

Représentation lexicale et TAO

E. Wehrli

Département de linguistique - LATL
Université de Genève

1 Introduction

L'établissement de lexiques, le choix des informations associées aux entrées lexicales, la structuration de la base de données, *etc.* figurent parmi les problèmes centraux en linguistique informatique. Souvent délaissée au profit de tâches plus "nobles" et moins fastidieuses, telles que le développement de nouveaux formalismes de représentation linguistique ou encore la spécification d'algorithmes d'analyse ou de génération syntaxique, la composante lexicale des systèmes de traitement du langage constitue pourtant la plate-forme sur laquelle s'articulent toutes les autres composantes.

Dans cet article nous présenterons la structure et la fonction de la composante lexicale du système interactif de traduction ITS-2, en cours de développement dans notre laboratoire. Après un rapide survol du système global, qui illustrera le rôle capital de l'information lexicale pour les trois composantes dynamiques du système de traduction (analyse, transfert, génération), nous aborderons le problème du choix de l'information lexicale, en particulier dans la perspective de traductions interactives pour monolingues.

2 Le système ITS-2

Le système ITS-2 (en anglais *Interactive Translation System*) est un prototype de système interactif de traduction de phrases en développement au Laboratoire d'analyse et de technologie du langage (LATL) de l'Université de Genève. Dans son état actuel, il accepte des phrases

prises dans un sous-ensemble de la grammaire du français pour les traduire en anglais. Le projet ITS-2 poursuit plusieurs objectifs, aussi bien théoriques que pratiques. Il s'agit de montrer, d'une part l'intérêt d'une approche interactive au problème de la TA, d'autre part les avantages d'un système basé sur le modèle grammatical chomskyen du Gouvernement et Liage (en anglais *Government and Binding Theory*, dorénavant GB). Pris conjointement, ces deux objectifs distinguent le projet ITS-2 de pratiquement tous les systèmes de traduction actuels ou en développement, puisque ceux-ci sont généralement conçus comme des systèmes autonomes (non-interactifs) et, à de très rares exceptions près, ne s'inspirent guère des travaux récents de l'école chomskyenne.

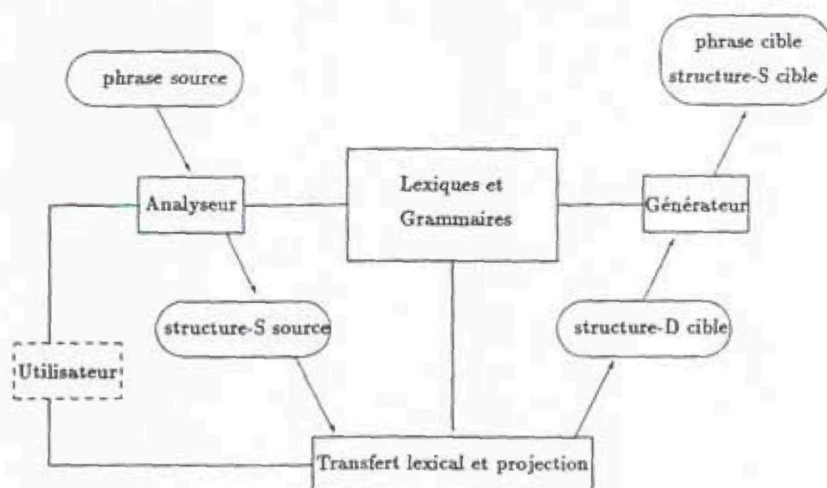


Figure 1: Architecture du système ITS-2

Le système ITS-2, dont l'organisation est donnée dans la figure 1, repose sur le modèle du transfert. On retrouve par conséquent les trois composantes fondamentales des systèmes basés sur le transfert, soit l'analyse, le transfert et la génération. En simplifiant quelque peu, on peut représenter le fonctionnement du système comme suit : une phrase en français (phrase source) est prise en charge par l'ana-

lyseur qui lui assigne une représentation syntaxique correspondant à une structure de surface enrichie de la grammaire GB (structure-S source)¹. Pour accomplir cette tâche, l'analyseur a besoin d'informations lexicales et grammaticales (lexiques et grammaires). Si nécessaire, il peut également demander des informations supplémentaires à l'utilisateur au moyen de dialogues de clarification (utilisateur).

La composante de transfert est chargée d'élaborer une structure syntaxique abstraite en langue cible (structure-D cible) sur la base de la structure-S source construite par l'analyseur, ou plus précisément sur la base de la structure-D contenue dans la structure-S². Le transfert consiste donc à mettre en correspondance à un certain niveau d'abstraction une structure source et une structure cible. Dans le système ITS-2, le niveau de représentation utilisé pour le transfert s'apparente à celui de la structure profonde (structure-D) de la grammaire GB. Comme on le verra plus en détail un peu plus bas, le transfert est dans une large mesure une question de correspondances lexicales entre items lexicaux de la langue source et items lexicaux de la langue cible, ainsi que d'un mécanisme de projection de structures sur la base d'informations lexicales.

La génération comporte deux sous-composantes, que nous appellerons respectivement la composante transformationnelle et la génération morphologique. La première a pour tâche la dérivation syntaxique de structures de surfaces (structures-S) à partir des structures-D par l'application de règles transformationnelles. Enfin, la génération morphologique est responsable du choix des représentations morphologiques des items lexicaux, ainsi que des éventuelles règles d'adaptation (élision, contraction, etc).

¹La théorie GB distingue deux niveaux de représentation syntaxiques, appelés respectivement **structure-D** et **structure-S**. Comme les structures profondes des versions antérieures de la théorie chomskyenne, les structures-D représentent, sous forme de structures canoniques, les relations entre prédicats et arguments. Plus généralement, ce niveau de représentation correspond à une projection de l'information lexicale (catégorie, sous-catégorisation, etc.) dans une structure \bar{X} . La structure-S représente l'organisation syntagmatique d'une phrase en tenant compte de l'ordre de surface des mots, ainsi que les traces des éléments déplacés par transformation.

²En effet, selon la théorie des traces, tout déplacement de constituant laisse derrière lui un constituant abstrait (*vide*), appelé trace, coïncidé avec le constituant déplacé. Il s'ensuit que la structure-S conserve la position d'origine de tous les constituants déplacés.

Ces différents processus sont repris en détail dans les sections suivantes. Mais avant d'aborder ces questions, il convient de mentionner deux principes fondamentaux sur lesquels repose ce système. Le premier, que nous appellerons principe de correspondance stipule que, pour des langues pas trop distantes dans leurs systèmes linguistiques et culturels, il existe pour chacune des phrases de la langue source une phrase cible qui lui soit équivalente du point de vue du sens. Le deuxième principe est celui de la préservation de la structure, en vertu duquel on cherche dans la mesure du possible à conserver pour la phrase cible le style et la forme syntaxique de la phrase source. Il est clair, cependant, que ce principe n'est pas absolu et que dans certaines circonstances on est obligé de s'éloigner quelque peu de la syntaxe de la phrase source pour choisir une autre construction; bref, à procéder à ce que les traducteurs appellent des transpositions.

2.1 L'analyse

Le rôle de l'analyseur syntaxique dans un système de traduction est capital. C'est à lui, en effet, qu'incombe la responsabilité de découper la phrase en syntagmes, d'établir le statut catégoriel des mots et des syntagmes, ainsi qu'à déterminer la fonction syntaxique des constituants (sujet, objet direct, etc).

Dans le contexte de la traduction, il est important de souligner le travail de désambiguïsation effectué par l'analyseur. En effet, une phrase consiste en une séquence de mots, et sa traduction passe, en partie au moins, par le choix de la traduction des mots qui la composent. Or ces derniers, pris hors contexte, sont pratiquement tous susceptibles d'interprétations multiples, conduisant la plupart du temps à des traductions fort différentes. Il est par conséquent fondamental de chercher à préciser autant que possible le sens d'un mot dans la phrase source. L'analyseur syntaxique se révèle être un filtre très efficace grâce aux nombreuses contraintes qu'il impose. Ainsi, de toutes les combinaisons potentielles des sens des mots d'une phrase, seules subsisteront celles qui satisfont le mieux aux contraintes syntaxiques imposées par l'analyseur.

Pour le système ITS-2, nous utilisons l'analyseur syntaxique

FIPS, dont on trouve une description détaillée dans Laenzlinger et Wehrli (1992).

2.2 Le transfert

C'est la composante de transfert qui est chargée d'établir la correspondance entre les structures sources et les structures cibles. Ainsi que nous l'avons déjà mentionné, le niveau de représentation utilisé pour décrire les structures syntaxiques source et cible correspond pour l'essentiel à une structure profonde reflétant les propriétés argumentales des éléments prédicatifs de la phrase cible (structures-D). Ainsi, la structure obtenue par transfert des phrases sources (1a, b) est-elle dans les deux cas celle décrite en (2).

- (1)a. le chat mange la souris.
 b. la souris est mangée par le chat.
- (2) [_S [_{NP} the cat] [_{VP} eat [_{NP} the mouse]]]

En effet, la structure d'arguments des deux phrases sources est la même, avec *manger* comme prédicat, *le chat* comme agent et *la souris* comme thème. Il est donc naturel que le transfert donne lieu à une même structure-D cible. Cependant, même si le transfert d'une phrase active ou de sa structure correspondante passive aboutit à la même structure-D cible, il n'est pas souhaitable que celle-ci donne lieu à la même phrase dans les deux cas. Du point de vue de la structure du discours, entre autres, l'opposition actif-passif peut être pertinente et devrait être maintenue par le processus de traduction. Cela correspond d'ailleurs au principe de la préservation de la structure que nous avons adopté. Pour ce faire, le mécanisme de transfert, outre la structure-D, détermine un ensemble d'opérations syntaxiques auxquelles sera soumise la structure dans la phase de génération. Ainsi, en ce qui concerne notre exemple, dans le cas de la traduction de la phrase (1b), on associera à la structure-D le trait [+passif], qui déclenche l'application de la transformation syntaxique du passif, pour aboutir à la phrase

(3) the mouse is eaten by the cat.

Comme on l'a vu, le transfert se fait à partir des structures-S produites par l'analyseur et engendre des structures-D. La génération consiste à transformer ces structures-D en structures de surface par l'application de règles transformationnelles, puis par l'application de règles morphologiques.

Bien que conceptuellement distinctes, les opérations de transfert et de génération sont largement imbriquées les unes dans les autres dans le système ITS-2, de façon à minimiser les parcours de structure, ainsi que la masse d'information à stocker.

La stratégie utilisée pour les opérations de transfert et de génération est de type descendant récursif. A partir du syntagme global de catégorie S (phrase) ou NP (syntagme nominal), on cherche l'élément clé (ou tête sémantique) du syntagme, respectivement le verbe principal de la proposition ou la tête nominale du syntagme nominal. Le transfert lexical s'effectue au niveau de cet élément clé. Les inévitables ambiguïtés lexicales sont partiellement levées sur la base de la structure d'arguments associée à la tête sémantique. Le parcours de structure qui mène à la tête du constituant calcule un certain nombre de paramètres morpho-syntaxiques (nombre, genre, détermination, etc.). Une fois achevé le transfert de la tête du syntagme, le même processus s'applique récursivement à ses éventuels compléments, ajouts et spécificateurs.

Le transfert se fait essentiellement au niveau des éléments lexicaux abstraits (lexèmes) sur la base de l'information contenue dans le dictionnaire bilingue. Ce dernier met en correspondance les items lexicaux sources et cibles. Cette correspondance se fait non pas au niveau des mots de la langue (trop ambigus), mais bien au niveau plus abstrait des lexèmes, c'est-à-dire des lectures particulières d'un item lexical. Nous reviendrons sur ces notions un peu plus bas.

Pour donner un exemple concret, dans le lexique français, le verbe *demande* est associé à deux lexèmes distincts (parmi d'autres), un pour la lecture pronominale, un autre pour la lecture non-pronominale. Dans le dictionnaire bilingue, le premier de ces deux lexèmes est mis en correspondance avec le lexème anglais *to wonder*, le second avec

to ask. Dans la mesure où l'analyse syntaxique permet, sur la base du contexte syntaxique, de déterminer avec précision la lecture particulière d'un item lexical, cette façon de procéder élimine un très grand nombre d'ambiguïtés de transfert potentielles.

2.3 Génération

Le processus de génération est chargé de la dérivation d'une structure bien-formée à partir des structures sous-jacentes produites par le transfert. Fortement couplée avec le transfert, la génération s'effectue en accord avec le principe du cycle. Ce principe, qui détermine le mode d'application des règles transformationnelles dans des phrases complexes, impose une application itérative des règles à l'intérieur de chacune des propositions S d'une phrase complexe, en commençant par la plus enchâssée pour finir par la proposition principale.

Selon l'hypothèse de la préservation de la structure, discutée au début de la section 3, les caractéristiques syntaxiques de la structure cible sont largement déterminées par celles de la phrase source. Il existe, cependant, des situations dans lesquelles le maintien de la structure source n'est pas possible, ou n'est pas souhaitable. Dans ces cas, on procède à des transpositions, ce qui signifie que l'on recourt à des structures cibles différentes des structures sources.

Ainsi, il se peut que la langue cible ne possède pas un certain type de structures. C'est le cas par exemple des structures clitiques, inconnues en anglais. Inversément, l'anglais permet le passif à partir de compléments prépositionnels, ce qui est impossible en français.

Dans le cas où une construction particulière existe dans les deux langues source et cible, il se peut pourtant que des préférences particulières distinguent les deux langues. C'est le cas, par exemple, pour la construction impersonnelle en *on*. Ainsi, une phrase comme (4a) se traduira en anglais par une formule passive comme en (4b) :

- (4)a. On a donné un livre à Paul.
- b. Paul was given a book.

Comme autre exemple, mentionnons encore le cas où des pro-

priétés sélectionnelles d'éléments lexicaux de la langue cible entrent en conflit avec l'hypothèse de la préservation de la structure. Cette situation est illustrée en (5) :

- (5)a. Jean veut que Marie parte.
 b. *Jean wants that Marie leaves.
 c. Jean wants Marie to leave.

Alors que le verbe *vouloir* sélectionne une phrase conjuguée, le verbe anglais *want* ne permet pas une telle construction, comme le montre (5b). Par contre, ce verbe sélectionne une phrase infinitive avec un sujet lexical (structure dite "à marquage de cas exceptionnel"), qui est impossible en français.

Les opérations de transfert appliquées à une structure source produisent une structure syntaxique abstraite (structure-D) en langue cible. Le rôle de la composante transformationnelle consiste à convertir cette structure-D en une structure-S bien-formée reflétant l'ordre des mots et la forme morphologique de la phrase cible.

L'exemple suivant illustre les différents niveaux de représentation impliqués dans le processus complet de traduction d'une phrase : A partir de la structure-S (6b) associée à une phrase telle que (6a), le transfert produit la structure-D cible (6c) qui, après l'application de la transformation passive et de la génération morphologique, donne la structure-S (6d).

- (6)a. Le livre a été volé.
 b. [_S [_{NP} le livre], [_{VP} a été volé [_{NP} e]_i]]
 c. [_S [_{NP} e] [_{VP} steal [_{NP} the book]]]
 d. [_S [_{NP} the book], [_{VP} has been stolen [_{NP} e]_i]]

La traduction de la phrase (7a) est un peu plus complexe. Le transfert lexical et l'opération de projection aboutissent à la structure-D (7c). On observe que l'ordre des arguments du verbe dans cette structure ne correspond pas à celui de la structure source (7b). En effet, le lexique bilingue spécifie que dans la paire *manquer* - *miss*

pertinente pour cet exemple le sujet du verbe *manquer* correspond à l'objet direct du verbe *miss* et que l'objet indirect de *manquer* correspond au sujet de *miss*. On retrouve donc dans la structure-D anglaise le pronom *they* (traduction du clitique *leur*) en position sujet et le pronom interrogatif *who* (traduction de *qui*) en position d'objet direct. Les transformations d'insertion d'auxiliaires, de mouvement-*wh* et d'inversion du sujet et de l'auxiliaire s'appliquent à la structure-D (7c) et produisent la structure-S (7d).

(7)a. Qui est-ce ce qui leur a manqué?

b. [_S [_{NP} qui]_i, est-ce qui [_S [_{NP} e]_i [_{VP} leur_j a manqué [_{PP} e]_j]]]

c. [_S [_{NP} they] [_{VP} miss] [_{NP} who]]

d. [_S [_{NP} who]_i, have [_S [_{NP} they] [_{VP} missed [_{NP} e]_i]]]

D'une façon générale, les transformations correspondent à des séquences d'opérations élémentaires, telles que le déplacement d'un constituant, l'insertion d'un constituant, la copie ou la vérification de traits, etc. Par exemple, la transformation du passif implique l'antéposition de l'objet direct dans la position sujet, l'insertion de l'auxiliaire *être*, et facultativement, l'insertion d'un complément d'agent.

Ce n'est qu'à l'issue du cycle transformationnel que peut s'appliquer la génération morphologique, puisque les transformations sont susceptibles de modifier des traits pertinents pour la morphologie, comme les traits d'accord.

3 Organisation et contenu des lexiques

Comme dans pratiquement tout système de traitement du langage naturel, la composante lexicale occupe une position centrale dans notre système. Ceci d'autant plus que la morphologie figure dans le lexique, comme nous le verrons plus loin. Il est facile de montrer que toutes les opérations impliquées dans la tâche de traduction d'une phrase font appel à la composante lexicale. Ainsi, le découpage d'un texte en phrases et la reconnaissance des mots ne peuvent être envisagés

sans l'information lexicale. De même, au plan syntaxique, le regroupement des syntagmes, la vérification des accords (au sens large), l'interprétation des catégories vides, *etc.* reposent également sur des traits lexicaux, tels que traits flexionnels, sous-catégorisation, traits d'assignation de cas et de rôles thématiques, *etc.* Le transfert, c'est-à-dire la mise en correspondance des structures sources et cibles au niveau abstrait des structures profondes (D-structures), repose également sur l'information lexicale. En effet le transfert d'une structure (syntagme) dans notre système prend la forme d'un parcours récursif impliquant (i) un transfert lexical de la tête du syntagme (ii) le transfert des arguments, des ajouts et des spécificateurs.

Tout transfert lexical est immédiatement suivi d'une projection maximale XP de la catégorie lexicale X. Toujours en fonction de la théorie \bar{X} paramétrée pour la langue cible, les spécificateurs et les ajouts sont attachés dans les positions respectives de spécificateurs et d'ajouts (cibles). Quant aux arguments, leur forme syntaxique ainsi que leur position d'attachement précise est déterminée sur la base d'une table de transferts d'arguments associée aux prédicats dans le lexique bilingue.

Le problème, en effet, est de rendre compte du fait que le sujet d'une phrase source ne doit pas forcément correspondre au sujet cible, ou qu'un objet direct peut fort bien devoir être réalisé sous la forme d'un complément circonstanciel.

(8)a. Paul manque à Pierre.

b. Pierre is missing Paul.

(9)a. fournir qqch à qqn.

b. provide so with sth.

3.1 Structure des lexiques monolingues

Les lexiques monolingues sont du type "à formes pleines", ce qui signifie que l'unité de base est le mot et non le morphème. De plus, toutes les variantes morphologiques (orthographiques) d'un mot sont

explicitement représentées dans le lexique. Toutes ces variantes constituent des entrées séparées mais reliées les unes aux autres, comme nous le verrons ci-dessous. Ainsi, les relations morphologiques (inflection) entre mots sont exprimées au moyen d'un ensemble de relations entre entrées lexicales.

Pour éviter une redondance inutile et surtout pour exprimer des niveaux conceptuels bien distincts, les lexiques monolingues sont constitués de deux types d'unités, les lexèmes et les mots. Un lexème correspond à la notion abstraite d'unité lexicale, alors qu'un mot n'est rien d'autre qu'une forme orthographique (ou phonétique). Autrement dit, à un lexème donné correspondront autant de mots qu'il y a de variantes orthographiques pour l'exprimer. En clair, cela signifie environ 40 mots distincts pour chaque verbe de la première classe, 2 mots pour chaque substantifs, etc. Les propriétés d'un item lexical qui varient en fonction de la morphologie sont associées aux mots, celles qui sont invariantes aux lexèmes. Ainsi, les traits de temps, nombre et personne des verbes apparaissent dans les entrées de mots, alors que les traits de sous-catégorisation et les restrictions selectionnelles figurent dans les entrées des lexèmes.

Comme l'information associée aux lexèmes est de nature syntaxico-sémantique (elle concerne en particulier la structure argumentale associée aux éléments prédicatifs) il est naturel de considérer que des lectures syntaxico-sémantiques différentes d'un verbe correspondent à des lexèmes distincts. Concrètement, et en simplifiant quelque peu, cela signifie que nous aurons un lexème distinct pour chaque combinaison possible de traits de sous-catégorisation pour un verbe donné.

La figure 2 ci-dessous donne une illustration des relations entre mots et lexèmes pour le verbe *avoir*.

3.2 Lexique bilingue

La distinction entre les concepts de mots et de lexèmes est fondamentale lorsque l'on aborde la notion de lexique bilingue. En effet, au niveau lexical, les correspondances ne s'établissent pas au niveau

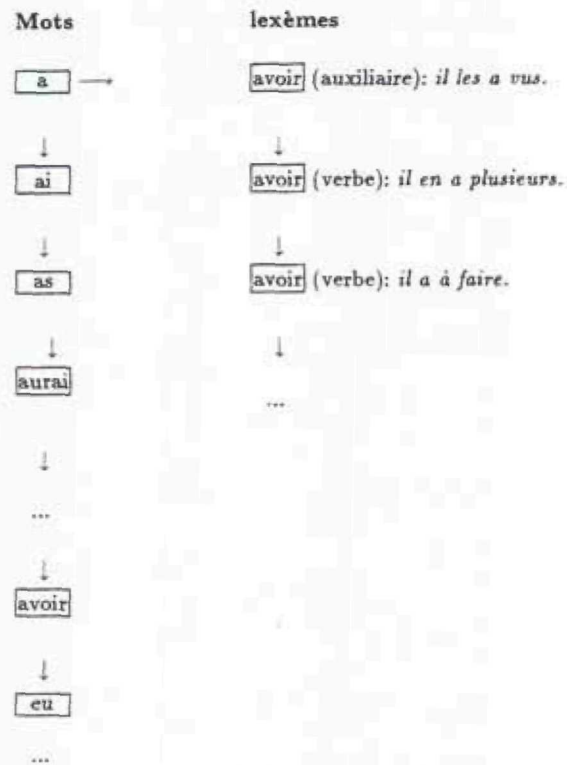


Figure 2: Exemple de relation entre mots et lexèmes

superficiel des mots orthographiques, mais bien à celui des lexèmes. Ainsi, il n'y aurait aucun sens à vouloir mettre en correspondance les mots *étaient* et *were*, *va* et *goes* comme l'illustrent les exemples suivants :

- (10)a. Ils étaient venus.
 b. They had come.
- (11)a. il va le faire.
 b. He's going to do it.

En effet, le choix d'une forme morphologique (orthographique) particulière dépend de plusieurs facteurs internes à la langue, sans

rappports avec la langue source. C'est bien au niveau plus abstrait des lexèmes, c'est-à-dire des lectures particulières d'un item lexical que peuvent s'établir les correspondances lexicales, et dans notre système, le lexique bilingue n'est rien d'autre qu'une table de correspondances entre lexèmes sources et lexèmes cibles. Chacune des entrées de ce lexique spécifie un lexème source et un lexème cible. Si, comme c'est très souvent le cas, un lexème source a plusieurs correspondants (ou inversement un lexème cible plusieurs correspondants), il y aura autant d'entrées que de correspondances.

Il y a, cependant, d'autres informations qu'un lexique bilingue se doit de spécifier. Par exemple, dans le cas des lexèmes susceptibles de prendre des arguments (verbes), il est nécessaire de spécifier les correspondances entre arguments. L'exemple suivant montre pourquoi.

- (12)a. Cet homme vous donnera tous les renseignements dont vous avez besoin.
 b. This man will provide you with all the information you need.

Le verbe *fournir* dans (12a) sélectionne un objet direct et un objet indirect. Cependant, dans (12b), l'objet de *provide* correspond à l'objet indirect de la phrase source, et c'est un complément prépositionnel qui correspond à l'objet direct de la phrase source.

3.3 Rôle et place de l'interaction

Dans le système ITS-2, l'interaction entre le processus de traduction et l'utilisateur prend place à plusieurs niveaux. Un premier niveau d'interaction se situe lors de l'analyse lexicographique, lorsque l'analyseur syntaxique lit la phrase source qui lui est soumise. A ce niveau, en effet, le programme peut buter sur des mots qui lui sont inconnus, c'est-à-dire des mots qui ne figurent pas dans le lexique. Il peut s'agir de noms propres ou de noms communs qui ne se trouvent pas dans ses dictionnaires, ou encore de mots mal orthographiés. Une fenêtre d'interaction est alors ouverte, exposant la phrase et mettant en évidence le ou les mots concernés. Un menu permet à l'utilisateur de choisir l'action la plus appropriée parmi les alternatives suivantes :

- corriger la phrase;
- valider le mot comme nom propre;
- créer une nouvelle entrée de dictionnaire;
- interrompre le processus.

Un deuxième niveau d'interaction prend place lors de l'analyse syntaxique. Il se peut que l'analyseur rencontre des difficultés qu'il est incapable de surmonter aux moyen de ses règles et heuristiques. Un exemple typique d'une telle situation est celui des ambiguïtés d'attachement de syntagmes prépositionnels, illustré en (13) :

- (13)a. Jean frappe le chien avec le collier.
 b. Jean frappe le chien avec les taches blanches.
 c. Jean frappe le chien avec le poing.

Dans la phrase (13a), le syntagme prépositionnel *avec le collier* peut se rattacher soit au syntagme nominal *le chien* et alors il est interprété comme un modificateur de ce syntagme nominal, soit au verbe *frapper* et alors il est interprété comme un complément de ce verbe. La première interprétation est sans doute la plus naturelle dans le cas de la phrase (13b), la deuxième dans la phrase (13c). Dans des cas de ce type, la fenêtre de dialogue peut afficher la chaîne de caractères correspondant au syntagme prépositionnel à attacher et les têtes des syntagmes auxquels il peut s'attacher. Dans notre exemple, cela reviendrait à afficher le syntagme prépositionnel [_{pp} avec le collier] et demander s'il faut l'attacher au syntagme nominal objet direct *chien* ou au verbe *frappe*. Comme toujours, l'utilisateur a également la possibilité de modifier la phrase ou d'interrompre la session.

Un autre type de problème qui exige une intervention de l'utilisateur est celui de l'interprétation des pronoms, comme dans les exemples (14). Contrairement aux exemples précédents, le problème de l'ambiguïté des pronoms ne se pose en général pas pendant la phase d'analyse, mais bien dans celle de la génération, lorsqu'il s'agit de déterminer la forme du pronom cible.

- (14)a. Jean parle à Marie de son dernier livre.
 b. Jean a appelé Paul parce qu'il lui doit encore beaucoup d'argent.
 c. Les enfants se sont lavés.

Le possessif *son* dans l'exemple (14a) peut renvoyer soit à *Jean*, soit à *Marie*, soit encore à une autre personne mentionnée précédemment dans le discours. Si la langue cible est l'anglais, la forme du pronom possessif dépendra du genre de son antécédent (*his/her*). Dans un tel cas, la fenêtre d'interaction va afficher la liste des antécédents potentiels et demander à l'utilisateur d'effectuer le bon choix.

La phrase (14b) contient deux pronoms, *il* et *lui*. Chacun de ces pronoms peut prendre comme antécédent *Jean*, *Paul* ou, éventuellement une tierce personne, non mentionnée dans la phrase. Les principes grammaticaux de la théorie du liage éliminent la possibilité d'un même antécédent pour les deux pronoms. Il reste donc deux combinaisons de référence s'il l'on exclut toute référence contextuelle, six sinon.

Enfin, dans le cas de la phrase (14c), l'ambiguïté concerne l'interprétation du pronom anaphorique *se*. Dans cette phrase, *se* peut être compris avec un sens réfléchi ou réciproque. La fenêtre d'interaction peut afficher l'information suivante :

- (15) Les enfants se sont lavés.
 a- les uns les autres ? (réciproque)
 b- eux-mêmes ? (réfléchi)

Enfin, de très nombreuses ambiguïtés se présentent au niveau du transfert, et notamment du transfert lexical, lorsque le système doit décider comment un élément lexical de la langue source doit être traduit. On sait que la correspondance entre items lexicaux de deux langues n'est pratiquement jamais une correspondance univoque. Autrement dit, il y a pratiquement toujours plusieurs possibilités de traduction pour un mot donné, même lorsque l'environnement syntaxique de ce mot a été bien dégagé. Cela est en partie une conséquence directe de l'ambiguïté lexicale, dans la mesure où les différents sens associés à un mot sont susceptibles de traductions différentes.

Il est légitime de penser que le dictionnaire bilingue contient toutes les correspondances possibles pour un item donné. On peut donc afficher cette liste dans la fenêtre d'interaction. Cette façon de faire nécessite, cependant, une connaissance de la langue cible. Si, comme nous le souhaitons, l'interaction doit être limitée à la langue source, la sélection d'une correspondance se fera sur la base des différentes acceptions du mot dans la langue source. On affichera alors le mot avec ses différents sens, en utilisant des synonymes. Par exemple, dans le cas (très simplifié) du mot *assurance*, on affichera

- *assurance* syn. confiance
- *assurance* syn. contrat

La sélection du sens le plus approprié d'un mot, qui permet au système de choisir une traduction correcte, est basée sur des exemples ou des synonymes associés aux entrées lexicales. Ce travail, qui reste à faire, nécessite d'abord de répertorier tous les termes d'une langue pour lesquels il existe plus d'une traduction possible, puis d'ajouter à ces entrées les informations susceptibles d'être affichées lors des dialogues de clarification (brèves définitions, synonymes, éventuellement exemples).

Bibliographie

- Berwick R., S. Abney et C. Tenny (éd.) (1991), *Principle-Based Parsing : Computation and psycholinguistics*, Kluwer Academic Publishers.
- Chomsky, N. (1986), *Knowledge of Language*, Praeger.
- Laenzlinger, C. et E. Wehrli (1992), "FIPS : un analyseur interactif pour le français", *TA Informations* 2, 35-49.