traitons le verbe terminologique (L'Homme, 2012) « infecter » en nous appuyant sur quelques travaux portant sur l'étude sémantico-syntaxique du lexique (Fillmore, 1982, Mel'čuk et al. 1995, Gross 2012) :

INFECTER, v. tr. Structure actancielle : X1 (Homme) ~ Y1 (Homme) ou X2 (STRUCTURE ANATOMIQUE) ~ Y2 (STRUCTURE ANATOMIQUE) ou X3 (AGENT INFECTIEUX) ~ Y3 (Homme) Construction syntaxique : X infecte Y ou X s'infecte.

La réalisation des actants X et Y sont divers dans le corpus, nous notons « agent infectieux », « bactérie », « virus » pour X et « Homme », « personne » et « sujet » pour l'actant Y.

Prenons le verbe « traiter » fréquemment utilisés dans le corpus. Voici ci-après les réalisations linguistiques des actants pour le verbe en question :

<p>[ENG] WHO has developed a comprehensive “event management system” to <u>manage</u> critical information about outbreaks and ensure accurate and timely communications between key international public health professionals, including WHO Regional Offices, Country Offices, collaborating centres and partners in the Global Outbreak Alert and Response Network.</p>	<p>[FRA] L'OMS a élaboré un système de gestion global permettant de <u>traiter</u> les informations cruciales sur les flambées et d'assurer leur communication rapide et fiable entre les principaux intervenants dans le domaine de la santé publique internationale (bureaux régionaux, bureaux de pays et centres collaborateurs de l'OMS, partenaires du réseau mondial d'alerte et d'action en cas d'épidémie). [~]</p>
<p>[ENG] Are antibiotics effective in preventing and <u>treating</u> the new coronavirus?</p>	<p>[FRA] Les antibiotiques sont-ils efficaces pour prévenir et <u>traiter</u> l'infection par le nouveau coronavirus ? [~]</p>
<p>[ENG] If you catch the disease, make sure you <u>treat</u> your symptoms.</p>	<p>[FRA] Si vous contractez la maladie, assurez-vous de <u>traiter</u> vos symptômes. [~]</p>
<p>[ENG] Are there any specific medicines to prevent or <u>treat</u> the new coronavirus?</p>	<p>[FRA] Existe-t-il des médicaments spécifiques pour prévenir ou <u>traiter</u> l'infection par le nouveau coronavirus ? [~]</p>

[ENG] To date, there is no specific medicine recommended to prevent or <u>treat</u> the new coronavirus (2019-nCoV).	[FRA] À ce jour, aucun médicament spécifique n'est recommandé pour prévenir ou <u>traiter</u> l'infection par le nouveau coronavirus (2019-nCoV). [~]
[ENG] Previous outbreaks have demonstrated that when health systems are overwhelmed, mortality from vaccine-preventable and other <u>treatable</u> conditions can also increase dramatically.	[FRA] D'autres épidémies ont montré que lorsque les systèmes de santé sont surchargés, la mortalité due à des maladies évitables par la vaccination et à d'autres affections que l'on peut <u>traiter</u> risque elle aussi d'augmenter considérablement. [~]
[ENG] Being <u>treated</u> with respect and dignity;	[FRA] être <u>traitée</u> avec respect et dignité ; [~]
[ENG] But ultimately, the best way for countries to end restrictions and ease their economic effects is to attack the virus, with the aggressive and comprehensive package of measures that we have spoken about many times before: find, test, isolate and <u>treat</u> every case, and trace every contact.	[FRA] Cependant, en définitive, la meilleure façon pour les pays de mettre fin aux restrictions et d'atténuer les effets économiques de celles-ci consiste à s'attaquer au virus, au moyen de la mise en place d'un ensemble de mesures très actives et complètes que nous avons déjà évoquées à plusieurs reprises, à savoir : détecter, tester, isoler et <u>traiter</u> les cas, et identifier les contacts. [~]
[ENG] Countries must continue to find, test, isolate and <u>treat</u> every case and trace every contact.	[FRA] Les pays doivent continuer à détecter, tester, isoler et <u>traiter</u> chaque cas, et à identifier tous les contacts. [~]

Nous pouvons observer que le verbe «traiter» renvoie à différentes acceptations dans le cadre des données recueillies. Dans notre thématique, il renvoie à trois conceptions selon les contextes précédents :

- Faire subir une certaine transformation à des données, à des informations.
- Mettre en œuvre pour combattre une maladie ou un problème de santé.

«Traiter», v. tr. Structure actancielle : X1 (réalisation ling. des actants : Homme)

- Y1 (réalisation ling. des actants : données)

«Traiter», v. tr. Structure actancielle : X2 (réalisation ling. des actants : Homme)

- Y2 (réalisation ling. des actants : cas)

«Traiter», v. tr. Structure actancielle : X3 (réalisation des actants ling. : Homme)

- Y3 (réalisation ling. des actants : infection, symptôme, maladie).

Construction syntaxique simplifiée : X traite Y ou X est traité.

La construction «un virus se propage» présente un degré de figement dans le cadre de ce discours, c'est à dire, une occurrence élevée du verbe «se propager» associant à un «un virus». La construction : «se propager/un virus» présente une unité verbale complexe figée. Un verbe comme «se propager» a été repéré en

146 occurrences (sur 328 pages) avec «un virus». Le verbe seul ne peut pas être traité comme vedette, mais l'ensemble de la construction syntaxique du verbe et de son argument entre plutôt dans le champ de l'analyse sémantico-syntaxique. Le «virus» est souvent défini comme un «microorganisme pathogène», contrairement à d'autres contextes où il est qualifié comme «logiciel malveillant» en informatique.

Conclusion

Nous avons constitué un corpus afin d'observer l'emploi linguistique des termes utilisés pour la COVID-19 plus particulièrement dans les textes à caractères normatifs issus de l'OMS. Conjointement, nous avons dépouillé les unités terminologiques existantes. Le corpus des textes rassemblés permet de formuler quelques observations sur le contact des langues, instruments de communication des communautés scientifiques, au sein desquelles se sont produits les objets, les phénomènes, les processus et les événements historiques d'un domaine donné. La situation de la COVID-19 ne cesse de progresser, il en est de même de sa terminologie : ainsi le corpus de textes de spécialité reste ouvert pour d'autres sources qui seront impliquées dans le cadre de notre recherche. Une partie de la terminologie s'est stabilisée telle que le terme COVID-19, ce terme est intégré aux dictionnaires et aux bases de données lexicales, mais un nombre de termes reste en voie de développement.

Le contexte constitue donc certainement l'une des questions fondamentales de l'analyse du verbe terminologique. Dans la mesure où plusieurs verbes peuvent être polysémiques, même à l'intérieur du domaine, le contexte permet de déterminer s'il s'agit d'un concept/terme lié à telle ou telle classe. Le contexte ne constitue donc pas une délimitation mécanique que l'on pourrait effectuer en sélectionnant un nombre déterminé de noms au hasard. Pour analyser les informations sémantico-syntaxiques du verbe spécialisé, il est nécessaire de prendre en considération son environnement.

Les résultats obtenus précédemment, basant sur un corpus en terminologie, pourraient tendre vers la création d'une base de données terminologiques riches, et enrichir les bases de données afin de prévenir l'utilisation d'une terminologie cohérente et stable sur la pandémie. Cette terminologie peut être destinée à l'extension des bases de données des institutions nationales et internationales.

Enfin, le travail sur un corpus de spécialité pourrait également répondre à un objectif pédagogique ou même méthodologique. L'objectif pédagogique s'inscrit dans le cadre de la création d'une méthode d'aide à l'acquisition de la terminologie

d'une thématique particulière. La liste de termes peut être ensuite destinée à l'acquisition de la terminologie spécialisée.

Bibliographie

- Biber, D. 1993. «Representativeness in Corpus Design». *Literary and Linguistic Computing* 8.4. p. 243-257.
- Cabré, M. T. 1998. *La terminologie : théorie, méthode et applications*. Traduit par M. C. Cormier, J. Humbley. Ottawa : Presse de l'université d'Ottawa & Paris : Armand Colin.
- Cabré, M. T. 2008. «Constituer un corpus de textes de spécialité». Université de Paris : *Cahier du CIEL*. p. 37-56.
- Condamines A., Rebeyrolle J. 1996. «Point de vue en langue spécialisée». *Meta : journal des traducteurs/ Meta : translators' Journal*. Montréal : PUM. N° 42. Vol. 1.p. 174-184.
- Condamines, A. 1993. «Un exemple d'utilisation de connaissances de sémantique lexicale : acquisition semi-automatique d'un vocabulaire de spécialité». *Cahiers de lexicologie*. Bruxelles : Duculot/Aupelf-UREF. N° 62, p. 25-65.
- Condamines, A. 2003. *Sémantique et Corpus spécialisés : Constitution de bases de connaissances terminologiques*. Mémoire d'Habilitation à diriger de recherches. Université Toulouse le Mirail.
- Condamines, A. 2005. «Linguistique du corpus et terminologie». *Langages*, Vol. 39. N° 157.p. 36-47.
- Dahlberg, I. 1981. «Les objets et les notions, les définitions et les termes ». Rondeau G., Felber H. (dir.), *Textes choisis de terminologie*. Québec : université Laval, Giresterm. p. 221-282.
- Drouin, P. 2003. «Term extraction using non-technical corpora as a point of leverage». L'Homme, M.C, Kaguera, k. [dir.]. *Terminology*. Vol. 9, issue 1. Amsterdam/ Philadelphia: John Benjamins publishing. p. 99-115.
- Fillmore, C. J. 1982. «Frame Semantics». *Linguistics in the Morning Calm*. Éd. The linguistics society of Korea. Seoul : Hanshin Publishing Co. p.111-137.
- Gross, G. 2012. *Manuel d'analyse linguistique : Approche sémantico-syntaxique du lexique (sens et structure)*. Villeneuve-d'Ascq : Presse universitaire du Septentrion.
- L'Homme, M. -C. 2004. *La terminologie : principes et techniques*. Montréal : PUM. Coll. Paramètres.
- L'Homme, M. -C. 2008. «Le dicoinfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés». *Traduire*. Paris : Société française des traducteurs. N° 217. p.141-153.
- L'Homme, M. -C. 2012. «*Le verbe terminologique : un portait de travaux récents*». Frank Neveu et coll. (Dir.), *Actes du troisième congrès mondial français*. Lyon : EDP sciences.
- Mel'čuk, I., Clas, A., Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Belgique: Louvain-la-Neuve, Duculot/Aupelf-UREF.
- Pearson, J. 1998. *Terms in context*. Amsterdam/Philadelphia: John Benjamins publishing.
- Tournier, N., Tournier, J. 2009. *Dictionnaire de Lexicologie française*. Préfacé par H. Walter. Paris : Édition Ellipses.

Références électroniques

- V-Médical : Une encyclopédie médicale sur le site suivant «vocabulary-medical.fr» [consulté le 28 octobre 2020].
- GDT : *Le Grand Dictionnaire terminologique* «<http://gdt.oqlf.gouv.qc.ca>» [consulté le 30 octobre 2020].

Notes

1. Les sigles étrangers prennent généralement le genre qu'aurait en français le mot de base qui les compose. En vertu de cette règle, COVID-19 est de genre féminin, car dans la forme longue du terme français, maladie à coronavirus 2019, le mot de base est maladie (GDT).
2. <https://youalign.com>