# Les langues contrôlées, fondements, besoins et applications

# Projet LiSE (Linguistique et Sécurité)

**Sylviane CARDEY[1], Laurent SPAGGIARI[2], Dominique VUITTON[3]**

[1]Centre Tesnière, UFR SLHS, Université de Franche-Comté, 30 rue Mégevand, 25030 Besançon Cedex

[2] Airbus-France, 316, route de Bayonne, 31060 Toulouse

[3] Laboratoire "Santé-Environnement Rural-Université de Franche-Comté", Faculté de médecine et pharmacie, place Saint Jacques, 25030 Besançon

sylviane.cardey@univ-fcomte.fr, laurent.spaggiari@airbus.com, dominique.vuitton@univ-fcomte.fr

**Résumé** – Sont présentés dans cet article, les besoins en langues contrôlées dans les domaines demandant une grande sécurité ainsi qu'un bref état de l'art. Ceci nous amènera à montrer les problèmes que rencontrent les traducteurs automatiques ou non. Nous expliquerons également notre choix d'une seule langue contrôlée versus plusieurs langues contrôlées en vue de la traduction en plusieurs langues malgré le défi et la difficulté que cela pose tant à l'écrit qu'à l'oral.

**Abstract** – In this paper we present the requirements in respect of controlled languages in domains in which safety and security are of mandatory importance together with a brief state of the art. This leads us to uncover and highlight the problems faced by translators, whether these be humans or machines. We also explain our choice for one single controlled language rather than several in view of translation to several languages, this despite the challenges and the difficulties that this poses for both written and oral applications.

## 1. Introduction

Le projet LiSe[1] (Linguistique et Sécurité) dont il est question dans cet article a pour objectif la mise en place d'une méthodologie générale reposant sur l'**analyse des normes** linguistiques. Cette méthodologie doit servir deux applications principales, la première étant non seulement l'extraction de données porteuses éventuellement d'informations mais également la bonne interprétation de ces informations, quant à la deuxième, elle aura pour but de générer de l'information. Concernant la première application sur des domaines demandant une haute sécurité, les méthodologies actuelles qui fonctionnent toutes par mots clés se révèlent très insuffisantes, d'où le besoin d'une nouvelle approche qui fera l'objet d'un autre article. Ici, nous parlerons de la deuxième application, c'est-à-dire de la méthode qui va permettre de générer l'information sans ambiguïté, rapidement et dans plusieurs langues ceci pour les cas d'urgence ou de crise.

[1] Le projet LiSe est subventionné par l'ANR

## 2. Information et langues contrôlées

### 2.1 L'information pour une meilleure interopérabilité

En cas de crise internationale, pour avertir les populations ou lorsque des protocoles ou des médicaments sont transmis ou envoyés dans un pays ou des pays, comment s'assurer que les messages, les procédures, les notices sont traduits rapidement et bien ? Comme, à l'exception de rares contextes, on ne peut pas prévoir tous les messages qu'il sera nécessaire d'envoyer et de traduire dans les aéroports, en cas de tsunami, pour avertir la population, ni les instructions à donner, il faut prévoir la façon de les écrire afin qu'ils soient compris par tout un chacun et traduisibles par une machine dans plusieurs langues sans risque d'erreurs.

Le seul moyen d'y parvenir est d'utiliser **une langue contrôlée.** Nous voudrions mettre à la disposition des personnes chargées d'écrire les messages (qui peuvent être des alertes, des protocoles médicaux ou des instructions de natures diverses) des règles à respecter, afin d'assurer lisibilité, compréhensibilité et traductibilité. De plus il est

possible de faire de l'aide à la rédaction et de la correction assistée automatique. Aussi la langue contrôlée élaborée saura éviter les pièges de la langue (des langues), de la traduction, en produisant des textes en langue source de structures simples et non ambigus tant en ce qui concerne le lexique que la syntaxe.

Les deux exemples suivants montrent le bien fondé de notre recherche :

(http://www.bpe.europresse.com/WebPages/Search/Doc.aspx?DocName=news%C2%B720070815%C2%B7LM%C2%B70Q1508_1552030_529457)

BERLIN. Quarante-sept patients opérés du genou dans un hôpital de Berlin ont été victimes d'une erreur médicale en raison d'une mauvaise traduction de la notice concernant leur prothèse, selon le *Tagesspiegel*. L'indication d'origine évoquait une prothèse *"non modular cemented"* (non modifiable et devant être cimentée), traduite par un auto collant *" prothèse ne nécessitant pas de ciment ". - (AFP.)*

Tue, 06 Mar 2007
www.earthtimes.org
Extract:
An error in translating English instructions for the use of software probably led to the death by an overdose of X-ray radiation of four patients at a French hospital.

## 2.2 Les langues contrôlées

### 2.2.1 Etat de l'art

Ainsi que l'a écrit Goyvaerts (voir [1]), "Industry does not need Shakespeare or Chaucer, industry needs clear, concise communicative writing - in one word Controlled Language". Les Langues Contrôlées (LCs) sont d'un intérêt vital pour l'industrie (tant pour des raisons de sécurité qu'économiques). Les LCs tendent à résoudre des problèmes tels que la compréhensibilité par le biais de restrictions sémantiques et syntaxiques du langage. Il est nécessaire pour les industries désireuses de créer une LC de connaître ce qui existe déjà. Cependant, il ne semble pas y avoir à l'heure actuelle un panorama fiable disponible dans le domaine, ce qui est confirmé par le projet américain de création d'un « National Consortium to Advance Controlled Language and Computer Aided Translation » ; ce consortium servirait l'organisation et la mise à disposition des informations via internet. Dans l'attente de la création d'un tel consortium, nous avons établi un Etat de l'Art des LCs. Un découpage linguistique nous a permis de situer les langues contrôlées.

Comme on peut le voir dans la figure 1, les langues contrôlées intéressent des domaines divers tels que l'aéronautique, la météorologie, les services d'urgences (police, pompiers, marine, ambulance, etc.). Un utilisateur souhaitant avoir des informations sur une langue contrôlée particulière peut les obtenir rapidement par un simple click. Des cartes ont en effet été établies, elles donnent par exemple, la date de création, la durée de l'étude, l'organisme (entreprise, université, etc., qui détient le produit), le concepteur, la personne qui est en charge de la LC, ses applications, son contenu, une bibliographie. Voir aussi [1], [2].

Les différentes langues contrôlées sont classées en guides de rédaction papier (rose), but informatique (bleu), guides de rédaction informatisés (jaune), le blanc est utilisé quand il nous a été impossible de classifier la LC dans une des 3 catégories citées.

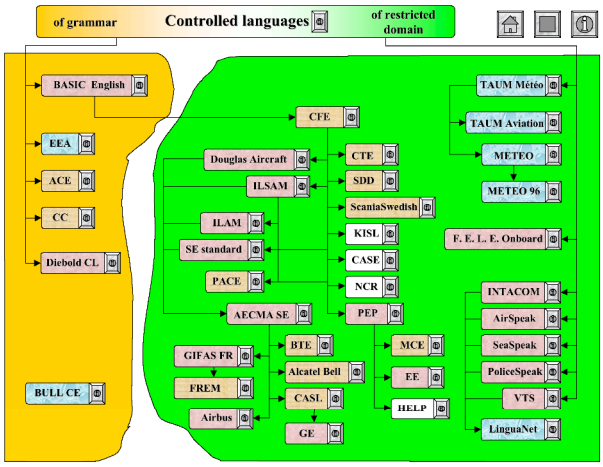


FIG. 1 : types de langues contrôlées

# 3. Notre langue contrôlée

## 3.1 Son but

Contrôler la langue signifie ainsi que nous l'avons dit, contrôler sa syntaxe et son lexique. Notre but ici est double et va vers des applications diverses.

Le premier objectif est d'aller vers une traduction fiable en plusieurs langues et le deuxième de réduire les interférences à l'oral dans une même langue ou encore entre deux ou plusieurs langues. Ce qui est nouveau, est qu'elle n'est pas orientée par le domaine mais tente d'être générale, quant aux interférences, celles-ci n'ont jamais été traitées. On peut également dire qu'il n'existe aucune langue contrôlée basée sur le français à part des débuts de travaux chez Airbus. La langue la plus utilisée comme base est l'anglais mais on trouve aussi du suédois, du chinois contrôlés depuis 1998.

### 3.1.1 Notre LC pour la traduction Automatique (TA)

Le but de notre langue contrôlée est de permettre à un rédacteur en cas d'urgence de rédiger un message, un protocole qui va éviter tous les pièges de la traduction. Pour ce faire, nous avons fait des études comparées entre plusieurs langues afin de dégager des normes. Que sont ces normes ? Partant du français, nous avons comparé plusieurs structures françaises qui actualisées dans le

discours vont donner un même sens. Nous avons ensuite retenu la plus simple et surtout la moins ambiguë à plusieurs points de vue. Premièrement toutes les ambiguïtés de structures ont été évitées, deuxièmement les ambiguïtés morphologiques et lexicales ont à leur tour été remplacées par des formes ou du lexique préconisés. Des structures équivalentes ont ensuite été recherchées en anglais, arabe, chinois, japonais et thaï et les mêmes travaux ont été effectués sur ces langues afin d'obtenir des structures équivalentes dans chacune des langues. Nous allons donner maintenant les étapes principales de cette première partie de la recherche sur notre langue contrôlée.

# 4. Etablissement des divergences et interférences entre les langues traitées

La phrase dans sa forme ne se présente pas de la même façon dans les différentes langues [3], [4]. Le concept de séparateur, le genre, le nombre, la flexion, le complexe verbal et sa construction, le classificateur, le nom sériel, la juxtaposition, la coordination, la propriété commune, l'anaphore entre autres sont différents ou absents selon les langues. En traduction automatique, doivent être traités les décalages : distinction sémantique plus fine dans la LC (Langue Cible) que dans la LS (Langue Source) ou l'inverse. Il faut tenir compte des divergences : les divergences de densité lexicale (une unité dans une langue pouvant correspondre à plusieurs « mots » combinés dans l'autre) ; les divergences catégorielles (un nom traduit par un verbe) ; les divergences syntagmatiques (un SN dans une langue pouvant se traduire en proposition conjonctionnelle en français par exemple) ; divergence de constructions syntaxiques (renversement de l'ordre des compléments) ; divergence anaphorique (zéro pour le chinois et pronom pour le français). Les déterminants varient ne serait-ce que lorsque l'on passe de l'anglais au français.

## 4.1 Structures divergentes

La première étape a donc consisté à détecter ce qui est commun et ce qui diverge dans les langues concernées ici. Nous avons par exemple ci-dessous deux structures syntaxiques, l'une est une structure arabe et l'autre une structure japonaise, elles sont complètement divergentes bien qu'elles soient utilisées pour exprimer la même chose :

arabe : opt(quest +يمكن+nver) + opt(ecd) + comps / compscc / compa / compacc + point

japonais : comps / compscc / compa / compacc + opt( ecd + opt(quest + vinf ) ) + point

Par contre on s'aperçoit que les deux structures suivantes l'une propre à l'arabe et l'autre à l'anglais sont identiques :

anglais : opt(neg) + pred + opt(mod) + opt(compacc/compa1) + opt(compacc/compa2) + opt(compacc/compa3)+ opt(mod) + opt(2pts+ liste) + point

arabe : opt(neg) + pred + opt(mod) + opt(compacc/compa1) + opt(compacc/compa2) + opt(compacc/compa3)+ opt(mod) + opt(2pts+ liste) + point

Ceci signifie qu'il a été possible de trouver dans le deuxième cas une structure similaire en anglais et en arabe mais que dans le premier cas, cela n'est pas possible. Pour ce dernier, il faut donc trouver des règles complémentaires qui vont permettre de passer d'une langue à l'autre.

## 4.2 Divergences lexicales

Voyons maintenant comment un même sens peut-être exprimé dans ces différentes langues et prenons le concept ou sème *risque* pour illustrer la variation de la forme dans l'expression. On découvre que sont utilisées différentes catégories grammaticales et du lexique spécifique, entre autres (extrait)

en français :
lexies simples spécifiques : attention/ danger
locutions verbales : prendre soin de, faire attention à…
adverbes, locution adverbiale : précautionneusement, prudemment, délicatement, avec délicatesse, avec soin, etc...
codes : lettres capitales, couleur rouge, point d'exclamation

en anglais :
lexique spécifique
adverbes
codes : rouge, lettres en majuscule, point d'exclamation

en arabe :
lexique : Nom
---> حذار\انتباه\احتراس\ملاحظة
Locutions verbales : V + Art.U + Nom
---> يرجى الانتباه

en chinois :
lexique :小心( attention !)
codes : rouge

en japonais :
lexique spécifique : 注意/危険
locutions verbales : 注意する/気をつける
adverbes : 注意して/気をつけて/注意を払って
codes : rouge, lettres en gras

en thaï :
lexique : □□□□□ (attention)
locutions verbales :□□□□□□□□□ (s'il vous plait, faire attention)

□□□□□□□□□□□ (ne pas être imprudent)

codes : lettres en gras, rouge

Ainsi des tables d'équivalences ont été dressées en vue de résoudre les problèmes de divergences et de trouver pour chaque concept un « équivalent forme » dans les autres langues.

## 4.3 Divergences grammaticales

Les divergences grammaticales sont représentées dans des tables également sur support informatique.

Certains problèmes tels que le *duel* ont débouché sur ce type d'explications que nous trouvons dans notre table d'équivalences

français : Ø

anglais : Ø

arabe : Le duel renvoie à deux individus. Il est formé à partir du suffixe ان pour les duels sujets (genre grammatical) et ين pour les duels compléments (genre grammatical)

chinois : marqueurs de duel :双/二/俩 (double)

japonais : (préfixe) '双'/'二'/'两' + N

thaï : marqueurs de duel : □□□ (double), □□□ (jumeaux) (ajouter ces mots après le nom)

Ce que nous faisons et qui n'existe pas est que nous contrôlons non seulement la langue de départ mais également la langue d'arrivée.

Pour ce faire nous avons eu besoin d'un grand corpus, environ mille textes au départ, nous en avons analysé plusieurs et ainsi au fur et à mesure des analyses, nous nous apercevons que les problèmes étant récurrents, nous avons de moins en moins de contrôle à faire sur les nouveaux textes. Ce corpus est constitué de protocoles, alertes, messages d'urgence de divers domaines et dans différentes langues.

## 4.4 Conclusion de la première étape

Ce que nous voulons c'est faciliter la tâche du rédacteur, nous ne pouvons de ce fait lui demander de rédiger son texte en français de plusieurs façons afin qu'il puisse être traduit dans différentes langues. Nous avons donc opté pour la rédaction d'un seul message mais ceci représente bien sûr un réel défi puisque ce message unique doit pouvoir se traduire automatiquement sans erreur dans plusieurs langues. Il faut effectivement gagner du temps dans ce genre de situation. Notre solution nous permet également de gagner en efficacité puisque nos règles de contrôle vont être quasiment uniques puisque qu'un seul contrôle doit nous permettre de passer à plusieurs langues.

# 5. Règles de contrôle

## 5.1 Exemple de règle

A partir de nos tables d'équivalences et de divergences, nous avons établi des règles de contrôle, celles-ci vont porter sur la syntaxe, la morphologie, le lexique.

Voyons un extrait de nos règles de contrôle

Règles :

**Comment rédiger une condition ?**

Cliquer sur le bouton "Condition".

Si vous groupez plusieurs conditions :

- commencer par la condition la plus spécifique ;
- terminer par la condition générale.

Exemple :

Si vous utilisez un cathéter d'hémodialyse :

- injecter 1mg/1mL d'Alteplase® dans chaque branche du cathéter ;
- ajouter le volume de NaCl 0,9% approprié au type de cathéter.

Sinon :

- injecter le volume d'Alteplase® approprié au volume du cathéter ;

Dans les deux cas

- ne pas laisser le médicament en contact plus de 72 heures.

## 5.2 L'interface utilisateur

Afin de faciliter la tâche du rédacteur, une interface a été pensée. Elle permet de guider l'utilisateur dans la rédaction de son message. Par exemple, dans certaines langues, il est impossible d'utiliser la traduction de *jeter* pour des objets liquides tels que l'eau. On doit utiliser un équivalent de *verser*, aussi le rédacteur ne devra pas écrire

*ne pas **jeter** d'eau sur de l'huile en feu*

mais

*ne pas **verser** d'eau sur de l'huile en feu*

Ce qui reste tout à fait correct en français et facilite grandement le passage vers les autres langues.

Aussi notre interface utilisateur lui indiquera

- **Verser** QQC[liquide] sur QQC/QQN
- **Verser** QQC[liquide] dans QQC

- **Jeter** QQC[solide] sur QQC/QQN
- **Jeter** QQC[solide] dans QQC
- **Jeter** QQC[solide] à la poubelle

| Négation | Verbe à l'infinitif |
|---|---|
| Ne pas | ~~jeter~~ verser |

| Quelque chose | Préposition |
|---|---|
| d'eau | Sur |

FIG. 2 : exemple d'interface utilisateur

Notre interface permet ainsi de faire appel à la sémantique par le biais de l'utilisateur. Elle aide aussi le rédacteur dans l'utilisation des règles pour créer son message. Elle permet de nous assurer de la conformité du message avec les règles préconisées et de la fiabilité des résultats.

# 6. Le contrôle de l'oral

Nous avons déjà dit que les ambiguïtés se situent à tous les niveaux, morphologique, lexical, syntaxique, sémantique et/ou phonétique.

Plusieurs mots de langues différentes peuvent partager une identité phonique ou graphique, sans pour autant avoir un sens commun ou avoir un sens commun mais se prononcer différemment (*passe, pass* (si on parle d'examen dans le cas du français, on ne sait pas si on a réussi, pour l'anglais, on a réussi)) ; *cul de sac* (existe en anglais et en français mais se prononce différemment).

Aussi la deuxième partie importante de notre langue contrôlée est qu'elle ne doit présenter aucune ambiguïté orale inter-langue ou intra-langue. Un exemple montrera la nécessité de telles recherches.

Une phrase ambiguë a été à l'origine de l'accident le plus meurtrier de l'histoire de l'aviation civile (Ténérife, 1977). La phrase ambiguë, adressée par un pilote hollandais (ce dernier ayant fait une confusion avec sa langue natale) à la tour de contrôle *We are at take offf* a été mal interprétée par cette dernière comme *We are waiting at takeoff point* au lieu de *We are already on the takeoff roll.* La conséquence a été que la tour de contrôle n'a pas dit au pilote d'abandonner le décollage ce qui a entraîné la collision avec un autre avion présent sur la piste. Cet accident a coûté la vie à 583 personnes.

## 6.1 Exemple d'un des problèmes que nous traitons

Le manque de différenciation entre certains phonèmes chez les locuteurs ou récepteurs peut provoquer des confusions qui peuvent parfois avoir de graves conséquences.

Le problème dont nous voulons parler ici est qu'un message peut être prononcé avec divers accents et écouté par des récepteurs qui n'ont pas forcément le même système phonétique de référence dans leur langue respective.

Voyons tout d'abord le système d'opposition des phonèmes consonantiques de l'anglais (l'anglais nous intéresse dans le projet car, entre autres, la langue utilisée en aéronautique est l'anglais américain [5], [6]). Nous utilisons les symboles SAMPA (système phonétique)

| p | t | k | f | T | s | S | tS |
|---|---|---|---|---|---|---|---|
| b | d | g | v | D | z | Z | dZ |

### 6.1.1 Quasi-homophones et reconnaissance

E. Gavieiro-Villatte, dans sa these [1], fait déjà mention de ces mots qui dans une même langue ont une graphie tellement proche qu'ils pourraient être confondus plus particulièrement en situation de stress. C'est le cas par exemple des mots qui ne diffèrent que par un seul de leurs phonèmes ou lettres. Nous nous sommes donc également penché sur ce problème afin d'éviter l'emploi d'homophones ou homographes dans une même langue mais également de ce que l'on peut appeler quasi-homophones et quasi-homographes. Nous pouvons définir ces deux derniers de la façon suivante :

des unités sont quasi-homophones ou quasi-homographes lorsque seulement un ou deux de leurs constituants (ici les lettres et phonèmes) sont différents. Cela signifie que l'une d'elles peut avoir un phonème en moins ou en plus (*aft-after*), un phonème différent (*check-deck*, *feed-feet*), ou encore deux phonèmes différents (*flap-slat*).

Aussi, nous devons, entre autres, déconseiller l'emploi de ce type de paires dans la même langue ou dans des langues différentes qui ne diffèrent que par un phonème ou deux lorsqu'on les prononce.

Par exemple on trouve l'opposition *d/t* dans *feed* et *feet*.

Des phonèmes qui occupent la même position dans des mots peuvent aussi être confondus comme *check* et *deck*, *flap* et *slap* et pour le français *dessus* et *dessous*.

Prenons un autre exemple en français, en effet, ce type de problème se retrouve dans toutes les langues. Si on dit à un pilote de *perdre de l'altitude*, il est important qu'il ne comprenne pas *prendre de l'altitude*. Les unités *altitude/attitude/latitude* présentent le même type de confusion et surtout pour l'œil les lettres sont très proches. Ce problème n'est pas surprenant si l'on sait que la lecture ne se fait pas de façon linéaire.

### 6.1.2 La lecture

On comprend mieux les problèmes cités précédemment en lisant les deux textes qui suivent. En effet tout individu capable de lire l'anglais ou le français pourra lire ce qui suit :

> it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a total mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the ...

l'odrre des ltteers dnas les mtos n'a pas d'ipmrotncae, la suele coshe ipmrotnate est que la pmeirère et la drenèire soit à la bnnoe pclae. Le rsete peut êrte dnas un dsérorde ttoal et vuos puoevz tujoruos lrie snas porlbème. C'est prace que le creaveu hmauin ne lit pas chuaqe ltetre elle-mmêe, mias ...

C'est une des raisons pour lesquelles en langue contrôlée on doit éviter l'emploi de termes qui présentent les mêmes lettres ou phonèmes que d'autres termes.

## 6.2 Génération et interprétation des sons du langage

Une mauvaise prononciation peut générer des homophones et changer le sens d'un message. Nous allons montrer ci-dessous quelques problèmes qu'une prononciation approximative peut engendrer.

### 6.2.1 Divergences phonétiques

Tout d'abord nous avons travaillé avec des linguistes de langue maternelle différente et nous avons noté également les phonèmes de l'anglais qui n'existent pas dans leur langue respective. Nous avons ensuite retenu les phonèmes qu'un non spécialiste de l'anglais emploie afin de remplacer les phonèmes anglais qui n'existent pas dans sa langue. Finalement nous pouvons montrer un cas extrême ci-dessous (figure 3), de mots anglais prononcés par un Thaïlandais et les ambiguïtés produites.

Les phonèmes suivants n'existent pas en thaï :
[T], [D], [f], [v], [s], [z], [l], [r]

| English phonemes | Sound in Thai | Ambiguities |
|---|---|---|
| [T] | [t], [d] | birth → bird |
| [D] | [d] | they → day |
| [f] | [p] | half → harp |
| [v] | [w] | vine → wine |
| [s] | [d] | bus → bud |
| [z] | [t] | buzz → but |
| [l] | [n] | ball → born |
| [r] | [l] | free → flee |

FIG. 3 : ambiguïtés résultant de l'absence de certains phonèmes anglais dans la langue thaï

Nous n'irons pas plus loin dans le détail ici, mais nous pouvons déjà constater que *half* par exemple se confond avec *harp* et que *ball* est confondu avec *born*.

### 6.2.2 Quelques remarques

Nous avons remarqué que les phonèmes anglais absents dans les autres langues étudiées sont [T] et [D]. Une table complète des interférences pourra servir à la langue contrôlée. Ce qui rend les choses encore plus complexes est que ces phonèmes prononcés de façon approximative ou incorrecte peuvent à leur tour être interprétés différemment encore selon la langue maternelle du récepteur [7].

# 7. Le sens pour la machine

## 7.1 Langue commune/langue de spécialité

Dans toute langue de spécialité les unités peuvent être :

- seulement éléments de la spécialité et ne présentent alors aucune difficulté, ni pour la reconnaissance, ni pour la traduction : *contre passation, crédit-bail, quasi-contrat ;*

- éléments de la spécialité mais aussi de la langue standard et peuvent avoir des sens différents en fonction de l'appartenance à l'une ou à l'autre ; elles peuvent aussi apparaître en langue de spécialité avec leur sens en langue standard.

On peut avoir la partition suivante :

- unité du domaine juridique par exemple avec un emploi standard dérivé *droit, justice, valide, arbitre ;*

- chaque terme peut aussi être membre d'une expression *droit de légation, droit mobilier, droit moral ;*

- unité dont le sens principal appartient à la langue standard avec un sens dérivé en langue de spécialité *siège, parquet,* que l'on peut aussi trouver dans des composés *parquet général*

Certaines unités ont le même sens dans les 2 langues, standard ou de spécialité

- *étude*
- *objection*
- *hypothèse*

Certaines unités peuvent avoir des sens divers à l'intérieur de la spécialité même

- *contrat (de gré à gré/aléatoire)*
- *force (de la chose jugée/majeure)*
- *action (coercitive/civile)*
- *libre (pratique/échange*)

ce dernier terme (*libre*) peut avoir jusqu'à 19 sens [8].

## 7.2 Les prépositions

Simplement en regardant la figure 4, on se rend compte des problèmes de traduction des propositions, elles ont donc été contrôlées dans leur emploi

| Langues | Préposition | Traductions possibles | *Exemples* |
|---|---|---|---|

| | | | |
|---|---|---|---|
| **arabe** | *à* | فى | *je vis **à** Toronto.* |
| | | على | *je te verrai **à** six heures* |
| | | ل | *ce jouet est **à** Gilles* |
| | | ø | *une cuillère **à** soupe* |
| | | ب | *des livres **à** dix dollars* |
| **chinois** | *par* | 从 | ***par** la fenêtre, par la poste,* |
| | | 每 | *une fois **par** an* |
| | | ø | *par exemple* |
| | | 用 | *prendre **par** la main.* |
| **japonais** | *sur* | にかけた | *faire attention à l'huile **sur** le feu et aux grille-pains* |
| | | の上の | *objets **sur** l'étagère* |
| | | に | *ne pas verser d'eau **sur** de l'huile en feu* |
| | | を | *ramper **sur** le sol.* |
| **thaï** | *par* | □□□□ | *remplacer qqn par qqn* |
| | | □□□□□□ | *finir par* |
| | | □□□□□□□□ | *passer par* |
| | | □□□ | *par courrier* |
| | | □□□ | *une fois par an* |

FIG. 4 : emploi des prépositions des langues étudiées

L'exemple suivant montre bien le problème :

*Nous avons convenu avec le secrétaire **de** l'agence **de** voyage **de** l'heure **de** départ **de** l'autobus.*

Où est le ***de*** qui s'attache à *convenir*

Pour la traduction automatique, ceci est loin d'être simple, il faut donc soit éviter ce type de construction soit le contrôler.

Il en est de même pour les ambiguïtés du type suivant :

*Elle ne rit pas parce qu'il est idiot*

où il y a deux interprétations, en effet dans un cas ***elle** rit* dans l'autre ***elle** ne rit pas*

# 8. La théorie

## 8.1 Approche systémique

Le langage naturel est complexe et bien que les langues soient des systèmes ouverts, la linguistique en tant que discipline progresse en étudiant de façon aussi exhaustive que possible les phénomènes linguistiques ; heureusement, les langues présentent des régularités mais il faut les mettre en évidence. Les techniques de modélisation pour le langage naturel ne doivent pas simplement admettre cette complexité, ces régularités et le fait que le langage n'est pas fixe mais aussi et surtout elles doivent mettre à notre portée la façon dont il fonctionne. Un modèle ne doit pas uniquement servir à la description d'un phénomène bien précis mais il doit pouvoir être une ressource permettant de résoudre d'autres phénomènes linguistiques couverts par ce modèle aussi bien que ceux non encore mis en évidence pour lesquels ce modèle et/ou d'autres pourront être utilisés.

## 8.2 De quoi a-t-on besoin ?

Nous avons besoin

- de techniques d'expression linguistique permettant la modélisation exhaustive analytique et compositionnelle et donc l'exploitation de phénomènes linguistiques intra et inter-langues ;
- de représentations économiques en intension plutôt qu'en extension qui regroupent ensembles et énumérations tout en étant explicites ;
- de rendre la validation aisée – les erreurs doivent être faciles à identifier et les « benchmarks » simples à construire ;
- de rendre la vérification aisée – avec des représentations bien formées (l'automatisation pouvant ainsi utiliser des techniques d'interprétation abstraites) ;
- de la possibilité de mesurer la qualité entre autres automatiquement ;
- d'optimisation pour obtenir une vitesse de calcul spécifique pour les applications qui en ont besoin ;
- que les applications ne soient pas limitées à l'utilisateur final mais utiles aux linguistes comme outils de recherche pour une extension des applications par exemple.

La théorie permet :

- de décrire les phénomènes linguistiques de façon aussi exhaustive que possible ;
- de traiter différentes langues ;
- diverses applications ;

- de faire des calculs.

## 8.3 La théorie en bref

La théorie préconise la décomposition ou recomposition de phénomènes linguistiques afin de mieux les analyser.

Au lieu de lister tous les éléments, il faut essayer de classer, d'ordonner, de ranger, de grouper afin de définir si certains de ces éléments qui font partie d'une langue peuvent fonctionner en tant que système à part entière ou comme systèmes en interrelation, ceci en fonction de ce qui doit être démontré ou résolu.

Face à un problème ou un phénomène précis, il faut choisir les éléments nécessaires et les structurer en système. Le problème devient ainsi manipulable car appréhendable. Le lexique, la morphologie, la syntaxe, la sémantique ne peuvent pas être séparés lorsque l'on traite de phénomènes de langues.

Comment faire une analyse systémique ?

Il faut :

- identifier le problème à traiter ;
- construire le système ou les systèmes, ces derniers pouvant être en interrelation.
- puis décrire le problème en utilisant le ou les système(s) afin de le résoudre si besoin.

## 8.4 Applications

- différentes langues ;
- différents problèmes ;
- différents domaines.

Nous donnons ci-dessous un exemple de réutilisation d'un domaine à l'autre – domaine culinaire que nous avons déjà traité en coréen [9] au domaine secours d'urgence que nous traitons dans le cadre du projet LiSe.

En français on aura le verbe *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]
a permis de marquer la différence dans les dictionnaires.

Les propriétés que nous utilisons sont conçues de manière relationnelle.

La propriété [liquide] s'oppose ainsi à la propriété [pulvérulent] et cette dernière comprend, par exemple, les sous-propriétés [poudreux] et [fragmentaire] (voir figure 5).

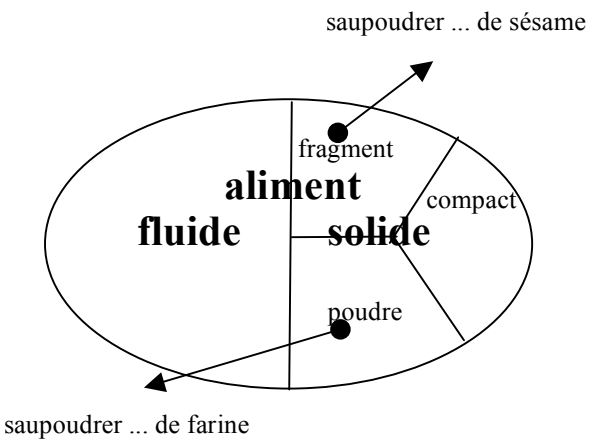


FIG. 5 : découpage systémique

Cette exemple de besoin de décompositions et classements se retrouve dans nos analyses, voir ci-dessus pour *verser/jeter* (paragraphe 5.2)

La théorie fera l'objet d'une publication à part entière en fin de projet avec les applications.

# 9. Validation

Des normes sont établies qui vont permettre d'écrire des messages, alertes, protocoles ou autres et améliorer la lisibilité, compréhensibilité et la traductibilité qui sont et seront les principaux critères d'évaluation. Des règles de morphologie sont conseillées ainsi que les structures préconisées et celles à éviter.

Nous avons besoin pour valider les résultats, au fur et à mesure de l'avancement des tâches, d'utilisateurs finals pour le data et sense-mining et la traduction automatique et de rédacteurs et utilisateurs finals pour les langues contrôlées. Les partenaires ont désigné ces utilisateurs [10]. Du côté aéronautique, ce seront les rédacteurs et pilotes de langues différentes ; pour les médecins, ce seront ceux qui écrivent des protocoles (chirurgiens et autres) et les utilisateurs de ces protocoles traduits automatiquement, ces derniers se trouvant dans les différents pays des langues traitées (France, Chine, Japon, Australie, Thaïlande, Liban). Pour la Sécurité en général, ils sont déterminés en fonction des besoins.

# 10. Conclusion

Nous avons essayé de montrer dans cet article l'utilité des langues contrôlées, ce qu'elles sont, les grandes étapes pour parvenir à les définir en général et dans nos applications actuelles et surtout comment nous résolvons les problèmes de communication internationale (diffusion d'informations et de messages).

La théorie ainsi que la méthodologie générale sont fondées sur un modèle micro-systémique original,

développé au Centre Tesnière [11], pour l'analyse et la génération de langues. Ce modèle est applicable sur toutes sortes de langues et a été implémenté pour des applications variées [12]. De par sa nature, la méthodologie répond au modèle de qualité systémique et permet validation et traçabilité, toutes deux essentielles pour des applications de sécurité critique. De nombreux domaines pourraient tirer parti de ces recherches.

# Références


[1] E. Gavieiro-Villatte, *Vers un modèle d'élaboration de la terminologie d'une langue contrôlée ; application aux textes d'alarmes en aéronautique pour les futurs postes de pilotage,* Thèse de doctorat, direction Sylviane Cardey, Centre Tesnière, Université de Franche-Comté, France, 1999.

[2] S. Cardey, P. Greenfield, S. Vienney (eds), *Machine Translation, Controlled Languages and Specialised Languages,* Lingvisticæ Investigationes, John Benjamins, ISNN 0378-4169/E-ISSN 1569-9927, 2005.

[3] K. Kuroda et H.-L. Chao (eds), *Divergence dans la traduction entre les langues orientales et le français*, BULAG n°30, ISSN 0758 6787, PUFC, CID, année 2005.

[4] S. Cardey, *Le miroir des peuples : phraséologie et traduction*, in Le français comme médiateur de la diversité culturelle et linguistique, Ministère de la culture et Ambassade de France en Thaïlande, 2007.

[5] AECMA., Association Européenne des Constructeurs de Matériel Aérospatial, *Simplified English, a guide for the preparation of aircraft maintenance documentation in the international aerospace maintenance language*, Bruxelles, Gulledelle, 1995.

[6] L. Spaggiari, F. Beaujard, E. Cannesson, *A controlled language at Airbus*. In Machine Translation, Controlled Languages and Specialised Languages, edited by Sylviane Cardey, Peter Greenfield, Séverine Vienney. Lingvisticæ Investigationes, tome XXVIII, John BenjaminsSN 0378–4169 / E–ISSN 1569–9927, 107-122, 2005.

[7] S. Cardey, *How to avoid interferences with other languages when constructing a spoken controlled language*, In Proceedings of the International Conference « La comunicazione parlata/Spoken Communication », 23rd-25th February 2006, Naples, Italy, 2007.

[8] H. Alsharal, S. Cardey, P. Greenfield, *French to Arabic Machine Translation: the Specificity of Language Couples*, in Proceedings of The European Association for Machine Translation (EAMT) Ninth Workshop, Malta, 26-27 April 2004, Foundation for International Studies, University of Malta, pp.11-17.

[9] S. Cardey, P. Greenfield, M.-S. Hong, *The TACT machine translation system: problems and solutions for the pair Korean-French*. In: Translation Quarterly, 27, The Hong Kong Translation Society, Hong Kong, pp. 22-44, 2003.

[10] P. Kern, H. Wen, N. Sato, D.-A. Vuitton, S. Bresson-Hadni, *WHO classification of alveolar echinococcosis: Principles and application.* Parasitol Int. 2006;55:S283-7, 2006.

[11] S. Cardey, P. Greenfield, *Systemic Linguistics with Applications, in Linguistics* in the Twenty First Century, Cambridge Scholars Press, United Kingdom, ISBN 1904303862, pp. 261-271, 2006.

[12] X. Wu, S. Cardey, P. Greenfield, *Some Problems of Prepositional Phrases in Machine Translation*. In: Proceedings of FinTAL 2006, 5th International Conference on Natural language Processing, Turku, Finland, 23-25 August 2006, Springer-Verlag – LNAI 4139, ISBN 3-540-37334-9*,* pp. 593-604, 2006.



Nous tenons à remercier tous les chercheurs du Centre Tesnière qui ont participé très activement à cette recherche, RENAHY Julie (tout particulièrement), GREENFIELD Peter, THOMAS Izabella, ANANTALAPOCHAI Raksi, BEDDAR Mohand, DEVITRE Dilber, JIN Gan, KURODA Kyoko, MIKATI Ziad, TORNABONI Morgane, WU Xiaohong, YENPAHT Jutharat, TECHANIYOM Chotika, DZIADKIEWICZ Aleksandra, LAMBERT Mathias et CANNESSON Emma d'Airbus.