

BULLETIN D'INFORMATIQUE APPROFONDIE ET APPLICATIONS

N° 37 MARS 1994

COMITE SCIENTIFIQUE

France Chappaz
M'hamed Charifi
Roger Cusin
Bernard Goossens
Patrick Isoardi
Jean - Philippe Lehmann
Nadia Mesli
Patrick Sanchez
Rolland Stutzmann

DIRECTEUR

Edmond Bianco

REDACTEUR EN CHEF

Jean - Michel Knippel

REDACTEUR ADJOINT

Sami Hilala

1 EDITORIAL,

par Edmond Bianco

3 LE FORMALISME CAT2,

par Nadia Mesli

**37 METHODOLOGIE DE DEVELOPPEMENT
DES SYSTEMES EXPERTS,**

par Lovska Popova

39 VOZZAVEDIBISAR,

par Jean - Michel Knippel

Publication gratuite trimestrielle de l'Université d'Aix - Marseille II
58, boulevard Charles Livon. F - 13284 Marseille Cedex 07
Téléphone : (33) 91 39 65 00 Télécopie : (33) 91 31 31 36

Edition 1996

ISSN 0291 - 5413

COMITE SCIENTIFIQUE

*France Chappaz
M'hamed Charifi
Roger Cusin
Bernard Goossens
Patrick Isoardi
Jean - Philippe Lehmann
Nadia Mesli
Patrick Sanchez
Rolland Stutzmann*

DIRECTEUR

Edmond Bianco

REDACTEUR EN CHEF

Jean - Michel Knippel

1 EDITORIAL,

par Edmond Bianco

REDACTEUR ADJOINT

Sami Hilala

3 LE FORMALISME CAT2,

par Nadia Mesli

**37 METHODOLOGIE DE DEVELOPPEMENT
DES SYSTEMES EXPERTS,**

par Lovska Popova

39 VOZZAVEDIBISAR,

par Jean - Michel Knippel

Publication gratuite trimestrielle de l'Université d'Aix - Marseille II
58, boulevard Charles Livon. F - 13284 Marseille Cedex 07
Téléphone : (33) 91 39 65 00 Télécopie : (33) 91 31 31 36

EDITORIAL, INTERNET.

Comment ? Je n'ai pas encore parlé d'INTERNET? Vraiment ? Ca alors!

Il est de ces éclairs lumineux tellement puissants, tellement réels, qu'ils vous éblouissent brutalement et vous laissent pantois. Voila, je suis pantois.

Désormais tout le monde va pouvoir - à bas prix, promis juré - communiquer avec tout le monde. **Co-mmu-ni-quer**. Sésame du troisième millénaire. Le troisième millénaire ne sera pas mystique comme le prétendait Malraux mais il sera communicant. Quand je pense Internet, il faut bien se faire une image, j'ignore pourquoi celle qui me monte à l'esprit ressemble à celle qu'on nous donne quelquefois du Palais Brongniart en pleine effervescence. Une masse compacte de gens hurlant, s'agitant, brandissant, brassant, jetant on ne sait quoi, éructant, interpellant, apostrophant on ne sait qui, en apparence vain tumulte dans un vacarme de gare les jours de grand départ, l'ambiance de fête en moins.

Les choses nouvelles ne le sont jamais totalement en ce sens que tant du point de vue de la conception que de celui de la perception, il faut bien partir de ce qui est connu. La simple recombinaison de choses banales peut être parfois d'un impact saisissant pour un esprit profane. L'effet de choc est d'ailleurs amplifié par les médias dont les hérauts, tout aussi profanes que le vulgum pecus, n'hésitent pas à claironner leur surprise et exulter leur enthousiasme. Ce qui ne manque pas de frapper les imaginations non averties. On ne compte plus le nombre de "révolutions" toutes plus fantastiques les unes que les autres que nous vécûmes depuis l'avènement de la télévision.

La concurrence des chaînes, en particulier, privées et publiques qui dansent sur l'audimat, comme les fakirs dansent sur le feu, pousse à toujours plus de sensationnel. Dénrée censée attirer le spectateur comme la viande avariée les fourmies ou les mouches. Pour parodier César, (celui de Pagnol), on voit très bien les grands manitous de ces chaînes serrant chacun les bornes du couillonnisme dans leurs bras et courant à toute vitesse pour agrandir leur domaine.

Parfois ce genre d'attitude donne des résultats que les amateurs apprécient. Ce fut le cas du charnier de Timisoara. Les Croque-nouvelles avaient été appâtées au sang.

Internet ?

On ouvre des yeux comme des soucoupes devant le prodige, jadis il fallait compter sur les sorciers pour entrer en relation à distance, maintenant on peut faire mieux avec son clavier. Bien sûr le petit côté mystérieux de l'irrationnel, lui aussi très porteur médiatique, en prend un coup, mais quelle victoire de la technique, on va pouvoir faire défiler sous ses yeux ébahis tout. Oui, tout. Tout ce qui est produit dans le monde de textes, d'images, de son, de bruits, de voyages sur terre et hors la terre... Tout on vous dit !

Et gourmandise suprême, chacun va pouvoir en remettre. Même le courrier va pouvoir se faire par ce canal. La Poste qui avait tout prévu, avait déjà commencé la distribution des prospectus, bientôt il ne lui restera plus que ça. Inutile de se déplacer vers les salons de peinture, toute l'exposition est là sur votre écran, vous n'avez qu'à faire défiler. Bien sûr on perd les petits gâteaux et le champagne des vernissages, mais on ne craint plus les grèves du métro. Les bistros eux-mêmes deviennent cyber. Fini l'assommoir au gros rouge. La solitude comblée au pichtogorn est remplacée par l'envolée dans les étoiles de la communication.

Quand le nuage de Tchernobyl nous est passé sur la tête, nos "politiciens" nous ont menti, sans doute n'étaient ils pas très certains que nos centrales valaient mieux. Pognon d'abord. Quand les vaches sont devenues folles, à peu près à la même époque, nos "politiciens" nous ont encore menti, il a fallu que la catastrophe prenne une dimension planétaire pour que l'information diffuse. Pognon d'abord. Après ces gens-là ont le front de s'étonner qu'on ne les croie pas quand ils viennent pérorer et plastronner au fenestron.

Il me reste à souhaiter que ceux qui ont des choses intéressantes à dire puissent les publier sur Internet. Que ceux qui savent qu'une centrale va exploser puissent avertir tout le monde, ou bien ceux qui se sont aperçu qu'on nourrissait des herbivores avec des cadavres avariés puissent protester devant tout le monde, que le citoyen soit enfin considéré comme un citoyen et non plus comme un minus à qui on raconte des fables pour pouvoir mieux lui soutirer son fric.

On peut toujours rêver ...

Edmond Bianco

Le formalisme CAT2. Un exemple
d'implémentation : les constructions à
verbe support

Nadia MESLI
IAI
Martin-Luther-Straße 14
D-6600 Saarbrücken 3
Allemagne

Le formalisme CAT2. Un exemple d'implémentation : les constructions à verbe support

Résumé

Très peu de systèmes de traduction automatique proposent un traitement particulier des constructions à verbe support du type "faire du sport", "donner un conseil" ou "prendre une décision". Le choix de la stratégie d'implémentation à adopter dans ce domaine n'a pas seulement un intérêt théorique. C'est celui-ci qui détermine le caractère "économique" ou non "économique" d'un système de traduction automatique, de même que son degré de finesse au niveau de l'analyse sémantique. Le système de traduction automatique utilisé à l'IAI de Sarrebruck dans le cadre du projet EUROTRA de la Communauté Economique Européenne peut être considéré comme l'un de ces systèmes. Le formalisme CAT2 dans lequel est effectuée l'analyse automatique des constructions à verbe support proposée par l'IAI est un descendant direct du prototype original d'EUROTRA. Implémenté en PROLOG, il se base sur le principe de l'unification. Tout comme le formalisme utilisé dans EUROTRA, le formalisme CAT2 est conçu pour des systèmes de traduction stratificationnels composés de générateurs et de translateurs assurant l'analyse, le transfert ainsi que la génération de phrases. Cet article montre ainsi l'intérêt, dans un système de traduction automatique, de l'analyse particulière des constructions à verbe support. Il présente le formalisme CAT2 ainsi qu'un exemple d'implémentation d'un phénomène linguistique : l'analyse automatique de constructions à verbe support ainsi qu'un système de traduction français-allemand de telles constructions.

Mots-clés : formalisme, traduction automatique, traitement informatique du langage naturel, constructions à verbe support.

LE FORMALISME CAT2. UN EXEMPLE D'IMPLEMENTATION : LES CONSTRUCTIONS A VERBE SUPPORT

1. Le formalisme CAT2

1.1. Présentation générale

Le formalisme CAT2, conçu et implémenté à l'IAI de Sarrebruck, se réfère au système de traduction développé dans le cadre du projet EUROTRA de la Communauté Economique Européenne. Démarré en 1982, le projet de recherche et de développement EUROTRA forme en effet un système de traduction automatique mettant en jeu les neuf langues officielles de la C.E.E. : le français, l'italien, l'espagnol, le portugais, l'anglais, l'allemand, le néerlandais, le danois et le grec. Le formalisme CAT2 s'inscrit ainsi dans le cadre des formalismes développés dans ce projet. Il est un descendant direct du prototype original d'EUROTRA basé sur les trois primitives suivantes : les "Constructeurs", les "Atomes" et les "Translateurs", et appelé de ce fait "<C,A>,T Framework" (D.J. Arnold et al. [1985, 1986]).

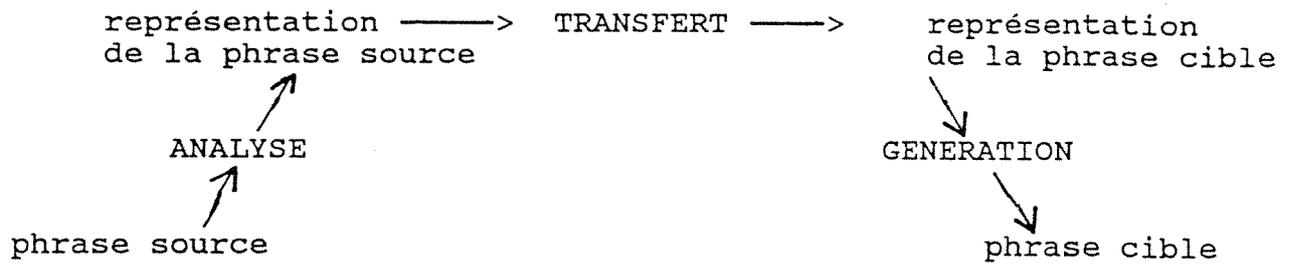
Tout comme les modèles d'analyse proposés dans EUROTRA, les modèles proposés dans le formalisme CAT2 se basent sur des critères très fortement liés à la nature du système de traduction EUROTRA, que nous nous proposons d'exposer ici brièvement.

1.1.1. Le modèle de traduction à transfert EUROTRA

Il existe principalement deux modèles de traduction automatique : 1. le modèle à *transfert* qui comporte une représentation de la phrase source (phrase à traduire) et un module de transfert qui transforme cette représentation en une représentation de la phrase cible (phrase traduite), 2. le modèle à *langage pivot* qui ne comporte qu'une représentation intermédiaire partagée par la phrase source et par la phrase cible (cf. L. Danlos [1989]). Le système EUROTRA auquel se réfère le formalisme CAT2 forme un système de traduction à transfert.

Un modèle à transfert est composé de trois modules :

- un module d'analyse qui produit une représentation de la phrase source,
- un module de transfert qui transforme la représentation de la phrase source en une représentation de la phrase cible,
- un module de génération qui traduit la représentation de la phrase cible en une phrase cible.



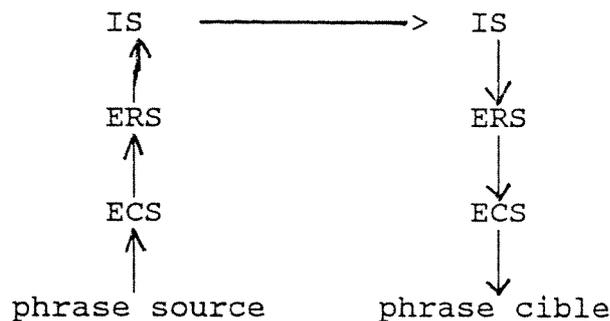
EUROTRA est donc composé de neuf modules d'analyse, de neuf modules de génération et de 72 modules de transfert. Chaque groupe national est chargé de l'analyse et de la génération de sa langue ainsi que du transfert des autres langues vers sa langue. Le nombre important de modules de transfert (72) impose la stratégie dite de "transfert simple" (ou *transfert lexical*). La représentation de la phrase cible doit être aussi similaire que possible à la représentation de la phrase source, la seule différence se situant au niveau du vocabulaire : le vocabulaire dans la représentation cible correspond alors simplement à la traduction du vocabulaire de la représentation source. Si ces représentations sont au contraire beaucoup trop éloignées l'une de l'autre (cas de non-isomorphie des représentations de la langue source et de la langue cible), le module de transfert sera important et fortement dépendant du couple de langues concernées. On parle dans ce cas-là de *transfert complexe* ou *structural* (changement de structure).

Le système de traduction EUROTRA est un système *stratificationnel*. L'analyse d'une phrase source se fait en effet en trois étapes ou à travers trois niveaux incluant un dictionnaire et un analyseur :

- Le niveau 1 ou niveau "ECS" (*EUROTRA Configurational Structure*) représente l'analyse *syntaxique*. Les mots de la phrase reçoivent la (ou leur) catégorie de discours (nom, verbe, déterminant, adjectif...) et sont regroupés en constituants (groupe nominal, verbal, prépositionnel...).
- Le niveau 2 ou niveau "ERS" (*EUROTRA Relational Structure*) représente l'analyse *relationnelle*. Les constituants ECS sont réorganisés selon leur fonction syntaxique : verbe, sujet, complément, modifieur.
- Le niveau 3 ou niveau "IS" (*Interface Structure*) représente l'analyse *sémantique*. Les fonctions ERS sont réorganisées ici selon leur rôle sémantique : prédicat (appelé "gouverneur"), argument ou modifieur.

Pour une langue donnée, les analyseurs et les dictionnaires de chaque niveau sont partagés par l'analyse et la génération. Le résultat de chaque niveau d'analyse est un arbre ou plusieurs arbres lorsqu'il y a ambiguïté. Le passage d'un niveau au suivant (ECS => ERS,

ERS => IS) est effectué au moyen de règles de translation et consiste en une modification d'arbres. La génération s'opère, elle, en effectuant les étapes de l'analyse en ordre inverse. La différence entre l'analyse et la génération d'une langue s'observe ainsi seulement dans les translateurs : la translation IS => ERS restitue les fonctions syntaxiques supprimées au niveau IS et la translation ERS => ECS défait l'ordre canonique des compléments imposé au niveau ERS. Le processus de traduction est donc une suite de transitions entre différents niveaux de représentation, ainsi que le montre le schéma ci-dessous :



Le système à transfert EUROTRA repose ainsi sur deux hypothèses fondamentales :

- l'hypothèse selon laquelle les phrases sont des *projections* de verbe : le verbe détermine en tant que "gouverneur" le nombre, la forme syntaxique ainsi que la nature sémantique des actants,
- l'hypothèse de la *compositionnalité* : la traduction d'une expression (ou phrase) correspond à la traduction de ses éléments constituants. Dans les cas de transfert simple où l'arbre IS de la phrase cible est isomorphe à l'arbre IS de la phrase source, les feuilles de l'arbre cible correspondent aux traductions des feuilles de l'arbre source.

Les analyseurs, les dictionnaires ainsi que les translateurs sont écrits dans ce système de traduction dans un même formalisme. Le formalisme utilisé par tous les groupes nationaux qui effectuent le travail linguistique est appelé "Engineering Framework" et est écrit en C-PROLOG par une équipe de la Commission. Il est de style déclaratif et est basé sur les grammaires d'unification.

1.1.2. Particularités du formalisme

Tout comme le formalisme "Engineering Framework" utilisé dans EUROTRA, le formalisme CAT2 est conçu pour des systèmes de traduction stratificationnels composés de générateurs et de translateurs assurant l'analyse, le transfert ainsi que la génération de phrases. Une grammaire consiste ainsi en CAT2 en un certain nombre de générateurs qui décrivent plusieurs niveaux d'analyse et de translateurs qui assurent le passage d'un niveau d'analyse à un autre, chaque niveau d'analyse représentant un ensemble de règles décrivant la

structure des phrases bien formées d'un langage. Les phrases d'un langage sont représentées dans un tel système par des structures arborescentes, appelées "objets".

Implémenté en PROLOG, il se base sur le principe de l'*unification* (unification de termes PROLOG et unification de graphes). Le formalisme CAT2 est très proche des formalismes LFG (*Lexical Functional Grammar*), FUG (*Functional Unification Grammar*), GPSG (*Generalized Phrase Structure Grammar*) ou HPSG (*Head-driven Phrase Structure Grammar*) de par ses structures de données qui incluent des ensembles de traits (*feature sets*) et qui autorisent différentes opérations sur ces traits (unification, disjonction, négation, existence de traits). Il s'en distance cependant de par ses structures de données arborescentes et donc ordonnées.

L'implémentation entièrement en PROLOG de CAT2 partage de nombreuses caractéristiques avec les implémentations des formalismes de la famille PATR (D-PATR, PATR-II, P-PATR). CAT2 exploite en effet le principe de l'unification ainsi que celui du backtracking. Il contient un interpréteur de commandes pour analyser, générer, comparer, éditer, détruire ou sauvegarder des données linguistiques, un débogueur qui permet de suivre pas à pas l'analyse ou la génération d'une donnée ainsi qu'un certain nombre d'autres routines (interaction avec le système UNIX, facilités d'édition, etc.).

L'implémentation actuelle de CAT2 est testée sur des modules de grammaires écrits pour le français, l'allemand, l'anglais, l'espagnol et le grec. Le système tourne sous le compilateur Prolog SICSTUS.

1.2. Description du formalisme

1.2.1. Les règles de générateurs

A) Syntaxe

Un objet (ou arbre) est produit dans le formalisme CAT2 par une règle de générateur dont la syntaxe est très proche de celle des listes de PROLOG :

TETE.QUEUE

La tête représente le noeud père d'un objet et la queue une partie d'objet, c'est-à-dire le premier niveau du reste de l'objet qui peut être vide dans le cas d'un noeud terminal (noté "[]"), ou non vide si le noeud père a un certain nombre de fils. Un objet dont la queue est vide est appelé "atome" et correspond à une entrée lexicale, par opposition aux objets non atomiques qui correspondent à des structures complexes (constituants, phrases).

Chaque noeud père comporte deux types d'informations représentées sous la forme suivante :

(PF,FL)

1) le trait privilégié caractéristique de son niveau d'analyse (PF = *privileged feature*). En effet, un trait privilégié est défini pour chaque générateur. Il s'agira par exemple de la catégorie syntaxique (*cat*) pour le générateur syntaxique (groupe nominal : "np", groupe verbal : "vp", groupe adjectival : "ap"...), de la fonction syntaxique (*sf*) pour le générateur relationnel (sujet, objet, modifieur...) et du rôle sémantique (*sr*) pour le générateur sémantique ("agent", "attribuant", "affected" qui sont des rôles proches de ceux proposés par les grammaires systémiques, cf. E. Steiner et al. [1988]).

2) la liste de traits associés (FL = *feature list*) consistant en un ensemble de couples "attribut-valeur" de la forme :

$$\{a_1=v_1, a_2=v_2, \dots, a_n=v_n\}$$

où chaque a_i représente un nom d'attribut et chaque v_i une constante, une variable ou une liste de couples "attribut-valeur".

Une de ces deux informations peut cependant ne pas être spécifiée : le noeud père comportera alors un point d'interrogation à la place du trait privilégié dans le cas où celui-ci n'est pas spécifié (TETE=(?,FL)) ou uniquement le trait privilégié dans le cas où la liste de traits associés est vide (TETE=(PF)). Ces deux informations peuvent également ne pas être spécifiées : la tête aura dans ce cas comme valeur un point d'interrogation (TETE=(?)). La tête d'un objet peut ainsi, dans le formalisme CAT2, contenir les informations suivantes (notation BNF) :

$$\text{TETE} ::= (\text{PF,FL}) \mid (\text{PF}) \mid (?,\text{FL}) \mid (?)$$

Les informations contenues dans la queue d'un objet non atomique sont de même nature que celles exposées précédemment et sont représentées telles que le montrent les notations BNF suivantes :

```

QUEUE_NONATOM ::= TETE
QUEUE_NONATOM ::= ^ QUEUE_NONATOM
QUEUE_NONATOM ::= + QUEUE_NONATOM
QUEUE_NONATOM ::= * QUEUE_NONATOM
QUEUE_NONATOM ::= (QUEUE_NONATOM ; QUEUE_NONATOM)
QUEUE_NONATOM ::= (QUEUE_NONATOM , QUEUE_NONATOM)

```

Le préfixe "^" dénote l'optionnalité (zéro ou une occurrence de l'expression), le préfixe "+" signifie une ou plusieurs occurrences de l'expression, préfixe "*" zéro ou plusieurs occurrences de l'expression, l'infixe ";" l'alternation et l'infixe "," la séquence d'expressions. Outre ces opérateurs, le formalisme CAT2 dispose de deux opérateurs supplémentaires qui le démarque des formalismes traditionnels basés sur les grammaires d'unification : l'opérateur d'inégalité "~=" et d'existence "==". Les informations contenues dans la queue d'un objet peuvent ainsi être représentées, dans le formalisme CAT2, de la manière suivante :

$$\text{QUEUE} ::= [] \mid [\text{QUEUE_NONATOM}]$$

B) Exemples

Le formalisme CAT2 comporte deux types de règles : des règles de "construction" (*building rules* ou *b-rules*) générant des objets, et des règles d'"attribut" (*attribute rules* ou *a-rules*) opérant sur ces objets, le critère d'applicabilité d'une *a-rule* sur un objet étant que cet objet ait le même trait privilégié exprimé dans la règle, et ce dans un même contexte. Les règles de construction générant des objets sont atomiques si elles se réfèrent à des entrées lexicales ou non atomiques si elles se réfèrent à des structures complexes.

Les règles d'attribut peuvent être de trois types :

- du type "défaut" (*default rules*) : ce type de règle permet d'ajouter des traits à une structure selon le principe de l'unification. Si l'unification échoue, c'est-à-dire si l'objet ne répond pas aux conditions exprimées dans la règle, celui-ci reste inchangé.
- du type "strict" (*strict rules*) : une règle du type "strict" exprime des conditions devant être remplies par un objet donné pour pouvoir être généré. Si l'unification échoue, la structure de l'objet est alors incorrecte et celui-ci n'est pas généré.
- du type "filtre" (*filter rules*) : ce type de règle exprime des conditions ne devant pas être remplies par un objet donné pour pouvoir être généré. Dans ce cas, l'objet est détruit lorsque l'unification se produit, c'est-à-dire lorsque l'objet répond aux conditions exprimées dans la règle.

Il faut noter que l'ordre est important dans les *a-rules*. Cette caractéristique fait du formalisme CAT2 un formalisme *non monotonique* qui se démarque des formalismes *monotoniques*.

La règle "défaut" suivante attribut la valeur "3" à la personne des verbes présents dans le lexique, ce qui permettra de ne coder que les personnes différentes de celle-ci (verbes à la première personne du singulier ou du pluriel par exemple). Nous noterons qu'une règle doit toujours être précédée dans le formalisme CAT2 d'un nom suivi du signe "=" et doit toujours

être terminée par un point, ainsi que la syntaxe de PROLOG l'exige (un *fait* PROLOG doit en effet toujours être terminé par un point) :

$$a1 = (v, \{per=3\}).[].$$

La règle "strict" suivante exprime, elle, des contraintes portant sur le nombre et la personne du sujet et du verbe d'une phrase donnée, ce qui permettra d'alléger les règles de construction générales :

$$a2 = s.[(np, \{per=P, num=N\}), (? , \{per=P, num=N\}), *].$$

La règle "filtre" suivante est relative au cas d'un groupe nominal et est chargée de détruire tous les objets comportant un groupe nominal de type non vide dont le cas n'est pas spécifié (de valeur "null") :

$$a3 = (?).[*, (np, \{type\sim=empty, case=null\}), *].$$

Une règle de générateur décrivant un objet pourra ainsi avoir la forme suivante :

- règles atomiques :

$$b1 = (v, \{lex=part, lu=partir, per=3, num=sing\}).[].$$

$$b2 = (v, \{lex=partent, lu=partir, per=3, num=plu\}).[].$$

- règles non atomiques :

$$b3 = (np, \{per=P, num=N\}).[^ (det, \{num=N\}), (n, \{per=P, num=N\})].$$

$$b4 = vp.[^ aux, v, * (np;pp), ^ s].$$

Les règles atomiques comportent dans ces exemples la forme conjuguée (*lex*), la forme de base (*lu*), la personne (*per*) ainsi que le nombre (*num*) du verbe considéré (*v*). Les règles non atomiques décrivent, elles, la structure d'un groupe nominal (*np*) ainsi que celle d'un groupe verbal (*vp*) . Alors que la règle b3 exprime des contraintes portant sur le nombre d'un déterminant éventuel (*det*) et celui d'un nom (*n*) entrant dans un groupe nominal (les traits de nombre doivent être identiques), la règle b4 ne contient aucune information codée sous la forme de listes de couples "attribut-valeur". Celle-ci ne fait que décrire la structure d'un groupe verbal comprenant éventuellement un auxiliaire (*aux*), un ou plusieurs groupes nominaux (*np*), un ou plusieurs groupes prépositionnels (*pp*) ou une phrase (*s*).

1.2.2. Les règles de translateurs

A) Syntaxe

Tout comme les règles de générateurs, les règles de translateurs (*t-rules*) peuvent être de construction (*b-t-rules*) ou d'attribut (*a-t-rules*). Leur syntaxe est la suivante :

$$TETE_S.QUEUE_S \Rightarrow TETE_C.QUEUE_C$$

La partie gauche de cette règle décrit la structure de l'objet source fourni par un niveau d'analyse et la partie droite celle de l'objet cible du prochain niveau d'analyse ou du même niveau sémantique IS lors d'un transfert. Ainsi, l'objet cible décrit dans une règle de translation ne sera créé que dans la mesure où la structure d'un objet source s'unifie avec la partie gauche de cette règle et que cet objet est lui-même défini dans le niveau cible.

B) Exemples

Le fonctionnement général des règles de translateurs repose sur le principe de la compositionnalité. En effet, la translation d'un objet décrit dans la partie gauche d'une règle de type "t" correspond à la translation de ses sous-objets (O_i) :

$$TETE_S.[O_{S1}, O_{S2}, \dots, O_{Sn}] \Rightarrow TETE_C.[O_{C1}, O_{C2}, \dots, O_{Cn}]$$

L'expression la plus simple du principe de la compositionnalité correspond au transfert lexical d'une langue vers une autre. Nous remarquerons qu'une règle de type "t" possède également un nom suivi d'un "=" de même qu'elle se termine par un point PROLOG :

$$b_t1 = (gov, \{lu=partir\}).[] \Rightarrow (gov, \{lu=gehen\}).[]$$

Ce principe ne pouvant toujours être respecté, le formalisme CAT2 met à la disposition de l'utilisateur des marqueurs, préfixés d'un "\$" et mis entre côtes, permettant de réordonner ou de supprimer des éléments d'une structure donnée, ou même d'insérer des éléments dans une structure donnée lors d'un passage d'un niveau à un autre :

- réarrangement d'éléments :

$$TETE_S.[\$1':O_{S1}, \$2':O_{S2}, \$3':O_{S3}] \Rightarrow TETE_C.[\$1', \$3', \$2']$$

- suppression d'éléments :

$$TETE_S.[\$1':O_{S1}, O_{S2}, \$2':O_{S3}] \Rightarrow TETE_C.[\$1', \$2']$$

-insertion d'éléments :

$$TETE_S.[\$1':O_{S1}, \$2':O_{S2}] \Rightarrow TETE_C.[\$1', O_C, \$2']$$

La règle de translation suivante exprimant le passage d'une structure de surface à une structure relationnelle de type "prédicat-arguments-modifieurs" est un exemple typique de telles opérations :

$$b_t2 = s.[\$1':np, vp.[\$2':v, \$3':(^np)], \$4':(^pp)] \\ \Rightarrow s.[\$2', \$1', \$3', \$4'].$$

Tout comme les règles d'attribut de générateurs, les règles d'attribut de translateurs opèrent sur des objets. La seule différence réside dans le critère d'applicabilité. Alors qu'il suffisait pour qu'une règle de générateur soit appliquée que le trait privilégié exprimé soit identique à celui de l'objet soumis, il est nécessaire ici que toute la partie gauche de la règle s'unifie avec la structure de l'objet source. Dans le cas d'un succès, cette règle aura les mêmes effets qu'une règle d'attribut de générateur. La règle "défaut" suivante fournit par exemple à chaque structure lexicale du niveau d'analyse cible sa catégorie associée :

$$a_t1 = (?,\{cat=C\}).[] \Rightarrow (?,\{cat=C\}).[].$$

Quant à la règle "strict" de transfert suivante, celle-ci assure que les rôles et les traits sémantiques soient identiques lors d'un passage d'une langue à une autre :

$$a_t2 = (?,\{role=R,semf=SF\}).[*] \Rightarrow (?,\{role=R,semf=SF\}).[*].$$

2. Les constructions à verbe support dans le domaine de la traduction automatique

2.1. Aperçu terminologique

Les constructions verbo-nominales consistant en l'association d'un verbe support (*Vsup*) et d'un nom prédicatif (*Npred*) telles que *faire du sport*, *donner un conseil* ou *prendre une décision* ont attiré l'attention, depuis les premiers travaux de Peter von Polenz [1963], de nombreux linguistes et grammairiens de langue allemande ou française. Les appellations ainsi que les définitions de telles constructions diffèrent d'un auteur à l'autre. Les appellations allemandes mettent en évidence la fonction du verbe support (*Funktionsverb*), "vide" sémantiquement (*Funktionsverbgefüge*, *Funktionsverbfügung*..., cf. W. Herrlitz [1973], H. Günther / S. Pape [1976], G. Helbig [1979], K.E. Sommerfeldt [1980], J. Yuan [1986], D. Bresson [1988a, 1988b] ou I. Starke [1989]), alors que les appellations françaises que l'on trouve principalement dans les travaux effectués dans le cadre du Laboratoire d'Automatique Documentaire et Linguistique de Paris (LADL) reposent à la fois sur la propriété du groupe nominal (*substantif / nom prédicatif*, *prédicat substantival / nominal*) et celle du verbe support ou opérateur (*construction à verbe support ou opérateur*, *construction ou phrase à prédicat nominal*, *construction nominale prédicative*, *construction nominale à verbe support*, *locution à verbe support*..., cf. J. Labelle [1974], J. Giry-Schneider [1986], G. Gross [1987] ou D. Bresson [1989]).

Un *prédicat nominal* est défini ici par le fait qu'il a des arguments ($N_{pred} (Argument_1, Argument_2... Argument_n)$) et par la nature de ces arguments, et non par un lien morphologique avec un prédicat verbal. Celui-ci est dans un emploi verbal "conjugué" ou "actualisé" par le verbe support porteur de marques de personne, de nombre ou de temps, et détermine le nombre, la forme syntaxique ainsi que la nature sémantique des arguments de l'ensemble de l'expression. Les propriétés d'un nom prédicatif dans un emploi nominal sont de ce fait supposées être identiques à celles d'un nom prédicatif dans un emploi verbal :

L'admiration de Max pour Léa
 => **admiration (Max, Léa)**
 => **Max a de l'admiration pour Léa**

2.2. Traitements automatiques existants des constructions étudiées

Il existe très peu de systèmes de traduction automatique qui proposent un traitement particulier des constructions à verbe support. En effet, si certains modèles ont été proposés (cf. M. Salkoff [1990]) ou si certains formalismes tels que le formalisme LTAG par exemple (*Lexicalized Tree Adjoining Grammar*, cf. A. Abeillé et al. [1990]) commencent à prendre en considération le phénomène des prédicats complexes illustré en grande partie par les constructions à verbe support, rares sont les systèmes de traduction automatique qui ont adopté une véritable stratégie d'implémentation pour ce type de constructions. Le système de traduction utilisé dans le cadre du projet EUROTRA peut être considéré comme l'un de ces systèmes.

Plusieurs modèles d'analyse de constructions à verbe support ont été proposés au sein d'EUROTRA, soit dans le cadre d'un projet de recherche sur les verbes supports et les noms prédicatifs dirigé par L. Danlos, soit dans le cadre d'une recherche effectuée par A. Rothkegel ayant fait l'objet d'une publication (cf. A. Rothkegel [1989]). Tous ces modèles, dont un seul a été retenu et implémenté, présentent des stratégies déterminées par les principes de base du système de traduction EUROTRA.

En effet, si l'hypothèse selon laquelle les phrases sont des *projections* de verbe et l'hypothèse de la *compositionnalité* sont deux principes qui s'appliquent parfaitement aux prédicats verbaux simples, ils ne conviennent généralement pas aux prédicats verbaux complexes (constructions à verbe support, expressions idiomatiques...) dont le sens est dans la plupart des cas semi ou non calculable.

Alors que la plupart des systèmes de traduction ne livrent aucune analyse particulière des constructions à verbe support considérées soit comme des constructions libres, formées de l'association d'un verbe plein et d'un complément, c'est-à-dire comme des constructions dont le sens est totalement calculable et pouvant être traduites compositionnellement, soit comme des expressions figées dont le sens n'est pas calculable et ne pouvant donc être traduites

compositionnellement, le système de traduction développé dans le cadre d'EUROTRA propose une analyse automatique basée sur la notion de *prédicat nominal* définie au sein du LADL, mettant ainsi en évidence la fonction sémantique des noms prédicatifs et par là-même la semi-calculabilité du sens de telles constructions.

Le choix du verbe support, lors d'une traduction d'une construction à verbe support d'une langue source en une construction à verbe support équivalente de la langue cible, est en effet non pas déterminé par le verbe support de la langue source, mais par le nom prédicatif de la langue cible, ainsi que le montrent les exemples suivants mettant en jeu différentes traductions anglaises du verbe *faire* :

| | |
|------------------------------|---------------------------------------|
| faire une promenade | => take a walk |
| faire du tennis | => play tennis |
| faire un miracle | => perform / work a miracle |
| faire la sieste | => have / take a nap |
| faire la guerre | => make / wage war |
| faire des réparations | => do repairs |

ou différentes traductions allemandes du verbe *prendre* :

| | |
|---------------------------------|--|
| prendre une décision | => einen Beschluß fassen |
| prendre l'initiative | => die Initiative ergreifen |
| prendre une mesure | => eine Maßnahme ergreifen / treffen |
| prendre peur | => Angst bekommen |
| prendre des risques | => Risiken eingehen |
| prendre de l'importance | => (an) Bedeutung gewinnen |
| prendre la peine (de) | => sich die Mühe machen |
| prendre en considération | => in Erwägung / in Betracht ziehen |

Il faut noter que le formalisme LTAG prend en considération ce type de problème relatif à la traduction automatique des constructions à verbe support. En effet, celui-ci ne traduit pas les constructions à verbe support selon le principe de la *compositionnalité*, c'est-à-dire en faisant correspondre à l'entrée lexicale du verbe support source l'entrée lexicale du verbe support cible :

take a walk => faire une promenade
*** take => faire, walk => promenade**

give a cry => pousser un cri
*** give => pousser, cry => cri**

mais plutôt selon le principe de la *non-compositionnalité* propre aux constructions figées : à une entrée lexicale de construction à verbe support source correspondra une entrée lexicale de construction à verbe support cible si celle-ci existe (*take a walk => faire une promenade*, *give a cry => pousser un cri*) et une entrée lexicale de verbe simple ou un autre type d'entrée

lexicale dans le cas contraire (*give a cough* => tousser, *faire un croc-en-jambe* => trip (up), *avoir soif* => be thirsty).

Le choix de la stratégie d'implémentation à adopter dans ce domaine (*compositionnalité*, *semi-compositionnalité* ou *non-compositionnalité*) n'a pas seulement un intérêt théorique. C'est celui-ci qui déterminera le caractère "économique" ou non "économique" d'un système de traduction automatique, de même que son degré de finesse au niveau de l'analyse sémantique. Si l'analyse des constructions à verbe support selon le principe de la *non-compositionnalité* rend un système de traduction automatique non économique, l'analyse de ces mêmes constructions selon le principe de la *compositionnalité* le rendra imprécis au niveau de l'analyse sémantique. Une phrase allemande comportant par exemple une construction à verbe support telle que *Probleme haben* (avoir des problèmes) avec topicalisation du nom prédicatif :

Probleme haben viele europäische Firmen

aura nécessairement deux interprétations au niveau sémantique au lieu d'une seule, étant donné le caractère non restreint du sujet du verbe *haben*, qui fourniront ainsi deux traductions françaises, dont l'une erronée :

**Ce sont des problèmes que beaucoup d'entreprises
européennes ont
* Des problèmes ont beaucoup d'entreprises européennes**

L'analyse automatique effectuée dans EUROTRA a ainsi le mérite, en adoptant le principe de la *semi-compositionnalité* et en soulignant de ce fait le rôle sémantique des noms entrant dans les constructions à verbe support, trop souvent négligé dans les études menées dans ce cadre, d'affiner l'analyse sémantique du système de traduction automatique utilisé.

Cette analyse présente cependant l'inconvénient majeur de reposer sur une définition trop stricte de ces expressions qui, en minimisant le rôle joué par les verbes supports, répond certes parfaitement aux exigences "économiques" de tout système de traduction automatique, mais exclut de la classe des constructions à verbe support un bon nombre de constructions qui semblent tout à fait en faire partie. Un nom jouant le rôle de prédicat sémantique dans une phrase peut être conjugué par un verbe support sans être lui-même porteur d'arguments :

**Paul fait de la natation / de la voile
* => La natation de Paul (natation (Paul))
* => La voile de Paul (voile (Paul))**

de même qu'un nom porteur d'arguments peut avoir une réalisation syntaxique différente selon qu'il se situe en position prédicative ou non, ainsi que le montre l'exemple en langue allemande suivant comprenant le nom prédicatif *Vorschlag* (proposition) :

Pauls Vorschlag an Marie
 (La proposition de Paul à Marie)
 => Vorschlag (Paul, Marie)
 => Paul macht Marie einen Vorschlag
 (Paul fait une proposition à Marie)

Dans ce cas-là, l'argument prépositionnel *an Marie* (à Marie) du nom prédicatif *Vorschlag* est réalisé dans la construction à verbe support *einen Vorschlag machen* (faire une proposition) non pas par un complément prépositionnel mais par un complément au datif (*Marie*), appelé incontestablement par le verbe support *machen*.

Si ce type de problème lié à l'emploi du datif de la langue allemande ne semble pas affecter les noms prédicatifs français, il présente l'intérêt de mettre en évidence l'influence, soulignée par K.E. Sommerfeldt [1980], que peut avoir le verbe support sur la nature et le nombre des actants des expressions étudiées, et permet ainsi de mettre en doute le rattachement systématique des compléments prépositionnels aux noms prédicatifs.

Ainsi, les modèles proposés dans le cadre d'EUROTRA tendent de différentes manières à refléter au mieux la sémantique des constructions à verbe support trop ignorée par les différents systèmes de traduction automatique existants. Ils alourdissent cependant considérablement les analyseurs, les lexiques et les modules de translation utilisés de par les représentations particulières exigées aux différents niveaux d'analyse. Ils ne conviennent pas toujours aux langues telles que l'allemand par exemple de même qu'ils n'expriment pas toujours la véritable sémantique de ces expressions (cf. N. MESLI [1991]).

3. Les constructions à verbe support dans le formalisme CAT2

3.1. Remarques préliminaires

3.1.1. Définition de la notion de "constructions à verbe support"

La question de savoir s'il faut considérer la construction toute entière (nom prédicatif + verbe support), en tant que verbe complexe, comme le principal porteur de valence de la phrase dans laquelle elle apparaît, ou uniquement le nom prédicatif, suscite de nombreuses discussions dans la recherche française et allemande. Toutes les analyses menées dans ce cadre montrent qu'il est difficile de parler en toute généralité uniquement de complément ou d'argument du nom de l'expression à verbe support et non de complément du verbe (cf. la suite *donner...à* suggérée par G. Gross [1987], selon lui plus complète que le simple verbe support *donner*) ou de la construction toute entière. Il nous semble alors plus raisonnable de considérer non pas le nom prédicatif comme le principal porteur de valence de la phrase dans

laquelle il apparaît, mais l'expression toute entière, tel que le suggèrent J. Korhonen [1978] ou J. Yuan [1986].

Nous entendrons donc ici par "nom prédicatif" un nom abstrait (un *Nomen actionis* au sens large, c'est-à-dire un nom exprimant un procès, une action, un état, un événement ou une qualité) qui peut jouer le rôle de prédicat sémantique dans une phrase. Plutôt que de "noms prédicatifs", vaudrait-il mieux parler, ainsi que le suggère D. Bresson [1989], de "groupes nominaux en position prédicative". Ce nom, lorsqu'il n'est pas "conjugué" ou "actualisé" par un verbe support, en d'autres termes en position non prédicative, peut être porteur d'arguments ou non :

L'influence néfaste de Paul sur Marie est connue de tous
 => **influence (Paul, Marie)**
 => **Paul exerce une influence néfaste sur Marie**

La natation est un sport qui plaît à Paul
 => **natation (0)**
 => **Paul fait de la natation**

Ce n'est pas le verbe qui porte l'information essentielle dans une construction à verbe support, mais le nom prédicatif. En tant que verbe support, il a considérablement réduit son contenu sémantique de verbe plein (on parle de verbe "vide" de sens ou "pauvre" en sens) et il ne sert qu'à conjuguer le nom ou nuancer sa signification. Il porte les marques grammaticales de personne, de nombre, de temps, de mode et de phase, et exprime des notions générales telles que l'entrée dans l'état ou le changement d'état qui relèvent du domaine de l'*Aktionsart* (cf. N. Mesli [1988, 1989]). L'expression toute entière forme cependant une unité sémantique de nature verbale à valeur de prédicat dans la phrase, et est dans son ensemble porteuse d'arguments. Le verbe support détermine en combinaison avec le nom prédicatif le nombre et la forme des arguments du complexe verbal :

Paul exerce une influence néfaste sur Marie
 => **exercer une influence (Paul, Marie)**

Paul fait de la natation
 => **faire de la natation (Paul)**

Cette définition détermine la stratégie adoptée à l'IAI de Sarrebruck pour l'analyse automatique des constructions à verbe support dans le formalisme CAT2. Cette analyse est ainsi basée sur l'hypothèse selon laquelle le nom prédicatif forme avec le verbe support une unité sémantique à valeur de prédicat dans la phrase, déterminant dans son ensemble le nombre, la forme syntaxique ainsi que la nature sémantique de ses arguments. Le nom prédicatif peut posséder ou ne pas posséder d'arguments, de même que la réalisation syntaxique de ses arguments peut être différente selon qu'il apparaît avec ou sans verbe support.

3) Npred au datif + Vsup

einer Prüfung / einem Verhör unterziehen
(soumettre à un examen / à un interrogatoire)

4) Npred au génitif + Vsup

guter Hoffnung / einer Meinung sein
(avoir des espérances, être du même avis)

5) Npred prépositionnel + Vsup

in Ordnung bringen, zu Hilfe kommen, im Zweifel lassen
(mettre en ordre, venir en aide, laisser dans le doute)

Si les constructions comprenant un nom prédicatif prépositionnel sont considérées par toutes les études allemandes menées dans ce cadre comme des constructions à verbe support, même s'il existe des divergences quant à la nature des prépositions entrant dans ces constructions, il n'en va pas de même pour une grande partie des constructions comprenant un nom prédicatif à l'accusatif ainsi que pour les autres types de constructions, considérées dans la plupart des cas comme des constructions libres.

L'analyse automatique effectuée dans le formalisme CAT2 est basée sur une conception large de la notion de construction à verbe support. En effet, elle considère ces constructions à la charnière des expressions libres, caractérisées par la calculabilité du sens et la variation, et des expressions idiomatiques à noms concrets métaphoriques, caractérisées, elles, par la non-calculabilité du sens et une stabilité souvent maximale. Le nom prédicatif est ici un substantif abstrait (*Nomen actionis* au sens large, désignant un procès, une action, une activité, un événement, un état, une qualité...) qui forme avec le verbe support un complexe verbal à valeur de prédicat dans la phrase. Le verbe support conjugue ou actualise le nom prédicatif - il porte les marques grammaticales de personne, de nombre, de temps, de mode et de phase de l'expression - et exprime des nuances diverses précisant le mode de déroulement du procès. Le verbe support est généralement "vide" de sens ou "pauvre" en sens dans la mesure où il perd en tant que verbe support considérablement son sens de verbe plein, mais peut également être plus chargé sémantiquement : il conserve dans ce cas-là plus ou moins son sens de verbe plein. Il apporte alors une nuance particulière à la construction ou impose son sens à celle-ci (constructions à verbe support existentiel ou événementiel où le nom prédicatif est en position sujet par exemple). Ces constructions ont alors un sens totalement calculable et ne posent pas de problème particulier dans un système de traduction automatique basé sur le principe de la compositionnalité.

L'étude menée à l'IAI de Sarrebruck n'a retenu ainsi que les types syntaxiques pertinents dans un système de traduction automatique. Elle a donc écarté les constructions à verbe support existentiel et événementiel, de même que les variantes stylistiques de verbes supports, plus chargées sémantiquement que ces derniers.

3.2. Le module d'analyse et de génération

L'avantage de l'analyse effectuée à l'IAI de Sarrebruck dans le formalisme CAT2 réside dans le fait qu'elle ne nécessite que très peu de modifications des grammaires existantes. Etant donné que cette solution conserve une représentation syntaxique des constructions à verbe support identique à celles des constructions libres, le module de translation entre le niveau syntaxique ERS et le niveau sémantique IS reste valable pour ces deux types de constructions. La reconnaissance et la représentation particulière des constructions à verbe support ne sont faites qu'au niveau sémantique seulement. Pour cela, le module de translation doit fournir au niveau IS pour chaque verbe apparu au niveau ERS deux interprétations :

- celle de verbe plein (la valeur lexicale du verbe est exprimée par la variable *LU*) :

```
(gov,{cat=v,lu=LU,luv=nil}).[]
```

- celle de verbe support (la valeur lexicale du verbe est mémorisée dans l'attribut *luv*) :

```
(gov,{cat=v,lu=vsup,luv=LU}).[]
```

Les règles ERS-IS suivantes effectuent ces opérations, à supposer que ces entrées lexicales existent au niveau IS :

```
atom_v      =   (gov,{cat=v,lu=LU}).[]
              => (gov,{cat=v,lu=LU,luv=nil}).[]
atom_vsup   =   (gov,{cat=v,lu=LU}).[]
              => (gov,{cat=v,lu=vsup,luv=LU}).[]
```

Pour pouvoir distinguer une position prédicative d'une position non prédicative d'un nom abstrait donné, le générateur sémantique présuppose un certain nombre d'informations pour ce nom. Ces informations sont codées, dans le dictionnaire associé, à l'intérieur d'un trait appelé "*vsup_frame*". Celui-ci aura en effet la valeur "nil" pour toutes les entrées lexicales des substantifs non prédicatifs ou des noms prédicatifs employés sans verbe support :

```
vsup_frame=nil
```

et les informations suivantes pour les entrées lexicales des noms en position prédicative :

```
vsup_frame={frame={...},vsup={...}}
```

1) des informations concernant les arguments de la construction à verbe support :

```
frame={arg1={...},arg2={...},arg3={...}}
```

Ces informations sont principalement sémantiques, la forme syntaxique de ces derniers ayant déjà été formulée au niveau ERS, soit dans l'entrée du nom prédicatif en tant qu'arguments de celui-ci, soit dans l'entrée du verbe en tant que compléments de verbe. Ces traits sémantiques se réfèrent respectivement aux rôles sémantiques proposés par E. Steiner et al. [1988] (*sr = agent / associated / affected...*) et à la classification sémantique des noms, verbes et adjectifs établie par C. Zelinski-Wibbelt [1988, 1989] (*semf = {abstract=concr} / {animate={anim={hum= _}}}*...). Certaines informations syntaxiques sont cependant conservées au niveau IS et exploitées afin d'éviter une "surgénération" d'objets : il s'agit des classes syntaxiques établies par P. Schmidt [1988] recensant les différentes fonctions syntaxiques possibles d'un argument donné :

| | | |
|----|---|--|
| c0 | : | fonction sujet |
| c1 | : | complément à l'accusatif |
| c2 | : | complément au génitif |
| c3 | : | complément au datif |
| c4 | : | complément prépositionnel |
| c5 | : | complément de lieu (locatif) |
| c6 | : | complément de temps |
| c7 | : | complément de lieu (directionnel) |
| c8 | : | attribut du sujet ou de l'objet |
| c9 | : | complément phrastique (infinitive, complétive ...) |

2) des informations concernant les verbes supports avec lesquels peut se combiner le nom prédicatif ou liées à l'emploi de ces verbes supports :

vsup={...}, avec :

- la valeur du verbe support (en allemand : *lu = haben / machen...*, en français : *lu = avoir / faire...*),
- l'*Aktionsart* du Vsup en combinaison avec le Npred (*akt = neut / incho / term / cont*),
- certaines caractéristiques de la construction à verbe support :
 - son caractère statique ou non statique (*vfeat = stat / nstat*),
 - son caractère causatif ou non causatif (*caus = yes / no*),
 - son caractère passif ou non passif (*pass = yes / no*),
 - son caractère converse ou non converse (*conv = yes / no*),
 - son caractère prépositionnel (en allemand : *pform = in / unter...*, en français : *pform = en / sous...*) ou non prépositionnel (*pform = nil*),
 - le cas du Npred dans une construction à verbe support prépositionnelle allemande (*case = acc / dat*).

ex : entrées lexicales IS du Npred **Angst** (peur)

- vsup_frame = nil :

angst =
(gov,{cat=n,lu=angst,vsup_frame=nil,

```

frame={arg1={sf=c2,sr=processor,
             semf={animate={anim={hum=__}}}},
       arg2=({sf=c4,sr=phen,pform=vor}
             ;{sf=c9,sr=phen} )},
semf={abstract={abstr={ment={mode=emot,cons=pass}}},
      gran=nil,animate=nil,temp={ext={unbounded=state}},
      bound=mass,complex=nil,homogen=homo}}.[]
- vsup_frame ~= nil :

```

(N.B. : les couples "attribut-valeur" *caus=no*, *pass=no*, *conv=no* et *pform=nil* sont attribués à ces entrées lexicales par des règles de type "défaut" et ne figurent donc pas dans celles-ci)

```

angst_vsup =
(gov,{cat=n,pred=yes,lu=angst,
  semf={abstract={abstr={ment={mode=emot,cons=pass}}},
        gran=nil,animate=nil,temp={ext={unbounded=state}},
        bound=mass,complex=nil,homogen=homo},
  vsup_frame=(
    {frame={arg1={sf=c0,sr=processor,
                  semf={animate={anim={hum=__}}}},
            arg2=({sf=c4,sr=phen,pform=vor}
                  ;{sf=c9,sr=phen} )},
            vsup=({lu=haben,akt=neut,vfeat=stat}
                  ;{lu=bekommen,akt=incho,vfeat=nstat}
                  ;{lu=verlieren,akt=term,vfeat=nstat} )}
    ; {frame={arg1={sf=c0,sr=phen},
              arg2={sf=c3,sr=processor,
                    semf={animate={anim={hum=__}}}},
              vsup={lu=machen,akt=incho,caus=yes,vfeat=nstat}}
    ; {frame={arg1={sf=c0,sr=processor,
                    semf={animate={anim={hum=__}}}},
              vsup={lu=geraten,akt=incho,pform=in,case=acc,vfeat=nstat}}
    ; {frame={arg1={sf=c0,sr=phen},
              arg2={sf=c1,sr=processor,
                    semf={animate={anim={hum=__}}}},
              vsup={lu=versetzen,akt=incho,pform=in,case=acc,
                    caus=yes,vfeat=nstat}}}}).[]

```

Nous constatons ainsi que le nom prédicatif *Angst* employé avec verbe support a une structure syntaxique et sémantique qui peut varier selon le verbe support et qui n'est donc pas toujours identique à celle de son emploi sans verbe support (informations contenues dans le trait *frame*).

La reconnaissance d'une construction à verbe support se fera ainsi au moyen des informations codées dans le trait *vsup_frame* et consistera :

1) à comparer la valeur lexicale du verbe apparu dans la construction, mémorisée dans le trait *luv*, ainsi que la valeur de la préposition du nom prépositionnel éventuellement apparu de même que son cas (cas du nom prépositionnel allemand), mémorisés eux dans les traits *pform* et *case*, avec les valeurs des traits respectifs codés dans l'entrée du nom prédicatif (*lu*, *pform*, *case*),

2) à vérifier que la forme syntaxique ainsi que la nature sémantique des actants apparus soient identiques à ceux décrits dans l'entrée du nom prédicatif.

Si toutes ces conditions sont réalisées, l'objet contenant la lecture de verbe plein du verbe apparu dans la construction ne survivra pas et l'analyseur sémantique générera pour la variante "vsup" de celui-ci une représentation à gouverneur verbal abstrait (*lu=vsup*) contenant toutes les informations relatives à la construction à verbe support (*Aktionsart*, caractère causatif, passif, converse...), directement suivi du nom prédicatif (appelé "pred") ainsi que des arguments et des modificateurs de la construction à verbe support :

constructeur pour les phrases comportant une construction à verbe support (*svc=yes*) :

```
cs1 = (?,{cat=s,semf=SEMF,svc=yes}).[
  (gov,{cat=v,lu=vsup,luv=LU,akt=AKT,vfeat=VF,
    caus=CA,conv=CO,pass=PA}),
  (pred,{cat=np,semf=SEMF,pform=PF,case=CS,
    vsup__frame={frame={arg1=A1,arg2=A2,arg3=A3},
      vsup=(lu=LU,pform=PF,case=CS,
        akt=AKT,vfeat=VF,
        caus=CA,conv=CO,pass=PA)}}),
  ^ (? ,A1),
  ^ (? ,A2),
  ^ (? ,A3),
  * mod,
  ^ neg ].
```

Dans le cas contraire, c'est l'objet contenant la lecture de verbe support du verbe apparu dans la construction qui ne survivra pas. L'analyseur sémantique générera alors pour la lecture de verbe plein de celui-ci une représentation normale à gouverneur verbal déterminant la forme syntaxique ainsi que la nature sémantique de ses arguments :

constructeur pour les phrases comportant une construction libre (*svc=no*) :

```
cs2 = (?,{cat=s,semf=SEMF,svc=no}).[
  (gov,{cat=v,luv=nil,semf=SEMF,
    frame={arg1=A1,arg2=A2,arg3=A3}}),
  ^ (? ,A1),
  ^ (? ,A2),
  ^ (? ,A3),
  * mod,
  ^ neg ].
```

Cette démarche présuppose deux règles pour les noms, selon la valeur du trait *vsup__frame*. Dans le cas où celui-ci vaut *nil*, la valeur sémantique des arguments provient du trait *frame* et aucune information particulière concernant ces derniers n'est communiquée aux relatives qui, dans le cas d'un nom prédicatif, pourraient comporter un verbe support :

```

cnp1 =
(?,{cat=np,semf=SEMF,vsup_frame=nil}).[
  (gov,{cat=n,semf=SEMF,vsup_frame=nil,
    frame={arg1=A1,arg2=A2,arg3=A3}}),
  ^ (? ,A1),
  ^ (? ,A2),
  ^ (? ,A3),
  * (mod,{cat~s}),
  ^ (mod,{cat=s,type=rel,
    topic={semf=SEMF,vsup_frame=nil}),
  ^ neg ].

```

Dans le cas d'un nom prédicatif (*vsup_frame~nil*), la valeur sémantique des arguments provient du trait *vsup_frame* et toute l'information qu'il contient est communiquée aux relatives afin de déterminer si celles-ci contiennent un verbe support ou non et si la nature des arguments du Npred qu'elles pourraient comporter est correcte :

```

cnp2 =
(?,{cat=np,semf=SEMF,vsup_frame=VSUP}).[
  (gov,{cat=n,semf=SEMF,vsup_frame=VSUP,
    vsup_frame={frame={arg2=A2,arg3=A3}}}),
  ^ (? ,A2),
  ^ (? ,A3),
  * (mod,{cat~s}),
  ^ (mod,{cat=s,type=rel,
    topic={semf=SEMF,vsup_frame=VSUP}),
  ^ neg ].

```

Le générateur sémantique décrit vaut aussi bien pour l'analyse que pour la génération d'objets. Il sert en effet de niveau cible lors de l'analyse pour créer des représentations sémantiques particulières (ERS => IS) et de niveau source lors de la génération pour restituer les représentations syntaxiques de la langue cible (IS => ERS). Ainsi, si la solution adoptée à l'IAI de Sarrebruck était avantageuse lors de l'analyse dans la mesure où le module de translation entre le niveau syntaxique ERS et le niveau sémantique IS restait valable pour les constructions à verbe support comme pour les constructions libres, cette solution l'est aussi lors de la génération dans la mesure où le module de translation inverse, c'est-à-dire entre le niveau sémantique IS et le niveau syntaxique ERS, reste également valable pour ces deux types de constructions. Celui-ci n'a pour tâche que de restituer la valeur lexicale des verbes supports, mémorisée dans le trait *luv*, ainsi que le montrent les règles de translation atomiques IS-ERS suivantes :

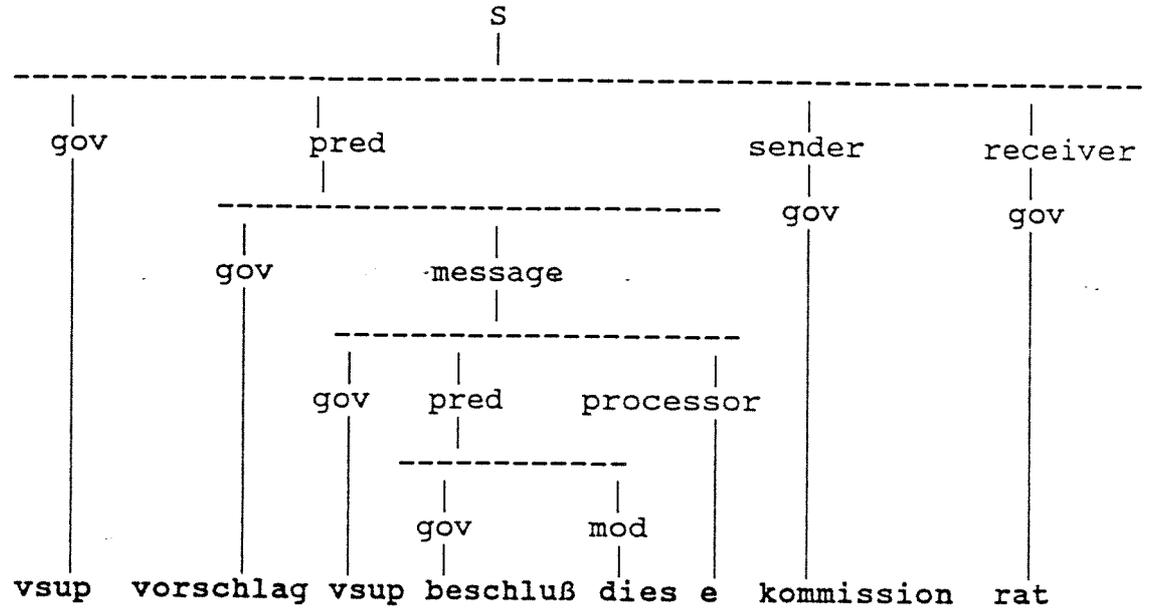
```

atom__v      =   (gov,{cat=v,lu=LU,luv=nil}).[]
              => (gov,{cat=v,lu=LU}).[]
atom__vsup   =   (gov,{cat=v,lu=vsup,luv=LU}).[]
              => (gov,{cat=v,lu=LU}).[]

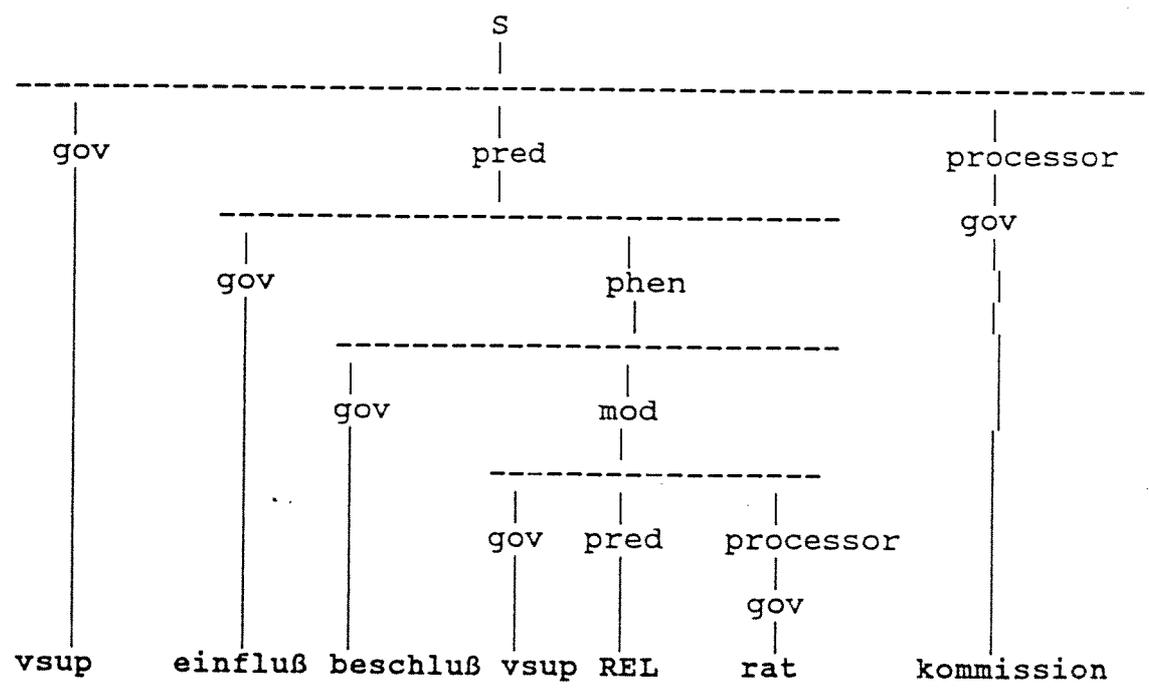
```

Exemples d'arbres IS allemands :

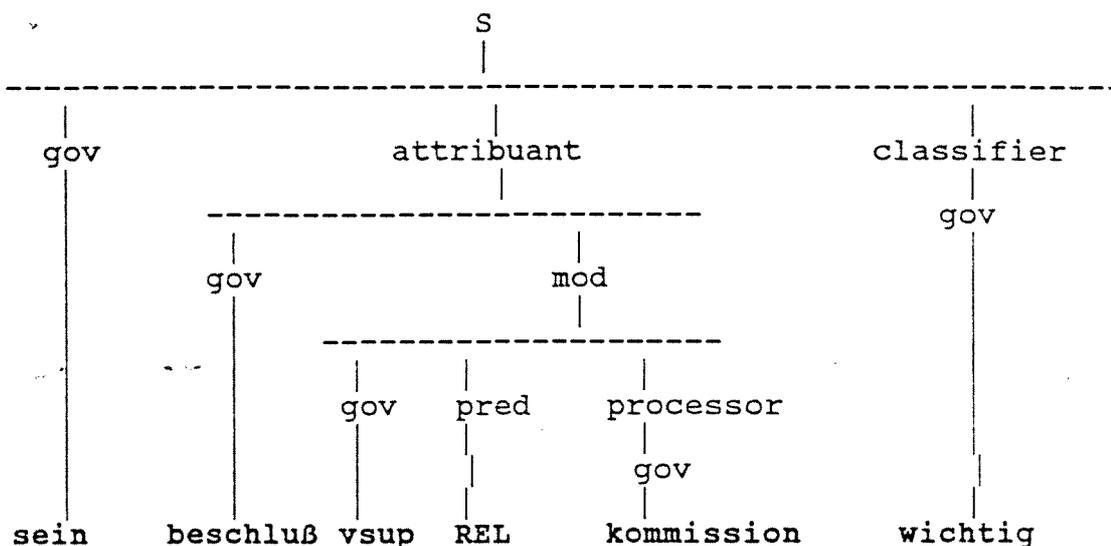
ex1 : Die Kommission macht dem Rat den Vorschlag, diesen
 Beschluß zu fassen
 (La Commission suggère au Conseil de prendre cette décision)



ex2 : Die Kommission ist unter dem Einfluß des Beschlusses,
 den der Rat faßt
 (La Commission est sous l'influence de la décision que le
 Conseil prend)



ex3 : Der Beschluß, den die Kommission faßt, ist wichtig
(La décision que la Commission prend est importante)



3.3. Le module de transfert

En ce qui concerne les modules de transfert, aucune des solutions proposées dans le cadre du système EUROTRA n'est plus avantageuse qu'une autre. En effet, aucune d'elles n'est susceptible de traiter le phénomène linguistique des constructions à verbe support dans son ensemble selon la stratégie dite de transfert simple. Quel que soit le modèle adopté, le transfert d'une construction à verbe support source ne pourra se réduire à un transfert lexical (traduction des feuilles de l'arbre) qu'à la condition qu'il existe une construction à verbe support cible équivalente, c'est-à-dire exprimant la même *Aktionsart* et les mêmes nuances (trait causatif ou non, nuance passive ou non passive etc.). Si cette condition n'est pas vérifiée, le transfert complexe ne pourra être évité. Ce transfert complexe pourra par exemple faire correspondre à une construction à verbe support source un verbe ordinaire, un verbe ordinaire accompagné d'une construction périphrastique exprimant la même *Aktionsart* que cette construction à verbe support, une construction périphrastique porteuse d'une *Aktionsart* identique à celle-ci suivie d'une construction à verbe support d'*Aktionsart* neutre, une construction adjectivale comportant un verbe copule de même *Aktionsart*, etc. Il y aura même transfert complexe si à une construction à verbe support cible correspond un verbe ordinaire, une construction adjectivale ou toute autre construction non isomorphe source.

Le transfert simple est exprimé dans le formalisme CAT2 par une règle unique transmettant toutes les structures sources au niveau cible :

$b_t1 = (?).[+'$'] \Rightarrow (?).['$']$.

Cette règle est donc valable aussi bien pour les constructions libres que pour les constructions à verbe support. Dans le cas de transfert simple, c'est-à-dire d'isomorphie des arbres IS, l'analyse effectuée à l'IAI de Sarrebruck ne présuppose ainsi, pour la traduction automatique des constructions à verbe support, que l'écriture de règles de transfert atomiques faisant correspondre aux noms prédicatifs de la langue source les noms prédicatifs de la langue cible ainsi qu'une règle transmettant par l'intermédiaire de la valeur abstraite "vsup" la nature de l'*Aktionsart* exprimée par la construction à verbe support de même que d'autres nuances de celle-ci, mémorisées au niveau IS, telles qu'un trait causatif, converse ou passif qui permettront d'obtenir une traduction plus juste de ces expressions :

```

b_vsup = (gov,{lu=vsup,akt=AKT,caus=CA,conv=CO,pass=PA}).[]
=> (gov,{lu=vsup,akt=AKT,caus=CA,conv=CO,pass=PA}).[]

% prendre une décision => einen Beschluß fassen
décision = (gov,{lu=décision}).[] => (gov,{lu=beschluß}).[]

% prendre une mesure => eine Maßnahme ergreifen / treffen
mesure = (gov,{lu=mesure}).[] => (gov,{lu=maßnahme}).[]

% avoir une influence => einen Einfluß haben
influence = (gov,{lu=influence}).[] => (gov,{lu=einfluß}).[]

```

Tous les cas que ne couvre pas la règle générale de transfert citée précédemment deviennent des cas particuliers qui doivent être résolus par l'écriture de règles de transfert complexe (non-isomorphie des arbres IS), ainsi que le montrent les règles suivantes faisant correspondre à des constructions à verbe support françaises des verbes ordinaires ou des constructions adjectivales allemandes, ou à des verbes ordinaires français des constructions à verbe support allemandes :

```

% faire partie => gehören (zu)
faire_partie =
(?,{cat=s}).[
    (gov,{cat=v,lu=vsup,luv=faire}),
    pred.[(gov,{lu=partie}),'$ARG'],
    '$SUBJ',
    *'$MODS':mod]
=>
(?,{cat=s}).[
    (gov,{cat=v,lu=gehören}),
    '$SUBJ',
    '$ARG':(?,(sf=c4,pform=zu)),
    '$MODS'].

% tirer profit => sich zunutze machen
tirer_profit =
(?,{cat=s}).[
    (gov,{cat=v,lu=vsup,luv=tirer}),
    pred.[(gov,{lu=profit}),'$ARG'],
    '$SUBJ',
    *'$MODS':mod]

```


système de traduction EUROTRA. Cet article a montré l'intérêt, dans un tel système de traduction automatique stratificationnel et à transfert, de l'analyse particulière des constructions à verbe support du type **faire du sport, donner un conseil ou prendre une décision**. Il a en effet donné un aperçu du traitement de ces constructions dans le domaine de la traduction automatique et a exposé un modèle basé sur une conception large de la notion de constructions à verbe support ainsi que sur des critères économiques : l'analyse automatique des constructions à verbe support effectuée à l'IAI de Sarrebruck dans le formalisme CAT2.

Cette analyse est cependant loin d'être exhaustive de même que tous les phénomènes traités dans le cadre du projet expérimental dans lequel elle s'insère. Un certain nombre de problèmes ne sont pas encore résolus ou que d'une manière partielle seulement. Il s'agit par exemple du problème des contraintes portant sur l'emploi des déterminants ou des expansions (modificateurs, arguments) des noms prédicatifs que l'on ne peut résoudre d'une manière satisfaisante que dans le cadre d'une étude sémantique et syntaxique générale des prédicats nominaux, ou bien du problème de la reconnaissance et de l'analyse de constructions à verbe support dans des structures que le formalisme ou la puissance du principe de l'unification ne nous permet pas pour l'instant de résoudre : la coordination de noms prédicatifs autour d'un même verbe support, les comparatives ou les phrases comportant deux constructions à verbe support autour d'un même nom prédicatif comme c'est le cas de phrases comportant une relative par exemple :

Die Kommission faßt den Beschluß, den der Rat faßt
(La Commission prend la décision que le Conseil prend)

Paul wurde viele Probleme los, die er einen Monat vorher noch gehabt hatte
(Paul fut débarrassé des nombreux problèmes qu'il avait eu un mois auparavant)

L'analyse et le système de traduction automatique de constructions à verbe support proposés par l'IAI de Sarrebruck restent ainsi expérimentaux et constituent une alternative intéressante car peu coûteuse au niveau des grammaires et des lexiques existants.

*
* *
*

Je voudrais remercier tout particulièrement mes collègues de l'IAI qui ont contribué à la réalisation du traitement automatique des constructions à verbe support effectué à Sarrebruck : Hans Haller et Paul Schmidt pour avoir conçu et implémenté des modules de grammaires cohérents et suffisamment complexes ayant assuré le bon déroulement de mon

travail, Randall Sharp pour ses conseils et l'aide qu'il m'a apportée lors de l'implémentation du phénomène étudié.

Bibliographie

ABEILLÉ, Anne, Y. SCHABES, A.K. JOSHI, 1990 : "Using Lexicalized Tags for Machine Translation", *Proceedings of COLING-90*, vol. 3, Helsinki, 1-6.

ARNOLD, Douglas J., L. JASPAERT, R. JOHNSON, S. KRAUWER, M. ROSNER, L. des TOMBES, G.B. VARILE & S. WARWICK, 1985 : "A MUI View of the <C,A>, T Framework in EUROTRA", *Proceedings of the Conference on Theoretical & Methodological Issues in Machine Translation of Natural Languages*, Hamilton, New York, 1-14.

ARNOLD, Douglas J., S. KRAUWER, M. ROSNER, L. des TOMBES & G.B. VARILE, 1986 : "The <C,A>, T Framework in EUROTRA : A Theoretically Committed Notation for MT", *Proceedings of COLING'86*, Bonn, 297-303.

BRESSON, Daniel, 1988a : "Der nominale Teil in den Funktionsverbgefügen", *Cahiers d'Etudes Germaniques* 14, 71-114.

BRESSON, Daniel, 1988b : "Classification des verbes supports (*Funktionsverben*) de l'allemand", *Cahiers d'Etudes Germaniques* 15, 53-65.

BRESSON, Daniel, 1989 : "La distribution du sens dans les locutions à verbe support", *Cercle Linguistique d'Aix-en-Provence*, travaux 7, 57-72.

DANLOS, Laurence, 1989 : "La traduction automatique", *Annales des Télécommunications*, Tome 44, n°1-2 janvier-février, 101-110.

GIRY-SCHNEIDER, Jacqueline, 1986 : "Les noms construits avec *faire* : compléments ou prédicats ?", *Langue française* 69, Paris, 49-63.

GROSS, Gaston, 1987 : *Etude syntaxique de constructions converses*, Thèse de Doctorat d'Etat, Université Paris XIII.

GÜNTHER, Heide / PAPE, Sabine, 1976 : "Funktionsverbgefüge als Problem der Beschreibung komplexer Verben in der Valenztheorie", in Helmut Schumacher (Hrsg.), *Untersuchungen zur Verbvalenz*, Forschungsberichte des Instituts für deutsche Sprache, Band 30, 92-128, Tübingen.

HELBIG, Gerhard, 1979 : "Probleme der Beschreibung von Funktionsverbgefügen im Deutschen", *Deutsch als Fremdsprache* 16, 273-285.

HERRLITZ, Wolfgang, 1973 : *Funktionsverbgefüge vom Typ "in Erfahrung bringen". Ein Beitrag zur generativ-transformationellen Grammatik des Deutschen*, Linguistische Arbeiten 1, Tübingen.

KORHONEN, Jarmo, 1978 : *Studien zu Dependenz, Valenz und Satzmodell. Teil II. Untersuchungen anhand eines Luther-Textes*, Europäische Hochschulschriften I/271, Bern / Frankfurt am Main.

LABELLE, Jacques, 1974 : *Etude de constructions avec opérateur AVOIR (nominalisations et extensions)*, Thèse de Troisième Cycle, LADL, Université Paris VIII.

MESLI, Nadia, 1988 : "Classification des verbes simples et composés de l'allemand fondée sur les *Aktionsarten*", *Cahiers d'Etudes Germaniques* 15, 29-51.

MESLI, Nadia, 1989 : *Les locutions verbales dans les écrits de M. Luther*, Thèse de Doctorat nouveau régime, Université d'Aix-Marseille I, Aix-en-Provence.

MESLI, Nadia, 1991 : "Funktionsverbgefüge in der maschinellen Analyse und Übersetzung : linguistische Beschreibung und Implementierung im CAT2-Formalismus", à paraître dans : *Eurotra-D Working Papers*, IAI-Saarbrücken.

von POLENZ, Peter, 1963 : *Funktionsverben im heutigen Deutsch. Sprache in der rationalisierten Welt*, Beihefte zur Zeitschrift *Wirkendes Wort*, Heft 5, Düsseldorf.

ROTHKEGEL, Annely, 1989 : *Polylexicalität. Verb-Nomen-Verbindungen und ihre Behandlung in Eurotra*, *Eurotra-D Working Papers* 17, IAI-Saarbrücken.

SALKOFF, Morris, 1990 : "Automatic translation of support verb constructions", *Proceedings of COLING-90*, vol. 3, Helsinki, 243-246.

SCHMIDT, Paul, 1988 : "A syntactic description of a fragment of German in the EUROTRA framework", in : E. Steiner, P. Schmidt & C. Zelinsky-Wibbelt (eds.), *From Syntax to Semantics : Insights from Machine Translation*, London : Frances Pinter, 11-39.

SHARP, Randall, 1988 : "CAT2 - Implementing a Formalism for Multi-Lingual MT", *Proceedings of the 2nd International Conference on Theoretical & Methodological Issues in Machine Translation of Natural Languages*, Carnegie Mellon University, Pittsburgh, Pennsylvania.

SHARP, Randall, 1989 : "CAT2 - A Formalism for Multilingual Machine Translation", *Proceedings of the International Seminar on Machine Translation*, Tblisi, Georgia, USSR.

- SHARP, Randall, 1990 : "Modelling GB in the CAT2 Machine Translation System", Workshop on GB-Parsing, University of Geneva.
- SOMMERFELDT, Karl-Ernst, 1980 : "Zur Valenz von Funktionsverbgefügen", *Deutsch als Fremdsprache* 17, 294-297.
- STARKE, Ingrid, 1989 : "Untersuchungen zur syntaktisch-semantischen Leistung von Funktionsverbgefügen im Deutschen (als Grundlage für eine automatische Analyse)", *studia grammatica* XXX, Akademie-Verlag Berlin, 78-114.
- STEINER, Erich, U. ECKERT, B. ROTH & J. WINTER-THIELEN, 1988 : "The development of the EUROTRA-D system of Semantic Relations", in : E. Steiner, P. Schmidt & C. Zelinsky-Wibbelt (eds.), *From Syntax to Semantics : Insights from Machine Translation*, London : Frances Pinter, 40-104.
- YUAN, Jie, 1986 : *Funktionsverbgefüge im heutigen Deutsch. Eine Analyse und Kontrastierung mit ihren chinesischen Entsprechungen*, Sammlung Groos 28, Heidelberg.
- ZELINSKY-WIBBELT, Cornelia, 1988 : "From cognitive grammar to the generation of semantic interpretation in machine translation", in : E. Steiner, P. Schmidt & C. Zelinsky-Wibbelt (eds.), *From Syntax to Semantics : Insights from Machine Translation*, London : Frances Pinter, 105-132.
- ZELINSKY-WIBBELT, Cornelia, 1989 : *Machine Translation Based on Cognitive Linguistics : What Lexical Semantics Contributes to the Semantic Unity of a Sentence*, EUROTRA-D Working Papers 16, IAI-Saarbrücken.

METHODOLOGIE DE DEVELOPPEMENT DES SYSTEMES EXPERTS

LOZKA POPOVA

Laboratoire d'Informatique de l'Université de Poitiers,
Faculté des Sciences Fondamentales et Appliquées
40, Avenue du recteur Pineau, 86022 Poitiers Cedex

e-mail popova@matpts.univ-poitiers.fr, tél. (33) 49.45.38.87

Résumé :

L'objectif de développement d'un système expert (SE) est de transformer l'expertise humaine en programme informatique.

Les systèmes experts sont distincts des systèmes informatiques traditionnels à la fois par le type de problème qu'ils abordent et la manière dont ils "vivent". La mise en place de la base de connaissance constitué un important problème lors de la conception de ces systèmes.

Le développement d'un SE, dans la méthodologie présente, est effectué sur la base d'une méthode structurale, qui donne un avantage lors de la spécification des différents composants du système. Lors de cette méthodologie chaque problème est divisé en sous-problèmes pour faciliter son étude. Chaque sous-problème est présenté comme un sous-système et il est défini indépendamment.

La conception du SE exige la séparation des modules pour le traitement des données et des modules spécifiques à l'intelligence artificielle. Chaque module est réalisé d'une manière relativement indépendante, en même temps en combinaison avec les autres.

La méthodologie présentée comporte les cinq phases : 1) étude; 2) analyse structurale; 3) projet de structure; 4) développement; 5) production du SE.

Les spécialistes effectuent à chaque étape un travail déterminé en fonction du temps, des moyens de nature financière, et aussi le recueil de la connaissance et le chargement de la base de connaissances.

L'exposé se situe sur un axe méthodologique, à mi chemin entre la théorie pure et la simple pratique.

GROUPE DE FORMATION DOCTORALE
INFORMATIQUE FONDAMENTALE ET SCIENCES DE LA COMPUTATION

===

| | | |
|------------------|-----------------------|--------------------------------|
| Edmond Bianco | Professeur | Université d'Aix-Marseille II |
| Roger Cusin | Professeur | Université d'Aix-Marseille II |
| Roger Dupuy | Professeur | Université de Paris VI |
| Louis Nolin | Professeur | Université de Paris VII |
| | | |
| A.Djabi | Maître Assistant | Université de Sétif |
| P.Isoardi | Maître de Conférences | Université d'Avignon |
| R.Jacquier | Maître de Conférences | Université d'Aix-Marseille III |
| B.Goossens | Maître de Conférences | Université de Paris VII |
| J.M.Knippel | Maître de Conférences | Université d'Aix-Marseille II |
| M.T.Laskri | Maître Assistant | Université d'Annaba |
| J.Ph.Lehmann | Maître de Conférences | Université d'Avignon |
| R.Stutzmann | Maître de Conférences | Université de Strasbourg |
| | | |
| E.Andriamasinoro | Chercheur | Université d'Aix-Marseille II |
| Ch.Berge | Chercheur | Université d'Aix-Marseille II |
| E.Boudiba | Chargé de cours | Université d'Aix-Marseille II |
| M.Boudjedir | Chercheur | Université d'Annaba |
| M.Charifi | Chef de projet | Handel |
| M.Derradji | Chercheur | Université d'Annaba |
| M.Djebali | Chercheur | Université d'Aix-Marseille II |
| Ch.Hurson | Chercheur | Université de Crète |
| K.Joëlson | Chercheur | Université d'Aix-Marseille II |
| M.Hadjali | Chargé de cours | Université de Damas |
| S.Hilala | Chargé de cours | Université d'Aix-Marseille II |
| A.Kacimi | Chercheur | Université d'Aix-Marseille II |
| F.Lorenzo | Chef de projet | Bouyghes |
| Ph.Maes | Ingénieur | Gemplus-Card International |
| P.Sanchez | Ingénieur | C.N.R.S.-Marseille |
| F.Serrano | Chef de projet | Conseil régional P.A.C.A. |

**Université de Provence
Atelier de Reprographie
Centre Saint Charles
3, place Victor Hugo
F - 13331 Marseille Cedex 3**