

Paraphrases sémantiques approximatives : par quels outils formels les décrire ?

Dans cette communication, je m'intéresserai à un type particulier de paraphrases, illustré en (1) :
[Question sous-jacente : *Est-ce que Max pourrait faire une chose comme ça ?*]

- (1) a. *Il pourrait (bien) faire ça (d'après moi).*
b. *Je m'attendrais à ça de sa part.*
c. *Ça ne m'étonnerait pas de lui.*
d. *Il serait capable de faire ça (, je crois).*
e. *(Faire) un truc pareil, ça serait son genre.*

Faciles à établir pour un locuteur, de telles paraphrases présentent un défi de taille pour le linguiste cherchant à en proposer une description formalisable (qui se prête à des utilisations en Traitement Automatique de la Langue, notamment).

Il n'existe pas de liens lexicaux évidents qui permettraient de relier les paraphrases en (1) par des règles du type relativement bien connu – règles de paraphrasage lexico-syntaxiques, formulées en termes de fonctions lexicales et fonctionnant au niveau syntaxique de représentation des énoncés. Ce sont des paraphrases sémantiques, dont la parenté n'est « visible » qu'avec recours à la décomposition sémantique – c'est-à-dire aux définitions lexicographiques des lexies impliquées – et qu'il s'agit de relier au niveau sémantique de représentation, au moyen de règles de paraphrasage sémantiques, encore insuffisamment explorées.

Ce sont aussi des paraphrases approximatives. Elles présentent des différences propositionnelles : *grosso modo*, les phrases (1a-c) parlent de la probabilité que le locuteur attribue à un fait P ayant à l'origine un individu X et les phrases (1d-e) parlent, toujours du point de vue du locuteur, de la capacité de X à causer P. Leur (micro-)organisation communicative (en termes d'oppositions communicatives « thème ~ rhème ~ spécificateur », « focalisé ~ non-focalisé », « avant-plan ~ arrière-plan », etc.) est différente.

Les paraphrases de type (1) soulèvent au moins les deux questions suivantes :

1) Quels moyens formels – les représentations et les règles linguistiques – sont nécessaires pour décrire leurs liens sémantiques « profonds »?

2) Comment rendre compte de différents degrés de parenté sémantique qu'elles présentent? Autrement dit, comment formaliser l'intuition des locuteurs, qui distinguent aisément entre paraphrases exactes et approximatives, d'une part, et entre paraphrases et non-paraphrases, d'autre part? Par exemple, les phrases (1a) et (1b) sont des paraphrases mutuelles plus proches que ne le sont (1a) et (1e). Et une phrase comme *Je serais prêt pour ça* est une paraphrase approximative de la seule (1b), cf. la définition de S'ATTENDRE ci-dessous, et d'aucune autre phrase en (1).

La théorie linguistique Sens-Texte, mon cadre de référence, a développé les outils puissants pour décrire, en premier lieu, la paraphrase lexico-syntaxique [Žolkovskij & Mel'čuk 1967, Mel'čuk 2013: 137-197] et, dans une moindre mesure, la paraphrase sémantique [Milićević 2007: 181-241]. Cependant, il reste du travail à faire, surtout en ce qui concerne la couverture des règles sémantiques de paraphrasage. Aussi, les représentations linguistiques – les réseaux sémantiques et les définitions lexicographiques Sens-Texte – pourraient bénéficier d'une meilleure structuration [Polguère 2002, Samson-Fauteux 2010].

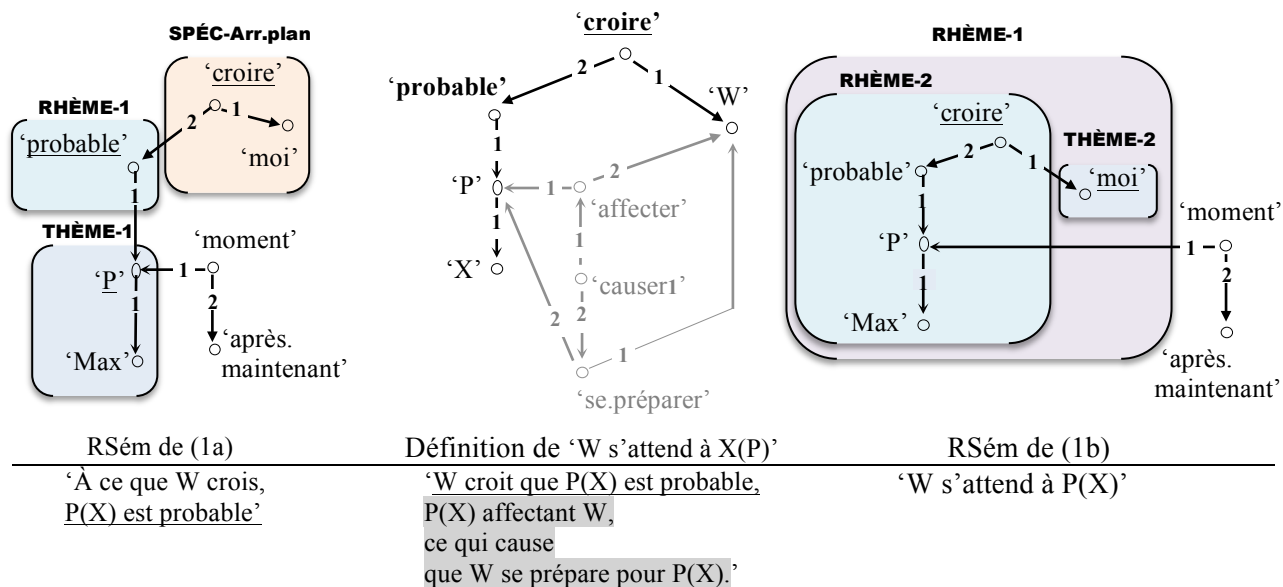
Je proposerai les représentations et les règles linguistiques nécessaires pour rendre compte des liens paraphrastiques unissant les phrases en (1) dans le but de dégager de cet exercice quelques conclusions d'ordre général.

Plus concrètement, il s'agira de distribuer les paraphrases dans les familles paraphrastiques et de dégager pour chacune des familles son invariant sémantique (= le noyau du sens commun) ; dans notre cas, expression de la probabilité ('P(X) est probable') pour les paraphrases (1a-c) et

expression de la capacité ('X est capable de P') pour les paraphrases (1d-e). Il faudra ensuite établir des liens paraphrastique (ou bien conceptuels ?) entre les invariants ; dans notre cas, on aura le lien suivant:

'X a la capacité/la tendance à faire P' \Rightarrow 'P(X) est possible/probable'.

Le lien entre deux paraphrases qui partagent le même invariant sera établi par un appariement flexible de leurs représentations sémantiques respectives [Lareau 200: 76], c'est-à-dire en ayant recours aux définitions lexicographiques. Avec les paraphrases (1a) et (1b), nous aurons:



La différence sémantique entre les deux phrases – les composantes ombrées dans la définition de S'ATTENDRE – est petite et peut être négligée. (Bien entendu, il faudra se doter de critères clairs permettant d'opérationnaliser ce genre d'intuition.) Il y a, en outre, une restructuration communicative, elle aussi jugée admissible étant donnée que l'orientation communicative générale est préservée, et devant être gérée par des règles de quasi-équivalence communicative.

Références

- Lareau, François (2002). *La synthèse automatique de paraphrases comme outil de vérification des dictionnaires et grammaires de type sens-Texte*. Mémoire de maîtrise. Département de linguistique et de traduction, Université de Montréal.
- Mel'čuk, Igor (2013). *Semantics. From Meaning to Text*, vol. 2. Amsterdam/Philadelphia: John Benjamins.
- Milićević, Jasmina (2007). *La paraphrase. Modélisation de la paraphrase langagière*. Berne: Peter Lang.
- Polguère, Alain (2002). Le sens linguistique, peut-il être visualisé? In D. Lagorgette & P. Larrivée, reds, *Représentations du sens linguistique*. Munich: Lincom Europa, p. 89-103.
- Samson-Fauteux, Mélissa (2010). *Représentations du sens lexical en Lexicologie explicative et combinatoire : étude de trois formalismes spécifiques*. Mémoire de maîtrise. Département de linguistique et de traduction, Université de Montréal.
- Žolkovskij, Aleksandr & Mel'čuk, Igor (1967). O semantičeskom sinteze. *Problemy kibernetiki*, 19: 177-238. [Traduction française: Sur la synthèse sémantique (1970). *TA Informations*, 2 : 1-85.