Communicative Structure of Sentences as a Means of Controlling the Generation of Paraphrases

Jasmina Milićević

The paper deals with the interaction of semantic and communicative information in the generation of paraphrases (= synonymous sentences) within the Meaning-Text theoretical framework. The language considered is French, but the results of the present study have universal implications, i.e. they are applicable to the generation of paraphrases in other languages as well.

Two main questions are addressed: 1) Well-formedness conditions for Semantic Representations (the starting point for the generation of paraphrases), i.e. constraints on combining—or pairing—Semantic and Communicative Structures; 2) Paraphrastic variation induced by variations of the Communicative Structure.

Section 1 (Introduction) shows the importance of paraphrasing in language as well as in formal linguistic models and its applications in Natural Language Processing, in particular in text generation.

Section 2 characterizes the paraphrasing in Meaning-Text linguistic models and introduces the problem of Semantic-Communicative pairings.

Section 3 examines Semantic Representations obtained by pairing one particular Semantic Structure with a series of different Communicative Structures, in an attempt to provide answers to the two above-mentioned questions.

Section 4 (Conclusion) sums up preliminary results.

Keywords: *formal semantics*, *paraphrasing*, *semantic-communicative pairings*.

1 INTRODUCTION

The paper deals with the interaction of semantic and communicative information in the generation of paraphrases within the framework of the Meaning-Text Theory [= MTT].[1]

Roughly speaking, PARAPHRASES are synonymous sentences, i.e. the sentences expressing (nearly) the same semantic content, cf. French sentences in (1):

(1) a. *Pénélope est sûre du retour d'Ulysse*.
'Penelope is sure of Odysseus's return'.

b. *Pénélope ne doute pas qu'Ulysse reviendra*.
'Penelope does not doubt that Odysseus will return'.

c. *Le retour d'Ulysse est une chose sûre pour Pénélope*.
'Odysseus's return is a sure thing [= a certainty] for Penelope'.

The semantic content [= propositional meaning] of these sentences is the same; they differ with respect to 1) the lexical and syntactic means chosen to express the semantic content, and 2) the way the content is 'packaged' [Chafe 1994] by the Speaker (i.e. they reflect his/her different communicative intents, see below). Paraphrases (1a) and (1b) are different in the first respect, while (1c) differs from both (1a) and (1b) in both respects.

The study of paraphrases consists in studying either 1) the PRODUCTION of paraphrases [= PARAPHRASING] from (a representation of) a common semantic content, or 2) the RECOGNITION of paraphrases, i.e. their 'reduction' to (a representation of) a common semantic content. In what follows, I will be concerned exclusively with the production [= synthesis] of paraphrases. For the recognition [= analysis] of paraphrases, see in particular Fuchs 1982, Fuchs et Le Goffic 1982, and Levrat 1993.

The relation of paraphrase, which is a particular case of synonymy (alongside lexical and syntactic synonymy), plays a crucial role in language. The mastery of paraphrasing is an important part of our linguistic competence [Žolkovskij & Mel'čuk 1967, Fuchs 1982], for at least the following three reasons:

- A speaker is capable of saying whether two sentences are paraphrases (in just the same way as s/he is able to judge the grammaticality of a sentence). This capacity is based on the intuitive notion of synonymy, central to all linguistic research.

---

[1] Familiarity with the MTT will be assumed; nevertheless, in order to facilitate the reading of the paper, some key concept will be introduced. For more on the MTT, see, for instance, Mel'čuk 1981, 1988: 43-103 et 1997.

- Lexical knowledge is crucially related to the notion of paraphrasing; thus, the lexicographic definition of a lexical meaning is its paraphrase in terms of other, simpler, meanings.

- Speaking can be thought of as virtual paraphrasing—a process involving multiple choices between possible paraphrases. (As we shall see, this is the viewpoint of the MTT.)

Besides the spontaneous, pre-conscious use of paraphrasing ('virtual paraphrasing'), there is also a deliberate, conscious use thereof, traditionally called reformulation; this is an operation which modifies an existing sentence. Paraphrasing as reformulation has a wide use in text editing, translation (= interlinguistic paraphrasing), as well as in everyday communication (where one often needs to change the style of expression, avoid repetitions, make oneself clear, etc).

The corollary of all this for linguistics is that a linguistic model of a language L should adequately reflect the paraphrasing power of L.

> A linguistic model of L has an ADEQUATE PARAPHRASING POWER if it is able to account for (nearly) all the paraphrastic choices available in L for expressing a given meaning.

In practical terms, this means that a linguistic model of L should include a paraphrasing component—a system of paraphrasing rules.

This fact has by now been widely recognized, both in theoretical and computational linguistics, and also in Natural Language Processing, in particular in text generation, where there has been a growing interest for building text generators having an adequate paraphrasing power and capable of producing high-quality texts (= meeting criteria of naturalness and variability). The paraphrasing provided by such systems can be used either 'full-strength', as in text reformulation/editing systems, or 'restrictedly', i.e. only in order to prevent dead-ends during text generation/translation.

However, few full-scale paraphrasing systems have been proposed so far, fewer still implemented. The reason for this scarcity is the scope and the complexity of the task. The main problems include, for instance:

- Linguistically rich and interesting paraphrasing presupposes that one has to start from a semantic representation, which means that the whole process of text synthesis is involved;

- Multiple mappings are possible between a given meaning and the texts expressing it; cf. the richness and diversity of synonymic means of a language (for an illustration see Mel'čuk 1988: 86-88; see also the Appendix to this paper); this can lead to a 'combinatorial explosion.'

- In order to describe the above mappings, one needs a sophisticated lexicon (the richest source of paraphrases is lexical synonymy; even the syntactic paraphrasing is triggered lexically), and such a lexicon is, of course, difficult to make.

Nevertheless, inspite of these difficulties, there are some paraphrasing systems, based on linguistic approaches in which paraphrasing plays an important role. Two such approaches are Systemic Functional Linguistics (Halliday 1985) and MTT (cf. in particular the paraphrasing system of Mel'čuk and Žolkovskij: Mel'čuk 1992).

For applications based on the ideas of Systemic Linguistics, see, for instance, the Penman Project 1989, Bateman & Wanner 1990 and Elhadad *et al.* 1997.

MTT-based applications include, for instance, paraphrasing systems incorporated into English/French text generators LFS (Iordanskaja *et al*. 1992) and RTS (CoGenTex 1992),[2] a large-coverage paraphrasing system for Russian (Apresjan & Cinman, 1999) and a reformulation system for (a restricted domain of) French (Nasr 1996).

This successful practical use of MTT ideas reflects the fact that paraphrasing is the core part of this theory; that is why it has been chosen here as the best possible frame of reference.

MTT identifies the following four sources of paraphrastic variation (all of which can of course be combined in the production of paraphrases):

- SEMANTIC VARIATION, i.e. (quasi-)equivalencies between (configurations of) meanings, e.g. ‘Penelope **believes** that Odysseus will return’ ≈ ‘**According to** Penelope, Odysseus will return’;

- LEXICAL VARIATION, i.e. different lexical expressions of a given meaning, e.g. *Penelope is **sure** <**certain**, **convinced**> that Odysseus will return*;

[2] For implementation details/problems, see Iordanskaja *et al*., 1996.

- SYNTACTIC VARIATION, i.e. different syntactic expressions of a given meaning, e.g. *The person **sure** <**who is sure**> of the return of Odysseus is Penelope* ;

- MORPHOLOGICAL VARIATION, i.e. different morphological expressions of a given meaning, e.g. Serbian *Penelopa veruje u povratak **svog** <**svoga**, **svojeg**, **svojega**> supruga* ‘P. believes in the return of her husband’.[3]

The system of rules that sets out to model the paraphrastic power of a language has to deal with two complementary problems: 1) the necessity to account for ALL available paraphrastic choices, thus achieving an adequate paraphrasing power; 2) the necessity to guide (= additionally constrain) paraphrastic choices, thus achieving the ability to produce ONLY THOSE paraphrases which are good in a particular context.

It is in connection with the second problem that comes into play the communicative orientation of the sentence—formally, the Communicative Structure: what is Rheme/Theme, what is Presupposed/Asserted, Focalized/Neutral, etc. (For more detail, see infra.) The Speaker, by specifying these communicative values, filters out the paraphrases which do not suit his/her specific communicative goals.

A straightforward illustration of how this constraint works can be found in translation (= interlingustic paraphrasing). Two examples will show how the communicative structure 'guides' the translation of two Serbian sentences into English by imposing lexical and syntactic changes in the target sentences. More specifically, they illustrate the use of one particular paraphrasing strategy: conversion; example (2) illustrates grammatical conversion (= passivization), and (3) is an illustration of lexical conversion. Both examples involve a basic communicative opposition between Theme [≈ Topic] and Rheme [≈ Comment] of a sentence. Bracketed Roman numerals indicate the Deep-Syntactic actants of the Main Verb.

(2) a. *Najveći* *deo* [= II] *sredstava*
biggest-MASC.ACC.SG part-ACC.SG ressources-GEN.PL

obezbedila *je* *država* [= I].
provide-PART.FEM.SG be$_{[aux]}$-PRES.3SG state-NOM.SG

lit. ‘The biggest part [= II] of the ressources provided the state [= I]’.

[3] In Serbian, the possessive adjectives *moj* ‘my’, *tvoj* ‘thy’, *njegov* ‘his’, and *svoj* ‘one's’ have four forms in the oblique cases of the singular, freely interchangeable in most contexts.

b. *The biggest part* [= I] *of the ressources has been provided by the state* [= II].

In (2a), the Direct Object—*deo* '[the] part'—, which corresponds to the second syntactic actant of the Main Verb [= MV], expresses the Theme and the Subject— *država* '[the] state'—, corresponding to the first actant of the MV, expresses the focal part of the Rheme. In the English translation, we would like to have the same communicative structure, i.e. *part* expressing the Theme, and *state* expressing the Rheme focus. Now, in English, the Theme is always expressed in the first position in the sentence, and it is only the syntactic Subject, i.e. the element corresponding to the first actant of the MV, which is allowed in this position. But, the first actant of *provide* is *state*, and not *part* (cf. the litteral gloss); therefore, in order to be able to use this verb here, we have to passivize it, and thus 'promote' *part* into its first actant, which will eventually surface as the syntactic Subject (2b).

(3) a. *Njegovoj* *porodici* [= II] *pripada*
his-FEM.DAT.SG family-DAT.SG belong-PRES.IND.3SG

nekoliko *kuća* [= I] *u* *gradu.*
several house-GEN.PL in city

lit. 'To his family [= II] belong several houses [=I] in city'.

b. *His family* [= I] *owns several houses* [= II] *in the city*.

(3b) uses the same strategy as (2b), this time with lexical means: the lexeme BELONG, the equivalent of Serbian PRIPADATI, cannot be used in this particular communicative configuration (with the Theme of the sentence expressed by an Indirect Object NP), cf. the literal gloss; the lexeme OWN, a lexical conversive of BELONG, has to be used instead.

2 PARAPHRASING IN MEANING-TEXT MODELS

As I have already said, in the MTT, the paraphrase has a place of honor. This theory considers natural language as a correspondence between meanings and texts, i.e. as a set of rules which map a given linguistic meaning to all texts capable of expressing it (and *vice versa*). To put it differently, MTT considers language to be a paraphrasing mechanism (cf. virtual paraphrasing, p. 2) and proposes, as a main tool for describing it, a Meaning-Text Model [= MTM].

An MTM is a synthesis-oriented functional model of language; it presupposes several levels of sentence representation and consists of sets of rules, or modules, which establish the correspondence between representations of adjacent levels.

The starting point for text synthesis in an MTM, the Semantic Representation, encodes a semantic/communicative content to be expressed (alternatively) by a set of synonymous output sentences, i.e. paraphrases. At present, the paraphrasing system of the model works only with the representation of the semantic content, out of which it can produce all Syntactic Representations which underlie the future paraphrases. However, to control the proliferation of paraphrases in the context of text generation, the model needs more stringent conditions on how the choice of communicative information blocks the production of whole sets of paraphrases. In the rest of this paper, I will consider one specific aspect of this task: a theoretical/formal problem of combining—or PAIRING—semantic and communicative structures. In other words, I will try to establish what communicative structures can be paired with what semantic structures, i.e. to identify constraints that underlie such pairings. In order to be able to state the problem in more formal terms, I first have to characterize formally the Semantic Representation [= SemR].

A SemR consists of the following three structures: Semantic Structure [= SemS] represents the propositional meaning of sentence(s) to be synthesized; Semantic-Communicative Structure [= Sem-CommS] encodes the communicative intent of the Speaker; Rhetorical Structure—which won't be considered in this paper—encodes the stylistic intent of the Speaker (should his/her text be neutral, formal, colloquial, poetic, etc.).

A SemS is a network whose nodes are labeled with semantemes (= lexical meanings of the language) and whose arcs are labeled with numbers indicating predicate/argument relations (cf. the reference SemS in **3.1** below).

A Sem-CommS is a division of the SemS into communicative areas, or subnetworks, marked by one of mutually exclusive values of one of eight communicative oppositions, such as {Rheme, Theme}, {Given, New}, {Asserted, Presupposed}, {Focalized, Neutral}, etc.[4] In each communicative area, one node is singled out as COMMUNICATIVELY DOMINANT: it is the node that 'sums up' the meaning of the area, i.e. the node to which the entire area can be semantically reduced.[5]

---

[4] For the full inventory of communicative oppositions used within the MTT framework, see Mel'čuk (forthcoming).

[5] Thus, the subnetwork 'Odysseus'←1—'return' (= *Odysseus's return*) can be reduced to 'return', whereas the reduction of the subnetwork 'Odysseus'←1—

A given SemS can be paired with several different Sem-CommSs, which gives rise to a series of SemRs:

$$\text{SemR}_1 : \langle \text{SemS}_1 + \text{Sem-CommS}_1 \rangle$$
$$\text{SemR}_2 : \langle \text{SemS}_1 + \text{Sem-CommS}_2 \rangle$$

...
$$\text{SemR}_n : \langle \text{SemS}_1 + \text{Sem-CommS}_n \rangle$$

Two questions can be asked at this point: 1) What are the constraints on SemS ~ Sem-CommS pairing, i.e. what are the corresponding well-formedness rules for Semantic Representations? 2) What kind of paraphrastic variation is induced as a function of varying communicative parameters of a given SemS? These questions have been probed in a controlled experiment, described in Section 3.

3 ASSOCIATING SEMANTIC AND COMMUNICATIVE STRUCTURES IN A MEANING-TEXT MODEL

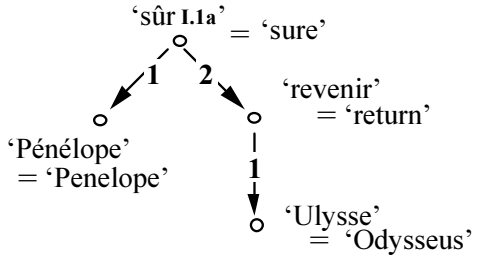
3.1 The notion of well-formed vs. ill-formed SemRs

I have examined fifteen SemRs (of French sentences), obtained by pairing the reference SemS[6] (opposite) with fifteen different Sem-CommSs logically possible for it. Only basic Sem-CommSs were considered, i.e. the Sem-CommSs consisting exclusively of a rhemo-thematic division, without any additional communicative marking, such as Focalization, Emphasis, etc.[7] (Numerical expressions accompanying French words are lexicographic numbers used to distinguish word-senses.)

'return' (= *Odysseus who returns*) gives 'Odysseus'. (The communicatively dominant node is underscored.)

[6] All representations given in this paper are simplified: inflectional meanings are not represented, since they are irrelevant for my present purposes. I will assume the future reading of the reference SemS, along the lines of *Pénélope est sûre***I.1a** *qu'Ulysse reviendra* 'Penelope is sure that Odysseus will return'.

[7] The number of binary partitions for a set of **n** elements is $2^{\mathbf{n}-1}$. Thus, for a semantic structure with 4 nodes, like ours, there are eight possible partitions into two communicative areas (Theme *vs.* Rheme): seven partitions allowing each two (inverse) distributions of the Theme and the Rheme (7 x 2 = 14), and one purely rhematic partition. If in each partition we consider only one choice of dominant nodes, we get a total of fifteen Sem-CommSs, which, superposed upon the reference SemS, yield fifteen different SemRs.

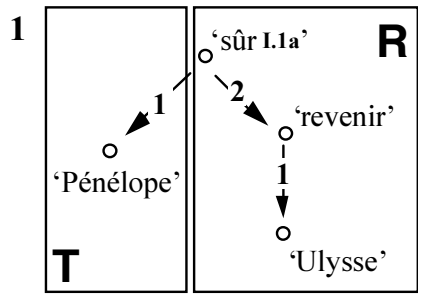


It was assumed that any well-formed SemR is implementable by grammatical sentences (and grammatical sentences only), given the appropriate rules. Setting aside the question of rule appropriateness until 3.2, let me quote two examples of well-formed SemRs, with their respective realizations.

Q = underlying question, providing a minimal context for a given SemR, which is implemented by a set of paraphrases; **R** = rheme; **T** = theme; $\mathbf{t_n}$ = sub-theme; communicatively dominant nodes are underscored; | = a pause.

Lexemes SÛR**I.1a** ('X is sure of Y') and SÛR**I.2a** ('Y is sure for X') are conversives realizing the same underlying semanteme, 'sûr**I.1a**'. Lexemes CERTITUDE**1** and CERTITUDE**2**, conversives themselves, are approximate nominalizations of SÛR**I.1a** and SÛR**I.2a**, respectively.

Q *Et Pénélope ?* ‘What about P.?’



Realizations:

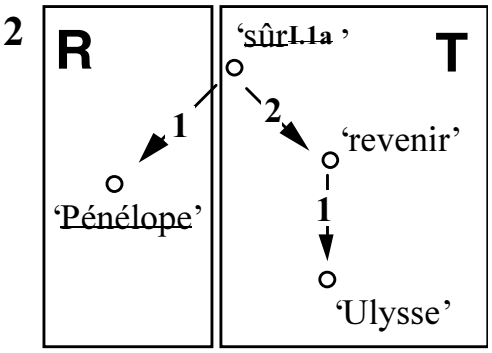
a. *P.* | *est sûre***I.1a** *qu'U. reviendra.* [8]
‘P. is sure that O. will return’

b. *P.* | *est sûre***I.1a** *du retour d'U.*
‘P. is sure of O.'s return’

c. *P.* | *a la certitude***1** *qu'U. reviendra.*
‘P. has the certainty that O. will return’

[8] Pauses are indicated for a later comparison with the SemR [6].

Q *Pour qui le retour d'U. est-il sûr***I.2a** *?*
'For whom is O.'s return certain?'

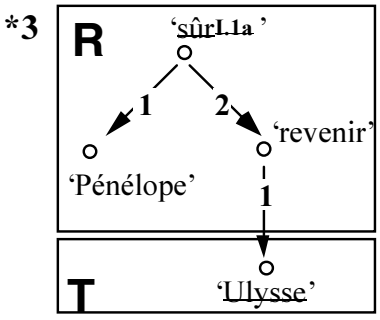


Realizations:

a. *Le retour d'U. est sûr***I.2a** *pour P.*
'O.'s return is certain for P.'
b. *Qu'U. reviendra est sûr***I.2a** *pour P.*
'That O. will return is certain for P.'
c. *Le retour d'U. est une certitude***2** *pour P.*
'O.'s return is a certainty for P.'
d. *Qu'U. reviendra est une certitude***2** *pour P.*
'That O. will return is a certainty for P.'

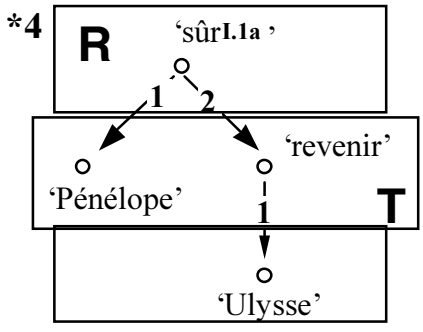
As for ill-formed SemRs, it was assumed that they either have no linguistic realization at all or have only ungrammatical realizations. Here are two examples of ill-formed SemRs:

Q *Et à propos d'U. ?*
'What about O.?'



No realization

Q **Et à propos de P. et du retour de qui ?*
'What about P. and the return of who?'



No realization

The well-formedness of SemRs can be captured by the following formal requirements:

- The dominant nodes of the Rheme and the Theme of a Sem-CommS must be directly connected by a semantic dependency [Polguère 1990]. This rules out the SemRs of type [*3].

- A Sem-Comm should not contain discontinuous areas—a Rheme or a Theme in which not all nodes are connected by semantic dependencies (which are equally included in the area). This rules out the SemRs of type [*4].

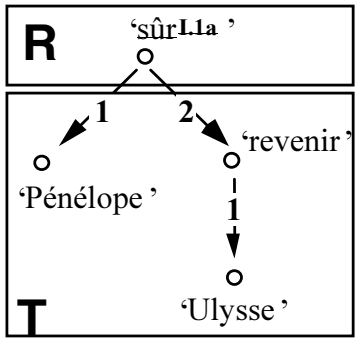
  A discontinuous communicative area is problematic in that it does not allow for a selection of a communicatively dominant node, i.e. cannot be semantically reduced to a single node (cf. p. 6). Thus, the discontinuous Theme of the SemR [*4] consists of two distinct semantic entities (‘Pénélope’ and ‘revenir’), which cannot be reduced in such a way (cf. the incongruous underlying question). The same considerations hold for the discontinuous Rheme of [*4].

Ten SemRs out of fifteen I examined turned out to be ill-formed because they violate these requirements.

However, a special case of discontinuity of Sem-Comm areas, involving discontinuous Themes, must be noted.

Q *Et à propos de P. et du retour d'U. ?*
‘What about P. and O.'s return?’
Et à propos du retour d'U. et de P. ?
‘What about O.'s return and P.?’

*5

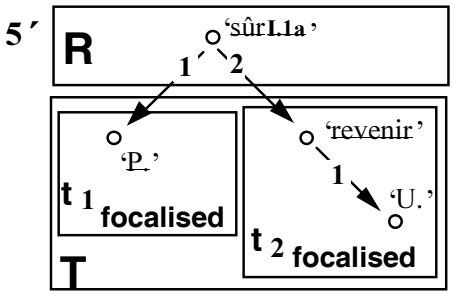


No realization

The SemR [5] is also ill-formed—it contains a discontinuous Theme (cf. two underlying questions)—and has no realization. Nevertheless, with additional communicative marking—specification of 'Sub-Themes' and Focalization—of

the ill-formed SemR [5], it is possible to obtain two well-formed representations, which can be implemented as sentences with dislocated Themes and resumptive clitic pronouns, frequent in colloquial French:

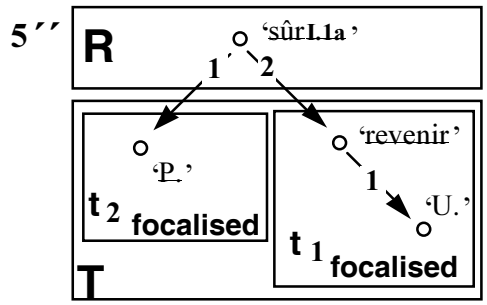
Q *Et à propos de P. et du retour d'U. ?*
'What about P. and O.'s return?'



Realization:

*P., le retour d'U., elle en est sûre***I.1a**.
'P., O.'s return, she is sure of it.'

Q *Et à propos du retour d'U. et de P. ?*
'What about O.'s return and P.?'



Realization:

*Le retour d'U., P., elle en est sûre***I.1a**.
'O.'s return, P. she is sure of it.'

This seems to indicate that at least some pairings giving rise to discontinuous communicative areas should be considered (i.e. should not be discarded outright).

The kind of 'repair' illustrated above does not work with all ill-formed SemRs. Thus, no additional communicative marking can salvage the ill-formed SemR [*4], p. 8, turning it into a well-formed representation.

3. 2 Constraining lexicalization rules

As mentioned above, a well-formed SemR cannot give rise to ungrammatical sentences unless the rules used to produce them are 'defective' in one way or another. I will now illustrate a case of ungrammaticality due to some insufficiently constrained lexicalization rules (belonging to a subcomponent of the semantic module of an MTM, which maps a given SemR to corresponding Deep-Syntactic Representations).

The available lexicalization rules cannot, in their present form, pre-empt the realization of the well-formed SemR [2], given on p. 7, by the following ungrammatical sentences:

a. **La certitude***1** *du retour d'U. est à P.* 'The certainty of O.'s return belongs to P.'
b. **La certitude***1** *qu'U. reviendra est à P.* 'The certainty that O. will return belongs to P.'

The ungrammaticality of these sentences may be due to the fact that the semanteme 'sûr**I.1a**' is a mental-state predicate. Presumably, if a mental-state predicate is found in a SemR of the above type, where the predicate itself is in the Theme and its first actant, i.e. Experiencer, in the Rheme, it cannot be lexicalized by a noun. Cf. ungrammatical realizations of SemRs containing mental-state predicates 'espoir' and 'opinion' in this position: **L'espoir* <**L'opinion*> *qu'Ulysse reviendra est à Pénélope* 'The hope <The opinion> that O. will return belongs to P.'. In contrast to mental-state predicates, predicates denoting acts do have nominal realizations when found in a SemR of this type; cf. the correct sentence *La communication concernant le retour d'Ulysse vient de Pénélope* 'The communication concerning O.'s return comes from P.', with the lexeme COMMUNICATION lexicalizing an act-denoting predicate (= 'communiquer').

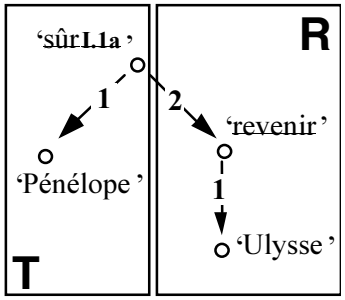
If this analysis is correct, lexicalization rules involving mental-state predicates must be supplemented with a condition banning nominal realization of these predicates from the pairings of the above type.

3. 3 Degree of paraphrastic variation

According to their similarity, the paraphrases produced from the five well-formed SemRs fall into two sets, the 'representatives' of which are sentences (1 a-c) and (6 a-d), on the one hand, and sentences (2 a-d) on the other. The paraphrases belonging to the same set are very close—they exhibit only negligible communicative differences—and should be interchangeable in context. The paraphrases belonging to different sets are more remote—they have an inverse Rheme/Theme distribution—and should be less freely interchangeable.

The SemR [1], p. 7, and the SemR [6] below have syntactically identical realizations, (1 a-c) and (6 a-c), which show only minor prosodic differences (cf. pauses).

6



Q *P. est sûre***I.1a** *de quoi ?*
'What is P. sure of?'

?*Et la certitude***1** *de P. ?*
'What about P.'s certainty?'

Realizations:

a. *P. est sûre***I.1a** | *qu'U. reviendra.*
b. *P. est sûre***I.1a** | *du retour d'U.*
c. *P. a la certitude***1** | *qu'O. reviendra.*

d. *La certitude***1** *de P.* | *porte sur le retour d'U.*
'The certainty of P. concerns O.'s return'.

The possibility of identical realizations of RSems having the same SemS raises the question of whether such RSems could be considered equivalent. Assuming that $SemR_1$ and $SemR_2$ are equivalent if their realizations are interchangeable in all contexts (barring rhetorical differences), my guess (further tests pending) is that the SemR [1] and [6] are equivalent.

However, the following two facts argue against their equivalence:

- The RSem [6] has an additional realization with respect to (1)—sentence (6 d).

- The SemRs [1] and [6] behave differently under Theme Focalization: with a Focalized Theme, the SemR [1] gives *Pénélope, elle est sûre***I.1a** *qu'Ulysse*

reviendra 'P., she is sure that O. will return', whereas the SemR [6] yields the questionable sentence ?*La certitude de Pénélope, elle porte sur le retour d'Ulysse* 'The certainty of P., it concerns O.'s return'.

3. 4 Paraphrastic means involved

The linguistic means involved in the production of paraphrases from the reference SemS include: conversion (e.g. (1 a) *vs.* (2 a)), nominalization (e.g. (1 a) *vs.* (1 c)), and different realizations of the government pattern (= sub-categorization frame) of a given lexeme (e.g., (1 a) *vs.* (1 b)). Among those, only conversion seems to be triggered by communicative factors (Rheme/Theme inversion), while in two other cases we are dealing with free variation of the expression of propositional meaning.

4 CONCLUSION

Well-formedness criteria for SemRs have been proposed (p. 8) and four related problems identified: 1) status and formal treatment of discontinuous communicative areas, 2) semantically/communicatively motivated constraints on lexicalization rules, 3) communicative equivalence of SemRs having (near-)identical realizations, and 4) linguistic means used in paraphrasing.

The role of Sem-CommS variation in the production of paraphrases has been considered; this variation has been linked to a single communicative opposition (Rheme/Theme distribution options). Future work should focus on 1) the role of other communicative oppositions in guiding the production of paraphrases, and 2) the ways in which different communicative oppositions interact in the process.

APPENDIX

This Appendix contains some more (by no means all!) paraphrases that can be generated from the reference SemS (and from (quasi-)equivalent SemSs).

Besides paraphrases obtained by varying only the Sem-CommS (of the reference SemS), already seen in Section 3, I cite here the paraphrases obtained by different lexicalizations of the SemS, some of which are triggered by communicative factors, and some occur independently of them. (These differences will not be indicated in the illustrations below.)

The astonishingly high number of paraphrases that can be generated from a 'minimal' SemS, such as the reference SemS, manifests convincingly the paraphrasing power of natural language.

1

P.	*est* ‘is’	*sûre* ‘sure’ *certaine* ‘certain’ *convaincue* ‘convinced’	*qu'U. reviendra.* ‘that O. will return’ *du retour d'O.* ‘of the return of O.’
	ne doute pas ‘does not doubt’		

2

P.	*a* ‘has’	*la certitude* ‘the certainty’	*qu'U. reviendra.* ‘that O. will return.’ *du retour d'U.* ‘of the return of O.’ *quant au retour d'U.* ‘concerning the return of O.’
	n'a aucun ‘has no’	*doute* ‘doubt’	

3

P.	*croit* ‘believes’ *pense* ‘thinks’	*qu'U. reviendra* ‘that O. will return’	*Sûrement* ‘surely.’ *à coup sûr.* ‘for sure.’ *Certainement* ‘certainly.’

4

P., *Quant à P.,* ‘As for P.,’	*elle* ‘she’	*est* ‘is’	*sûre* ‘sure’ *certaine* ‘certain’ *convaincue* ‘convinced’	*qu'U. reviendra.* ‘that O. will return’ *du retour d'U.* ‘of the return of O.’
		a ‘has’	*la certitude* ‘the certainty’	
		ne doute pas ‘does not doubt’		

5

Pour ‘For’ *Selon* ‘According to’ *D'après* ‘According to’	*P.,*	*le retour d'U.* ‘the return of O.’	*est* ‘is’	*sûr.* ‘sure’ *une certitude.* ‘a certainty’ *un fait certain.* ‘a fact’ *une chose sûre.* ‘a sure thing’
			n'est pas ‘is not’	*douteux.* ‘doubtful’

6

Pour 'For' *Selon* 'According to' *D'après* 'According to'	*P.,*	*(il est)* 'it is'	*sûr* 'sure' *certain* 'certain'	*qu'U. reviendra.* 'that O. will return'
		(il n'y a) aucun '(there is) no'	*doute* 'doubt'	

7

PÉNELOPE 'PENELOPE'	*est* 'is'	*sûre* 'sure' *certaine* 'certain' *convaincue* 'convinced'	*qu'U. reviendra.* 'that O. will return' *du retour d'U.* 'of the return of O.'
C'est Pénélope 'It is Penelope'	*qui est* 'who is'		
	qui ne doute pas 'who does not doubt'		

8

La personne 'The person'	*qui est* 'who is'	*sûre* 'sure' *certaine* 'certain' *convaincue* 'convinced'	*qu'U. reviendra* 'that O. will return' *du retour d'U.* 'of the return of O.'	*est P.* 'is P.'
	qui ne doute pas 'who does not doubt'			

9

Le retour d'U. 'The return of O.' *Qu'U. reviendra* 'That O. will return'	*est* 'is'	*sûr* 'sure' *une certitude* 'a certainty' *un fait certain* 'a fact' *une chose sûre* 'a sure thing'	*pour* 'for' *d'après* 'according to' *selon* 'according to'	*P.*
	n'est pas 'is not'	*douteux* 'doubtful'		

10

Le retour d'U., 'The return of O.'	*P.*	*en* 'of it'	*est* 'is'	*Sûre.* 'sure.' *Certaine* 'certain.' *Convaincue* 'convinced.'
			a 'has'	*la certitude.* 'the certainty.'

			n'en doute pas. 'of it does not doubt.'		

11

U., *Quant à U.,* 'As for O.'	*son retour* 'his return'	*est* 'is'	*sûr* 'sure' *une certitude* 'a certainty' *un fait certain* 'a fact' *une chose sûre* 'a sure thing'	*Pour* 'for' *d'après* 'according to' *selon* 'according to'	*P.*
		n'est pas 'is not'	*douteux* 'doubtful'		

Some more:

(1) *Il est sûr pour Pénélope qu'Ulysse reviendra*, lit. 'It is certain for Penelope that Odysseus will return';
(2) *Sûr pour Pénélope qu'Ulysse reviendra*, lit. 'Certain for Penelope that Odysseus will return';
(**3**) *Le retour d'Ulysse ne soulève chez Pénélope aucun doute*, lit. 'Odysseus's return does not arouse any doubt in Penelope';
(**4**) *Le retour d'Ulysse est ce dont Pénélope est sûre*, lit. 'Odysseus's return is what Penelope is sure of'; (**5**) *La certitude de Pénélope est qu'Ulysse reviendra*, lit. 'The certainty of Penelope is that Odysseus will return';
(**6**) *La certitude de Pénélope porte sur le retour d'Ulysse*, lit. 'The certainty of Penelope concerns Odysseus's return';
(**7**) *La certitude du retour d'Ulysse, c'est Pénélope qui l'a*, lit. 'The certainty of Odysseus's return, it is Penelope who has it';
(**8**) *Ulysse reviendra : c'est ce dont Pénélope est sûre*, lit. 'Odysseus will return: this is what Penelope is sure of';
(**9**) *Ce dont Pénélope est sûre, c'est qu'Ulysse reviendra*, lit. 'What Penelope is sure of is that Odysseus will return';
(**10**) *La chose sûre, pour Pénélope, c'est qu'Ulysse reviendra*, lit. 'Sure thing, for Penelope, is that Odysseus will return';
(**11**) *Ulysse, Pénélope est sûre à propos de son retour*, lit. 'Odysseus, Penelope is sure about his return';
(**12**) *Ulysse, Pénélope est sûre qu'il reviendra*, lit. 'Odysseus, Penelope is sure that he will return';
(**13**) *Le retour d'Ulysse, Pénélope, elle en est sûre*, lit. 'The retun of Odysseus, Penelope is sure about it';
(**14**) *Pénélope ne remet pas en question le retour d'Ulysse*, lit. 'Penelope does not question Odysseus's return';
(**15**) *Pour Pénélope, il n'est pas probable qu'Ulysse ne revienne pas*, lit. 'For Penelope it is not probable that Odysseus will not return';

(**16**) *D'après Pénélope, il n'y a pas lieu de douter du retour d'Ulysse*, lit. ‘According to Penelope, there is no reason to doubt about Odysseus's return’;
(**17**) *À ce que Pénélope croit, Ulysse ne manquera pas de revenir*, lit ‘According to what Penelope believes, Odysseus will not fail to return’;
(**18**) *Pénélope SAIT qu'Ulysse reviendra*, lit. ‘Penelope KNOWS that Odysseus will return’;
(**19**) *Pénélope est sûre du retour d'ULYSSE*, lit. ‘Penelope is sure of ODYSSEUS' return’;
(**20**) *Ulysse est la personne dont Pénélope est sûre du retour*, lit. ‘Odysseus is the person about whose return Penelope is sure’;
(**21**) *La personne dont le retour est sûr pour Pénélope est Ulysse*, lit. ‘The person about whose return Penelope is sure is Odysseus’.

REFERENCES

APRESJAN, Y. and CINMAN, L. (1999), “Perifrazirovanie na kompjutere”. In : *Semiotika i informatika* 36, Moskva, pp. 177-202.

BATEMAN, J. and WANNER, L. (1990), “Towards a Lexicon for German Organized by Communicative Function: An Application of Lexical Functions”. In: *Proceedings of the 14th. German Workshop on Artificial Intelligence* (GWAI), September 1990.

CHAFE, W. (1994), *Discourse, Consciousness*, *and Time*, Chicago/London: The University of Chicago Press.

CoGenTex (1992), *Bilingual Text Synthesis System for Statistics Canada Database Reports: Design of Retail Trade Statistics (RTS) Prototype*. Technical Report 8, CoGenTex, Montreal.

ELHADAD, M., McKEOWN, K. and ROBIN, J. (1996) : “Floating Constraints in Lexical Choice”. In : *Computational Linguistics*, Vol. 22, nº 4, pp. 1-45.

FUCHS, C. (1982), *La paraphrase*, Paris : Presses Universitaires de France.

FUCHS, C. et LE GOFFIC, P. (1982), “Esquisse d'une approche énonciative des liens entre ambiguïté et paraphrase”. In: *Linguistische Berichte 77/82*, pp. 32-45.

HALLIDAY, M.A.K. (1985), *An Introduction to Functional Grammar*, London: Edward Arnold.

IORDANSKAJA, L., KITTREDGE, R. and POLGUÈRE, A. (1991), “Lexical Selection and Paraphrase in a Meaning-Text Generation Model”. In: Paris, C. L., Swartout, W. R. et Mann, W. C. (eds.), *Natural Language Generation*

in Artificial Intelligence and Computational Linguistics, Boston: Kluwer, pp. 293-312.

IORDANSKAJA, L., KIM, M., KITTREDGE, R., LAVOIE, B. and POLGUÈRE, A. (1992), "Generation of Extended Bilingual Statistical Reports". In: *COLING-92*, Nantes, pp. 1019-1022.

IORDANSKAJA, L., KIM, M. and POLGUÈRE, A. (1996), "Some Procedural Problems in the Implementation of Lexical Functions for Text Generaton". In: Wanner, L. (ed.), *Lexical Functions in Lexicography and natural Language Processing*, Amsterdam: Benjamins, pp. 279-297.

LEVRAT, B. (1993), *Le problème du sens dans le systèmes de traitement automatique du langage naturel : une approche alternative au travers de la paraphrase*, thèse de doctorat d'état (spécialité informatique), Paris: Université Paris-Nord.

MEL'ČUK, I. (1981), "Meaning-Text Models: A Recent Trend in Soviet Linguistics", *Annual Review of Anthropology 10*, pp. 27-62.

MEL'ČUK, I. (1988), *Dependency syntax: Theory and Practice*, Albany: State University of New York Press.

MEL'ČUK, I. (1992), "Paraphrase et lexique: la théorie Sens-Texte et le *Dictionnaire explicatif et combinatoire*". In: Mel'ãuk, I., Arbatchewsky-Jumarie, N., Iordanskaja, L. et Mantha, S., *Dictionnaire explicatif et combinatoire du français contemporain*. Recherches lexico-sémantiques III, Montréal: Presses de l'Université de Montréal, pp. 9-59.

MEL'ČUK, I. (1997), *Vers une linguistique Sens-Texte*. *Leçon inaugurale*, Paris: Collège de France.

MEL'ČUK, I. (forthcoming), *Communicative Organization in Natural Languages*, Amsterdam/Philadelphia: Benjamins.

NASR, A. (1996), *Un modèle de reformulation automatique fondé sur la théorie Sens-Texte: applications aux langages contrôlés*, thèse de doctorat en informatique, Paris: Université Paris 7, UFR informatique.

Penman Project (1989). *PENMAN Documentation: the Primer, the User Guide, the Reference Manual, and the Nigel Manual*. Technical Report, USC/Informations Sciences Institute, Marina del Rey, California.

POLGUÈRE, A. (1990), *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*, Ph.D. thesis, Montréal: Université de Montréal.

ŽOLKOVSKIJ, A. and MEL'ČUK, I. (1967), *O semantiãeskom sinteze*. In: In: *Problemy kibernetiki*, nº 19, pp. 177-238. [French translation "Sur la synthèse sémantique" (1970), *TA Informations*, nº 2, pp. 1-85.]

Jasmina Milićević is a doctoral student at the Observatoire de linguistique Sens-Texte (OLST), Université de Montréal, Département de linguistique et traduction, C. P. 6128, Succ. Centre-Ville, Montréal (Qc), H3C 3J7, Canada. She can be reached at milicevj@ere.umontreal.ca.

I express many thanks to Richard Kittredge, Igor Mel'čuk, Jean-Yves Morin and Alain Polguère, who were kind enough to read and comment upon this paper. I would also like to acknowledge financial support from the *Fonds pour la formation des chercheurs et l'aide à la recherche* (FCAR), received during the time this paper was being written.