

Traduction Automatique et texte juridique arabe en Algérie

BOUHADIBA Farouk
Université d'Oran

Résumé :

De par ses caractéristiques morpho-phonologiques, morphosyntaxiques, lexicales et autres systèmes et sous-systèmes de son fonctionnement, la langue arabe représente un système de non-concaténation (ou non-enchaînement des morphèmes). Elle diffère dans ce sens des langues Indo-européennes – à systèmes de concaténation – et demeure pour ainsi dire, une langue assez complexe à gérer dans le domaine du Traitement Automatique des Langues (TAL). Ceci, surtout lorsqu'il s'agit de traduire automatiquement des faits de langue porteurs d'éléments culturels propres à cette langue.

L'une des difficultés premières dans l'automation et la numérisation de la langue arabe a été, à premier abord, l'instabilité de ses caractères graphiques, sa segmentation et sa modélisation. Ses diacritiques tels que l'allongement vocalique, l'allongement consonantique (gémination), l'emphase et surtout l'absence de vocalisation (voyellation) de ses consonnes (connue en Grammaire arabe sous le terme de 'harakât' ou 'mouvements consonantiques' et 'sukun' 'absence de mouvement consonantique') représentaient autant de sources de difficultés dans sa programmation.

Cette question de la représentation graphique de la langue arabe sur ordinateur a été soulevée dès les années 70. Certaines de ces difficultés ont trouvé solution puisqu'il est tout à fait possible de nos jours de procéder au traitement automatique de textes arabes, de faire appel à des dictionnaires et vérificateurs électroniques d'orthographe en arabe, voire même d'envoyer des messages en arabe sur Internet et à travers la téléphonie mobile. Ceci est révélateur de grands pas franchis depuis les balbutiements des années 70 face à la complexité, voire même l'impossibilité - disait-on- de procéder au traitement de l'écrit

arabe et plus tard de la reconnaissance de la parole par ordinateur.

Dans cet article, nous utilisons l'expression «texte juridique arabe» au lieu de «texte juridique **en** arabe» pour distinguer les textes juridiques rédigés en arabe selon la juridiction arabomusulmane (voir *infra*) des textes juridiques d'origines diverses (français, anglais, espagnol, etc.) et traduits en langue arabe.

La recherche sur les origines et la filiation des langues à des fins de traduction de textes –anciens, surtout- est assez vaste. La philologie et la linguistique comparée ont relevé des similitudes et des différences entre plusieurs langues. Les langues d'origine Indo-européennes représentaient ainsi un domaine de prédilection où des comparaisons étaient établies à plusieurs niveaux d'analyse (phonétique, phonologique, morphologique, syntaxique, lexical, sémantique, etc.). Le but étant de classifier et de catégoriser les langues selon leurs origines et leurs (degré de) parentés. Les études actuelles dans la fouille de texte (data mining) semblent moins investies dans la traduction par ordinateur d'ensembles phraséologiques tels que les proverbes et les dictons ou bien les textes juridiques arabes.

Nous essayerons d'examiner le texte juridique arabe à travers des prototypes de jugements rédigés en arabe en Algérie. Ceci pour essayer de comprendre comment le Traitement Automatique des Langues Naturelles (TALN) peut-il s'accaparer du sens de ces textes qui sont souvent porteurs de différences culturelles. De ce fait, la question serait de voir si la machine pourrait prendre en charge des faits culturels propres à chaque langue qui sont incorporés dans le texte juridique et qui se manifestent à travers le mot ou l'expression.

A cet effet, deux notions nous serviront de base pour la compréhension de ces différences, à savoir :

a. La relativité linguistique (ou « Linguistic Relativity »)¹ telle que développée chez W. Von Humbolt, reprise dans l'hypothèse de Sapir et Whorf² et selon laquelle la pensée est soumise à la vision du monde à travers le langage – *language shapes thought* –. Cette vision du monde à travers le langage diffère d'une langue à une autre. Les natifs d'une langue donnée perçoivent donc la réalité selon leur culture ou « force propre » que véhicule leur langue.

b. Les Universaux du langage tels que définis chez N. Chomsky (1964, 1965) en termes de « Language Universals »³ pour expliquer que les langues partagent des traits communs tels que le genre et le nombre.

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

Ces deux visions du langage représentent deux tendances: les Relativistes et les Universalistes. Elles semblent, à premier abord, contradictoires. En fait, elles convergent en ce sens

¹ Concept développé au 19^{ème} siècle par des penseurs tels que W. Von Humbolt et selon lequel le langage est l'expression de 'l'esprit d'une nation', de sa vision du Monde et de sa perception de la réalité. Concept repris et développé au 20^{ème} siècle à travers la fameuse Hypothèse de Sapir et Whorf et plus récemment par le courant cognitiviste selon lequel le langage influence des processus cognitifs (Casasanto, D. 2008; Drivoniko et.al. (2007).

² . Edward **Sapir**: Lauenburg, Allemagne, 1884 - New Haven, Connecticut, 1939
Linguiste américain. Il a dégagé la notion de phonème et proposé une nouvelle typologie des langues, fondée sur des critères formels (syntaxe et sémantique) et non plus historiques. C'est l'un des initiateurs du courant structuraliste.

Benjamin Lee **Whorf**: Winthrop, Massachusetts, 1897 - Wethersfield, Connecticut, 1941. Linguiste américain. Disciple d'E. Sapir, il a émis l'hypothèse que le langage est en relation causale avec le système de représentation du monde. Le Petit Larousse, 2009

³ . Dans le courant cognitiviste et la Théorie universaliste, Chomsky (1964, 1965) a développé le concept de « Language universals » pour stimuler la recherche d'une Grammaire Universelle (Universal Grammar).

qu'elles ont comme dénominateur commun le *langage* et la *pensée*. Les Relativistes insistent sur les idiosyncrasies (irrégularités) que l'on retrouve dans les langues et les différents dispositifs qui les gouvernent alors que les Universalistes insistent sur les traits et systèmes communs aux langues. Ceci représente en fait le premier stade de l'acte traductif.

La traduction est basée, à notre sens, sur des hypothèses tirées implicitement de ces deux visions du langage humain. La genèse même de la traduction présuppose l'existence de certains paramètres universels qui rendent toutes les langues traduisibles. Certains éléments inhérents ou propres à chaque langue deviennent idiosyncratiques et posent le problème de l'intraduisibilité / l'intraductibilité. Par conséquent, une traduction totale, juste, fiable, ou parfaite ne semble pas, en fin de parcours, une opération possible et réalisable. Il existe par contre différentes stratégies de la traduction telles que la transposition culturelle pour gérer des cas d'exotisme ou de calque, la transplantation, l'emprunt culturel, la translittération, la compensation, la traduction par omission -comme pour le cas des doublets sémantiques en arabe-, l'addition, l'équivalence et toute une batterie de stratagèmes auxquels pourrait faire appel le traducteur humain -par opposition au traducteur machine⁴- pour aboutir à une représentation adéquate de *l'intention* de l'auteur du texte source sous forme d'*intention* du traducteur du texte cible.

Selon Steiner, G. (1975:149) la logique des relativistes sous-entend qu'aucun acte complet de traduction entre différents champs sémantiques n'est possible et que toute traduction n'est qu'approximative et ontologiquement réductrice du sens. D'autre part, une grammaire universelle postulera que l'inter-

⁴ . Le terme 'traducteur machine' est utilisé pour désigner l'ensemble des opérations à des fins de traduction automatique et dictionnaire tels que les algorithmes, les systèmes filtres, les systèmes experts, la segmentation, la modélisation, etc.

traduisibilité dans les langues naturelles est possible de par la nature même de la Proto-Langue (Proto-language), source ou Mère de toutes les langues.

Catford, J.C. (1965) met l'accent sur l'aspect situationnel et contextuel en présentant le concept d'équivalence en traduction. Par conséquent, le texte de la Langue Source (SL)⁵ et celui de la Langue Cible (TL) doivent être en rapport étroit avec les traits pertinents et fonctionnels d'une situation (contexte) donnée afin d'aboutir à l'équivalence en traduction. Nida, E. (1964) explique 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 de la structure du texte qui devrait être visée dans l'acte traductif.

1. La Traduction Automatique (TA)

Ce n'est qu'après la Seconde Guerre Mondiale que la TA s'est développée de manière observable. L'Université de Georgetown (USA) a réalisé une traduction entièrement automatique en 1954. Le projet portait sur une soixantaine de phrases figées en russe et traduites en anglais. Cette première expérience a déclenché une recherche accélérée grâce à des fonds de soutien pour la recherche en TA. Malgré de lourds investissements financiers, le progrès dans ce domaine restait limité. Le rapport de 1966 du Comité Consultatif sur le Traitement Automatique du Langage, ALPAC (Automatic Language Processing Advisory Committee) a été plus que négatif quant aux progrès de la TA aux Etats-Unis. Ceci a réduit au maximum les subventions allouées dans ce domaine. L'euphorie des années cinquantes s'est alors transformée en désagrément amer et en de profondes illusions (from euphoria to bitter criticism).

⁵ . SL renvoie à Source Language (en anglais) par opposition au TL (Target Language) ou Langue Cible (en français).

Ce n'est que bien plus tard que les développements de l'informatique, des statistiques et de la linguistique computationnelle des années 80 ont permis la conception de plusieurs modèles de traduction automatique mis à la disposition du grand public sur le marché de l'informatique. Ces programmes et outils informatiques à des fins de TA présentaient des défaillances. Ils ont été raffinés et rendus plus performants au fil des ans et ils sont accessibles à divers utilisateurs et organismes tels que la CEE qui utilise une version adaptée (users-need) du système commercial SYSTRAN pour traduire un grand nombre de documents à usage interne. Microsoft a développé à son tour un système hybride de la TA dans l'élaboration d'une base de données de soutien technique Anglais-Espagnol. Le système a été conçu par le groupe de recherche «Natural Language» de Microsoft (2003). Ceci a été suivi par l'élaboration d'un système de traduction automatique Anglais-Japonais sur la base de systèmes Anglais-Français et Anglais-Allemands développés auparavant⁶. On retrouve actuellement ces logiciels sur le marché de l'informatique sous forme de produits finis ou bien sous forme de prototypes de traducteurs et de dictionnaires mis en ligne au service du grand public.

2. La TA, le TAL, le TALN et le TALA

La Traduction Automatique (TA) est un terme englobant qui réfère à l'automation de la langue en général. Elle se ramifie en Traitement Automatique des Langues (TAL) et ce suite au développement des technologies des langues, puis en Traitement Automatique des Langues Naturelles (TALN) pour distinguer ces dernières des langues dites artificielles. TALA – une ramification récente du TAL/TALN- réfère au Traitement Automatique de la Langue Arabe. La TA, le TAL, le TALN et le TALA dérivent de la linguistique l'informatique (Computational

⁶ .cf. Ghenimi, A. (2007 : 18)

Linguistics) à des fins de conception et d'élaboration de logiciels de traduction de textes et de reconnaissance vocale.

A l'origine, la TA se basait essentiellement sur la substitution de mots atomiques d'une langue par des mots ou cognats (équivalents) dans d'autres langues. En utilisant des techniques de fouille de textes grâce à la Linguistique de corpus (Corpus Linguistics) et aux Statistiques entre autres, des traductions plus complexes pouvaient être effectuées moyennant une meilleure manipulation et une bonne maîtrise des différences typologiques entre les langues naturelles, l'identification d'expressions et d'idiomes communs à deux ou plusieurs langues et en procédant à l'isolement des différences entre les langues. Cependant et malgré la rapidité de la traduction machine qui est le résultat de l'implication de compétences transversales et multidisciplinaires en linguistique, en informatique, en mathématiques et en statistiques, entre autres, les systèmes actuels ne peuvent offrir le même rendement avec la même qualité que le traducteur humain. Ceci est d'autant plus évident lorsqu'il s'agit de traduire des documents spécifiques tels que le texte juridique. Il existe, certes, sur le marché de la programmation informatique (software) des dictionnaires spécialisés performants et des traducteurs automatiques très itératifs qui sont d'une aide précieuse pour le traducteur humain. Tout comme il existe des programmes de rédaction technique dans des domaines sensibles tels que celui de la santé et de la sécurité (publique, routière, industrielle, ferroviaire, aérienne, etc.) par exemple. Des logiciels 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 ils améliorent ainsi le rendement du traducteur humain en limitant au maximum le facteur temps nécessaire à l'opération de traduction. Des techniques nouvelles et particulièrement efficaces font appel aux décisions (sélections, instructions) de l'utilisateur. Le rendement dans la

traduction peut ainsi être amélioré qualitativement et quantitativement lorsque l'utilisateur agit sur le système (le logiciel) en identifiant les mots segments dans le texte qui ont la fonction de nom propres par exemple. Avec l'aide de ces techniques, la TA s'est révélée utile comme outil pour aider les traducteurs humains. Elle peut même, dans certains cas, produire le rendement escompté. Mais cela est très rare et dépend essentiellement des systèmes filtres et des systèmes experts du logiciel en question.

La Traduction Automatique, parfois dénommée TAO (Traduction Assistée par Ordinateur), qui représentait initialement un objectif plutôt difficile à atteindre, offre de nos jours des programmes accessibles à tout un chacun et elle aboutit à des résultats de meilleure qualité qu'auparavant. Des traducteurs automatiques et dictionnaires électroniques (sous forme de produits finis ou en ligne) sont disponibles dans des domaines de spécialités aussi variés que celui de la santé, de la sécurité, du droit juridique, de la chimie, de l'architecture, etc.⁷

De façon très générale, le processus de la reconnaissance textuelle en vue de la traduction se fait essentiellement sur la base d'un décodage de la signification du texte source et d'un encodage de cette signification dans la langue cible à travers une batterie d'algorithmes et de paramètres morphosyntaxiques et mathématiques. Ce procédé simple en apparence suscite des opérations cognitives très complexes. Ainsi, pour décoder la signification du texte source dans sa totalité, le traducteur doit interpréter et analyser tous les dispositifs du texte. Ceci exige une connaissance détaillée de la syntaxe des langues à traduire et de leur sémantique, entre autres, ainsi que la culture des langues en question avant de procéder au décodage et à

⁷. Cf. à ce sujet, Hutchins, W. J. & Somers, H. (1992): Introduction to Machine Translation (London : Academic Press).

l'encodage. C'est là, précisément que réside tout le défi dans la traduction automatique. Comment pourrait-on 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.

Il existe plusieurs systèmes pour aboutir à une automatisation des langues tels que la Traduction automatique à base de dictionnaires électroniques ou Dictionary-Based Machine Translation (DBMT). Celle-ci utilise des bases d'entrées lexicales. Les mots sont traduits comme dans un dictionnaire au format papier. Ils sont présentés sous forme de mot-à-mot et sont généralement démunis de corrélation particulière quant à la signification des mots de la Langue Source par rapport à ceux de la Langue Cible.

La Traduction automatique par Statistiques ou Statistical Machine Translation (SMT) est un autre procédé d'automatisation des langues. Son but est de produire des traductions en utilisant des méthodes statistiques basées sur des corpus bilingues de textes. Il s'agirait dans ce cas d'élaborer des corpus de textes spécifiques bilingues pour aboutir à des résultats probants.

La Traduction automatique à base d'exemples ou Example-Based Machine Translation (EBMT) se caractérise par l'utilisation d'un corpus bilingue en tant que base de données principale. C'est essentiellement une traduction par analogie et elle peut être considérée comme l'application de l'approche du raisonnement au cas par cas.

La Traduction automatique inter-langue ou Interlingual Machine Translation fonctionne à base de règles génératrices. Ainsi, le texte à traduire est transformé en une représentation qui est indépendante de la langue source et de langue cible. Le texte de la langue cible est alors généré à partir de cette représentation intermédiaire.

Il existe également des prototypes de traducteurs développés dans des universités et centres de recherche en TAL. Le Centre

Lucien Tesnière de Besançon, par exemple, a développé des dictionnaires et traducteurs automatiques tels que le Multilingual Co-built Dictionary (MultiCodict) ou le Labelgram.

Le système que nous essayons de développer actuellement est basé sur un composant instruit de génération de langue en utilisant des bases de données avec une mémoire de traduction d'un millier de phrases et ce dans les domaines des Hydrocarbures et du Tourisme.⁸

3. Quelques aspects de la segmentation en arabe

Nous essayerons dans ce qui suit de mentionner brièvement des phases que nous jugeons nécessaires à la segmentation et à la modélisation de la langue arabe et ce par rapport aux caractéristiques et fonctionnements d'autres langues. La comparaison lexicologique est l'une des bases qui permettent de mieux saisir les différences et similitudes de la langue arabe par rapport à d'autres langues.

A titre d'exemple, l'anglais possède des unités lexicologiques pour les articles {a, an, the} ou bien l'ensemble vide { $\emptyset$ } pour l'indéfini pluriel, des prépositions {in, on, out, and, about, ...}, des pronoms non affixés {I, You, Him, Her, ...} et des auxiliaires et modaux {Be, Have, Can, May, ...} comme marqueurs du temps et de l'aspect. L'arabe génère certaines de ces fonctions soit en liant les mots et morphèmes affixés (préfixés, infixés, suffixés) et en les incorporant à d'autres formes telles que les noms et les verbes, soit en séparant ces derniers. Ceci étant gouverné par des règles morphosyntaxiques qui opèrent dans la langue. Nous optons pour une segmentation où les inflexions par exemple sont comptabilisées comme des unités lexicologiques indépendantes. Ainsi, l'article indéfini en arabe est représenté par l'ensemble vide { $\emptyset$ }. Une forme arabe isolée

⁸. cf. BOUHADIBA, F. & ROUABHI, M., RML6, 2008 : 111-141

telle que وَلَدٌ /walad(un)/⁹ (vocalisée) que l'on retrouve dans un texte non vocalisé sous la forme graphique ولد et qui ne serait reconnue comme telle, c'est-à-dire 'un garçon' que si l'on insérerait dans la chaîne la structure {Ø + وَلَد} afin de ne pas confondre cette dernière avec d'autres formes isolées telles que le passif وَلِدَ 'il est / fut né' ou le causatif وَلَدَ 'générer' par exemple.

Il en est de même pour les pronoms personnels implicites comme dans (أَنَا) كَتَبْتُ / (anâ) katabt(u)/ 'J'ai écrit' que nous comptabilisons comme deux unités lexicales indépendantes même si le pronom personnel أَنَا n'apparaît pas dans le texte mais il est sous entendu comme dans la phrase كَتَبْتُ الْقِصَّةَ /katabtu l qiqṣṣa(ta) / 'J'ai écrit l'histoire (le conte)',¹⁰.

Une autre caractéristique de l'arabe par rapport l'anglais c'est la profusion de noms pour désigner un concept alors que l'anglais contient un plus grand nombre d'adjectifs que l'arabe. La comparaison syntaxique permet également de mieux saisir les différences et similitudes entre l'arabe et d'autres langues. Le nombre d'unités syntaxiques (grammaticales) est plus élevé en anglais. L'arabe a tendance à utiliser des coordinations syndétiques et asyndétiques là où l'anglais fait appel à des phrases mixtes et complexes avec divers procédés de coordination.

La coordination en arabe représente un dispositif très productif 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

⁹ . Les parenthèses à l'intérieur de la représentation phonologique indiquent la forme pausale en arabe et l'on entend [walad].

¹⁰ . Nous ne pouvons, dans le cadre de cet article, détailler les procédures qui nous permettent de comptabiliser des unités lexicologiques indépendantes. Ceci fera l'objet d'un prochain article dans ce sens.

français et l'anglais utilisent d'autres constructions grammaticales sans le «et» ou le «and». Ainsi, une phrase arabe telle que {qâṭaʿaha wa ṣâḥa} (Mustafa Lutfi Al-Manfaluti, 1952: 236) ne se traduit pas 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. Quelques aspects 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 résumés comme suit :

- 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 ponctuation 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 ensembles des mots ou des expressions par coordination : {et, ou, ...} en français {and, or, ...} en anglais et {waw} en arabe.
- 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 ou toute équivoque dans la lecture du procès.
- Des sources religieuses en tant que sources bien fondées, sont généralement présentes dans le texte juridique. Il s'agirait d'instruire la machine en vue d'une traduction automatique dans ce sens.
- 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 dans la juridiction. Sur le plan grammatical, l'une des particularités du texte juridique anglais est qu'il est basé sur des modaux tels que {can, could, may might, will, ...} et dont les nuances en anglais sont parfois difficiles à traduire en arabe même pour un traducteur averti. La traduction du texte juridique de l'arabe vers le français (ou vice-versa) et encore plus de l'arabe vers l'anglais (ou vice-versa) 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 en France, aux Etats Unis ou en Grande Bretagne (divorces 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 الوقائع ‘al waqâʔiʔ’¹¹, les **attendus** par الحثيات ‘al ḥajθijât’ du fait que la copule arabe حيث {ḥajθu} et l’expression حيث أن {ḥajθu anna} (vu que, étant donné, etc.) sont très productives dans le corps d’un jugement donné حكم ‘ḥukm’ et la **décision** devient المنطوق ‘al mantûq’ parce qu’elle est produite oralement.

En voici quelques exemples :

بيان وقائع الدّعوة

الوقائع:

بعريضة افتتاحية مذاعة لدى كتابة ضبط القسم المدني بتاريخ
رفع المدّعي المباشر للخصام بوسيلة وكيله الأستاذ
بدعوة ضدّ المدّعي عليه المباشر للخصام بنفسه موضوعها تثبيت
حجز تحفظي هذا ملخصها.....

Que nous transcrivons ci-dessous :

bajân waqâʔiʔ addaʔwa

Les faits : bi ʔaridâ iftitâʔijja mudâʔa ladâ kitâbat dâbt al qism al madani bitarîḫ rafaʔa al muddaʔi al mubâʔir lil ḫiṣâm biwasîṭati wakîlihi al ustað bidaʔwa dâdda al muddaʔi ʔalaih al mubâʔir lilḫiṣâm bi nafsîhi mawðuʔuha taθbît fiagz tafiaffuðî hâða mulaḫḫaṣuha...

En donnant le texte arabe ci-dessous, ce dernier est traduit en français par des outils de traduction sur le Net de la façon suivante :

¹¹ . Nous utilisons les symboles phonétiques de l’API (IPA) dans nos exemples.

الوقائع :
 بعريضة افتتاحية مذاعة لدى كتابة ضبط القسم المدني بتاريخ
 رفع المدّعي المباشر للخصام بوسيلة وكيله الأستاذ
 بدعوة ضدّ المدّعي عليه المباشر للخصام بنفسه موضوعها تثبيت
 حيز تحفظي هذا ملخصها.....

Exposé des faits de l'appel

Faits:

Pétition écrit une section de contrôle éditorial diffusé sur ci-vile..... Le demandeur Argument direct à l'argument et son agent, M. Contre le défendeur invité à diriger son propre thème querelle de fixation de cette synthèse installés

que nous traduisons 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 Professeur une plainte à l'encontre du défendant lui-même dont le sujet est la mise en conservation provisoire dont voici le contenu ...

En traduisant ce texte arabe en Anglais, ces outils nous produisent :

Statement of the facts of the call

Facts:

Petition in writing an editorial broadcast control section on civil The plaintiff..... Direct argument for argument and his agent, Mr. Against the defendant invited to direct his own quarrel theme installed attachment summary of this.....

Systran (version 6 avec la sélection du dictionnaire juridique) nous produit la 'traduction' suivante:

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: baʿda l mudâwala qânûnan haythu anna mawduʿ annizâʿ al maṭrûḥ jadûru ḥawla taḥbît al ḥadẓz attaḥaffðî l wâriḥ ʿala

الحثيات:

بعد المداولة قانونا حيث أن موضوع النزاع المطروح يدور حول تثبيت الحجز التحفظي الوارث على.....

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

Cet extrait de texte juridique arabe est traduit en français par des outils de traduction sur le Net comme suit :

Attendus

Ensuite après avoir lu les affaires dessus de délibéré au sujet des dont le sujet du conflit en présence tourne autour de la conservation temporaire.....

Systran (version 6 avec dictionnaire juridique) le traduit en anglais comme suit:

Waited

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

La decision: ḥakamat al maḥkama faṣlan fil mawâd al madaniyya ḥukman ʿalâniyyan ḥuḍuriyyan wab tidâʿiyyan ...

1. qabûl addaʿwa ṣaklan
2. al ḥukmu bi taḥbît al ḥadẓz attaḥaffuḍi al wâriḥ ʿal....

المنطوق:

حكمت المحكمة فصلا في المواد المدنية حكما علانيا حضوريا وابتدائيا

1. قبول الدّعوة شكلا
2. ألحكم بتثبيت الحجز التحفظي الوارث على

Traduit en français (par nous-mêmes) :

Le Tribunal a décidé en matière de droit civil et en procès public et ce en présence des parties concernées et en premier ressort :

1. L'affaire recevable en la forme
2. Confirmation des mesures conservatoires provisoires concernant....

Cet extrait de texte juridique arabe est traduit en français par des outils de traduction sur le Net comme suit :

Parlé / Oral

La cour décidée en ce qui concerne la loi civile et dans le public de procès et ceci en présence des pièces concernées et dans la première ressource:

1. Les affaires admissibles sous la forme
2. Confirmation de C des mesures d'académies S temporaire au sujet de....

Traduit en Anglais par Systran version 6 avec dictionnaire juridique :

Decision

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, non seulement des irrégularités dans les correspondances plus spécialement aux niveaux syntaxique et lexical entre les langues en question mais aussi 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 socioculturelles et autres que même l'humain trouve parfois difficiles à traduire. Qu'en serait-il alors pour la machine? Nous remarquons, à titre d'exemple, que dans les traductions de l'arabe vers le français ou l'anglais, l'article défini de l'arabe

est parfois omis et ce surtout lorsqu'il fait partie d'un mot isolé tel que الوقائع traduit en français par **Faits** et en Anglais par **Facts** (voir supra les exemples de الحثيات et de المنطوق). 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 automatique français-arabe¹² 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

النظام الإداري يحكم في النزاعات المرتبطة بالتنظيم و آلية الخدمات العامة
وبالعقود الإداري

Nous avons été maintes fois surpris de constater que pour passer d'une langue à une autre certains systèmes empruntaient une 3^{ème} langue qui est souvent l'anglais; ce qui fait que les résultats ne peuvent pas être nécessairement en adéquation entre SL et TL en passant par une langue intermédiaire (l'anglais). Par exemple, pour le mot « libre » en français ou « free » en anglais, nous avons trouvé dix neuf entrées lexicales dans la fouille de textes juridiques arabes. Ceci en plus du sens commun de حر. Il s'agirait donc de répertorier ces différentes entrées lexicales arabe, de situer ou de localiser les contextes dans lesquels elles sont susceptibles d'apparaître et d'instruire la machine dans ce sens. Tâche tout à fait réalisable par une recherche lexicographique et sémantique dans ce domaine.

¹² . Cf. «ALSHARAF, H. *et al.* (2004)»

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.

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 TAL (Traitement Automatique des Langues)», *Mémoire de Maîtrise d'anglais* mention Industries de la Langue, Besançon, 1999.

BOUHADIBA, F. & ROUABHI, M. (2008) : A propos d'un Dictionnaire Electronique Arabe /Anglais/ Français dans le Domaine des Hydrocarbures, *Revue Maghrébine des Langues*, pp. 111-141. Université d'Oran, Algérie.

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

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

CASASANTO, DANIEL (2008), "Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought", *Language Learning* **58** (1): 79

COOK, VIVIAN & BENEDETTA BASSETTI (2010), *Language and Bilingual Cognition*, Hove: Psychology Press

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

CHOMSKY, N. (1964) : *Current Issues in Linguistic Theory*, The Hague : Mouton

CHOMSKY, N. (1965) : *Aspects of the Theory of Syntax*, Cambridge, Mass. M.I.T. Press.

CHOMSKY, N. (1968) : *Language and Mind*, New York : Harcourt Brace.

CHOMSKY, N. (1971) : *Selected Writings*, edited by Allen, J.P.B. and Van Buren, P. London, Oxford University Press.

DRIVONIKOU, G. V., P. KAY, T. REGIER, R. B. IVRY, A. L. GILBERT, A. FRANKLIN & I. R. L. DAVIES (2007): "Further evidence that Whorfian effects are stronger in the right visual field than the left", *Proceedings of the National Academy of the Sciences USA*. **104** (3), pp. 1097–1102

FISHMAN, J.A. (1978), "Positive bilingualism: Some overlooked rationales and forefathers", in J. E. Alatis (Ed.), *International dimensions of bilingual education*, Washington, D.C.: Georgetown University Press, pp. 42–52.

FISHMAN, J.A. (1982), "Whorfianism of the third kind: Ethnolinguistic diversity as a worldwide societal asset", *Language in Society* 11, pp. 1–22

FRANK, MICHAEL C.; EVERETT, DANIEL L.; FEDORENKO, EVELINA; GIBSON, EDWARD (2008), "Number as a Cognitive Technology: evidence from Pirahã language and cognition" *Cognition* 108 (3), pp. 819–824.

GHENIMI, A. (2007): *Conception d'un modèle de traduction Arabe-Français appliqué au domaine des Mathématiques*, Thèse de Doctorat, Centre Lucien Tesnière, Université de Franche Comté, Besançon, France.

GUMPERZ, JOHN, & STEPHEN LEVINSON EDS. (1996), *Rethinking Linguistic Relativity*, Cambridge: Cambridge University Press

HUTCHINS, W. J. & SOMERS, H. (1992): *Introduction to Machine Translation*. London: Academic Press.

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

SAPIR, E. (1983), David G. Mandelbaum, ed., *Selected Writings of Edward Sapir in Language, Culture, and Personality*, University of California Press

STEINER, George (1975): *After Babel: Aspects of Language and Translation*, London, OUP

WHORF, B. (1956), John B. Carroll (ed.), ed., *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, MIT Press