

# L'analyse prédicative : un outil pour l'évaluation de la traduction des textes bilingues alignés

Inès Sfar  
Université de Tunis  
TIL (00/UR/0201)-Université de Manouba



Synergies Tunisie n° 2 - 2010 pp. 99-113

**Résumé :** *Le présent article est centré sur une méthode proposée pour l'alignement des corpus bilingues ou multilingues dans le cadre d'une théorie spécifique, celle des trois fonctions primaires des unités lexicales. Après avoir présenté les principes de cette théorie lexicaliste appliquée à des corpus monolingues, nous dégagons les principes de la méthode basée sur l'analyse prédicative. Nous exposons ensuite les résultats d'une telle démarche et son application à un corpus bilingue en montrant qu'il est possible d'extraire les unités prédicatives, argumentales et actualisatrices à partir de corpus bilingues alignés et de construire des correspondances très fiables. Le croisement de l'analyse prédicative et de l'alignement de corpus se présente comme un outil d'évaluation de la qualité de la traduction.*

**Mots-clés :** *alignement, analyse prédicative, lexique bilingue, terminologie linguistique, fonctions primaires*

**Abstract:** *The present article focuses on a method suggested for the alignment of bilingual or multilingual corpora within the framework of a specific theory, that of the three primary functions of lexical units. After having presented the principles of this lexicalist theory applied to monolingual corpora, we withdraw the principles of the method based on the predicative analysis. We then describe the results of such an approach and its application to a bilingual corpus by demonstrating that it is possible to extract predicates, arguments and actualisers from aligned bilingual corpora and to build highly reliable correspondences. The junction of predicative analysis and corpus alignment presents itself as a tool for the evaluation of translation quality.*

**Keywords:** *alignment, predicative analysis, bilingual lexicon, linguistic terminology, primary functions*

## Introduction

L'utilité des corpus alignés bilingues ou multilingues n'est plus à démontrer puisqu'ils permettent d'extraire automatiquement des équivalents de traduction, des collocations et des lexies complexes. Ce qui permet de normaliser la terminologie dans un champ donné, notamment en science du langage, discipline qui souffre énormément d'une situation de flou terminologique. La tâche de construction d'un grand corpus bilingue aligné est très importante pour la traduction automatique. Ce corpus est décrit comme un

ensemble de paires de phrases bilingues. Aujourd'hui, un tel corpus aligné français-arabe n'est pas disponible, mais les travaux de traduction de textes linguistiques du français vers l'arabe<sup>1</sup>, ont permis de rassembler dans des bases de données (lexicographiques, terminologiques et textuelles) un matériau comparable très important constitué de documents traduits dans plusieurs domaines.

Nous proposons dans cet article une nouvelle méthode pour l'extraction de lexiques bilingues de corpus alignés à partir des fonctions assignées aux unités lexicales et terminologiques dans le discours. Pour ce faire, nous détaillons les spécificités de l'approche basée sur les trois fonctions primaires en corpus monolingue. Ensuite, nous évaluons la méthode de l'analyse prédicative sur un corpus bilingue français-arabe<sup>2</sup>, et montrons comment la combinaison de l'analyse prédicative et de l'alignement des textes permet d'évaluer la qualité de la traduction et d'améliorer de façon significative les résultats.

## 1. Les outils théoriques et méthodologiques

### 1.1. Les trois fonctions des unités lexicales

Les unités linguistiques, qu'elles soient lexicales ou terminologiques, peuvent avoir trois fonctions dans la phrase : en tant que prédicats, arguments ou actualisateurs. Ce constat est valable pour le français comme pour l'arabe. Le texte arabe étant la traduction du texte français, le repérage des unités linguistiques se fait sur la même base. Nous présentons dans ce qui suit les principes méthodologiques de l'analyse prédicative des textes spécialisés en français.

#### 1.1.1. La fonction prédicative

Tout comme les énoncés du discours général, les discours spécialisés peuvent être abordés sous l'angle de la prédication. On leur applique alors les mêmes critères préconisés pour l'analyse des discours de la langue générale. C'est dans ce sens que nous faisons la distinction entre :

- la prédication élémentaire, qu'on retrouve au niveau de la phrase simple et qui concerne le prédicat élémentaire chargé de sélectionner ses arguments, comme dans :

« Les définitions peuvent **varier** considérablement pour un même vocable »

- la prédication de second ordre, qui est spécifique aux phrases complexes et qui repose sur des arguments phrastiques, comme dans :

« Une phrase est **un lieu de flou** si elle comporte un ou plusieurs **prédicats flous**, c'est-à-dire dont l'intension n'est que partiellement **spécifiable**, **de telle sorte que** l'extension correspondante n'est pas univoquement **déterminée** ».

Le discours spécialisé, comme tout discours, comporte des prédicats spécifiques : il se construit à partir d'un ensemble de mots lexicaux (nom, adjectif, verbe, etc.) et grammaticaux (articles, pronoms personnels, déterminants démonstratifs, etc.) qui construisent un réseau de rapports anaphoriques et cataphoriques formant l'échafaudage qui le soutient. Par conséquent, nous avons relevé un nombre important d'unités lexicales qui fonctionnent comme des prédicats anaphoriques :

« En revanche *Les freins de ma voiture ont été refaits* a du sens parce qu'on peut dire dans quelles conditions **cela** est vrai ».

Une deuxième spécificité du discours linguistique, et non des moindres, est celle de la relation caractérisante ou définitionnelle qui unit les notions, les termes et les concepts. Tout le texte en linguistique est structuré sur la base d'un jeu définitionnel dans lequel les deux arguments de la relation se transforment en prédicats :

« La vérité langagière *est* une vérité prise en charge par un sujet »

« Les relations entre phrases que l'on cherche à calculer, indépendantes de la situation discursive et valables quel que soit le locuteur dont on a sollicité le jugement, ne *sont* rien d'autre que des relations qui s'établissent en vertu du sens ».

### 1.1.2. La fonction argumentale

Etant donné que les arguments sont sélectionnés par les prédicats, la nature du prédicat a donc un impact sur la nature de l'argument et sur sa forme. Si la prédication est élémentaire, les arguments sont sélectionnés en conséquence. Par exemple :

« (...) d'incontestables liens *unissent* le sens et la vérité ».

Si la prédication est complexe, les arguments sélectionnés sont de nature prédicative, notamment dans le cas de la prédication de second ordre :

« Un énoncé a du sens *dès lors qu'*il est possible d'énumérer les conditions dans lesquelles il peut être déclaré vrai et, *conséquemment*, dans lesquelles il peut être déclaré faux ».

Dans d'autres cas, plus spécifiques au discours des sciences du langage, nous relevons des arguments formés d'une ou plusieurs phrases successives. En effet, la spécificité des textes linguistiques est qu'ils sont structurés d'une manière binaire : exemples et commentaires. Il arrive que les exemples soient entièrement repris dans le commentaire, comme dans :

« *Deux et deux font quatre* ou bien *La racine de 81 est 9* ou encore *Le plus grand dénominateur commun de 18 et 24 est 6* sont des propositions vraies par les seules règles que l'arithmétique se donne ».

Par conséquent, ces phrases sont reconnues comme faisant partie des arguments du prédicat *propositions*. Les exemples de ce type sont très nombreux. On peut en citer un autre :

« Ce type de vérité est fort éloigné de la vérité de *Les chimpanzés sont des singes* »

Il arrive aussi que ce même argument se transforme en prédicat (cf. § 2.2.).

### 1.1.3. La fonction actualisatrice

Les actualisateurs sont des éléments qui permettent d'inscrire les prédicats et les arguments dans l'énonciation en véhiculant les différentes catégories grammaticales de la personne, du genre, du nombre, du temps, de la diathèse, de la modalité et de l'aspect. Ils peuvent être de nature grammaticale (préposition, déterminant, flexion verbale, nominale ou adjectivale, etc.) ou lexicale (adjectifs modifieurs, adverbes, morphèmes dérivationnels, etc.). Un texte du domaine des sciences du langage représente un pourcentage élevé de présence d'actualisateurs. La raison est la suivante : le texte linguistique est un texte à contenu informatif, scientifique, de type argumentatif et

illustratif. Toutes ses propriétés rhétoriques font de l'usage des actualisateurs (simples ou complexes) une particularité du discours des sciences du langage. L'exemple qui revient très souvent est celui du verbe *être*, employé comme verbe support de prédicats nominaux ou adjectivaux. Cette fréquence d'emploi peut s'expliquer par le fait qu'il s'agit d'un texte basé sur des relations définitionnelles « caractérisantes ». Le deuxième exemple est celui des prépositions composées et des conjonctions de subordination.

Il n'en demeure pas moins que la notion d'analyticité est **centrale** pour notre propos.

و لا يمنع ذلك من اعتبار أنّ مفهوم التّحليليّة مركزيّ لموضوع حديثنا

Préd : centrale (arg 1 : notion d'analyticité ; arg 2 : propos)

مركزيّ (مفهوم التّحليليّة/موضوع حديثنا)

## 1.2. La technique d'alignement automatique

L'alignement automatique de textes consiste en la mise en correspondance entre des unités logiques de textes alignés, c'est-à-dire un ensemble de textes de langues différentes, constitué d'un texte original et de sa ou ses différentes traductions. L'alignement peut être réalisé à différents niveaux selon plusieurs méthodologies (cf. Kraif, 2002 ; Zimina, 2002-2004 ; Langlais, 1997, Véronis, 2000 ; etc.).

Selon Véronis (2000 : 158), deux méthodes principales ont été utilisées pour l'alignement de phrases : l'ancrage lexical et la longueur des phrases. Elles reposent toutes les deux sur une information interne, c'est-à-dire que l'information nécessaire à l'alignement se trouve dans les textes à aligner eux-mêmes, et utilisent des algorithmes de calcul. Dans un cas, il s'agit d'aligner les mots entre les deux textes. Démarche difficile, certes, mais qui peut conduire à un alignement satisfaisant au niveau des phrases. Dans l'autre cas, il s'agit d'utiliser la longueur des phrases comme base pour l'alignement. Ce modèle probabiliste est construit à partir de la mesure de dissimilarité entre les phrases des textes à aligner. Ces méthodes demeurent malheureusement « grossières » et imparfaites puisqu'elles utilisent des critères strictement formels. La méthode des *cognats*, par exemple, donne certainement de bons résultats pour les textes français-anglais, mais elle trouve toutefois ses limites lors de l'alignement de couples de langues non apparentées, telles que le français et l'arabe.

L'introduction de connaissances linguistiques dans l'alignement automatique<sup>3</sup>, même si elle semble relativement coûteuse et difficile à réaliser puisqu'elle n'est pas indépendante des langues, peut améliorer les méthodes purement ou principalement statistiques.

## 2. L'analyse prédicative dans les textes bilingues alignés

Nous montrons dans ce qui suit l'importance du croisement des textes bilingues alignés et de l'analyse prédicative ; ce qui permettrait de vérifier l'alignement des unités prédicatives, argumentales et actualisatrices dans les textes bilingues et d'affiner le degré de granularité de l'alignement.

### 2.1. Alignement des unités prédicatives

L'exploration d'une nouvelle méthode d'alignement et d'extraction part d'un souci de précision et d'exhaustivité. En effet, le seul repérage des unités lexicales et terminologiques à partir des catégories grammaticales auxquelles elles appartiennent et de leur forme présente un outil insuffisant pour deux raisons : la première est que la catégorie

grammaticale reste tributaire de l'emploi de l'unité dans le discours et que le passage d'une langue à une autre peut impliquer un changement de catégorie ; la seconde est que la forme de l'unité lexicale ou terminologique est variable selon les emplois, puisqu'elle peut être simple ou composée, monolexicale ou polylexicale (plus particulièrement si on prend le cas des collocations spécialisées ou des unités phraséologiques qui représentent dans la plupart des cas une extension de l'unité monolexicale, par exemple *accent*, *accent tonique*, *accent atone*, *accent aigu*, *accent grave*, etc.<sup>4</sup>).

Notre objectif étant de combiner les méthodes structurelles d'alignement et la description linguistique, on cherche à dépasser le cadre des mots et des catégories grammaticales afin d'optimiser les techniques d'alignement et les résultats escomptés. En effet, on se base sur un principe fondamental selon lequel les langues catégorisent différemment les objets du monde. Ce qui importe pour l'alignement, ce sont les unités logiques sélectionnées pour la mise en correspondance. Or, les propriétés structurelles, même si elles présentent quelques diversités, sont fondamentalement et dans la plupart des cas identiques. Si nous prenons l'exemple suivant, nous remarquons que nous avons la même relation prédicative et le même contenu informatif dans les deux phrases, qu'on peut représenter schématiquement de la manière suivante :

Préd : caractères de la relativité (arg 1 : vérité ; arg 2 : langage naturel)

صفات النسبية (حقيقة / لغة طبيعية)

Français	Arabe
Dans le langage naturel, la vérité a tous les caractères de la relativité.	إنَّ للحقيقة في اللغة الطبيعيَّة جميع صفات النسبيَّة

Tableau 1 : Alignement de phrases

Cette méthode nous permet d'atteindre un degré de finesse très important au niveau de la granularité de la correspondance lexicale entre ces deux corpus bilingues étant donné que le niveau de structuration de l'énoncé choisi comme base pour l'alignement est la phrase. Le recours aux informations linguistiques impliquant l'analyse et la comparaison du contenu permet d'obtenir un alignement phrastique de qualité.

Français	Arabe
Le flou peut être dû aussi à des phénomènes d'« expansion sémantique », ainsi pour les prédicats « transformatifs », ceux qui marquent le passage d'un état à un autre. Ces prédicats ne sont strictement vérifiés que si l'état nouveau est effectivement atteint. Mais, par expansion sémantique, ils peuvent être tenus pour appropriés à des distances fort variables du seuil séparateur. Ainsi, le train Strasbourg-Paris, selon l'humeur du contrôleur, est déclaré arriver en gare de Paris-Est quand les quais sont en vue ou bien quand le train passe, en pleine lancée, à la Villette, à Pantin ou même à Noisy-le-Sec.	إنَّ الضبابيَّة يمكن أن ترجع أيضا إلى ظواهر "توسُّع دلالي" كما هو الثَّمَان بالنسبة إلى المسندات "التحويليَّة" و هي التي تدلُّ على الانتقال من حالة إلى أخرى. و هذه المسندات لا يمكن إثباتها بدقَّة إلا إذا تمَّ فعلا بلوغ الحالة الجديدة. و لكن يمكنها بالتوسُّع الدلالي أن تلازم و مسافات جَدَّ متنوِّعة من الحدِّ الفاصل. مثال ذلك قطار ستراسبورج-باريس الذي يتمُّ الإعلان عن وصوله محطة باريس الشرقيَّة حسب مزاج المراقب عندما تظهر المناظر الأرضيَّة أو عندما يمرُّ القطار مسرعا بلافاليات أو بثمان أو حتَّى نوازي لي ساك.

Tableau 2 : Alignement de paragraphe

De nombreuses équivalences traductionnelles, observées en corpus, ne peuvent guère être isolées de leur contexte, surtout si on se situe à un niveau de granularité fin, comme celui des unités lexicales. En effet, de nombreuses unités lexicales perdent

leur autonomie et se restructurent différemment dans le texte traduit, de manière à respecter les lois de la signification, de la transparence et de la fonctionnalité, définitoires d'un texte scientifique.

Or, pour que les corpus bilingues alignés puissent constituer un matériau fiable pour l'observation contrastive des langues, il faut réussir à répertorier les problèmes traductionnels et les confusions qu'ils impliquent au niveau du décodage du texte source. En effet, un des avantages du travail sur corpus est celui d'avoir accès à un très grand nombre de textes, qui peuvent constituer un bon outil de comparaison. De cette confrontation naissent des régularités significatives au niveau des textes traduits. La présence régulière et massive de ces phénomènes traductionnels dans les textes cibles montre qu'il ne s'agit pas de phénomènes accidentels liés à une théorie linguistique particulière, à un style donné ou à une méthodologie spécifique. Nous en fournissons quelques exemples relevés à partir de l'analyse de notre corpus dans ce qui suit.

## 2.2. Elaboration de concordanciers bilingues<sup>5</sup>

Les concordanciers permettent d'explorer l'utilisation d'un mot ou d'une expression dans un type de texte donné. Un même mot, ou une même expression, peut avoir différents sens selon les emplois, et par conséquent plusieurs traductions qui correspondent aux diverses acceptions. Les concordanciers bilingues permettent à la fois de représenter les différents emplois d'une unité lexicale ou terminologique dans les deux langues, et de déterminer les différentes fonctions assignées à cette unité lexicale dans le texte. Etant donné qu'ils sont dans la plupart des cas adossés à des corpus textuels bilingues, ils servent à donner une vue d'ensemble sur les particularités du discours en question et sur les variations subies.

Nous fournissons dans ce qui suit deux exemples de concordanciers bilingues d'unités prédicatives et d'unités argumentales. Il serait possible d'élaborer des concordanciers des unités actualisatrices.

Il s'y ajoute aussi que bon nombre d'énoncés, pourtant bien formés, n'ont par nature aucune <u>valeur de vérité</u>	و يُضَافُ إِلَى ذَلِكَ أَنَّ عَدِيدَ اللَّفِظَاتِ رَغْمَ جُودَةِ صِيَاقِهَا لَيْسَ لَهَا بِطَبِيعَتِهَا أَيُّ قِيَمَةٍ حَقِّ
Pourtant on accordera aussi, toujours en vertu du « bon sens », qu'un assemblage qui n'a pas de sens ne peut pas avoir non plus de <u>valeur de vérité</u>	و مَعَ ذَلِكَ نَقِيلُ أَيْضًا، بِحُكْمِ "الْبَهَاةِ" دَوْمًا، بَأَنَّ وَصْلًا بَدُونَ مَعْنَى لَا يُمْكِنُ أَنْ تَكُونَ لَهُ كَذَلِكَ قِيَمَةٌ حَقِّ
Un contexte est déclaré opaque quand il se prête à une lecture telle que la substitution d'expressions co-référentielles n'en préserve pas la <u>valeur de vérité</u>	إِنَّ سِيَاقًا مَا يُوَصَفُ بِأَنَّهُ لِاشْفَافٍ عِنْدَمَا يَسْمَحُ بِقِرَاءَةٍ لَا يَحْفَظُ فِيهَا تَعْوِيزُ التَّعَابِيرِ ذَاتِ الْمَرْجِعِيَةِ الْمَشْتَرَكَةِ عَلَى قِيَمَةٍ حَقِّقَتِهَا
On appellera <i>univers actuel</i> d'un locuteur donné à un moment déterminé du temps l'ensemble des propositions auxquelles ce locuteur attribue effectivement une <u>valeur de vérité</u> .	يَكُونُ الْمَحِيطُ التَّقْدِيرِي لِمَتَكَلِّمٍ مَا فِي فِتْرَةٍ زَمَنِيَّةٍ مَحْدَدَةٍ مَجْمُوعَ الْأَقْوَالِ الْقَابِلَةِ لِلتَّقْرِيرِ مِنْ قِبَلِهِ، أَيَّ إِنَّ بِإِمْكَانِهِ أَنْ يَدْفُقَ ظُرُوفَ حَقِّقَتِهَا.
une proposition attribuée à quelque « image d'univers », que je me hâte de distinguer de mes propres croyances, où la proposition en question ne saurait prendre aucune <u>valeur de vérité</u>	قَوْلًا مَنسُوبًا إِلَى "صُورَةٍ مَحِيطٍ" مَا، أَسْرَعَ إِلَى تَمْيِيزِهَا عَنِ مَعْتَقَدَاتِي الْخَاصَّةِ، حَيْثُ لَا يُمْكِنُ لِلْقَوْلِ الْمَعْنَى أَنْ تَكُونَ لَهُ قِيَمَةٌ حَقِّقَةٌ
Quoique les conditions de vérité ne soient pas en cause, il m'est impossible d'attribuer à cet énoncé aucune <u>valeur de vérité</u>	فَرِغَ مِنْ ظُرُوفِ الْحَقِيقَةِ لَيْسَتْ مَحَلَّ شَكِّ، يَسْتَحِيلُ عَلَيَّ مَنَحُ هَذَا اللَّفِظِ أَيُّ قِيَمَةٍ حَقِّ
Au lieu de conférer lui-même à une proposition une <u>valeur de vérité</u> , le locuteur peut aussi la situer dans quelque univers qu'il évoque.	يُمْكِنُ كَذَلِكَ لِلْمَتَكَلِّمِ أَنْ يَمُوضِعَ قَوْلًا مَا فِي مَحِيطٍ مَا يَذْكُرُهُ بَدَلًا أَنْ يَضْفِي عَلَيْهِ بِنَفْسِهِ قِيَمَةٌ حَقِّ.

Tableau 3 : Concordancier bilingue français-arabe de la fonction prédicative de la séquence « valeur de vérité »

Outre le sens, le concordancier permet également d'étudier la syntaxe des unités lexicales. En effet, à partir du concordancier présenté ci-dessous, nous pouvons dégager une série d'observations au sujet des propriétés du texte source et de sa traduction :

- toutes les occurrences de l'unité polylexicale prédicative *valeur de vérité* ont le même correspondant en langue cible *قيمة حق* (nous reviendrons sur les cas de variations dans la dernière partie de cette étude § 3) ;
- les verbes supports employés avec cette unité sont différents. Nous pouvons donc en faire un inventaire exhaustif, si nous nous appuyons sur la totalité de l'ouvrage étudié : *avoir, préserver, attribuer, prendre, conférer, etc.* ;
- les arguments sélectionnés par ce prédicat monolexical sont les mêmes : *énoncé, proposition, assemblage, contexte, etc.*

Il s'y ajoute aussi que bon nombre d'énoncés, pourtant bien formés, n'ont par nature aucune valeur de vérité	و يُضَاف إلى ذلك أن عديد اللفيظات رغم جودة صياغتها ليس لها بطبيعتها أي قيمة حق
Qu'est-ce, en effet, pour un énoncé qu'avoir du sens ?	و فعلا فعلام يدل قولنا إن اللفيظ معنى ؟
Un énoncé a du sens dès lors qu'il est possible d'énumérer les conditions dans lesquelles il peut être déclaré vrai et, conséquemment, dans lesquelles il peut être déclaré faux.	فاللفيظ يكون له معنى، بمجرد أنه يمكن تعداد الظروف التي يتسنى فيها أن يُقال إنه حق، و تبعاً لذلك الظروف التي يمكن فيها أن يُقال إنه باطل.
Tout langage produit des énoncés valides en raison des seules règles que ce langage comporte.	إن كل لغة تصوغ تصوغ لفيظات صحيحة بحكم القواعد التي تتضمنها تلك اللغة دون غيرها.
Un énoncé est vrai pour quelqu'un.	فاللفيظ حق بالنسبة إلى شخص.
Quoique les conditions de vérité ne soient pas en cause, il m'est impossible d'attribuer à cet énoncé aucune valeur de vérité	فرغم أن ظروف الحقيقة ليست محل شك، يستحيل عليّ منح هذا اللفيظ أي قيمة حق
De tels énoncés sont pour moi disconvenants.	فمثل هذا القول غير ملائم بالنسبة إليّ.

Tableau 4 : Concordancier bilingue français-arabe de la fonction argumentale du terme « énoncé »

L'élaboration du concordancier permet de réaliser trois tâches :

- étudier la distribution de l'unité argumentale en question (ex : *énoncé*) en effectuant un repérage systématique de toutes les unités prédicatives qui la sélectionnent : nominale (*valeur de vérité*), adjectivale (*faux, vrai, disconvenant*) ou verbale (*produire*), etc. ;
- séparer les emplois de l'unité argumentale monolexicale et ses emplois en tant qu'unité collocationnelle (*énoncé injonctif, énoncé générique, énoncé contrastif*, etc.) ;
- vérifier la traduction du terme en question : la dernière ligne du concordancier nous montre clairement que le terme *énoncé* a été traduit en arabe par deux termes différents (*لفيظ/قول*).

### 2.3. Principe de l'analyse prédicative bilingue

L'analyse prédicative consiste en la description systématique de toutes les relations prédicatives à la fois au niveau intra-phrastique et inter-phrastique. A partir des trois fonctions primaires présentées ci-dessus, il est possible de construire un réseau de relations logico-sémantiques entre les phrases qui constituent le texte spécialisé et de sa traduction :

<p><u>La notion de grammaire repose traditionnellement sur le concept de compositionnalité. Cela veut dire que le sens d'une séquence est le produit de celui des éléments composants. Ainsi, le sens d'une phrase est facteur de celui de son prédicat et de celui de ses arguments. Le sens de la phrase suivante est compositionnel : L'enfant a mangé le gâteau.</u></p>	<p>يرتكز مفهوم النحو تقليدياً على متصور التركيبية. يعني ذلك أنّ معنى وصلة ما هو حاصل معنى العناصر المركبة لها. هكذا يكون معنى الجملة نتيجة معنى مُسنّدها و معنى مَعْمولاتها. إنّ أكل الولد كعكة معنى الجملة التّالية تركيبي :</p>
--	---

[gras : prédicat / souligné : argument / italique : actualisateur]

Tableau 5 : Analyse prédicative d'un fragment textuel bilingue

L'appariement de ces deux fragments et l'application du principe de l'analyse prédicative nous permet de comparer les deux textes (source et cible). Une première constatation peut être faite : elle concerne l'emboîtement des trois fonctions : les arguments, eux-mêmes actualisateurs, les prédicats eux-mêmes arguments, les actualisateurs présents systématiquement dans les syntagmes arguments, etc. La complexité du contenu prédicatif est représentée à travers ce réseau de relations. La seconde remarque concerne le degré de fidélité du texte cible par rapport au texte source. En effet, tous les prédicats repérés dans le texte en français sont rendus tels quels dans le texte arabe, excepté le prédicat anaphorique *cela*, suffisamment explicité en français mais rendu d'une manière implicite en arabe. Nous reviendrons sur cette question dans la troisième partie de cette étude.

La deuxième difficulté qui fait face au principe de l'analyse prédicative est celle de la hiérarchie prédicative. On parle de hiérarchie prédicative quand la phrase comporte plus d'un prédicat. Or, l'objectif du discours spécialisé est de présenter des concepts, de les commenter et d'en fournir des illustrations. La succession de phrases simples et complexes fait de la trame textuelle un ensemble de prédicats, se situant à des niveaux hiérarchiques différents. Cette notion de hiérarchie prédicative est observée particulièrement au niveau des phrases complexes. Par exemple, dans :

« Un énoncé a du sens **dès lors qu'**il est possible d'énumérer les conditions dans lesquelles il peut être déclaré vrai et, conséquemment, dans lesquelles il peut être déclaré faux »,

nous avons deux arguments phrastiques, qui représentent eux-mêmes des prédicats [arg 1 : énoncé ayant du sens / arg 2 : énumérer les conditions]. Ces deux phrases sont reliées par un prédicat hiérarchiquement supérieur *dès lors que* qui exprime une valeur tempo-causale.

Le deuxième cas de figure qui est très fréquent dans notre corpus concerne les hypothétiques en *si*. On pourra donc appliquer cette même analyse à l'exemple suivant :

Français	Arabe
Une phrase est un lieu de flou si elle comporte un ou plusieurs prédicats flous	الجملة تكون موضع ضبابية إذا تضمّنت مسندا ضبابياً أو أكثر



La relation entre les prédicats peut être plus complexe :

Français	Arabe
<p>Les « mondes possibles » eux-mêmes se prêtent à des conceptions diverses. On peut envisager un « monde possible » comme une totalité inconditionnée de faits non contradictoires, auquel cas le monde effectif apparaît comme un monde possible parmi une infinité d'autres. Ce critère de la non-contradiction confère au possible une extension infinie.</p>	<p>إنّ "العوالم الممكنة" ذاتها تخضع لتصورات متنوّعة. فيمكننا أن نتصور "عالما ممكنا" ككلّ لا مشروط من أحداث غير متناقضة. وفي هذه الحالة يبدو العالم الحقيقي كعالم ممكن من بين عدد لا متناه من العوالم. إنّ هذا المعيار، معيار عدم التناقض، يجعل الممكن لانهائي التأمط.</p>

Tableau 6 : Exemple de hiérarchie prédicative

La hiérarchie prédicative qui découle du principe de l'analyse prédicative joue un rôle très important dans la mise en relation du texte source et de sa traduction. En effet, si le traducteur n'arrive pas à défaire la relation prédicative dans sa complexité dans le texte source, le transfert dans la langue cible en sera inéluctablement affecté. C'est ce que nous montrerons dans ce qui suit.

### 3. Evaluation de la traduction

Les corpus alignés fournissent un moyen de vérification de la qualité de la traduction. En effet, la confrontation systématique d'un texte source et d'un texte cible permet de détecter les usages linguistiques qui résultent de l'activité de traduction. On peut fournir dans ce qui suit un exemple parfait de traduction : respect de la forme, de la syntaxe, du sens et du style :

Français	Arabe
<p>Certes, selon le locuteur, ces conditions seront plus ou moins précises. Pour mon garagiste, refaire les freins a des conditions de vérité beaucoup plus spécifiques que pour moi : il énumère là toute une série d'opérations (dépose des roues, démontage du disque...) dont je dois bien avouer que je me désintéresse; pour moi <i>refaire les freins</i>, c'est conduire ma voiture chez le garagiste et la reprendre le soir, non sans avoir réglé une facture de quelque 1 000 ou 2 000 francs qui porte, pour ma consolation, la mention « réparation des freins ». Si d'aventure mon garagiste est honnête, j'ai de surcroît la satisfaction de n'avoir plus à appuyer des deux pieds sur la pédale pour obtenir un semblant de résultat. Voilà qui fait voir toute la relativité du sens, à quoi nous reviendrons en évoquant la notion d'univers.</p>	<p>إنّ هذه الظروف تكون بالتأكيد متفاوتة الدقة حسب المتكلم. فبالنسبة إلى صاحب الورشة، مقارنة بي، لعملية إصلاح الفرامل ظروف حقيقة أكثر تخصيصاً: فهو يعدد في ذلك سلسلة كاملة من العمليات (نزع العجلات، فك الاسطوانة...) و هو ما يجب أن أعتزف أنني لا أعتد به. فبالنسبة إلى إصلاح الفرامل يعني الذهاب بسيّارتي إلى الورشة واسترجاعها مساءً. ولا يحصل ذلك بدون أن أكون قد دفعْتُ فاتورة بالف أو ألفي فرنك تتضمن -عزاًء لي- عبارة "إصلاح فرامل". وإذا صادف أن كان صاحب الورشة من الثقات، كان لي أيضاً الرضا بأن لا أكون في حاجة إلى الضنظ على دواصة الفرامل بالرّجلين معاً للحصول على ما يشبه النتيجة. إنّ هذا لمّا يظهر كامل نسبية المعنى، و هو ما سنعود إليه عند التّعرّض إلى مفهوم المحيط.</p>

Tableau 7 : Exemple de traduction fidèle au texte source

Toutefois, de pareils exemples ne sont pas très fréquents. Il est vrai que cette hypothèse reste tributaire de la démarche adoptée par le traducteur : « sourcier » ou « cibliste », selon la terminologie de J.-R. Ladmiral. Nous classerons les écarts par rapport au texte source en trois catégories : des variations formelles, des déperditions sémantique et syntaxique et un écart stylistique.

### 3.1. Variations formelles

Les textes alignés étant basés sur un texte source et sa traduction, la langue source peut influencer la langue cible et par conséquent engendrer des phénomènes de variations formelles puisqu'il arrive que les mêmes syntagmes soient traduits différemment dans le texte :

Français	Arabe
On dit aussi qu'il est « <b>analytiquement vrai</b> », c'est-à-dire vrai en vertu de son sens.	فمثل هذا اللفيظ يختصن بأنه لا يمكن أن يكون باطلا. فهو حقّ بفضل معناه.
Est <b>analytiquement vrai</b> ce qui l'est pour tout locuteur en vertu de sa compétence linguistique.	يكون حقًا تحليليًا ما يكون كذلك لدى كلّ متكلّم بحكم مقدّره اللسانية. وبتعبير آخر، يكون تحليليًا حقًا ما ينتمي إلى جميع العوالم الممكنة مهما كان المحيط المعقدي.
En d'autres termes, est <b>analytiquement vrai</b> ce qui, quel que soit l'univers de croyance, appartient à tous les mondes possibles.	

Par exemple dans la première phrase, le prédicat *analytiquement vrai* a été traduit en arabe par une forme négative « ne peut pas être faux » ; tandis que dans les deux dernières phrases, le syntagme adverbe + adjectif en français est traduit par un syntagme équivalent ou inversé (adjectif + adverbe). De telles variations peuvent être expliquées par les contraintes de la langue. Ce phénomène est très fréquent puisque nous avons constaté que la plupart des collocations adverbe + adjectif relevées pour le français ont subi des transformations lors du transfert vers l'arabe. Nous en citons quelques exemples :

Français	Arabe
partiellement spécifiable	محدّدة بصفة أحادية
rigoureusement tracé	يُحطّ (... ) بكلّ دقّة
nécessairement vrai	حقّ بالضرّورة

Ces transformations sont de nature diverse, même si elles relèvent toutes de contraintes à la fois morphologiques et combinatoires. En effet, le traducteur peut avoir

recours, selon les cas, à des combinaisons propres à la langue arabe. Au lieu de la combinaison adverbe + adjectif, disponible en français, on aura respectivement la combinaison adjectif + syntagme prépositionnel, verbe + syntagme prépositionnel, adjectif + adverbe.

### 3.2. Déperditions syntaxiques et sémantiques

Plusieurs cas de figures représentent des écarts par rapport au texte source lors du transfert. Nous en citons à titre d'illustration :

- le cas où le prédicat est éliidé en français. Il arrive que la construction utilisée en français soit spécifique et que sa traduction en arabe implique automatiquement un ajout. Tel est le cas de la construction « ne pas relever de ... mais de ... ». Nous constatons que le prédicat *relever* = [ يهيم ] a été explicité en arabe, contrairement à la règle en français, qui impose que le deuxième prédicat soit elliptique afin d'éviter la répétition.

Français	Arabe
Il faut remarquer aussi que la détermination ne <b>relève</b> pas, comme on le laisse entendre de façon implicite, du seul groupe nominal mais de la phrase entière.	علينا أن نلاحظ أيضا أن التحديد لا <b>يهتم</b> فقط المجموعة الاسمية كما يذهب إلى الظن ضمنياً، وإنما <b>يهتم</b> كامل الجملة.

- le cas où le prédicat est omis en arabe. Contrairement au premier cas, nous remarquons qu'il y a eu omission du prédicat *cependant* du texte traduit. La valeur d'opposition exprimée par ce prédicat demeure absente du texte traduit.

Français	Arabe
Entre l'inintelligible et l'intelligible la rupture n'est pas brutale. Pour moi qui suis fort ignorant de la botanique, l'énoncé <i>Les rutacées sont des phanérogames angiospermes</i> n'est cependant pas totalement impénétrable. D'abord j'y reconnais aisément qu'il s'agit de botanique : le champ des interprétations possibles s'en trouve déjà singulièrement limité.	إن القطيعة ما بين المبهم و البين ليست حادة. فبالنسبة إليّ أنا الجاهل بعلم النباتات. فأنا أتبين بسهولة بادئ الأمر أن الموضوع ليس اللفيظ المتأبّيات هي باديات زهر كاسيات البزر منغلقة تمام الانغلاق يتصل بعلم النبات : فذلك يضيق إلى حد كبير مجال التأويلات الممكنة.

- le cas où la négation n'a pas la même portée dans l'énoncé source et l'énoncé cible. Dans l'extrait suivant, la négation porte sur le verbe modal *falloir*, alors que dans la traduction arabe, la négation est portée sur le verbe *séparer*, qui le suit.

Français	Arabe
Il ne faut pas <b>séparer</b> , dans l'analyse, la détermination des substantifs de la nature sémantique des prédicats, des temps et de l'aspect qui actualisent ces prédicats et des adverbiaux qui y participent <sup>6</sup> .	<b>فيجب</b> ألا <b>نفصل</b> في التحليل بين تحديد الأسماء و الطبيعة الدلالية للمسانيد، و طبيعة الأظرف التي تساهم في هذا التحيين.

- le cas où l'on substitue un prédicat à un autre.

Français	Arabe
Nous ne <b>parlerons</b> pas à leur sujet de figement, <b>puisque</b> ces restrictions sont de nature syntaxique ou sémantique (mais non lexicale), même dans le cas particulier de certaines locutions verbales comme <i>casser sa pipe</i> , où le possessif est <b>coréférent</b> au sujet.	ولن <b>تحدث</b> في هذا الصدد عن تكلس، حيث أن هذا الحصر من طبع تركيبي أو دلالي (و ليس معجمياً)، و ذلك رغم الحالة الخاصة لعدد من التعابير من مثل وجد نفسه بين المطرقة و السندان حيث <b>يعود</b> الضمير ه على الفاعل.

### 3.3. Impact au niveau du style

Les corpus alignés sont un excellent outil de contrôle de variation au niveau du style. La différence est relativement importante entre le texte source et le texte traduit.

Français	Arabe
Ce n'est pas tout : je sais qu'en grec <i>phaneros</i> signifie « apparent », et que <i>gamos</i> désigne le mariage. Le mot <i>sperme</i> m'est évidemment connu. De toute évidence, il est donc question des organes de fructification. En quoi sont-ils « apparents » ? C'est-là que s'arrête ma science.	(فانيروس) تعني phaneros بالإضافة إلى هذا : أعلم أنّ (قاموس) تعني "التزاوج". Gamos باليونانية "ظاهر" و أنّ أما كلمة بزر فهي طبعاً مألوفة لديّ. لذا فمن الواضح أنّ الأمر يتعلّق بأعضاء الإخصاب. فما الذي يجعلها "ظاهرة" ؟ إنّ علمي يقف عند هذا الحدّ.

### 3.4. Variations terminologiques

Les textes spécialisés sont fortement constitués d'occurrences complexes : mots composés, locutions, phraséologie, etc. Aucun alignement ou extraction ne peut sérieusement être fait au niveau lexical sans prendre en compte ses phénomènes. S. Mejri affirme à ce propos qu'« un discours spécialisé se définit d'abord par sa phraséologie, c'est-à-dire ses termes monolexicaux, le plus souvent impliqués dans les regroupements syntagmatiques que sont les collocations, et ses termes polylexicaux, c'est-à-dire les séquences polylexicales » (2011 : 134). Cette importance accordée au phénomène phraséologique dans les textes spécialisés ne peut qu'être accrue dans les textes bilingues alignés. Ces unités phraséologiques vont servir d'indice de stabilité terminologique dans la traduction. Si, dans notre cas, le texte arabe est traduit du français, le transfert des contenus conceptuels et des unités terminologiques doit obéir aux critères de fixation et d'ancrage de la langue source. C'est pourquoi, nous remarquons une certaine stabilité des traductions puisqu'il y a peu de variations dans la traduction des termes et collocations de la linguistique. C'est ce que Duchet, Castillo et Kraif appellent « équivalences génériques » (2008 : 134), qui, étayées par de nombreuses occurrences vérifiées sur des corpus textuels massifs, et en faisant appel à des techniques d'extraction automatique de correspondances lexicales, permettent d'obtenir des lexiques bilingues de bonne qualité à partir de corpus alignés. Les phraséologismes fonctionnent comme un outil d'alignement mais également comme un indice de la finesse de l'alignement. En effet, l'extraction de termes polylexicaux se réalise facilement parce que les unités phraséologiques représentent des signifiants stables et fixés dans la langue, qui correspondent à des définitions précises dans le domaine en question. Nous pouvons en citer quelques exemples :

Français	Arabe
proposition analytique	قول تحليلي
univers de discours	محيط الخطاب
valeur de vérité	قيمة حقّ
monde contrefactuel	عالم مصطنع

Ces formations syntagmatiques peuvent également prendre la forme de séries fermées où il y aurait une base fixe et une expansion variable.

La terminologie s'en sert pour dénommer les différents types du même phénomène, l'exemple du nom *sémantique* :

Français	Arabe
Sémantique interprétative	الدَّلَالِيَّةُ التَّأْوِيلِيَّةُ
Sémantique lexicale	الدَّلَالِيَّةُ المعجمية
Sémantique logique	الدَّلَالِيَّةُ المنطقية
Sémantique phrastique	الدَّلَالِيَّةُ الجمليَّة
Sémantique historique	الدَّلَالِيَّةُ التاريخية

Cependant, ces unités terminologiques telles qu'elles sont employées dans les corpus textuels peuvent subir quelques modifications. Elles peuvent perdre leur statut de terme lorsqu'elles sont employées

en dehors de l'univers technique qui leur a assigné ce statut terminologique. Par exemple, si on prend la notion de *valeur de vérité*, chez R. Martin, qui est elle-même empruntée à la logique, elle peut acquérir une autre acception, dans la plupart des cas un glissement sémantique ou une déspecialisation, quand elle est reprise dans un autre contexte différent du contexte d'origine. Les termes polylexicaux spécialisés se caractérisent par un degré de figement variable. De ce fait, ils ne résistent pas aux forces cohésives qui s'exercent sur eux dans le texte spécialisé. Ils se voient souvent transformés par un certain nombre d'opérations syntaxiques qui visent à adapter leur linéarité aux caractéristiques de la phrase ou de la suite de phrases qui la contiennent. Pour optimiser la qualité de l'extraction, on est amené à étendre la liste des unités phraséologiques terminologiques et à l'enrichir par les formes « modifiées » ou les « variantes ».

## Conclusion

Même si le point de départ de cette étude était de présenter une nouvelle méthodologie d'analyse des textes linguistiques, permettant leur alignement et l'appariement d'éléments équivalents dans les deux langues (français et arabe), ses perspectives sont plus étendues pour trois raisons : la première est que nous avons pris conscience de l'intérêt de la constitution de corpus bilingues alignés exploitables par des concordanciers qui donnent des résultats aussi satisfaisants que perfectibles au niveau de l'extraction automatique et semi-automatique des lexiques bilingues ; la deuxième est que notre expérience des corpus bilingues alignés nous a clairement montré qu'en faisant varier la granularité de l'observation sur des phénomènes plus ou moins précis, tels que les trois fonctions primaires des unités lexicales, on est en mesure de concevoir des outils permettant l'amélioration de la qualité des traductions par la confrontation systématique du texte source et du texte cible ; la troisième est qu'il résulte de cette comparaison des régularités significatives et des structures récurrentes dans la langue de spécialité qui ne pouvaient pas être analysées de manière approfondie que grâce à des corpus textuels alignés permettant de recueillir un nombre suffisant d'illustrations et de contextes pour aboutir à une analyse linguistique pertinente du domaine en question : les aspects morphosyntaxiques de la langue de spécialité, les propriétés stylistiques du texte et de sa traduction et ses aspects lexicaux et terminologiques.

## Bibliographie

Déjean H. & Gaussier E., 2002 « Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables », *Lexicometrica*, Numéro spécial, Jean Veronis (dir.), disponible sur le web à l'adresse : <http://lexicometrica.univ-paris3.fr/thema/thema6.htm>

Duchet J.-L., Kraif O. et Castillo M.-T., 2008, « Corpus massifs et corpus bilingues alignés : leur impact sur la recherche linguistique », *Bulletin de la société linguistique de Paris*, t. CIII, fasc. 1, p. 129-150.

Kraif O., 2002, « Méthodes de filtrage pour l'extraction d'un lexique bilingue à partir d'un corpus aligné », *Lexicometrica*, Numéro spécial, Jean Veronis (dir.), disponible sur le web à l'adresse : <http://lexicometrica.univ-paris3.fr/thema/thema6.htm>

Langlais Ph., 1997 : « Alignement de corpus bilingues : intérêts, algorithmes et évaluation », *Bulletin de Linguistique Appliquée et Générale, numéro Hors Série*, Université de Franche-Comté, p. 245-254.

Martin R., 1982, *Pour une logique du sens*, PUF.

Mejri S. & Baccouche T., 2006, traduction de *Pour une logique du sens*, de R. Martin, 1982, PUF, OAT, Beyrouth.

Mejri S. & Neveu F., (dirs.), 2007 : *Langages 167*, La tradition grammaticale, Paris, Larousse.

Mejri S., 2011, « Phraséologie et traduction des textes spécialisés », *Estudios y análisis de fraseología contrastiva: lexicografía, traducción y análisis de corpus*, P. Mogorron Huerta et C. Gonzalez, p. 125-138, Publication de l'Université d'Alicante.

Nakamura-Delloye Y., 2007, *Alignement automatique de textes parallèles Français-Japonais*, Thèse de Doctorat, Université Paris-Diderot, Paris VII, C. Fuchs (dir.).

Neveu F., 2006 : « Un aspect de l'apport des corpus parallèles à la terminologie linguistique : l'alignement », *Mots, termes et contextes*, p. 381-390, AUF.

Sfar I., 2010 : « Figement et incorporation. L'établissement d'un concordancier bilingue français-arabe », *Meta* 55, 1, Hommages au Professeur André Clas, sous la direction de Salah Mejri et Gaston Gross, *Le parcours du sens*, Presses de l'Université de Montréal, p. 158-167.

Sfar I., 2011, « Le traitement des collocations dans les corpus parallèles des textes spécialisés », *Estudios y análisis de fraseología contrastiva: lexicografía, traducción y análisis de corpus*, P. Mogorron Huerta et C. Gonzalez, p. 215-228, Publication de l'Université d'Alicante.

Véronis J., 2000 : « Alignement de corpus multilingues », J.-M. Pierrel (dir.), *Ingénierie des langues*, chap. 6, p. 313-334, Paris : HERMES.

Zimina M., 2002 : « Repérage lexicométrique des équivalences à basse fréquence dans les corpus bilingues », *Lexicometrica*, « Alignement lexical dans les corpus multilingues », Jean Véronis (dir.), revue électronique disponible à l'adresse suivante : <http://www.cavi.univ-paris3.fr/lexicometrica/thema/thema6.htm>

Zimina M., 2004 : « Alignement textométrique des unités lexicales à correspondances multiples dans les corpus parallèles », *JADT*, p. 1195-1202.

## Notes

<sup>1</sup> Ces travaux de traduction, menés dans le cadre de l'équipe de recherche TIL (00/UR/0201) sous la direction de T. Baccouche et S. Mejri, ont donné lieu à une base de données importante de textes traduits en science du langage, nous en citons particulièrement la traduction du *Dictionnaire des sciences du langage* de F. Neveu (Armand Colin, 2004) par S. Mejri (à paraître, 2011) et qui nous a servi de corpus de base.

## L'analyse prédicative : un outil pour l'évaluation de la traduction des textes bilingues alignés

<sup>2</sup> La traduction de l'ouvrage *Pour une logique du sens*, de R. Martin (1982, PUF) par Salah Mejri et Taieb Baccouche, 2006, OAT, Beyrouth.

<sup>3</sup> On peut citer à titre d'exemple la thèse de Nakamura-Delloye Y. intitulée *Alignement automatique de textes parallèles Français-Japonais*, où elle expose une méthodologie d'alignement au niveau des propositions.

<sup>4</sup> Pour plus de détails, voir S. Mejri (2011) et I. Sfar (2011).

<sup>5</sup> Les concordanciers présentés dans cette partie sont extraits du chapitre premier « Sémantique et vérité : notions fondamentales » de l'ouvrage de R. Martin, *Pour une logique du sens* et de sa traduction en arabe. La même méthodologie sera appliquée ensuite sur la totalité du livre et de sa traduction.

<sup>6</sup> La traduction de l'ouvrage *Les expressions figées en français*, de G. Gross (1996, Ophrys), par Salah Mejri et Béchir Ouerhani, 2008, CERES, Tunis.