
Texte juridique et Traitement Automatique des Langues

BOUHADIBA Farouk
Université d'Oran

Résumé :

L'un des aspects les plus difficiles à traduire d'une langue vers une autre – surtout lorsque les langues en questions sont d'apparemment génétique différent, comme c'est le cas pour l'arabe par rapport à des langues telles que le français ou l'anglais – est à notre avis la traduction de traits culturels qui font la « force propre » d'un peuple. Ceux-ci sont, pour ainsi dire, ancrés dans des faits de langue tels que les proverbes, formes d'adresse, les textes juridiques, etc. La traduction humaine procéderait alors en commençant par la phase de traduction littérale, suivie de la traduction équilibrée pour aboutir à une traduction communicative. Ceci en faisant appel à des stratégies de transposition culturelle et autres. La question se pose à savoir si la machine serait capable d'être programmée pour traduire ces traits. Nous essayerons dans cet exposé de relever certaines difficultés dans la traduction automatique avec des illustrations de fragments de textes juridiques et autres où il est encore précoce de s'annoncer quant à la faisabilité d'une telle entreprise par le biais du traitement automatiques des langues naturelles.

1. Introduction

Les recherches et approches concernant la filiation des langues en vue de la traduction sont nombreuses et variées. Depuis Tolède et même bien avant, la traduction de textes portant sur plusieurs domaines de la connaissance ont suscité de nombreuses divergences. La philologie, et la linguistique comparée se sont attelées à découvrir les similitudes et les différences entre plusieurs langues. Celles-ci se sont souvent penchées sur des langues d'origine Indo-européennes où des comparaisons ont été faites aux niveaux lexical, syntaxique, morphologique, etc. afin de classer et de catégoriser les langues et leur filiation. Ces études se sont rarement intéressées aux ensembles de type phraséologiques (proverbes, dictons), textes juridiques et autres types de discours spécifiques. Nous essayerons d'explorer trois langues (arabe, anglais, français) pour observer le poids du sens à la lumière de différentes cultures et civilisations qui sont révélatrices de diverses façons de penser. De ce fait, la question serait de savoir comment la machine, dans le cadre d'un traitement automatique des langues (TAL), pourrait prendre en charge de telles faits culturels propres à chaque langue et qui sont incorporés dans les textes juridiques, les phraséologies et autres et qui se manifestent à travers le mot.

A cet effet, deux notions de taille pourraient servir de base pour la compréhension des différences de langues :

1. « Linguistic relativism » ou la *relativité linguistique* telle que développée par W. Von Humbolt et plus tard par l'hypothèse de Sapir et Whorf et qui suggère que toute pensée est relative par rapport à la vision du monde que sous-tend le langage (vision du Monde qui diffère d'une langue par rapport

Texte juridique et Traitement Automatique

à une autre et qui pousse les natifs de différentes langues à percevoir le monde différemment et ce en fonction de leur culture ou « force propre » d'un peuple

2. La notion de Chomsky sur les "Linguistic Universals" ou 'universaux du langage (genre, nombre, etc.) et selon laquelle le langage est un phénomène universel.

Ainsi, en mettant en perspective les éléments universels du langage, on pourrait parvenir à une meilleure compréhension du fonctionnement du langage et par voie de conséquence une vision plus claire de la relation langage/pensée, langage/société.

Ces deux visions du langage représentent deux écoles différentes de pensée linguistique (les Relativistes / les Universalistes) et semblent à premier abord contradictoires. En fait, elles convergent en ce sens qu'elles ont comme dénominateur commun le *langage* et la *pensée*. Les Relativistes nous enseignent sur les idiosyncrasies des langues et les différents dispositifs qui les gouvernent alors que les Universalistes nous poussent à chercher les traits et principes communs aux langues. N'est-ce pas là l'objectif primaire du traducteur ? L'étude de la traduction est basée, à notre sens, sur des hypothèses tirées implicitement de ces deux visions linguistiques du langage humain. La nature même de la traduction présuppose l'existence de certains paramètres universels qui, somme toute, nous mènent à rendre toutes les langues traduisibles. Elle reconnaît, par voie de fait, que certains éléments inhérents à chaque langue sont idiosyncratiques (problème de l'intraduisibilité/ intractabilité) et qu'une traduction totale, fiable, juste, parfaite, n'est finalement pas possible et qu'il faudrait avoir recours aux différentes stratégies de la traduction telles que la transposition culturelle –exotisme, calque,

transplantation, emprunt culturel, translittération, la compensation, la traduction par omission (doublets sémantiques en arabe), l'addition, l'équivalence, et autres pour aboutir à ce que l'on appelle communément une représentation fiable de *l'intention* de l'auteur du texte source sous forme d'*intention* du traducteur dans le texte cible.

Selon Steiner (1975 : 149) la logique des relativistes sous-entend qu' aucun acte complet de traduction entre différentes champs sémantiques n'est possible et que toute traduction n'est qu'approximative et ontologiquement réductrice du sens. D'un autre côté, une grammaire universelle affirmera que l'inter-traduisibilité dans les langues naturelles est possible de par la nature même de la Proto-Langue, Mère de toutes les langues. Catford (1965) à son tour met l'accent sur l'élément situationnel et contextuel en parlant d'équivalence en traduction. Il insiste sur le fait que les textes de la Langue Source (SL) et ceux de la Langue Cible (TL) doivent être en rapport étroits avec les traits pertinents et fonctionnels qui sont propres à une situation / contexte afin d'aboutir à l'équivalence en traduction. Eugène Nida (1964) nous rappelle que la reproduction du message de la Langue Source (SL) doit être reflétée par l'équivalent le plus proche de la Langue Cible (TL). Ainsi, c'est la *conservation* du message plutôt que la *conversion* de la forme et structure de l'expression qui devrait être visée dans tout acte ou processus de traduction.

1. La TA

2.1 Quelques repères historiques

Pour l'historique, la traduction automatique a vu naître ses premiers balbutiements dans les années 50 après la deuxième guerre mondiale. L'expérience de l'Université de Georgetown aux USA en 1954 a porté sur la traduction entiè-

Texte juridique et Traitement Automatique

rement automatique de plus de soixante phrases russes en anglais (phrases figées, bien sûr !!!!). L'expérience a été un grand succès et a déclenché une ère de subventions gouvernementales et autres pour la recherche en traduction automatique. Il s'est avéré par la suite que la recherche dans ce domaine était trop lente et qu'il n'y avait pas de progrès réel dans ce domaine en dix années. Le rapport ALPAC de 1966 sur les progrès de la TA aux Etats-Unis a réduit au maximum les subventions et l'euphorie des années cinquantes s'est transformée en désagrément amer et en illusions démesurées. Ce n'est que vers la fin des années 80 avec les progrès en informatique et où l'accès généralisé à l'outil informatique et au software devenait moins cher que l'intérêt pour les modèles statistiques de traduction automatique a repris de l'essor.

En dépit de leurs limitations inhérentes, des programmes de la TA sont actuellement employés par divers organismes à travers le monde. Le plus grand utilisateur institutionnel est probablement la Commission Européenne, qui emploie une version fortement adaptée aux besoins du client du système commercial SYSTRAN pour faire de la traduction automatique d'un grand nombre d'avant-projets de documents à usage interne.

En avril 2003 Microsoft a commencé à employer un système hybride de la TA pour la traduction d'une base de données de documents de soutien technique Anglais-Espagnol. Le système a été développé intérieurement par le groupe de recherche Natural Language de Microsoft. Le groupe examine actuellement un système Anglais-Japonais sur la base de systèmes Anglais-Français et Anglais-Allemands en ligne.

2.2 TA et TAL

En termes généraux, la Traduction Automatique (TA), parfois appelée Traitement Automatique des Langues (TAL)

est une branche de la linguistique l'informatique (Computational Linguistics) qui vise l'utilisation de logiciels d'ordinateur à des fins de traduction de langues naturelles sous forme de textes ou de parole (reconnaissance vocale). A l'origine, la TA se base essentiellement sur la substitution simple de mots atomiques d'une langue naturelle par des mots dans d'autres langues naturelles. En utilisant des techniques de corpus, des traductions plus complexes peuvent être effectuées moyennant une meilleure manipulation et maîtrise des différences typologiques, l'identification d'expressions, et la traduction linguistiques des idiomes, aussi bien que l'isolement des anomalies. Malgré la rapidité dans le processus de traduction, les systèmes actuels ne peuvent aboutir au même rendement avec la même qualité que le traducteur humain. Ceci est particulièrement vrai quand il s'agit de textes à traduire portant sur des domaines spécifiques tels que le texte juridique et ce malgré les dictionnaires spécifiques assez productifs et performants de traducteurs automatiques disponibles sur le marché de la programmation informatique (software). Des logiciels performants de traduction automatique tels que le SYSTRAN, IBM, NASHQ, Power Translator et autres tiennent compte de la spécificité du domaine grâce à leurs dictionnaires par domaine et améliorent ainsi le rendement en limitant la portée des substitutions autorisées. Cette technique est particulièrement efficace dans les domaines où la langue formelle est utilisée.

L'amélioration de la qualité du rendement peut également être réalisée par interposition humaine: par exemple, quelques systèmes peuvent traduire plus exactement à condition que l'utilisateur ait clairement identifié quels mots segments dans le texte sont des Noms propres. Avec l'aide de ces techniques, la TA s'est révélée utile comme outil pour aider les traducteurs humains, et dans certains cas elle peut même produire le rendement escompté. Mais cela est très rare et dé-

Texte juridique et Traitement Automatique

pend essentiellement des systèmes filtres et des systèmes experts du logiciel en question.

Historiquement parlant, la TA représentait un objectif plutôt insaisissable et intangible, voire même fuyant, mais de nos jours un certain nombre de systèmes sont disponibles pour produire des rendements non négligeables et de qualité suffisante pour être utilisés dans un certain nombre de domaines spécifiques (domaine des mathématiques, de la chimie, et de l'architecture).(1)

De façon très générale, le processus de la reconnaissance textuelle en vue de la traduction se fait essentiellement sur la base de deux axes :

1. Décodage de la signification du texte source,
- et
2. Encodage de cette signification dans la langue cible.

Derrière ce procédé simple à première vue, se trouve une opération cognitive très complexe. Par exemple, pour décoder la signification du texte source dans sa totalité, le traducteur doit interpréter et analyser tous les dispositifs du texte, un processus qui exige la connaissance détaillée des deux grammaires, sémantiques, syntaxes, idiomes et autres de la langue source, aussi bien que la culture des deux langues. Le traducteur a besoin de la même connaissance détaillée pour encoder la signification dans la langue cible. C'est là, précisément que réside tout le défi dans la traduction automatique: comment programmer un ordinateur pour "saisir" ces données sous-jacentes d'un texte (tel que le fait le traducteur humain) et "créer" un nouveau texte dans la langue cible avec la même signification. Plusieurs procédés sont actuellement en cours de réalisation et d'amélioration pour aboutir à une traduction automatique plus performante. Nous citerons à titre

d'exemple :

- La Traduction Automatique à base de dictionnaires électroniques (Dictionary-based machine translation). Celle-ci utilise une méthode basée sur des entrées lexicales. Cela signifie que les mots seront traduits comme dans un dictionnaire, mot-à-mot, généralement sans corrélation particulière de la signification entre les mots de la Langue Source et ceux de la Langue Cible

- La Traduction automatique statistique (Statistical Machine Translation) qui vise essentiellement à produire des traductions en utilisant des méthodes statistiques basées sur les corpus bilingues des textes, tels que le corpus canadien de Hansard, le compact Anglais-Français du parlement canadien. Là où de tels corpus sont disponibles, des résultats probants peuvent être réalisés en traduisant des textes spécifiques. Malheureusement, de tels corpus sont très rares.

- Traduction automatique à base d'exemples (Example-based machine translation). Appelée en anglais Example-Based Machine Translation (EBMT), elle est caractérisée par son utilisation d'un corpus bilingue en tant que base de données principale, au moment de l'exécution du programme (moins de systèmes filtres et de systèmes experts). C'est essentiellement une traduction par analogie et elle peut être considérée comme l'application de l'approche du raisonnement au cas par cas.

- Traduction automatique inter langue (Interlingual machine translation) qui est basée sur des règles génératrices. Selon cette approche, la langue source, c'est-à-dire le texte à traduire, est transformée en une représentation qui est indé-

Texte juridique et Traitement Automatique

pendante de la langue source et de langue cible. Le texte de la langue cible est alors généré à partir de cette représentation intermédiaire.

Au Centre Lucien Tesnière de Besançon, nous examinons actuellement dans le cadre du MultiCodict (Multilingual Co-built Dictionary) et d'un Projet CMEP un traducteur Automatique arabe /français/ anglais dans le domaine des mathématiques et des Hydrocarbures et ce sur la base de traducteurs automatiques anglais français, français japonais, français-thai, etc. déjà développés par le Centre en question. Le système que nous essayons de développer est basé sur un composant instruit de génération de langue en utilisant des bases de données de plus de mille phrases de mémoire de traduction.

3. Aperçu sur les spécificités de la segmentation en arabe et la TA

3.1. Comparaison Lexicologique

Bien que l'anglais ait des unités lexicologiques pour les articles, les prépositions comme {à, dans, ...}, pour les pronoms personnels et pour les verbes auxiliaires qui marquent le temps et l'aspect, l'arabe tend à incorporer ces fonctions dans les noms ou les verbes. Pour ce qui est de l'arabe, les inflexions pour les parties du discours ci-dessus sont comptabilisées comme unités lexicologiques indépendantes. A titre d'exemple, l'article indéfini en arabe est représenté par le symbole {Ø}. Il en est de même pour les pronoms personnels implicites comme dans {ana katabtu}. Une autre spécificité du texte arabe est qu'il a un plus grand nombre de noms que le texte anglais et peu de noms en arabe sont modifiés (par d'autres noms ou adjectifs) alors que le texte anglais contient un plus grand nombre d'adjectifs que le texte arabe.

3.2. Comparaison Syntaxique

Le nombre d'unités T (unités traduites) est plus élevé dans le texte anglais. Le texte arabe utilise beaucoup plus la coordination syndétique et asyndétique que le texte anglais qui fait appel à des phrases mixtes et complexes. La coordination étant un dispositif très productif du modèle arabe alors que le système de ponctuation est employé de façon non fonctionnelle. Nous citerons à titre d'exemples des cas tels que celui du *Wāw* qui exprime la simultanéité en arabe là où le français et l'anglais utilisent d'autres constructions grammaticales sans le « et » ou le « and ». Ainsi, une phrase arabe telle que {*qāṭaʿaha wa ṣāḥa*} (MuSTafa LuTfi al-ManfalūTi, 1952 : 236) ne peut se traduire en français par « * Il l'a interrompue et il a crié » ou en anglais par « *He interrupted her and he shouted » mais par « Il l'a interrompue tout en criant » et « He interrupted her, shouting ... » respectivement. Nous remarquons que l'anglais utilise une virgule pour indiquer la simultanéité dans les actions là où le français a recours à la double préposition « tout en ».

4. Spécificités du texte juridique

La loi est une profession où chaque mot est supposé être porteur d'un sens bien particulier, voire même précis. Les dispositifs généraux de la langue légale qui s'appliquent à l'arabe, à l'anglais ou au français sont les suivants :

- Le discours législatif est différent de la langue écrite sur le plan lexical et sur le plan de la construction de phrases qui sont généralement très longues. La différence entre l'arabe et le français ou l'anglais sur ce dispositif est que l'arabe utilise moins de ponctuation (non fonctionnelle en arabe) alors que les deux autres langues usent de la

Texte juridique et Traitement Automatique

punctuation comme mécanisme fonctionnel et producteur de sens dans la langue.

- La profusion de phrases longues et complexes est due non seulement à l'archaïsme du discours juridique mais aussi et surtout dans un but d'éviter toute ambiguïté dans les textes du procès en plaçant le maximum d'information et de données sur une unité phraséologique particulière.
- Un autre dispositif typique relie ensemble des mots ou des expressions par coordination ("et, ou" en français « and, or » en anglais et waw en arabe ayant la signification "et," "ou").
- Une profusion dans l'utilisation de structures phrastiques peu communes ou très rarement utilisées par les natifs des deux langues.
- La loi est toujours exprimée d'une façon impersonnelle afin de s'adresser à plusieurs instances (le législateur, l'exécutif, le mandataire, la sentence, la Cour, le Juge, l'Avocat, etc).
- Le texte juridique évite en général toute allégorie, résonance, allitération, exotisme, synonymie, sens allusif, métaphore et autres aspects du langage humain pour éviter toute ambiguïté dans le texte juridique ou tout équivoque dans la lecture du procès.
- Des sources religieuses telles que la Bible ou le Coran en tant que sources bien fondées, sont généralement présentes dans le texte juridique et qu'il s'agirait d'instruire la

machine en vue d'une traduction automatique.

- Le texte juridique est par définition un texte conservateur (donc non susceptible de changement) du fait que la phraséologie a déjà été testée, approuvée et appliquée.

Sur le plan grammatical, l'une des particularités du texte juridique anglais est qu'il est basé sur des modaux (can, could, may might, will, etc.) qu'il est difficile de traduire en arabe, même pour un traducteur averti. La traduction du texte juridique (de l'arabe vers le français, encore plus de l'arabe vers l'anglais) doit surtout prendre en charge des problèmes étroitement liés à la nature de la langue légale et aux dispositifs spécifiques des systèmes légaux des langues française et arabe ou anglaise et arabe. Ceci apparaît souvent dans des jugements arrêtés aux USA ou en Angleterre (divorce par exemple) à traduire en arabe et inversement.

A prime abord, ces spécificités du texte juridique de chaque langue en question se manifestent dans la constitution même d'un jugement donné. A titre d'exemple, citons le cas des *faits*, des *attendus* et de la *décision*. Les *faits* sont généralement traduits en arabe par {al waqaa ?iε} (الوقائع), les *attendus* par {al ĥayθijaat} (الحيثيات) du fait que la copule arabe {ĥajθu / ĥajθu anna} (vu que, étant donné, etc.) est très productive dans le corps d'un jugement donné (ĥukm) الحكم et la *décision* المنطوق devient {al manṭuuq} parce qu'elle est produite oralement.

En voici quelques exemples :

bayaan waqaa ?iε adda?wa

بيان وقائع الدعوة

Texte juridique et Traitement Automatique

Les faits : bi 9aridha iftitaaHijja mudaa9a ladaa kitaabat DhabT al qism al madani bitariikh....

rafa9a al mudda9ii al mubaashir lil khiSaam biwaasitati wakiilihi al ustadh Bida9wa Dhidda al mudda9i 9alayh al mubaashir lilkhissam bi nafsii mawdho9uha tathbiit Hadgz taHaffudii haada mulaxxassuha...¹

Traduit en français (par nous-mêmes) par : Résumé des faits Suite à une requête déposée auprès du Greffe de la Section Civile en date du sous le N° , le défendeur a déposé par son avocat le professeurune plainte à l'encontre du défendant lui-même dont le sujet est la mise en conservation provisoire dont voici le contenu...

Traduit en Anglais par Systran version 6 avec dictionnaire juridique :

Following a request deposited auprès of Greffe of Section Civile on...>. > under N° , the defendant deposited by its lawyer the professor...>. > a complaint with opposition of the defendant itself whose subject is the setting in provisional conservation of which here contents...

Les attendus : ba9da l mudaawala qaanuunan haythu anna mawdoo9 annizaa9 al maTruuH

Yaduuru Hawla Tathbiit al Hadgs attaHaffudii al waarit 9ala

Traduit en français (par nous-mêmes) par : Après avoir lu l'affaire en délibéré dont le sujet du conflit en présence tourne autour de la conservation provisoire concernant

Traduit en Anglais par Systran version 6 avec dictionnaire juridique :

Afterwards to have read the business in deliberated of which the subject of the involved conflict turn around the provisional conservation concerning.....

La decision: Hakamat al maHkama faSlan fil mawaad al ma-daniyya Hukman 9alaaniyyan HuDhuuriyyan wab tidaa?iyyan

...

1. qabuul adda9wa shaklan
2. al Hukm bi tathbiit al Hadjz attaHaffudhi al waarit 9ala

Traduit en français (par nous-mêmes) par : Le Tribunal a décidé en matière de droit civil et en procès publique et ce en présence des parties concernées et en premier ressort :

1. L'affaire recevable en la forme
2. confirmation des mesures conservatoires provisoire concernant....

Traduit en Anglais par Systran version 6 avec dictionnaire juridique :

The Court decided out of matter of civil law and in lawsuit public and this in the presence of the parts concerned and in the first resort:

1. The admissible business in the form
2. Confirmation of academies measurements concerning....

Ces exemples sont révélateurs, nous semble-t-il, de la difficulté dans la transposition d'un texte chargé de tout un poids culturel, traditionnel et civilisationnel vers un texte n'ayant pas nécessairement les mêmes références socio-culturelles et autres que même l'humain trouve difficile à traduire. Qu'en serait-il alors pour la machine? Ceci n'empêche pas pour autant des tentatives de recherche dans ce domaine pour instruire la machine à procéder à des prototypes de traduction de textes juridiques. Le système TACT (Traduction Automatique Centre Tesnière, France) par exemple est un système de traduction

Texte juridique et Traitement Automatique

automatique français-arabe² de textes juridiques. Appliqué au domaine du droit, ce traducteur automatique arrive à 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

Nous avons été maintes fois surpris de constater que certains systèmes pour passer d'une langue à une autre empruntaient une 3^{ème} 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 faudrait 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 l'on pourrait envisager de travailler dans l'autre sens, ce qui nécessite inévitablement la création 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.

NOTES :

¹ . Voir à ce sujet les travaux de Hutchins, W. J. and Somers, H. (1992) *Introduction to Machine Translation* (London : Academic Press)

² . Nous utiliserons dans ces fragments de texte juridique une transcription générale au lieu de la transcription API (telle qu'utilisée dans nos exemples isolés) pour permettre une lecture plus facile du texte transcrit.

³ . Voir « ALSHARAF H. *et al.* (2004) »

RÉFÉRENCES

ALSHARAF H., CARDEY, S., GREENFIELD, P. (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., GREENFIELD, P. (2001) : « The analysis and design of computerised lexicons », in *Lexis in Contrast* (B. Altenberg & S. Granger eds.), Benjamins, pp. 231-249

Texte juridique et Traitement Automatique

CARDEY, S., GREENFIELD, P. (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

CATFORD, J.C. (1965): *A Linguistic Theory of Translation*, London, Oxford University Press.1965)

CHAN NG, R. (1999) : « Prototype informatisé de dictionnaire du tourisme français-espagnol-

NIDA, Eugene A. (1964): *Towards a Science of Translating*, Leiden, E.J. Brill, 1964

STEINER, George (1975): *After Babel: Aspects of Language and Translation*, London, Oxford University Press.1975