

Un Modèle Informatisé pour la Traduction Automatique des Locutions Arabes et Françaises selon les Principes de la TST

Racha Mohammad Salem

Département de Français, Faculté des Lettres, Université d'Alexandrie, Egypte
salem.racha@live.com

Résumé : *Vu leur construction complexe, les locutions constituent une des énigmes des langues naturelles que la plupart des systèmes informatiques n'arrivent pas à déchiffrer jusqu'à nos jours. A travers l'examen des outils de traduction automatique gratuits disponibles en ligne, nous décrirons en détail la problématique, puis nous proposerons comme solution l'élaboration d'un modèle informatisé basé en principe sur l'analyse lexico-sémantique des locutions. Nous adopterons comme cadre théorique la Théorie Sens-Texte qui paraît la plus capable à surmonter le défi des locutions grâce à un système formel fondé sur une double analyse linguistico-informatique des langues naturelles où la sémantique est prioritaire.*

Mots-clés : *Locution, paraphrase, fonction lexicale, la Théorie Sens-Texte, traduction automatique, modèle informatisé, représentation sémantique.*

1 INTRODUCTION

La modélisation des locutions représente un vrai défi face aux logiciels de traduction automatique. Une locution est un syntagme non libre qui se compose de plusieurs unités lexicales et dont le sens provient de la combinatoire de ces dernières. Autrement dit, le sens d'une locution est implicite, il est impossible de le détecter à travers une simple étude lexico-syntaxique de ses composantes. Une analyse approfondie des différents contextes employant la locution soutenue par une étude plutôt lexico-sémantique de ses composantes sont nécessaires afin de dévoiler le sens profond. Le présent article aborde le traitement des locutions par les outils de traduction automatique disponibles en ligne, tels que *Google*⁽¹⁾, *Reverso*⁽²⁾, *Bing*⁽³⁾, *Yandex*⁽⁴⁾ et *Babylon*⁽⁵⁾.

A travers l'analyse de cinq locutions arabes et quatre françaises, une étude de cas sera avancée en cernant les points négatifs qui rendent le modèle informatique appliqué actuellement par ces outils incapable de les gérer et de les traduire correctement. Ensuite, nous proposerons un modèle informatisé qui est basé sur les principes de la Théorie Sens-Texte et qui est apte à offrir une traduction plus performante des locutions.

2 La TST ET L'ÉQUILIBRE ENTRE LA LINGUISTIQUE ET L'INFORMATIQUE

La Théorie Sens-Texte a été élaborée à Moscou dans les années soixante par Igor Mel'čuk, Alexandre Zholkovsky et Juri Apresjan dans le but de créer un modèle informatique qui est capable à transformer la langue naturelle en langage formel qui pourrait être utilisé par la suite dans les différentes applications du TALN surtout la traduction automatique. Pour aboutir à cette fin, ils élaboraient un système rigoureux de modélisation des langues naturelles qu'ils appellent le « Modèle Sens-Texte » (MST). Ce dernier se compose de deux volets : la linguistique et l'informatique. La TST propose une analyse linguistique multistratale : sémantique, syntaxique, morphologique et phonologique. En outre, l'unité lexicale n'est pas considérée comme une entité indépendante ou isolée, au contraire elle n'acquiert son sens qu'à travers les liens tenus avec les autres unités lexicales. Une langue naturelle est une chaîne qui se compose d'un nombre infini d'anneaux attachés l'un à l'autre. Pour la comprendre, il est indispensable d'étudier les rapports sémantico-syntaxiques qui sont esquissés entre les différents mots. Quant au deuxième volet du Modèle Sens-Texte, il concerne la formalisation de la langue naturelle. Au cours des lignes suivantes, nous expliquerons avec plus de détails les principes du MST.

A. Les Atouts de la TST

La TST se distingue des autres théories de linguistique informatique par l'intérêt qu'elle accorde aux études langagières. En fait, la plupart de ces théories s'intéressent au côté formel au détriment de la linguistique. Les résultats ne sont pas souvent satisfaisants. A l'encontre, les fondateurs de la TST se sont rendus compte que la modélisation d'une langue naturelle n'est possible qu'à travers une analyse linguistique méticuleuse de toutes les unités lexicales, d'une part et des relations paradigmatiques et syntagmatiques qui les unissent, d'autre part. Cette analyse s'effectue à travers deux processus représentant les bases fondamentales de la TST : la lexicologie explicative et combinatoire (LEC) et la paraphrase. La première constitue la composante lexicale de la Théorie Sens-Texte. Elle définit l'unité lexicale et analyse sa combinatoire avec les autres unités de la langue naturelle : la combinatoire lexicale étudie les collocations construites à partir d'un lexème⁽⁶⁾ alors que la combinatoire syntaxique étudie les relations syntaxiques qu'un lexème maintient avec les autres lexèmes qu'il gouverne et / ou dont il dépend [Racha Salem, 2017].

Le deuxième atout de la TST est l'instauration d'un système de paraphrasage multistratal qui débute par la sémantique, passe par la syntaxe et la morphologie pour terminer par la phonologie. Il est constitué d'un groupe de règles lexico-syntaxiques qui gèrent les différentes relations paradigmatiques et syntagmatiques liant les unités lexicales ainsi que les phénomènes linguistiques cruciaux d'une langue naturelle, tels que la synonymie et l'ambiguïté.

D'autre part, pour rendre notre apport plus clair, il paraît nécessaire de jeter la lumière sur les différents types de phrasèmes déterminés par les fondateurs de la TST.

B. Types de Phrasèmes Selon la TST

Un phrasème est un syntagme non libre dénotant toute combinatoire qui se fait d'une manière régulière et presque stable afin de répondre à une situation communicative particulière ou d'exprimer un sens précis. Igor Mel'čuk détermine dans son article intitulé « *Collocations dans le dictionnaire* », publié en 2003, trois formes de phrasèmes selon le degré de figement :

- Le phrasème complet (locution figée) est une expression idiomatique dont le sens est saisi à travers la combinatoire de ses composantes. Si un phrasème complet se compose de deux éléments (A + B), le sens n'est inclu ni dans (A) seul ni dans (B) seul, mais il résulte du fusionnement des deux composantes. Citons à titre d'exemple la locution « **donner carte blanche** » qui signifie « **donner la permission** ». Le sens ne se trouve ni dans le verbe 'donner', ni dans le nom 'carte', ni dans l'adjectif 'blanche'. En fait, l'amalgame de ces trois lexèmes donne naissance à un sens nouveau. La suppression d'un de ces trois éléments causera la perte du sens de la locution.
- Le semi-phrasème (collocation) est construit à partir de deux lexies ⁽⁷⁾ (A + B) dont l'une est communicativement dominante. Cette dernière qui est librement choisie par le locuteur constitue la base de collocation. L'autre lexie qui est sélectionnée en fonction de la première représente le collocatif. Le sens de la collocation est saisi d'après sa base. Exemples : la collocation « **le feu des critiques** », la lexie 'critiques' représente la base de collocation alors que la lexie 'feu' constitue le collocatif. Le sens réside dans la deuxième, si nous enlevons 'feu', l'idée restera claire : « **le fait d'attaquer une personne** ».
- Le quasi-phrasème (locution quasi-figée) est un cas intermédiaire entre le premier et le deuxième types. Le sens est inclu dans les deux lexies formant le syntagme. Si un quasi-phrasème se compose de deux éléments (A + B), le sens inclut celui de (A) et celui de (B). Ce qui veut dire que ni (A) ni (B) n'est communicativement dominant. Exemple : le quasi-phrasème « **donner le sein au bébé** » ⁽⁸⁾ signifie « **faire nourrir le bébé** ». C'est la combinatoire des trois lexies ['donner' + 'sein' + 'bébé'] qui communique le sens au destinataire. La suppression d'une lexie causera la perte d'une partie intégrale du sens du phrasème : on ne peut dire ni « donner au bébé » ni « le sein au bébé ».

C. La TST au Service du TAL

Le modèle informatisé de la TST reçoit à l'entrée un sens pour faire sortir toutes les phrases qui l'expriment. Comme nous l'avons signalé ci-dessus, le MST se compose de quatre représentations linguistiques : la sémantique, la syntaxe, la morphologie et la phonologie. Les trois dernières sont divisées en deux niveaux : niveau profond et niveau de surface. Les niveaux profonds sont orientés vers le sens ou le contenu qu'on cherche à exprimer, tandis que les niveaux de surface sont orientés vers le texte ou la forme du contenu exprimé.

Le volet formel du MST modélise les relations paradigmatiques et syntagmatiques qui lient les unités lexicales d'une langue naturelle à travers le système de fonctions lexicales (FL). Les fonctions lexicales paradigmatiques modélisent les relations sémantico-lexicales telles que la synonymie, la conversion et l'antonymie ainsi que les dérivations sémantiques (nom typique de l'actant, nom typique de l'instrument) et syntaxiques (la nominalisation, la verbalisation et l'adjectivisation). Les fonctions lexicales syntagmatiques modélisent les syntagmes non libres. Les FL sont toutes de type standard – universel – vu leur capacité à formaliser les relations linguistiques communes dans toutes les langues naturelles. Néanmoins, il existe un autre type de fonctions lexicales dit non standard ou non universel qui varient d'une langue à une autre. Elles formalisent les liens linguistiques spécifiques à chaque langue.

Dès son élaboration, le système des FL syntagmatiques a été appliqué aux collocations dont le processus de formalisation est beaucoup plus simple que celui des locutions. Des études abordant la modélisation des locutions ont vu le jour au cours des dernières années. Contrairement aux autres théories de linguistique formelle, la TST modélise non les unités lexicales, mais les relations lexico-sémantiques qui les lient. Car l'objectif est la modélisation du sens (le signifié) et non de la forme (le signifiant) d'un signe linguistique, ce qui remédie à maints problèmes linguistiques faisant face aux applications du TAL. Citons à titre d'exemple, la polysémie : une unité lexicale polysémique (ou vocable selon la terminologie de la TST) sera codée au sein d'un système informatique par plusieurs FL dont chacune représentera un sens. Revenons aux phrasèmes figés ou les locutions, leur modélisation est effectivement une modélisation du sens profond et non des unités lexicales dont elles se composent. C'est pourquoi la TST paraît plus efficace à traiter ce type de phrasèmes.

3 LE TRAITEMENT DES LOCUTIONS PAR LES OUTILS DE TRADUCTION AUTOMATIQUE

A. Les « Belles Infidèles »

Parmi les outils de traduction automatique qui sont actuellement utilisés par les internautes nous en avons sélectionné cinq afin d'analyser les traductions qu'ils proposent aux locutions arabes et françaises ciblées par la présente recherche. Au cours des pages suivantes, nous démontrons que ces outils ne donnent pas l'équivalent exact de chaque locution parce qu'ils sont incapables à détecter le phrasème figé. Ils le traitent comme si c'était un syntagme libre. C'est pour cette raison que la traduction obtenue est « belle » mais « infidèle » : il s'agit plutôt d'une traduction littérale de chaque unité lexicale dont la locution se compose. Une traduction qui ne reflète pas son vrai sens ⁽⁹⁾.

B. Vers une Traduction Plus Performante des Phrasèmes Figés et Quasi-Figés

Comme solution à ce problème, nous proposons la création d'un modèle informatisé qui intègre deux volets (lexical et informatique) à l'image du MST. Nous y accordons le nom (*MTLIL / Modèle du Traitement Lexico-Informatique des Locutions*). Il pourrait être introduit dans tout programme de traduction automatique. Le volet linguistique comporte une base de données qui détermine les locutions de chaque langue selon les critères Sens-Texte dont nous avons parlé ci-dessus. Une fois déterminées, ces locutions seront analysées sémantiquement pour décoder leurs sens profonds et par conséquent les traduire correctement. Le volet informatique du MTLIL modélise les locutions selon le système des fonctions lexicales afin de faciliter leur intégration dans les programmes informatiques.

Le module sémantique dans un MST fonctionne selon un groupe de règles formelles qui décortiquent le sens (ou le contenu sémantique) d'une unité lexicale (L), puis le paraphrasent en cernant toutes les formes lexicales et / ou les structures syntaxiques exprimant le sens de (L). Ces règles sont divisées en règles de transition (ou d'expression) et en règles d'équivalence (ou de paraphrasage). La première catégorie comporte trois types de règles : de lexicalisation, d'arborisation et de morphologisation. La deuxième catégorie se compose de deux types de règles : règles de paraphrasage d'équivalence et règle de paraphrasage lexico-syntaxique.

Le processus du déchiffrement d'une locution via le MTLIL est réalisé à travers les étapes suivantes : premièrement, déterminer tous les contextes où la locution en question figure. Deuxièmement, analyser sémantiquement ces contextes afin d'avancer une définition lexicographique de la locution en appliquant la **règle locutionnelle** / **R^{LEX3}** (troisième règle de lexicalisation qui explique les syntagmes non libres). Troisièmement, préciser, grâce aux règles de paraphrasage, les syntagmes libres ou les collocations qui

expriment le même contenu sémantique de la locution. Quatrièmement, traduire l'équivalent sémantique déjà précisé dans la langue cible. Ainsi le MTLIL facilite-t-il la tâche du traducteur automatique en remplaçant le phrasème figé par un groupe d'unités lexicales libres qui font surgir son sens caché. Cinquièmement, implémenter la locution dans le MTLIL à travers une FL reflétant son contenu sémantique explicité au préalable.

4 EXEMPLES DE LOCUTIONS TRAITÉES SELON LE MTLIL

Au cours des pages suivantes, nous présentons quelques exemples de locutions arabes et françaises qui sont traduites par les outils de traduction [**Bing traduction**] et [**Yandex**] respectivement. Elles seront traitées par le MTLIL dans le but d'y proposer d'autres traductions plus pertinentes.

A. Les Locutions Arabes

Il est à noter que les cinq locutions étudiées sont basées sur des images. Notre tâche consistera à décortiquer chaque image pour comprendre le sens profond dissimulé.

1) La Locution « يسد عين الشمس » : La locution arabe « يسد عين الشمس » { *yasidu 'ayna 'al- šams* } est traduite automatiquement par [**Bing**] par le syntagme verbal « obstruent l'œil du soleil ».

أكد الوزير أن لدينا كميات من السكر تسد عين الشمس.

'akadda 'al-wazîr 'anna ladayna kimiyât *tasidu 'ayna 'al- šams*

* Le ministre a confirmé que nous avons des charges de sucre qui obstruent l'œil du soleil ⁽¹⁰⁾.

Cette traduction littérale qui, bien qu'il soit grammaticalement correct, n'a aucun sens en langue française. Dans ce cas, nous faisons face à un double problème : l'outil de traduction n'arrive ni à détecter la locution ni la collocation autour de laquelle cette dernière a été construite. Expliquons davantage : le syntagme arabe « عين الشمس » { *'ayna 'al- šams* } est une collocation qui équivaut la collocation « قرص الشمس » { *qurş 'al- šams* } qui, à son tour, équivaut la collocation française « le disque solaire ». « عين الشمس » { *'ayna 'al- šams* } se combine avec le verbe « يسد » { *yasidu* } formant ainsi la locution figée « يسد عن الشمس » { *yasidu 'ayna 'al- šams* } qui dénote le sens « كبير جدا » { *kabîr jidan* } en langue arabe ou 'très grande quantité' en langue française. Elle est utilisée en tant qu'adjectif pour qualifier un nom. Nous pourrions la paraphraser par :

"س" يسد عين الشمس ≡ "س" له حجم يتجاوز الحجم المعتاد

(« X » a un volume qui dépasse les normes)

Ainsi la phrase arabe citée ci-dessus est plutôt traduite par :

Le ministre a affirmé que nous avons de **très grandes quantités** de sucre.

Formellement, cette locution est modélisée par la fonction lexicale **Magn** qui exprime l'intensification.

Magn (يسد عين الشمس) = كبير جدا

Notons que le vocable arabe « شمس » { *šams* } est riche en locutions qui reflètent l'importance de cet astre dans la vie les Arabes. Les différentes lexies de « شمس » sont souvent intégrées dans des phrasèmes complètement figés afin d'exprimer une situation extrême qui atteint son apogée.

2) *La Locution* « ذهب وراء الشمس » : Cette locution est basée sur une image : « aller derrière le soleil » est un acte fictif qui est utilisé en langue arabe pour désigner le fait de « تعرض لعقاب شديد » {ta'arada li 'iqâb šadîd} ou « subir une forte punition ».

صديقي ذهب وراء الشمس بسبب حاكم ظالم.

šadîqi ḡahaba warâ 'al- šams bisabab ḡakim zâlim

* *Mon ami est allé après le soleil à cause d'un souverain injuste.*

Cette traduction n'exprime que le sens superficiel du phrasème tout en négligeant le sens profond. La locution « ذهب وراء الشمس » pourrait être paraphrasée par la collocation « تعرض لعقاب شديد ».

"س" ذهب وراء الشمس ≡ "س" تعرض لعقاب شديد بواسطة "ص"

(« X » a subi une forte punition déterminée par « Y »)

C'est ainsi que nous pourrions traduire la phrase mentionnée de la façon suivante :

صديقي ذهب وراء الشمس بسبب حاكم ظالم.

Mon ami a subi une forte punition à cause d'un souverain injuste.

Du point de vue formel, la locution « ذهب وراء الشمس » est modélisée par l'amalgame de deux fonctions lexicales, ou une configuration de FL, selon la terminologie de la TST.

[Oper +Magn] تعرض لعقاب شديد = (ذهب وراء الشمس)

La première (Oper)⁽¹¹⁾ formalise l'action et la deuxième (Magn) formalise le sens d'intensification :

Oper (عقاب) = تعرض ~

Magn (عقاب) = شديد ~

3) *La Locution* « أطفأ نور الشمس » : Comme c'est le cas avec les deux locutions précédentes, l'outil de traduction automatique avance une traduction littérale de « أطفأ نور الشمس » {atfa' nûr 'al- šams}.

لا تطفنوا نور الشمس في حياتكم بعقوق الوالدين.

lâ tuḡfi'û nûr 'al- šams fi ḡyâtakum bi 'iqûq 'al-wâlidayn

* *N'éteignez pas la lumière du soleil dans votre vie en désobéissant les parents.*

En fait, le soleil symbolise la vie et l'espoir dans la culture arabe aussi bien que française. La disparition de la lumière du soleil symbolise par conséquent la fin ou la destruction de la vie. Lorsque nous disons :

(« س » أطفأ نور الشمس) (« X » a éteint la lumière du soleil)⁽¹²⁾, nous visons le fait que « X » a corrompu sa vie à cause des faits illégaux, injustes ou malhonnêtes qu'il a commis.

"س" أطفأ نور الشمس ≡ "س" فعل "ص" و الذي يؤثر سلبا على حياة "س".

« X » commet « Y » qui influe négativement la vie de « X ».

C'est ainsi que nous suggérons comme traduction à la locution « أطفأ نور الشمس » :

لا تطفنوا نور الشمس في حياتكم بعقوق الوالدين.

Ne corrompez pas votre vie en désobéissant à vos parents.

La fonction lexicale (**Degrad**) connote le sens 'dégradation' ou 'échec'. Elle pourrait modéliser la locution « أطفأ نور الشمس ».

Degrad (أفسد حياته = أطفأ نور الشمس)

4) La Locution « وقف على قلب رجل واحد » : La quatrième locution arabe que nous traitons dans le présent article est « وقف على قلب رجل واحد » {*yaqif 'ala qalb rajul wâhid*}. Le traducteur automatique traite chaque unité lexicale comme si c'était une entité indépendante. Ceci paraît clairement à travers la traduction littérale avancée par [Bing] :

يقف الشعب كله على قلب رجل واحد

yaqif 'al- ša 'b 'ala qalb rajul wâhid

* *Tout le peuple est au coeur d'un seul home.*

La locution « وقف على قلب رجل واحد » désigne en fait l'union ou l'alliance qui pourrait avoir lieu entre un groupe de personnes au niveau des idées et des opinions adoptées. Ce qui veut dire que cette locution pourrait être paraphrasée par le syntagme « تبنى موقفا موحدًا ».

وقف على قلب رجل واحد ≡ "س" و "ص" لهم نفس الرأي بشأن "ي"

« X » et « Y » adoptent la même opinion à propos de « W ».

Afin d'exprimer son vrai sens, la locution pourrait être traduite par :

Tout le peuple adopte la même opinion

En outre, la locution « وقف على قلب رجل واحد » est modélisée par une fonction lexicale non standard :

الرأي / الفكر

الوقوف على قلب رجل واحد : وحدة الرأي / الفكر

5) La Locution « العين لا تعلق على الحاجب » : La dernière locution arabe que nous abordons est souvent employée dans des contextes traitant du respect de la hiérarchie sociale ou professionnelle : La locution « العين لا تعلق على الحاجب » {*al-'ayn la ta'lû 'an 'al-ḥâjib*} est construite à partir des parties du corps humain : comme l'œil ne peut pas dépasser les sourcils, le plus jeune doit respecter le plus âgé et le subalterne doit respecter son directeur. Cependant, le traducteur automatique la traduit littéralement en « l'œil ne se substitue pas au front ». Examinons le contexte suivant :

المعارضة تدرك بأن العين لا تعلق على الحاجب و بأن صورة و هيئة السلطة خطان أحمران مقدسان لا يحق لأحد تجاوزهما.

'al-mu'âraḍa tudrik bi'anna 'al-'ayn la ta'lû 'an 'al-ḥâjib wa bi'anna šûrat wa haybat 'al-šulṭa ḥaṭân 'aḥmarân muqdsân la yaḥiq tajâwizhumâ

* *L'opposition se rend compte que l'œil ne substitue pas au front et que l'image et le prestige du pouvoir sont deux lignes rouges sacrées que personne n'a le droit de dépasser.*

Etant donné que la locution « العين لا تعلق على الحاجب » est employée pour dénoter le respect de la hiérarchie comme nous l'avons signalé ci-dessus, elle serait paraphrasée par « احترام المقام » {*iḥtirâm 'al-mqâm*}.

العين لا تعلق على الحاجب ≡ "س" يحترم مقام "ص"

« X » respecte la position (situation) élevée de « Y ».

Le syntagme français « respecter la hiérarchie » reflète le sens profond de la locution arabe :

L'opposition se rend compte qu'il faut respecter la hiérarchie et que l'image et le prestige du pouvoir sont deux lignes rouges sacrées que personne n'a le droit de dépasser.

La locution est modélisée par La FL (**Oper**) :

Oper (احترم المقام = (العين لا تعلق على الحاجب))

B. Les Locutions Françaises

Parmi les locutions françaises nous en avons sélectionné quatre pour les traiter selon le MTLIL. Nous présentons au départ les traductions suggérées par l'outil de traduction automatique disponible en ligne [*Yandex*].

1) *La Locution « Commencer une Famille »* : Comme c'est le cas avec les locutions arabes, l'outil de traduction considère « commencer une famille » comme étant un syntagme libre.

Commencer sa famille sur le tard a ses avantages, mais aussi ses inconvénients.

* *بدءا من الأسرة الخاص بك في وقت متأخر مزاياء، ولكن أيضا سلبياته.*

bad'an min 'al-' usra 'al-hâš bik fi waqt muta'ahir mazayâh, wa lakin 'aydan lahu salbiyâtuhu

En fait, il s'agit d'une locution quasi-figée dont le sens provient de l'amalgame des sens de ses deux éléments ('commencer' **A** + 'famille' **B**). Elle est paraphrasée par : (« X » est uni juridiquement à « Y »). La locution « commencer une famille » est synonyme du verbe « se marier » qui se traduit en arabe par « تزوج ».

Commencer une famille sur le tard a ses avantages, mais aussi ses inconvénients.

الزواج المتأخر له مميزات، ولكن أيضا له سلبيات.

'al-zwâj 'al-muta'ahir lahu mumayazât wa lakin 'aydan lahu salbiyât

Le sens de cette locution est modélisé par la FL (**Fact₀**) qui formalise les verbes de réalisation.

Fact₀ (commencer une famille) = « X » se marie

2) *La Locution « Attendre une Famille »* : Si nous observons la phrase suivante, nous remarquerons que la traduction proposée contredit le sens visé du phrasème français : D'après le traducteur automatique on comprend qu' « Elise attend l'arrivée de sa famille ».

Elise, dans son jupon plein de froment qui se gonfle en crinoline, a l'air d'attendre famille.

* *إليز في ثوب نسائي لها الكاملة من القمح تورم في قماش قطني، يبدو أنها تنتظر الأسرة.*

'îlîz, fi tawb nisâ 'î lahâ 'al-kamila min 'al-qamhî tawarum fi qumâš 'al-qṭnî, yabdû 'annahâ tantaḡîr 'al-' usra

Pourtant, « attendre une famille » est une locution quasi-figée qui signifie « être enceinte ».

Attendre une famille ≡ être enceinte

Elise, dans son jupon plein de froment qui se gonfle en crinoline, a l'air d'attendre famille.

إليز، في ثوبها النسائي المليء بالقمح و المنتفخ بقطع القماش القطني، تبدو وكأنها حامل.

'ilīz, fī ṭawbihâ 'al-nisâ'î 'al-malî' bi 'al-qamḥ wa 'al-muntafiḥ biqīṭa' 'al-qumâš 'al-quṭnî, tabdû wa ka'anahâ **ḥâmil**

Cette locution verbale est employée pour décrire un état, c'est pour cette raison qu'elle pourrait être modélisée par la FL d'adjectivisation (A₀).

A₀(attendre une famille) = être enceinte

3) *La Locution « Cordon Bleu »* : Par opposition aux deux locutions précédentes, « cordon bleu » est une locution complètement figée dont le sens résulte de la combinaison de ses deux éléments. Une combinaison qui mène à l'effacement total des sens premiers des éléments de la locution pour faire naître un nouveau sens entièrement différent. La locution « cordon bleu » est employée souvent dans un contexte gastronomique afin de qualifier un cuisinier habile. Mais comme nous apercevons par la suite que le traducteur automatique n'arrive pas à exprimer ce sens. Il présente plutôt la translittération arabe de la locution française.

Ma femme est un véritable cordon bleu.

* زوجتي الحقيقي كوردون بلو

Zawjatî 'al-ḥaqîqî **kurdun blû**

Parfois, c'est la traduction littérale qui est mentionnée, comme c'est le cas avec l'outil [Bing Traduction].

Ma femme est un véritable cordon bleu.

* زوجتي هي الطوق الأزرق الحقيقي.

zawjatî hiya 'al-ṭûq 'al-'azraq 'al-ḥaqîqî

Effectivement, la locution « cordon bleu » signifie « une personne qui est habile en cuisine ». Autrement dit, elle pourrait être paraphrasée par la collocation « cuisinier habile » qui est traduite en arabe par « الطاهي الماهر ». Reprenons la phrase précédente :

Ma femme est un véritable **cordon bleu**.

زوجتي طاهية ماهرة حقيقية.

zawjatî ṭâhiya mâhira ḥaqîqiyya

Sur le plan formel, la locution « cordon bleu » est formalisée par la FL (Ver) qui dénote le sens 'tel qu'il le faut'.

Ver (cordon bleu) = cuisinier habile

4) *La Locution « Coq du Village »* : Le dernier phrasème traité au cours de la présente recherche est « coq du village » qui est une locution complètement figée. Il est employé pour désigner « l'homme admiré par les femmes ». Il est paraphrasé par « le galant homme ».

Un coq du village ≡ un galant homme

Cependant, ce sens n'est pas exprimé par [**Yandex**], examinons l'exemple suivant :

Le Coq du village est un film italien réalisé par Alessandro Blasetti sorti en 1964.

* *الديك القرية الإيطالية الفيلم من إخراج اليكسندر بلاستي وقد صدر في عام 1964.*

'al-dîk 'qariya 'al-îtâlîyya 'al-film min 'hrâj 'aliksandar blâstî wa qad şadara fî 'âm 1964

Comme c'est clair, le phrasème « le coq du village » est traduit littéralement par « ديك القرية ». Ce qui souligne que la base de données lexicographique de l'outil de traduction n'arrive pas à gérer cette locution. En employant le MTLIL, la phrase citée ci-dessus sera alors traduite par :

Le Coq du village est un film italien réalisé par Alessandro Blasetti sorti en 1964.

معيود النساء هو فيلم إيطالي إخراج اليكسندر بلاستي وقد ظهر عام 1964.

ma 'abûd 'al-nisâ' huwa film 'îtâlî 'hrâj 'aliksandar blâstî wa qad zahara 'âm 1964

Sur le plan informatique, la locution « coq du village » est modélisée par la FL (**Bon**) qui connote le sens 'bon'.

Bon (coq du village) = galant homme

5 L'IMPLÉMENTATION DES LOCUTIONS AU SEIN DU MTLIL

Dernièrement, nous avançons quatre exemples dont deux arabes et deux français de l'implémentation informatique des locutions dans le MTLIL.

	Langue source	Langue cible
Locution	يسد عين الشمس	Très grande quantité
Paraphrase	"س" له حجم يتجاوز الحجم المعتاد	Le volume de « X » dépasse les normes
Fonction lexicale	كبير جدا = (يسد عين الشمس) Magn	Magn (quantité) = très grande ~

Figure 1 : La modélisation de la locution « يسد عين الشمس »

	Langue source	Langue cible
Locution	ذهب وراء الشمس	Subir une forte punition
Paraphrase	"س" تعرض لعقاب شديد بواسطة "ص"	« X » a subi une forte punition déterminée par « Y ».
Fonction lexicale	Oper + Magn (ذهب وراء الشمس) = تعرض لعقاب شديد	Oper (punition) + Magn (punition) = subir une forte punition

Figure 2 : La modélisation de la locution « ذهب وراء الشمس »

	Langue source	Langue cible
Locution	Cordon bleu	الطاهي الماهر
Paraphrase	« X » cuisine d'une manière très bonne qui plaît à « Y ».	"س" يطهو بطريقة جيدة جدا تعجب "ص".
Fonction lexicale	Ver (cordon bleu) = cuisinier habile	Ver (طاهي) = ماهر ~

Figure 3 : La modélisation de la locution « cordon bleu »

	Langue source	Langue cible
Locution	Coq du village	معبود النساء
Paraphrase	« X » est un homme. « Y » est une femme. « Y » aime beaucoup « X ».	"س" رجل. "ص" أنثى. "ص" يحب كثيرا "س".
Fonction lexicale	Bon (homme) = galant ~	Bon (رجل) = معبود النساء ~

Figure 4 : La modélisation de la locution « coq du village »

Comme nous apercevons dans les Figures de (1) à (4), l'implémentation des locutions dans le MTLIL consiste à créer des fiches où la locution avec la paraphrase qui l'explique et la fonction lexicale qui la formalise dans la langue source ainsi que la langue cible sont mises en parallèle.

6 CONCLUSION

Pour conclure, la traduction automatique des locutions est un des problèmes linguistiques mettant en question la performance et l'efficacité des applications du TALN. Le système informatique adopté actuellement par les outils de traduction automatique disponibles en ligne n'arrive pas à gérer les phrasèmes figés et quasi-figés. Leur construction compliquée nécessite un traitement spécifique pour déchiffrer leur sens caché. Nous avons démontré au cours des pages précédentes que la Théorie Sens-Texte paraît la plus capable à remédier au problème des locutions grâce à son système basé sur la double analyse linguistique et informatique des unités lexicales et des relations qui les unissent.

En adoptant les principes du modèle Sens-Texte (MST), nous avons conçu un modèle de traitement lexico-informatique des locutions (MTLIL). Neuf locutions arabes et françaises ont été traitées par le MTLIL au cours de la présente recherche. Les atouts du modèle proposé résident dans l'analyse linguistique multistratale prenant la sémantique comme point de départ, l'étude du contexte et la modélisation du sens plutôt que de la forme.

Finalement, reste à souligner qu'à travers cette étude nous avons essayé d'attirer l'attention des chercheurs vers un domaine de recherche si riche, celui de l'informatisation des syntagmes non libres dans les langues naturelles. Ce domaine ne groupe pas seulement les locutions, mais aussi les proverbes et les clichés linguistiques. Nous invitons alors les chercheurs à se lancer dans ce champ afin de créer un langage formel universel qui soit apte à gérer les différents phénomènes et structures linguistiques dans les différentes langues humaines.

NOTES

(1) <https://translate.google.com/?hl=fr>

(2) http://www.reverso.net/text_translation.aspx?lang=FR

- (³) <https://www.bing.com/translator>
- (⁴) <https://translate.yandex.com/>
- (⁵) <https://traduction.babylon-software.com/>
- (⁶) Le « lexème » est un mot monosémique qui n'a qu'un seul signifié et un seul signifiant.
- (⁷) Le terme « lexie » désigne chaque acception d'une unité lexicale polysémique (ou « vocable » selon la terminologie de la TST) ayant un sens bien précis.
- (⁸) Cet exemple a été cité par Mel'čuk dans « Collocations dans le dictionnaire ».
- (⁹) L'adjectif possessif renvoie à « locution ».
- (¹⁰) L'astérisque signale la fausse traduction.
- (¹¹) La FL(**Oper**) modélise les verbes supports.
- (¹²) « X » renvoie à une personne.

RÉFÉRENCES

- [1] S. Bouchaddakh, *Applications des principes de la lexicologie explicative et combinatoire à la construction d'un dictionnaire bilingue informatisé français – arabe*. Thèse de doctorat. Université de Sherbrooke, 2010.
- [2] A. Clas, I. Mel'čuk et A. Polguère, *Introduction à la lexicologie explicative et combinatoire*. Editions Duculot, 1995.
- [3] Sous la direction de F. Corblin et C. Gardent, *Interpréter en contexte*. Lavoisier, 2005.
- [4] P. Enjalbert, *Sémantique et traitement automatique du langage naturel*. Lavoisier, 2005.
- [5] H. Hamzé, *Collocations et maîtrise des langues*, in www.vob.edu.bh/uob_files/436/issue17/17_227_232.pdf
- [6] A.L. Jousse, *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales*. Thèse de doctorat. Université de Montréal et Université Paris Diderot, 2010.
- [7] I. Mel'čuk et J. Milicévic, *Introduction à la linguistique*, volume 1. Hermann, 2014.
- [8] I. Mel'čuk, *Collocations dans le dictionnaire*, in <http://olst.ling.umontreal.ca/pdf/Collocations-szende.pdf>.
- [9] I. Mel'čuk, *Vers une linguistique Sens-Texte*, in <http://olst.ling.umontreal.ca/pdf/melcukColldeFr.pdf>
- [10] J. Milicévic, *La paraphrase Modélisation de la paraphrase langagière*. Editions Peter Lang, 2007.
- [11] A. Nasr, *Un modèle de reformulation automatique fondé sur la Théorie Sens – Texte*. Thèse de doctorat. Université Paris7, 1996.
- [12] M. S. Pause, *Structure lexico-syntaxique des locutions du français et incidence sur leur combinatoire*. Thèse de doctorat. Université de Lorraine, 2017.
- [13] S. Popovic, *Paraphrasage des liens de fonctions lexicales*. Mémoire. Université de Montréal, 2003.
- [14] R. Salem, *Modélisation de la polysémie : approche contrastive arabe – français basée sur la Théorie Sens – Texte*. Thèse de doctorat. Université d'Alexandrie, 2017.
- [15] أحمد مختار عمر، صناعة المعجم الحديث، عالم الكتب، 2009.
- [16] حلمي خليل، مقدمة لدراسة علم اللغة، دار المعرفة الجامعية، 2000.
- [17] مجدي إبراهيم محمد، بحوث في علم الدلالة بين القدماء و المحدثين، الهيئة المصرية العامة للكتاب، 2013.

BIOGRAPHY**Racha Mohammad Salem**

Lecturer, French Language Department, Faculty of Arts, Alexandria University. Computer linguistics. She holds a doctorate in 2017 and the subject of "Automatic treatment of multiple meanings: a comparative study between Arabic and French." While in France on a scientific mission in 2016, she participated in several workshops and training courses on computer meanings taught by the University of Lorraine's ATILF.

A Computerized Model for Automatic Translation of Arabic and French Locations According to TST Principles

Racha Mohammad Salem

Department of French, Faculty of Arts, Alexandria University, Egypt

Salem.racha@live.com

Abstract: *Given their complex construction, the phrases are one of the natural language riddles that most computer systems can not decipher until today. Through the review of the free automatic translation tools available online, we will describe the problem in detail, then we will propose as a solution the development of a computer model based in principle on the lexico-semantic analysis of the locutions. We will adopt as theoretical framework the Meaning-Text Theory which seems the most able to overcome the challenge of locutions thanks to a formal system based on a double linguistic-computer analysis of natural languages where semantics is a priority.*

Keywords : *Phrase, paraphrase, lexical functions, computer model, automatic translation, semantic representation.*

نموذج حاسوبي للترجمة التلقائية للمتلازمات اللفظية العربية و الفرنسية وفقا لمبادئ نظرية المعنى-النص

رشا محمد سالم

قسم اللغة الفرنسية، كلية الآداب، جامعة الإسكندرية، مصر

salem.racha@live.com**ملخص :**

بالنظر إلى بنيتها المعقدة ، فإن المتلازمات اللفظية هي واحدة من الغاز اللغات الطبيعية التي لا تستطيع معظم أنظمة الكمبيوتر فكها حتى اليوم. من خلال مراجعة أدوات الترجمة التلقائية المجانية المتوفرة على الإنترنت ، سنصف المشكلة بالتفصيل ، ثم سنقترح كحل تطوير نموذج حاسوبي يستند مبدئياً إلى التحليل اللغوي للدلالي لهذه المتلازمات. سوف نعتمد كإطار نظري نظرية المعنى-النص التي تبدو الأكثر قدرة على التغلب على هذه المشكلة من خلال نظام حاسوبي قائم على تحليل لغوي مزدوج للغات الطبيعية تكون الأولوية فيه للدلالة الخاصة بكل متلازم لفظي.

الكلمات المفتاحية : عبارة ، إعادة صياغة ، وظائف معجمية ، نموذج حاسوبي ، ترجمة تلقائية ، تمثيل دلالي.