

Traduction, restructurations syntaxiques et grammaires de correspondance

Sylvain Kahane

Modyco, Université Paris 10 - Nanterre & CNRS

sk@ccr.jussieu.fr

<http://www.kahane.fr>

Résumé Cet article présente une nouvelle formalisation du modèle de traduction par transfert de la Théorie Sens-Texte. Notre modélisation utilise les grammaires de correspondance polarisées et fait une stricte séparation entre les modèles monolingues, un lexique bilingue minimal et des règles de restructuration universelles, directement associées aux fonctions lexicales syntaxiques.

Abstract This paper presents a new formalisation of transfer-based translation model of the Meaning-Text Theory. Our modelling is based on polarized correspondence grammars and observes a strict separation between monolingual models, the bilingual lexicon and universal restructuring rules, directly associated with syntactic lexical functions.

Mots-clés : Traduction automatique, paraphrase, restructuration syntaxique, TST (Théorie Sens-Texte), grammaire de dépendance, fonction lexicale, lexique bilingue, GUP (Grammaire d'Unification Polarisée), grammaire de correspondance, grammaires synchrones.

Keywords: Machine translation, paraphrase, syntactic restructuring, MTT (Meaning-Text Theory), dependency grammar, lexical function, bilingual lexicon, PUG (Polarized Unification Grammar), correspondence grammar, synchronous grammars.

1 Introduction

Cet article s'intéresse à la modélisation des restructurations syntaxiques dans un système de traduction par transfert. L'architecture que nous considérons a été introduite dès les premiers travaux en traduction automatique [TA] (Kulagina & Mel'čuk 1967) et reprise par différents systèmes comme la plateforme Etap-2 (Apresian *et al.* 2003), le système TransLex (Nasr *et al.* 1997, Lavoie *et al.* 2000) ou le projet Eurotra (Arnold & des Tombes 1987, Danlos & Samvelian 1992). Cette architecture se caractérise par une réduction au minimum du lexique bilingue ; les règles qui permettent les restructurations sont universelles et séparées des grammaires monolingues.

Notre contribution se situe uniquement à un niveau théorique et formel. Les exemples que nous étudions sont très bien connus (Tesnière 1959 (chapitres sur la métataxe), Lindop & Tsujii 1993, Dorr 1994) et notre travail n'a fait l'objet d'aucune implémentation. Il nous semble néanmoins que même si la plupart de ces phénomènes sont considérés dans de nombreuses études, il est encore possible d'améliorer la modélisation des traductions mettant en jeu de telles restructurations syntaxiques. Bien que de nombreux systèmes à transfert aient été développés, peu de travaux proposent un formalisme propre pour l'écriture du module de transfert. La plupart des systèmes utilisent processus de transfert ad hoc basé sur de la réécriture d'arbre et des transformations d'arbres. Les formalismes les plus achevés que je connaisse utilisent des grammaires synchrones (Abeillé *et al.* 1990, Nesson *et al.* 2006 ou Ding & Palmer 2005), mais de telles architectures n'opèrent pas une séparation forte entre lexicale bilingue et grammaires monolingues, puisque ce sont les règles des grammaires monolingues qui sont alignées.

Notre modélisation s'inscrit dans le cadre de la Théorie Sens-Texte [dorénavant TST] (Žolkovskij & Mel'čuk 1967, Mel'čuk 1988a), laquelle théorie linguistique a d'ailleurs été initialement élaborée dans la perspective de TA. La modélisation des restructurations en TST a été traitée en détail dans Mel'čuk 1988b et plus récemment dans Mel'čuk & Wanner à paraître. La formalisation que nous proposons, utilisant les grammaires de correspondance polarisées développées dans Kahane 2004 et Kahane & Lareau 2005 nous semble à la fois plus simple, plus rigoureuse et plus aboutie.

Nous commencerons par présenter, dans la Section 2, la notion centrale d'unité significative, une nouvelle définition de la structure syntaxique profonde et le principe de la traduction par commutation d'unités significatives. La Section 3 sera consacrée aux restructurations syntaxiques et au lexique bilingue dont on a besoin en conséquence. Les règles de restructuration et la grammaire de correspondance polarisée qui les met en jeu seront introduites à la Section 4. La Section 5 présentera le problème des restructurations multiples et l'algèbre des fonctions lexicales syntaxiques. Nous finirons avec le problème classique des verbes de mouvement à la Section 6.

2 Traduction, unités significatives, structure syntaxique profonde

Traduire, c'est exprimer à peu près le même sens dans une autre langue. Dans la quasi-totalité des cas, il est possible d'obtenir une traduction convenable en remplaçant chaque unité significative de l'énoncé dans la langue cible par une unité significative de la langue source exprimant un sens similaire (et a priori rarement identique, d'où l'impossibilité de trouver une interlingua satisfaisante, même pour un unique couple de langues). En fait, les seuls cas où ceci n'est pas possible sont les cas où l'une des unités significatives de l'énoncé source ne peut pas être traduit de façon satisfaisante, cas qui échappent pour l'instant à toute tentative de traduction automatique.

Les *unités significatives* [US] (Martinet 1960 : chap. 4 ; Ducrot 1995) sont les signes linguistiques indécomposables dans leur signifié¹ : elles se répartissent en *unités lexicales* (y compris les locutions), *unités grammaticales* et *constructions* (au sens des grammaires de construction ; Goldberg 1995). Les US sont les unités de choix ; par exemple dans l'énoncé *La moutarde me monte au nez*, au sens de 'Je sens la colère m'envahir', il y a 4 choix faits par

¹ Les signes linguistiques indécomposables dans leur signifiant sont les morphes.

le locuteur et autant d'unités significatives : la locution « LA MOUTARDE MONTER AU NEZ¹ », le pronom MOI (sous sa forme atone *me*) qui est l'unique actant de cette locution, le temps présent et la construction déclarative (nous ne considérerons pas cette US dans la suite, mais elle est bien réelle et s'oppose par exemple à la construction interrogative).

Bien que la *structure syntaxique profonde* [SSyntP] soit présentée dans de nombreuses études en TST (Mel'čuk 1988a), elle nous semble devoir faire l'objet d'une meilleure définition et de quelques aménagements. La SSyntP d'un énoncé est la structure qui indique comment les US de cet énoncé se combinent. Elle peut être vue comme la structure de dérivation de l'interface syntaxe-sémantique (Kahane 2003). La syntaxe de surface [SyntS] impose une structure hiérarchique, squelette sur lequel nous indiquerons les numéros différenciant les différents actants. Relations syntaxiques de surface, régimes et accords n'apparaissent pas dans la SSyntP et peuvent être recalculés à partir du lexique et de la grammaire. Les unités lexicales [UL] (que nous notons en petites capitales, en les faisant suivre de leur partie du discours [pdd]) occupent les nœuds de l'arbre SyntP. Les unités grammaticales [UG] sont indiquées en indice de l'UL avec laquelle elles se combinent. Une des difficultés pour la SSyntP est que les US peuvent se combiner différemment au niveau sémantique [Sém] et syntaxique. Or la SSyntP, qui sert justement d'interface entre Sém et SyntS, se doit d'encoder simultanément les deux. Prenons l'exemple d'un verbe à montée comme SEMBLER dans *Max semble dormir*. Au niveau sémantique, en raison de la synonymie avec *Il semble que Max dorme*, on peut affirmer que 'sembler' n'a qu'un argument qui est 'dormir', qui a lui-même pour argument 'Max'. Par contre, au niveau syntaxique, MAX est le sujet du verbe SEMBLER et pas celui de DORMIR (d'où la métaphore de la « montée » du sujet potentiel de DORMIR en sujet de SEMBLER). Dans la SSyntP, nous indiquerons donc que SEMBLER a deux actants, que nous numérotons 1 et 2, mais que le 1^{er} actant est « additionnel », ce que nous notons par un +, c'est-à-dire qu'il n'apparaît qu'en SyntS. Par ailleurs, nous indiquons, par un exposant sur la pdd (V¹), que le 1^{er} argument de DORMIR n'est pas réalisé comme actant de DORMIR et, par un 1 en exposant sur la dépendance entre son gouverneur et lui (2¹), qu'il est réalisé comme 1^{er} actant de son gouverneur.

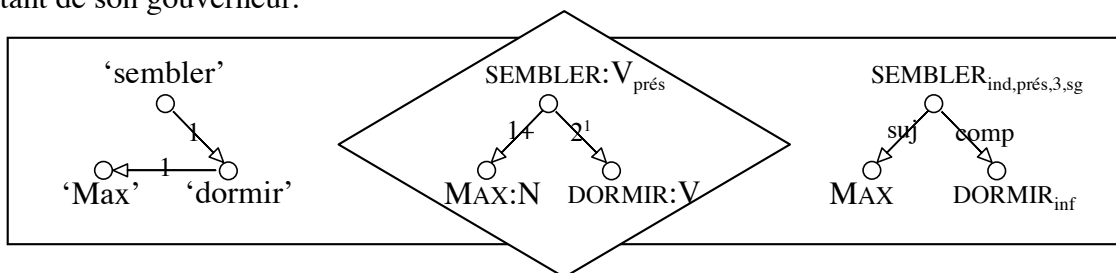


Fig. 1 Les structures Sém, SyntP et SyntS de *Max semble dormir*

En remplaçant dans un énoncé, une US par une US équivalente (c'est-à-dire ayant à peu près le même sens), on obtient une paraphrase. Une traduction sera obtenue en commutant simultanément toutes les US d'un énoncé de la langue source par des US de la langue cible équivalentes². Lorsque chaque US de l'énoncé source est traduite par une US de construction

² Nous laissons de côté la question des collocatifs, par ailleurs fort bien traitée dans le cadre de TST. Un *collocatif* est une UL dont le choix est lexicalement contraint par une autre UL, appelée la *base* de la *collocation*. Par exemple pour traduire *gros fumeur*, il faut savoir que GROS est un collocatif de FUMEUR. Le lexique du français devra donc indiquer que $\text{Magn}(\text{FUMEUR}) = \text{GROS}$, où Magn est la fonction lexicale qui associe à une base un intensifieur. Pour une traduction en anglais, on traduira FUMEUR par SMOKER et on consultera le lexique de l'anglais pour connaître $\text{Magn}(\text{SMOKER})$: $T(\text{gros fumeur}) = T(\text{FUMEUR} \oplus \text{Magn}) = T(\text{FUMEUR}) \oplus T(\text{Magn}) = \text{SMOKER} \oplus \text{Magn} = \text{heavy smoker}$.

équivalente, la traduction est particulièrement simple : les deux énoncés ayant exactement la même SSyntP. Nous illustrons par la traduction anglais-français suivante :

$$(1) \quad \begin{array}{l} \mathbf{a. He slept.} \\ \mathbf{b. Il a dormi.} \end{array} \quad T \left(\begin{array}{c} \text{SLEEP:V}_{\text{perf}} \\ \downarrow \\ \text{HE:N}_{\text{sg}} \end{array} \right) \equiv T(\text{SLEEP}):V_{T(\text{perf})} \equiv \text{DORMIR:V}_{\text{passé-c}} \\ \downarrow \\ T(\text{HE}):N_{T(\text{sg})} \equiv \text{LUI:N}_{\text{sg}}$$

Cette traduction est réalisée en deux étapes. La première étape modélise notre postulat : la traduction d'un énoncé peut être réalisé en traduisant chaque US séparément. La deuxième étape commute chaque US avec une de ses traductions possibles en utilisant le lexique bilingue anglais-français :

$\mathcal{L}_{E-F} = \begin{array}{l} (\text{SLEEP:V}, \text{DORMIR:V}) \\ (\text{HE:N}, \text{LUI:N}) \\ (:V_{\text{perf}}, :V_{\text{passé-c}}) \\ (:N_{\text{sg}}, :N_{\text{sg}}) \end{array}$	$\begin{array}{l} T_{F-E}(\text{SLEEP}) = \text{DORMIR} \\ T_{F-E}(\text{HE}) = \text{LUI} \\ T_{F-E}(\text{perf}) = \text{passé-c} \\ T_{F-E}(\text{sg}) = \text{sg} \end{array}$
--	--

Le lexique bilingue (colonne de gauche) a pour ces cas « simples » la forme d'une liste de couples d'unités significatives qui sont la traduction l'une de l'autre. Autrement dit, dire que (SLEEP:V, DORMIR:V) est dans le dictionnaire English-Français revient à dire que $T_{E-F}(\text{SLEEP}) = \text{DORMIR}$ ou que $T_{F-E}(\text{SLEEP}) = \text{DORMIR}$, où T_{E-F} et T_{F-E} sont des fonctions de traduction³.

3 Restructurations et lexique bilingue

Considérons les exemples classiques suivants :

- (2) **a. I miss you.**
b. Tu me manques.
- (3) **a. Ich schwimme gern.** (litt. je nage volontiers)
b. J'aime nager.
- (4) **a. Zoe needs a book.**
b. Zoé a besoin d'un livre.

Pour la traduction (2), on doit non seulement indiquer que MISS et MANQUER sont la traduction l'un de l'autre, mais aussi qu'une conversion des actants est opérée. Pour (3), le lexique allemand-français devra indiquer que GERN est un adverbe⁴ dont le gouverneur verbal correspond au deuxième actant de AIMER, tandis que le sujet de ce gouverneur verbal correspond au sujet de AIMER. Nous pouvons modéliser cela de deux façons : soit en indiquant les diathèses dans le lexique bilingue (à la Nasr *et al.* 1997), soit en exprimant ce changement de diathèse par une fonction lexicale (à la Mel'čuk & Wanner 2006). Pour la diathèse, nous utilisons les notations proposées par Kahane & Polguère 2001 pour l'encodage explicite des fonctions lexicales : $[x,y]$ désigne la liste des actants par ordre de saillance croissante (x est

³ T_{E-F} et T_{F-E} ne sont pas réellement des fonctions, car la correspondance bilingue est en fait multivoque. Par ailleurs, les dictionnaires bilingues ne sont généralement pas considérés comme symétriques. Nous pensons pour notre part qu'ils doivent l'être et que le fait qu'une traduction dans un sens soit plus utile que dans l'autre provient d'informations monolingues sur l'usage des unités significatives en correspondance.

⁴ Plus précisément, GERN est ce que nous appelons un Adv', c'est-à-dire un Adv qui contrôle le sujet de son gouverneur et a donc deux arguments sémantiques.

donc le premier actant ; pour un verbe c'est le sujet). Pour un Adj ou un Adv, le premier « actant » est en fait le gouverneur syntaxique, ce que nous signalons par le symbole \wedge (à lire comme une flèche vers le haut).

(MISS:V[1,2], MANQUER:V[2,1])
 (GERN:Adv \wedge [2[1] \wedge], AIMER:V[1,2])
 (NEED:V, BESOIN:N)

$T_{E-F}(\text{MISS}) = \text{Conv}_{12}(\text{MANQUER})$
 $T_{D-F}(\text{GERN}) = \text{Adv}_2^1(\text{AIMER})$
 $T_{E-F}(\text{NEED}) = V_0(\text{BESOIN})$

Conv_{12} effectue la conversion des actants 1 et 2. Adv_2^1 adverbialise son mot-clé : le 2 en indice indique que cet Adv va modifier le 2^{ème} actant du mot-clé, le 1 en exposant indique qu'il contrôle aussi le 1^{er} actant de son gouverneur⁵. V_0 est une verbalisation sans changement de diathèse. Pour (4), le lexique bilingue indique que la traduction du verbe NEED n'est pas un verbe mais le nom BESOIN et c'est tout ; c'est le lexique du français qui indique par ailleurs que ce nom possède le verbe support : $\text{Oper}_1(\text{BESOIN:N}) = \text{AVOIR}$.

Voyons maintenant comment s'effectuent les restructurations nécessaires pour ces trois exemples. A chaque fonction lexicale syntaxique est associée une *règle de restructuration* (Mel'čuk 1988). Nous utilisons pour écrire ces règles le formalisme des Grammaires d'Unification Polarisées [GUP] (Kahane 2004) et plus précisément les grammaires de correspondance polarisées (Kahane & Lareau 2005). Trois types d'objets⁶ sont considérés : les nœuds syntaxiques représentés par des ronds, les dépendances syntaxiques entre nœuds représentées par des flèches et les liens de correspondance entre nœuds, représentés par des traits pointillés assortis d'un losange pour la polarisation. Chaque objet reçoit une polarité sous la forme d'une couleur noire ou blanche : les objets noirs sont ceux qui sont manipulés par la règle, tandis que les objets blancs expriment des besoins et doivent obligatoirement être fusionnés avec un objet noir. La correspondance est achevée lorsque, après fusion, tous les objets sont noirs.

Les règles associées aux fonctions lexicales Conv_{12} et Adv_2^1 manipulent uniquement des dépendances⁷ : les nœuds syntaxiques, ainsi que les liens de correspondance entre eux sont donc blancs et doivent être construits par d'autres règles. La règle pour Oper_1 introduit le nœud supplémentaire pour le verbe support Oper_1 et construit un lien⁸ (Fig. 2).

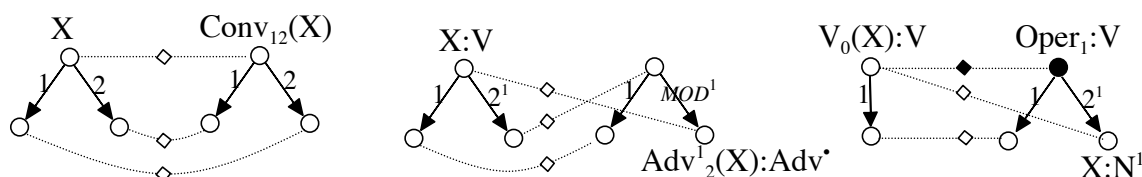


Fig. 2. Règles de restructuration associées à Conv_{12} , Adv_2^1 et Oper_1

⁵ Donnons un autre exemple monolingue : $\text{Adv}_2^1(\text{PRECIPITER}) = \text{PRECIPITAMMENT}$ (*Max a précipité son départ* = *Max est parti précipitamment*).

⁶ Par souci de simplification, nous n'avons pas représenté les grammèmes comme des objets, mais il le faudrait. Par ailleurs, bien qu'elle l'assure en partie, la grammaire de correspondance n'a pas pour objet de vérifier la bonne formation des SSyntP qu'elle met en correspondance. Ceci peut et doit être assuré par une grammaire indépendante. Voir Kahane & Lareau 2005 pour l'interfaçage de grammaires de correspondance et de grammaires de bonne formation.

⁷ Nous notons MOD la relation modificative (plutôt que ATTR comme il est usuel en TST). Il s'agit généralement d'une relation prédicative dont le prédicat est le dépendant. Le 1 en exposant indique que le 1^{er} actant du gouverneur est également un argument du dépendant.

⁸ Le fait que $V_0(X)$ soit liés à la fois à X et à Oper_1 est utile. En effet, un modifieur peut aussi bien aller sur X (*Max a **grand** besoin d'un livre*) que sur Oper_1 (*Max a **vraiment** besoin d'un livre*).

La correspondance entre deux SSyntP s'effectue en combinant les règles de restructuration et des *règles de copie* assurant la copie des unités lexicales, unités grammaticales et dépendances qui n'ont pas encore été prise en compte (Fig. 3)



Fig. 3. Règles de copie pour AIMER, le présent et une dépendance 1

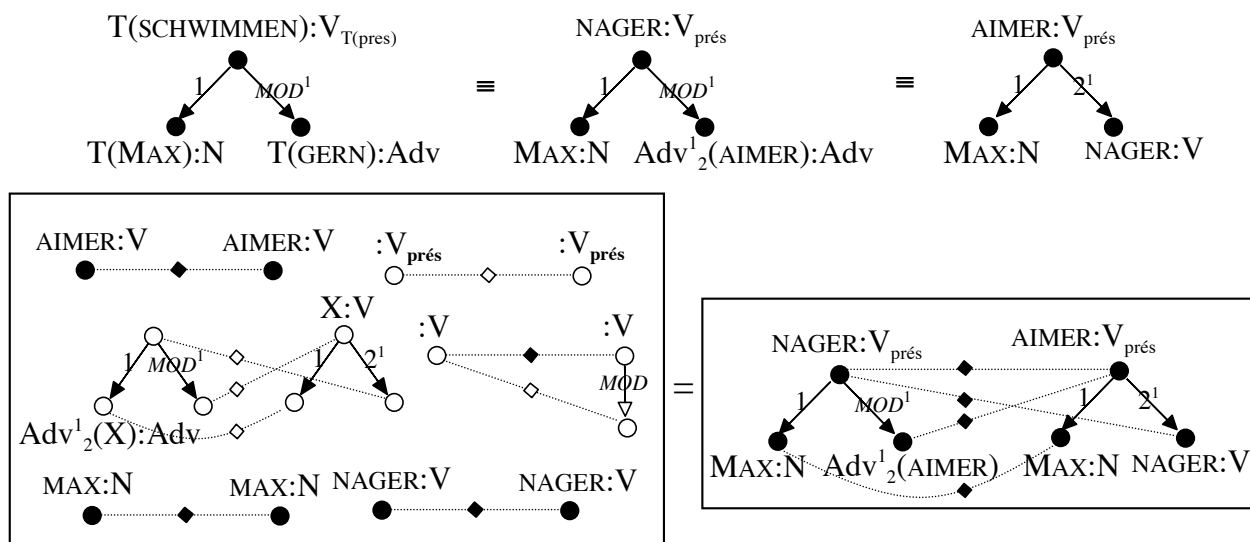


Fig. 4. La traduction (4)

La Fig. 4 montre le fonctionnement complet du système sur la traduction (3) : la première étape utilise uniquement le lexique bilingue sans modifier la structure ; la deuxième étape (détaillée dans la partie basse de la figure) effectue la restructuration. On notera qu'en plus de la règle de restructuration associée à Adv^1_2 et des règles de copie une règle permet de créer un lien entre deux verbes si l'un des deux est associé à un modifieur de l'autre. Cette règle est essentielle pour permettre la copie du temps présent de NAGER sur AIMER. Une telle règle revient à autoriser un changement de tête pendant la traduction : le sens lexicalisé comme tête de la phrase n'est plus le même⁹.

Comme on le voit PUG permet de modéliser très simplement la combinaison des règles pour une simple fusion des objets noir et blanc (des nœuds, comme des liens de correspondance) et offre ainsi un moyen élégant d'écrire des règles de transfert.

4 Restructurations multiples

Une des propriétés les plus intéressantes du formalisme est qu'il permet de prendre en compte des restructurations multiples simultanées. Nous allons montrer ça sur un exemple monolingue (ce qui ne change rien à la question) :

- (5) **a.** Max a été malade pendant 2 semaines.
b. La maladie de Max a duré deux semaines.

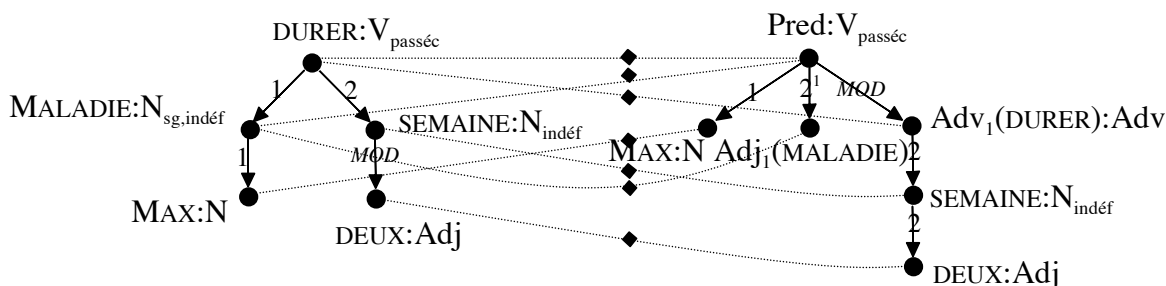
⁹ Il faudrait indiquer que ce lien est de nature « secondaire », bloquant ainsi le transfert d'actant entre les deux têtes. Le formalisme GUP permet aisément ceci en typant différemment les deux liens, à la différence des grammaires synchrones (Nesson *et al.* 2006) où les liens sont encodés par une coïndexation.

Pour réaliser, ce paraphrasage deux règles lexicales sont nécessaires :

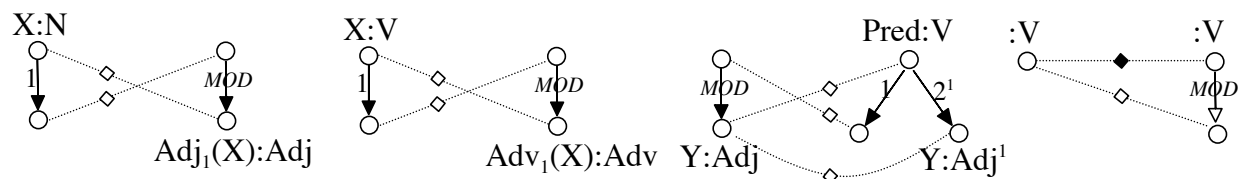
MALADE = Adj₁(MALADIE)

PENDANT = Adv₁(DURER)

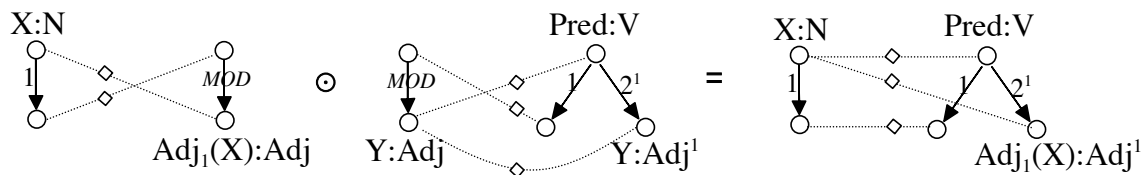
La fonction Adj₁ indique que MALADE est un Adj qui a le même sens que MALADIE et dont le gouverneur est le 1^{er} actant de MALADIE. Idem pour Adv₁ (le 2^{ème} actant de DURER devient le 2^{ème} actant de PENDANT, lequel est une préposition, c'est-à-dire un averbe « transitif »). Par ailleurs, nous introduisons la fonction lexicale Pred, qui associe à chaque Adj un verbe support, lequel est la copule pour tous les Adj en français : Pred = ETRE. La paraphrase est réalisée par l'équivalence suivante :



Notons enfin que les règles de transfert doivent s'appliquer simultanément : il n'est pas possible de changer DURER en PENDANT puis MALADIE en MALADE, car en même temps que DURER devient PENDANT, son premier actant qui est un nom (MALADIE) doit devenir un verbe (en fait il devient un adjectif, MALADE, translaté en verbe par Pred). Pour ce transfert, quatre règles de restructuration vont devoir être agir conjointement. Les trois premières sont les règles structurelles associées aux fonctions lexicales Adj₁, Adv₁ et Pred. Les règles pour Adj₁ et Adv₁ disent que ce sont des modificateurs dont le gouverneur correspond au 1^{er} actant du mot-clé ; la règle pour Pred associe à un Adj une tournure verbale dont le 1^{er} actant est le gouverneur de l'Adj. La dernière est la règle de changement de tête vu précédemment¹⁰.



Les règles associées à Adj₁ et Pred ne vont pas s'appliquer telle quelle. Elles doivent d'abord être combinées entre elle pour donner une nouvelle règle permettant le passage direct du N MALADIE à la structure Pred+Adj ETRE MALADE.



Les règles structurelles et les fonctions lexicales forment ainsi un système algébrique avec une opération de composition, que nous notons ⊙ (Kahane & Polguère 2001). La composition

¹⁰ En changeant de tête syntaxique, on modifie fortement la structure informationnelle de la phrase, la partition sujet-verbe correspondant souvent en français à la partition thème-rhème. Une autre raison de la différence sensible de sens entre (5)a et b est le fait que le remplacement d'un adjectif par un nom entraîne l'introduction d'un nombre et d'une détermination sur ce nom et donc de deux nouvelles unités significatives que sont le singulier et le défini.

des règles structurelles est assurée par le « recollement » de la partie gauche de la 1^{ère} règle sur la 2^{ème}, comme on le voit sur la figure précédente.

5 Verbes de mouvement

Les verbes de mouvement représentent un problème classique de traduction :

- (6) a. *Zoe swam accross the river.*
b. *Zoé a traversé la rivière à la nage.*
- (7) a. *Zoe drove to the hospital.*
b. *Zoé est allé à l'hôpital en voiture.*
- (8) a. *Max crawled out of the den.*
b. *Max est sorti de la tanière en rampant.*

Comme on le sait, pour reprendre la terminologie de l'article fondateur de Talmy 1976, l'anglais est *satellite-frame*, tandis que le français est *verb-frame*, c'est-à-dire que l'anglais indique la trajectoire du mouvement par un satellite (et la « manière » par le verbe), tandis que le français l'indique par le verbe principal (et la « manière » par un circonstanciel). Ceci se traduit lexicalement par le fait qu'un verbe comme SWIM ou CRAWL, à la différence de ses équivalents français, NAGER et RAMPER, possède un 2^{ème} actant qui est un complément locatif. En conséquence, lorsque ce 2^{ème} actant est instancié, le verbe anglais ne peut généralement pas être traduit par un verbe français et une restructuration est nécessaire pour attribuer ce 2^{ème} actant à un autre verbe. C'est ce que fait la règle de la Fig. 5, qui remplace la préposition locative Y par la tournure verbale GO Y en même temps qu'elle remplace le verbe de mouvement X par un Adv¹⁺ (le fait que X soit un verbe de mouvement assure que l'introduction de GO n'ajoute pas de sens). Un Adv¹⁺ peut être exprimé par le gérondif (Adv¹⁺(RAMPER) = *en rampant*). Le + indique que la relation MOD est additionnelle : il n'y a pas de relation prédicative entre un Adv₊ et son gouverneur ; les deux forment plutôt une coprédication. Certains verbes de mouvement possèdent un Adv¹⁺ spécifique : Adv¹⁺(NAGER) = 'A LA NAGE', Adv¹⁺(CONDUIRE) = 'EN VOITURE'.

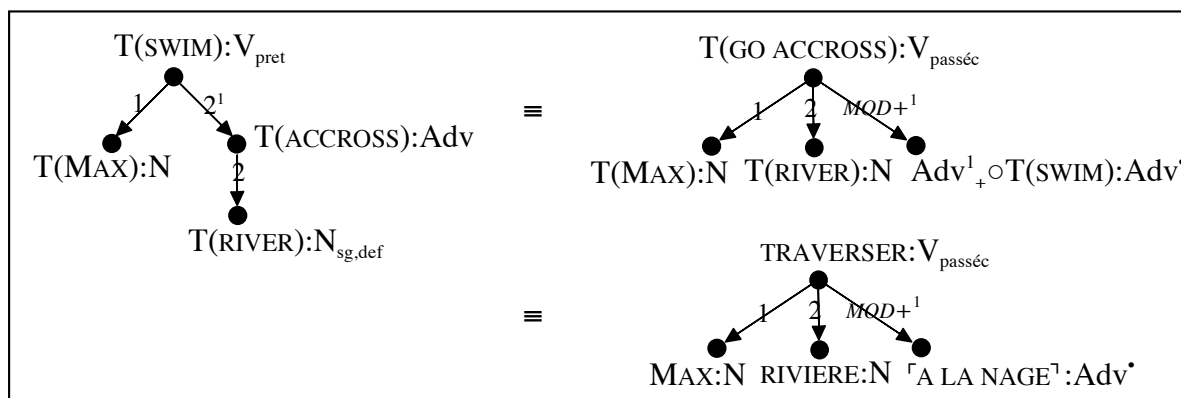
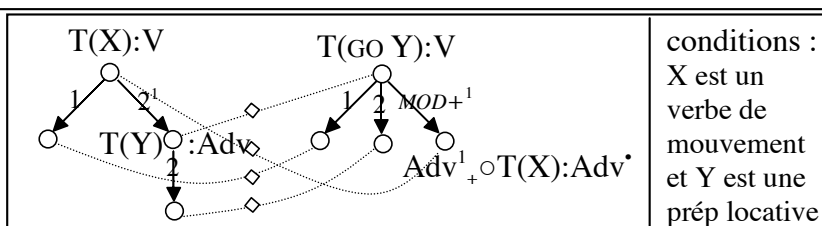


Fig. 5. La traduction (6) et la règle de restructuration correspondante.



Dorr 1994 argumente qu'un transfert direct est difficilement applicable dans de tels cas et propose une solution utilisant une représentation plus profonde supposée être une interlingua. Nasr *et al.* 1997 propose une solution avec transfert basée décomposition sous-jacente des UL et mettant en jeu un échange des traits sémantiques correspondant à la trajectoire et la manière du mouvement. Notre solution est plus simple et notre lexique bilingue contient simplement des entrées telles que (GO OUT:V, SORTIR:V), (GO ACROSS:V, TRAVERSER:V), (GO TO:V, ALLER (à/jusqu'à):V) et la règle de restructuration de la Fig. 5 permet à ces règles de s'appliquer dans tous les cas.

6 Conclusion

L'architecture que nous avons présentée (et qui, rappelons-le, a été proposée dès les années 60) est particulièrement économique : la plus grande partie des règles utilisées appartiennent aux modèles monolingues (en particulier la description des collocations), les restructurations sont assurées par des règles universelles et le lexique bilingue est ainsi ramené à une taille minimale. Chaque règle de restructuration est ancrée par une fonction lexicale syntaxique qui contrôle quand la règle peut être utilisée, c'est-à-dire quand la règle de restructuration peut se combiner avec une correspondance entre deux UL modélisée par cette fonction lexicale dans les lexiques monolingues ou bilingue. Le formalisme des grammaires de correspondance polarisées permet une écriture élégante des règles et un contrôle facile de ce que manipule exactement chaque règle. L'implémentation du modèle doit faire l'objet d'études supplémentaires.

Notons qu'il est possible d'effectuer le transfert directement au niveau SyntS (voir par ex. Schubert 1987). Mais dans ce cas, on devra se faire correspondre des portions des SSyntS, portions qui correspondent en fait à des US avec leur régime. Un tel mécanisme (équivalent aux grammaires synchrones à la Ding & Palmer 2005 ou Nesson *et al.* 2006) revient donc à intégrer dans le système de transfert une description de la SyntS des US et rend le développement du lexique bilingue extrêmement plus coûteux (d'où le recours à une extraction automatique)¹¹.

Remerciements

Je remercie Laurence Danlos, Igor Mel'čuk et Alain Polguère et les trois relecteurs pour leurs commentaires sur la version préliminaire de cet article, ainsi que Leo Wanner et Igor Mel'čuk à nouveau pour les discussions acharnées à propos de leur travail sur le même sujet.

Références

- ABEILLE A., SCHABES Y., JOSHI A. (1990) Using Lexicalized TAGs for Machine Translation, *Proceedings of COLING*, Helsinki, Finland, 1-6.
- APRESIAN J., BOGUSLAVSKY I., IOMDIN L., LAZURSKY A., SANNIKOV V., SIZOV V. & TSINMAN L. (2003) ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT, *Proceeding of MTT*, Paris, 279-288.

¹¹ La question du recours à un modèle statistique, elle, n'est pas remise en cause. Des informations sur la fréquence (et les conditions d'utilisation) des US sont nécessaires pour le meilleur ajustement des traductions.

- ARNOLD D., DES TOMBES L. (1987) Basic theory and methodology in Eurotra. In S. Nirenburg (ed.), *Machine Translation: Theoretical and Methodological Issue*, Cambridge: Cambridge University Press, 114-135.
- DING Y., PALMER M. (2005) Machine Translation Using Probabilistic Synchronous Dependency Insertion Grammars, Actes de *ACL*, Ann Arbor.
- DANLOS D., SAMVELIAN P. (1992) Translation of the predicative elements of a sentence: category switching and aspect, in *Proceedings of Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, Montréal, Canada.
- DORR B. J. (1994) Machine Translation Divergences: A Formal Description and Proposed Solution, *Computational Linguistics*, 20(4):597-633.
- DUCROT O. (1995) Les unités significatives. In Ducrot O. & J.-M. Schaeffer, *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris : Seuil, 358-365.
- GOLDBERG A. (1995) *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- KAHANE S. (2004) Grammaires d'Unification Polarisées, Actes de *TALN*, Fès, 10 p.
- KAHANE S., LAREAU F. (2005), Grammaire d'Unification Sens-Texte : modularité et polarisation, Actes de *TALN*, Dourdan, 23-32.
- KAHANE S., POLGUERE A. (2001) Formal foundations of lexical functions, Actes du *Workshop on Collocation, ACL*, Toulouse, 8 p.
- KULAGINA O.S., MEL'CUK I.A. (1997) Automatic translation: some theoretical aspects and the design of a translation system, in A.D Booth (ed.) *Machine translation*, Amsterdam: North-Holland, 131-171.
- LAVOIE B., KITTREDGE R. KORELSKY T., RAMBOW O. (2000) A Framework for MT and Multilingual NLG Systems Based on Uniform Lexico-Structural Processing, *Proceedings of ANLP/NAACL 2000*, Seattle, Washington.
- LINDOP J., TSUJII. J. (1993) Complex Transfer in MT: A Survey of Examples. Technical report, num 91, 5, Center for Computational Linguistics, Manchester, UMIST.
- MARTINET A. (1960) *Éléments de linguistique générale*, Paris.
- MEL'CUK I. (1988a) *Dependency Syntax: Theory and Practice*, SUNY.
- MEL'CUK I. (1988b) Paraphrase et lexique dans la théorie linguistique Sens-Texte — Vingt ans après, *Revue internationale de lexicologie et lexicographie*, Vol. 52/53, pp. 5-50/5-53.
- MEL'CUK I, WANNER L. (2006) Syntactic mismatches in machine translation, *Machine Translation*, 20:2, 81-138.
- NASR A., RAMBOW O., PALMER M., ROSENZWEIG J. (1997) Enriching Lexical Transfer With Cross-Linguistic Semantic Features. In *Proceedings of the AMTA/SIG-IL First Workshop on Interlingua*, San Diego, CA. Published as New Mexico State University Computing Research Laboratory technical report MCCS-97-314.
- NESSON R., SHIEBER S., RUSH A. (2006) Induction of probabilistic synchronous tree-insertion grammars for machine translation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA)*, Boston, Massachusetts.
- SCHUBERT K. (1987) *Metataxis: Contrastive Dependency Syntax for Machine Translation*. Dordrecht/Providence: Foris
- TALMY L. (1976) Semantic Causative Types. In M. Shibitani (ed.) *Syntax and Semantics 6: The Grammar of Causative Constructions*, Academic Press : N.Y, 43-116.
- TESNIERE L. (1959) *Éléments de syntaxe structurale*, Paris: Klincksieck.
- VIEGAS E. (1997) Mismatches and divergences: the continuum perspective, *Proceedings of TMI*, Santa Fe.
- ŽOLKOVSKIJ A., MEL'CUK I. (1967) O semantičeskom sinteze [Sur la synthèse sémantique (de textes)], *Problemy Kybernetiki [Problèmes de Cybernétique]*, 19, 177-238. [trad. franç. : 1970, *T.A. Information*, 2, 1-85.]