Représentation logico-linguistique des questions/réponses suscitées par un texte

Yllias Chali

Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier
118, route de Narbonne
F-31062 Toulouse Cedex
chali@irit.fr

Résumé

Le travail présenté ici s'inscrit dans le cadre d'un projet commun à des chercheurs en neuropsycholinguistique, en linguistique et en informatique. Il consiste en la mise au jour d'une représentation formelle du texte, permettant d'obtenir, à partir de celui-ci, un ensemble de questions qui lui sont adressables, ainsi que l'ensemble des réponses qu'il contient. Cette représentation est constitué à la fois des structures des grammaires de Harris, et d'une structure de questions/réponses. On conclura sur les perspectives de cette étude au niveau de la formalisation, mais aussi dans le domaine de la génération automatique de textes.

Abstract

The work presented in this paper is lying within the framework of the common project to neuropsycholinguistic, linguistic and computer science researchers. It consists in the up to date of the formal representation of text which allows the underlying of the wholle set of questions about a text and the wholle set of answers to these generated questions. This representation is based at the once time on the structures of Harris's grammar and the question/answer structures established here. We conclude with some perspectives according to the formalization of the question/answer relations and the application of this model of text to an automated text generation.

Représentation logico-linguistique des questions/réponses suscitées par un texte

1 Introduction

Ce travail s'inscrit dans le cadre d'un projet commun à des chercheurs en neuropsycholinguistique, en linguistique et en informatique. Il en est la facette logico-linguistique. Le problème est le suivant : en neuropsycholinguistique, une classe de protocoles expérimentaux, dont le but est d'évaluer l'appréhension des textes par des sujets, est basée sur un système de questions posées aux sujets. La pertinence des réponses à ces questions permet de mesurer la compréhension/mémorisation du texte par ces sujets [Nespoulous et Virbel, 1991].

Or les modèles de texte sur lesquels ces protocoles reposent, manquent d'une caractérisation formelle ; ils ne permettent pas de maîtriser les caractéristiques des questions posées, ni de définir l'univers des questions posables. Afin de répondre à ce problème, on a recourt à deux principes théoriques : d'une part, la décomposition du texte en un ensemble de phrases élémentaires et en un ensemble de caractérisations associées, selon l'approche de Z. Harris dans *Structures Mathématiques du Langage* [Harris, 1971]. D'autre part, l'association aux éléments de l'ensemble précédent d'une structure de questions/réponses, en relation avec la logique des questions [Harrah, 1984].

2 La base textuelle

Harris, dans [Harris, 1971], formule en termes de système mathématique toutes les propriétés et les relations nécessaires et suffisantes du langage naturel, et donne une caractérisation des expressions de la langue. Il démontre que les phrases possèdent, sauf certains cas de dégénérescence, la propriété d'être décomposables, de façon univoque et calculable, en phrases élémentaires. Ainsi *(1)* peut être décomposé en *(2)* :

(1) L'homme s'est retrouvé coincé sur le toit de la maison, alors qu'il venait de détruire un nid de guêpes. Car en voulant redescendre très vite du toit, il a fait tomber l'échelle qu'il avait posée en équilibre instable.

(2) a. L'homme s'est retrouvé coincé sur le toit
b. La maison comportait un toit
c. L'homme venait de détruire le nid
d. Des guêpes avaient fait un nid
e. Très vite, l'homme voulait redescendre du toit
f. L'homme a fait tomber l'échelle
g. L'homme avait posé l'échelle en équilibre instable

À chaque phrase élémentaire est associé un schéma syntaxique qui permet de procéder à certains calculs (voir section suivante). Par exemple, à la phrase élémentaire *(2)a* est associé le schéma syntaxique suivant : N_0 ê$tre\ V_{pp}\ Prep\ N_1$

Le produit de cette décomposition en un ensemble de phrases élémentaires constitue la base textuelle. Le niveau de décomposition en phrases élémentaires est lui-même

passible de variations. Ainsi [Harris, 1971], la détermination du substantif (un/le) ou le temps verbal peuvent être considérés comme des primitives, ou au contraire donner lieu à des décompositions supplémentaires qui fixent la référence des entités ou des intervalles temporels. De manière à optimiser la naturalité des phrases manipulées, nous avons pris le premier parti dans cette étude.

3 La structure questions/réponses

L'ensemble des questions qui peuvent être adressées à un texte sont de deux catégories[1] [Virbel, 1993] : les questions internes (i.e. les réponses à ces questions sont comprises dans la phrase élémentaire support de ces questions), et les questions externes (i.e. les réponses à ces questions sont externes à la phrase élémentaire support de ces questions, elles font intervenir d'autres phrases élémentaires que celle-ci).

Les questions internes découlent syntaxiquement de la structure du schéma de phrase associé à la phrase élémentaire support de questions. Ainsi au schéma syntaxique associé à une phrase élémentaire : $N_0\ V\ N_1\ prep\ N_2$ peuvent être associées les questions suivantes :

(3) a. $Qui\ V\ N_1\ prep\ N_2$? $\quad$ $r\acute{e}ponse : N_0$
b. $Que\ V\ N_0\ prep\ N_2$? $\quad$ $r\acute{e}ponse : N_1$
c. Int_{prep}[2] $N_0\ V\ N_1$? $\quad$ $r\acute{e}ponse : N_2$

Le traitement des questions externes est abordé comme suit : selon le principe provenant de la logique des questions [Harrah, 1984], et en poursuivant la méthodolologie de la *réduction SA (Set-of-Answers)* des questions, dont l'idée est d'identifier une question par l'ensemble de ses réponses, l'ensemble des phrases élémentaires issues de la décomposition du texte est muni de relations interrogatives. Cette relation formalise le mécanisme de questions/réponses entre les phrases élémentaires : elle relie une phrase élémentaire support de question à l'ensemble de ses réponses. Quatre types de relations sont distinguées dans ce contexte [Virbel, 1993] : lieu (répondant à la question *où ?*) ; manière (répondant à la question *comment ?*) ; temporel (répondant à la question *quand ?*) ; implicationel (répondant à la question *pourquoi ?*) qui est subdivisé en implicationel-motif (réponse en *parce que*), et en implicationel-but (réponse en *dans le but de*).

Les relations de questions/réponses reliant cet ensemble de phrases élémentaires sont résumées dans le tableau TAB. 1. Par exemple, les phrases *c, d, e, f, g* sont des réponses au pourquoi-motif de la phrase *a*.

4 Perspectives

La base textuelle, et la structure questions/réponses associée, peuvent permettre, à court terme, de générer des textes parfaitement contrôlés par la sélection de leur contenu informatif : on peut évoquer une histoire par ses éléments essentiels, mais aussi on peut

[1] Notons que les questions absolues, formulées par l'interrogative est-ce-que ... ?, auxquelles on répond par oui ou par non, ne sont pas considérées dans cette étude. De même pour les questions relatives aux verbes.

[2] Int_{prep} est la forme interrogative associée à la préposition, par exemple, Int_{prep} de *sur* est *où ?*, Int_{prep} de *par* est *par où ?*, etc.

Phrases supports de questions	Phrases réponses		
	Pourquoi-motif	Comment	Quand
a	c, d, e, f, g		c, e, f
b			
c	d		
d			
e	c		
f	e, g	e	e

TAB. 1 - : Relations questions/réponses entre les phrases élémentaires *(2)* issues de la décomposition du texte *(1)*

raconter la même histoire en introduisant plus de détails. Ainsi, on peut définir une relation d'ordre sur les phrases, à partir de la structure questions/réponses. Cet ordre établit un degré d'importance de la phrase par rapport à l'essence de l'histoire narrée. La formalisation de ce procédé constitue un de nos centres d'intérêts actuels. Par ailleurs, l'organisation du discours sera réalisée par la capture des relations rhétoriques (définies par [Mann et Thompson, 1987]), celles-ci assurant la cohérence du texte. Les questions externes, qui sont exprimées par une structure reletionnelle à travers les quatre types de relations identifiées, restent à formaliser. Ces relations étant de nature sémantique, cette formalisation est un problème complexe.

Références

[Gross, 1990] Gross, M. (1990). Sur la notion harisienne de transformation et son application au français. *Revue Langages, Les grammaires de Harris et leurs applications*, par Daladier, A., Gross, M., Harris, Z. A., Lentin, A., Ryckman, T., (99):39-56, Larousse, Paris.

[Harrah, 1984] Harrah, D. (1984). The Logic of Questions. In D. Gabbay and F. Guenthner, editors. *Handbook of Philosophical Logic*, volume 2, pages 715-764, by D. Reidel Publishing Company.

[Harris, 1971] Harris, Z. S. (1971). *Structures Mathématiques du Langage*. Édition Dunod, Paris.

[Mann et Thompson, 1987] Mann, W. C., Thompson, S. A. 1987. Rhetorical Structure Theory: a Theory of Text Organization. Technical Report of the Information Sciences Institute, ISI/RS-87-190.

[Nespoulous et Virbel, 1991] Nespoulous, J. L., Virbel, J. (1991). Compréhension/mémorisation de textes de différentes structures par sujets normaux et pathologiques. Rapport technique du Programme de Recherche en Sciences Cognitives de Toulouse.

[Virbel, 1993] Virbel, J. (1993). *Décomposition du texte en un ensemble de phrases élémentaires et caractérisations associées*. Rapport technique du Programme de Recherche en Sciences Cognitives de Toulouse.