



HAL
open science

CATEGORISATION SEMANTICO-ENONCIATIVE DU LEXIQUE A PARTIR D'UN DICTIONNAIRE ELECTRONIQUE

Pierre-André Buvet

► **To cite this version:**

Pierre-André Buvet. CATEGORISATION SEMANTICO-ENONCIATIVE DU LEXIQUE A PARTIR D'UN DICTIONNAIRE ELECTRONIQUE. Os di.ci.o.nă.rios Fontes, métodos e novas tecnologias, 2011, pp. 75-96. hal-00677290

HAL Id: hal-00677290

<https://hal.science/hal-00677290>

Submitted on 7 Mar 2012

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.

CATEGORISATION SEMANTICO-ENONCIATIVE DU LEXIQUE A PARTIR D'UN DICTIONNAIRE ELECTRONIQUE

Pierre-André BUVET
laboratoire Lexiques Dictionnaires Informatique
UMR 7187 CNRS-Université Paris 13

Nous présentons une étude en cours qui a pour objectif d'établir, d'une part, comment la modalité et les relations logiques contribuent au mode d'organisation des textes et, d'autre part, de combiner les résultats de l'analyse avec ceux de l'analyse sémantique du lexique afin de modéliser des inférences donnant lieu à la compréhension globale d'un texte. La mise en place d'un prototype de catégorisation des documents sert de cadre à une expérimentation qui fait l'hypothèse que la structuration textuelle joue un rôle de premier plan dans le processus inférentiel. Nous discutons ici principalement du traitement de la modalité et de son impact sur le processus inférentiel.

Après avoir précisé le cadre méthodologique de l'étude, nous présentons les catégorisations énonciative et sémantique du lexique ainsi que leur mode d'articulation et nous montrons ensuite comment exploiter ces catégorisations pour interpréter un texte.

1 Cadre méthodologique

Nous discutons de la notion d'approche intégrée qui sous-tend l'étude des catégories sémantico-énonciatives du lexique, puis nous présentons successivement les principaux travaux sur l'inférence, la modalité et les relations logiques et nous montrons les spécificités de notre approche concernant ces trois problématiques.

1.1 L'approche intégrée des faits de langue

L'analyse syntactico-sémantique du lexique permet d'interpréter les contenus propositionnels constitutifs d'un discours en les rapportant à autant de structures prédicat-argument. Il en a résulté, dans la perspective du Traitement Automatique des Langues, la création de dictionnaires électroniques et de grammaires locales [Buvet 2009a]. L'interprétation d'un discours n'est pas uniquement fondée sur les contenus propositionnels qu'il incorpore, elle repose également sur des facteurs énonciatifs.

D'une façon générale, la conception du traitement sémantique des textes diffère selon qu'il porte sur des unités lexicales ou sur des unités énonciatives. Les deux types de traitement sémantique sont souvent présentés comme antagonistes dans la mesure où le premier procéderait de la linguistique de la phrase, le second de la linguistique de l'énoncé [Reboul & Moeschler 1998]. Nous proposons de dépasser ce clivage en tenant compte de tous les facteurs sémantiques, c'est-à-dire aussi bien d'ordre lexico-syntaxique que d'ordre énonciatif, et de déterminer en quoi leur combinaison contribue à l'interprétation ciblée des textes.

Quatre pistes de recherche sont explorées : l'actualisation, la catégorisation énonciative du lexique, les relations anaphoriques et les relations logiques. L'enjeu est la mise en place d'une véritable linguistique du discours, c'est-à-dire la structuration textuelle, qu'elle soit de nature

lexicale, sémantique ou syntaxique, occupe une place centrale dans l'analyse linguistique. Le traitement de l'inférence est également un enjeu majeur du fait que la compréhension d'un texte découle d'un processus inférentiel [Ganier, Gombert & Fayol 2001].

L'analyse sémantique est distinguée selon trois niveaux discursifs : le niveau logico-sémantique ; le niveau énonciatif ; le niveau interprétatif. Le premier niveau concerne la langue ; il est postulé qu'elle est structurée en termes de prédicat et d'argument. Le second niveau est centré sur la position du locuteur par rapport à ce qu'il énonce¹. Le troisième niveau a trait à la compréhension du discours, c'est-à-dire à la représentation qu'en a l'interlocuteur. Les niveaux logico-sémantique et interprétatif ne sont pas directement observables, seul le niveau énonciatif l'est car il implique la production d'un discours. L'étude des discours occupe une place centrale dans les études linguistiques pour trois raisons : il s'agit de matériaux non préconstruits ; ils constituent l'unique moyen d'appréhender le fonctionnement de la langue ; ils sont révélateurs des mécanismes qui caractérisent l'intention de leur auteur et, ce faisant, ils permettent d'expliquer le processus inférentiel sous-jacent à l'interprétation des messages.

L'analyse des faits de langue est fondée sur les trois fonctions primaires : la fonction prédicative, la fonction argumentale et la fonction actualisatrice [Mejri 2009]. Elles permettent de catégoriser les unités linguistiques sur le plan syntactico-sémantique et d'expliquer leur rôle dans la construction d'un énoncé. Les deux premières fonctions rendent compte notamment du contenu d'un énoncé ; la fonction prédicative stipule quel est l'élément structurant et la fonction argumentale, relativement à l'élément structurant, quels éléments sont mis en relation ou quel élément est qualifié. La fonction actualisatrice fait état des autres éléments, c'est-à-dire ceux dont dépend l'instanciation de la structure prédicat-argument dans l'énoncé. La fonction prédicative et la fonction argumentale concernent également la structure des textes ; la première caractérise les connecteurs logiques, explicites ou implicites, la seconde les énoncés qui sont dans la portée des connecteurs.

La linguistique du discours se caractérise par une approche intégrée : les différentes facettes du traitement sémantique sont considérées comme autant de parties intégrantes car elles se rapportent conjointement au lexique, à l'énonciation et à la compréhension. Pour ce qui est du dernier point, l'analyse du processus inférentiel est également une partie intégrante dans la mesure où elle porte sur le transfert entre deux représentations sémantiques, l'une centrée sur le locuteur, l'autre sur l'interlocuteur.

Il s'agit d'obtenir comme résultat final une typologie des éléments linguistiques qui entrent en ligne de compte dans le processus inférentiel. D'un point de vue applicatif, il s'agit d'implémenter la typologie dans un système dédié à la compréhension ciblée des textes.

1.2 État de l'art

Nous présentons des travaux sur la modalité, les relations logiques et l'inférence.

L'approche syntaxique de la modalité utilise la notion d'opérateur, et intègre la modalité au langage formel de la syntaxe sous la forme du modalisateur M opérant sur une prédication Px

¹ Signalons que l'énonciation ne se limite pas à la question de la modalité ; elle implique également la deixis, cf. BOSREDON, B. & FISHER, S. 1992.

avec, par défaut, une valeur assertive. Les modalités forment alors des listes fermées, calculées à partir des formes verbales et de certains types d'opérateurs (semi-auxiliaires, adverbes, etc.) [de Jongh & Veltman, 1988].

L'approche sémantique de la modalité privilégie l'idée d'opération sur celle de marqueur [Martin 1983]. La modalité est conçue comme le résultat d'une opération de calcul de la valeur de vérité d'un énoncé, et non pas la simple combinatoire d'un prédicat et d'un modalisateur. L'assertion est considérée comme une modalité inhérente, que l'on définit comme une prise en charge de l'énoncé par le locuteur et peut être affectée par des opérations de suspension ou de modification de la valeur de vérité (interrogation, subordination, prédicats de second ordre). Dans une telle perspective, l'interprétation modale d'un énoncé se construit à partir de la valeur modale décelée dans certains de ses éléments.

Des études des relations logiques telles que la cause, la finalité, l'hypothèse, etc., il ressort que le terrain privilégié de l'inférence n'est pas seulement le sous-codage, mais aussi le codage simple, de telle sorte qu'on développe un contenu plus riche à partir du contenu codé [Gross & Prandi 2004] et [Hybertie 1996]. Plusieurs critères ont été mis en place pour modéliser le fonctionnement de l'enrichissement inférentiel, comme la compatibilité sémantique entre prédicats, la cohérence des structures conceptuelles et cognitives partagées, la fréquence, etc. [Prandi 2004].

Les travaux linguistiques classiques partagent l'idée que le mécanisme de l'inférence est une propriété inhérente aux langues naturelles, et que tout dispositif théorique qui ne lui accorde pas de place particulière est limité, sinon voué à l'échec. Les sous-types d'inférence qui sont particulièrement étudiés sont la présupposition et l'implication. La théorie argumentative fournit un modèle d'analyse générale de la langue centré sur l'inférence [Ducrot 1972, Anscombe et Ducrot 1983]. Il s'agit d'élucider les systèmes implicites qui régissent les enchaînements inférentiels à partir des orientations qui servent à construire le sens. Ces orientations sont portées entre autres par des connecteurs comme *pourtant*, *donc*, *alors*, etc.

D'autres travaux d'inspiration cognitiviste s'intéressent aux stratégies d'enrichissement inférentiel [König et Traugott 1988] [Hopper et Traugott 1993] et [Kortmann 1997]. Parmi les acquis de ces recherches, on note l'articulation et le partage de travail entre le codage linguistique et l'inférence.

Plusieurs travaux proposent des formalismes qui rendent compte de l'organisation hiérarchique des discours et des différentes relations qui les structurent, tant du point de vue de la reconnaissance que de la génération. Le formalisme D-STAG est dédié à l'analyse automatique de la structure discursive des textes [Danlos 2009], tandis que le formalisme G-TAG est dédié à la génération de textes à partir de l'extension de la grammaire TAG [Abeillé 2002]. Ces deux formalismes s'inspirent de la théorie des représentations discursives SDRT (Segmented Discourse representation Theory) [Asher 1993], qui a pour objectif non seulement de rendre compte du contenu sémantique d'une information, mais surtout de déterminer la forme logique d'un discours sous forme de système d'inférences.

1.3. Objectifs

L'objectif général de l'étude est d'analyser les modes d'articulation entre les niveaux discursifs afin de décrire le processus inférentiel sous-jacent à l'interprétation ciblée d'un

texte. De ce point de vue, sont particulièrement analysées la modalité, les relations logiques et l'inférence.

La modalité est définie comme la prise en charge par le locuteur du contenu de son discours. Elle est traitée en termes d'actualisation, d'une part, de catégorisation lexico-énonciative, d'autre part.

L'actualisation étant le moyen de produire des énoncés bien formés à partir de structures prédicat-argument, elle implique les catégories énonciatives (personne, temps, aspect, modalité, etc.) qui permettent l'ancrage des prédicats et de leurs arguments dans un discours en fonction de la position du locuteur par rapport à ce qu'il énonce. Les actualisateurs participent directement ou indirectement à l'instanciation d'une structure prédicat-argument selon qu'ils sont nécessaires pour construire des phrases simples ou bien qu'ils ne le sont pas. Les modalisateurs sont le plus souvent des actualisateurs du type indirect dans la mesure où ils ne sont généralement pas obligatoirement constitutifs de phrases simples.

Les modalisateurs sont décrits en fonction de leurs propriétés syntactico-sémantiques. Il s'agit de les analyser en fonction de leur combinatoire avec les structures prédicat-argument instanciées afin de préciser leurs conditions d'occurrences et de déterminer l'impact de leurs valeurs sur celles des contenus propositionnels. Par exemple, l'adverbe *incroyablement* est un modalisateur qui renforce la valeur du contenu propositionnel dans *Il est incroyablement compétent* alors que la locution adverbiale *soi-disant* est un modalisateur qui l'atténue dans *Il est soi-disant compétent*.

La catégorisation lexico-énonciative résulte des propriétés sémantiques des prédicats car elles les impliquent dans un type de modalité. Par exemple, l'adjectif *triste* peut être constitutif d'une assertion, *On pleure quand on est triste*, et, en tant que description subjective, contribuer à l'expression d'une modalité élocutive, *C'est triste de faire cela*. De même, le verbe *applaudir* peut être constitutif d'une assertion, *La presse a applaudi sa performance*, et, en tant que description interindividuelle, peut contribuer à l'expression d'une modalité allocutive, *Je vous applaudis des deux mains*, cf. *infra*.

Les relations logiques sont des valeurs allouées aux rapports entre au moins deux propositions : la cause, l'opposition, la comparaison, etc. Le codage linguistique de ces relations n'est pas homogène ; il peut aller du surcodage au sous-codage en passant par un codage complet ou adéquat. Moins l'expression de la relation est explicite, plus elle exige un effort d'interprétation de la part de l'interlocuteur. Dans le cas du surcodage, les connecteurs, en plus d'exprimer de façon explicite la relation logique, enrichissent celle-ci de valeurs supplémentaires. C'est ce qui se produit dans *Ma commande est arrivée à temps grâce à votre collaboration* du fait que le connecteur *grâce à* code une relation de cause enrichie par un sens positif. Le codage complet correspond à une relation logique codée de façon explicite et non ambiguë (*Ma commande a été annulée car une erreur technique est survenue dans votre site*). Enfin, le sous-codage correspond à l'absence de connecteur (*Ma commande a été annulée ; une erreur technique de plus survenue dans votre site*).

Les emplois des connecteurs sont décrits en fonction de leurs propriétés syntactico-sémantiques. Par exemple *comme* est une comparaison dans *Le tarif de l'écran comme celui des frais d'expédition est très intéressant* et une cause dans *Comme j'avais un code de réduction, le frais de livraison était quasiment négligeable*. Les environnements syntaxiques sont ensuite sémantiquement décrits. Les travaux sur l'environnement syntaxique des

connecteurs servent de point de départ aux travaux sur les relations logiques exprimées sans connecteur. Il s'agit d'établir si les configurations de propositions sémantiquement décrites pour chaque connecteur s'observent également en leur absence et, de ce fait, permettent d'établir des règles qui rendent compte de ces relations logiques.

L'inférence est définie comme le moyen d'interpréter un contenu discursif à partir des éléments de nature lexico-syntaxique, énonciative et contextuelle qu'il comporte, de leurs interactions et de leur mode d'organisation. L'analyse de l'inférence découle de la conception d'un modèle inférentiel. Elle fait l'objet d'une expérimentation afin d'avoir une meilleure compréhension des mécanismes de l'inférence ; les données analysées sont fournies par un moteur d'inférence conçu dans le contexte de la catégorisation automatique des textes.

Le modèle inférentiel est fondé sur l'exploitation de connaissances pragmatiques et métalinguistiques. Il s'agit d'établir l'apport de chaque type de connaissance. L'expérimentation en cours implique la prise en compte de plusieurs paramètres qu'il conviendra de régler de différentes façons pour mettre en avant certains des phénomènes étudiés au dépend d'autres. Pour ce faire, on intervient sur les étiquettes de façon à les neutraliser temporairement, les combiner, les pondérer, etc. Les paramètres étudiés impliquent que l'on ait fait une hypothèse sur le processus inférentiel : s'il dépend à des degrés divers des connaissances stipulées en termes de structure prédicat-argument, de modalité, de relation logique et de situation d'énonciation, les connaissances relatives à la structuration textuelle jouent un rôle de premier plan. On mesure l'influence de chaque type de connaissance sur le processus inférentiel. L'interprétation ciblée des textes sert de cadre de référence pour l'expérience.

Au final, dans le contexte applicatif envisagé, il s'agit de faire une description des mécanismes de l'inférence du point de vue de la sémantique lexicale, de la sémantique énonciative et de leurs interactions afin de valider l'intérêt d'une approche intégrée. La part de l'importance des facteurs extralinguistiques par rapport aux facteurs linguistiques est également étudiée.

2. Catégorisation sémantico-énonciative du lexique

Nous établissons comment la description lexicographique de la prédication peut contribuer à intégrer la sémantique lexicale et la sémantique énonciative dans une même approche des faits de langue. Après avoir présenté succinctement la notion d'emploi prédicatif, nous présentons rapidement la microstructure du dictionnaire des prédicats adjectivaux, puis nous faisons état de la dimension lexicale de l'énonciation.

2.1 La notion d'emploi prédicatif

Le fonctionnement des prédicats relève du niveau logico-sémantique, celui des emplois prédicatifs du niveau énonciatif, *cf. supra*. Un prédicat est instancié dans un énoncé en tant qu'emploi prédicatif. Il peut correspondre à plusieurs emplois prédicatifs. Les particularités des prédicats sont inférées de celles des emplois prédicatifs qu'ils subsument.

La parenté morphologique ne donne pas nécessairement lieu à une parenté prédicative. D'autres facteurs, de nature sémantique et syntaxique, entrent également en ligne de compte. Ainsi, si le verbe *couper*, les noms *coupure* et *coupe*, les adjectifs *coupé* et *coupant* sont morphologiquement apparentés dans la mesure où ils partagent radical *coup-*, ils ont un mode

de fonctionnement beaucoup plus complexe du point de la prédicativité. Considérons les énoncés suivants, extraits de documents trouvés sur le Web :

- (1) *De tout temps, l'homme **a coupé** du bois.*
- (2) *Un katana ça **coupe***
- (3) *Il **s'est coupé** avec des tessons de verre*
- (4) *La présidente des Jardins de Maria-Chapdelaine, Mme Nicole Fortin, a procédé à la **coupe** du ruban marquant l'inauguration du Centre de transformation alimentaire ...*
- (5) *La **coupe** est franche nette et précise, de l'ordre du dixième de mm...*
- (6) *Elles lui ont fait des **coupures** avec des lames de rasoir.*
- (7) *Il avait des **coupures** sur ses deux épaules et des lésions sur la main gauche.*
- (8) *Les ongles sont **coupés** net, précisément et sans danger.*
- (9) *Qu'est-ce qui fait qu'un couteau est **coupant**?*

D'un point de vue sémantique, deux valeurs sont clairement distinguées : la valeur SEPARATION stipulée dans (1), (4), (5) et (8) et la valeur ENTAILLE stipulée dans (3), (6) et (7). Chacune de ces deux valeurs est par ailleurs associée à une valeur de nature causative dans (2) et (9). Il s'ensuit, en premier lieu, l'existence de deux racines prédicatives [Sfar 2007], c'est-à-dire deux prédicats caractérisés par la même forme, *coup-*, mais par des valeurs différentes, SEPARATION et ENTAILLE. Les adjectifs, les noms et les verbes morphologiquement apparentés dans (1) à (9) le sont de deux façons différentes du point de vue de la prédication ; ceux de (1), (4), (5) et (7) sont apparentés en fonction de leur rattachement à la racine prédicative caractérisé par le formant *coup-* et la valeur SEPARATION et ceux de (3), (6) et (7) sont apparentés en fonction de leur rattachement à la racine prédicative caractérisé par le formant *coup-* et la valeur ENTAILLE.

Par ailleurs, au sein d'une même parenté prédicative, un adjectif, un nom ou un verbe peut correspondre à plus d'un emploi. C'est le cas du nom *coupure* ; le premier emploi a une interprétation processive dans (6) et le second une interprétation stative dans (7) [Anscombe 2000]. Le verbe de (2) et l'adjectif de (7) sont des emplois à caractère causatif du fait de la spécificité de leur construction ; l'argument qui sature leur position sujet est différent de celui qui sature celle autre des emplois ; il s'agit du moyen de réaliser l'action, ou d'aboutir à l'état en rapport avec le prédicat. Il en résulte que *couper* et *coupant* ont comme valeur : CAUSATIF_SEPARATION/ENTAILLE. Les emplois de *couper* dans (1) et de (2) et ceux de *coupure* dans (6) et (7) illustrent la nécessité de prendre en compte la construction et la distribution pour caractériser d'une façon générale les emplois.

Remarquons qu'il existe d'autres emplois prédicatifs des adjectifs, des noms et verbes qui partagent le radical *coup-* ; ils sont rattachés à des prédicats différents : *La compagnie lui a coupé le courant, Le courant est coupé, Il y a eu une coupure de courant, Deux droites se coupent en un point, Son ton était très coupant*, etc. Il s'agit également de constituants d'expressions figées [Mejri 2003] :

Un emploi prédicatif est caractérisé par deux sortes de propriétés [Buvet 2000b]. Les premières ont trait à : (i) sa racine prédicative ; (ii) sa classe sémantique ; (iii) son type sémantique ; (iv) son aspect inhérent. Les secondes portent sur : (v) sa construction ; (vi) sa distribution morphosyntaxique ; (vii) sa distribution sémantique. Un prédicat est caractérisé par les propriétés suivantes : (a) une racine ; (b) une classe sémantique ; (c) un domaine d'arguments. Les propriétés (a) et (b) des prédicats sont communes avec les propriétés (i) et (ii) des emplois prédicatifs qui leur correspondent. Lorsque plusieurs emplois prédicatifs se

rapportent à un même prédicat, ils partagent tous ces propriétés. La propriété (c) du prédicat résulte de la propriété (vii) de son unique emploi ou de l'union des propriétés (vii) de ses divers emplois.

Le tableau ci-dessous montre comment ces propriétés distinguent des emplois.

emploi propriété	<i>couper-1</i>	<i>couper-2</i>	<i>coupure-1</i>	<i>coupure-2</i>
(i)	<i>coup-</i>	<i>coup-</i>	<i>coup-</i>	<i>coup-</i>
(ii)	SEPARATION	CAUSATIF_SEPARATION/ ENTAILLE	ENTAILLE	ENTAILLE
(iii)	action	état	action	état
(iv)	duratif accompli	permanent	duratif accompli	provisoire
(v)	X0 V X1 (PREP2 X2)	X0 V	X0 faire DET N PREP1 X1 (PREP2 X2)	X0 avoir DET N PREP1 X1
(vi)	X0= GN X1 = GN	X0 = GN X0 = GN	X0 = GN X1 = GN X2 = GN	X0 = GN X1 = GN
(vii)	X0 = HUMAIN X1 = OBJET ₁ X2 = OBJET ₂	X0 = OBJET ₂	X0 = HUMAIN/OBJET ₂ X1 = HUMAIN X2 = PARTIE_CORPS	X0 = HUMAIN X1 = PARTIE_CORPS

2.2 Le dictionnaire des prédicats adjectivaux

Les emplois adjectivaux ont fait l'objet d'un recensement exhaustif ; plus de 20 000 emplois sont listés et décrits dans un dictionnaire électronique pour ce qui est des adjectifs simples. Un dictionnaire électronique, dit également dictionnaire automatisé, est une ressource linguistique conçue dans la perspective du Traitement Automatique des Langues [Martin 2001]. La macrostructure d'un tel dictionnaire doit être totalement normalisée de telle sorte que les informations spécifiées dans le dictionnaire le soient d'une façon homogène et, de ce fait, puissent faire l'objet d'un traitement informatique.

Chaque emploi adjectival correspond à une entrée du dictionnaire ; elle est associée à 7 types de descripteurs faisant état des particularités de la structure argumentale de l'adjectif (le nombre d'arguments, leur spécificité morphosyntaxique et leur nature sémantico-lexicale), de son mode de rection (s'il y a lieu), de sa racine prédicative et de sa classe sémantique ; deux autres types de descripteur, dits catégorie 1 et catégorie 2, permettent de stipuler la catégorisation sémantico-énonciative du lexique.

Le tableau ci-dessous montre comment la polysémie est traitée dans le dictionnaire :

champ 1 entrée	champ 2 X0	champ 6 classe
--------------------------	----------------------	--------------------------

<i>abrupt</i>	<i>penne</i>	INCLINAISON
<i>abrupt</i>	<i>montagne</i>	INCLINAISON
<i>abrupt</i>	<i>versant</i>	INCLINAISON
<i>abrupt</i>	HUMAIN	BRUSQUERIE
<i>abrupt</i>	<i>comportement</i>	BRUSQUERIE
<i>abrupt</i>	<i>ton</i>	MEPRIS
<i>abrupt</i>	<i>style</i>	STYLE_MAUVAIS

Les prédicats adjectivaux du dictionnaire sont décrits par environ 1100 classes sémantiques. Une classe sémantique caractérise des prédicats apparentés en fonction de leurs valeurs équivalentes ou du fait qu'ils sont tous hyponymes d'un même hyperonyme. Les emplois prédictifs ont la même classe sémantique que les prédicats qui les subsument. Les adjectifs qui correspondent à des prédicats définis la classe METEO_MAUVAISE ont tous une valeur similaire. Ils n'ont pas tous les mêmes contraintes de sélection :

*Un(e) (*brouillard + *ciel+ *climat + hiver+ *horizon + *journée + *mai+ *matin + *région+ *saison+ *soir+ *temps+ vent) âpre*
*Un(e) (brouillard + ciel+ *climat + *hiver+ *horizon + *journée + *mai+ *matin + *région+ *saison+ *soir+ temps+ *vent) automnal*
*Un(e) (*brouillard + ciel+ *climat + *hiver+ horizon + *journée + *mai+ *matin + *région+ *saison+ *soir+ temps+ *vent) bouché(e)*
*Un(e) (*brouillard + *ciel+ *climat + *hiver+ *horizon + journée + *mai+ matin + région+ *saison+ soir+ temps+ *vent) brouillardeu(x/se)*
*Un(e) (*brouillard + ciel+ *climat + *hiver+ *horizon + *journée + *mai+ *matin + *région+ *saison+ *soir+ temps+ *vent) brouillé*
 etc.

Les adjectifs qui correspondent à des prédicats définis par la classe MALADIE_PSYCHIATRIQUE ont tous la particularité d'occuper la position **A** dans la construction *être A est une maladie psychiatrique*.

être (abandonnique + boulique + aliéné + amnésique + anorexique + autiste + autodestructeur + boulimique + caractériel + catatonique, + confabulant + cyclothymique + dément + dépressif + dépressif + déprimé + ...) est une maladie psychiatrique

Ils ont tous un nom d'humain comme argument :

Le patient est (abandonnique + boulique + aliéné + amnésique + anorexique + autiste + autodestructeur + boulimique + caractériel + catatonique, + confabulant + cyclothymique + dément + dépressif + dépressif + déprimé + ...)

2.3 De la sémantique lexicale à la sémantique énonciative

Pour ce qui est des prédicats et des emplois prédictifs, les classes sémantiques sont donc des informations métalinguistiques, relatives à leur signification, définies en fonction de critères formels [Buvet et Grezka 2007]. Elles permettent, dans le cadre du traitement automatique des langues, de faire de l'étiquetage lexical. Lorsque les dictionnaires électroniques sont associés à des grammaires locales, *i.e.* lorsque les environnements syntaxiques des dictionnaires sont

décrits [Gross 1995], les ambiguïtés sémantiques afférentes à la simple projection des ressources lexicales sur les textes sont très souvent levées.

Afin d'expliquer quel est le rôle des prédicats dans la structuration textuelle, les classes sémantiques qui leur sont rattachées ont fait l'objet d'une première catégorisation notionnelle. Les classes JOIE, PEUR, TRISTESSE, ... ont été rattachées à la catégorie AFFECT parce que les emplois prédicatifs qu'elles décrivent permettent d'exprimer un ressenti psychologique centré sur l'intériorité d'un individu. Les classes ABERRATION, CAPRICE COURAGE, ... ont été rattachées à la catégorie APPRECIATION parce qu'elles décrivent permettent d'exprimer un ressenti psychologique centré sur l'extériorité. Les classes CONTIGUITE, DEPEUPLEMENT, ELOIGNEMENT_SPATIALE, FORTE_INCLINAISON... sont rattachées à la catégorie DESCRIPTION_LIEU parce que les emplois prédicatifs qu'elles décrivent permettent de faire état des particularités des lieux, qu'ils soient naturels comme *montagne, mer, forêt* ou bien qu'ils ne soient pas comme *ville, maison, stade*. En l'état actuel de nos travaux, les classes sémantiques sont ainsi réparties dans une trentaine de catégories notionnelles.

Nous avons rapporté ensuite les catégories notionnelles à trois catégories majeures: DESCRIPTION_SUBJECTIVE², DESCRIPTION_INTERINDIVIDUELLE³ et DESCRIPTION_OBJECTIVE⁴. La première concerne tous les prédicats en rapport avec l'intériorité d'un être humain, la seconde tous les prédicats en rapport avec une relation entre deux êtres humains et la troisième tous les autres prédicats. Les trois catégories majeures sont dites énonciatives car elles permettent d'expliquer l'implication des prédicats dans une modalité lorsqu'ils ont des environnements syntaxiques particuliers.

3. Interprétation automatique des textes

Nous montrons rapidement comment nous prévoyons d'exploiter les descriptions sémantiques des emplois prédicatifs dans les dictionnaires électroniques pour effectuer un calcul inférentiel sur un texte et, ce faisant, l'interpréter automatiquement.

En premier lieu, nous considérons que la modalité est un phénomène discursif que l'on peut rapporter à une analyse syntactico-sémantique du lexique. Un discours est structuré par des modalités qui expriment le degré d'engagement du locuteur par rapport à ce qu'il énonce [Le Querler 1996]. Les modalités sont prises en charge par des modalisateurs (des verbes auxiliaires comme *devoir, pouvoir*, les marques de conjugaison relative au conditionnel, etc.) ou impliquées par des emplois prédicatifs et leurs environnements syntaxiques (*j'ai la certitude qu'il pleut*). Seul le second point nous intéresse ici. Trois types de modalité sont distingués : la modalité allocutive (*je t'avertis qu'il pleut*), la modalité élocutive (*j'imagine qu'il pleut*) et la modalité délocutive (n'importe quelle assertion qui ne comporte aucune des deux autres modalités, par. ex. *Il pleut*) [Charaudeau 1992]. On peut identifier les modalités en fonction des catégories énonciatives des prédicats lorsque les cotextes de leurs emplois s'y prêtent. Par exemple, la modalité élocutive est spécifiée par un emploi prédicatif rattaché à la

² La catégorie DESCRIPTION_SUBJECTIVE subsume les catégories AFFECT, APPRECIATION, COGNITION, ...

³ La catégorie DESCRIPTION_INTERINDIVIDUELLE subsume les catégories AUTORISATION, AVERISSEMENT, CONTRAINTE, ...

⁴ La catégorie DESCRIPTION_OBJECTIVE subsume les catégories DESCRIPTION_ANIMAL, DESCRIPTION_HUMAIN, DESCRIPTION_LIEU, ...

catégorie DESCRIPTION_SUBJECTIVE si son argument en position sujet est une mention du locuteur, typiquement un pronom à la première personne.

En second lieu, il faut allouer un poids sémantique à chaque type de modalité. Le poids sémantique est une valeur, comprise entre 0 et 1, rattachée à des éléments textuels ou des ensembles d'éléments textuels en fonction de leur rôle dans la compréhension d'un texte. Les interactions entre les différents poids alloués font l'objet d'un calcul qui permet d'inférer des informations d'un texte. Le calcul se présente sous la forme d'une formule mathématique intégrant les différents paramètres pris en compte pour analyser le texte et l'interpréter. Il s'agit d'un travail en cours de réalisation. Une expérimentation sur la modalité doit expliquer comment les mécanismes de l'inférence la prennent en compte. Parallèlement, il faut établir qu'elle est sa portée, c'est-à-dire sur quelle portion du texte, elle s'applique.

Nous terminons par un exemple authentique, qui fait état de deux modalités (en rouge et en gras) et des segments textuels qui sont dans leur portée (en gras et souligné).

Un petit tour sur le site officiel du Festival plus tard, **je suis immédiatement soulagé : il y a bien un endroit pour déposer ses valises**. Ouf ! Me voilà donc parti pour un court trajet en train entre Antibes et Cannes. L'émotion n'est évidemment plus la même cette année au moment de gravir les fameuses marches et de m'installer dans la grande salle, **c'est plutôt de la joie de me retrouver à nouveau plongé au cœur de ce tourbillon permanent**.

Dans l'immédiat, les résultats de nos travaux permettent de rajouter automatiquement des étiquettes comme suit :

Un petit tour sur le site officiel du Festival plus tard, ((je) [LOCUTEUR] suis immédiatement (soulagé) [AFFECT_POSITIF_APAISEMENT]) [MODALITE_ELOCUTIVE] : il y a bien un endroit pour déposer ses valises. Ouf ! Me voilà donc parti pour un court trajet en train entre Antibes et Cannes. L'émotion n'est évidemment plus la même cette année au moment de gravir les fameuses marches et de m'installer dans la grande salle, c'est plutôt de la ((joie) [AFFECT_POSITIF_JOIE] de (me)) [LOCUTEUR] [MODALITE_ELOCUTIVE] retrouver à nouveau plongé au cœur de ce tourbillon permanent.

Références bibliographiques

- Anscombe J.-C., 2000, « Éléments de classification des noms processifs », *BULAG* Hors Série, pp. 153-165.
- Anscombe J-C et O. Ducrot, 1983, *L'argumentation dans la langue*, P. Mardaga, Bruxelles.
- Abeillé A., 2002, *Une grammaire électronique du français*, CNRS Editions, Paris.
- Asher N.; 1993, *Reference to Abstract Objects in Discourse*, Dordrecht, Kluwer.
- Bosredon, B. et S. Fisher, 1992. «Étiquetage et objets de représentation ou 'Ce N' impossible», in M.-A. Morel - L. Danon-Boileau (éds.), *La deixis*.
- Buvet P-A., 2009a, « Quelles procédures d'étiquetage pour la gestion de l'information textuelle électronique ? », *L'information grammaticale*, 122, Louvain, Peeters, pp. 40-48.
- Buvet P.-A. 2009b, « Des mots aux emplois : la représentation lexicographique des prédicats », *Le Français Moderne*, 77, 1, Paris, CILF, pp. 83-96.
- Buvet P-A et Grezka A., 2007, . « Élaboration d'outils méthodologiques pour décrire les prédicats du français », *Linguisticae Investigationes* 30 (2) , pp. 217-245.
- Charaudeau P. 1992, *Grammaire du sens et de l'expression*, Paris, Hachette.
- Danlos L. 2009 « D-STAG: a Formalism for Discourse Analysis based on SDRT and using Synchronous TAG », *Proceedings of the 14th Conference on Formal Grammar (FG'09)* pp 1-20.

- de Jongh D et F. Veltman, 1988, *Intensional Logic Lecture Notes*, Amsterdam, Presses de l'Université d'Amsterdam.
- Ducrot O., 1972, *Dire et ne pas Dire*, Hermann, Paris.
- Ganier F et J.-E. Gombert, M. Fayol, 2001, « Discours procédurales et activités mentales : de la compréhension d'instructions complexes à la planification de l'action », *Langages*, 141, pp. 47-63.
- Gross G. et M. Prandi, 2004, *La finalité : fondements conceptuels et genèse linguistique*, Louvain-la-Neuve, Champs linguistiques Duculot.
- Gross M., 1995, « Une grammaire locale de l'expression des sentiments », *Langue française*, 105, pp. 70-87
- Hooper P. et C. Traugott, 1993, *Grammaticalization*, Cambridge, Cambridge University Press.
- Hybertie, C., 1996, *La conséquence en français*, Paris, Ophrys.
- König E. et E.-C. Traugott, 1988, « Pragmatic strengthening and semantic change : the conventionalizing of conventional implicature », in W. Hüllen, R. Schulze (eds), *Understanding the Lexicon. Meaning, Sense and World Knowledge in Lexical Semantics*, Niemeyer, Tübingen, pp. 110-124.
- Kortmann B., 1997, *Adverbial Subordination*, Berlin-New York, Mouton De Gruyter.
- Le Querler N., 1996, *Typologie des modalités*, Caen, Presses universitaires de Caen
- Martin R., 1983, *Pour une logique du sens*, Paris, Presses Universitaires de France.
- Martin, R. 2001, *Sémantique et automate, L'apport du dictionnaire informatisé*, Paris, Presses Universitaires de France.
- Mejri S. (éd), 2003, *Le figement lexical, Cahiers de lexicologie*, 82.
- Mejri S., 2009, « Le mot : problématique théorique », *Le Français Moderne* 77 (1), pp. 68-82.
- Prandi M., 2004, *The Building Blocks of Meaning*, Amsterdam-Philadelphie, John Benjamins.
- Reboul A. et J Moeschler 1998, *La pragmatique aujourd'hui*, Seuil, Paris.