



**L'expression affective dans les textes
Étude des prédicats Adjectivaux, nominaux et verbaux**

**Affective expression in texts
Study of Adjective, Nominal and Verbal Predicates**

Pierre-André Buvet

Université Sorbonne Paris Cité Université Paris 13, France,
pierreandre.buvet@gmail.com

Article information

History of the article- Historique de l'article

Received: 31/10/2018

Accepted : 25/05/2019

Published : 31/12/2019

Abstarct

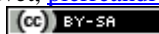
We present an analysis of the emotional expression. Our objective is to supply descriptions formalized of the predicates of affect with the prospect of the automatic understanding of texts.

Keywords: discourse, affect predicates, nouns, adjectives, verbs

Résumé

Nous présentons une analyse de l'expression émotionnelle. Notre objectif est de fournir des descriptions formalisées des prédicats d'affect avec la perspective de la compréhension automatique des textes.

Mots clés : discours, prédicats d'affect, noms, adjectifs, verbes



Introduction

Nous présentons une analyse de l'expression affective en discours. Il s'agit d'un travail qui a pour objectif de fournir des descriptions normalisées des prédicats d'affect dans le cadre général du traitement automatique des textes. La théorie des trois fonctions primaires est la méthodologie sous-jacente à tous ces travaux, *cf.* Buvet 2016 et Mejri 2016.

1. Ressources linguistiques pour le traitement Automatique des langues

1.1. Préalables théoriques

De nombreux travaux en linguistique s'affranchissent de l'analyse des contenus propositionnels et de l'étude du lexique et de la syntaxe qu'elle implique pour ne tenir compte que du discours afin d'expliquer, notamment, comment les contenus propositionnels s'y insèrent. La justification de cette position épistémologique est essentiellement de nature sémantique et fondée sur l'opposition entre langue et discours.

La langue a "comme éléments de départ, un lexique et des règles de grammaire, et comme produit final, des phrases" et le discours "est une manifestation concrète de la langue, et il se produit nécessairement dans un contexte particulier, où entrent en ligne de compte non seulement les éléments linguistiques mais aussi les circonstances de leur production : interlocuteurs, temps et lieu, rapports existant entre ces éléments extralinguistiques" Todorov 1978. Il s'ensuit que l'interprétation des énoncés n'est pas réductible à la signification des mots et de leur combinatoire. L'interprétation dépend également des conditions de production des énoncés. Parmi les concepts étudiés dans les travaux sur l'énonciation, nous prenons en compte uniquement les concepts directement exploitables en Traitement Automatique des Langues : la modalité, les relations inter-propositionnelles et le discours rapporté. Ils constituent autant de paramètres d'analyse qui doivent améliorer l'interprétation automatique des textes.

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, *cf.* infra. L'interprétation d'un discours n'est pas uniquement fondée sur les contenus propositionnels qu'il incorpore, elle repose également sur des facteurs énonciatifs.

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 présentés comme antagonistes dans la mesure où le premier procéderait de la linguistique de la phrase, le second de la linguistique de l'énoncé, *cf.* Reboul & Moeschler 1998. Nous rendons caduque 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 en établissant en quoi leur combinaison contribue à l'interprétation des textes. Nous appelons linguistique de discours cette approche intégrative, *cf.* Buvet 2011. Elle doit permettre d'exploiter les mêmes ressources linguistiques pour utiliser les outils qui analysent les structures prédicat-argument, d'une part, les discours qui les incorporent, d'autre part.

La linguistique du discours se caractérise par une approche intégrée telle que les différentes facettes de l'analyse sémantique sont considérées comme se rapportant conjointement au lexique, à l'énonciation et à la compréhension. Il ne s'agit pas d'une accumulation de traitements sémantiques composites mais d'un traitement homogène qui prend appui sur le lexique pour analyser les autres phénomènes langagiers. L'approche intégrée est fondée sur une distinction du sens en trois niveaux : 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 énonce. Le troisième niveau a trait à la compréhension du discours, c'est-à-dire à la représentation qu'en a l'interlocuteur. Les premier et troisième niveaux ne sont pas directement observables, seul le second niveau l'est car il implique la production d'un discours. Il s'ensuit que l'analyse sémantique consiste en premier lieu à étudier l'articulation entre le premier niveau et le second afin d'identifier les contenus propositionnels et leur mode d'insertion dans les énoncés. Elle implique ensuite d'étudier l'articulation entre le second niveau et le troisième afin d'expliquer le processus inférentiel qui permet le transfert entre deux représentations sémantiques, l'une centrée sur le locuteur, l'autre sur l'interlocuteur.

Le traitement automatique des textes fondé sur une approche intégrée paraît plus adapté pour analyser les verbatims que les traitements entièrement statistiques ou basés sur la projection d'ontologies car le caractère irrégulier et hétérogène des textes les rend réfractaires aux analyses qui s'appuient sur des phénomènes récurrents et font abstraction de la diversité des écrits. Le développement de nouveaux outils pour l'interprétation des textes s'appuie sur des ressources qui résultent de l'approche intégrée telle que nous l'avons définie. Autrement dit, les bases de données lexicales et les banques de graphes qui permettent l'étiquetage sémantique des structures prédicat-argument dans les textes sont complétées avec de nouvelles informations métalinguistiques de telle sorte que de nouvelles étiquettes, faisant état des phénomènes discursifs étudiés, enrichissent les textes traités afin d'en améliorer la compréhension automatique. La conception de bases de règles pallie les difficultés que posent des descriptions incompatibles avec le format des dictionnaires électroniques et celui des grammaires locales.

1.2. Ressources linguistiques

Nous présentons les trois sortes de ressources linguistiques nécessaires à l'étiquetage sémantique des contenus prépositionnels identifiés dans les textes.

1.2.1. Les corpus

Les corpus exploités en Traitement Automatique des Langues sont généralement des textes bruts ou des textes bruts annotés, cf. Mitkov 2003. On peut les distinguer selon leur finalité : corpus de travail, corpus d'apprentissage, corpus d'expérimentation, corpus d'évaluation, etc. Chaque type de corpus correspond à une tâche spécifique. Le corpus de travail est dédié aux recherches qui fournissent des outils linguistiques, le corpus d'apprentissage aux recherches qui impliquent une analyse statistique des textes, le corpus d'expérimentation aux recherches menées dans un cadre expérimental et le corpus d'évaluation a pour fonction de vérifier la qualité des résultats d'une analyse automatique.

L'exploitation d'un corpus de travail nécessite qu'il soit bien profilé. Par exemple, dans le cadre de l'acquisition automatique de vocabulaire, il est impératif d'analyser préalablement les sources de l'information traitée afin de vérifier que les documents textuels qui constitueront le corpus comportent un lexique suffisamment riche en ce qui concerne les thématiques étudiées, cf. Tromeur 2011. De ce point de vue, le recours à des mots-clefs et l'analyse des méta-informations contenues dans les balises qui structurent les textes s'avèrent très insuffisants car, s'ils permettent d'identifier la thématique du document, ils ne garantissent pas d'accéder à un vocabulaire abondant et diversifié en rapport avec le domaine étudié. Typiquement, une majorité d'articles de journaux récupérés à partir des termes 'automobile' et 'industrie de l'automobile' ne sont pas constitués en rapport avec les voitures (*moteur, roue, portière, arbre à came*, etc.). Il est donc impératif d'effectuer en amont de la tâche de constitution de corpus, un travail documentaire et linguistique pour s'assurer que les textes traités conviennent au traitement envisagé.

L'élaboration d'outils performants pour effectuer les analyses linguistiques est fondée sur l'exploitation de trois sortes de corpus de travail : (i) le corpus d'investigation ; (ii) le corpus de test ; (iii) le corpus de validation. Chaque sous-type corpus porte sur des contenus spécifiques pour éviter les solutions ad hoc.

Le corpus d'investigation permet d'identifier les phénomènes linguistiques qui seront pris en compte et traités automatiquement. Compte tenu de ces phénomènes, il doit indiquer en premier lieu quelles ressources lexicographiques sont nécessaires au bon fonctionnement des outils d'analyse. Il doit permettre en second lieu de calibrer les transducteurs à états finis existants et, le cas échéant, de concevoir de nouveaux transducteurs.

La fonction du corpus de test est d'expérimenter les outils calibrés ou nouvellement développés afin de les corriger au fur et à mesure lorsqu'ils donnent lieu à du bruit ou à du silence. Il doit permettre également de pointer d'éventuelles défaillances lexicographiques. L'exploitation du corpus de test doit accompagner au plus près la mise en place des outils afin d'anticiper les difficultés que peut entraîner leur utilisation après leur intégration dans une plateforme.

Le corpus de validation fournit des résultats qui permettent de vérifier la qualité de l'ensemble des outils créés ou paramétrés. En cas d'invalidation, d'autres corpus d'investigation et de test sont utilisés pour améliorer les outils.

Les corpus de test et de validation sont des éléments essentiels à l'élaboration des dictionnaires et les grammaires locales. Ils doivent permettre d'établir puis de mesurer l'adéquation entre ces ressources et les phénomènes traités aussi bien au niveau intra-phrastique qu'au niveau inter-phrastique.

1.2.2. Les dictionnaires électroniques

Les dictionnaires électroniques que nous utilisons sont de deux sortes :

- (i) des dictionnaires morphosyntaxiques au format MORFETIK, *cf.* Mathieu-Colas 2009 ;
- (ii) des dictionnaires syntactico-sémantiques au format PRED-DIC, ARGU-DIC, ACTU-DIC et ETHU-DIC, *cf.* Buvet 2009a.

Le format PRED-DIC concerne les prédicats, le format ARGU-DIC les arguments, le format ACTU-DIC les actualisateurs et le format ETHU-DIC les êtres humains. Nous présentons rapidement le premier format.

La macrostructure des dictionnaires du type PRED-DIC est constituée des différents emplois prédictifs. Les emplois prédictifs sont les différentes instances des prédicats dans les phrases. Le niveau d'analyse des emplois prédictifs diffère de celui des prédicats, *cf.* Buvet 2009b.

La microstructure des dictionnaires du type PRED-DIC est constituée d'une entrée et de deux catégories de descripteurs formels : les descripteurs de définition ; les descripteurs de conditions. La fonction des descripteurs de définition est d'indiquer les propriétés sémantiques des emplois, celle des descripteurs de conditions est de vérifier que les emplois prédictifs ont les propriétés remarquables qui les caractérisent.

Les descripteurs de définition donnent lieu à quatre sortes de spécification : la racine prédictive, la classe sémantique, le type sémantique et l'aspect inhérent. Les spécifications de racine et de classe permettent de préciser la nature de l'articulation entre les emplois prédictifs et les prédicats. Les interprétations des emplois prédictifs résultent des quatre spécifications.

Les descripteurs de conditions donnent lieu à trois sortes de spécifications : la construction ; la distribution morphosyntaxique ; la distribution syntactico-sémantique. Les spécifications de construction permettent de faire état du mode de structuration des arguments par rapport à l'emploi prédictif, du nombre d'arguments spécifiés et de leur caractère obligatoire ou facultatif en position complément. Les spécifications de distribution morphosyntaxique portent sur les particularités formelles des arguments. Les descripteurs consistent à préciser quelles constructions occupent les positions argumentales (groupes nominaux, complétives ou infinitives). Les spécifications de distribution syntactico-sémantique portent sur les particularités sémantiques des arguments. Les descripteurs sont des classes sémantiques.

Les descripteurs de conditions ne désambigüisent pas toujours une forme prédictive donnée. Les propriétés combinatoires des emplois prédictifs peuvent contribuer à la levée d'ambigüité. Elles sont décrites dans des grammaires locales.

1.2.3. Les grammaires locales

Une grammaire locale décrit le cotexte d'une unité lexicale donnée en tant qu'ensemble de configurations de mots, *cf.* Gross 1995,. Elle est représentée par un graphe comportant : un nœud initial, un nœud final et un ensemble de nœuds intermédiaires ; des arcs qui relient les nœuds en fonction des configurations de mots de la grammaire locale. Les nœuds intermédiaires sont associés à des mots, à des lemmes, à des catégories grammaticales ou à d'autres sortes d'informations métalinguistiques

Un automate à états finis est un outil informatique qui permet de définir une grammaire locale et d'analyser une séquence de mots. Un automate à états finis permet de décider si la séquence de mots analysée correspond à l'une des configurations de la grammaire locale.

Un transducteur à états finis est un automate à états finis qui permet d'associer une nouvelle information à de l'information reconnue. Il permet également de remplacer les items d'une séquence reconnue par d'autres items¹.

Qu'il s'agisse d'automates ou de transducteurs, les informations métalinguistiques enregistrées dans les graphes au niveau des nœuds sont celles qui sont encodées dans les dictionnaires électroniques associés au graphe.

3. Les Prédicats d'affect

3.1. Définition

Les prédicatifs d'AFPECT se rapportent à l'intériorité de l'homme du point de vue de ce qu'il ressent psychologiquement mais aussi, de façon parallèle, physiologiquement et cognitivement, *cf.* Buvet & Alii 2005, Tutin & Alii 2006. En tant que prédicats, il s'agit de relations telles que leur nature affective implique d'avoir une entité humaine comme point d'ancrage de la relation. Les domaines d'arguments des prédicats d'affect sont unaires exemple *maussade*) ou binaires (par exemple *hair*) de telle sorte que les relations ont un caractère réflexif dans le premier cas de figure et sont orientés dans le second cas de figure. Lorsque le prédicat est dyadique, l'orientation s'effectue vers l'entité humaine qui est donc à la fois le point d'ancrage et le point d'arrivée de la relation. La notion d'expérienceur rend compte de cette double nature de l'individu affecté, *cf.* Anscombe 1995 et 1996. L'autre argument concerne le point de départ de la relation et, en tant que tel il s'interprète toujours comme une cause, *cf.* Van De Velde 1995. Sa nature est de toute sorte et varie selon prédicats d'affect.

Un prédicat monadique, par définition, ne comporte aucune cause dans son domaine d'arguments dans la mesure où celle-ci implique une autre entité qu'elle détermine et explique, *cf.* Nazarenko 2000. L'argument étant limité à

¹ La première fonctionnalité est fondamentale pour l'étiquetage. La seconde fonctionnalité permet toutes sortes de manipulations des textes. Par exemple, on peut utiliser des transducteurs pour faire des représentations métalinguistiques, remplacer n'importe quelle phrase par sa construction canonique, etc.

une entité humaine, celle-ci a trois propriétés : elle est le point d'ancrage d'un affect spécifique, le point de départ et le point d'arrivée de la relation, ce qui explique son caractère réflexif.

La dénomination des affects a donné lieu en français à un lexique varié qui s'avère parfois très imagé, par exemple, *avoir les boules* comme équivalent d'*éprouver de la contrariété*. Les conditions d'occurrences des unités lexicales sont déterminés par les emplois prédicatifs qu'elles impliquent, cf. *infra*. Avant de préciser ce dernier point, remarquons que le lexique relatif à des affects a deux caractéristiques essentielles : (i) la polymorphie de la plupart des unités monolexicales (*aimer, amour* et *amoureux* dans *Il l'aime, Il éprouve de l'amour pour elle* et *Il est amoureux d'elle* sont trois formes correspondant à un même prédicat; il est en de même pour *énervement, énervant*, et *énervé* dans *Il ressent un certain énervement, Il est énervant, Il l'énerve*); (ii) le figement des unités polylexicales *Il prend facilement la mouche*. Pour ce qui est de ce dernier, divers tests transformationnels permettent de le constater, cf. Mejri 1997.

3.2. Prédicat et emplois prédicatifs

La théorie des trois fonctions primaires est fondée sur deux catégorisations des unités linguistiques : l'une est de nature morphosyntaxique, l'autre de nature syntactico-sémantique. Les deux catégorisations ne portent pas toujours sur les mêmes objets : la première décrit des mots, la deuxième des mots et des morphèmes². L'articulation entre les deux catégorisations s'appuie sur la notion d'emploi. La notion d'emploi est fondamentale en sémantique lexicale. On ne peut pas attribuer systématiquement une valeur à un mot à partir de sa forme ; il faut tenir compte également de ses occurrences dans les énoncés, c'est-à-dire ses conditions d'emploi.

La théorie des trois fonctions primaires postule que les énoncés minimaux sont des structures prédicat-argument actualisées. Par exemple, dans *Ce garçon déteste son père*, le verbe et les deux noms correspondent à une structure prédicat-argument tel que l'emploi verbal du prédicat est actualisé par la marque du passé et les deux arguments nominaux par les déterminants. Il est posé que les prédicats prédominent les arguments car les occurrences des arguments sont subordonnées à celles des prédicats ; le sujet et le complément du verbe *détester* sont obligatoirement des groupes nominaux formée d'un nom d'HUMAIN. Par contre, les occurrences des prédicats ne procèdent pas nécessairement de celles de leurs arguments ; ni *garçon* ni *père* n'exigent de se combiner avec le verbe *détester*, cf. Harris 1971. La structure prédicat-argument conduit ensuite à distinguer les actualisateurs selon qu'ils se rapportent aux prédicats ou aux arguments, cf. Buvet 2013³.

² Par exemple, les actualisateurs sont des mots, les verbes supports entre autres ou des morphèmes, les flexions entre autres. Les prédicats sont des racines ou des mots à part entière, cf. *infra*.

³ La catégorisation syntactico-sémantique ne correspond pas à la catégorisation grammaticale. Par exemple, les noms s'interprètent comme : (i) des prédicats (*faim* dans *Luc a faim*) ; (ii)

Dans le premier cas, les occurrences des actualisateurs dépendent uniquement de celles des prédicats, par ex. l'indication temporelle est imputable au seul verbe. Dans le second cas, elles dépendent des relations entre les prédicats et leurs arguments, la combinatoire des noms avec le verbe est un des éléments d'explication de leur combinatoire avec le déterminant.

Les emplois prédicatifs sont les différentes instances des prédicats dans les énoncés. Il peut s'agir de verbes comme *hérissier* dans *Cela m'hérisse*, de noms comme *liesse* *La foule manifestait sa liesse*, d'adjectifs comme *grognon* dans *Il est grognon*. Certains prédicats sont polymorphiques ; le verbe *aimer*, l'adjectif *amoureux* et le nom *amour* sont trois formes correspondant à un même prédicat dans *Il l'aime*, *Il est amoureux d'elle* et *Il éprouve de l'amour pour elle*⁴. Les emplois prédicatifs sont des unités monolexicales comme *énervé* dans *Cela m'énervé* ou des unités polylexicales comme *porter sur les nerfs* dans *Cela me porte sur les nerfs*, cf. Mejri 2003.

La définition sémantique des emplois est corrélée à leurs propriétés syntaxiques. Par exemple, *peur* a deux significations différentes dans *Il a peur* et *Il a peur que le train n'arrive pas*. On le constate par le biais de la substitution synonymique ; seul le premier exemple autorise le remplacement de *peur* par *frousse*, *Il a (peur + la frousse)* et *Il a (peur + *la frousse) que le train n'arrive pas à l'heure*, tandis que le deuxième autorise le remplacement de *avoir peur* par *redouter* *Il (a peur + *craint)* et *Il (a peur + redoute) que le train n'arrive pas à l'heure*.

Le premier emploi de *peur* est caractérisé par la construction **X0 avoir N (E+ (PREP1) X1)**, le second emploi par la construction **X0 avoir N (PREP1) X1**. Autrement dit, si les deux prédicats nominaux ont la même actualisation, ils se combinent avec l'article zéro et le verbe *avoir*, ils diffèrent pour ce qui est du complément : il est facultatif pour le premier emploi, obligatoire pour le second emploi : *Il a (peur + la frousse)* et *Il a (peur + *la frousse) que le train n'arrive pas à l'heure* *Il a peur (E + du noir)* (premier emploi) ; *Il a peur (*E+ que le train soit en retard)* (deuxième emploi). L'adjectif *peureux* est un emploi prédicatif qui a la même racine prédicatif que le premier emploi de *peur* ; ils relèvent de la même classe sémantique. Pour autant, la différence n'est pas uniquement morphologique car *peureux* n'accepte pas de complément de telle sorte que sa construction est **X0 être A**. Cette différence se traduit du point de vue l'aspect inhérent ; l'emploi nominal est du type passager alors que l'emploi adjectival est du type permanent, cf. Anscombe 2000. La combinatoire avec *continuellement* atteste de cette distinction : *Luc a continuellement peur* vs **Luc est continuellement peureux*

La théorie des trois fonctions primaires donne lieu à des descriptions linguistiques à large couverture sous forme de dictionnaires électroniques et de grammaires dans la perspective du traitement automatique des langues.

des arguments élémentaires (*olive* dans *Luc mange une olive*) ; (iii) des actualisateurs (*kyrielle* dans *Luc a lu une kyrielle de romans*).

⁴ Un prédicat peut correspondre à une préposition comme *sur* dans *Le livre est sur la table*

Ces deux ressources ont comme particularité de décrire tous les emplois des mots avec des informations métalinguistiques standardisées. Celles-ci suffisamment régulières et explicites pour faire l'objet de procédures informatisées. Les emplois prédicatifs des exemples présentés ci-dessus sont traités dans le dictionnaire des prédicats, PRED-DIC, cf. supra.

3.3. Profil combinatoire

L'actualisation d'un prédicat est le fait d'un certain nombre d'unités linguistiques en rapport avec sa morphologie ou bien sa syntaxe : les actualisateurs, i.e. divers morphèmes flexionnels et affixaux, les verbes supports, des prépositions et une partie des déterminants. L'une des particularités de l'actualisation est la signification grammaticale : elle consiste en valeurs sémantiques spécifiques qui sont sus-jacentes ou sous-jacentes aux différents niveaux de structuration des phrases où se situent les unités lexicales, cf. Blanco & Buvet 2005. Les actualisateurs se distinguent selon que leur signification grammaticale est faible ou forte. En ce qui concerne les prédicats, la signification forte est prise en charge essentiellement par des marqueurs aspectuels et intensifs.

L'aspect cotextuel concerne les valeurs suivantes : le duratif, le fréquentatif, l'inchoatif, l'itératif, le ponctuel, le progressif, le semelfactif et le terminatif⁵. Il s'agit de propriétés externes aux prédicats qui sont spécifiées, entre autres, par des déterminants nominaux ou des verbes supports : *Luc a ressenti une bouffée de jalousie* (aspect ponctuel) ; *Luc s'est mis en colère contre Tom* (aspect inchoatif) ; *Luc a ressassé sa rancune envers Tom* (aspect fréquentatif). Divers verbes auxiliaires sont également des marqueurs aspectuels : lorsque les prédicats qu'ils caractérisent correspondent à des adjectifs ou à des noms, ils se combinent obligatoirement avec des verbes supports qui n'ont pas la même signification grammaticale :

Luc a commencé à aimer Zoé ce jour-là (aspect inchoatif); *Luc a commencé à (*devenir + être) amoureux de Zoé ce jour-là* (aspect inchoatif) ; *Luc a commencé à (éprouver de + *succomber à) l'amour pour Zoé* (aspect inchoatif).

La catégorie de l'intensité est binaire : l'intensité forte, d'une part, l'intensité faible, d'autre part. Ce sont également des propriétés externes aux prédicats qui sont prises en charge, notamment, par des verbes supports, des déterminants ou des adverbes⁶ : *Luc est transporté de joie*; *Luc ressent une joie extrême*; *Luc est extrêmement joyeux*. Les réseaux métaphoriques en rapport avec les emplois prédicatifs prennent également en charge l'aspect. Par exemple, la métaphore du feu : *Luc se consume de honte*; *Luc ressent une honte brûlante*.

⁵ Les aspects ponctuel et duratif correspondent tantôt à des propriétés internes tantôt à des propriétés externes.

⁶ L'intensité peut être spécifiée lexicalement, par ex. Il y a une gradation entre les trois verbes de GOUT dans *Luc (apprécie + aime + adore) le vin*.

4. Du lexique au discours

4.1. A propos de la modalité

La modalité est traitée en termes d'actualisation, d'une part, de catégorisation lexico-énonciative, d'autre part. L'actualisation est 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 modalisateurs sont des actualisateurs qui n'ont généralement pas un caractère obligatoire. L'objectif est de les décrire 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*. L'objectif ici est de décrire une modalité à partir de la classe sémantique du prédicat cf. *infra*.

4.2. Modalité et lexique

Les classes sémantiques sont des informations métalinguistiques normalisées et formellement définies. Elles sont constituées à partir des relations de synonymie ou d'hyponymie des items qu'elles caractérisent⁷. La classe COLERE décrit les emplois prédicatifs qui sont synonymes du substantif éponyme de la classe :

Colère/colèreux, courroux/courroucé/courroucer, rogne, etc

La classe PRATIQUE_ALIMENTAIRE est constituée des prédicats hyponymes *anthropogie/anthropophage*,

autophagie/autophage, cannibalisme/cannibale, carnassier, carnivore, etc.

dont l'hyponyme est le substantif éponyme de la classe. Les classes sémantiques sont fondamentales pour procéder à l'analyse d'un texte. Plus de

⁷ Une classe sémantique est nécessaire mais non suffisante pour caractériser un emploi prédicatif. Deux emplois prédicatifs peuvent avoir des interprétations différentes lorsqu'ils héritent de la classe sémantique de leur prédicat commun. Le nom prédicatif **gastrectomie** correspond à deux emplois différents : l'un est de nature processive, *Luc fait une gastrectomie à Tom* ; l'autre est de nature événementielle, *La gastrectomie a eu lieu à 9 heures*.

20 000 emplois adjectivaux sont déjà décrits en termes de classe sémantique. La description des emplois nominaux et verbaux est cours de réalisation. Pour expliquer comment les emplois prédicatifs participent à l'expression d'une modalité, les classes sémantiques qui les caractérisent ont fait l'objet d'une première catégorisation notionnelle. Ainsi, les classes sémantiques JOIE, PEUR et TRISTESSE caractérisent respectivement des adjectifs prédicatifs comme : (i) *allègre, joyeux* ; (ii) *effrayant, peureux* ; (iii) *attristé, désolant*. Elles sont rattachées à la catégorie AFFECT car les emplois adjectivaux qu'elles décrivent dénotent un ressenti psychologique centré sur l'intériorité d'un individu. De même, les classes ABERRATION, COURAGE et SALETE caractérisent respectivement des adjectifs prédicatifs comme : (i) *hurluberlu, illogique* ; (ii) *courageux, hardi* ; (iii) *cloacal, crade*. La catégorie APPRECIATION les subsume parce qu'elles se rapportent à des prédicats adjectivaux qui dénotent un ressenti psychologique centré sur l'extériorité d'un individu. Enfin, les classes CONTIGUITE, DEPEUPLEMENT et INCLINAISON caractérisent respectivement des adjectifs prédicatifs comme (i) *accolé, adjacent* ; (ii) *abandonné, déserté* ; (iii) *escarpé, incliné*. Ces classes sémantiques 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 artificiels comme *ville, maison, stade*. En l'état actuel de nos travaux, plus de 2000 classes sémantiques sont réparties dans une trentaine de catégories notionnelles.

Les catégories notionnelles sont rapportées à trois catégories énonciatives. La première, dite DESCRIPTION_SUBJECTIVE concerne tous les prédicats en rapport avec l'intériorité d'un être humain. La seconde, dite DESCRIPTION_INTERINDIVUELLE, s'applique à tous les prédicats en rapport avec une relation entre deux êtres humains. La troisième, dite DESCRIPTION_OBJECTIVE, a trait tous les autres prédicats. Ces trois catégories sont dites énonciatives car elles permettent d'expliquer l'implication des prédicats dans une modalité lorsqu'ils ont des environnements syntaxiques particuliers. Le tableau ci-dessous rend compte de l'articulation entre les classes notionnelles et les classes énonciatives. Chaque classe notionnelle subsume un ensemble de classes sémantiques.

DESCRIPTION_INTERINDIVIDUELLE	AUTORISATION
	AVERTISSEMENT
	CONTRAINTE
	DEMANDE_ACTION
	DEMANDE_INFORMATION
	DIRE
	ECHANGE
	ENGAGEMENT
	INJONCTION
	LOUANGE
	PROMESSE
REPROCHE	
DESCRIPTION SUBJECTIVE	AFFECT
	APPRECIATION
	APPRECIATION_POSITIVE
	APPRECIATION_NEGATIVE
	CAPACITE
	COGNITION
	DEVOIR
	OPINION
	PERCEPTION
VOLITION	
DESCRIPTION OBJECTIVE	EVENEMENT
	DESCRIPTION_ANIMAL
	DESCRIPTION_ARTEFACT
	DESCRIPTION_HUMAIN
	DESCRIPTION_LIEU
	DESCRIPTION_ODEUR
	DESCRIPTION_VEGETATION

4.3. Résultats

Nous indiquons maintenant comment les descriptions sémantiques des emplois prédicatifs dans les dictionnaires sont exploitées pour analyser automatiquement la modalité

Rappelons que les modalités sont prises en charge par des modalisateurs ou impliquées par des emplois prédicatifs et leurs environnements syntaxiques Seul le second point nous intéresse ici. On identifie les trois types de modalité (la modalité allocutive, la modalité élocutive et la modalité délocutive, *cf.* CHARAUDEAU1992) en exploitant la catégorisation énonciative 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 catégorie DESCRIPTION_SUBJECTIVE si son argument en position sujet est une mention du locuteur, typiquement un pronom à la première personne. Parallèlement, il faut établir qu'elle est la portée de la modalité spécifié, c'est-à-dire sur quelle portion du texte, elle s'applique.

Pour ce faire, on fait appel à des grammaires locales qui étiquettent des structures prédicat-argument et des bases de règles pour spécifier quels cotextes sont adéquats.

Dans l'exemple suivant, il y a deux modalités (en gras et en italique) 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**.

Les ressources élaborées et les outils qui les exploitent rajoutent automatiquement des étiquettes sémantiques aussi bien du type lexical que du type énonciatif :

Un petit tour sur le site officiel du Festival plus tard, ((je) [LOCUTEUR] suis **immédiatement (soulagé) [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) [JOIE] de (me)) [LOCUTEUR] [MODALITE_ELOCUTIVE] retrouver à nouveau plongé au cœur de ce tourbillon permanent.

5. Conclusion

Les travaux sur la modalité, les relations inter-propositionnelles (relations anaphoriques et relations logiques) et le discours rapporté améliorent la qualité des étiquettes sémantiques qui enrichissent les textes. A moyen terme, ces différents jeux d'étiquettes permettront de développer un système dédié à la catégorisation globale et ciblée des textes

Bibliographie

Anscombre, Jean-Claude, 1995, « Morphologie et représentation événementielle : le cas des noms de sentiment et d'attitude », *Langue française* 105, Paris, Larousse, pp. 40-54.

Anscombre, Jean-Claude, 1996, « Noms de sentiment, noms d'attitude et noms abstraits », in *Les noms abstraits, histoire et théories*, Flaux, N., Glatigny, M. et Samain, D. (eds.), Lille, Presses

Universitaires du Septentrion, pp. 257-273.

Anscombre, Jean-Claude, 2000, « "Éléments de classification des noms processifs", BULAG, numéro hors série, *Lexique, Syntaxe et Sémantique, Mélanges offerts à Gaston Gross à l'occasion de son 60e anniversaire*, Presses Universitaires de Franche-Comté, Besançon, pp. 345-364.

- Blanco, Xavier. et Buvet, Pierre.-André, 2004, « Verbes supports et significations grammaticales. Implications pour la traduction espagnol-français », *Linguisticae Investigationes* XXVII : 2, pp. 327-34.
- Buvet, Pierre-André, 2009a, « Quelles procédures d'étiquetage pour la gestion de l'information textuelle électronique ? », *L'information grammaticale*, 122, Louvain, Peeters, pp. 40-48.
- Buvet, Pierre-André, 2009b, « Des mots aux emplois : la représentation lexicographique des prédicats », *Le Français Moderne*, 77, 1, Paris, CILF, pp. 83-96.
- Buvet, Pierre-André, 2011, « Catégorisation sémantico-énonciative du lexique à partir d'un dictionnaire électronique » in *Os di.ci.o.nã.rios Fontes, métodos e novas tecnologias*, Instituto de Letras da Universidade Federal da Bahia.
- Buvet, Pierre-André, 2013, *La dimension lexicale de la détermination en français*, Champion
- Buvet, Pierre-André, 2016, « La fonction argumentale au regard des noms d'artefact », *Actes du colloque CMLF 2016*.
- Buvet, Pierre-André, Girardin, C., Gross, G., et Groud, C., 2005, « Les prédicats d'<affect> ». *LIDIL* 32, Grenoble : Presses Universitaires de Grenoble, pp. 125-143.
- Charaudeau, Patrick, 1992, *Grammaire du sens et de l'expression*, Paris, Hachette.
- Gross, Maurice, 1995, « Une grammaire locale de l'expression des sentiments », *Langue française*, 105, pp. 70-87
- Harris, Zellner Steven, 1971, *Structures mathématiques du langage*, Dunod, Paris.
- Mathieu-Colas, Michel, 2009, « Morfetik : une ressource lexicale pour le TAL », *Cahiers de lexicologie*, 2009-1, 94, pp. 137-146.
- Mejri, Salah, 1997, *Le figement lexical : descriptions linguistiques et structuration sémantique*, Faculté des lettres de la Manouba, Tunis.
- Mejri, Salah, 2003, *Le figement lexical*, *Cahiers de lexicologie*, 82, Champion.
- Mejri, Salah, 2009, « Le mot : problématique théorique », *Le Français Moderne* 77 (1), p. 68-82.
- Mejri, Salah, 2016, « Le prédicat et les trois fonctions primaires », in *Nos caminhos do léxico*, Editora UFMGS, Campo Grande do Sul, Brésil, pp.313-337.
- Mitkov, Ruslan, 2003, *the oxford handbooks of computational linguistics*, Oxford University Press.

- Nazarenko, Adeline, 2000, La cause et son expression en français, *Ophrys*
- Reboul, Anne. et Moeschler, Jacques, 1998, *La pragmatique aujourd'hui*, Paris, Seuil.
- Todorov, Tzvetan, 1978, *Les genres du discours*, Paris, Seuil.
- Tromeur, Laurent, 2011, Mise en place d'une interface en langue naturelle pour la plateforme Ontomantics, *Thèse de doctorat, Université Paris 13*.
- Tutin, Agnès., Novakova, Iva., Grossmann, Francis et Cavalla, Cristelle, 2006, « Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires », *Langue française*, 150, pp. 32-49.
- Van de Velde, Danielle, 1995, *Le spectre nominal*, Peeters, Louvain-Paris.