

THÈSE de DOCTORAT d'ÉTAT

spécialité Informatique,

présentée à l'Université Paris-Nord

par Bernard LEVRAT,

pour obtenir le grade de Docteur ès Sciences,

sur le sujet :

**Le problème du sens dans les systèmes de traitement automatique
du langage naturel : une approche alternative au travers de la
paraphrase**

Soutenue le 26 Février 1993 devant le jury composé de :

MM	Igor Mel'čuk	<i>Président et rapporteur</i>
	Michel Chein	<i>Rapporteur</i>
	Jean-Marie Pierrel	<i>Rapporteur</i>
	Daniel Kayser	
	Guy Chaty	
	Daniel Coulon	
Mme	Catherine Fuchs	
M	Gérard Sabah	

Thèse préparée au Laboratoire d'Informatique de Paris-Nord
Université de Paris-Nord
Avenue Jean-Baptiste Clément
93430 Villetaneuse

Abstract :

Paraphrasis may be an issue to the difficult problem of the *representation*, or more adequately *the capture*, of *meanings* expressed in natural language. Trying to characterize what *kind of entity* is "*sense*" through different paradigms used for lexical definitions, allow us to emphasize crucial, and we think unsolvable, problems in semantics.

The inability to define a relevant *meta-language* for natural language semantics, leads us to search solutions besides traditional ones. Finally *reformulation* seems us to be a good way.

Two experiments are used to illustrate this approach :

- First we sketch a question answering system which uses *rewriting rules* between surface structure of natural language sentences to express inferences.
- Second, a more recent application concern PARAPH a *paraphrasis diagnosis system* to be used where it is the case to *render fine semantics phenomena* such as the meaning of *grammatical categories* (temporality, aspect, modality etc.).

PARAPH is a *conditional tree rewriting system*, in which conditionnal rules implement *elementary paraphrastic transformations*. Conditions labelling rules, represent *choices and approximations* which legitimate paraphrastic transformations they are associated with.

As a first application of PARAPH, we aim at approaching the *modality* expressed by the french modal verb "*pouvoir*", depending on the context of its occurrence. The rule set is to be build from a real corpus collected by psycholinguist researchers.

More general utilisation are expected for PARAPH for it seems to be well suited for applications where a *conditional tree rewriting system* is needed. This is particularly the case for modelization in computational linguistics.

Key-words : Computational linguistics, semantics, semantic representations, paraphrasis, paraphrasis diagnosis, tree conditional rewriting system, modality.

Table des matières

INTRODUCTION	1
LE SENS ET SA REPRÉSENTATION	7
I) UN SENS INDÉPENDANT DU LANGAGE	10
I-1) Les justifications d'origine psychologique.....	10
I-2) Les justifications d'origine linguistique.....	11
I-3) Les justifications d'origine logique	15
II) SENS LANGAGIER ET INFÉRENCE DE SENS COMMUN	20
III) LES TENTATIVES DE REPRÉSENTATION DU SENS	24
III-1) La nature des unités sémantiques	25
III-1-1) la persistance de la langue.....	25
III-1-2) Les approches paraphrastiques :.....	28
III-2) Définitions cycliques ou à base de p.s.?.....	31
III-2-1) Les systèmes à primitives sémantiques	31
III-2-2) Canonicité vs unicité	34
III-3) Le problème de la catégorisation.....	35
III-3-1) Les définitions par C.N.S.....	35
III-3-2) Les théories prototypiques	38
III-3-3) La sémantique du lien de " <i>sorte de</i> ".....	41
IV) NOS CHOIX EN SÉMANTIQUE	44
PARAPHRASE ET REFORMULATION	53
A) LA PARAPHRASE	
I) L'APPROCHE SYNTAXIQUE	57
I-1) Une caractérisation.....	57
I-2) Les opérateurs paraphrastiques.....	60
I-3) Une conclusion	63

II

II) LES APPROCHES SYNTAXICO-SÉMANTIQUES	65
II-1) Paraphrases linguistiques vs pragmatiques.....	65
II-2) Quelle identité sémantique ?.....	68
II-2-1) Le problème de l'identité sémantique	70
II-2-2) Les variations sémantiques acceptables.....	72
II-2-2-a) D'ordre connotatif	72
II-2-2-b) Dans les présupposés.....	73
II-2-2-c) Dans les degrés d'oppositions.....	74
 B) LA REFORMULATION	
I) LA LOGIQUE DU SYSTÈME	75
I-1) PLANES	75
I-2) SAPHIR.....	77
I-3) INTERFRANCE	78
I-4) En conclusion	79
II) UNE REFORMULATION FINALISÉE	79
II-1) CO-OP	79
II-2) POLLY.....	82
III) ENRICHIR LES CONNAISSANCES DE L'UTILISATEUR	85
III-1) INTERIX.....	86
III-2) INTERFACILE.....	87
 C) UNE CONCLUSION	88
I) SUR LA PARAPHRASE EN GÉNÉRAL.....	88
II) SUR LA REFORMULATION :.....	89
 LA THÉORIE SENS-TEXTE	99
I) UNE DESCRIPTION DU MODÈLE SENS TEXTE	101
I-1) La représentation sémantique.....	104
I-1-1) Le graphe sémantique.....	104
I-1-2) L'organisation communicative.....	106
I-1-3) Un exemple.....	107

III

I-2) La représentation syntaxique profonde.....	108
I-2-1) La structure syntaxique profonde.....	108
a) Les dépendances syntaxiques profondes.....	108
b) Les lexèmes généralisés.....	109
c) Un exemple :.....	110
I-3) La composante sémantique	110
I-3-1) Les règles sémantiques	111
I-3-2) Les règles sémantiques de paraphrase.....	112
II) LE SYSTÈME DE PARAPHRASE	113
II-1) La classe des paraphrases traitées.....	113
II-2) Les règles lexicales de paraphrase.....	115
II-2-1) Les fonctions lexicales	115
a) Une typologie des fonctions lexicales :.....	116
b) Quelques problèmes.....	118
II-2-2) Les règles	120
II-2-3) Une typologie des transformations.....	121
A) Les équivalences sémantiques.....	121
1) Les substitutions synonymiques.....	121
2) L'introduction de la négation.	122
3) Les transformations spéciales.....	122
a) Dérivatives adverbiales	122
b) Métaphoriques ou métonymiques.....	123
B) Les règles d'inférences	123
II-2-4) Quelques remarques.....	124
III) DISCUSSION :	124
III-1) Sur la théorie elle-même.....	125
III-1-1) Sur la définition du sens.....	125
III-1-2) Sur la fragmentation des significations.	126
III-1-3) Sur la détermination des primitives.....	127
III-1-4) Sur le choix de la synthèse :.....	128
III-2) Sur le système de paraphrase.....	129
III-2-1) Le point de vue de Mel'čuk	129
a) L'absence de précision sémantique :.....	129
b) L'absence de conditions d'application :.....	130
c) Le champ couvert par le système	131
III-2-2) Des remarques complémentaires.....	131
LE DIAGNOSTIC DE PARAPHRASE	139
I) LES FONDEMENTS LINGUISTIQUES	143

IV

I-1) Un processus interprétatif.....	143
I-2) Une approche polysémique du sens.....	144
I-3) Polysémie et paraphrase.....	146
I-4) Le jugement de paraphrase.....	150
II) LA DESCRIPTION DU SYSTÈME	152
II-1) Les principes	152
II-1-1) Un diagnostic	152
II-1-2) Un corpus réel.....	154
II-1-3) Un exemple de diagnostic	154
II-1-4) Des règles conditionnelles.....	155
II-1-5) Des règles élémentaires.....	156
II-1-6) Des règles très contraintes	156
II-2) La structure générale du système.....	157
II-2-1) Les règles	157
a) La partie gauche	157
b) La partie droite	158
c) Les conditions interprétatives.....	159
d) Les pointeurs.....	160
e) Quelques exemples de règles.....	160
II-2-2) Le filtrage	164
a) Les types de filtrages.....	164
b) L'instanciation des variables.....	164
II-2-3) Le programme.....	164
III) QUELQUES LIGNES DE REFORMULATION	165
III-1) Exemple 1	165
III-2) Exemple 2.....	169
III-3) Exemple 3.....	170
IV) LA CONSTITUTION D'UN JEU DE RÈGLES	172
IV-1) Les sémantismes de " <i>pouvoir</i> "	173
IV-1-1) Les approches atomisantes.....	173
IV-1-2) Les approches unifiantes.....	174
IV-2) Les reformulations de " <i>pouvoir</i> "	174
IV-3) La construction des règles.....	178
V) EN CONCLUSION	183
CONCLUSION	191
BIBLIOGRAPHIE	199

ANNEXES	215
---------------	-----