Relations intersubjectives dans les discours rapportés

Agata Jackiewicz

▶ **To cite this version:**

Agata Jackiewicz. Relations intersubjectives dans les discours rapportés. Revue TAL : traitement automatique des langues, 2006, Discours et document : traitements automatiques, 47 (2), pp.65-87. halshs-01839037

HAL Id: halshs-01839037
https://shs.hal.science/halshs-01839037

Submitted on 13 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Relations intersubjectives dans les discours rapportés

Agata Jackiewicz

Université de Paris-Sorbonne (Paris 4)
Maison de la Recherche, 28, rue Serpente, 75006 Paris
LaLIC, Paris-Sorbonne

Agata.Jackiewicz@paris4.sorbonne.fr

RÉSUMÉ. L'article décrit les relations énonciatives et sémantiques constitutives des discours rapportés, dégagées à partir de l'analyse des marques linguistiques de l'activité citationnelle dans des médias écrits. Les catégorisations et les ressources constituées sont destinées à différentes tâches d'annotation de textes, notamment dans le contexte du web sémantique.

ABSTRACT. This article studies the enunciative and semantic relations in reported speech, based on the analysis of the linguistic markers in the written media's quotation practices. The categorisations and linguistic resources are to be used in several text annotations' tasks, in the Semantic Web field, among others.

MOTS-CLÉS : linguistique du discours, énonciation, typologie des citations, relations intersubjectives, sémantique des verbes introducteurs de discours rapportés, annotation de textes, ontologie du discours, web sémantique, fouille d'opinion.

KEYWORDS : discourse linguistics, utterance, reported speech typology, intersubjective relations, reported speech introducing verbs semantics, texts annotation, discourse ontology, Semantic Web, opinion mining.

1. Introduction

Dans la hiérarchie, aujourd'hui devenue classique, des niveaux d'analyse à laquelle il est nécessaire de soumettre un texte pour permettre son interprétation, le niveau énonciatif[1] constitue l'étape ultime. On y associe habituellement le recours à certains repères contextuels et co-textuels (spatiaux, temporels, interpersonnels…) pour calculer la référence, ainsi que l'incidence de l'implicite (du présupposé ou du sous-entendu) sur les valeurs de succès des actes de discours. De nombreux exemples illustratifs (dont certains sont également devenus classiques) en sont donnés, le plus souvent pour montrer la nécessité (et la difficulté) de tenir compte des connaissances extralinguistiques. L'inventaire des systèmes informatiques qui assurent ces différents types d'analyse s'arrête, pour sa part, au niveau sémantique, pour lequel relativement peu de réalisations effectives peuvent être exhibées. La dimension énonciative des productions textuelles, quant à elle, n'est pratiquement jamais considérée en tant que telle comme objet de traitement.[2]

Parallèlement, on note un développement significatif d'applications directement intéressées par certains aspects de la textualité, tels que la cohésion thématique, référentielle, anaphorique ou relationnelle pour la segmentation en unités de traitement et la navigation textuelle (Bilhaut *et al.*, 2003), (Couto *et al.*, 2004), (Hernandez et Grau, 2005) ou l'identification de certains « objets » textuels comme les titres, les annonces thématiques, les annonces de plan, les énoncés conclusifs, les références bibliographiques… à des fins de résumé ou de fouille (Minel *et al.*, 2001), (Laublet *et al.*, 2002), (Abdalla and Teufel, 2006), (Jacques et Rebeyrolle, 2006), (Teufel, 2006). Parfois en marge des recherches sur le discours, souvent pour répondre à des besoins concrets de veille ou d'analyse d'image, certains travaux touchent à des phénomènes énonciatifs quand ils cherchent à capter des contenus subjectifs (opinions, émotions, sentiments…) exprimés dans des textes (Wiebe *et al*., 2005), (Prasad *et al.*, 2006). Nous reviendrons sur ce point dans le paragraphe 2.2.2.

L'étude que nous présentons ici s'intéresse résolument à l'énonciation textuelle, et ce dans une double approche : théorique et appliquée. Elle prend pour objet les discours rapportés (DR) dont il s'agit de mettre en évidence les opérations

---

[1] Généralement inclus dans l'intension du terme *pragmatique*.

[2] Notons cependant qu'il existe des réalisations informatiques qui se réclament du champ énonciatif et pragmatique. L'outil Tropes, accessible sur le web (http://www.acetic.fr/), est destiné à l'analyse de discours « en apportant une approche sémantique et pragmatique ». Toutefois, son acception des notions *énonciation* et *prise en charge*, au vu des résultats produits effectivement, ne semble pas correspondre à ce qu'elles recouvrent en linguistique. Par ailleurs, « le marquage textuel » et « la détection de faits et d'événements » (« wrapping » où sont incluses les citations et les opinions) font partie des modules réalisés par la société Lingway (Coch, 2006).

énonciatives et sémantiques constituantes. Ce travail est réalisé grâce à l'analyse des marques linguistiques, témoins de l'activité citationnelle observée dans des médias écrits. Sur le plan linguistique, outre une meilleure compréhension du discours rapporté, notre étude apporte une contribution aux recherches sur les relations de discours. Une réflexion sur la manière dont les relations que nous dégageons pourraient enrichir le formalisme de certaines théories qui cherchent à rendre compte des représentations discursives (telles la SDRT, par exemple, (Busquets *et al.,* 2001)) constitue un prolongement possible de ce travail. Quant aux applications visées, les catégorisations proposées ainsi que les ressources linguistiques correspondantes sont destinées à différentes procédures d'annotation de textes, dans des contextes de l'enseignement, de la recherche d'informations et du web sémantique.

Après avoir rappelé que les discours contiennent de façon constitutive des indications relatives à l'acte de parler et aux interactions entre les usagers de la langue, nous montrons qu'une prise en considération effective des organisations énonciatives dans des textes conduirait à améliorer notablement la qualité des sorties de systèmes de résumé, de filtrage ou de fouille d'opinion (section 2). La section 3 présente un modèle énonciatif et sémantique des discours rapportés, ainsi qu'une étude sur les ressources linguistiques qui lui sont associées. La question des applications possibles et prévues de ce travail est abordée dans la section 4, avant une conclusion générale qui résume les grandes lignes de l'étude (section 5).

2. L'énonciation textuelle et son importance pour le TAL

2.1 La dimension énonciative des textes

Tout discours se construit à travers l'énonciation. L'énonciation prend en charge un énonçable en le plongeant dans un système de référence intersujets (ou intersubjectif), en relation avec l'espace-temps. C'est un phénomène langagier complexe qui témoigne également de la façon dont le sujet parlant s'approprie la langue pour organiser et présenter son discours. Organiser son discours, c'est le segmenter et articuler ses unités (qu'il faut délimiter, lier, enchâsser, regrouper dans des unités supérieures). Cette organisation formelle peut être donnée à voir (avec plus ou moins d'implication de la part de l'énonciateur) par des moyens typo-dispositionnels et discursifs (découpage en paragraphes, sections, mots de liaison, diverses « balises », marqueurs d'intégration linéaire…). Présenter son discours, c'est se situer par rapport à ce qu'il dit sur le monde, par rapport aux paroles choisies pour le dire, par rapport à l'interlocuteur. On peut même parler d'une véritable dynamique énonciative, car l'énonciateur, selon les termes de (Culioli, 2002, p. 187), est amené à ajuster en permanence son discours : « La notion d'ajustement est fondée elle-même sur la conception qu'il n'y a pas des énoncés tout prêts, des vérités toutes prêtes, mais des énoncés produits par un sujet, de telle manière que l'autrui va lui-même reconstruire à partir des marqueurs des

représentations. Entre cette construction et cette reconstruction, nous allons donc avoir un ajustement ». Ces différents ajustements (intersujets, sur soi-même, à son propre discours…) renvoient, au moins partiellement, à ce que (Authier-Revuz, 1995) appelle des non-coïncidences[3]. Paraphrases, reformulations, commentaires autonymes, diverses annonces, qui laissent transparaître de manière variable la subjectivité de l'énonciateur, font partie des traces discursives les plus manifestes de cette « hétérogénéité montrée ». Les discours rapportés condensent, quant à eux, plusieurs relations intersubjectives différentes. Ce phénomène énonciatif et discursif complexe mérite à nos yeux une attention toute particulière, tant sur le plan théorique, qu'en raison de ses retombées possibles pour différentes applications finalisées.

2.2 L'énonciation dans quelques applications de type TAL

Les travaux actuels en TAL et en web sémantique se donnent pour objet d'une part l'aide à la construction de modèles conceptuels ou d'ontologies par l'identification de candidats termes, et d'autre part l'indexation et l'annotation de documents à l'aide de la recherche d'entités nommées et, plus rarement, de relations structurantes. Cette approche référentielle voit un texte principalement comme un lieu où sont déposées des connaissances sur le monde, sous forme d'étiquettes représentant des notions extralinguistiques. La notion du contenu est ainsi abordée en faisant abstraction de la dimension intersubjective des discours analysés. Cet état des lieux doit toutefois être nuancé et complété. On peut rétorquer en effet que ce n'est pas parce que les éléments énonciatifs ne sont pas explicitement désignés comme objets d'analyses ou de traitements, qu'ils ne sont pas dans les faits considérés par des applications informatiques. Nombre d'entre elles y recourent, souvent sans les identifier comme tels. Toutes ces solutions, souvent très empiriques, gagneraient de toute évidence à exploiter d'une manière plus poussée et plus systématique certaines propriétés des organisations énonciatives des textes. En voici quelques illustrations.

2.2.1. *Le résumé automatique*

Citons tout d'abord les systèmes de résumé automatique par extraction de phrases, basés sur la rhétorique de l'auteur du texte, c'est-à-dire sur l'identification d'annonces thématiques, de soulignements d'importance, d'annonces de plan ou de conclusions, tels que ContextO (Minel *et al.*, 2001). La qualité des sorties de ces systèmes dépend d'une part de la présence des indices « dialogiques » par lesquels l'auteur signale explicitement ses intentions communicationnelles (une part importante de ces indices sont de nature énonciative), et d'autre part de la capacité

[3] (1) Non-coïncidence dans l'interlocution (écart entre les co-énonciateurs : intelligibilité, bonne transmission du message) ; (2) non-coïncidence du discours à lui-même (renvoi à une autre source énonciative, emprunt de mots) ; (3) non-coïncidence entre les mots et les choses (les mots ne correspondent pas exactement à la réalité à laquelle ils sont censés référer) ; (4) non-coïncidence des mots à eux-mêmes (le sens des mots est équivoque).

des règles à les reconnaître et à les interpréter. Comme le montre (Blais, 2004), ce deuxième point est loin d'être trivial[4]. On constate que parmi les sources d'erreurs commises par le système, il y a une interprétation trop rapide (appuyée sur peu d'éléments) ou bien trop locale (car limitée à la phrase) des indices pertinents considérés (certains déictiques, verbes présentatifs, formes adverbiales détachées…).

Une annonce thématique ou une conclusion doivent se rapporter au document en cours et non à un autre document qu'il peut être question de présenter ou de résumer (1). Un soulignement d'importance, pour être révélateur des idées de l'auteur, doit être pris en charge par ce dernier, ce qui conduit soit à écarter tout ce qui appartient aux discours rapportés, soit à vérifier que l'auteur adhère aux contenus qu'il cite. Une reformulation qui suit une citation concerne le plus souvent le discours rapporté, et non le discours de l'auteur lui-même, elle ne peut donc être intégrée au résumé censé refléter les idées de l'auteur. Dans (2), nous sommes en présence d'une reformulation d'un discours rapporté avec une distanciation (non prise en charge) de la part de l'énonciateur.

> (1) En guise de conclusion, le rapport relève que, « *sans doute parce que son coût n'est pas couvert par l'impôt mais par des redevances aéronautiques, le service public français de contrôle aérien, dont l'efficacité technique n'est pas en cause, s'est trop longtemps exonéré d'une gestion administrative et financière rigoureuse et économe* ». (*Le Monde*, 7 novembre 2002, p. 23)
>
> (2) Selon M. Dominique Strauss-Kahn, ministre français de l'économie et des finances, alors que ce solde financier était négatif de 149,2 milliards de francs en 1990, il était devenu positif de 134,7 milliards de francs en 1996. Ce qui revient à dire que, chose extrêmement rare, non seulement les entreprises françaises n'ont pas besoin d'emprunter pour financer leurs investissements, mais qu'elles sont en mesure de... prêter. (*Le Monde diplomatique*, avril 1998, p. 18)

Pour pouvoir statuer sur la pertinence d'un énoncé en vue de son intégration dans le résumé, il faudrait idéalement savoir découper le texte en unités homogènes du point de vue de leur source énonciative, attribuer un statut sémantique à chaque unité, puis associer une interprétation textuelle aux enchaînements d'objets identifiés (exemples : séquence « citation par discours direct, reformulation » ou séquence « soulignement d'importance par auteur, reformulation »). Une telle solution est de toute évidence très (ou trop) coûteuse. Une voie plus réaliste serait d'étendre l'espace de recherche au-delà de la phrase et d'enrichir les configurations de marques pertinentes, en tenant compte des traces linguistiques des citations. Par exemple, les introducteurs de cadres médiatifs (*selon*…) peuvent être retenus comme des indices pertinents pour permettre une interprétation énonciative des marques de reformulation (*ce qui revient à dire*…), comme on le voit dans (2).

---

[4] « Actuellement sur la plate-forme ContextO, une grande partie des règles ne distinguent pas si les formes qu'elles étiquettent proviennent de l'auteur ou non » (Blais, 2004, p. 62).

Nous insistons sur le fait que la question des discours rapportés ne peut être occultée par les systèmes de résumé automatique, car les textes à résumer n'incombent pas nécessairement à une seule instance énonciative. Pour certaines applications concrètes, ce point peut revêtir une importance capitale, comme cela a été le cas dans le résumé des jurisprudences, étudié par (Farzindar, 2005) dans le cadre de sa thèse. Ce travail montre clairement le besoin de départager dans les textes de jurisprudences l'auto et l'hétéro-discours, avant de procéder à l'interprétation sémantique de leurs composantes. Cela a conduit à intégrer de plein droit l'identification des discours autres (articles de lois applicables, par exemple) dans la chaîne de traitements opérés par le système[5].

2.2.2 *Fouille d'opinions (opinion mining*)

Depuis quelques années, le besoin d'identifier des contenus subjectifs exprimés sur le web ou dans des bases textuelles conduit à la réalisation d'applications informatiques dites de fouille d'opinions (opinion mining), d'analyse d'émotions, de sentiments ou d'images[6]. Rechercher des commentaires de consommateurs à propos d'un produit ou d'un service donné, interpréter la nature des sollicitations des clients dans des courriers électroniques, analyser l'image d'un parti politique, d'une personnalité publique ou d'une institution, trouver des personnes partageant les mêmes centres d'intérêt sont quelques exemples de tâches qui relèvent de ces problématiques. Les enjeux sociaux et économiques en sont considérables. Parmi les travaux linguistiques menés dans ce domaine, citons à titre d'exemple les recherches de (Bestgen, 2002), (Beauchêne *et al.*, 2002) ou (Nava, 2003).

L'approche proposée dans (Nava, 2003) a attiré notre attention car elle présente la particularité de s'intéresser à des verbes. Plus précisément, il s'agit de verbes (ou syntagmes verbaux) qui permettent d'exprimer des « sollicitations » (demande d'information, demande d'intervention, plainte, constat, engagement, refus, contrainte, explication). La proximité de ces verbes avec les termes du domaine considéré est à la base du classement et du routage des courriers électroniques. Cette

---

[5] Comme le précise Farzindar, le traitement des citations devrait faire l'objet d'extensions supplémentaires pour pouvoir travailler sur d'autres types de jugements, et notamment ceux de la cour d'appel, lesquels restituent non seulement des articles de lois, mais aussi des raisonnements conduits par d'autres juges lors de procès précédemment traités par des cours de niveau inférieur.

[6] Citons *Opinmind* (http://www.activeille.net/index.php/archives/2005/12/13/opinmind-opinion-en-ligne/), l'outil en ligne destiné à collecter des opinions exprimées dans la « blogosphère » ; voir les actes de *AAAI Spring 2006 Symposia* : Computational Approaches to Analysing weblog (http://www.aaai.org/Library/Symposia/Spring/ss06-03.php). Un autre système *OpinionFinder* (http://www.cs.pitt.edu/mpqa/) « is a system that processes documents and automatically identifies subjective sentences as well as various aspects of subjectivity within sentences, including agents who are sources of opinion, direct subjective expressions and speech events, and sentiment expressions. OpinionFinder was developed by researchers at the University of Pittsburgh, Cornell University, and the University of Utah.»

solution, bien qu'elle soit présentée comme délibérément empirique pour répondre à des besoins très précis, cible et exploite judicieusement un sous-ensemble de verbes performatifs, dont la capacité à exprimer des actes de langage pourrait servir dans plusieurs autres contextes appliqués. Le travail sur les verbes de parole que nous présentons dans la section 3, en constitue une généralisation.

Sans rentrer dans les particularités distinctives des différentes méthodologies utilisées, on constate que sur le plan linguistique les ressources utilisées sont essentiellement de nature lexicale (nominale et adjectivale en priorité). Les mots sont classés selon l'attitude exprimée (positive, négative, neutre) ou parfois plus finement (dans le cas des sentiments, par exemple), selon la direction émotive (bonheur, tristesse, colère, crainte, dédain, surprise) qui peut leur être associée. Leur traitement est généralement de nature statistique. Le problème qui se pose avec acuité est de savoir selon quelles méthodes sont constituées ces ressources ? Quelle est leur couverture ? Comment sont choisies les catégories ou les valeurs sémantiques pertinentes ? Comment associe-t-on aux mots leur orientation et leur force émotive ou évaluative…? Car dans le discours, on le sait, avec un éclairage contextuel approprié, il est possible de charger de valeur positive ou négative des termes qui a priori sont neutres ou possèdent une valeur opposée.

2.2.3 *Filtrage selon des points de vue*

Annoter sémantiquement des objets textuels (une définition, une séquence causale…), pour en extraire certains conformément à un critère sémantique ou à un jugement d'importance, nécessite au préalable une étude de la manière dont les frontières de ces objets et leurs modes d'organisation sont signalés dans et par le discours. Il est en effet important de savoir extraire de l'information relativement complète, pour être en mesure de l'interpréter correctement et de statuer sur sa légitimité. Une des difficultés auxquelles sont confrontés les systèmes de fouille est de fournir le contexte de validité (par exemple de vérité ou de prise en charge) de la phrase placée dans le fragment textuel. Par exemple, une information causale ou une description d'un événement n'auront pas le même intérêt pour un utilisateur donné, suivant qu'elles sont entièrement prises en charge par l'auteur du texte, attribuées à un autre énonciateur avec une marque de distanciation de la part de l'auteur, ou plus simplement placées à l'intérieur d'un univers spatial ou temporel limité (Jackiewicz, 2004). Isoler une connaissance de son cadre de validité revient à occulter une part essentielle de son contenu, au risque de lui faire perdre sa valeur opératoire.

Comme on le voit bien dans l'extrait qui suit, l'explication causale de la détermination du sexe (« c'est une plus forte chaleur qui fait concevoir un mâle ») y est indissociable des cadres médiatifs (*dans les deux cas, Pour Anaxadore, Pour Empédocle*) dans lesquels elle prend sens, puis devient ou non opératoire.

> (3) Pour Anaxadore, la détermination du sexe vient du père : les garçons proviennent du testicule droit, le plus chaud, les filles du gauche. Pour Empédocle, c'est la plus forte chaleur de la matrice, selon l'état du sang menstruel, qui fait naître un garçon ou une fille. Comme on le voit, dans les deux cas, c'est une plus forte chaleur qui fait

> concevoir un mâle. Aristote critique ses devanciers sur certains points, car ce n'est pas une mince affaire, dit-il, que de démontrer que la mise en présence du froid déclenche la production d'un utérus sur le fœtus, dans la matrice de la mère, bien qu'il conserve la vertu de l'opposition entre chaud et froid. (F. Héritier, Masculin/féminin, 1996, p. 192)

Enfin, la problématique même d'assignation de valeurs sémantiques à des objets textuels est concernée par l'organisation discursive et énonciative du texte, car la levée de certaines ambiguïtés s'opère souvent à l'intérieur des unités supérieures à la phrase. Nous l'avons montré notamment dans le cas de l'interprétation sémantique d'énoncés causaux (Jackiewicz, 2004).

3. Étude énonciative des discours rapportés

La problématique des discours rapportés, se trouvant de fait au cœur des recherches sur l'énonciation, concerne également les analyses de discours intéressées par les organisations textuelles. En effet, comme on le verra dans les paragraphes qui suivent, l'objet « citation » est une entité discursive multiforme, composée de segments fonctionnellement hétérogènes, non tous nécessaires, mais participant chacun d'une manière précise à l'intelligibilité de l'ensemble. La question de l'identification et de l'annotation de cet objet aux multiples facettes est donc éminemment complexe, car plusieurs angles d'approche semblent possibles. Le rôle d'une analyse linguistique préalable est d'indiquer les éléments de solution ayant une certaine validité théorique.

3.1 *Apports des travaux précédents, rôle du corpus et démarche suivie*

On ne peut résumer en quelques mots les recherches linguistiques sur le discours rapporté, compte tenu de leur longue tradition et de la diversité des approches, dont l'une des plus récentes s'intéresse à l'effacement énonciatif dans les discours rapportés (Rabatel, 2004). Nous avons privilégié et exploité les travaux qui, contrairement à ce dernier courant, mettent clairement l'accent sur le signalement de l'activité citationnelle en surface du texte. D'une part, il existe de nombreuses descriptions, souvent minutieuses et abondamment exemplifiées, des modes de citation (discours direct, discours indirect, discours narrativisé et de multiples variantes de ces modes de base) que nous devons, sans être exhaustive, à (de Gaulmyn, 1983), (Authier-Revuz, 1995), (Charaudeau, 1997), (Kerbrat-Orecchioni, 1999), (Rosier, 1999), ainsi qu'à la vaste communauté Ci-dit. D'autre part, plusieurs auteurs ont proposé des analyses syntaxiques et/ou sémantiques des verbes de parole, dont (Gross, 1975), (Vanderveken, 1988), (Eshkol, 2002) et (Lamiroy et Charolles, 2003).

Cependant, il est frappant de constater qu'il n'existe pas de classifications sémantiques de verbes potentiellement introducteurs de DR ayant pour visée de catégoriser systématiquement l'ensemble de ces verbes. On note l'existence de nombreuses ébauches de catégories, illustrées pour la plupart par quelques verbes

prototypiques, ce qui rend leur portée théorique difficile à cerner. On constate que les catégories sémantiques qui reviennent le plus souvent (« la fidélité » / « l'infidélité » du DR, « le discours du vrai » / « le discours du faux », « l'objectivité » / « la subjectivité » du locuteur rapportant, « la prise en charge » / « la non-prise en charge ») touchent globalement à l'attitude que l'énonciateur rapportant adopte par rapport aux paroles rapportées. Inversement, on trouve au fil des analyses consacrées aux citations des listes de verbes caractérisés selon les cas par leurs propriétés sémantiques ou énonciatives. Par exemple, (Rosier, 99) fait mention de « verbes descripteurs du déroulement énonciatif », « verbes de mouvement ou de mimiques gestiques », verbes de « vocifération », verbes d' « interrogation indirecte »…

Prenant appui sur les travaux mentionnés, nous avons choisi de travailler également sur corpus, afin de pouvoir examiner des séquences citationnelles, telles qu'elles fonctionnent dans des situations réelles de communication. Par ailleurs, il est important à nos yeux de pouvoir travailler sur des séquences citationnelles « complètes », c'est-à-dire interprétables (et exploitables) indépendamment de leur co-texte. La question de la délimitation de l'objet citation, composé éventuellement de plusieurs segments, devient alors importante, alors qu'elle n'est pas soulevée par les études linguistiques sans visée applicative. Les marques linguistiques pertinentes pour l'étude deviennent alors des configurations d'indices où les expressions introductrices de DR interagissent avec d'autres éléments contextuels également impliqués dans les repérages énonciatifs de la citation.

Nous avons choisi comme observables des textes de presse quotidienne, « naturellement » riches en citations[7]. Les médias tels que *le Monde* rapportent des événements énonciatifs fort variés : présentations orales, conférences de presse, interviews tenues par divers acteurs de la vie politique, sociale ou culturelle, diverses réactions venant des syndicats, des associations… mais aussi, livres et critiques de livres, extraits ou synthèses de différents rapports ou publications. Mais, leur mission n'est pas seulement de transmettre ou reproduire des paroles qui circulent dans l'espace social, mais également (ou surtout) de les rendre intelligibles et de montrer l'impact qu'elles produisent dans cet espace. Il s'agit de représenter d'une manière synthétique et lisible les aspects saillants des discours originels et, préférentiellement, les interactions qui s'opèrent dans et à travers l'activité linguistique des sujets communicants. Pour cette raison, nous avons focalisé notre attention sur des relations intersubjectives qui s'expriment dans l'activité citationnelle[8].

---

[7] Notre corpus d'étude rassemble les articles du *Monde* parus entre le 1er juillet 2002 et le 30 juin 2004. Le guide rédactionnel « Le style du Monde » (*Le Monde*, 2002) nous a fourni des indications précieuses sur les recommandations que doivent suivre les journalistes en matière de citations.

[8] Dans la presse, comme dans le roman analysé par Bakhtine, « Le discours de l'auteur représente et enchâsse le discours d'autrui, crée pour lui une perspective, distribue ses ombres

Concernant les ressources linguistiques préexistantes et exploitables dans notre étude, nous avons retenu la liste des verbes de parole (512 verbes classés en transitifs et intransitifs) constituée par (Lamiroy et Charolles, 2003), ainsi que les prédicats de parole (1 130 prédicats, dont 637 verbes) de (Eshkol, 2002). Même si elles ont été constituées en dehors de la problématique des discours rapportés, ces deux listes fournissent de fait les principaux marqueurs aptes à introduire des paroles étrangères. Concernant la couverture de ces ressources et leur adéquation à nos besoins, plusieurs remarques s'imposent. Si tous les verbes de parole répertoriés dans les dictionnaires faisant autorité y sont présents, la décision d'inclure dans la base des néologismes (tels que *litoter, euphémiser, pléonasmer, xéroxer)* inconnus de ces dictionnaires, mais très présents dans des forums ou des blogs, et pour certains utilisés couramment dans la presse (*euphémiser*), fait l'objet d'un choix à assumer. Certains verbes de parole n'introduisent des discours rapportés que dans des constructions très particulières et donc très rares[9]. Inversement, plusieurs classes de verbes ne rentrent pas directement dans la catégorie des verbes de parole, mais peuvent introduire des discours rapportés, il s'agit des verbes ayant trait à (i) certaines activités intellectuelles, (ii) certaines attitudes corporelles, (iii) des attitudes propositionnelles, (iv) des situations de communication de nature non spécifiée, (v) de la perception, (vi) des cris d'animaux. À l'exception de la classe (vi), tous ces verbes ont été rassemblés par nos soins sur corpus, il est donc impossible de garantir leur complétude[10]. Difficile (voire impossible) à atteindre en pratique, la complétude des ressources n'est pas visée par notre travail. Nous cherchons en revanche à dégager une catégorisation sémantique de ces ressources, permettant de savoir quels types de marques peuvent en faire partie.

Au final, une base de 722 verbes a été constituée pour l'étude. Ces verbes ont été classés dans un premier temps à l'aide de catégories sémantiques issues de la littérature consultée. Les représentants de chaque classe, associés éventuellement à d'autres marques de nature énonciative, ont permis de sélectionner des citations dans le corpus de référence[11]. L'observation de ces extraits a débouché sur la

---

et ses lumières, crée sa situation et toutes les conditions de sa résonance, enfin y pénétrant de l'intérieur, y introduit ses accents et ses expressions, crée pour lui un fond dialogique » (Bakhtine, 1978, p. 175). En dépit de l'exigence de neutralité et d'objectivité qui s'impose aux écrits journalistiques, ces derniers sont constamment traversés de modalisations qui à des degrés divers guident et influencent le lecteur dans l'interprétation des paroles citées.

[9] Frantext livre des attestations prouvant que même les verbes intransitifs génériques (*causer, parler, tchatcher*) ou les verbes qui renvoient à une activité non planifiée et durative du discours (*bavarder, converser, discourir…)* peuvent entrer dans des expressions introductrices de discours étrangers (*profitent de la halte pour venir bavarder avec les fantassins : « nous, on est de la brigade légère… »* ou *Car voilà comment parle le Prophète: « Voici dorénavant le nom de la cité… »*)

[10] Les classes (i) à (v) représentent actuellement 52 verbes.

[11] Précisons qu'une vérification systématique a été menée sur Frantext et sur le web pour s'assurer de la capacité de chacun de ces verbes à introduire un DR. Cette base d'attestations tout-venant a également aidé à décrire la sémantique de ces verbes. Remarquons toutefois que

proposition d'un premier ensemble de relations énonciatives permettant de distinguer plusieurs types de séquences citationnelles. L'analyse sémantique systématique de l'ensemble des verbes (hors corpus, mais avec consultation d'attestations pour de nombreux verbes) a permis de dégager une organisation sémantique cohérente (en plusieurs catégories et sous-catégories, voir § 3.3), débouchant en retour sur une structuration plus fine des relations énonciatives. Les relations énonciatives ainsi dégagées ont servi à consolider la typologie initiale des discours rapportés.

3.2 *Typologie des discours rapportés*

Le problème de la citation est comment restituer d'une manière juste et synthétique toutes les informations qui disent dans un contexte initial (oral ou écrit) comment il faut comprendre un discours. Citer consiste donc, au minimum, à transcrire ou reproduire l'ensemble ou une partie du discours d'un autre (rapporter ce qui a été dit). Mais, la compréhension de ces paroles s'appuie de toute évidence sur la connaissance des paramètres essentiels de la situation d'énonciation les caractérisant (par qui, quand, où, à l'attention de qui cela a été dit). De plus, l'interprétation des intentions du locuteur (si elle ne découle pas clairement des paroles citées), celle des attitudes manifestées par le locuteur en parlant, facilitent notablement, voire conditionnent dans certains cas, l'accès au sens de la citation. Enfin, certains discours appellent, en raison notamment de l'absence d'un bagage culturel ou d'un savoir commun entre le locuteur d'origine et le co-énonciateur, un commentaire, une explication ou une reformulation mettant en évidence leur teneur effective. Nous dirons donc que, selon la nature de ses composants, une citation peut être qualifiée de minimale, authentifiée, interprétée ou dialogique.

L'extrait (4) exemplifie une séquence citationnelle composée qui restitue des paroles étrangères, authentifie l'énonciation et après avoir porté un jugement d'euphémisation sur la forme d'expression du discours rapporté, rend intelligible son contenu. Si la citation minimale correspondant au discours entre guillemets (« *pas toujours l'exemple que l'on pourrait attendre d'elle* ») est difficilement interprétable, on constate que le segment plus large exprimant une citation authentifiée (*selon…, mercredi 26 mai*) n'est pas encore suffisamment informatif. La teneur réelle de cette citation est révélée par l'énonciateur dans le segment qui suit la citation à proprement parler (*Un euphémisme*…).

> (4) La France ne serait-elle plus la « *patrie des droits de l'homme* » ? Selon un rapport d'Amnesty International, rendu public mercredi 26 mai, elle ne donne, en tout cas, « *pas toujours l'exemple que l'on pourrait attendre d'elle* ». Un euphémisme, cachant des critiques sévères sur certaines législations et pratiques des autorités publiques, mises en lumière par l'organisation non gouvernementale (ONG). (*Le Monde*, 27 mai 2004)

---

pour certains verbes de parole, il est tout aussi difficile d'attester que d'exclure la possibilité d'introduire des paroles rapportées. C'est notamment le cas des verbes métalinguistiques tels que *adjectiver, substantiver, postposer*…

3.2.1 *Citation minimale*

Les paroles rapportées à proprement dit constituent le noyau de toute citation. Le seul fait de les délimiter par des signes typographiques conventionnels dédiés à cet usage (guillemets, italique…)[12] permet de signifier qu'elles ont une origine différente de celle du discours dans lequel elles sont plongées. Si aucune autre marque explicite ne vient s'ajouter aux balises marquant les frontières des paroles rapportées, c'est que leur auteur et le contexte d'énonciation d'origine : (i) sont supposés connus du lecteur, faisant partie d'un savoir partagé (5) ; (ii) ne sont plus disponibles dans la mémoire collective ou ne sont pas importants, comme c'est le cas d'un grand nombre de « particitations » au sens de (Maingueneau, 2004) (4) ; (iii) sont accessibles au lecteur dans un contexte plus large, il s'agit des citations minimales contenues dans des titres et dans des accroches, qui sont généralement issues des discours rapportés plus amplement dans le corps de l'article ; (iv) ne sont pas connus, c'est le cas des bruits et des rumeurs (6) qui doivent alors être signalés comme tels[13]. L'énonciateur, quant à lui, s'efface pour laisser résonner les paroles étrangères.

(5) Les 35 heures : bilan globalement positif (*Le Monde*, 21 septembre 2002 ; titre de l'article)

(6) Les rumeurs, et elles sont nombreuses, assurent que, dans la période de succession qui s'ouvre, certains membres de la famille royale auraient une attitude ambiguë. *(Le Monde*, 22 juin 2004)

3.2.2 *Citation authentifiée*

Si les paroles enregistrées ou transcrites sont souvent déjà interprétables (c'est le cas de bon nombre d'écrits anciens), elles restent néanmoins coupées de leur contexte d'origine. Pour mieux comprendre leur teneur, nous avons besoin de les authentifier c'est-à-dire connaître les principaux paramètres de l'acte d'énonciation[14], à savoir qui a parlé ? où ? quand ? à l'attention de qui ? (7), ainsi que certaines propriétés saillantes de ces paramètres (8). Différentes combinaisons de ces paramètres peuvent être linguistiquement exprimées et chacun d'entre eux

---

[12] Aux marqueurs formels agissant comme des démarcateurs, s'ajoutent toutefois des marqueurs de cohérence et de cohésion interne du discours rapporté, lesquels jouent également un rôle dans le tracé de la frontière entre le discours citant et le discours cité.

[13] « *Le Monde* ne publie pas de citations non sourcées, sauf cas exceptionnel » (*Le Monde*, 2002, p. 8)

[14] « Un contexte possible d'énonciation d'une interprétation sémantique est composé de cinq aspects ou ensembles d'aspects fondamentaux différents ; lesquels sont : un *locuteur*, un ou plusieurs *allocutaires* (ceux auxquels le locuteur s'adresse), un *ensemble fini d'énoncés* (qui est l'ensemble des énoncés utilisés par le locuteur dans ce contexte), un *moment* et un *lieu* (qui sont le moment et le lieu de l'énonciation de ces énoncés), et une série de propriétés ou d'attributs de ce locuteur, de ces allocutaires, de ce moment et de ce lieu d'énonciation qui sont pertinents pour déterminer d'abord la nature exacte, et ensuite les valeurs de succès et de satisfaction des actes de discours exprimés dans ce contexte » (Vanderveken, 1988, p. 52)

peut être spécifié d'une manière plus ou moins riche. Le statut privilégié revient toutefois au locuteur, à la fois auteur de l'énoncé rapporté et source des repérages spatio-temporels et personnels au sein de cet énoncé. Individuelle ou collective, cette entité peut être présentée avec différents degrés de détermination, de manière à renseigner sur sa notoriété et, par voie de conséquence, sur la légitimité et l'impact possible de ses paroles.

> (7) « *Cette grève de la faim est un choix personnel de Mme Nasraoui. Elle n'a qu'à en assumer les conséquences* », a déclaré, vendredi 26 juillet, le responsable du dossier des droits de l'homme auprès du ministre de la Justice, Ridha Khemakhem, en réponse à une question du *Monde* concernant la dégradation de l'état de santé de l'avocate. (*Le Monde*, 29 juillet 2002)
>
> (8) Au bout du fil, un homme s'exprimant en anglais, avec un fort accent arabe, qui lui annonce qu'il a été recruté pour faire sauter, dans les prochaines heures, un camion bourré d'explosifs devant l'ambassade des États-Unis, près de la place de la Concorde. (*Le Monde*, 10 décembre 2003)

Notons qu'une citation qui restitue son énonciation fait clairement référence à un acte de parole effectif. Cet événement, s'il est oral, peut être caractérisé également par les propriétés physiques et acoustiques de la parole du locuteur. À l'écrit, c'est la forme graphique de la parole, indissociable de son support physique, qui témoigne de son existence réelle. On distingue de cette façon une parole réellement prononcée ou écrite d'un acte de communication réalisé par d'autres biais (un dessin, un signe gestuel, une mimique… qui font comprendre une intention).

3.2.3 *Citation interprétée*

Dans la plupart des cas, les paroles citées ne font qu'illustrer un discours plus large dont le but communicatif est restitué grâce à un travail interprétatif de l'ensemble de la situation d'énonciation d'origine. L'énonciateur prend alors en charge ce qu'il a identifié et reconstruit comme essentiel dans l'intention communicative manifestée par le locuteur. C'est le rapport subjectif de ce dernier à sa propre parole et à l'effet qu'elle est censée produire qui se trouve ainsi mis en évidence, sans qu'aucun jugement n'y soit porté. Premièrement, la compréhension d'un discours rapporté exige que l'on détecte la valeur d'illocution que ce discours portait dans son contexte d'énonciation originel[15]. Dans certains cas, les paroles rapportées suffisent à reconstituer la nature de l'acte de discours réalisé (9) ; autrement, l'énonciateur se doit de l'exprimer explicitement (10). Deuxièmement, il est parfois pertinent de restituer l'état d'esprit et l'attitude affective de celui qui a pris la parole, ses croyances, ses réserves, l'importance qu'il accorde à son propos, son rapport à la valeur de vérité de ce qu'il avance. Troisièmement, qualifier le style, le mode d'argumentation ou la manière de s'exprimer[16] du locuteur s'impose parfois, notamment quand celui-ci emploie des figures du style où ce qu'il veut dire

---

[15] « La compréhension de l'illocutoire est une condition nécessaire […] pour comprendre un énoncé » (Lyons, *Sémantique linguistique*, 1990, vol. 2, p. 731).

[16] « Le style de la personne est respecté et restitué au plus près. » (*Le Monde*, 2002, p. 8)

diffère nettement de la signification conventionnelle de son énoncé (l'ironie, le sarcasme, le sous-entendu…) (11) ou s'exprime dans un style professionnel (éloquence juridique, politique…) ou social (dialecte social, régional…) particulier.

(9) Dans son discours, le ministre a, par ailleurs, affirmé : « *Je déclare devant vous que l'Irak est libre de toute arme nucléaire, chimique et biologique.* » (*Le Monde*, 21 septembre 2002)

(10) Les enquêteurs du SRPJ de Marseille, dans leur rapport de synthèse, innocentent totalement Bernard Tapie des accusations de Pierre Dubiton. *« Les mouvements (achats-ventes) de joueurs ont été nombreux durant l'intersaison 2001-2002 et le mercato d'hiver,* écrivent-ils, et, comme l'a reconnu M. Dubiton, *aucun document engageant le club n'a été signé par Bernard Tapie*. » (*Le Monde*, 14 juin 2003)

(11) Et le quotidien Trybuna, héritier de Trybuna Ludu, jadis organe du parti communiste, toujours au service de ses amis postcommunistes, conclut avec sarcasme : « *La nervosité s'installe dans les cercles politiques et médiatiques. Le ministre Kolodko ose réclamer une déclaration de patrimoine !* » Avant de démasquer, dans son meilleur style, les fraudeurs : « *Pour satisfaire Kolodko, chaque honnête businessman et sa businesswoman, M. et Mme le juge, M. et Mme le médecin, M. et Mme le député, Mgr l'évêque, devraient déclarer ce qu'ils possèdent. Quel culot !* » (*Le Monde*, 4 septembre 2002)

3.2.4 *Citation dialogique*

La quatrième composante du discours rapporté met en lumière la double relation dialogique qui s'établit d'une part entre celui qui rapporte (l'énonciateur) et le locuteur d'origine, et d'autre part entre l'énonciateur et celui à l'attention de qui le discours est rapporté (le co-énonciateur). Dans la première, l'énonciateur affiche explicitement le jugement qu'il porte sur le discours rapporté. Son appréciation peut concerner (i) l'adéquation des choix des moyens d'expression ou des modes d'organisation du discours au contenu du discours ou à la situation d'énonciation (12), (ii) l'attitude du locuteur face à son allocutaire, (iii) le positionnement du locuteur face au contenu exprimé.

Dans la deuxième relation, l'implication affichée de l'énonciateur traduit son souci de rendre intelligible le discours rapporté à son propre co-énonciateur. En effet, certains énoncés n'expriment pas d'une manière claire et directe ce qu'ils veulent vraiment dire : ils impliquent, insinuent, suggèrent autant (ou plus) qu'ils (n') affirment, déclarent ou définissent ouvertement. De plus, certains discours ne conduisent à leurs significations réelles qu'à condition d'accéder à certaines connaissances extralinguistiques précises (12, 13). Par conséquent, de nombreuses citations ne peuvent être interprétées sans un commentaire, une reformulation ou une explication rédigés en fonction de la personne à qui elles sont destinées. Cet aspect de la citation est très saillant dans les articles de presse où la fonction de guidage est importante. Le journaliste se positionne en médiateur qui traduit en langage ordinaire, reformule, paraphrase, met en images, métaphorise et recontextualise le discours autre (scientifique, technique, politique, appartenant à une culture particulière ou peu connue…) pour le rendre accessible au plus grand

nombre[17]. La séquence 12 illustre parfaitement cette double intervention de la part de l'énonciateur : l'appréciation sur la manière de s'exprimer (*le doux euphémisme*) du locuteur cité, y est suivie d'une reformulation (*en clair*) adressée à des lecteurs a priori peu habitués à la rhétorique communiste[18].

> (12) Li Qinfu sait parfaitement qu'une cellule communiste stipendiée peut lui être d'une grande utilité dans ses relations avec les autorités. « *Le grand problème en Chine dans les affaires*, dit-il, *c'est l'absence de transparence des politiques officielles. Ces politiques changent tout le temps. Le comité du parti nous aide à mieux comprendre.* » Plus important encore est l'ordre social que fait régner ce même parti. « *Il nous aide à évaluer la performance des employés*. » Le doux euphémisme. En clair, le parti, c'est le syndicat maison. Sous le dôme du Capitole de Pinghu, le néo-capitalisme chinois croisé de tradition communiste germe. Une culture en serre d'un nouveau type, idéologiquement modifié. (*Le Monde*, 15 janvier 2002)
>
> (13) Deux jours après s'en être entretenu avec Nicolas Sarkozy, Jacques Chirac a fait savoir, jeudi 24 juin, par les services de l'Élysée, qu' « *un ministre ne peut pas être en même temps président du principal parti de la majorité* ». En clair, M. Sarkozy devra abandonner Bercy, s'il se présente et est élu à la présidence de l'UMP. (*Le Monde*, 26 juin 2004)

3.3 *Relations énonciatives et sémantiques constitutives des discours rapportés*

La typologie énonciative des discours rapportés que nous venons d'esquisser est construite à partir des modes d'intervention de l'énonciateur dans le discours qu'il rapporte[19]. « Absent » dans la *citation minimale*, restituant les paramètres fondamentaux de l'énonciation rapportée dans la *citation authentifiée*, observateur et interprétant neutre des différents aspects de l'énoncé et de l'énonciation rapportés dans la *citation interprétée*, enfin, véritable médiateur investi, dialoguant à la fois avec la parole rapportée et avec le co-énonciateur, dans la *citation dialogique*.

---

[17] « En effet, les journalistes se représentent comme devant affronter des logiques de masquage auxquelles ils opposeraient des logiques de dévoilement, mais ils se représentent également bien souvent comme faisant face à des logiques techniques dont ils assureraient le déchiffrement (à travers la "traduction" des " jargons", par exemple) » (Krieg-Planque, 2004, p. 70)

[18] Nous avons analysé d'une manière plus approfondie les discours rapportés qualifiés d'euphémistiques par leurs énonciateurs (82 occurrences dans notre corpus). Le jugement d'euphémisation porté sur les mots d'autrui permet, dans un souci de médiation, de traduire une réalité que l'autre nomme imparfaitement ou, d'une manière ouvertement critique, de dénoncer l'autre qui nomme mal. 52 % de ces citations comportent une reformulation.

[19] Pour une discussion approfondie des différents modes d'intervention de l'énonciateur dans le discours qu'il rapporte voir (Jackiewicz, A. « Objets de la prise en charge dans le discours rapporté », Colloque *Prise en charge/Commitment*, Anvers, 11-13/01/2007).

Chaque catégorie de citation est définie par un ensemble de relations énonciatives constitutives (tableau 1). Nous y symbolisons le locuteur d'origine[20] (l'énonciateur rapporté) par Je1, son allocutaire (le destinataire initial du propos rapporté) par Tu1, le lieu par S1, le moment par T1, l'énonciateur rapportant par Je2, le co-énonciateur (le destinataire du texte comportant la citation) par Tu2, la forme du propos par dire (P), le contenu du propos par dit (P), l'existence d'une relation explicite par →.

Les relations énonciatives de niveau 0 sont relatives aux repérages élémentaires, constitutifs de toute énonciation. Les relations énonciatives de niveau 1 sont celles que l'énonciateur restitue, en observateur ou en interprétant objectif (symbolisé par <Je2)> :) ; elles concernent les choix et les positionnements du locuteur à l'égard de son allocutaire, de la forme d'expression et du contenu exprimé. Les relations énonciatives de niveau 2 intègrent explicitement l'énonciateur, lequel exprime un jugement sur les choix et les attitudes manifestés par le locuteur. Au niveau 3, l'énonciateur explique ou reformule le discours rapporté à l'attention de son propre co-énonciateur, lequel dans la plupart des cas n'est pas identifié (<Tu2 >).

Type de citation		**Relations énonciatives**
Citation minimale		
Citation authentifiée		(0) Je1 → Tu1 Je1 → S1 Je1 → T1 Je1 → dire (P)
Citation interprétée		(1) <Je2> : (Je1 → Tu1) <Je2> : (Je1 → dire (P)) <Je2> : (Je1 → dit (P))
Citation dialogique	appréciative	(2) Je2 → (Je1 → Tu1) Je2 → (Je1 → dire (P)) Je2 → (Je1 → dit (P))
	intelligible	(3) Je2 → <Tu2 >

Tableau 1 : *Relations constitutives des quatre types de discours rapportés*

À chaque relation énonciative est associée une famille de valeurs sémantiques relatives aux propriétés saillantes à l'œuvre dans la relation. Ces valeurs, dégagées essentiellement à partir de l'étude des verbes introducteurs de DR (voir le § 3.4), peuvent être catégorisées de manière de plus en plus fine. Au niveau le plus général, trois aspects saillants semblent se dégager dans la manière de présenter des paroles d'autrui : (i) faire référence à leur forme d'expression originelle, (ii) focaliser

[20] Nous appellerons, suite à (Desclés et Guentchéva, 2000), *locuteur* le dernier *énonciateur* qui prend en charge directement la relation prédicative (ce qui est dit).

l'attention sur leur contenu, (iii) insister sur la relation (visée) interpersonnelle dont elles étaient porteuses dans leur énonciation d'origine. Chacun de ces aspects peut, à son tour, être décrit par un certain nombre de catégories plus précises. Par exemple, la forme d'expression qui réfère au choix des moyens d'expression et des modes d'organisation du discours, peut être caractérisée selon les « rubriques » suivantes : 'mode argumentatif', 'figure de style', 'genre', 'gestion du discours', 'manière de s'exprimer', 'support', 'type d'objet textuel'. La rubrique 'type d'objet textuel' correspond aux différentes valeurs exprimées par les verbes tels que *introduire, conclure, titrer, résumer, exemplifier…*

Pour montrer d'une manière précise comment les valeurs sémantiques donnent corps à une relation énonciative, considérons la relation Je2 → (Je1 → dire (P)). Dans cette relation, l'énonciateur porte un jugement sur l'adéquation des moyens d'expression ou des modes d'organisation du discours choisis par le locuteur au contenu du discours ou à la situation d'énonciation. Cette attitude critique peut prendre de nombreuses nuances sémantiques : (i) dénoncer une écriture rapide et dépourvue de soin de textes se réclamant d'un certain genre (*écrivasser, écrivailler, rimailler, criticailler*…), (ii) signaler des défauts sur le plan de l'argumentation : s'écarter du thème (*digresser*), le fait d'opposer des arguments excessivement subtils et captieux (*ergoter*) ou d'avoir un souci excessif de détails (*pinailler*), (iii) pointer l'excès de forme cachant l'absence de contenu (*blablater, laïusser*), (iv) révéler l'usage de l'implicite, le fait de laisser entendre quelque chose sans l'exprimer directement (*insinuer*), (v) critiquer le style employé : excès d'emphase (*pérorer*), langage obscur (*jargonner*) (14), (vi) porter une appréciation sur la manière de parler (élocution, force, violence…) (*criailler, invectiver*), le fait d'émettre des sons (souvent désagréables) faisant penser à des cris d'animaux (*beugler, bramer…*), (vii) porter une appréciation sur la gestion du discours (longueur, répétitions…) (*discourir, palabrer, rabâcher, radoter, rebattre, ressasser*),

> (14) « *L'appropriation politique du pacte par les États membres a diminué* », jargonne la Commission, ajoutant que les gouvernements n'ont « *pas voulu tirer les conséquences de l'introduction de l'euro dans la conduite des politiques budgétaires nationales* » (*Le Monde*, 27 novembre 2002)

Les différentes rubriques sémantiques peuvent apparaître dans plusieurs relations énonciatives. En effet, les relations énonciatives s'organisent dans un système où de nombreuses caractéristiques du discours rapporté, touchant à la forme, au contenu ou au rapport social entre interlocuteurs, peuvent être restituées par l'énonciateur d'une manière neutre ou faire l'objet de sa part d'un jugement explicite. En voici un exemple. La trace de la sollicitation de l'autre qui est une des valeurs possibles de la relation interpersonnelle <Je2> : Je1→ Tu1, peut s'enrichir d'une appréciation de la part de l'énonciateur, dans la relation Je2→(Je1→Tu1), mettant en évidence son caractère abusif (*racoler, bassiner*…), la volonté d'imposer un modèle de pensée (*endoctriner*…) ou le fait de chercher par de belles paroles à obtenir des avantages personnels (*enjôler, fayoter*…) (15, 16).

(15) Parallèlement, le juge s'est attaqué aux finances du mouvement indépendantiste basque et de ses possessions présumées à l'étranger. Il a sollicité, jeudi, du président du Parlement européen Pat Cox, « *la suspension des aides et subventions que reçoit Batasuna* » (*Le Monde*, 7 septembre 2002)

(16) Simple militant du PS aujourd'hui, il estime que la « *motion Hollande, qui nous a bassiné en disant :* " attention à la gauche du parti *" fait aujourd'hui exactement la même chose* ». (Le Monde, 6 juin 2003)

Au fond, cette organisation permet de rendre compte du fait que les citations ne reflètent pas seulement les rapports, s'exprimant en termes d'adhésion ou de distanciation, que le locuteur et l'énonciateur entretiennent avec le contenu propositionnel des paroles rapportées, mais aussi de nombreuses relations sémantiquement riches et nuancées touchant tant à des caractéristiques formelles du discours rapporté, qu'à des relations interpersonnelles établies par ce discours.

3.4 *Marqueurs pertinents pour la typologie des discours rapportés*

Les valeurs énonciatives qui fondent notre classification des discours rapportés ont été dégagées à partir de l'analyse des configurations de marques impliquées dans le signalement de l'activité citationnelle. En font partie les expressions introductrices de discours rapportés (verbes, syntagmes nominaux et expressions adverbiales), les marques relatives aux principaux paramètres de l'énonciation rapportée, ainsi qu'une vaste famille de marques permettant à l'énonciateur de commenter, d'expliquer ou de reformuler le discours d'autrui. Afin de valider la pertinence de ces catégories, nous avons analysé 722 verbes introducteurs de DR, dont l'aptitude à superposer la fonction d'attribution de dire avec diverses modalisations, en fait les principaux « porteurs » de la valeur sémantique de la citation[21]. Plus de 660 verbes ont pu être classés dans les catégories énonciatives proposées, ce qui confirme leur adéquation à la modélisation des discours rapportés. Le classement des verbes selon les valeurs sémantiques possibles pour chaque relation énonciative, se heurte à l'épineux problème posé par la multiplicité des nuances de sens exprimables par les verbes de parole. Nous avons proposé une organisation sémantique selon trois niveaux, cohérente avec l'organisation énonciative, sans pour autant prétendre d'avoir trouvé le « bon » grain dans la distinction des différentes catégories.

Nous pensons qu'il serait souhaitable de réaliser un travail analogue sur certaines expressions adverbiales qui exercent, elles aussi, une incidence significative sur l'interprétation des paroles rapportées (*d'une voix forte, sur un ton monocorde, en deux mots, dans un anglais parfait, non sans ironie, avec précision et rigueur, avec mépris, avec raison…).* Spécifiant, selon les cas, certaines caractéristiques du locuteur, du discours produit ou de la relation interlocutive, voire exprimant des appréciations de l'énonciateur sur ces différents aspects du discours rapporté, ces expressions semblent jouer sur le plan énonciatif et sémantique un rôle

[21] Les détails de cette analyse sont présentés dans (Jackiewicz 2007, à paraître).

comparable à celui des verbes introducteurs (17). Il s'agirait de construire un protocole précis de collecte automatique de ces expressions sur corpus. Puis, de vérifier si les valeurs énonciatives et sémantiques dont elles sont porteuses au sein des discours rapportés, permettent de valider les catégories que nous avons constituées à partir du classement des verbes. Une étude sémantique comparative entre ces deux ensembles de marques pourrait alors être entreprise, pour expliquer quel type de valeurs sont préférentiellement portées par quel type de marque.

(17) D'un ton monocorde, parfois vindicatif, Dutroux s'est appliqué à démontrer maladroitement qu'il n'était qu'un « *fusible* », que le procès visait en réalité à cacher « *ce que le peuple ne doit jamais comprendre, afin que la haine bloque la raison* ». Il a aussi dit souhaiter, par moments, qu'un infarctus « *efface les fautes* ». (*Le Monde*, 12 juin 2004)

4. Catégorisations discursives, construction d'ontologies et annotation de textes

Comme nous l'avons montré dans la section 2, l'exploitation avertie des propriétés énonciatives des discours peut conduire à améliorer sensiblement la qualité des résultats obtenus par des applications de résumé, de filtrage ou de fouille de textes. Nous pensons que la catégorisation des opérations discursives, produite entre autres par nos recherches dans le champ de l'énonciation textuelle, pourrait trouver toute son utilité dans diverses tâches d'annotations de segments textuels (pour l'enseignement ou pour la recherche d'information, par exemple).

En particulier, il nous semble intéressant et productif de situer ce travail d'annotation dans le contexte du web sémantique (WS) (Handschuh and Staab, 2003) qui permet de bénéficier de langages standardisés et d'outils exploitant de différentes manières les représentations produites. Un des atouts majeurs du WS est l'utilisation d'ontologies dans le sens spécifique donné à cette notion par l'ingénierie des connaissances (IC) qui les définit, pour schématiser, comme des modèles conceptuels partagés et (ré) utilisables par des communautés de pratique de dimension extrêmement variable. Dans la plupart de ces travaux du WS, les annotations utilisées correspondent à des ontologies d'un domaine applicatif. Pour notre part, nous pensons que pour différentes tâches, il est intéressant d'associer des annotations portant sur des aspects discursifs et permettant de qualifier des portions de texte en fonction de leur rôle dans l'énonciation et par rapport à d'autres segments textuels. On pourra par exemple repérer un certain type de discours rapporté, ainsi que son locuteur avec son type (personne, gouvernement, etc.), en tirant profit de la structuration de l'ontologie.

Avec cet objectif, nous avons entamé récemment une réflexion avec des spécialistes de l'IC, et notamment avec Philippe Laublet. Cela nécessite de réfléchir à la constitution d'une ontologie qui modélise certains aspects de la linguistique du discours. Certains travaux sur les ontologies de la linguistique vont déjà dans ce sens (Farrar and Langendoen, 2003), en notant toutefois que les plus abouties se focalisent essentiellement sur la morphosyntaxe et sur des aspects grammaticaux,

même si certaines réalisations en cours se consacrent à la modélisation de relations rhétoriques en suivant telle ou telle théorie, comme la RST (Goecke *et ali*., 2005). En s'inspirant de théories ou travaux plus récents (encadrement du discours…), on se propose de s'intéresser, de manière cohérente avec nos travaux**,** à l'énonciation textuelle. Une telle ontologie, à laquelle nous entendons contribuer sans prétendre couvrir l'intégralité du champ, peut trouver son utilité dans différentes tâches : annotation d'articles scientifiques de linguistique ou de leurs résumés, indexation de ressources linguistiques et interopérabilité entre descriptions linguistiques, aide à l'enseignement de la linguistique.

De manière plus générale en association avec des ontologies spécifiques, on peut s'intéresser à des situations d'annotation de textes où des utilisateurs cherchent à repérer différents types d'énoncés et leur contexte (journalistes, veilleurs…). L'utilisation des formalismes du WS permet alors, en utilisant des catégories énonciatives et sémantiques bien définies ainsi que des ontologies de domaine, de faciliter les réponses à des requêtes de différents types, la présentation des textes et la navigation intra ou intertextuelle.

5. Conclusion

Les discours, quelle que soit leur nature, sont emprunts de la subjectivité de leurs auteurs et fortement orientés en fonction de l'horizon d'autrui. Cette dimension énonciative est source d'instructions précieuses pour la gestion de l'activité de compréhension, tant par l'humain que par la machine. Ainsi, pouvoir capter et interpréter correctement ces instructions devient aujourd'hui un enjeu à part entière pour plusieurs types d'applications informatiques. On constate toutefois que les stratégies effectives sont souvent très empiriques et ne font que rarement le lien avec des travaux linguistiques touchant à l'énonciation. Du côté de la linguistique, peu de consensus existe sur les catégories pertinentes propres à ce domaine et peu d'études livrent des ressources (marqueurs et connaissances) systématisées exploitables informatiquement.

Le travail présenté ici se situe dans la lignée des recherches sur l'énonciation, accordant une priorité à l'analyse systématique des marques de surface considérées comme traces d'opérations constitutives des objets de discours. Les catégories que nous proposons pour typer et décrire les discours rapportés traduisent la diversité des relations intersubjectives dans lesquelles sont impliqués non seulement les énonciateurs rapportant et rapporté, mais aussi leurs co-énonciateurs respectifs. Elles permettent de dépasser les oppositions binaires « habituelles » : entre l'opinion positive et l'opinion négative d'un côté, et entre la prise en charge et l'absence de prise en charge des contenus, de l'autre. Ces dernières écrasent de toute évidence l'étendue des significations dont peut être chargé l'acte de transmission de paroles étrangères. D'une part, l'intervention de l'énonciateur rapportant peut prendre plusieurs formes d'implication différentes, ce qui nous amène à distinguer quatre

grands types de séquences citationnelles, d'autre part, trois aspects saillants semblent se dégager dans la manière de présenter des paroles d'autrui : (i) faire référence à leur forme d'expression originelle, (ii) focaliser l'attention sur leur contenu, (iii) insister sur la relation (visée) interpersonnelle dont elles étaient porteuses dans leur énonciation d'origine.

Les ressources et les catégorisations proposées demandent de toute évidence à être validées à la lumière des travaux linguistiques similaires et par confrontation à des objectifs précis, notamment pour différentes tâches de recherche d'information (pour la fouille d'opinions ou la veille sociale, par exemple). Sur le plan théorique, elles pourront alimenter les réflexions sur la constitution d'une ontologie linguistique relative au domaine du discours, dont l'utilité immédiate est d'abord d'ordre didactique. Mais, comme nous l'avons dit plus haut, toute une gamme d'applications informatiques qui relèvent du web sémantique pourraient combiner efficacement les ontologies de domaine avec une ontologie des faits de discours, pour améliorer la qualité des annotations textuelles et des traitements qui en dépendent.

Remerciements

Merci à Philippe Laublet, Marie-Paule Péry-Woodley et aux relecteurs anonymes pour leurs remarques et suggestions sur les premières versions de cet article.

6. Bibliographie

Abdalla R., Teufel S., « A bootstrapping approach to unsupervised detection of cue phrase variants ». *Proceedings of ACL/COLING 2006.*

Athier-Revuz J., *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, t.1 et t.2, Larousse, 1995.

Bakhtine M., *Esthétique et théorie du roman*, Tel, Gallimard, 1978.

Beauchêne D., Million-Rousseau C., Rieu Ch., « Détection automatique de l'insatisfaction du client dans un contexte de commerce électronique », *CIFT'2002*, pp.107-117, 2002.

Bestgen Y., « Détermination de la valence affective de termes dans de grands corpus de textes », *CIFT'2002*, pp.81-94, 2002.

Bilhaut F., Ho Dac M., Borillo A., Charnois T., Enjalbert P., Le Draoulec A., Mathet Y., Miguet H., Péry-Woodley M.-P., Sarda L., « Indexation discursive pour la navigation documentaire : cadres temporels et spatiaux dans l'information géographique », *Actes de TALN 2003*, pp. 315-320, 2003.

Blais A., Évaluation de la tâche Résumé dans la plate-forme informatique Context0. Proposition d'étiquetage pour la classe Annonce thématique, Mémoire DEA MIASH, Université de Paris-Sorbonne, 2004.

Busquets J., Vieu L., Asher N., « La SDRT : une approche de la cohérence du discours dans la tradition de la sémantique dynamique », *Verbum*, t. XXIII, n°1, 2001.

Charaudeau P., *Le discours d'information médiatique*, Nathan, 1997.

Coch J., Candito M.-H., « Les outils d'extraction d'information développés par le projet Watson et disponibles dans Lingway KM », *TALN06*, pp.901-909, 2006.

Couto J., Ferret O., Grau B., Hernandez N., Jackiewicz A., Minel J.L., Porhiel, S., « RÉGAL, un système pour la visualisation sélective de documents », *Revue d'Intelligence Artificielle*, vol. 18, n° 4/2004, pp.481-514, 2004.

Culioli A., *Variations sur la linguistique*, Klincksieck, 2002.

Desclés J.-P., Guentcheva Z., « Enonciateur, locuteur, médiateur », *Les rituels du dialogue*, A. Becquelin et Ph. Erikson (eds), Editions de l'Harmattan, 2000.

Eshkol I., Typologie sémantique des prédicats de parole, Thèse de l'Université de Paris XIII, 2002.

Farrar S., Langendoen T., « A linguistic ontology for the Semantic Web». *Glot International* Vo.7, N°3, pp.97-100, mars 2003.

Farzindar A., Résumé automatique de textes juridiques, Thèse de l'Université de Montréal et de l'Université de Parisnne, 2005.

Gaulmyn (de) M.M., Les verbes de communication dans la structuration du discours. Essai sur la réflexivité du langage. Thèse d'Etat, Paris VIII, 1983.

Goecke D., Lüngen H., Sasaki F., Witt A., Farrar S., « GOLD and Discourse. Domain – and Community – Specific Extensions », Beitrag der Projekte A2 & C1 vorgestellt bein *E-MELD Workshop 2005*, Cambridge, MA, 2005.

Gross M., *Méthodes en Syntaxe*. Paris, Hermann, 1975.

Handschuh S., Staab S., *Annotation for the Semantic Web (Frontiers in Artificial Intelligence and Applications)*, Hardcover, 2003.

Hernandez N., Grau B., « Détection automatique de structures fines du discours », *TALN2005*, Dourdan, 2005.

Jackiewicz A., « La causalité dans les textes : une question de point de vue », *Intellectica*, n°38, 2004.

Jackiewicz A., « Transmission dialogique des paroles d'autrui », *Verbum*, 2007, à paraître.

Jacques M.-P., Rebeyrolle J., « Titres et structuration des documents », *ISDD2006*, Caen, 2006.

Kerbrat-Orecchioni C., *L'énonciation. De la subjectivité dans le langage*, A. Colin, 1999.

Krieg-Planque A., « Souligner l'euphémisme : opération savante ou acte d'engagement ? Analyse du 'jugement d'euphémisation' dans le discours politique », *Semen*, n°17, pp.59-80, 2004.

Lamiroy B., Charolles M., « Les verbes de parole et la question de l'(in)transitivité », *Colloque Lexique et Grammaire*, 22 pages, Thessaloniki, 18.09.2003.

Laublet P., Naït-Baha L., Jackiewicz, A., Djioua B., « Collecte d'informations textuelles sur le web selon différents points de vue », *Interaction Homme-Machine et Recherche d'Informations*, pp.261-293, éditions Hermès, Paris, 2002.

Le Monde, *Le style du Monde*, Guide rédactionnel, 2002.

Maingueneau D., « Hyperénonciateur et "particitation"», *Langages*, n°156, pp.111-126, décembre 2004.

Minel J.-L., Cartier E., Crispino G., Desclés J.-P., Ben Hazez S., Jackiewicz A. « Résumé automatique par filtrage sémantique d'informations dans des textes », *Technique et Science Informatiques*, n°3, pp.369-396, Hermès, Paris, 2001.

Nava M., Hypertexte dynamique pour la navigation dans une mémoire de rédaction. Contextualisation automatique de syntagmes nominaux. Thèse de doctorat, Université de Paris-Sorbonne, 2003.

Prasad R., Dinesh N., Lee A., Joshi A., Webber B., « Annotating Attribution in the Penn Discourse TreeBank », *Proceedings of the COLING/ACL Workshop on Sentiment and Subjectivity in Text*, Sydney, Australia. July 2006.

Rabatel A., « Effacement énonciatif dans les discours rapportés et ses effets pragmatiques », *Langages*, n°156, pp.3-17, décembre 2004.

Rosier L., *Le discours rapporté. Histoire, théories, pratiques*. Bruxelles, Paris, Duculot, 1999.

Teufel S., « Discourse structure in scientific articles: argumentation and citation », *ISDD2006*, Caen, 2006.

Vanderveken, D., *Les actes de discours*, Mardaga, Liège, 1988.

Wiebe J., Wilson T., Cardie C., « Annotating Expressions of Opinions and Emotions in Language », *Language Resources and Evaluation*, volume 39, issue 2-3, pp.165-210, 2005.