

Polguère, Alain (2011) Classification sémantique des lexies fondée sur le paraphrasage. *Cahiers de lexicologie* 98, « Du lexique aux dictionnaires en passant par la grammaire. Hommages à Chai-song Hong », Seong Heon Lee (dir.), Paris, Classiques Garnier, 197–211.

CLASSIFICATION SÉMANTIQUE DES LEXIES FONDÉE SUR LE PARAPHRASAGE¹

Résumé

Cet article s'appuie sur le travail de construction et d'exploitation lexicographique du système des étiquettes sémantiques du français mené depuis plus de quinze ans dans le contexte de la Lexicographie Explicative et Combinatoire. Il offre un recul théorique sur la notion, maintenant qu'elle a été intensivement exploitée, tout en permettant une mise à jour de la dernière présentation détaillée qui en avait été faite (en 2003). Après quelques considérations sur le problème de la classification des lexies, nous offrons une caractérisation de la notion d'étiquette sémantique. Nous examinons ensuite la structure globale du système des étiquettes sémantiques du français et concluons par un retour sur l'historique de la notion afin d'éclairer mieux sa portée et ses applications potentielles.

Abstract

This article builds on the work relating to the creation and lexicographic exploitation of the system of French semantic labels conducted for over fifteen years in the context of Explanatory Combinatorial Lexicography. It offers a theoretical step back on the notion, now that it has been intensively exploited; at the same time, it updates the latest detailed presentation (made in 2003). Following a few preliminary considerations on the problem of classifying lexical units, we offer a characterization of the notion of semantic label. We then examine the global structure of the system of French semantic labels, and conclude by returning to the origins of the notion so as to better clarify its scope and potential applications.

Mots-clés : étiquette sémantique, classe sémantique, définition analytique, Lexicographie Explicative et Combinatoire

Keywords: semantic label, semantic class, analytical definition, Explanatory Combinatorial Lexicography

1. Remarques préliminaires sur la classification des lexies

Pourquoi classer les lexies ? Tout d'abord, parce qu'elles forment un ensemble bien trop vaste pour être véritablement appréhendé dans sa totalité. Classer les lexies, c'est en partie donner un sens au lexique lui-même en mettant de l'ordre dans un désordre apparent. Confrontés au gigantisme du lexique, il est normal que nous cherchions à le rationaliser, et que nous procédions comme nous avons appris à le faire depuis notre plus tendre enfance : nous classons pour connaître et reconnaître. Et le lexique n'est pas une entité qui peut se concevoir sous un angle unique. De

¹ Nous sommes extrêmement reconnaissant à Igor Mel'čuk pour ses commentaires sur une première version de cet article.

multiples systèmes de classification des lexies existent, parce que le lexique est tissé de multiples structures : il est multidimensionnel. Rappelons quelques-uns des systèmes de classification lexicale les plus courants, en commençant par les plus « superficiels », pour terminer par les plus « profonds ».

Tout d'abord, les dictionnaires classent les lexies simplement pour permettre un accès à l'information qu'ils encodent. Si le système d'écriture est alphabétique, on classe en fonction de l'ordre canonique des lettres de l'alphabet qui composent le signifiant écrit de la lexie ; on classe aussi par racines consonantiques dans les dictionnaires de langues sémitiques. En l'absence d'alphabet, on classe par « clés » de caractères ou codes numériques inférés des configurations de traits en écriture chinoise, etc. Bref, on peut classer simplement pour ranger en fonction de critères formels fondés sur les signifiants lexicaux.

Ensuite, on peut classer selon la logique grammaticale, en utilisant les parties du discours. Ces classes regroupent les lexies avant tout en fonction de leurs propriétés de combinatoire grammaticale : valence passive, combinatoire morphologique, etc. C'est sur cette classification qu'opèrent les règles de grammaire. Ces dernières ne font pas référence à des lexies individuelles, mais bien à des classes (verbe, nom commun, adjectif qualificatif, etc.), les comportements réguliers qu'elles énoncent s'appliquant alors à n'importe quelle lexie pour peu qu'elle appartienne à la classe en question.

On peut finalement effectuer une classification sémantique des lexies. En effet, il est important de rappeler que la classification par parties du discours n'est aucunement sémantique et que les propriétés sémantiques partagées par toutes les lexies d'une partie du discours donnée sont tellement peu spécifiques qu'elles n'ont guère de pertinence en regard des propriétés grammaticales. Bien entendu, tous les verbes sont des prédicats sémantiques. Mais cela vaut aussi pour tous les adjectifs, tous les adverbes et une bonne partie des noms. Certains noms peuvent être des quasi-prédicats – sens liants dénotant des entités, comme 'ami', 'voisin', 'extrémité'... – ou être des noms sémantiques – sens non liants dénotant des entités, comme 'caillou', 'colline'... –, mais puisque les prédicats nominaux existent, cela n'est aucunement une propriété définitoire des noms². Pour classifier sémantiquement les lexies, il faut créer des systèmes véritablement fondés sur le contenu exprimé par les lexies, totalement ou en grande partie autonome de leurs propriétés grammaticales. Deux types de classements sont alors utilisés.

Premièrement, on peut avoir recours à un classement aristotélicien, ou ontologique. C'est un classement qui structure le lexique de façon hiérarchique, en faisant appel, directement ou indirectement, à la relation d'hyponymie. Il s'agit d'ordonner le lexique du haut en bas : des unités les plus vagues aux unités les plus spécifiques. Parfois, la classification est simultanément sémantique et grammaticale, comme dans le cas de WordNet (Fellbaum 1998)³ – qui scinde la hiérarchisation du lexique selon les parties du discours profondes de Verbe, Nom, Adjectif et Adverbe. Parfois, elle ignore les partitions grammaticales, comme dans le cas de la hiérarchie des étiquettes sémantiques du français, dont il sera question ici.

2 Sur les notions de prédicat, quasi-prédicat et nom sémantiques, voir Mel'čuk et Polguère (2008).

3 <http://wordnet.princeton.edu/>.

Deuxièmement, on peut fonctionner par « champs », en s'attachant à la présence de composantes sémantiques spécifiques dans la définition des lexies. Ainsi, un champ sémantique S – champs sémantiques de l'alimentation, des animaux, des sentiments, etc. – sera une classe de lexies possédant toutes un sémantème correspondant à S en position stratégique dans leur définition, que ce sémantème soit ou non la composante centrale de la définition en question. Plutôt que de regrouper les lexies – unités lexicales –, on peut aussi regrouper les vocables, c'est à dire les « mots polysémiques », selon la même logique. On construit alors des champs lexicaux (et non des champs sémantiques) : des classes de vocables dont la lexie de base appartient à un champ sémantique donné. Selon cette distinction, on dira que le champ sémantique des sentiments contient, notamment, la lexie DOULEUR B ('souffrance psychique'), mais que le champ lexical des sentiments ne contient pas le vocable DOULEUR, parce que la définition de la lexie de base de ce vocable – DOULEUR A ('souffrance physique') – ne met pas en jeu le sémantème 'sentiment'.

CONVENTION 1 – Nous nous appuyerons par défaut, tout au long de l'article, sur la numérotation du *Trésor de la Langue Française informatisé* ou TLFi (Dendien et Pierrel 2003)⁴.

Comme l'illustre le cas de DOULEUR ci-dessus, la notion de champ lexical est opératoire dans un contexte comme celui du travail lexicographique, où l'on doit explorer la structure polysémique des vocables avec un maximum d'homogénéité, alors que la notion de champ sémantique a une valeur dans un contexte où prime la prise en considération des buts communicatifs du locuteur.

Toutes ces observations visent à montrer qu'il n'y a pas et ne peut pas y avoir un unique système de classification des lexies. De par la nature du lexique et de par la nature de l'ensemble du système de la langue (interaction lexique-grammaire), plusieurs classifications sont nécessaires, chacune ayant une finalité donnée. Bien plus, ces différents systèmes de classification doivent être conçus pour interagir. On ne peut ainsi concevoir une classification sémantique des lexies, sans se poser la question de son interaction avec le système de classification grammaticale porté par les parties du discours.

La suite de cet article est consacrée à un système particulier de classification sémantique, celui des étiquettes sémantiques, qui a originellement été conçu dans le contexte d'un travail strictement lexicographique, mais a maintenant développé des ramifications allant bien au-delà de l'agenda initial. Nous voulons notamment revenir sur la dernière présentation du système en question (Polguère 2003a), pour à la fois effectuer une mise à jour de la notion elle-même et en élargir la perspective. Nous procédons pour cela en trois étapes. Nous proposons tout d'abord, section 2, une définition de la notion d'étiquette sémantique, après avoir introduit plusieurs notions fondamentales de lexicologie sur lesquelles elle s'appuie. Ensuite, section 3, nous caractérisons de façon globale le système formé par l'ensemble des étiquettes sémantiques d'une langue comme le français. Finalement, section 4, nous expliquons brièvement, en guise de conclusion, comment nous avons été amené à

4 Rappelons que le TLFi est consultable en ligne : <http://atilf.atilf.fr/tlf.htm>.

élaborer les étiquettes sémantiques dans le cadre d'une entreprise de nature purement lexicographique, alors qu'il nous semble à présent qu'elles peuvent jouer un rôle bien en dehors de ce contexte initial de développement.

Notons que la notion d'étiquette sémantique est universelle, mais que chaque langue possède son propre système d'étiquettes sémantiques. Nos exemples seront tous fondés sur le système des étiquettes du français, qui est, à notre connaissance, le seul étudié avec une certaine systématisme à ce jour⁵.

Nous dédions ces modestes réflexions au Professeur Hong Chai Song, en souvenir de deux séjours mémorables effectués grâce à lui à l'Université Nationale de Séoul, en 2005 et 2006.

2. Notion d'étiquette sémantique

La présentation la plus à jour des étiquettes sémantiques disponible sous forme publiée est Polguère (2003a). Elle est limitative dans la mesure où elle est exclusivement centrée sur le travail fait sur la base lexicale DiCo (Polguère 2000), qui est le projet dans le cadre duquel la notion a été développée. Elle s'embarrasse ainsi de nombreuses idiosyncrasies de l'étiquetage utilisé dans le DiCo à l'époque et ne représente qu'un contexte possible d'utilisation des étiquettes sémantiques parmi d'autres. La présentation de 2003 est aussi en partie dépassée et il convient de tenir compte d'affinements qui ont été apportés, notamment dans le contexte de l'élaboration du *Lexique actif du français* (Mel'čuk et Polguère 2007), désormais désigné sous le nom de *LAF*. Il convient en fait de remettre à plat la notion et de la présenter en elle-même. Nous ne disposons malheureusement pas ici de suffisamment d'espace pour entreprendre une telle présentation détaillée, et nous nous contentons d'introduire les éléments essentiels qui, d'une part, définissent la notion et, d'autre part, révisent certains aspects de celle-ci présentés antérieurement et qui sont maintenant périmés.

2.1. Notions préliminaires

La notion d'étiquette sémantique est purement linguistique, en ce sens qu'il n'est pas nécessaire d'avoir recours à des notions autres que celles de la linguistique pour la construire. Il n'est notamment pas nécessaire de faire appel à des notions relevant de la psychologie ou de la cognition. Bref, les étiquettes sémantiques ne se conçoivent pas relativement au domaine des concepts, mais relativement au système de la langue : signes et règles linguistiques⁶. Les notions fondamentales dont nous

5 On pourra se reporter à Blanco (2007) pour une application de la notion d'étiquette sémantique au lexique verbal de l'espagnol.

6 Il existe bien entendu des connexions entre étiquettes sémantiques et concepts structurant la cognition. Une étiquette sémantique, comme toute entité lexicale, est associée à un concept correspondant. Elle s'identifie cependant strictement à partir du système linguistique et n'est opératoire que dans celui-ci. Si des réflexions doivent être menées sur un concept associé à une étiquette sémantique, alors ces réflexions concernent le concept lui-même. Notamment, ce concept peut être classifiant, comme l'est une étiquette sémantique, mais il ne classe alors que d'autres concepts. L'étiquette, elle, comme nous le verrons plus bas, classe exclusivement les lexies.

avons besoin pour définir ce que sont les étiquettes sémantiques sont les six notions linguistiques de lexème, locution, unité lexicale = lexie, vocable, entité lexicale et définition analytique. Elles seront brièvement introduites ci-dessous, des présentations détaillées pouvant être trouvées dans les textes d'introduction à l'approche de la Lexicologie Explicative et Combinatoire (Mel'čuk et coll. 1995 ; Mel'čuk 2006 ; Polguère 2008), qui est notre cadre théorique de référence.

1) Un **lexème** – par exemple, COUCHER¹I.A [Pour pouvoir coucher le tableau sur le sol et ainsi pouvoir le nettoyer, enlever son cadre.] – est un regroupement de mots-formes (coucher¹I.A, [je] couche¹I.A, [tu] couches¹I.A...), mais aussi de syntagmes ([j'] ai couché¹I.A, [tu] as couché¹I.A...), que seule distingue la flexion.

2) Une **locution** – par exemple, COUPER COURT [Son arrivée coupa court à la conversation.] – est un regroupement de syntagmes sémantiquement non compositionnels que seule distingue la flexion.

3) Les lexèmes et les locutions sont les unités lexicologiques et lexicographiques. Ce sont donc des **unités lexicales**, appelées ici **lexies**.

4) Un **vocable** – par exemple, COUCHER¹ – est un regroupement de lexies – COUCHER¹I.A... COUCHER¹III.A.3 [Faute de place, elle couchera cette nuit avec sa grand-mère.]... – s'exprimant par les mêmes signifiants et possédant une intersection significative de sens.

5) Une **entité lexicale** est une entité linguistique qui est soit une lexie, soit présente une forme d'analogie avec les lexies. La notion d'entité lexicale ratisse très large. Ainsi, les mots-formes, en tant que signes linguistiques lexicaux, sont des entités lexicales. La notion d'entité lexicale va aussi englober l'ensemble des syntagmes phraséologiques compositionnels, comme les collocations (*mettre au lit*) et les clichés linguistiques (*Je vous en prie*), de même que les fonctions lexicales (Mel'čuk 1996) – comme, par exemple, l'intensificateur Magn ou le verbe support Oper₁ –, qui peuvent être considérées comme des métalexies (Polguère 2003b).

CONVENTION 2 – Dans cet article, nous écrivons en police non proportionnelle les éléments de métalangage descriptif, tels que les noms de fonctions lexicales – Syn, Magn... – et les étiquettes sémantiques – sentiment, individu qui exerce un métier...

6) Finalement, une **définition analytique** – ou définition par genre prochain et différences spécifiques – est une modélisation du sens d'une lexie L qui possède les trois propriétés fondamentales suivantes : (i) c'est une paraphrase exacte de L, (ii) elle est constituée de lexies sémantiquement plus simples que L, (iii) elle est structurée en une composante centrale, le **genre prochain**, qui est la paraphrase minimale de L, et en un ensemble de composantes périphériques, les **différences spécifiques**, qui particularisent le sens de L relativement à son genre prochain et aux autres lexies ayant le même genre prochain.

Voilà deux exemples de définitions analytiques :

(1) *X couche*^{I.A} *Y sur Z*

=

L'individu X **dispose** l'objet physique Y sur Z

- horizontalement sur la superficie de Z
- dans le sens de la longueur de Y

(2) *X couche*^{III.A.3} *avec Y*

=

L'individu X **passé la nuit** avec l'individu Y

- pour dormir
- en étant couché aux côtés de Y

Chacune des définitions (1) et (2) est constituée d'un *definiendum* – qui identifie la forme propositionnelle du lexème prédicatif défini – et d'un *definiens* – la paraphrase définissante, elle-même structurée en genre prochain (ici, en gras) et différences spécifiques. Notons que, pour être rigoureux, il nous faudrait identifier chaque sens lexical utilisé dans les définitions ci-dessus au moyen de numéros d'acceptations⁷.

Comme on le voit, le genre prochain peut être l'hyperonyme de L – DISPOSER hyperonyme de COUCHER^{I.A} –, si une telle lexicalisation existe, ou un syntagme compositionnel (syntagme libre ou collocation) – la collocation *passer la nuit* en (2) –, si la langue n'offre pas d'hyperonyme suffisamment spécifique. Ce dernier point est très important. Il est clair que, dans la hiérarchie des lexies du français, PASSER [une période de temps] est hyperonymique relativement à COUCHER^{III.A.3}. Son sens est cependant bien trop vague pour fonctionner comme genre prochain dans la définition (2) : le genre prochain doit être suffisamment proche du contenu sémantique de la lexie définie pour pouvoir être considéré comme une paraphrase minimale, cette dernière notion étant prise comme primitive et non définie dans notre système notionnel.

2.2. Définition de la notion d'étiquette sémantique

À partir des notions préliminaires présentées ci-dessus, il est aisé de définir celle d'étiquette sémantique.

DÉFINITION – L'étiquette sémantique É de la lexie L est la forme normalisée du genre prochain de la définition de L qui établit l'appartenance de L à une classe sémantique de lexies d'au moins deux éléments.

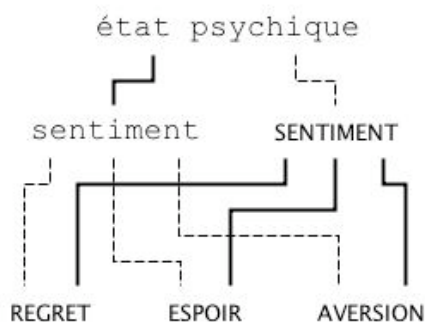
Comme on le voit, la notion d'étiquette sémantique repose presque entièrement sur celle de genre prochain d'une définition. Elle s'en distingue en

⁷ Faute de place, nous ne pouvons justifier en détail le contenu de ces deux définitions. Mentionnons que, dans le cas de (2), nous nous focalisons sur une acception du vocable COUCHER^I tout à fait distincte de celle employée dans une phrase comme *Luce n'a jamais pardonné à son mari d'avoir couché avec une amie d'enfance*. L'acception traitée en (2) dénote une situation où X et Y passent la nuit ensemble (principalement) pour dormir. De plus, nous n'avons pas indiqué que X et Y passaient la nuit dans le même lit, dans la mesure où X peut tout à fait coucher^{I III.A.3} avec Y dans le foin d'une étable ou, si il/elle est moins fortuné(e), à même le sol.

partie par le fait qu'il n'est pas nécessaire que l'étiquette sémantique d'une lexie soit « textuellement » le genre prochain de sa définition ; elle est la variante unique, normalisée, d'une famille d'expressions linguistiques paraphrasant de façon minimale au moins deux lexies de la langue. Des écarts, morphologiques notamment, peuvent donc exister entre la formulation de l'étiquette sémantique et celle du genre prochain. De plus, si le genre prochain est la paraphrase minimale d'une lexie, son étiquette sémantique est, elle, le nom de la classe sémantique lexicale dont la lexie est une instance, dans la hiérarchie des classes sémantiques lexicales de la langue. Nous reviendrons en détail sur ce point, dans la section 3. Il nous faut tout d'abord identifier trois types formels d'étiquettes sémantiques.

2.3. Types formels d'étiquettes sémantiques

Tout genre prochain d'une définition, à moins qu'il s'agisse d'un primitif sémantique, peut être défini ; il possède ainsi lui-même un genre prochain et, donc, une étiquette sémantique. La conséquence logique de cet enchaînement de faits est que chaque étiquette sémantique est elle-même en relation d'hyponymie avec une étiquette plus vague. Par exemple, l'étiquette *sentiment* est hyponyme de l'étiquette *état psychique*. Le rapport entre l'étiquetage sémantique et la subordination hyper-/hyponymique des étiquettes sémantiques et des lexies est illustré dans la figure ci-dessous, où les arcs pointillés indiquent l'étiquetage sémantique et les arcs pleins la relation d'hyper-/hyponymie.



Ce graphe hybride, qui mélange les connexions entre deux types de nœuds – étiquettes sémantiques et lexies –, met en évidence le fait qu'une étiquette « normale » est toujours subordonnée à une autre dont elle est l'hyponyme, cette hiérarchisation fonctionnant en parallèle avec la hiérarchisation correspondante à l'œuvre dans le lexique (= l'ensemble des lexies de la langue). Une **étiquette simple** est une étiquette qui, comme *sentiment*, est hyponyme d'une et une seule autre étiquette du système des étiquettes sémantiques.

Toutes les lexies de la langue ne peuvent cependant pas se classer au moyen d'une étiquette sémantique de ce type. Il existe en effet des lexies dites *sémantiquement ambivalentes*, dont le genre prochain est « bicéphale » : ces lexies peuvent avoir une double dénotation.

Prenons le cas de la lexie *DOCUMENT* ; il est bien connu qu'elle peut être utilisée dans une phrase pour dénoter soit un contenu informationnel, soit le support

physique de ce contenu, soit simultanément les deux, comme l'illustre la phrase ci-dessous⁸ :

- (3) *Les services de sécurité ont saisi de nombreux documents compromettants pour le pouvoir.*

L'ambivalence sémantique de la lexie DOCUMENT doit être reflétée dans le genre prochain de sa définition ; cette propriété sera donc nécessairement reflétée au niveau de l'étiquetage sémantique de la lexie, qui se verra attribuer une **étiquette disjonctive inclusive**, c'est-à-dire une étiquette possédant la forme É1 et/ou É2. Dans le cas de la lexie DOCUMENT, il s'agit de l'étiquette contenu informationnel qu'on communique et/ou support physique de ce contenu.

Au-delà cette formulation particulière, l'étiquette disjonctive inclusive va se distinguer de l'étiquette simple par le fait qu'elle est subordonnée non à une, mais à deux étiquettes hyperonymiques (É1 et É2).

On trouve en langue un autre cas d'ambivalence sémantique, fondé lui sur une disjonction exclusive à l'intérieur du genre prochain de la définition : la lexie peut dénoter deux choses distinctes, mais jamais simultanément dans le même énoncé. Les étiquettes correspondantes sont donc appelées **étiquettes disjonctives exclusives** et ont, par défaut, la forme É1 ou É2 ; comme les étiquettes disjonctives inclusives, elles sont subordonnées à deux étiquettes hyperonymes. Ainsi, la lexie CHAT est étiquetée par animal domestique ou espèce animale correspondante, étiquette subordonnée aux deux étiquettes simples réunies par la conjonction *ou*.

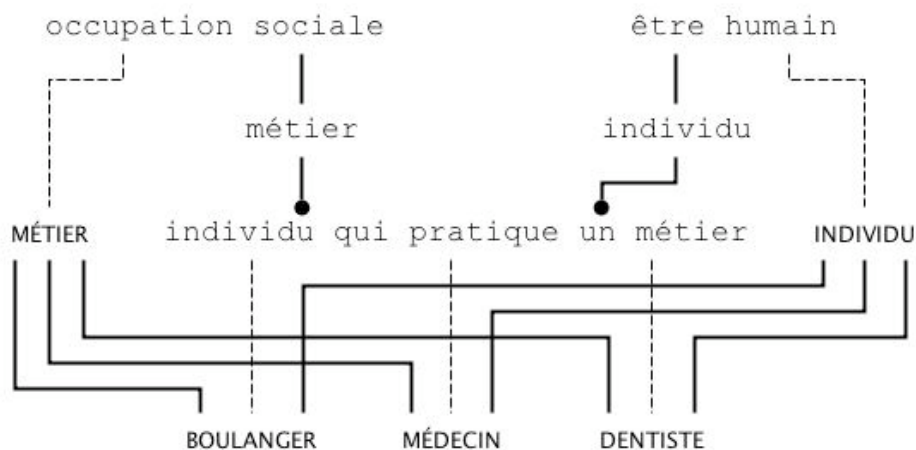
Certaines ambivalences sémantiques sont, pour une langue donnée, non seulement régulières, mais aussi absolument systématiques. Par exemple, en français, tout nom dénotant individu ayant un certain métier – par exemple, MÉDECIN – peut être utilisé pour dénoter le métier en question (mais jamais pour dénoter les deux simultanément) :

- (4) a. *Luce trouve son médecin particulièrement sympathique.*
b. *Médecin est un métier éprouvant pour la santé mentale et physique.*

Pour refléter le fait que cette ambivalence est systématique en français, le LAF a introduit dans un tel cas une étiquette qui « banalise » l'ambivalence en ne la soulignant pas au niveau de la structure formelle (sans disjonction *ou*) : individu qui pratique un métier. Le LAF indique par là même que le genre prochain de la définition correspondante devrait aussi posséder une structure de ce type (fondée sur une subordonnée relative plutôt que sur un syntagme coordonné). Bien entendu, l'ambivalence se trouve tout de même prise en compte par le fait que l'étiquette en question est caractérisée dans le système d'étiquetage comme disjonctive exclusive, subordonnée aux deux étiquettes simples individu et

⁸ Cette propriété est héritée par toutes les lexies dont DOCUMENT est le genre prochain de la définition (directement ou par transitivité), c'est-à-dire par toutes les lexies qui dénotent des types de documents.

métier. La figure ci-dessous illustre le système de connexions gravitant autour d'une étiquette disjonctive exclusive. (La nature disjonctive exclusive, et non inclusive, de l'étiquette est indiquée par les petits cercles sur les arcs hyperonymiques entrants.)⁹



Cette figure montre bien que l'étiquetage sémantique est un véritable système de classification parallèle au système des liens hyperonymiques entre lexies, dont il est hérité. Ainsi, alors que la lexie MÉTIER pourrait être considérée comme un hyperonyme classifiant pour BOULANGER, MÉDECIN, etc., la lexie INDIVIDU est clairement trop vague pour classifier ces lexies ambivalentes sous l'angle de leur autre dénotation possible. C'est alors que l'étiquette sémantique compositionnellement construite individu qui pratique un métier entre en jeu. C'est elle qui offre la véritable classification sémantique, cette dernière devant reposer sur un système formel hérité du lexique, mais autonome vis-à-vis de lui.

L'examen des trois types formels d'étiquettes sémantiques nous a permis d'introduire indirectement plusieurs aspects importants de la structure globale du système d'étiquetage sémantique de la langue, notamment sa nature de hiérarchie à héritage multiple. Nous allons maintenant nous centrer sur ce système dans sa globalité.

3. Structure globale du système des étiquettes sémantiques

3.1. Classification sémantique et parties du discours

Pour présenter la structure globale du système des étiquettes sémantiques, il convient d'examiner le rapport qu'entretiennent ces dernières avec un autre outil de classification des lexies : les parties du discours. Commençons par énoncer, à cet égard, une propriété fondamentale des étiquettes sémantiques.

⁹ On trouvera dans Milićević et Polguère (2010) une présentation de l'ambivalence sémantique et de sa modélisation lexicographique au moyen, notamment, des étiquettes sémantiques. Comme indiqué dans Polguère (2007a), il existe une parenté assez forte entre les étiquettes sémantiques disjonctives inclusives et les types pointés (*dotted types*) de l'approche du lexique génératif.

PROPRIÉTÉ – L'étiquette sémantique d'une lexie L, puisqu'elle correspond au genre prochain de la définition de L, possède nécessairement la même partie du discours profonde que L – Verbe, Nom, Adjectif, Adverbe ou Clausatif (= Énoncé).

Le système des étiquettes sémantiques est structuré autour des étiquettes nominales, car le Nom est justement l'entité lexicale de « nommage » ; c'est lui qui permet le mieux de désigner les classes de lexies. L'attribution d'une étiquette à une lexie, cependant, est entièrement fondée sur le paraphrasage (genre prochain en tant que paraphrase minimale) et le système d'étiquetage doit donc offrir, pour chaque étiquette nominale, une étiquette verbale, adjectivale, adverbiale et clausative correspondante. Ces étiquettes « dérivées » permettent de classer ce que l'on désigne en termes de fonctions lexicales comme étant les V_0 (équivalents verbaux), A_0 (équivalents adjectivaux) et Adv_0 (équivalents adverbiaux) des lexies nominales étiquetées par les étiquettes sémantiques de base. (Les clausatifs dérivés – par exemple, la locution clausative MES EXCUSES ! dérivée de la lexie EXCUSE – peuvent être considérés comme des cas particuliers de V_0 performatifs.)

En plus des étiquettes de lexies unies par les liens d'équivalence sémantique qui viennent d'être examinés, il est aussi nécessaire de connecter directement les étiquettes de lexies liées par des dérivations sémantiques pleines, mais très directement impliquées dans le paraphrasage. Il s'agit notamment des trois types de liens suivants.

1. Fonction lexicale S_1 : ADMIRATION \rightarrow ADMIRATEUR = 'personne qui éprouve de l'admiration' [= S_1 (ADMIRATION)]
2. Fonction lexicale A_1 : ADMIRATION \rightarrow ADMIRATIF = 'qui éprouve de l'admiration' [A_1 (ADMIRATION)]
3. Fonction lexicale $Able_1$: ADMIRATION \rightarrow ADMIRABLE = 'envers qui/quoi on tend à éprouver de l'admiration' [$Able_2$ (ADMIRATION)]

Tout cela fait que chaque étiquette sémantique est en fait élément d'une **métaclasse sémantique** réunissant l'étiquette nominale de base et ses variantes paraphrastiques dans les autres parties du discours. Par convention, nous écrivons {É} la métaclasse regroupant l'étiquette nominale É et ses variantes associées. Par exemple, la métaclasse {sentiment} est, minimalement, le regroupement des étiquettes suivantes : sentiment, éprouver un sentiment (V_0), relatif à un sentiment (A_0), en éprouvant un sentiment (Adv_0), individu qui éprouve un sentiment (S_1), qui éprouve un sentiment (A_1), qui tend à éprouver un sentiment ($Able_1$), envers qui/quoi on tend à éprouver un sentiment ($Able_2$).

3.2. Hiérarchisation hyper-/hyponymique des étiquettes sémantiques

Les observations qui précèdent montrent que, contrairement à une approche de type WordNet, le système des étiquettes sémantiques – qui est structuré hiérarchiquement grâce à la relation d'hyper-/hyponymie – n'est aucunement la réunion de plusieurs sous-hiérarchies, chacune étant associée à une partie du discours profonde. Il est essentiel pour nous que les étiquettes de lexies comme

ADMIRER, ADMIRATION, ADMIRATIF, ADMIRABLE, etc. soient toutes positionnées dans une même aire du système des étiquettes sémantiques, parce que toutes ces lexies ont une définition dont le genre prochain est structuré autour du même sémantème 'admiration'. Les arguments psycholinguistiques avancés à l'origine du projet WordNet pour justifier la compartimentalisation du lexique selon les parties du discours profondes (Miller et coll. 1990) sont à notre avis non pertinents, au moins pour une classification sémantique des lexies. Ils reposent sur des expérimentations de nature tout à fait behavioriste, qui ne font rien de plus que confirmer le fait qu'un locuteur aura plutôt tendance à faire une association immédiate entre un verbe et un verbe, un nom et nom, etc. Cela n'indique aucunement une fragmentation intrinsèquement grammaticale du lexique, puisqu'il s'agit tout simplement d'une conséquence du fait que la relation lexicale fondamentale est la synonymie et que la synonymie fonctionne à l'intérieur d'une même partie du discours.

La bonne approche, selon nous, est d'imposer qu'une étiquette sémantique possède bien la même combinatoire grammaticale générale que les lexies qu'elle classe, mais de s'assurer en même temps que la structure du système de classification sémantique permette de refléter la proximité directe de lexies qui appartiennent à des parties du discours distinctes tout en étant en relation de (quasi-)équivalence sur le plan sémantique. C'est ce « tour de magie » que la structure complexe du système des étiquettes sémantiques présenté ici permet de réaliser. C'est ce qui le rend, selon nous, à la fois tout à fait original et opératoire en lexicologie, ainsi que dans tous les domaines d'application de cette discipline.

Il nous faut maintenant faire une mise au point par rapport à ce qui était jusqu'à présent notre approche de la structuration formelle du système des étiquettes sémantiques. Dans la pratique, même si ce sont les étiquettes elles-mêmes qui sont subordonnées les unes aux autres par la relation d'hyper-/hyponymie, le système des étiquettes sémantiques devrait être implémenté informatiquement non comme une hiérarchie d'étiquettes (\mathbb{E}), mais comme une hiérarchie de métaclasse sémantiques ($\{\mathbb{E}\}$). Ce n'est pas ce qui a été effectué jusqu'à présent dans le travail sur le DiCo/LAF – système implanté grâce à l'éditeur d'ontologies Protégé et présenté dans Polguère (2003a) –, mais c'est l'approche que nous comptons adopter dorénavant dans le cadre de l'élaboration d'une ressource lexicale de plus grande envergure : le *Réseau Lexical du Français* (Lux-Pogodalla et Polguère à paraître).

Bien entendu, comme il avait déjà été signalé dès l'origine de l'élaboration du système des étiquettes, la structure de ce dernier est et demeurera une hiérarchie non strictement arborescente, qui modélise un système de classes sémantiques à héritage multiple, du fait du phénomène d'ambivalence sémantique des lexies et de l'existence d'étiquettes disjonctives (section 2.3, ci-dessus).

4. En guise de conclusion : retour sur l'origine et l'évolution de la notion

Pour terminer notre présentation de la notion d'étiquette sémantique, il nous semble intéressant de revenir sur son origine. Il s'agit d'une notion qui, comme les *synsets* de WordNet (Fellbaum 1998) ou les classes d'objets (Gross 1994, 2008 ; Buvet et Grezka 2009), est intimement liée à une problématique de modélisation formelle de type lexicographique. Très périphérique au départ, elle est en quelque

sorte montée en grade et a pris une importance que nous n'anticipions pas. C'est ce phénomène que nous tentons brièvement d'expliquer maintenant.

La notion d'étiquette sémantique n'a pas été construite originellement comme une fin en soi, mais « faute de mieux », à partir des trois constatations suivantes.

1) Toute description de nature lexicographique (dictionnaire ou base lexicale) doit intégrer une modélisation du sens de chaque lexie, qui idéalement est le centre de gravité de la modélisation en question.

2) La modélisation du sens lexical la plus opératoire et la plus linguistique par nature est la définition lexicographique, telle qu'on la trouve dans les « bons » dictionnaires de langue ou dans les dictionnaires théoriques, de nature plus formelle, comme ceux élaborés dans le contexte du *Natural Semantic Metalanguage* (Wierzbicka 1985, 1987) ou de la Lexicographie Explicative et Combinatoire (Mel'čuk et coll. 1984, 1988, 1992, 1999).

3) L'état de l'art en traitement automatique de la langue ne permet pas d'effectuer de véritable calcul sémantique fondé sur les définitions lexicographiques, soit parce que celles-ci ne sont pas formalisées du tout (cas des dictionnaires de langue), soit parce qu'elles sont formalisées de façon trop approximative et, surtout, ne couvrent pas un éventail suffisamment riche du lexique de chaque langue.

Il s'agissait alors pour nous de constatations, issues de la pratique des domaines de la lexicologie, de la lexicographie et du traitement automatique des langues. Le terme de *constatations* présupposant que les trois points en question sont valides, le lecteur pourra lui substituer celui de *postulats*, s'il n'adhère pas à ce qui vient d'être énoncé. On peut notamment être en désaccord avec le fait que la définition lexicographique est la meilleure modélisation du sens lexical, et lui préférer par exemple des structures logiques, comme celles du Lexique Génératif (Pustejovsky 1995, 1998). On peut aussi soutenir que, malgré leur absence de formalisation, les définitions des dictionnaires de langue – comme le TLFi et le *Petit Robert*, pour le français, ou le *Longman Dictionary of Contemporary English*, pour l'anglais – sont suffisamment bien construites pour permettre un calcul sémantique véritable en traitement automatique de la langue. Nous ne pouvons entreprendre ici une réfutation argumentée de ces points de vue alternatifs, et nous nous contentons donc de rendre au moins explicites les présupposés de l'approche qui est la nôtre.

Lorsque nous avons décidé avec I. Mel'čuk d'adjoindre à la pratique de la Lexicographie Explicative et Combinatoire un travail de construction d'une base lexicale du français suffisamment formalisée pour pouvoir servir de support au traitement automatique de la langue, les trois constatations ci-dessus ont été prises en considération pour explorer une alternative à la définition lexicographique, alternative qui devait ménager à la fois la chèvre du traitement automatique et le chou de la sémantique définitionnelle. Les étiquettes sémantiques se sont alors imposées comme un outil descriptif approprié, outil que nous n'avons cessé d'affiner et améliorer. Cela s'est fait à la fois par la pratique descriptive, dans le cadre de l'élaboration de la base DiCo et du LAF, et par des réflexions de nature plus théoriques sur l'organisation elle-même du système des étiquettes sémantiques (Milićević 1997, Polguère 2003a, Milićević et Polguère 2010).

Il est important de mentionner que, si des ajustements formels ont été apportés

au système des étiquettes sémantiques, nous ne nous sommes jamais écarté des trois principes suivants dans le cours de son élaboration.

1) Une étiquette sémantique est fondamentalement déterminée, non par un ensemble de propriétés de combinatoire des lexies qu'elle classe (même si de telles propriétés existent), mais par la relation de paraphrase minimale qu'elle entretient avec ces lexies.

2) Chaque langue possède son propre système d'étiquettes sémantiques, qui reflète la structuration particulière de son lexique.

3) Les étiquettes ne doivent pas être posées *a priori*, mais être identifiées strictement de façon inductive, dans le cadre du processus de modélisation lexicographique du lexique de la langue.

En suivant ces trois principes, il a été possible d'élaborer un système d'étiquettes sémantiques pour le français encore très incomplet, mais suffisamment riche pour que l'on identifie des régularités permettant une certaine systématisation du traitement lexicographique de cette langue. Notons que, si le deuxième principe mentionné ci-dessus affirme la non-universalité du système des étiquettes sémantiques, nous pensons que le sommet de la hiérarchie est vraisemblablement équivalent dans toutes les langues. La hiérarchie du français se déploie en effet à partir d'une classe unique {quelque chose}, qui domine immédiatement les deux sous-classes {entité} et {fait}. Ce triumvirat s'est révélé être le fondement des notions centrales en sémantique de prédicat, quasi-prédicat et nom sémantiques – voir en tout début d'article, section 1 ; il donc est probable qu'on le retrouve (en traduction) dans la hiérarchie sémantique de toute langue.

Pour conclure ce bref historique, mentionnons que la notion d'étiquette sémantique ne s'est pas imposée d'elle-même parmi nos collègues praticiens de la Lexicographie Explicative et Combinatoire, de fortes réserves ayant été émises sur, à la fois, sa pertinence et sa validité. Ces réserves se sont peu à peu atténuées dans le cours de la mise en pratique de la notion, notamment dans le cadre de l'élaboration du LAF, où les étiquettes sémantiques ont démontré un véritable potentiel descriptif et théorique (Polguère 2007b).

Malgré cela, il est sain de continuer à considérer les étiquettes sémantiques pour ce qu'elles sont : un système formel hérité de la structuration sémantique du lexique, qui n'est elle-même que la résultante des interconnexions entre définitions lexicographiques. Ce sont ces dernières, et non les systèmes de classification sémantique, qui représentent le seul modèle linguistique véritable du sens lexical¹⁰.

Alain POLGUÈRE
Nancy-Université & ATILF CNRS
alain.polguere@univ-nancy2.fr

BIBLIOGRAPHIE

10 Sur la définition en lexicologie, voir le lumineux article d'A. Wierzbicka (1992).

- BLANCO Xavier (2007) : « Les étiquettes sémantiques comme genre prochain : le cas des verbes », *Verbum*, XXIX(1-2), numéro *Verbes et classes sémantiques* coordonné par Aude Grezka et Françoise Martin-Berthet, Nancy, Presses Universitaires de Nancy, p. 113-125.
- BUVET Pierre-André et GREZKA Aude (2009) : « Les dictionnaires électroniques du modèle des classes d'objets », *Langages*, 176, Paris, Larousse, p. 63-79.
- DENDIEN Jacques et PIERREL Jean-Marie (2003) : « Le Trésor de la Langue Française informatisé : un exemple d'informatisation d'un dictionnaire de langue de référence », *Traitement Automatique des Langues (TAL)*, 44(2), Paris, Lavoisier, p. 11-37.
- FELLBAUM Christiane (1998) : *WordNet : An Electronic Lexical Database*, Cambridge, MA, The MIT Press.
- GROSS Gaston (1994) : « Classes d'objets et description des verbes », *Langages*, 115, Paris, Larousse, p. 15-30.
- (2008) : « Les classes d'objets », *Lalies*, 28, Paris, Éditions Rue d'Ulm, p. 111-165.
- LUX-POGODALLA Veronika et POLGUÈRE Alain (à paraître) : « Construction of a French Lexical Network : Methodological Issues », *Proceedings of the International Workshop on Lexical Resources (WoLeR 2011)*, Ljubljana.
- MEL'ČUK Igor (1996) : « Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon », in Leo Wanner (dir.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia, Benjamins, p. 37-102.
- (2006) : « Explanatory Combinatorial Dictionary », in Giandomenico Sica (dir.), *Open Problems in Linguistics and Lexicography*, Monza, Polimetrica, p. 225-355.
- MEL'ČUK Igor, CLAS André et POLGUÈRE Alain (1995) : *Introduction à la lexicologie explicative et combinatoire*, Paris/Louvain-la-Neuve, Duculot.
- MEL'ČUK Igor et coll. (1984, 1988, 1992, 1999) : *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Volumes I-IV*, Montréal, Les Presses de l'Université de Montréal.
- MEL'ČUK Igor et POLGUÈRE Alain (2007) : *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, Bruxelles, De Boeck & Larcier, coll. « Champs linguistiques ».
- (2008) : « Prédicats et quasi-prédicats sémantiques dans une perspective lexicographique », *Revue de linguistique et de didactique des langues (Lidil)*, 37, Grenoble, Ellug, p. 99-114.
- MILIĆEVIĆ Jasmina (1997) : *Étiquettes sémantiques dans un dictionnaire formalisé du type Dictionnaire explicatif et combinatoire*, mémoire de maîtrise, Département de linguistique et de traduction, Université de Montréal.
- MILIĆEVIĆ Jasmina et POLGUÈRE Alain (2010) : « Ambivalence sémantique des noms de communication langagière du français », *Actes de la section « Lexique et morphologie » du 2^e Congrès Mondial de Linguistique Française (CMLF'10)*, La Nouvelle Orléans, p. 1029-1050.
- MILLER George A., BECKWITH Richard, FELLBAUM Christiane, GROSS Derek et MILLER Katherine J. (1990) : « Introduction to WordNet : An On-line Lexical Database », *International Journal of Lexicography*, 3(4), Oxford, Oxford University Press, p. 235-244.
- POLGUÈRE Alain (2000) : « Une base de données lexicales du français et ses applications possibles en didactique », *Revue de linguistique et de didactique des langues (Lidil)*, 21, Grenoble, Ellug, p. 75-97.
- (2003a) : « Étiquetage sémantique des lexies dans la base de données DiCo », *Traitement Automatique des Langues (TAL)*, 44(2), Paris, Lavoisier, p. 39-68.
- (2003b) : « Collocations et fonctions lexicales : pour un modèle d'apprentissage », in Francis Grossmann et Agnès Tutin (dir.), *Les collocations : analyse et traitement*, Amsterdam, De Werelt, coll. « Travaux et Recherches en Linguistique Appliquée » [*Revue Française de Linguistique Appliquée*, série E, n° 1], p. 117-133.
- (2007a) : « Is "Meaning-Text Generative Lexicon" a Swearword? », *Proceedings of the Fourth International Workshop on Generative Approaches to the Lexicon (GL'2007)*, Paris.
- (2007b) : « Lessons from the Lexique actif du français », *Proceedings of the Third International Conference on the Meaning Text Theory (MTT'2007)*, Klagenfurt.
- (2008) : *Lexicologie et sémantique lexicale. Notions fondamentales*, nouvelle édition, Montréal, Les Presses de l'Université de Montréal, coll. « Paramètres ».
- PUSTEJOVSKY James (1995) : *The Generative Lexicon*, Cambridge, MA & London, MIT Press.
- (1998) : « Generativity and Eplanation in Semantics: A Reply to Fodor and Lepore », *Linguistic Inquiry*, 29(2), Cambridge, MA & London, MIT Press, p. 289-311.
- WIERZBICKA Anna (1985) : *Lexicography and Conceptual Analysis*, Ann Arbor, MI, Karoma.
- (1987) : *English Speech Act Verbs : A Semantic Dictionary*, Sydney, Academic Press.

— (1992) : « Back to definitions : Cognition, semantics, and lexicography », *Lexicographica*, 8, Tübingen, Max Niemeyer, p. 146-174.