

**Syntaxe Comparative et Langues Contrôlées en vue de la
Traduction Automatique des Alertes et des Protocoles de Sécurité
du Français vers l'Arabe
Système TACT Contrôlé Français-Arabe**

Mohand BEDDAR

Centre Lucien Tesnière, Université de Franche-Comté
mohand.beddar@edu.univ-fcomte.fr

Résumé

Le recours à la méthode comparative en traduction automatique est inévitable et constitue l'essence même de ce domaine. Quel que soit le degré de parenté entre les deux langues source ou cible, l'étude contrastive s'impose, car elle met en lumière certains aspects qu'il est difficile de déceler sans une démarche comparative et une analyse parallèle entre langue de départ et langue d'arrivée. Dans le système TACT contrôlé français-arabe appliqué aux protocoles de sécurité et qui est développé au laboratoire Lucien Tesnière dans le cadre du Projet LiSe, la méthode comparative a été d'un grand apport puisqu'elle a permis de mettre en exergue les convergences et les divergences entre les deux systèmes linguistiques français et arabe et de mesurer le dispositif de contrôle à mettre en place.

Abstract

The use of comparative method in machine translation is essential and constitutes the essence of this domain. Whatever the degree of similarity between the source language and the target language, a contrastive study is necessary in order to highlight some aspects of divergences that are difficult to detect without a comparative approach and a parallel analysis between the two languages. In the controlled French to Arabic machine translation system TACT applied to security protocols, developed in Lucien Tesniere laboratory as part of LiSe project, the comparative method has been very important; it has allowed to highlight the similarities and differences between the two linguistic systems, French and Arabic, and to determine the set of rules that need to be implemented.

Mots-clés : méthode comparative, traduction automatique, système TACT français-arabe, protocoles de sécurité, convergences, contrôle, règles de transfert

Keywords: comparative method, machine translation, French to Arabic machine translation system TACT, control rules, transfer rules

1. Langue contrôlée LiSe et système TACT

La langue contrôlée LiSe et le système TACT sont indissociables. Le contenu de la langue contrôlée LiSe est le reflet de l'architecture du système TACT français-arabe, il correspond au travail d'analyse et de réflexion mené pour sa conception. Inversement, le mécanisme de fonctionnement du système est en rapport direct avec chaque règle de contrôle et chaque modélisation structurale. Chaque contrainte linguistique qui vient renforcer le dispositif de contrôle est justifiée (Cardey, 2008). Elle provient directement de l'analyse menée sur deux axes différents, un axe monolingue qui concerne chaque langue, indépendamment l'une de l'autre, et un axe contrastif qui met l'accent sur les convergences et divergences entre les deux langues.

1.1 Contrôle par restrictions

1.1.1 Contrôle statique

Le contrôle statique s'effectue sur chaque langue indépendamment l'une de l'autre. Il s'applique dans la majorité des cas sur la langue source mais peut concerner aussi la langue cible. Dans ce dernier cas, le contrôle sert plutôt à éviter des complications liées à la formation morphologique et syntaxique surtout dans le cas de la langue arabe. La morphosyntaxe est un phénomène qui prend de l'ampleur une fois la phrase traduite vers la langue arabe. De ce fait, l'interdiction de certaines catégories ou certaines structures diminue la complexité, sans nuire à l'effet communicatif de la langue. L'exemple 1(a) montre une règle de contrôle statique sur l'arabe. Elle est établie exclusivement grâce à des connaissances linguistiques sur la langue arabe. Elle n'est donc en aucun cas motivée par les particularités de la langue source, le français. L'interdiction d'utiliser certains pronoms vise à alléger le système de conjugaison arabe très flexionnel si nous prenons en compte toutes les catégories de verbe qui existent en arabe et leur morphologie sensible à plusieurs facteurs (genre, temps, négation,...). Cependant, il est clair qu'une telle restriction impacte la langue source, puisque toute restriction sur l'arabe conduit à une restriction indirecte sur le français.

Le contrôle statique dans la langue source, quant à lui vise plutôt à désambiguïser l'énoncé par des restrictions linguistiques mais aussi à accroître la lisibilité des textes par des restrictions de format. Il sert aussi à donner au rédacteur des moyens de rédaction facile d'utilisation et légers en informations linguistiques. Dans l'exemple 1(b) la règle de contrôle interdit l'utilisation dans les protocoles de sécurité de plusieurs verbes dans la même phrase. Cela revient à tolérer un seul verbe dans une condition ou une instruction.

(1) a. Seule la conjugaison à la 3^{ème} personne féminin/masculin du singulier هو/هي, la 3^{ème} personne féminin/masculin du pluriel «هن/هم» et la 2^{ème} personne masculin pluriel «أنتم» est autorisée

b. utiliser une seule action par phrase c'est-à-dire un seul verbe

1.1.2 Contrôle dynamique

Le contrôle dynamique s'effectue pendant la mise en contact des deux langues. Ce sont les problèmes qui ressortent du va-et-vient qu'effectue le traducteur entre langue source et langue cible. Il vise à aplanir des écarts qui nécessiteraient des moyens de formalisation complexes pouvant contribuer à l'alourdissement du système. L'exemple 2 montre une règle de contrôle

dynamique qui est due à un problème de traduction des adjectifs numériques du français vers l'arabe. La règle exige l'utilisation des chiffres arabe au lieu des chiffres en lettres, car l'adjectif numéral cardinal est sensible en arabe à l'aspect genre, il est au masculin quand il précède un nom féminin et il est au féminin quand il précède un nom masculin.

(2) a. Refroidir la peau sous l'eau courante pendant *quinze** minutes

b. Refroidir la peau sous l'eau courante pendant 15 minutes

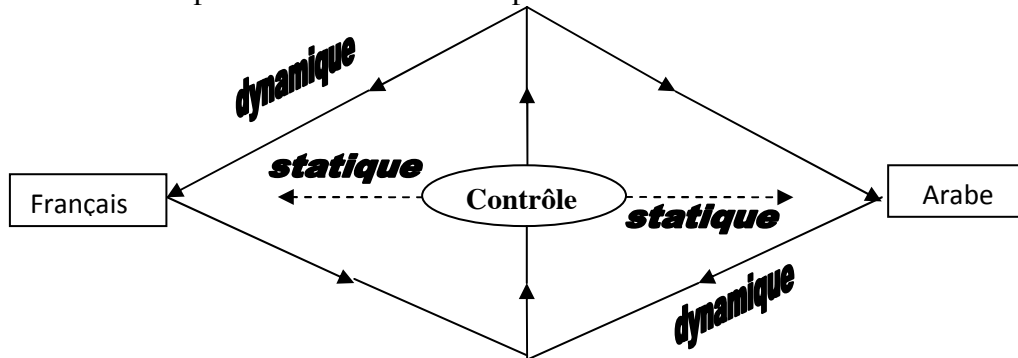


Figure 1 : Contrôle statique et contrôle dynamique

1.2 Contrôle par transformation

1.2.1 Contrôle bilingue par nominalisation

La nominalisation n'est pas une simple modification morphologique sur le verbe ou l'adjectif en vue d'obtention de syntagmes nominaux mais un moyen de production phrastique (Dubois, 1969). Les phrases générées par ce moyen deviennent des structures argumentales stables appelées macrostructures-miroir. Ces macrostructures correspondent au produit final du contrôle au moyen de la nominalisation. C'est le point extrême de l'isomorphisme entre l'arabe et le français une fois la désambiguïisation atteinte et les divergences, que la méthode contrastive a mis en exergue, aplanies. (Voir figure 2)

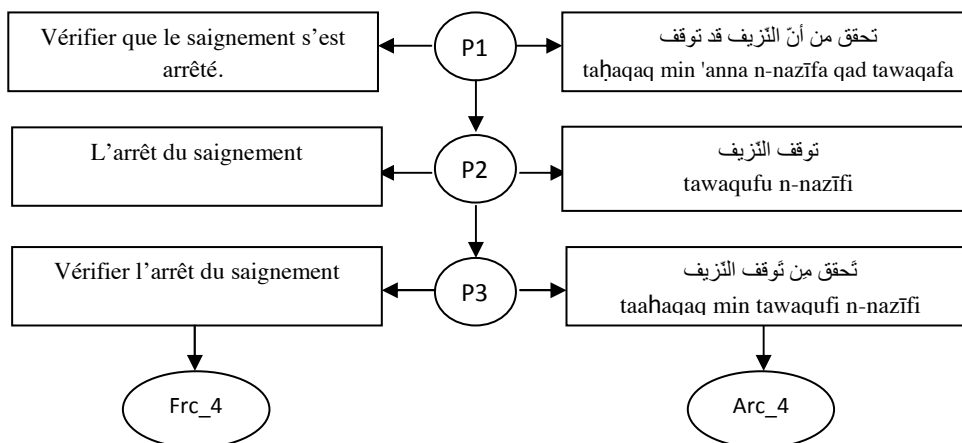


Figure 2 : contrôle bilingue par nominalisation

2. Modèle syntactico-sémantique du système TACT

L'approche repose sur la catégorie du verbe et son comportement dans la phrase, en particulier sa capacité valentielle. Le verbe est un invariant monosémique (voir figure3),

autrement dit un noyau capable de faire graviter autour de lui des variables aussi bien obligatoires qu'optionnelles qui constitueront la structure argumentale appelée macrostructures-miroir. Les prépositions rattachées directement au verbe deviennent systématiquement des atomes à cause de leur degré d'instabilité car la même préposition en français peut avoir plusieurs équivalents en arabe (voir exemple 3).

(3) a. Eloigner + (...) + *de* + (...) --> (...) + عن + (...) + إبعاد

b. Sortir + (...) + *de* + (...) --> (...) + من + (...) + خروج

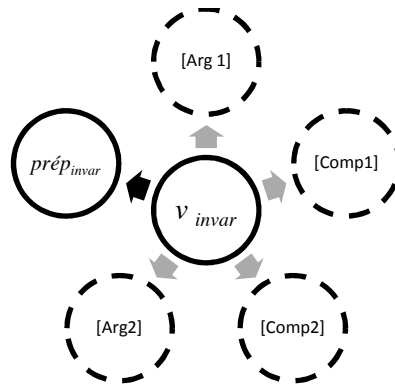


Figure 3 : verbe noyau et structure argumentale

2.1 Macrostructures-miroir à noyau verbal

La notion de la *macrostructure-miroir* telle qu'elle est établie dans le système TACT est une notion nouvelle à la fois par la démarche et la réflexion menée pour sa mise en place et par le résultat lui-même. Elle tire son originalité du modèle de représentation syntactico-sémantique qui prend en compte les divergences et les convergences entre les deux langues source et cible mais aussi les règles de contrôle qui s'y rapportent. Ce mariage entre méthode comparative et contrôle vise à ramener les deux langues sur un plan plus au moins isomorphe en fixant des structures argumentales stables et équivalentes. Lorsque les écarts sont importants au moment de la traduction des textes non contrôlés, le contrôle intervient par un ensemble de transformations considérées comme autant de règles de contrôle, pour essentiellement désambiguïser et rapprocher les deux langues. Le résultat final de ce contrôle et du travail autour du verbe est donc des structures contrôlées non ambiguës valables autant pour la langue source le français que la langue cible, l'arabe.

La macrostructure dans les deux langues contrôlées correspond à la structure type de chaque verbe. En effet, chaque verbe est muni d'une seule et unique structure argumentale et d'une clé argumentale qui renvoie à cette structure. Les verbes sont rangés ensuite par groupes verbaux selon qu'ils portent la même clé ou non. À chaque macrostructure française correspond une macrostructure arabe et l'équivalence s'effectue par la clé argumentale. En effet, le verbe monosémique en français porte une clé et renvoie à une structure argumentale unique. Il en est de même en arabe. Les verbes ainsi autorisés dans la langue contrôlée sont réparties dans des groupes selon leur structure argumentale. (Beddar, 2009)

Les rôles sémantiques des arguments dans les macrostructures-miroir se définissent systématiquement par rapport au sens porté par le verbe. Cette approche, choisie pour la représentation du sens en langue source et son transfert vers l'arabe, est motivée par la fiabilité qu'elle donne à la traduction et aussi parce qu'elle présente un moyen efficace pour surmonter les divergences entre le français et l'arabe. La traduction des verbes français vers

