

***Contrôler le français pour mieux traduire
vers l'arabe***

**BEDDAR Mohand
Centre Lucien TESNIERE**

La traduction automatique a connu et connaît encore un fulgurant essor avec les nouvelles technologies. La volonté de traduire plus facilement et correctement avec une machine devient, le dialogue des cultures et l'économie du marché l'obligent, une nécessité impérieuse et un objectif incontournable. Cependant, le langage naturel de part sa complexité pose de nombreux problèmes, les concepts à traduire sont complexes et s'avèrent difficiles à traduire par une machine. L'arabe à l'image des autres langues naturelles ne déroge pas à cette contrainte. Son système linguistique n'étant pas en phase avec le système linguistique de la langue française, on peut mettre en exergue lors de la traduction automatique français arabe plusieurs obstacles à franchir ou à contourner. Aussi, le langage contrôlé s'impose comme la solution la plus appropriée pour s'affranchir des difficultés et ainsi réaliser des traductions rapides et pertinentes. Il faut bien contrôler le français pour mieux traduire.

Notre recherche sur la traduction automatique des protocoles de sécurité du français vers l'arabe démontre les limites d'une traduction qui prend le langage naturel général comme point de départ et prône les langues contrôlées. L'écart est tel entre les deux systèmes que le passage entre les deux langues devient pernicieux et compliqué et ce à tous les niveaux.

Un système de traduction tout à fait fiable n'est plus un rêve aujourd'hui mais une réalité avec l'apport des langues contrôlées. Les divergences syntaxiques et morphosyntaxiques sont ainsi dépassées. Les contraintes de style et de rédaction imposées au rédacteur du français comme langue de départ évitent les ambiguïtés de tout ordre et facilitent ainsi le transfert vers l'arabe avec des règles nettement définies.

La langue écrite, et ce quelle que soit la langue, est d'une profonde richesse. Toutefois si cette richesse est importante, elle est particulièrement problématique pour la traduction. Il faut par conséquent améliorer la clarté des langues écrites par l'utilisation d'un langage contrôlé, ainsi l'informatique aura la possibilité de traiter les langues dans des limites définies par cette dernière.

L'objectif est alors de pouvoir traduire le français vers l'arabe sans accumulation de règles de transfert, bien au contraire le nombre de règles se réduit et la précision des transferts qu'elles effectuent augmente.

Les règles de la langue contrôlée sont ensuite informatisées afin d'être utilisées par le traducteur que nous mettons au point en vue de traductions fiables. Ces règles donnent des indications sur la syntaxe, la sémantique et la structure. Elles peuvent également s'intégrer dans un processus de chaîne documentaire pour la rédaction de textes destinés à la traduction par ordinateur. Ainsi on obtient les indications nécessaires à la préparation d'un texte source contrôlé, on peut alors : éviter toutes les ambiguïtés sémantiques, simplifier la syntaxe et éliminer totalement ou en partie le travail de post-édition.

La langue contrôlée n'est pas fixe, il y a en particulier le choix du vocabulaire qui fait moduler celle-ci. Certains lexiques sont déterminés en fonction du domaine d'activité comme par exemple la médecine, l'architecture, la biologie, les protocoles,...

L'idée de cette simplification des langues naturelles en langue contrôlée ne date pas d'hier, dans le passé, elle a été envisagée comme une manière de faciliter la communication d'idées complexes par des idées simplifiées. En effet, le langage naturel chargé d'ambiguïté et de complexité a toujours été un défi pour l'ingénierie linguistique qui peine à trouver des solutions efficaces.

Sur le plan linguistique, on peut dire que la stylistique comparée a permis de mettre en évidence certains traits communs à toutes les langues comme la notion de temps. Ces universaux linguistiques sont importants pour faciliter la tâche de la traduction. Au fur et à mesure que les langues se rapprochent, la traduction semble être plus aisée. Ceci pour dire que la traduction du français vers l'arabe ne constitue pas une tâche facile du fait de leur éloignement du point de vue linguistique : syntaxique, morphologique,...

L'arabe est aujourd'hui l'une des dix principales langues de la planète. L'arabe est une langue sémitique dont l'histoire est extrêmement riche et complexe. On constate que chaque période (conquête, islamisation, etc.) que cette langue a traversée à apporter sa contribution à ce qu'elle est devenue actuellement.

Mais même si cette langue possède de nombreux dialectes à l'orale, elle ne possède qu'une écriture, celle que l'on trouve dans le Coran.

L'arabe est une langue qui a une grammaire très riche et très complexe. Et pour pouvoir traduire une langue, il faut déjà saisir toutes ses spécificités grammaticales pour voir les points de convergences et de divergences avec celle de la langue source qui est le français.

Le contrôle de notre langue de départ A se fait en fonction de la langue d'arrivée B le but étant d'éviter des difficultés de traduction pour aller de la langue A à la langue B. En effet une langue prise dans son ensemble montre l'impossibilité d'une traduction complètement fiable, les systèmes qui sont assez éloignés l'un de l'autre, ne permettent pas toujours des correspondances stables. Je parle de stabilité car c'est l'une des bases de la langue contrôlée.

La stabilité dans une langue contrôlée signifie tout simplement des structures immuables selon lesquelles toutes les phrases pourraient être formulées. La langue française contrôlée en fonction de l'arabe fixe ainsi un schéma de correspondance stable vers cette même langue basé sur le verbe noyau de la phrase, et selon Tesnière avec des actants qui gravitent autour.

Le structuralisme est souvent un pilier de la langue contrôlée. Définir la structure d'une langue, c'est connaître son système afin de pouvoir le manipuler. « ...Notre organe [Acta Linguistica] se met au service d'une discipline qui considère la parole comme subordonnée à la langue. Il n'acceptera des recherches sur la parole que dans la mesure où ces recherches se fondent sur le système sur la structure de la langue et visent directement à élucider celle-ci. »

« ... L'espèce langue est le véritable et principal objet de la linguistique structurale. Une langue particulière est

subordonnée au type, et le type à l'espèce. On voudrait donc que les recherches portant sur quelques langues particulières se fondent sur la structure du type ou de l'espèce langue et visent directement à élucider celle-ci. » (Hjelmslev, Ibid, p 24-25)

Nous remarquons dans cette question deux points essentiels : la langue est l'objet principal de la linguistique structurale et non la parole ce qui est aussi l'objet de mon projet linguistique, travailler sur les corpus écrits et essayer de dégager une structure stable. La langue est une manifestation parmi plusieurs manifestations de l'espèce c'est-à-dire le langage dont l'étude constitue à la fois le point de départ et le but ultime de la recherche de la linguistique.

On dirait clairement que la pensée de Hjelmslev est une pensée glossématique qui se propose d'établir ce qui est commun à toutes les langues, quelles que soit leurs diversités, de façon à déterminer les caractéristiques générales de l'espèce, c'est-à-dire ce qui fait l'universalité du langage à travers ses manifestations les plus variées. Ensuite seulement, on analysera des langues particulières, en espérant trouver un pivot à deux axes, convergence et divergence. C'est ce pivot prenant en compte les deux axes qui nous permettra un passage facile et sans risque d'une langue à une autre, dans mon cas du français vers l'arabe.

Afin de clarifier cette idée d'universalité d'un langage c'est-à-dire la glossématique qui à partir d'un système linguistique universel de référence, effectue une étude comparative de deux langues particulières issues de ce même système glossématique, j'ai choisi de prendre un exemple probant en syntaxe qui montre que une telle

théorie pourrait sans aucun doute contribuer à l'élaboration d'une langue contrôlée. Le français et l'arabe, du point de vu lexico-syntaxique possèdent le même système pour exprimer ses idées. Chacune des deux langues utilise les mêmes fonctions syntaxiques dans leurs phrases pour exprimer idées et concepts. La fonction du sujet, du verbe et de l'objet sont tout à fait semblables dans les deux langues du point de vue glossématique. Toutefois, une étude comparative s'appuyant sur cette observation laisse apparaître une différence importante entre le français et l'arabe, c'est en effet la syntaxe. Celle-ci au premier abord semble commune, néanmoins les fonctions syntaxiques quoique identiques dans les deux langues, ne sont pas disposées de la même manière au sein de la phrase. Le français étant une langue plutôt de l'ordre syntaxique SVO (sujet – verbe - objet) diffère complètement de l'arabe qui est essentiellement une langue VSO.

L'analyse des différents corpus en langue française (protocoles de sécurité civile, aéronautique, protocoles médicaux) a permis d'établir une liste des structures récurrentes. L'analyse est fondée sur le verbe qui constitue le pilier de mon travail et le noyau de mes phrases autour duquel gravitent des actants obligatoires et facultatifs. Le recensement de ces structures m'a permis de créer un schéma de correspondance vers les structures des verbes arabes et qui pose pour chaque structure française la structure arabe la plus adaptée. Cela me permet une transition rapide et efficace en termes de temps et de fiabilité. Dans le domaine que nous traitons, c'est-à-dire les protocoles de sécurité, civile, aéronautique et médicaux, nous avons pu mettre en avant deux types de

grandes structures qui sont les structures en verbes infinitifs (*protéger la victime*) et les structures en « si » (*si la personne saigne*). Ces deux principales structures sont exprimées en arabe par deux structures pratiquement équivalentes avec des remaniements à l'intérieur des arguments. Ces équivalences entre les structures françaises et arabes vont permettre au traducteur automatique d'avoir une syntaxe prédéfinie en langue cible. Ces structures sont, de ce fait, l'assise du contrôle de la langue française car en mettant pour chaque verbe français une structure précise, on limite les ambiguïtés et on facilite ainsi à la fois la rédaction des protocoles à l'utilisateur et le passage du français vers l'arabe.

Parmi les verbes qui posent beaucoup de problèmes dans la traduction du français vers l'arabe, on trouve les verbes « être » et « avoir » qui même dans la langue de départ présente des ambiguïtés.

Les verbes « être » et « avoir » sont des verbes d'état. On peut également prêter aux verbes « être » et « avoir » le rôle de soutien des prédicats sémantiques d'état, ce qui en fait alors aussi des verbes support.

« Etre » et « avoir » n'indique pas le même état. Avec « être », on a un état de celui qui est quelque chose (état de l' « étant ») ; avec l'autre celui à qui quelque chose est (état de l' « ayant »). « Etre » établit un rapport intrinsèque d'identité entre les termes qu'il joint, « avoir » un rapport extrinsèque entre les deux termes qui restent distincts.

Les verbes d'état (même s'ils ne réalisent pas le même découpage sémantique) convergent vers l'expression de l'état de quelqu'un ou de quelque chose.

Comme nous l'avons vu ci-dessus ces deux verbes sont aussi envisageables comme des verbes supports. Il fonctionne avec des noms abstraits : les noms d'état.

Les verbes supports se distinguent des verbes prédicatifs par certaines caractéristiques : ils servent à actualiser des prédicats nominaux et n'ont pas eux-mêmes de fonction prédicative, donc ils peuvent être enlevé dans la phrase sans que celle-ci perde son statut de phrase (réduction qui se fait après la relativation. Par contre s'il s'agissait de supprimer un verbe prédicatif, la phrase disparaîtrait de facto.

De plus, les verbes prédicatifs peuvent changer de catégorie (nominalisation) mais pas les verbes supports. En effet, « être », « faire », « avoir », qui sont des verbes support, n'ont pas de forme nominale. Les verbes peuvent être nominalisés seulement lors de leurs emplois prédicatifs.

Ceci dit les verbes « avoir » et « être » devraient être analysés avec précaution car l'écart de traduction entre le français et l'arabe est tellement important qu'il faut des règles de transfert ou de substitution précises.

Le verbe « être » dans la structure en « si » ne pose pas de problèmes particuliers car dans ce type de structure, le temps de traduction utilisé vers la langue arabe est l'accompli. Le verbe « être » trouve donc naturellement sa place dans la traduction du français vers l'arabe et ne pose par conséquent pas de problème pour ce type de phrase. Par contre avec les structures en « vinf », la traduction est beaucoup plus problématique. Ces phrases servent la plupart du temps dans nos corpus (protocole) à donner des instructions afin d'adopter des comportements précis. Pour traduire « être » dans ces phrases à structures en « vinf »,

la première chose à faire est d'associer le verbe « être » avec son attribut car pour réaliser une bonne traduction on ne peut pas prendre le verbe être séparément de son attribut, car celui-ci ne peut pas être traduit en arabe seul. Le verbe « être » n'existant pas à l'infinitif.

Le verbe être dans les structures en « vinf » est souvent suivi d'un adjectif attribut. Pour traduire « être » dans ces phrases à structures en « vinf », la première chose à faire est d'associer le verbe « être » avec son attribut car pour réaliser une bonne traduction on ne peut pas prendre le verbe être séparément de son attribut. « Etre » et son attribut constitue une unité lexicale à part entière donc une seule entrée. Le système prendra l'unité la plus longue c'est à dire repérer déjà le verbe « être » et ensuite chercher s'il y a un adjectif attribut. Enfin il faudra donner la traduction équivalente en arabe qui est un nom verbal.

Les règles de transfert pour ce verbe sont les suivantes :

La structure en « être » dans les phrase avec un argument 1 qui est soit adjectif attributif soit nom d'état, s'accordent tous les deux avec l'argument 0 en genre et en nombre. Le verbe « être » entre dans la liste des verbes d'état. Ceci dit, et à défaut d'une structure précise en langue française qui régit la reconnaissance en langue-source et la génération en langue-cible, ma règle de transfert sera la suivante : si, dans la structure en « si » de type argument 0(Arg0) + verbe conjugué (Vconj), le verbe est un verbe d'état y compris le verbe « être », l'argument 1 qu'il soit adjectif attribut ou un nom d'état s'accordera en genre et en nombre avec l'argument 0.

Si l'argument 1 est un groupe nominal simple, indéfini, singulier ou pluriel interne, il sera toujours à l'accusatif c'est à dire au cas direct qui exige un *tanwîn* à la fin.

L'adjectif épithète, attaché au nom, sachant qu'il s'accorde en genre et en nombre avec le nom qu'il qualifie se comportera de la même manière et donc prendra un *tanwîn*.

L'argument 1 qui dans le cas des verbes d'état est soit un adjectif soit un nom d'état, au pluriel externe, sera toujours au cas direct.

Le verbe « avoir » à l'image de la copule « être », pose énormément de problèmes dans la traduction français arabe. Le verbe avoir n'existe pas dans la langue arabe. Que ce soit dans les structures en « vinf » ou en « si », le verbe « avoir » nécessite un traitement local c'est à dire un traitement particulier à l'intérieur de la phrase que ce soit dans les structures en « si » ou dans les structures en « vinf ». Avec la structure en « si » ou en « vinf », le verbe « avoir » peut avoir, du point de vue sémantique, deux acceptations : l'une, indique la possession (*si vous avez une plaie ouverte*), l'autre, indique un état physique ou psychologique (*si vous avez froid*). Selon ces deux types d'acceptations, la traduction ne se fera pas de la même manière. Des règles de transfert précises régiront alors la traduction de la langue française vers la langue arabe.

Avec la structure en « si », et Dans le cas de la possession, le verbe avoir se traduira isolément, il sera invariable, la traduction ne se fera pas de façon littérale, le verbe avoir n'ayant pas d'équivalent arabe, mais par le verbe être à l'inaccompli et une préposition de possession. Des règles de transfert régiront cette traduction du verbe « avoir » de possession **كان** :

Avec le pronom vous, on ajoute un pronom possessif juste après le verbe et si le sujet est un groupe nominal, la

deuxième partie du verbe avoir qu'est la préposition de possession s'attachera au nom qu'elle précède.

Dans le cas de la description d'un état (quel qu'il soit), le verbe « avoir » ne se traduira pas isolément mais avec le nom d'état qu'il introduit. Le verbe, dans ce cas là, s'accordera en genre et en nombre avec le sujet.

Avec la structure en « vinf », on contourne le problème avec un autre verbe équivalent au verbe *avoir* français quand il s'agit d'une possession. Quand « avoir » désigne un état, la traduction se fera de la même manière que lors de la traduction avec la structure en « si ». Le verbe « avoir » ne se traduira donc séparément mais avec le nom d'état qu'il introduit.

D'autres problèmes sont éliminés grâce à l'utilisation de la langue contrôlée notamment celui de la négation dans les structures en « si » où le traitement des verbes défectueux rend la traduction laborieuse. Il en va de même pour les pronoms possessifs qui sont ambigus non seulement en arabe mais aussi en français avec les anaphores qu'ils peuvent générer. Les prépositions constituent aussi un obstacle lors du passage du français vers l'arabe, leur traduction est flexible selon le verbe qui les introduit. La langue contrôlée intervient à ce niveau là pour limiter l'utilisation de ces prépositions et imposer des prépositions particulières par rapport d'autres.

Ceci dit, la langue contrôlée s'impose alors aujourd'hui comme une alternative salvatrice, elle réduit les ambiguïtés par des règles de styles rigoureuses et adaptés à la capacité technologique des systèmes de traductions sans nuire à la qualité de celle-ci. C'est l'ensemble de ces règles, qui constituent la langue contrôlée, qui est nécessaire pour un traitement correct du langage naturel.

Bibliographie

Ouvrages:

Eléments de l'arabe classique, Régis Blachère, 1946,
Edition Librairie Orientale et Américaine

La systémique, Daniel Durand, 2002, Edition Que sais-je ?

Théorie générale des systèmes, Ludwig von Bertalanffy,
1972, Edition DUNOD

Linguistique et Traitements Automatique des Langues,
Catherine Fuchs, 1993, Edition Hachette supérieurs

BULAG La traduction et le Traitement Automatique des
Langues n°25, coordonné par Sylviane Cardey, 2000,
Edition de l'Université de Franche-Comté

L'arabe, 2002, Edition Que sais-je ?

La traductique, Pierrette Bouillon et André Clas, 1993,
Edition des Presses de l'Université de Montréal
AUPELF/UREF

Problèmes de linguistique générale, Emile Benveniste,
1966

Conception d'un modèle de traduction arabe-français
appliqué au domaine des mathématiques, session 2006,
Abdelouafi GHENIMI

Vers un système de traduction automatique du langage
juridique du français vers l'arabe, janvier 2003, Haytham
ALSHARAF

Les problèmes théoriques de la traduction, 1963,
Gallimard, Georges MOUNIN

Éléments de syntaxe structurale, 1959, Lucien TESNIERE

Articles:

La traduction automatique du français vers l'arabe,
Haytham Alsharaf, in BULAG La traduction et le

Traitement Automatique des Langues n°25, coordonné par Sylviane Cardey, 2000, Edition de l'Université de Franche-Comté, p 109

Disambiguating and tagging using systemic grammar, Sylviane Cardey et Peter Greenfield, in Actes du 8^e International Symposium on Social Communication, 2003, Santiago de Cuba Actas I, p 559

Recherche pour le traitement automatique de la langue française et ses différentes applications, Sylviane Cardey, in Actes des journées linguistique franco-asiatique, 2003, Université Naresuan Phitsanulok de Thaïland, p 54

Systemic Linguistics with applications, Sylviane Cardey et Peter Greenfield, in the Proceedings of the 9^e International Symposium on Social Communication, 2005, Santiago de Cuba Actas II, p 649