

Outils génériques et transfert hybride pour la traduction automatique sur Internet

Sébastien L'haire, Juri Mengon et Christopher Laenzlinger

Laboratoire d'Analyse et de Technologie du Langage
Université de Genève
2, rue de Candolle, 1211 Genève 4

Résumé

Dans cet article, nous décrivons un système de traduction automatique pour l'allemand, le français, l'italien et l'anglais. Nous utilisons la technique classique analyse-transfert-génération. Les phrases d'entrée sont analysées par un analyseur générique multilingue basé sur la théorie «Principes & Paramètres» de la grammaire générative chomskienne. Le mécanisme de transfert agit sur des représentations hybrides qui combinent des éléments lexicaux avec de l'information sémantique abstraite. Enfin, un générateur inspiré de la même théorie linguistique engendre des phrases de sortie correctes. Nous décrivons également brièvement les différentes interfaces envisagées sur Internet.

1. Introduction

Le projet ITS-3 a été développé au Laboratoire d'Analyse et de Technologie du Langage de l'Université de Genève. Il s'agit d'un prototype de traduction automatique en cours de développement pour le français, l'anglais, l'allemand et l'italien, basé sur la technique analyse-transfert-génération. L'analyseur IPS, inspiré de la théorie «Principes & Paramètres» (Chomsky & Lasnik, 1995), traite le texte en langue source. Il fournit des structures syntaxiques riches, qui contiennent des informations lexicales, syntagmatiques, grammaticales et thématiques. IPS est développé pour fournir une analyse profonde et robuste pour une large couverture de la langue. Ces qualités sont essentielles pour les applications génériques à domaine non restreint. Le transfert agit au niveau de représentations hybrides, appelées *structures pseudo-sémantiques* (PSS) qui combinent les éléments lexicaux à des informations sémantiques abstraites. Enfin, à partir des PSS, le générateur produit des phrases correctes. Nous allons maintenant discuter les avantages et inconvénients de divers formalismes grammaticaux. Puis nous détaillerons le fonctionnement de notre prototype à travers un exemple concret de traduction. Enfin nous décrirons comment notre traducteur sera mis à disposition sur Internet.

2. Les formalismes grammaticaux

Les grammaires basées sur des règles, telles que GPSG (Gazdar *et al.*, 1985) ou LFG (Kaplan & Bresnan, 1982), décrivent la structure de surface des phrases à l'aide de règles de réécriture syntagmatiques indépendantes du contexte. L'utilisation de ce type de règles est facilitée par la disponibilité de différents algorithmes d'analyse. En outre, elles décrivent la grammaire de manière uniforme et leurs propriétés mathématiques et computationnelles sont bien connues. Néanmoins, les grammaires basées sur des règles atteignent souvent des limites importantes lors de la maturation des systèmes informatiques. Parmi les inconvénients les plus grands figurent leur nature spécifique aux constructions et leur dépendance face à la langue en question. Pour cela, d'autres formalismes restreignent le nombre de règles par des contraintes, afin d'éviter les problèmes de dépendance de la langue et de constructions spécifiques. Ainsi, HPSG (Pollard & Sag, 1994) est une grammaire d'unification basée sur les contraintes.

La théorie de «Principes & Paramètres» (Chomsky & Lasnik, 1995) propose une approche différente: la grammaire est considérée comme un ensemble de principes interactifs de grammaticalité, qui sont indépendants de la langue (projections syntagmatiques, interface lexicque-syntaxe, attachement gauche-droite, interprétation thématique), combinés à des paramètres spécifiques à chaque langue (cliticisation du pronom en français, insertion du «do» en anglais etc.). Une des questions les plus traitées au sujet des analyseurs inspirés de la théorie chomskienne est la manière d'appliquer efficacement ces principes. Comme ces derniers sont exprimés de manière déclarative dans la théorie, leur réalisation procédurale doit préalablement être définie dans un système informatique (Berwick, 1991). Une approche naïve procéderait en deux étapes: la première consisterait à générer les structures possibles, la seconde à éliminer celles qui violent au moins un des principes. Pourtant, cette approche est inefficace et a tendance à surgénérer. Différentes solutions ont été proposées pour faire face à ces deux problèmes. Ainsi, Fong (1990) prône un ordre optimal pour l'application des différents principes, tandis que l'analyseur de Dorr (1991) utilise un modèle de co-routine: un module construit un squelette de structures syntaxiques, alors qu'un autre les modifie selon les principes et les contraintes de la théorie du Gouvernement & Liage (Chomsky, 1981). Enfin, dans le système PRICIPAR (Lin, 1993), une distinction est faite entre les constituants et leur description compacte. Au lieu de l'opération coûteuse qui consiste à traverser les structures pour appliquer les principes, un simple test est fait sur ces descriptions compactes. Ainsi, le système ne génère de nouvelles structures que si tous les principes sont respectés.

Dans l'analyseur IPS présenté dans cet article, le problème de la surgénération est traité par le module d'attachement, qui sera décrit à la section 3.1. Nous verrons que ce modèle s'est avéré efficace pour le développement d'un système de traduction multilingue à large couverture.

3. L'analyseur IPS

3.1. La modularité d'IPS

L'analyseur IPS (*Interactive Parsing System*) est basé sur la théorie de «Principes et Paramètres» et diffère des analyseurs à base de règles par son architecture. Des modules génériques implémentent les principes généraux de la grammaire et sont valables pour toutes les langues. D'autres modules traitent des propriétés spécifiques à une langue. Le défi lors de la réalisation de ce type d'analyseur consiste à parvenir à une modularisation qui tienne compte de la distinction entre modules génériques et modules spécifiques.

Au coeur du système, on trouve le module \bar{X} , qui traite de la forme générale des structures syntaxiques. Tous les constituants sont basés sur le schéma \bar{X} donné en 1.

- (1) [Spécificateur_{liste} X⁰ Complément_{liste}]_{XP}

Pour des raisons de simplicité, le niveau barre n'a pas été représenté. Spécificateur et Complément sont des listes éventuellement vides de projections maximales. X⁰ peut être soit une tête lexicale (Adv, Adj, N, V, P) ou une catégorie fonctionnelle (C, I, D, F)¹. Toutes ces têtes projettent une projection maximale XP. Ainsi, le schéma \bar{X} garantit une structure uniforme pour tous les syntagmes. Si l'on prend l'exemple de l'allemand, les objets peuvent précéder ou suivre le verbe selon leur catégorie (nominale/prépositionnelle vs. phrastique). Ainsi, elles peuvent apparaître tant dans la liste des spécificateurs que dans celle des compléments, à gauche ou à droite de la tête verbale². C'est le module d'attachement qui détermine quels compléments et quels spécificateurs peuvent être attachés à une tête spécifique. Cette procédure attache des structures selon les propriétés de sélection des têtes et les filtre d'après les contraintes d'accord.

Comme outils génériques, les modules de projection et d'attachement ont tendance à surgénérer. Pour faire face à ce problème, IPS utilise des contraintes de filtre à stratégie descendante, comme le module thématique, qui établit des relations sémantiques/thématiques entre le prédicat et ses arguments (agent, thème, but etc.), et le module d'assignation de cas, qui exige que chaque syntagme nominal lexicalisé reçoive un cas morphologique ou abstrait (nominatif, accusatif, datif etc.).

Parmi les paramètres spécifiques aux langues, nous aimerions en mentionner certains qui sont réalisés pour l'allemand: IPS applique, entre autres, la contrainte du verbe à la seconde place (V2), l'ordre objet-verbe vs. verbe-objet, le réagencement des constituants (*scrambling*), ainsi que d'autres contraintes sur des constructions typiques de l'allemand (extraposition des relatives, *infinitivus pro participio*, possibilités de coordination etc.).

3.2. Un exemple d'analyse

A titre d'exemple, voyons le processus d'analyse de la phrase suivante:

- (2) Dann hatte Hans es dieser Frau geschenkt.
 Ensuite a Hans le à cette femme offert
 'Ensuite Hans l'a offert à cette femme.'

L'analyseur IPS procède tout d'abord à une analyse lexicale en segmentant la phrase d'entrée en unités lexicales et en les insérant comme des arcs dans un graphe pour chaque lecture possible (Fig. 1, étape 1). A chaque unité lexicale dans le graphe correspond une projection maximale. Toutes les caractéristiques lexicales sont ainsi transférées au niveau syntaxique (XP) en application du Principe de projection (Fig. 1, étape 2) qui exige que l'information lexicale soit représentée à tous les niveaux de la syntaxe. Ensuite, le module d'attachement vérifie tous

1. Pour représenter les constituants, nous utilisons les abréviations suivantes: Adv(erbe), Adj(ectif), N(om), V(erbe), Prep(osition), D(éterminant), C(omplémenteur), I(nflection) et F(onctionnel).

2. Les listes des Spécificateurs et des Compléments n'ont donc pas une fonction d'interprétation en tant que telles. Ainsi, un complément de verbe peut apparaître dans la liste des spécificateurs, tandis qu'un ajout (donc un élément qui agit comme un spécificateur) peut être placé dans la liste des compléments.

les attachements possibles en position Spec ou Compl des diverses projections XP^3 d'après les règles de sélection et les relations thématiques. Le squelette de la structure des propositions est $CP \prec IP \prec VP$, comme illustré à la Fig. 1, étape 3. IP est le complément à droite de CP et VP est le complément droit de IP. Dans les phrases principales, le premier constituant, ici l'AdvP *dann* («ensuite»), occupe la première position en Spec-CP. Le verbe conjugué est déplacé en C^0 , la deuxième position dans la proposition principale. C'est une manière simple d'appliquer la contrainte V2. Comme les auxiliaires conjugués sont générés en position de base I^0 , le module de formation de chaîne lie l'auxiliaire déplacé en C^0 avec sa trace en I^0 (Fig. 1, étape 4). Dans notre analyse, le DP sujet *Hans* est attaché en Spec-VP⁴. La configuration OV en allemand provient de l'attachement des DP⁵ compléments *es* («le») et *seiner Frau* («à sa femme») en Spec-VP. Le participe passé *geschenkt* («offert») occupe la position V^0 . A cette étape, les modules interprétatifs de filtrage vérifient si chaque argument de la clause reçoit un rôle thématique du verbe (module thématique), et que tout syntagme nominal est dans une position correcte pour recevoir un cas (module de cas, Fig. 1, étape 5). La structure de la phrase est donnée à la Fig. 1, étape 3.

3.3. Caractéristiques du système IPS

Outre la modularité et la généricité, le système IPS présente deux autres caractéristiques majeures : la robustesse et l'utilisation d'analyses riches et profondes. Un système doit être robuste pour être exploitable dans des applications indépendantes du domaine et à large couverture. Dans ce but, l'analyseur IPS peut traiter les mots inconnus et comporte des «micro-grammaires» pour reconnaître les expressions idiomatiques, les expressions parenthétiques et temporelles, etc. La robustesse est améliorée par une stratégie dite «sans échec» selon laquelle les analyses incomplètes sont encore exploitables par des résultats partiels. De plus, les structures profondes et riches permettent d'utiliser l'analyseur dans de nombreuses applications multilingues et indépendantes du domaine, notamment en traduction automatique. Cette approche du traitement du langage est très éloignée de l'approche assez répandue adoptée par les analyseurs «superficiels» (*shallow parsers*) ou les identificateurs de syntagmes nominaux, qui consiste à utiliser une sorte de pré-traitement linguistique adapté à des applications spécifiques.

4. Le projet de traduction ITS-3

Le projet ITS-3 (*Interactive Translation System*) vise à développer un outil de traduction en utilisant des structures d'interface abstraites. Avant de présenter le prototype, nous allons tout d'abord aborder la question des techniques de transfert en traduction automatique.

3. IPS est un analyseur à stratégie ascendante qui poursuit toutes les alternatives en parallèle. Sa stratégie est appelée «attachement au coin droit» (Wehrli, 1997). Une description plus complète de la stratégie d'IPS peut être trouvée dans cet ouvrage.

4. Contrairement aux analyses proposées pour l'anglais et le français, le sujet n'occupe pas la position Spec-IP en allemand. Ceci est dû au phénomène du «brassage» (*scrambling*), grâce auquel l'ordre des mots pour le sujet, les objets et les ajouts éventuels est relativement libre en allemand. La position Spec-VP nous sert donc de domaine de réagencement à l'intérieur duquel ces composants sont générés à la base.

5. Nous suivons l'hypothèse d'Abney (1987) pour les DP, selon laquelle le déterminant est la tête des expressions nominales et prend un NP comme complément.

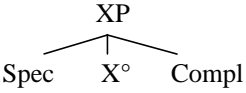
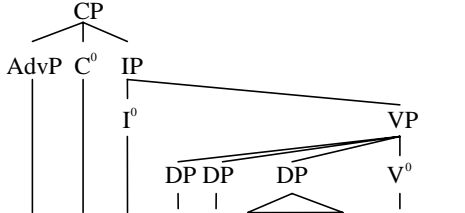
Étapes	Processus
1	Analyse lexicale: [dann _{Adv}] [hatte _v] [Hans _D] [es _D] [dieser _D] [Frau _N] [geschenkt _v]
2	Projection au niveau syntaxique: Projeter une projection maximale XP pour chaque élément dans le graphe de l'analyse lexicale  où $X^\circ \in \{N, V, A, P, D, Adv, C, I\}$
3	Vérification des attachements:  <i>Dann hatte_v t_i Hans es dieser Frau geschenkt.</i>
4	Formation de chaînes: (hatte, t) _i
5	Interprétation des arguments: Arg 1: <i>Hans</i> CAS = nominatif; ROLE-θ = agent Arg 2: <i>es</i> CAS = accusatif; ROLE-θ = thème Arg 3: <i>dieser Frau</i> CAS = datif; ROLE-θ = bénéficiaire

FIG. 1 – Le processus d'analyse consiste en 5 étapes principales: analyse lexicale, projection au niveau syntaxique, vérification des attachements, formation de chaînes et interprétation des arguments.

4.1. Traduction automatique et techniques de transfert

Dans le domaine de la traduction automatique, le transfert lexical et syntaxique est actuellement la technique la plus utilisée. Parmi les systèmes les plus connus qui utilisent cette technique, on peut citer Systran (Ryan, 1989) et Ariane (Boitet, 1989). Le transfert lexico-structural est à mi-chemin entre la traduction mot-à-mot et la technique de l'interlangue, comme nous pouvons le voir à la Fig. 2 ou chez Hutchins & Somers (1992). La procédure de traduction se déroule comme suit : dans un premier temps, une phrase d'entrée est traitée par un analyseur syntaxique. Une structure fonctionnelle est ensuite dérivée de cette analyse avec un degré d'abstraction variable. De telles structures intermédiaires sont aussi appelées «structures d'interface». A l'étape suivante, la composante de transfert dérive une structure fonctionnelle correspondante pour la langue cible à travers un transfert lexical et des adaptations de structure (Tucker, 1987). Enfin, le générateur produit une phrase grammaticale en langue cible à partir de la structure d'interface. Le transfert lexico-structural présente trois inconvénients majeurs, que nous avons déjà rencontrés dans une version antérieure de notre traducteur, le système ITS-2 (Ramluckun & Wehrli, 1993). Tout d'abord, il demande un lexique très étendu qui contient de multiples correspondances lexicales associées à des informations syntaxiques, sémantiques et contextuelles. Ensuite, il est nécessaire d'utiliser une analyse sémantique et pragmatique au-delà de la phrase pour obtenir des résultats de traduction satisfaisants. Enfin, le transfert lexico-structural est largement dépendant de la langue. Ainsi, un système de traduction multilingue qui traiterait par exemple 9 langues nécessiterait un module de transfert pour chaque paire de langue, ce qui fait

72 modules en tout. Un tel système deviendrait très vite ingérable.

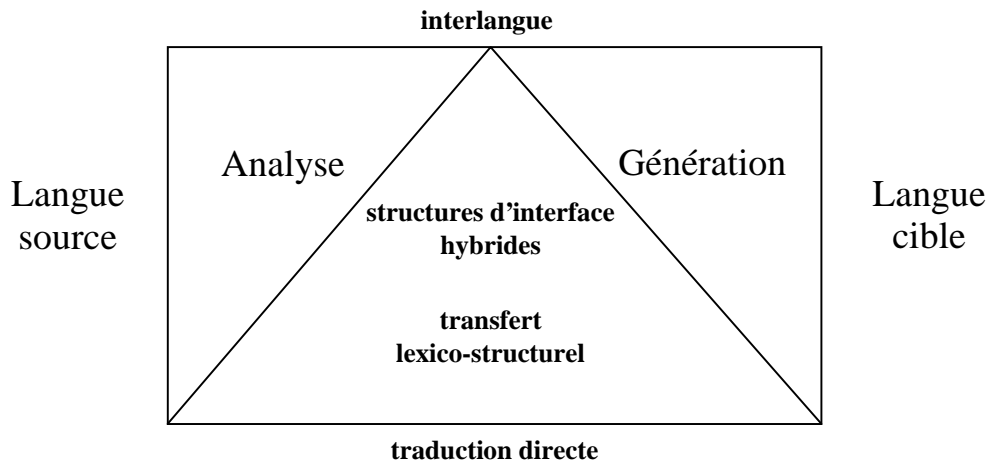


FIG. 2 – Pyramide représentant le degré d'abstraction des différentes techniques de traduction.

L'approche de l'interlangue est située au sommet des méthodologies schématisée à la fig. 2 ou chez Hutchins & Somers (1992). L'interlangue consiste en des représentations abstraites d'entités et de fonctions linguistiques qui fonctionnent comme une "troisième langue" qui relie la langue source à la langue cible. Le principal avantage de l'interlangue est d'être indépendante des langues. Ainsi, elle peut être utilisée dans des systèmes multilingues qui traitent des paires de langues typologiquement distinctes. De plus, l'interlangue peut encoder des informations contextuelles et extralinguistiques comme l'inférence, l'anaphore ou la présupposition. Néanmoins, la méthode de l'interlangue a de sérieux inconvénients : la complexité et l'abstraction de ses représentations linguistiques. Il est actuellement impossible de construire un ensemble de représentations abstraites universelles pour tous les substantifs (notations) et tous les verbes (événements) qui restent valables pour toutes les langues. A cause de la complexité et de la lourdeur de représentations linguistiques et contextuelles totalement abstraites, cette approche de la traduction automatique est très difficile, voire impossible à utiliser pour une application à large couverture. Pourtant, des efforts considérables sont actuellement effectués dans le projet UNL, dont le but est d'atteindre une interlangue indépendante du domaine d'application.

Comme solution, on peut faire appel à des structures d'interface hybrides, combinant un transfert lexical restreint à des représentations sémantiques abstraites. Ce type de composante d'interface a déjà été utilisé de manière variée dans le domaine de la traduction automatique. Par exemple, citons les structures d'interface utilisées dans le *Core Language Engine* (Alshawi, 1991), qui traduit des phrases anglaises dans des représentations formelles appelées formules quasi-logiques (*quasi-logical forms*). Ces représentations constituent des formules logiques dépendantes du contexte capables d'interprétation extra-linguistique (inférence, anaphore, présupposition etc.) et adaptées à une application trans-linguistique en traduction automatique. A la section suivante, nous présentons un autre type de représentations hybrides appelées structures pseudo-sémantiques (*pseudo-semantic structures*).

4.2. Traduction à l'aide de structures pseudo-sémantiques

Le prototype ITS-3 est un outil de traduction qui se sert des structures d'interface abstraites appelées structures pseudo-sémantiques (*pseudo-semantic structures*, PSS) (Etchegoyhen & Wehrli, 1998). Les PSS ont une nature hybride car elles combinent des représentations

sémantiques abstraites avec des éléments lexicaux et constituent les données d'entrée pour le générateur syntaxique GBGen (Etchegoyhen & Wehrle, 1998).

Pour comprendre comment la procédure globale de traduction fonctionne, nous allons poursuivre la traduction en français de l'exemple allemand donné en 3.2. La première étape après l'analyse syntaxique (Fig. 1, étape 3) est le transfert de ce résultat dans une PSS d'interface. Une PSS contient des informations sur la proposition, à savoir son mode (réel ou irréel), son temps et des spécifications d'aspect. Pour déterminer les informations sur le temps dans les PSS, nous utilisons des relations de continuum entre les valeurs de temps de l'événement (E) et temps d'énonciation (S)⁶ de Reichenbach (1947), mêlées à des spécifications aspectuelles (progressif, perfectif, imperfectif). On spécifie également des informations sur la voix, la négation et le type d'énonciation de la proposition. Comme les PSS impliquent un transfert lexical restreint aux classes lexicales ouvertes (verbe, substantif, adjectifs et adverbes), le prédicat est spécifié sous la forme d'un élément lexical. Chaque PSS peut avoir un certain nombre de satellites qui dépendent d'elle. Ainsi, les arguments non phrastiques sont représentés par des structures DP (*DP-Structures*, DPS), de type opérateur-propriété, qui correspondent à la relation syntaxique déterminant-substantif. Les AdvP sont représentés dans des «structures de caractéristique» (*Characteristics Structures*, CHS). Ainsi, pour la phrase 3.2, une PSS est dérivée comme à la Fig. 3.

La représentation sémantique donnée dans la figure 3 constitue la structure d'entrée du générateur GBGen. Selon ces informations, la phrase de sortie sera une phrase déclarative active affirmative. Le temps correspondra à l'indicatif plus-que-parfait. Le prédicat verbal «offrir» prend trois arguments et un modifieur temporel portant sur la phrase. Comme l'argument externe (l'agent) est généré comme sujet de la phrase, il sera réalisé en DP attaché en Spec-IP. Le second argument est un pronom personnel en objet direct, qui devra être cliticisé à l'auxiliaire en I⁰ et sera lié par une trace à sa position de base en Compl-VP. Le troisième argument sera réalisé comme un objet indirect «datif» avec la préposition sous-catégorisée *à* et sera exprimé par le PP *à cette femme*. Ce complément prépositionnel sera attaché en Compl-VP. Enfin, le quatrième satellite sera généré syntaxiquement comme un AdvP attaché en Spec-VP. La phrase et la structure qui en résultent sont données en Fig. 4.

5. Interfaces d'ITS-3 sur Internet

Internet se développe à toute vitesse et fournit des documents dans une multitude de langues. L'information recherchée, ou la meilleure information, n'est pas forcément dans une langue que l'on comprend ou que l'on maîtrise suffisamment. C'est pourquoi la traduction sur Internet est une des applications les plus recherchées de la traduction automatique. Le système le plus connu sur Internet est Systran⁷ (Ryan, 1989), mais il existe aussi d'autres traducteurs comme Reverso⁸. A cause de sa typologie particulière et de la difficulté de traitement, l'allemand est une langue peu traitée par les logiciels de traduction automatique. A notre connaissance, seuls les deux systèmes cités ci-dessus proposent une traduction allemand-français. Notre prototype sera disponible sur Internet à travers trois interfaces :

- un formulaire où l'utilisateur tape une phrase ;

6. Les valeurs possibles sont E<S, E=S et E>S.

7. Systran est disponible sur Internet à l'adresse <http://babelfish.altavista.digital.com/>.

8. Développé par Softissimo, <http://proto.softissimo.com/reverso/asp/textonly/default.asp>.

Informations sur la proposition:

PSS	
Mode	: réel (= indicatif)
Temps	: E<S (= passé)
Aspect	: (non progressif, perfectif)
Voix	: actif
Négation	: pas de négation
Type d'énonciation	: déclaration
Prédicat	: <i>schchenken</i> ⇔ <i>offrir</i>

Information sur les satellites:

1.) DPS		2.) DPS	
Rôle thématique	: agent	Rôle thématique	: thème
Propriété	: <i>Hans</i>	Propriété	: delta
Opérateur	: delta	Opérateur	: delta
		Genre	: neutre
		Personne	: 3 ^{ème} personne
		Nombre	: singulier
3.) DPS		4.) CHS	
Rôle thématique	: bénéficiaire	Valeur	: quand
Propriété	: <i>Frau</i> ⇔ <i>femme</i>	Portée	: phrase
Opérateur	: démonstratif	Caractéristique	: <i>dann</i> ⇔ <i>ensuite</i>
Nombre	: singulier		

FIG. 3 – Structure pseudo-sémantique : 'delta' indique que les valeurs assignées ici sont laissées non spécifiées ou sous-spécifiées.

- la traduction d'une page web avec restitution de la mise en page originale ;
- la traduction de courriers électroniques.

La traduction de pages web et la restitution de leur mise en page est loin d'être triviale. En effet, un lien hypertexte est souvent placé sur un mot ou un groupe de mots. Pourtant, leur position peut varier d'une langue à l'autre comme le montre l'exemple 3a et sa traduction en 3b, dans lesquels l'ordre des mots se révèle bien différent. D'ailleurs, le même problème se pose pour les mots en italiques ou en gras.

- (3)a. Heute kann der Programmierer seine Arbeit beginnen.
- b. Le programmeur peut commencer son travail aujourd'hui.

Dans notre système, un automate à états finis traite la page à traduire. Il convertit les caractères accentués et fournit des phrases ou des segments de phrase au traducteur. Les balises HTML sont stockées dans des listes et réécrites en sortie. Une liste gère en particulier les balises devant être insérées dans une phrase. Pour cela, les positions des tags qui portent sur des parties de phrase, comme les liens et les mises en forme (gras, italiques etc.) sont recalculés pour les phrases en langue cible.

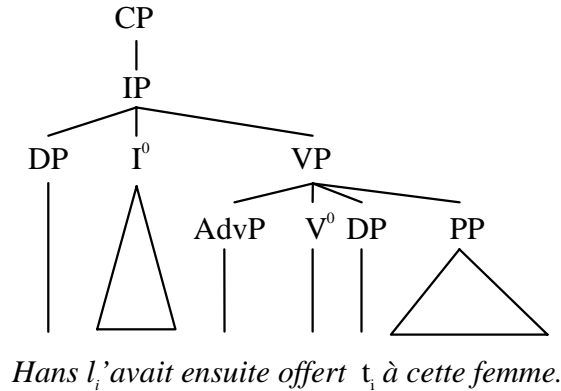


FIG. 4 – Structure syntaxique de la phrase en langue cible générée par le système GBGen

Le traitement des courriers électroniques est plus simple. Les caractères accentués doivent également être traduits et quelques caractères doivent être éliminés, comme le signe > qui signale en début de ligne d'une réponse que l'auteur cite le message original. L'utilisateur qui désire traduire un message devra simplement le transmettre à une adresse électronique qui correspond à la paire de langue souhaitée.

6. Conclusion

Nous avons décrit un système de traduction automatique basé sur l'architecture classique analyse-transfert-génération. Nous avons illustré son fonctionnement à travers un exemple avec l'allemand en langue source et le français en langue cible. Sur la base de la théorie des Principes & Paramètres, un analyseur générique et modulaire a été développé pour des applications multilingues de traitement automatique du langage naturel. Cet analyseur fournit des structures syntaxiques riches. A partir des résultats de l'analyse, une composante de transfert dérive des représentations lexico-sémantiques hybrides, appelées structures pseudo-sémantiques (*pseudo-semantic structures*, PSS), qui combinent un transfert lexical avec des informations fonctionnelles et sémantiques abstraites. Cette technique de transfert mixte tire parti à la fois de la simplicité de la procédure de transfert lexical et de l'abstraction de l'approche de l'interlangue. De plus, les PSS sont particulièrement adaptées aux informations linguistiques contenues dans les analyses syntaxiques profondes fournies par l'analyseur. En dernier lieu, le module de génération prend les PSS en entrée et restitue des phrases correctes en sortie. Différentes interfaces proposées sur Internet répondent aux besoins des utilisateurs.

7. Remerciements

Nous remercions à Eric Wehrli, Paola Merlo et Thierry Etchegoyhen pour leur précieux commentaires et leur discussions utiles de cet article. Cette recherche est soutenue par un subside du Fonds National suisse pour la Recherche Scientifique (n° 1214-053792.98/1).

Références

- ABNEY S. (1987). *The English Noun Phrase in its Sentential Aspect*. Ph. D. Thesis. Cambridge Mass.: MIT Press.
- ALSHAWI H. (1991). *The Core Language Engine*. Cambridge, Mass.: MIT Press.
- BERWICK R. (1991). Principles of principle-based. In R. BERWICK, S. ABNEY & C. TENNY, Eds., *Principle-based Parsing: Computation and Psycholinguistics*, p. 1–37. Dordrecht: Kluwer Academic Press.
- BOITET C. (1989). Geta project. In M. NAGAO, H. TANAKA & T. MAKINO, Eds., *Machine Translation Summit*, p. 54–65. Tokyo: Ohmsha, Ltd.
- CHOMSKY N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris Publications.
- CHOMSKY N. & LASNIK H. (1995). The theory of principles and parameters. In N. CHOMSKY, Ed., *The Minimalist Programm*, p. 13–127. Cambridge, Mass.: MIT Press.
- DORR B. (1991). Principle-based parsing for machine translation. In R. BERWICK, S. ABNEY & C. TENNY, Eds., *Principle-based Parsing: Computation and Psycholinguistics*, p. 153–183. Dordrecht: Kluwer Academic Press.
- ETCHEGOYHEN T. & WEHRLE T. (1998). Overview of gbgen: A large-scale, domain independent syntactic generator. In *Proceedings of the 9th International Workshop on Natural Language Generation*, p. 288–291. Niagara-on-the-Lake: Association for Computational Linguistics.
- ETCHEGOYHEN T. & WEHRLI E. (1998). Traduction automatique et structures d'interface. In *5^e Conférence sur le Traitement Automatique du Langage Naturel*, p. 2–11. Paris: ATALA.
- FONG S. (1990). *Computational Properties of Principle-based Grammatical Theories*. Ph. D. Thesis. Cambridge, Mass.: MIT Press.
- GAZDAR G., KLEIN E., PULLUM G. & SAG I. (1985). *Generalized Phrase Structure Grammar*. Oxford: Blackwell.
- HUTCHINS W. J. & SOMERS H. L. (1992). *An Introduction to Machine Translation*. Cambridge: Academic Press Ltd.
- KAPLAN R. M. & BRESNAN J. (1982). Lexical functional grammar : A formal system for grammatical representation. In J. BRESNAN, Ed., *The Mental Representation of Grammatical Relations*, p. 173–281. Cambridge, Mass.: MIT Press.
- LIN D. (1993). Principle-based parsing without overgeneration. In *Proceedings of the 31st Annual Meeting of ACL*, p. 112–120. Columbus, Ohio: Ohio State University.
- POLLARD C. & SAG I. (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University Press of Chicago.
- RAMLUCKUN M. & WEHRLI E. (1993). Its-2: An interactive personal translation system. In *Proceedings of the sixth Conference of EACL*, p. 476. Utrecht: OTS-Research Institute for Language and Speech, University of Utrecht.
- REICHENBACH H. (1947). *Elements of Symbolic Logic*. New York: Free Press.
- RYAN J. P. (1989). Systran: A machine translation system to meet user needs. In M. NAGAO, H. TANAKA & T. MAKINO, Eds., *Machine Translation Summit*, p. 116–121. Tokyo: Ohmsha, Ltd.
- TUCKER A. B. (1987). Current strategies in machine translation research and development. In S. NIRENBURG, Ed., *Machine Translation. Theoretical and Methodological Issues*, p. 22–41. Cambridge: Cambridge University Press.
- WEHRLI E. (1997). *L'analyse syntaxique des langues naturelles. Problèmes et méthodes*. Paris: Masson.