Extraction de contextes riches en connaissances à partir d'un corpus comparable de textes médicaux (français-arabe)

Rim Abouwarda

Extraction de contextes riches en connaissances à partir d'un corpus comparable de textes médicaux

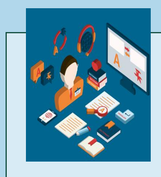
Rim ABOUWARDA – Doctorante en cotutelle (Université d'Alexandrie / Université Grenoble Alpes)
rim.abouwarda@univ-grenoble-alpes.fr / rim.abouwarda@alexu.edu.eg

INTRODUCTION

Les langues de spécialité suscitent actuellement un grand intérêt du fait que les sciences et les techniques connaissent une évolution continue, ce qui est reflété par les besoins linguistiques accrus.

La présente analyse, inscrite dans le cadre de notre projet de thèse, est basée sur l'étude d'un corpus comparable de textes médicaux (français-arabe) **dans le but de répondre aux besoins d'ordre traductionnel dépendant du contexte.** En effet, les sources d'informations auxquelles ont recours les traducteurs sont souvent les dictionnaires ou les glossaires offrant un nombre limité de contextes. Il s'ensuit donc que la consultation des **bases de données terminologiques** présentant des informations lexicales et conceptuelles est fortement recommandée.

OBJECTIFS DE L'ÉTUDE



L'étude s'adresse aux **traducteurs** en tant qu'« *utilisateurs prioritaires* » de la terminologie (Cabré, 1998). Notre recherche s'est donc fixée comme objectif de fournir aux traducteurs une **contextualisation des termes** dans les deux langues source et cible leur permettant de mieux saisir le sens des termes et d'en faire un usage à bon escient.

CADRE THÉORIQUE

Terminologie textuelle (Bourigault et Slodzian, 1999) → Optique lexico-sémantique (L'Homme, 2020) → Contextes riches en connaissances CRC (Meyer 2001)

Dans la littérature, plusieurs études se sont intéressées à l'identification et à l'extraction des CRC en corpus comparable. Notons, à ce propos, la thèse de Hmida (2017) consacrée à l'étude des patrons de connaissances pour l'identification des définitions dans des corpus comparables portant sur l'oncologie et la vulcanologie. Cette étude a mis l'accent sur l'instabilité des marqueurs de relations due à la polysémie. Hypothèse corroborée par d'autres travaux comme Condamines (2002) et Marshman (2014). De plus, les marqueurs conceptuels ont été également étudiés dans une perspective diachronique par Picton (2009) pour définir des marqueurs d'évolution servant à identifier des CRC dans le domaine spatial.

Bien que la littérature sur le sujet soit assez riche, la dimension contrastive axée sur la combinaison linguistique français-arabe n'a pas été explorée. Dans ce sillage, il nous semble donc intéressant de se pencher sur les marqueurs lexico-syntaxiques en langue arabe permettant d'identifier des CRC. Pour ce faire, et en se basant sur les études menées sur le sujet, nous avons pu formuler l'hypothèse **qu'une variation au niveau des patterns linguistiques entre le français et l'arabe pourrait être observée.**

PROBLÉMATIQUE

Comment, à partir de l'interrogation d'un corpus de textes comparable dans le domaine médical, extraire des contextes définitoires et identifier les marqueurs lexico-syntaxiques dans une perspective contrastive dans les deux langues français et arabe ?

CORPUS COMPARABLE

Critères	Corpus français	Corpus arabe
Période	2000-2021	1980-2021
Type de documents	Articles scientifiques Manuels de psychiatrie	Ouvrages scientifiques Articles scientifiques
Degré de spécialisation	Rédigés par des experts Destinés à des spécialistes Destinés au grand public	Rédigés par des experts Destinés à des spécialistes Destinés au grand public
Taille	18 revues scientifiques (n° de 2000 à 2021) + 7 manuels de psychiatrie	20 ouvrages scientifiques + 5 revues scientifiques (n° de 2014 à 2021)
Format	XML et TXT	TXT

CADRE MÉTHODOLOGIQUE

Extraction des contextes définitoires en français: liste MAR-EL pour les marqueurs conceptuels (Lefeuvre, 2017)

Traitement outillé du corpus: le Lexicoscope (Kraif, 2019)

Absence d'une liste existante de marqueurs linguistiques en arabe: Extraction et analyse des contextes où apparaissent les candidats-termes grâce aux patterns productifs prédéfinis dans le cadre de notre thèse sur la base d'une liste de fréquence de référence.

Il nous revient de noter que Lehmann et Martin-Berthet (2008) ont classé les types de contextes en trois catégories : définitoires, encyclopédiques et linguistiques. Nous nous focalisons, aux fins de notre analyse, sur les **contextes définitoires renfermant des éléments permettant de se renseigner sur la signification du terme**.

ANALYSES ET RÉSULTATS

• Convergences dans l'emploi de certains marqueurs linguistiques:

Français	Terme +	V. être +	Déterminant + Nom +	Caractérisé(e) +	par
Arabe	المصطلح +	هو +	اسم نكرة+	يتميز +	ب

❑ *La schizophrénie est une affection caractérisée par un ensemble de signes particuliers, incluant des idées délirantes, des hallucinations, un discours désorganisé, (...)*

❑ *اضطراب وجداني ثنائي قطبي هو اضطراب يتميز بنوبات متكررة يضطرب فيها مزاج الشخص ومستوى نشاطه بشكل عميق.*

اضطرا ب	وجدان ي	ثنائي قطبي	هو	اضطرا ب	يتميز	ب	نوب ات	متكررة	يضطر ب فيها	مزاج	الشخص	و	مستوى	نشاطه	بشكل عميق
Le troubl e	affect if	bipolai re	es t	un troubl e	caract érisé	par	des crise s	récurre ntes	qui troubl ent	l'hume ur	de l'indivi du	et	Son niveau	d'acti vité	profondé ment

• Prédominance des indices typo-dispositionnel comme « : » dans le corpus arabe:

❑ *اضطراب التوافق: هي حالات من الضيق الذاتي والضيق الانفعالي (...)*

اضطرا ب	التوافق :	هي	حالات	من	الضيق	و	الذاتي	الضيق	الانفعالي
Le trouble	d'adaptation :	es t	des états	de	malaise	et	personnel	malaise	émotionn el

• Emploi du mot « التعريف » (la définition) précédant le contexte définitoire :

❑ *الهذيان الارتعاشي:*
التعريف: هو مرض عقلي ذهاني ونوع خاص من الهذيان الحاد (...)

الهذيان	الارتعاشي:	التعريف:	هو	مرض	عقلي	ذهاني	و	نوع	خاص	من	الهذيان	الحاد
Le delirium	tremens	La définition :	es t	une maladi e	mental e	délirant e	et	un typ e	particul ier	du	délire	grave

• Emploi plus fréquent des indices lexicaux « comme, c'est-à-dire » dans le corpus français:

❑ *L'autisme est défini comme une maladie complexe du développement du SNC et est associé à une étiologie multifactorielle.*

❑ *L'anorexie mentale chez les adolescentes se présente comme un syndrome qui associe simultanément ou progressivement : une perte d'appétit ; un amaigrissement ; l'arrêt des règles (aménorrhée) ; (...)*

❑ *La survenue d'attaques de panique régulières peut aussi conduire à l'installation d'une agoraphobie, c'est-à-dire la crainte de se trouver dans un lieu ou une situation d'où il sera impossible de s'échapper en cas de survenue d'une attaque de panique.*