

**Modélisation de la polysémie : approche contrastive arabe – français basée
sur la Théorie Sens-Texte**

Thèse de doctorat ès-Lettres

Travail présenté par

Racha MOHAMMAD SALEM

Sous la direction de

Mme le professeur docteur / Jehanne KIRA

Mme docteur / Samah NASR

Mme docteur / Inas AL-HADIDY

Novembre 2017

La pluralité de sens est un des phénomènes linguistiques fondamentaux qui marque l'évolution et la richesse des langues naturelles, nous en distinguons deux aspects : l'homonymie et la polysémie. Certains linguistes arabes aussi bien que français marginalisent le rôle de la pluralité de sens ou nient sa présence. Ce point de vue contredit l'objectif principal de la linguistique qui décrit la communication langagière entre le locuteur et le destinataire et la transforme en un outil qui aide uniquement le deuxième actant de la communication à analyser le message émis par le premier actant sans s'intéresser à offrir à ce dernier les moyens de formuler son message.

Grâce à l'évolution sociale, scientifique et industrielle, le lexique subit des changements importants au fil des siècles. Certains termes monosémiques acquièrent de nouveaux sens et deviennent donc polysémiques. Avec l'avènement du traitement automatique des langues naturelles (TALN), une nouvelle branche d'études linguistiques voit le jour, c'est la linguistique formelle qui se charge de la création d'outils informatiques aptes de gérer le lexique humain avec toutes ses complexités et à leur tête l'homonymie et la polysémie.

Les théories de la linguistique formelle sont multiples. Nous choisissons comme cadre théorique la Théorie Sens-Texte fondée par Igor Mel'čuk. L'atout de la TST est la place intermédiaire qu'elle occupe entre la linguistique classique et la linguistique formelle, ce qui la rend plus fiable à décoder les énigmes des langues naturelles.

La présente recherche étudie la polysémie dans le but de :

- proposer une nouvelle méthode de description lexicologique des polysèmes qui diffère de celle adoptée actuellement dans les dictionnaires arabes et qui délimite les frontières séparant les homonymes des polysèmes.
- Présenter l'exemple d'un modèle informatisé apte de remédier aux problèmes linguistiques qui résultent de l'ambiguïté d'une grande partie du

lexique arabe et qui, par conséquence, mettent en question la performance des logiciels du traitement automatique de la langue arabe.

Pour réaliser ces objectifs, l'étude s'effectue selon deux axes : lexicologique et informatique.

Nous étudions le patron polysémique [phénomène naturel – phénomène social] qui est constitué de six vocables en langue arabe et de quatre en langue française. La recherche se compose de six chapitres qui présentent le processus d'analyse linguistique et de modélisation de la polysémie dans les deux langues. Le registre langagier ciblé par la recherche est le registre standard. Nous étudions la langue arabe égyptienne standard et la langue française courante. Pour l'extraction des exemples, nous nous sommes référés à des corpus journalistiques égyptiens et français.

Le premier chapitre (*Problématique et état de l'art*) présente un bref compte rendu des travaux français et arabes qui ont été effectués à propos de la pluralité de sens. Nous commençons par Michel Bréal qui emploie le terme « polysémie » pour la première fois dans son ouvrage **Essai de sémantique**, en passant par Ferdinand De Saussure qui n'évoque ni l'homonymie ni la polysémie dans ses **Cours**. Au contraire, il élabore la règle de biunivocité du signe linguistique qui met en évidence le rapport unique et indissociable liant le signifié au signifiant. Parmi les linguistes qui se sont intéressés à l'étude de la polysémie, nous citons à titre d'exemple : André Martinet, Jacqueline Picoche et Georges Kleiber.

Les linguistes arabes avaient pris conscience de l'importance de la pluralité de sens même avant les Français. Au huitième siècle, (سيبويه) Sîbawayhi, dans son ouvrage **الكتاب** ('*al-Kitâb* / Le Livre), classait le lexique arabe selon trois catégories dont la troisième groupe les mots ayant plusieurs sens. Au quinzième siècle, (السيوطي) 'al-Siyûtî décrivait le phénomène de pluralité de sens en

l'appelant « المشترك اللفظي ». Les linguistes arabes contemporains présentent une nouvelle conception du phénomène de pluralité de sens plus précise que celle avancée par les anciens.

Le deuxième chapitre (*Cadre théorique*) aborde les principes de la TST en se focalisant sur la paraphrase, la lexicologie explicative et combinatoire (LEC) ainsi que le système de fonctions lexicales (FL). La TST lance une nouvelle conception de la langue naturelle et du signe linguistique qui est basée principalement sur la combinatoire et la dépendance. Igor Mel'čuk considère le signe linguistique comme étant un triangle qui se compose de trois angles : le signifié, le signifiant et le syntactique. Ce dernier étudie la combinatoire du signe avec les autres signes de la langue.

L'importance du syntactique s'accroît en cas d'ambiguïté lexicale. L'analyse du caractère combinatoire facilite la distinction entre les homonymes d'une unité lexicale ou les lexies à l'intérieur d'un vocable car un signe linguistique n'a qu'un seul syntactique. Le deuxième chapitre explique en détails le modèle linguistique de la Théorie Sens-Texte (MST) qui se compose de sept niveaux.

Le troisième chapitre (*Description lexicologique des vocables du patron [phénomène naturel – phénomène social]*) présente la description lexicologique des vocables appartenant au patron polysémique [phénomène naturel – phénomène social]. Ce dernier s'est construit de l'usage des lexies qui désignent un phénomène naturel pour caractériser un phénomène social qui influence positivement ou négativement une collectivité d'humains. Il se compose, en langue arabe, de six vocables : « شمس », « نور », « فجر », « نجم », « نار » et « شرارة ». Cependant, le même patron, en langue française, ne comporte que quatre vocables : « LUMIERE », « AUBE », « FEU » et « ETINCELLE ». Malgré leur caractère polysémique, les vocables « SOLEIL » et « ETOILE » en sont exclus car le sens du phénomène social n'existe pas en français. Nous commençons par expliquer la

définition lexicologique selon la TST qui s'effectue selon certaines règles dont les plus importantes sont la règle de décomposition sémantique et la règle du bloc maximal. D'autre part, le patron [phénomène naturel – phénomène social] se caractérise par une palette de liens polysémiques variés dont la plupart sont tissés de la même manière en français aussi bien qu'en arabe. Ce qui souligne la quasi-équivalence qui se trouve entre les deux langues sur le plan lexicologique.

Le quatrième chapitre (*Etude de la combinatoire lexicale et syntaxique dans un vocable*) aborde les différents types de phrasèmes qui sont formés à travers les unités lexicales étudiées : Les collocations, les phrasèmes à double nature et les greffes collocationnelles.

Le cinquième chapitre (*Le caractère polysémique des collocations arabes*) traite de la composition des collocations polysémiques qui résultent de la similarité existant entre les syntactiques de deux signes linguistiques. Contrairement aux vocables, les collocations polysémiques ne sont pas toutes analysables au niveau sémantique. La syntaxe joue un rôle crucial dans le déchiffrement de ce type de collocations. L'étude englobe les quatre vocables qui permettent la formation des collocations polysémiques. Ces vocables sont : « شمس », « نار », « نجم » et « فجر ».

Le sixième chapitre (*Modélisation des relations sémantico-lexicales à travers le système de fonctions lexicales*) explique le formalisme de la TST. Le système de fonctions lexicales est réparti en deux rubriques : les fonctions paradigmatiques et les fonctions syntagmatiques. La deuxième partie du sixième chapitre aborde l'implémentation des vocables dans le réseau lexical de l'arabe ; projet qui a été lancé en 2014 par Dr. Dina Al-Kassas. Nous présentons au cours de l'étude des exemples de la structure des vocables dans le RL-ar en appliquant les principes adoptés dans le réseau lexical du français.

Dans la dernière partie du sixième chapitre, nous avançons les articles lexicographiques des lexies arabes qui construisent le patron polysémique

[phénomène naturel – phénomène social]. Parallèlement, nous exposons l'équivalente française de chaque lexie arabe afin de mettre au clair les similarités et les différences entre les deux langues.

En guise de conclusion, nous avons abordé à travers la présente recherche le phénomène de la polysémie qui constitue un des aspects de pluralité de sens et qui représente, jusqu'à nos jours, un défi pour les modèles d'informatisation des langues naturelles. En élaborant les définitions, nous nous sommes référée aux corpus et aux dictionnaires. En fait, l'étude de la polysémie nécessite la référence à plusieurs ressources afin de distinguer les sens lexicalisés d'un polysème de ceux non lexicalisés. L'importance de cette variété de ressources est évidente surtout avec les vocables arabes dont l'étude reflète un décalage entre le dictionnaire et le contexte. Un sens qui est lexicalisé dans l'un, ne l'est pas dans l'autre. Ainsi, la polysémie met-elle en question le rapport entre la théorie et l'usage. De même, la polysémie dans le patron [phénomène naturel – phénomène social] met en question le rapport entre la culture et le lexique.

Les spécificités linguistiques de la langue arabe ont été relevées grâce à l'étude contrastive comparant cette dernière à la langue française dans le but de cerner les convergences et les divergences concernant la définition lexicologique ainsi que l'analyse sémantico-syntaxique des vocables dans les deux langues. L'étude des cooccurrents lexicaux des polysèmes nous a permis de distinguer quatre types de phrasèmes en langue arabe : les collocations, les greffes collocationnelles, les locutions et les phrasèmes de type bivalent. L'analyse des collocations et des greffes collocationnelles qui sont construites à partir des polysèmes a mis en relief l'existence d'une similarité au niveau sémantico-syntaxique entre l'arabe et le français qui sont effectivement considérées comme deux langues grammaticalement différentes. C'est ainsi que le patron polysémique

étudié a montré la possibilité de trouver des zones communes entre les langues naturelles dissemblables.

Reste à souligner qu'en domaine d'études contrastives, il est nécessaire de penser à la conception de réseaux lexicaux bilingues qui seront basés sur un système informatique composé de réseaux monolingues avec une technique qui organise l'échange linguistique entre les langues. C'est pour cette raison que nous espérons que la présente recherche ouvre la voie vers l'élaboration d'un réseau lexical bilingue français – arabe qui répondra aux besoins des chercheurs en domaine de la linguistique contrastive ainsi qu'en domaine de la linguistique formelle. Une ressource lexicale bilingue informatisée servira à la création d'outils informatiques pour la traduction automatique et la génération multilingue de textes.