

# Langues et cultures, systèmes et traduction

SYLVIANE CARDEY (France), HELENA MORGADINHO (France), ALEKSANDRA DZIADKIEWICZ (Pologne), SOMBAT KHRUATHONG (Thaïlande), HSIANG-I LIN (Taiwan), KYOKO KURODA (Japon), GINA MELIAN (Roumanie), FAROUK BOUHADIBA (Algérie), DUYGU CAN (Turquie), EUN SOON YU (Corée), XIAOHONG WU (Chine), IGOR SKOURATOV (Russie), VALENTINE GROSJEAN (France), GABRIEL SEKUNDA (Pologne), IZABELLA THOMAS (Pologne), YVES GENTILHOMME (France), NAGA ANURADHA CHINTALAPUDI (Inde), ROSITA CHAN (Panama)

*Université de Franche-Comté, Besançon, France*

[sylviane.cardey@univ-fcomte.fr](mailto:sylviane.cardey@univ-fcomte.fr)

## RESUME

Nous essaierons dans cet article de mettre en regard des langues de même et diverses origines afin de montrer leurs points communs et leurs différences (concernant leur fonctionnement dans un but de traduction). Ceci nous amènera à revoir la notion de « mot », de « parties du discours ». Nous pourrions montrer aussi combien la perception du monde à travers les civilisations joue son rôle dans l'organisation des langues (les traces du passé dans la pensée en sont des témoins comme le montrent les proverbes et autres composés). L'arabe, le chinois, le coréen, l'espagnol, le français, l'italien, le japonais, le polonais, le portugais, le roumain, le russe, le sanskrit, le thaï et le turc serviront de base à nos remarques et études. Toutes ces remarques nous conduiront, à travers des exemples, à la traduction en général et à la traduction automatique ou aide (dictionnaires) à la traduction en particulier.

## ABSTRACT

In this paper we compare languages having the same origin and others with different roots so as to demonstrate what they have in common and how they differ for the purpose of machine translation. In doing so, we will revisit the notions of 'word' and 'part of speech'. These comparisons demonstrate how different are the views of the world through civilisations, and their impact on the structure of languages (compounds, idioms, proverbs will served our demonstration). Arabic, Chinese, French, Italian, Japanese, Korean, Polish, Portuguese, Romanian, Russian, Sanskrit, Spanish, Thai and Turkish will be at the basis of our studies. These comparisons aim at showing through examples how dictionaries should be organised and how to obtain acceptable translations made by machine.

## MOTS-CLES/KEYWORDS

traduction multilingue, dictionnaires, parties du discours, segments, phraséologie

## 1. Introduction

Les recherches sur l'origine et la filiation des langues sont nombreuses et ont provoqué maintes polémiques. Les linguistes, il est vrai travaillent souvent sur très peu de langues à la fois. Des comparaisons ont été faites au niveau du lexique (comparaison des formes et évolution phonétique), de la syntaxe, de la morphologie en vue d'essayer de retrouver la filiation des langues. Les études n'ont pas beaucoup porté sur des unités de types phraséologismes (proverbes, dictons et autres) que nous verrons également dans le cadre de cet article. Notre but est de comparer 14 langues. Nous verrons ainsi que le mot, unité de sens, n'a plus grand sens quand on le met à la lumière de différentes civilisations avec diverses façons de penser. Les catégories grammaticales, parties du discours, que nous utilisons pour les langues latines voire indo-européennes ont souvent du mal à trouver un équivalent dans d'autres langues et vice versa. Pour terminer nous proposerons une façon de voir les dictionnaires multilingues et la traduction automatique. Des langues telles que l'arabe, le chinois, le coréen, l'espagnol, le français, l'italien, le japonais, le polonais, le portugais, le

roumain, le russe, le sanskrit, le thaï et le turc vont être étudiées à travers différents filtres ou analyses.

## 2. Langues et familles de langues, que trouve-t-on dans la littérature ?

Si nous reprenons rapidement les travaux faits pour classer ces langues, on peut à peu près dire que le turc appartiendrait au groupe altaïque qui inclut aussi le coréen et le japonais. L'italien, l'espagnol, le français, le roumain, le portugais sont classés parmi les langues romanes dérivées du latin de la branche italique de la famille indo-européenne à laquelle serait aussi rattachés le russe et le polonais par la branche slave elle-même dérivée du balto-slave. Le chinois, langue de la famille sino-tibétaine est la plus ancienne après le sumérien et la plus parlée au monde (environ un milliard de personnes parleraient le chinois mandarin). Le thaï appartient à la famille austro-thaï. L'arabe est une langue sémitique de la famille afro-asiatique. Le sanskrit est la langue connue la plus ancienne de la famille indo-aryenne, branche de la famille indo-européenne. L'origine et les groupements par familles étant toujours en débat, les théories diverses, nous ne poursuivrons pas cette description qui n'est pas l'objet de ce travail.

## 3. La notion de mot

Selon le niveau de description dans lequel on se place, niveau phonétique-phonologique, niveau orthographique-graphémique, niveau morphologique, niveau lexico-sémantique ou encore syntaxique, un « mot » peut prendre des caractéristiques variées voire très souvent contradictoires. Au lieu donc d'essayer de trouver la seule et unique bonne définition du « mot », nous montrerons à travers différentes langues comment cette notion apparaît à l'écrit. Aussi pourrions-nous peut-être reprendre les notions de mots et de mots formes (le mot étant une représentation canonique de mots-formes ou variations), mais ceci sera-t-il suffisant ? Nous verrons comment les unités de sens sont représentées à travers les langues. Nous nous placerons dans le cadre de la traduction automatique qui va demander des règles de description précises pour analyser la langue avec une machine qui elle ne comprend rien. Tout ceci nous amènera à parler plutôt de « segments » que de « mots ». Nous savons que le problème en traduction relève du sens plus que de la notion de mot, cependant en traduction automatique tout part de représentations tangibles que la machine va devoir lire et calculer, il en est de même d'ailleurs au départ pour le traducteur humain devant son texte. Quelle que soit la langue étudiée dans un but de traduction, nous avons besoin avant tout de reconnaître des formes, des segments à partir desquels la traduction va pouvoir commencer et uniquement commencer. À travers des exemples, nous allons essayer de voir comment certaines langues s'organisent afin de mettre en évidence les segments sur lesquels travailler lors de la traduction automatique ou assistée.

### 3.1. Les segments ou mots d'une langue à l'autre

3.1.1. Dans les langues romanes, si l'on n'y prend pas garde,

*pomme de terre*

n'est pas reconnu par l'ordinateur comme un ensemble mais comme trois unités indépendantes. À l'inverse la langue espagnole est dotée d'unités composées qui ressemblent à des unités simples :

*Saltamontes* : composé formé par la concaténation du verbe *salta* et du substantif *montes*

On retrouve ces deux types de compositions en roumain :

*untdelemn* : *unt* (beurre) + *de* (de) + *lemn* (bois) qui signifie en français : huile  
*floarea-soarelui* la fleur du soleil, tournesol

L'enclise ou la contraction dans les langues romanes sont autant de phénomènes qu'il faut pouvoir reconnaître.

3.1.2. Les langues slaves présentent la particularité de rattacher des morphèmes marqueurs de personnes (*-m, -ś, -śmy, -ście*) aux différentes parties du discours :

*Ciekawiśmy.* = Nous sommes curieux

Le morphème *-śmy*, première personne du pluriel, se rattache directement à l'adjectif, en l'occurrence, masculin du pluriel *ciekawi*. Les morphèmes marqueurs de personnes peuvent se rattacher de la même manière à d'autres parties du discours (pronoms personnels, adverbes, pronoms interrogatifs...).

3.1.3. En turc les mots composés forment un seul mot ; lorsqu'une voyelle se trouve à la fin du premier mot et au début du second, l'une d'elles disparaît pour donner une seule forme

Cuma ertesi  
*Cumartesi*

3.1.4. En arabe, un « mot » est constitué à la base par une racine composée essentiellement de deux (atypique), trois (trilitère), majoritairement, quatre (quadrilitère) consonnes au sein desquelles sont insérées des voyelles appelées voyelles thématiques. Cette racine avec les voyelles insérées vont constituer un « schème » ou un « thème ». Pour illustrer cette composition morphologique du mot en arabe, à la racine <ktb> qui représente le sens global d' « écrire », des voyelles thématiques vont être insérées pour construire des schèmes tels que :

*kataba* : il a écrit

*kitaab* : livre

Chaque racine nous donne un sens :

<ktb> 'idée d'écrire'

<rnn> 'idée de sonner'

Une des particularités de la langue arabe est de faire usage de doublets sémantiques qui sont des synonymes utilisés généralement en fin de phrase et qui se traduisent souvent par un seul mot dans la langue cible.

*la tu9add wa la tuhSa* (pas comptée et pas évaluée)  
traduction : « qui ne peut être évaluée »  
et non pas « qui ne peut être comptée ni évaluée ».

3.1.5. Dans les langues asiatiques analysées, cerner la notion de mot équivaut à s'interroger sur la notion de syllabe. Cette notion de syllabe est difficile à définir car elle varie selon la langue.

En thaï une syllabe a besoin de 3 éléments pour former une unité sémantique : une consonne initiale, une voyelle et un tonème. La consonne finale n'est pas obligatoire.

En chinois deux termes sont liés à la notion de mot : 字 (*zì*) et 詞 (*cí*). Un 字 (*zì*) est un caractère (donc monosyllabique), tandis qu'un 詞 (*cí*) peut être monosyllabique, dissyllabique ou polysyllabique. Autrement dit, un 詞 (*cí*) peut être composé d'un ou de plusieurs 字 (*zì*). Un 詞 (*cí*) est une unité sémantique indépendante, pouvant être traduit par « mot » en français alors qu'un 字 (*zì*) est parfois un mot, parfois un morphème (dans ce dernier cas il ne peut pas fonctionner seul). De plus, 詞 (*cí*) englobe deux notions : 單純詞 (*dān chún cí* : mot simple) et 合成詞 (*hé chéng cí* : mot composé). La plupart des mots simples chinois sont monosyllabiques.

En thaï, les mots simples ou mots fondamentaux sont à dominante monosyllabique. Chacun d'eux représente un concept simple. Un mot composé se constitue par juxtaposition de concepts simples, sans aucun espace entre eux.

*/baj-1/* (feuille) + */klua:j-3/* (bananier) = */baj-1 + klua:j-3/* (feuille de bananier)

Le vocabulaire du thaï se forme principalement par juxtaposition de concepts simples. Deux noms thaï peuvent être juxtaposés pour composer un nouveau nom. En général, les composés thaï ont le « mot tête » à gauche. Les mots dépendants s'étendent à droite. Cependant, il existe une masse de termes empruntés ultérieurement au Pâli et au Sanskrit qui utilisent la procédure inverse. Comme en chinois, en thaï, si l'ordre des mots est inversé alors apparaissent de nouvelles significations.

*/pva:n-3//lu:k-3/* : ami/amié de (ma) fille ou de (mon) fils  
*/lu:k-3//pva:n-3/* : enfant de (mon) ami ou de (mon) amie

Les langues asiatiques s'écrivent de gauche à droite, mais le chinois (à Taiwan), le japonais et le coréen peuvent aussi s'écrire de haut en bas. Il n'y a pas de marques de séparation de mots sauf en coréen ; les mots composés ne sont pas séparés. Il n'y a pas de majuscules, ce qui représente une difficulté majeure dans la segmentation automatique, la distinction entre noms propres et noms communs n'est pas simple.

Finalement nous terminerons simplement en disant qu'un « segment », un « mot » a une fonction et un sens lorsqu'il peut être reconnu, analysé et traduit.

#### 4. Les parties du discours

Dans la grammaire traditionnelle des langues latines, on cite en général les 9 parties du discours suivantes :

adjectif, adverbe, article, conjonction, interjection, nom, préposition, pronom, verbe.

Cependant il existe des divergences selon les langues :

- en Sanskrit, il y a 5 parties du discours : nom, verbe, adjectif, adverbe et préposition. C'est une langue privée d'articles et de pronoms relatifs comme le chinois, le japonais et le coréen ;
- en turc, il y a deux grandes catégories, le verbe et le nom. Les noms peuvent prendre de multiples fonctions, adverbe, interjection, pronom, conjonction, etc. ;
- en chinois, en japonais et en coréen, il n'y a pas d'adjectifs possessifs ni de pronoms relatifs ;
- en thaï, le pronom relatif est calqué sur la grammaire anglaise ;
- en coréen pour traduire

*la neige qui tombait*

on utilise la racine du verbe à laquelle on adjoint une terminaison relative. Le pronom relatif est exprimé à l'aide de terminaisons correspondantes qui s'ajoutent après le radical (racine) des verbes d'action, de qualité ou d'état (adjectif).

- en thaï pour exprimer la possession, on peut employer la préposition « kh@:ng-1 » précédée d'un nom et suivie d'un nom ou pronom.
- dans les langues asiatiques, il n'existe pas d'articles. Dans certains cas l'absence est comblée par l'emploi de classificateurs pour indiquer la classe sémantique à laquelle appartient un terme.

Cas sans article :

*/chan-5+mi ; -1+nang-5sv/*  
 moi + avoir+livre(s)  
 Pron+V+N  
 S+V+COD

*/chan-5+mi ; -1+nang-5sv :-5+lem-3+nvng-2moi+avoir+livre+1+CI/*  
 Pron+V+N+classificateur pour livre+Nbr  
 S+V+COD+classificateur pour livre+ Nbr

- les langues slaves ont majoritairement les mêmes parties du discours que le latin. Par rapport aux langues romanes, on note l'absence d'articles définis et indéfinis, ce qui pose problème en traduction automatique. Les unités qui ne correspondent pas aux catégories grammaticales traditionnelles sont regroupés dans la catégorie appelée particules :

- les particules interrogatives (*czy* en polonais, *ли* en russe)
- les particules expressives (*no, że* en polonais, *ну, же* en russe) qui sont difficiles à traduire puisqu'elles expriment des attitudes, des sentiments  
*Иди же сюда!* Viens donc ici !  
*No, no!* Dis-donc ! Ca alors !

## 5. Remarque conclusive

Nous voyons clairement que ce qui pose problème dans le cadre de la traduction automatique est le repérage des « segments » et la grande divergence de catégories de discours d'une langue à l'autre. Les moyens linguistiques utilisés pour remplir certaines fonctions sont très variés et parfois certains éléments qui n'existent pas dans une langue sont indispensables dans une autre.

## 6. La phraséologie

### 6.1. Les phraséologismes pragmatiques

On peut dire à peu près qu'un phraséologisme est une formule avec une structure linguistique fixée par l'usage et utilisée conventionnellement dans des situations pragmatiques déterminées comme par exemple :

*Auriez-vous l'amabilité / la bonté de...?*  
*Y a moyen que tu (me prêtés 10 euros)?* (registre familier)  
*Merci de (bien vouloir)+infinitif*

#### 6.1.1. l'acte de requête

L'acte de requête est une manifestation verbale du désir d'obtenir de la part du destinataire du message une réaction qui soit bénéfique pour l'émetteur. Chaque langue a un stock de formules indirectes conventionnelles et de termes d'adresse qui permettent d'atténuer la requête et de réaliser

les principes de la politesse. Nous allons voir comment ces formules sont construites d'une langue à l'autre.

#### 6.1.1.1. en français

*Peux-tu me passer le sel ?*  
*Tu veux pas me passer le sel ?*  
*Veuillez me confirmer votre participation ?*  
*Auriez-vous la gentillesse + infinitif + ?*  
*Je vous serais très reconnaissant de (bien vouloir) + infinitif.*

#### 6.1.1.2. en espagnol/italien/portugais

espagnol

*Puedes pasarme la sal?*  
*Quieres pasarme la sal ?*  
*Puede / pueden / podéis confirmarme su / su / vuestra llegada?*

Les formes varient en fonction du destinataire : *usted / ustedes / vosotros*

italien

*Puoi passarmi il sale ?*  
*Non vuoi passarmi il sale ?*  
*Voglia / vogliono / vogliono confermarmi*  
*la sua / la loro / la vostra partecipazione.*

(opposition : *lei / loro / voi*)

portugais

*Podes passar-me o sal?*  
*Não te importas de me passar o sal? : cela ne te dérange pas de me passer le sel*  
*Faça a favor de confirmar a sua participação : faites moi la faveur de confirmer*

#### 6.1.1.3. en roumain

*Poți să îmi dai sarea ?*  
*Nu vrei să îmi dai sarea ? (ne sera pas employé)*  
*Tu ne veux pas me passer le sel?*  
*Vă rog să îmi confirmați participarea dumneavoastră.*  
*Je vous prie de me confirmer votre participation*

Le système des pronoms de politesse est plus complexe : *dumneata* (tu), *dumneavoastră* (tu ou vous), *dumnealui* (il), *dumneaei* (elle), *dumnealor* (ils ou elles)

#### 6.1.1.4. en russe

Le vouvoiement fonctionne de la même façon qu'en français. Pour atténuer la requête, on utilise la particule *бы* ('by') qui complète le verbe au passé :

*Не могли бы вы передать мне соль?*  
*Ne pourriez-vous pas me passer le sel ?*

*Пожалуйста, подтвердите мне ваше участие.*  
S.V.P., confirmez-moi votre participation.  
*Будьте добры дать мне...*  
Soyez bon de me donner...

6.1.1.5. en polonais

*Chcesz podać mi sól ?*  
Veux-tu me passer le sel ?  
Forme imperfective de « vouloir », interprétée comme une simple question  
*Zechcesz / zechciałbyś podać mi sól ?*  
Forme perfective de « vouloir » qui sert à formuler une requête  
*Uprzejmie proszę Pana / Panią / Panów / Panie / Państwa o potwierdzenie*  
*uczestnictwa.*  
Je prie gentiment Monsieur / Madame / Messieurs / Mesdames / Mesdames et  
Messieurs de confirmer sa/leur participation.

6.1.1.6. en turc

On utilise surtout le modélisateur *pouvoir* *-ibilmek* (infinitif), qui s'attache à la fin du verbe concerné comme un suffixe.

*Tuzu uzatabilir misin ?*  
Tu peux me passer le sel ?  
*Tuzu uzatamaz misin ?*  
Tu ne peux pas me passer le sel ?  
la forme négative est impolie  
*Tuzu uzatırsan çok sevindir*  
Je serais très contente si tu me passais le sel

6.1.1.7. en thaï

Les structures avec *pouvoir* sont à éviter pour ne pas mettre en doute la capacité du destinataire. On trouvera plutôt :

*/chua:y-3+song+2+klva:-1+haj:-3+chan-5+n@:j-1/*  
(*n@:j-5* = particule de demande)  
Aide-moi à me procurer du sel  
*/ka-2ru-4 na:-1+?ja:-2+c@:t-2+rot-4+khwa:ng-5+pra-2tu:-1/*  
Vous auriez pitié de ne pas garer la voiture devant la porte

6.1.1.8. en japonais

On utilise surtout le modélisateur *kudasai*.

*Kudasai* : impératif contracté de l'expression '*kudasari + mase*' (s'il vous plait ou faites\_moi la faveur de)  
*Shio wo to-tte kudasai.*  
(Sel + particule COD + passer + *kudasai*)

6.1.1.9. en chinois

Plus la phrase est longue, plus le locuteur est poli ou humble par rapport à son interlocuteur. Ainsi quelques mots ou unités lexicales s'ajoutent au noyau de la phrase.

請你搬一下這張桌子好嗎？謝謝

*prier-tu-déplacer-un peu-ce-zhang/Cls-table-d'accord ?-merci*

Je te prie de déplacer un peu cette table d'accord ? Merci

(我)可不可以麻煩您搬一下這張桌子，謝謝

*(je)-pouvoir ou pas-déranger-vous-déplacer-un peu-ce-zhang/Cls-table,-merci.*

Je peux ou pas vous déranger pour déplacer un peu cette table, merci.

#### 6.1.1.10. en thaï/japonais

Il existe un grand nombre d'expressions honorifiques adaptées aux divers types de rapports sociaux. Ces formes affectent le vocabulaire, la conjugaison des verbes et certaines tournures de phrases.

#### 6.1.2. Conclusion

L'acte de requête est un acte de langage universel. Ses réalisations varient d'une langue à l'autre. Dans la traduction, il faut donc tenir compte de ces variations culturelles pour éviter les fautes pragmatiques : les contresens (requête, offre, question) et les impolitesses. Les dictionnaires doivent donc aussi tenir compte de toutes ces particularités.

### 6.2. *Proverbes et expressions figées*

La langue est un mode d'expression de la réalité, pour cette raison elle tente de refléter les choses qu'elle nomme comme la nature, l'homme, sa vie. L'expression figée et le proverbe jouent entre autres le rôle de « miroir du peuple ». Cependant les peuples ne se voient pas tous à travers le même miroir.

#### 6.2.1. La vie quotidienne a enrichi les locutions.

en français : avoir un oeil à la poêle et l'autre au chat : surveiller deux choses en même temps.

en russe : *V tesnote da ne v obide* : quand on est serré, on n'est pas lésé.  
Trad. Litt.: Pas de place mais pas de vexation non plus  
(esprit communautaire russe !).

en chinois : 人一走，茶就凉。  
Trad.Litt : Dès qu'on part, le thé devient froid  
Trad. Loin des yeux, loin du coeur

en japonais : *Kaze no tayori.*  
Trad. Litt. : les nouvelles du vent  
Trad. Le petit doigt me l'a dit

#### 6.2.2. Le corps est source de nombreuses locutions

en français : *tirer une épine du pied* avec deux sens  
- ôter l'épine qui s'est plantée dans le pied  
- se délivrer d'une situation embarrassante.

#### 6.2.3. Les animaux apparaissent souvent dans les locutions



*les murs ont des oreilles* donnera en coréen l'oiseau écoute ce qui se dit le jour et la souris écoute ce qui se dit la nuit.

#### 6.2.4. Institutions, techniques, coutumes et mœurs

Par exemple : la chasse et la pêche

en français : *donner le change* (tromper)

Vient de la chasse à courre où il arrive que le cerf change sa voie en suivant la trace d'une autre bête dont l'odeur dérouté les chiens en les entraînant sur une autre piste. De même :

en espagnol : *Cazar largo* : voir loin, avoir du flair

Deux exemples de chengyus chinois.

(*chengyu* : construction figée chinoise quadrisyllabique)

混水摸魚

*hùn shuǐ mō yú* : ① pêcher en eau trouble ② tirer au flanc trouble-eau-pêcher-poisson

Trad. Litt. : pêcher des poissons dans de l'eau trouble

大海撈針

*dà hǎi lāo zhēn* : chercher une aiguille dans une botte de foin grand-mer-pêcher-aiguille

Trad. Litt. : pêcher une aiguille dans la mer

撈 : *lāo* : *pêcher* signifie « essayer de récupérer quelque chose dans l'eau »

#### 6.2.5. La religion avec son droit, sa liturgie et ses coutumes est à l'origine de locutions

en chinois : (平時不燒香, )臨時抱佛腳

(*píng shí bù shāo xiāng*), *lín shí bào fó jiǎo* : se presser au dernier moment  
Trad. Litt. : (en temps normal ne brûler pas d'encens) au dernier moment, en cas d'urgence, serrer les pieds de Bouddha dans les bras

en français : *avoir voix au chapitre* (avoir le droit de donner son avis)

*ne savoir à quel saint se vouer* : ne savoir à qui recourir

*S'en laver les mains* : dégager sa responsabilité (comme le fit Ponce Pilate après la condamnation de Jésus par les juifs)

en espagnol : *comulgar con ruedas de molina* : croire, avaler n'importe quoi

en polonais : *Jak u Pana Boga za piecem* : en sécurité, dans de bonnes conditions ; comme un coq en pâte

Trad. Litt. : Comme chez Seigneur Dieu derrière son fourneau

*Bać się jak diabeł święconej wody* : avoir une peur bleue

Trad. Litt. avoir peur comme le diable (a peur) de l'eau bénite.

en japonais : *Mutsu-go no tamashii hyaku made-mo* :

Trad. Litt. L'âme à l'âge de six ans ne change plus jusqu'à l'âge de cent ans : dans sa peau mourra le loup

*Tamashii* : âme

#### 6.2.6. Autres comme l'équitation, la guerre, les jeux, la médecine :

Lorsque l'on tombe de cheval on peut *se remettre en selle*, si on veut retenir l'animal, *on tire sur la bride*

en espagnol : *salirle a uno el tiro por la culata*

en prendre plein la figure

Expression qui évoque la mésaventure du tireur qui voit son arme lui exploser au visage.

*brouiller les cartes* : semer la confusion

*se faire de la bile* : 1) s'inquiéter 2) se désopiler la rate

#### 6.2.7. et pour finir : le folklore

en espagnol : *ser más agarrado que un chotis* : être radin

*Chotis* est une danse folklorique.

### **6.3. Comment faire coïncider les miroirs de chaque peuple, phraséologie et traduction**

« Les civilisations différentes se transmettent des idées et des expressions au cours des échanges dans des aspects divers. L'origine d'une expression peut ainsi être la traduction littérale d'une expression étrangère. A titre d'exemple, « jeter l'éponge » est une expression française provenant d'une traduction littérale de l'expression anglaise « *throw up the sponge* » dans le contexte de la boxe (Rey et Chantreau, 1997 : 366) « LIN H.-I. (2004) ».

L'objectif est de trouver l'équivalent de sens exact à un proverbe ou à une expression figée même si la forme syntaxique n'est pas respectée.

Plusieurs cas se présentent :

- Equivalence sémantique, lexicale et structurelle totale
- Equivalence sémantique mais équivalence lexicale et structurelle partielle
- Equivalence sémantique mais totale différence lexicale et structurelle

#### 6.3.1. Equivalence sémantique, lexicale et structurelle totale entre proverbes des diverses langues

français : *Œil pour œil, dent pour dent*

espagnol : *Ojo por ojo, diente por diente*

portugais : *Olho por olho, dente por dente*

italien : *Occhio per occhio, dente per dente*

roumain : *Ochi pentru ochi, dinte pentru dinte*

russe : *Око за око, зуб за зуб*

polonais : *Oko za oko, ząb za ząb*

turc : *Göze göz, dişe diş*

coréen : *nunenîn nunîlo ienîn ilo*

chinois : *以眼還眼, 以牙還牙。*

arabe : *al 9aynu bi l 9ayni wa ssinnu bi ssinnu wal baadi'u aDlamu*

Remarques : en arabe *wal baadi'u aDlamu* a été ajouté justement pour donner au proverbe le « cachet arabe »

japonais : *Me niwa me wo, ha niwa ha* (*wo* : œil, *niwa* : pour, *ha* : dent)

on obtient une traduction produite par le calque avec la langue française

### 6.3.2. Equivalence sémantique mais équivalence lexicale et structurelle partielle

français : *loin des yeux, loin du cœur*  
espagnol : *ojos que no ven, corazón que no siente*  
portugais : *longe da vista, longe do coração*  
italien : *lontano dagli occhi, lontano dal cuore*  
roumain : *Ochii care nu se văd, se uită*  
russe : *С глаз долой – из сердца вон* : *S glas doloj - is serdtsa von*  
polonais : *Co z oczu to i z serca*  
turc : *Gözden ýrak, gönülden ýrak*  
coréen : *nunesô môlôjimyôn maumesô môlôjimda*  
thaï : */klaj+I+ta:-I+klaj+I+caj+I/* loin+yeux+loin+cœur  
chinois : 人一走, 茶就凉。 Trad. Litt. : Dès qu'on part, le thé devient froid  
arabe : *ba9iidun 9ani al 9yuni, ba9iidun 9anil qalbi*  
sanskrit : *nethra door satyam, hrudhaya door satyam*  
japonais : *Saru mono wa hibi ni utoshi*

On remarque que le domaine corporel (œil et cœur) est présent dans toutes ces expressions à l'exception du chinois (thé)

Concernant la syntaxe, la structure syntaxique française est identique à la structure portugaise et italienne.

ADV+ART+SUBST+ADV+ART+SUBST  
et aussi très proche du thaï.

Le coréen et le turc ont inversé les éléments.

L'espagnol, le roumain, le polonais ont paraphrasé le sens en employant une relative (introduites par *que, care, Co*), ce qui est différent du français, de l'italien et du portugais qui ont une équivalence sémantique et lexicale totale.

En arabe il s'agit d'une expression très récurrente dans le langage courant et qui a la même charge sémantique qu'en langue française.

### 6.3.3. Equivalence sémantique mais totale différence lexicale et structurelle

français : *Dans sa peau mourra le loup*  
espagnol : *Muda el lobo los dientes y no las mientes*  
portugais : *O que o berço dá a tumba o leva*  
italien : *Il lupo perde il pelo ma non il vizio*  
roumain : *Lupul își schimbă părul dar năravul ba*  
Trad. Litt. : le loup change son poil mais son caractère non  
russe : *Какое в колыбельку, такое и в могилку* : *Kakov v kolybel'kou, takov i v moguilkou*  
Trad. Litt. : Tel au berceau, tel au tombeau  
polonais : *Natura ciagnie wilka do lasu*  
Trad. Litt. : la nature appelle le loup à la forêt  
coréen : *akinîn yônjekajina akinida* : Le mauvais ne s'amende pas  
Trad. Litt. Le mauvais est toujours le mauvais  
thaï : */cha:t-3+sva:-5+maj-3+thing-4+la:j-I/*  
espèce+tigre+ne pas+abandonner+trait

chinois : 狗改不了吃屎 *gǒu gǎi bù liǎo chī shǐ*

chien-changer-non-entièrement-manger-merde

Trad. Litt.(fam) : Un chien ne changera pas (de son habitude) de manger de la merde

japonais : *Mutsu-go no tamashii hyaku made-mo*

Trad. Litt. : l'âme à l'âge de six ans ne change plus jusqu'à l'âge de cent ans

Dans les langues française, espagnole et italienne, les traductions s'appuient sur une image animale et corporelle.

*lobo, lupo*, loup ;

*dientes, pelo*

En portugais, on a la comparaison de l'homme à sa naissance (*berço* : berceau) et à sa mort (*tumba* : sépulture)

Pour les langues turque et arabe, cette expression figée n'existe pas.

En thaï, on attribue une qualité aux tigres, grands chasseurs de nature. Le thaï se donne une image des tigres qui meurent avec dignité sans « changer de peau ».

Nous avons donné ici plus particulièrement des traductions littérales (Trad. Litt.) en français afin de mettre en évidence les grandes divergences dans la formulation linguistique d'une même idée. On voit aussi qu'une formule de départ en langue soutenue peut donner comme résultat une traduction plutôt familière.

D'un point de vue sémantique, toutes ces traductions veulent bien dire la même chose mais ici le lexique est fondé sur d'autres images et la forme syntaxique est différente.

## **7. Quelques propositions pour la traduction automatique et assistée**

### **7.1 Les dictionnaires**

Comme le montre le rapport Pointer « POINTER (1996) » il existe des outils qui permettent des accès interactifs de manière conviviale aux banques de données terminologiques. Mais c'est précisément l'organisation de telles banques de données de collocations au sens large (non seulement les termes simples) qui nous intéresse. Des besoins comme la convivialité, l'appartenance des données à telle ou telle norme, sont, à notre avis, bien sûr importantes, mais elles ne servent qu'à accompagner toute cette organisation qui représente le véritable travail de fond du linguiste formaliste.

Les dictionnaires spécialisés interactifs de collocations (terminologiques et autres) multilingues devraient mettre en évidence de façon explicite des phénomènes linguistiques qui sont d'une importance primordiale pour le traducteur.

MultiCoDiCT « GREENFIELD, P *et al.* (1999) » est un prototype du noyau de ce type de dictionnaires sous la forme d'une plate-forme utilisée, programmée linguistiquement (ex. affichage de parties du discours, entre autres) et perfectionnée (pour les besoins de la représentation spécialisée du lexique) par de nombreux chercheurs. MultiCoDiCT prend en compte les phénomènes linguistiques « CARDEY, S., GREENFIELD, P. (2001) » et repose sur une spécification formelle « GREENFIELD, P. (1998) ».

Dans MultiCoDiCT, la représentation spécialisée du lexique permet le codage de collocations dans un contexte d'équivalences de significations dans un domaine donné et à travers un ensemble de langues. Cette représentation prend en compte les exigences d'intégrité de données, l'explicitation de contraintes du domaine (telles que l'interdiction de synonymie, de polysémie), et permet aussi, à partir des collocations, la génération automatique d'articles ainsi que celle de l'indexation des entrées. Il faut rappeler que MultiCoDiCT n'est pas qu'un système de dictionnaires de « mots », c'est aussi un système de dictionnaires (multilingue) de collocations et « segments ».

Plusieurs dictionnaires prototypes de collocations ont été conçus dans des domaines divers avec MultiCoDiCT dans différentes langues : *anglais-français*(6), *français-espagnol*(2), *anglais-coréen-français*, *arabe-français*, *espagnol-français-portugais*, *français-latin-portugais*, *chinois-français*, *thai-français*. Ces dictionnaires ont été élaborés comme prototypes de nature « archétypal » pour aider les recherches portant par exemple sur l'apprentissage, la terminologie, le figement, la traduction assistée et automatique.

Nous donnons ci-dessous un extrait du récapitulatif des dictionnaires créés et renvoyons le lecteur à l'article de GREENFIELD, P. (2003) pour une présentation plus détaillée des particularités non seulement des dictionnaires mais aussi et surtout des langues décrites.

Domaine	Type	Langues	Particularités	Auteur et référence
T.A.L.	Terminologie	anglais-français	Alternatives internes	ACHECHE, S. (1999)
Tourisme	Terminologie	français-espagnol	Mots polycatégoriels, Variantes de langue	CHAN NG, R. (1999)
Recherche clinique	Terminologie	anglais-français	Non réversibilité	GAVIEIRO, E. (1997)
Parties du corps	Expressions figées	français-espagnol	Multiples entrées pour une collocation, Polysémie, Synonymie	MORGADINHO, H. (1999)
Animalier	Expressions imagées	français-thai	Type de traductions	WERTHE-HENGSA, V. (2001)

Le dernier dictionnaire « WERTHE-HENGSA, V. (2001) » de ce tableau présente des particularités intéressantes que l'on ne rencontre dans aucun dictionnaire actuel et qui portent sur les types de traductions à partir des expressions imagées dans le domaine animalier. Les types de traductions sont les suivantes :

- Eq équivalent – fourni à condition que l'on puisse en trouver un
- TL traduction littérale – mot à mot
- SA sens analytique – traduction littérale reformulée en français correct
- SF sens fonctionnel – le sens véritable de l'expression traduite

## 7.2. La traduction automatique

Nous parlerons brièvement ici de 2 systèmes de traduction automatique, l'un nous servira à montrer les manipulations sémantiques nécessaires en vue d'obtenir des traductions correctes, l'autre donnera simplement des résultats lorsqu'il est appliqué à un domaine particulier.

### 7.2.1. Le système TACT coréen/français

Afin de traiter le problème d'entrées de dictionnaire lorsqu'à un verbe dans une langue correspondent 2 verbes dans une autres tel que :

*arroser* dans  
*arroser le poisson de jus de citron*

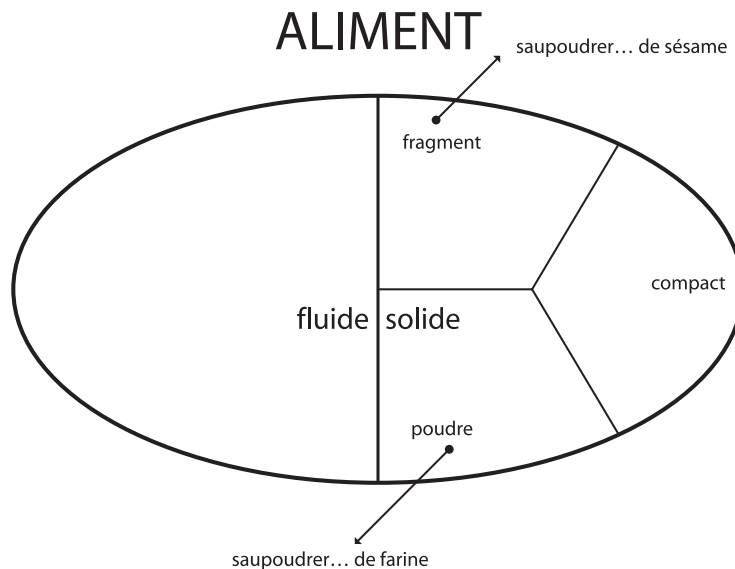
et

*saupoudrer* dans  
saupoudrer le poisson de farine

qui se traduisent par la même unité en coréen, l'opposition  
[liquide/pulvérulent]  
va permettre de marquer la différence dans les dictionnaires.

Les propriétés que nous utilisons sont conçues de façon relationnelle, nous utilisons la théorie SyGuLAC (Systemic Grammar using a Linguistically Motivated Algebra & Calculus) « CARDEY, S., GREENFIELD, P. (2005) » qui va permettre dans un premier temps de partitionner le lexique ici en fonction de traits sémantiques de la façon suivante :

TACT  
(Traduction Automatique Centre Tesnière)  
Coréen → Français  
Mi-Seon HONG



La propriété [liquide] s'oppose ainsi à la propriété [solide] et cette dernière comprend, par exemple, les sous-propriétés [poudreux], [fragmentaire] et [compact].

### 7.2.2. Le système TACT français-arabe

Dans le même ordre d'idée nous avons conçu un système de traduction automatique français-arabe « ALSHARAF H. *et al.* (2004) » qui va permettre lorsqu'il est appliqué par exemple au domaine du droit de générer le type de traduction suivante :

*l'ordre administratif connaît des litiges relatifs à l'organisation et au fonctionnement des services publics et aux contrats administratifs*

ال نظام ال اداري يحكم في ال نزاعات ال مرتبطات ب التنظيم و آلية ال خدمات ال عامة و ب ال عقود ال ادارية

et lorsqu'il est appliqué au domaine de la météo des traductions telles que

*ailleurs sur la moitié nord-ouest du pays, la matinée sera peu nuageuse et brumeuse sur le sud-ouest*

الى موضع آخر على ال قسم ال شمال-غربي ل ال  
بلد ال صبحية ستكون غائمة قليلا و معتمة على ال  
جنوب-غرب

## 8. Conclusion

Nous avons été maintes fois surpris de constater que certains systèmes pour passer d'une langue à une autre empruntaient une 3<sup>ème</sup> langue, l'anglais, ce qui fait que les résultats ne peuvent en aucun cas être bons. Par exemple, pour le mot « libre » en français nous avons trouvé 19 sens, qui peuvent se retrouver dans la langue du droit avec un sens spécifique au droit mais aussi un sens plus commun. Comment faire pour obtenir la bonne traduction ? Pour conclure, nous dirons simplement, que si l'on veut de bonnes traductions automatiques, il faut se limiter à un domaine précis, de plus notre expérience nous a montré que les systèmes doivent être conçus de façon unidirectionnelle dans un premier temps et seulement ensuite on peut envisager de travailler dans l'autre sens, ce qui va demander la création en fait d'un nouveau système. Si l'on cherche des résultats de qualité, l'analyse de la langue source devrait se faire en fonction de la langue cible, en sachant que ces analyses ne sont pas forcément réversibles.

## RÉFÉRENCES

- ALSHARAF, H., CARDEY, S. and P. GREENFIELD (2004) : « French to Arabic Machine Translation: the Specificity of Language Couples », *Actes de The European Association for Machine Translation (EAMT) Ninth Workshop*, Malta, 26-27 April 2004, Foundation for International Studies, University of Malta, pp.11-17.
- ACHECHE, S. (1999) : « Système de traduction semi-automatique des termes relatifs au T.A.L. (Traitement Automatique des Langues) », *Mémoire de Maîtrise d'anglais* mention Industries de la Langue, Besançon, 1999.
- CARDEY, S. and P. GREENFIELD (2001) : « The Analysis and Design of Computerised Lexicons », in *Lexis in Contrast* (B. Altenberg & S. Granger eds.), Benjamins, pp. 231-249
- CARDEY, S. and P. GREENFIELD (2005) : « Systemic Linguistics with Applications », *Proceedings of the 9th International Symposium on Social Communication*, Santiago de Cuba, January 24-28, 2005, Actas II, pp. 649-653
- CHAN NG, R. (1999) : « Prototype informatisé de dictionnaire du tourisme français-espagnol-français », *DEA*

*Sciences du langage*, Besançon, 1999.

- GAVIEIRO, E. (1997) : « Elaboration d'un prototype de lexique informatisé anglais/français, application au domaine de la recherche clinique », *DEA Sciences du langage*, Besançon, 1997.
- GREENFIELD, P. (1998) : « L'espace de l'état et les invariants de l'état des dictionnaires terminologiques spécialisés de collocations multilingues ». In *actes de la 1ère Rencontre Linguistique Méditerranéenne, Le Figement Lexical*, Tunis, les 17-18 et 19 septembre 1998, pages 271-283.
- GREENFIELD, P., CARDEY, S., ACHÈCHE, S., CHAN NG, R., GALLIOT, J., GAVIEIRO, E., MORGADINHO, H. et E. PETIT (1999) : « Conception de systèmes de dictionnaires de collocations multilingues, le projet MultiCoDiCT ». *Actualité Scientifique, AUPELF-UREF, actes du Colloque international AUPELF Réseau Lexicologie, Terminologie, Traduction, L'éloge de la différence : la voix de l'autre*, Beyrouth, 11-13 novembre 1999, pages 103-115.
- GREENFIELD, P. (2003) : « Le rôle de l'informatique dans le traitement et l'enseignement des langues », *Actes du Congrès international : Journées linguistiques franco-asiatiques*, Université Naresuan, Phitsanulok, Thaïlande, 20-22 août 2003, pp.69-84.
- LIN, H.-I. (2004) : « Vers une traduction automatique des expressions figées françaises en chinois : la traduction canonique », *Thèse de doctorat*, Besançon, France.
- MORGADINHO, H. (1999) : « Dictionnaire électronique français-espagnol. Expressions figées », *Mémoire de Maîtrise d'espagnol mention Industries de la langue*, Besançon, 1999.
- POINTER (1996) : « Pointer (Proposal for an Operational Infrastructure for Terminology in Europe) » 1996, <http://www.mcs.surrey.ac.uk/Research/CS/AI/Pointer/report/contents.html>
- WERTHE-HENGANGA, V. (2001) : « Etude de la traduction automatique en français des expressions imagées de la langue thaï (domaine animalier) », *DEA Sciences du langage*, Besançon, 2001.