



ESSAI D'UNE THEORIE SEMANTIQUE APPLICABLE
AU TRAITEMENT DE LANGAGE
(MODELE "SENS - TEXTE").

A. JOLKOVSKY - I. MEL'CUK
Institut de Linguistique
MOSCOU

RESUME

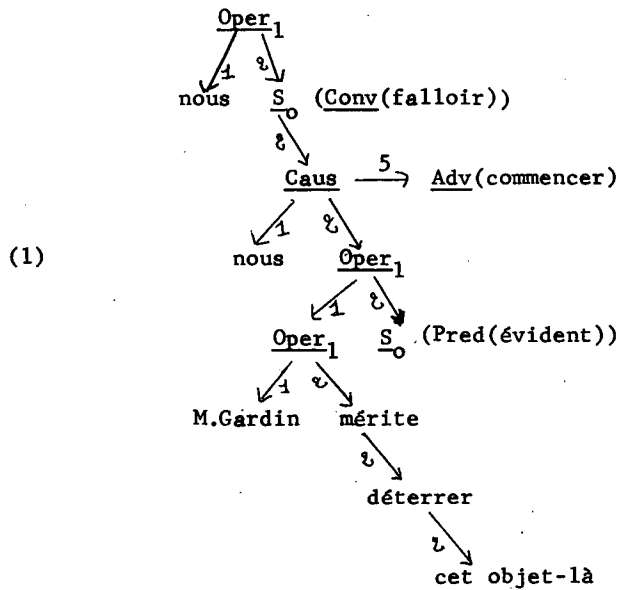
1. Le problème central de la linguistique est conçu comme la construction d'un modèle fonctionnel (susceptible d'être réalisé sous forme d'un programme pour ordinateur) du langage, c'est-à-dire, d'un système qui transforme n'importe quel sens donné en textes correspondants ou vice-versa. Partant du fait fondamental qu'en parlant on ne fait qu'exprimer un contenu préconçu, le modèle ici présenté ne se veut pas un système génératif dans le sens strict du terme, mais plutôt un système translatif avec entrée et sortie : si l'on introduit à l'entrée une "portion" de sens on obtient à la sortie tous les textes (de la langue modelée) qui portent le sens donné (synthèse) ; fonctionnant inversement le modèle produira pour chaque texte la notation symbolique de son sens (analyse).

Le modèle "Sens - Texte" (dorénavant -MST) comporte trois niveaux principaux - morphologie, syntaxique et sémantique - dont seulement le dernier sera esquissé ici (sous l'aspect synthétique exclusivement).

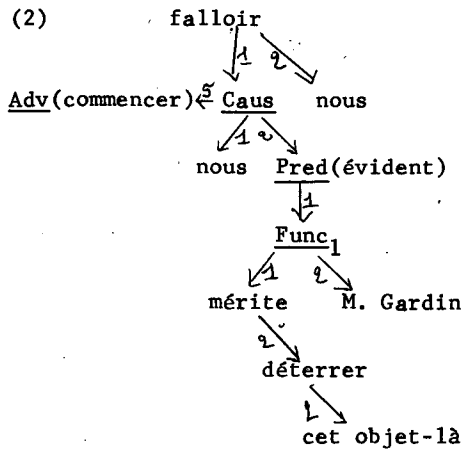
2. Le composant sémantique du MST a pour but de transformer le sens (la notation sémantique donnée) en structures syntaxiques superficielles, c'est-à-dire, en ensembles de mots reliés par de liens syntaxiques, telles que "sujet-verbe fini" ou "préposition-nom régi" etc. Cette transformation s'opère en trois temps successifs : 1) de la "notation sémantique" (contenu d'un texte sous forme d'un graphe complexe constitué par des entités sémantiques élémentaires) vers les Basic-structures lexico-syntaxiques (BSLS, cf. 3) ; 2) d'une BSLS vers toutes les SLS synonymes (entre elles, aussi bien qu'avec cette BSLS), ou "structures profondes" ; 3) d'une SLS vers toutes les structures syntaxiques superficielles incarnant cette SLS. C'est de la deuxième opération que nous allons nous occuper ici, et notamment du mécanisme qui en est responsable et que nous

désigner par système de périphrases. On notera que sous périphrase nous entendons ici la transition d'une SLS aux SLS synonymes/ (et non d'une phrase aux phrases synonymes!) une BSLS étant toujours une SLS).

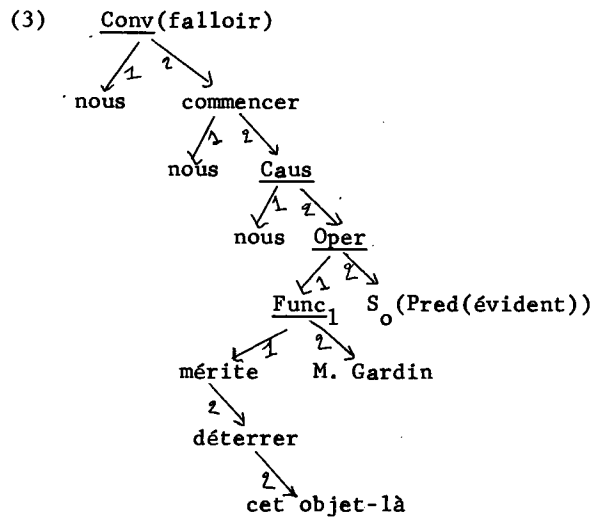
3. Une SLS (structure profonde d'une phrase) est une arborescence dont les branches dénotent des relations syntaxiques profondes et dont les noeuds sont des lexèmes profonds. La syntaxe profonde comporte six relations : 4 relations "actantielles" (cf. L. Tesnière), 1 relation attributive et 1 relation coordinative. Le lexique profond comprend des mots autonomes, ou mots clés, et des fonctions lexicales (cf. 4). Exemples de SLS synonymes (les fonctions lexicales sont imprimées en caractères gras) :



(Nous avons le devoir de mettre d'abord en évidence [le fait que] M. Gardin a le mérite d'avoir détérré cet objet-là.)



(Il nous faut rendre d'abord évident
[le fait que] le mérite d'avoir dé-
terré cet objet-là revient à
M. Gardin.)



(Nous devons commencer par mettre en
évidence / le fait que / le mérite
d'avoir déterré cet objet-là revient
à M. Gardin.)

Dans l'ensemble des SLS synonymes, la SLS la plus simple et universelle (ces notions ne sont pas expliquées ici) est choisie pour représenter cet ensemble. Cette SLS est appelée Basic-SLS (sur son rôle, v. 7).

4. Une fonction lexicale est une dépendance sémantique entre un mot-clé (l'argument de la fonction) et un autre mot qu'on appellera le corrélat lexical (la valeur de la fonction).

Ainsi :

<u>Fonction</u>	<u>Argument</u>	<u>Valeur</u>
<u>Conv</u> (l'inversement des actants)	précéder	suivre
<u>Adv</u> (l'adverbe dérivé)	falloir	devoir
<u>So</u> (le nom dérivé)	commencer	d'abord
<u>Oper₁</u> ("faire", "être l'agent de")	déterrer	déterrement
<u>Func₁</u> ("être l'action ou la propriété de")	<u>Pred</u> (évident)	évidence
<u>Pred</u> ("être + ...")	tomber	chute
<u>Caus</u> ("créer", "faire avoir lieu")	coup	porter
<u>Magn</u> ("très", "un grand degré de")	accueil	réserver
	devoir, mérite	avoir
	<u>S₀</u> (<u>Pred</u> (évident))	être de [toute évidence]
	défaite	essuyer
	mérite	revenir à
	peur	posséder, gagner
	porteur	porter
	liste	dresser
	symphonie	composer
	<u>Func₁</u> (peur)	faire à
	<u>Pred</u> (évident)	rendre
	<u>Oper₁</u> (<u>S₀</u> (<u>Pred</u> (évident)))	mettre en [évidence]
	peur	grande, extrême, horrible, terrible
	défaite	entière, complète, grande
	mérite	grand, éminent, supérieur
	besoin	grand, extrême, urgent, pressant

Jusqu'à présent une trentaine de fonctions lexicales ont été dégagées. Elles forment deux classes :

- a) Les substituts - mots ou locutions susceptibles d'être employés

dans le texte à la place du mot-clé, cette substitution n'exigeant que des modifications purement syntaxiques visant à sauvegarder le sens original (Conv, Adv, S_o dans la liste des exemples ci-dessus) ;

b) Les paramètres sémantiques - mots ou locutions susceptibles d'être employés avec le mot-clé pour exprimer d'une façon idiomatique certains sens ("très", "faire", "créer" etc. dans la liste ci-dessus).

Les valeurs de toutes les fonctions lexicales sont indiquées aux articles (mots-clés) respectifs d'un dictionnaire spécial ce qui permet une expression idiomatique des sens.

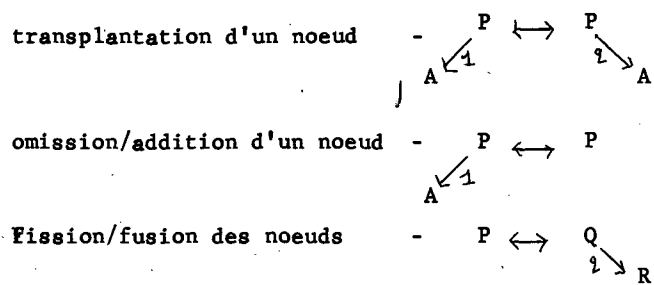
5. Les transformations réciproques des SLS sont effectuées par deux systèmes de règles :

a) Règles lexicales (une cinquantaine) qui formulent des équivalences sémantiques en termes de fonctions lexicales. Ainsi, les transitions entre les SLS (1)-(3) sont fondées sur les règles suivantes :

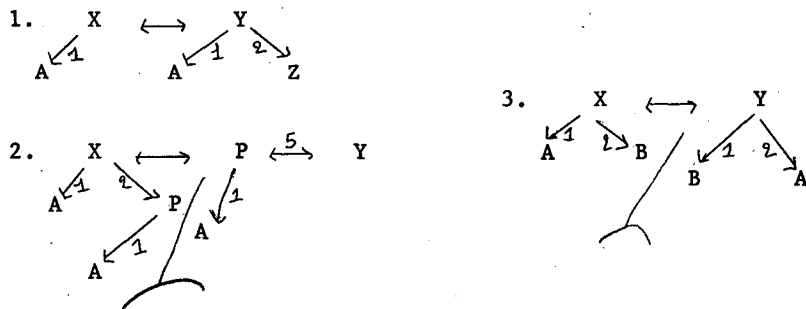
	X	Y	Z		
1.	i_o (mot-clé)	\leftrightarrow <u>Oper₁</u>	<u>S_o</u> (i_o)	(1)	Nous accueillons... - Nous réservons un accueil à...
2.	i_o	\leftrightarrow <u>Adv</u> (i_o)	-	(2)	Nous commençons par écrire ... - D'abord nous écrivons...
3.	i_o	\leftrightarrow <u>Conv</u> (i_o)	-	(3)	Il nous faut lire... - Nous devons lire...
4.	<u>Oper₁</u>	\leftrightarrow <u>Func₁</u>	-	(3)	M. Gardin a le mérite... - Le mérite... revient à M. Gardin...

b) Règles syntaxiques (une vingtaine) qui ne sont en fait que les transformations possibles définies sur des arborescentes ^{ou} (sous-arborescentes) "profondes". Déclenchée par une règle lexicale à laquelle elle est rattachée

(v. les numéros dans la dernière colonne des règles lexicales), une règle syntaxique sert à remédier aux perturbations produites par les changements de vocabulaire dus à la dite règle lexicale. Toute règle syntaxique est une combinaison de trois opérations-type primitives :



Les règles syntaxiques qui tiennent aux règles lexicales 1 - 4 sont :



6. Les règles mentionnées forment une espèce de générateur qui produit pour une SLS d'entrée le maximum des SLS synonymes. Ce générateur est couplé avec un sélecteur qui n'admet à la sortie que des SLS représentant des phrases tout à fait grammaticales. Le sélecteur contrôle les résultats de l'application des transformations synonymiques (formulées comme valables universellement) dans des cas concrets. Il consiste en un système assez ramifié de filtres - contraintes, exigences, limitations de toute sorte etc. - disposés ça et là dans le système de périphrases et associés à des unités de dictionnaire, aux classes de mots etc. Les filtres visent l'absence d'une forme grammaticale d'un mot, le régime fort des verbes (tel complément est indispensable ou, au contraire, incompatible avec tel verbe), l'impossibilité d'une telle classe lexicale dans une telle construction etc. Les filtres sont absolument indépendants

des règles et excluent certains emplois d'une unité, quelle que soit la règle qui pourrait conduire à ces emplois.

7. Appliqué au traitement automatique de textes, le MST est en mesure de pourvoir à l'incongruence des structures des phrases des langues diverses, incongruence dont les recettes ad hoc ne sont qu'un remède de fortune. Le MST peut être incorporé dans un système, disons, de traduction automatique, de la façon suivante : la phrase de la langue-source doit être ramenée à sa SLS ; le système des périphrases transformera cette SLS dans la Basic-SLS correspondante ; celle-ci sera remplacée par la Basic-SLS (de la langue-cible) portant le même sens - ce qui doit être beaucoup plus facile que la traduction directe au niveau des phrases ; la BSL ainsi obtenue entrera, à son tour, dans le système des périphrases (de la langue - cible) qui va générer pour elle, une à une, les SLS synonymes jusqu'à ce qu'on aboutisse à une SLS acceptée par tous les filtres ; cette SLS sera développée en phrase par les mécanismes suivants du système. On peut espérer que le processus esquissé ici permettra d'éviter les discrédances interlinguistiques d'une manière naturelle et produire une traduction exacte, souple et assez idiomatique.

Le système des périphrases pour le russe, programmé par l'ordinateur "Ural-4" (par N. G. Arsentjeva et al.), est en train d'être expérimenté.

8. Quelques conclusions théoriques :

a) Le problème du sens est traité d'une façon opérationnelle, le sens n'étant dans le MST que l'invariante des transformations synonymiques.

b) L'opposition "profond - superficiel" est étendue au vocabulaire (fonctions lexicales vs. leurs valeurs verbales au niveau superficiel).

c) Les fonctions lexicales contribuent à la description systématique du vocabulaire révélant des proportions du type protestation : vigoureuse = défaite défaite : complète = brouillard : épais = dormir : comme un plomb, à poings fermés etc. ou avoir : mérite = porter : coup = faire : pas = prêter, porter : aide etc. (cf. /p/ : /b/ = /k/ : /g/ = /t/ : /d/...).

d) Un dictionnaire de type nouveau devient le pivot de la description linguistique. Chacun de ses articles doit contenir une liste de valeurs (30-50 environ) des fonctions lexicales pour le mot-clé. Les fonctions lexicales peuvent être considérées comme une sorte de questionnaire lexicographique universel.

e) La synonymie apparaît comme un phénomène beaucoup plus répandu qu'on ne le croit d'habitude, si l'on est prêt à la considérer à une échelle d'unités plus grande que celle des mots isolés. La synonymie n'est que l'interchangeabilité aux niveaux profonds du système. Quant à l'antinomie "équivalence vs. interchangeabilité réelle dans un contexte concret", celle-ci est résolue par la stratification du modèle : des expressions résultant des substitutions sans limites à un niveau profond peuvent être exclues par des filtres d'un niveau superficiel.