



**HAL**  
open science

## Parcours interprétatifs et terminologie

Thierry Poibeau

► **To cite this version:**

| Thierry Poibeau. Parcours interprétatifs et terminologie. 2005, pp.12. hal-00005508

**HAL Id: hal-00005508**

**<https://hal.science/hal-00005508>**

Submitted on 11 Jul 2006

**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.

# Parcours interprétatifs et terminologie

Thierry Poibeau

Laboratoire d'Informatique de Paris-Nord, Université Paris 13 et CNRS UMR 7030

99, avenue Jean-Baptiste Clément, 93430 Villetaneuse

Thierry.poibeau@lipn.univ-paris13.fr

<http://www-lipn.univ-paris13.fr/~poibeau>

---

## Mots clés

Langues de spécialités et langue générale, domaine, référence

## Résumé

Cet article esquisse quelques réflexions liées au renouveau de l'analyse terminologique de ces dernières années. Alors que l'analyse reposait auparavant essentiellement sur les connaissances d'experts du domaine, la linguistique textuelle a proposé un nouveau cadre pour l'acquisition de terminologie. Cette évolution renouvelle la réflexion sur le statut des connaissances obtenues et modélisées : en quoi la langue rend-elle compte de la notion de domaine ? Quelle est la relation entre terminologie et monde physique ? Le terme a-t-il une dimension référentielle ? Nous proposons quelques pistes de réflexion, en nous inspirant largement des travaux théoriques récents dans le domaine.

---

## 1 Introduction

Dans la théorie classique de la terminologie telle que l'a définie Wüster [1981], le terme est une étiquette linguistique sur une unité du monde. Cette conception fixiste du terme est encore largement répandue, par exemple dans la définition qu'en donne l'organisme des standards internationaux (ISO [Sager 1990]) pour qui les termes sont des identifiants uniques et non ambigus (*ideal and unambiguous identifiers whose textual utterances are systematically identical to their base-form*<sup>1</sup>).

L'analyse de corpus a montré que les unités linguistiques qui apparaissent dans les textes sont souvent différentes des formes canoniques enregistrées dans les glossaires ou les terminologies d'entreprises. La théorie classique a alors été progressivement remise en question : les termes sont apparus comme des entités linguistiques complexes, pouvant subir des variations de différentes natures (morphologiques, syntaxiques ou sémantiques) et ayant une valeur dénotationnelle complexe [Gross, 1988 ; Habert et Jacquemin, 1993 ; Jacquemin, 1991, 2001]. L'approche fondée sur le corpus a par ailleurs permis de caractériser les termes en tant qu'entités linguistiques, indépendamment de leur statut cognitif. Cette approche n'est toutefois pas sans poser elle-même de nouvelles questions : en quoi la langue rend-elle compte de la notion de domaine ? Le terme a-t-il une dimension référentielle ? Quelle place accorder à la dimension interprétative ?

Cet article propose quelques pistes de réflexion à partir de ces questions. Nous nous inspirons de recherches récentes en terminologie. Les travaux émanant du groupe

---

<sup>1</sup> Cité par [Jacquemin, 2001]

Terminologie et Intelligence Artificielle ont ainsi inspiré certaines des idées exposées ici (nous pensons en particulier à l'exposé de M. Slodzian et D. Bourigault intitulé « Pour une terminologie textuelle » lors de la conférence TIA'1999). Certaines idées s'en écartent toutefois, notamment en ce qui concerne la réutilisabilité des ressources obtenues à l'issue du travail terminologique. Nous nous situons aussi dans un cadre sémantique large, qui fait la part belle aux théories de F. Rastier ainsi qu'aux théories sémantiques cognitives. Une partie des notions employées le sont toutefois dans un sens légèrement différent du sens classique, nous pensons en particulier à la distinction que nous opérons entre référence et dénotation<sup>2</sup>.

L'article est organisé comme suit. La section 2 décrit brièvement quelques éléments clés du renouveau de l'analyse terminologique, fondée sur l'analyse de corpus. Nous revenons dans la section 3 sur certaines notions essentielles à la théorie terminologique, comme la notion de domaine, de langue de spécialité ou de sous-langage. La section 4 aborde ensuite les questions de modélisation et de réutilisabilité des ressources ainsi développées. Nous proposons enfin dans la section 5 de revenir sur les aspects sémantiques sous-jacents au travail d'analyse terminologique, en traitant notamment de la question de référence du terme.

## 2 Le renouveau de l'analyse terminologique

Dans cette section, nous examinons rapidement quelques-uns des traits qui ont amené à un renouvellement du travail terminologique fondé sur l'analyse de corpus et non plus sur l'introspection de l'expert.

### 2.1 *Processus d'analyse terminologique*

Le travail terminologique peut schématiquement être décrit à travers trois étapes principales : extraction des termes, structuration de la terminologie et modélisation du domaine (mise en relation des termes suivant un modèle formel du domaine).

De nombreux outils d'extraction de terminologies existent. Jacquemin [2001] en fait une large recension et énumère leurs principales caractéristiques<sup>3</sup>. On peut sommairement catégoriser les outils en trois grands types : ceux qui se fondent sur une analyse linguistique (par exemple Lexter [Bourigault 1994]), ceux qui se fondent sur une analyse fondée sur l'apprentissage (ANA [Enguehard 1993]) et enfin ceux qui mêlent ces approches au sein d'une analyse hybride (Acabit [Daille 1994]). Tous ces outils visent à extraire les groupes nominaux qui sont des termes au sein d'un corpus donné. L'usage de patrons et de critères statistiques permet de filtrer une partie des groupes nominaux pour essayer de proposer une liste de candidats termes pertinents. Les résultats demeurent cependant fortement bruités et exigent une validation manuelle qui peut être fastidieuse si les données sont importantes.

La deuxième partie des années 1990 a vu l'apparition de nombreux outils de structuration de terminologie, permettant d'esquisser des liens entre les termes (voir [Grabar et Hamon 2004] pour un aperçu des outils et des techniques utilisées). Le premier travail consiste à regrouper ensemble les séquences qui ne sont que des variantes d'un même terme. Les travaux de Jacquemin [1991, 2001] portent sur cet aspect, sur le plan tant morphologique que syntaxique et sémantique. Au-delà de la variation terminologique, le lien privilégié est bien évidemment le lien taxinomique, qui permet de classer les termes en hiérarchies suivant un lien *est-un* [Sager 1990]. D'autres types de relations ont été étudiées comme la méronymie

---

<sup>2</sup> Cet article doit beaucoup aux remarques avisées de D. Bourigault, A. Condamines, M. Slodzian et I. Tellier, sur une première version de ce texte. Les deux relecteurs anonymes de la conférence ont également fourni des indications qui m'ont aidé à améliorer le contenu de l'article. Enfin, A. Korhonen m'a suggéré la traduction anglaise du titre.

<sup>3</sup> Voir aussi [Cabré *et al.* 200].

[Séguéla, 2001] ou la synonymie [Hamon, 1998]. Des chercheurs ont enfin cherché à compléter les terminologies ainsi élaborées par l'extraction d'autres types de connaissances comme les définitions [Cartier *et al.* 1997] [Rebeyrolle *et al.* 2000] ou les relations de causalité [Jackiewicz *et al.* 2003]. Les techniques permettant de repérer ces liens sont fondées soit sur une analyse syntagmatique (typiquement une batterie de patrons permettant de mettre en relation deux termes distincts) ou sur une analyse distributionnelle (pour faire apparaître des cadres d'occurrence communs).

La mise en relation de la terminologie structurée avec un modèle du domaine est une tâche de représentation des connaissances. Dès les débuts de l'intelligence artificielle, divers formalismes de représentation des connaissances sont apparus. De très nombreux outils ont été développés avec des fortunes diverses (Protégé [Noy *et al.* 2000] est certainement le plus célèbre d'entre eux ; si cet outil est né du développement de systèmes à base de connaissances à la fin des années 1980, il connaît depuis plusieurs années un très grand succès dans le sillage du développement du web sémantique). Nous ne détaillerons pas cet aspect dans cette étude. Citons toutefois Terminae [Biebow et Szulman, 1999] qui est un outil de modélisation conçu pour permettre une modélisation assistée par l'intégration des résultats de plusieurs outils du type de ceux décrits précédemment. Le couplage de l'analyse terminologique et des outils de modélisation permet ici un gain indéniable.

## **2.2 Importance du corpus**

Comme nous l'avons mentionné en introduction, tous ces travaux ont en commun de reposer avant tout sur l'analyse de corpus. L'introspection et l'interview d'expert sont laissés de côté au profit de l'attesté, du corpus assimilé à un « réservoir » de connaissances.

De très nombreux travaux ont porté, au cours de la dernière décennie, sur la notion de corpus (voir [Habert *et al.* 1997] pour une synthèse). Ils s'inspirent généralement de travaux anglo-saxons plus anciens et les définitions classiques (par exemple celles de [Sinclair 1996] ou [Atkins *et al.* 1992]) soulignent le fait que le corpus est une donnée construite en fonction de critères explicites et qu'il doit être représentatif de l'application visée. Une question fortement débattue est celle de la représentativité du corpus, voire de sa clôture [Nazarenko 2004]. Une position dominante semble toutefois se dégager pour affirmer que l'obligation de représentativité amène à élaborer un corpus en prenant en compte les notions de genre, de registre et de point de vue [Bouveret, 1998]. Il semble toutefois impossible de complètement couvrir un domaine à travers un corpus : celui-ci ne peut qu'être représentatif d'une partie des phénomènes visés (pour un point de vue contrasté, voir [Condamines 2003]).

## **3 Notion de domaine et de langue de spécialité**

Nous nous intéressons dans cette section aux liens qui peuvent exister entre les notions de domaine, de langue de spécialité et de sous-langage. Nous essayons tout d'abord de définir ces notions puis nous examinons les conséquences des choix de définition effectués.

### **3.1 Domaine, terminologie et langues de spécialité**

La terminologie reflète l'organisation des unités linguistiques pertinentes, propres à un domaine. La notion de domaine désigne un champ d'activités homogènes mais sa réalité tangible en linguistique est ténue. Elle se manifeste essentiellement à travers la terminologie qui *rend compte* du domaine<sup>4</sup>. On a en conséquence une circularité entre les définitions de

---

<sup>4</sup> Si l'on adopte une perspective relativement classique en terminologie, on peut dire, avec Bruno de Bessé [2000] que « pour exister en tant que terme, une forme linguistique doit désigner un concept appartenant à un domaine et déterminé par une définition ». Le renouveau des travaux en terminologie met en avant la notion de

terminologie et de domaine. Cette circularité n'est pas gênante dans la mesure où la notion de domaine n'intervient qu'à l'arrière plan de l'activité terminologique. D'un certain point de vue, le domaine n'est pas un élément déterminant de la théorie terminologique, il s'agit d'une notion empirique qui reflète le champ de savoir d'un expert (on parle à ce propos d'*expert du domaine*).

L'expert est généralement la personne chargée de valider le travail effectué, en l'occurrence la terminologie proposée par un outil informatique. Or, il est fort probable qu'il n'existe pas d'expert en tant que tel : il s'agit le plus souvent d'un professionnel (scientifique, veilleur, documentaliste, indexeur) qui a une plus ou moins bonne connaissance du domaine visé et, surtout, une connaissance parfois approximative des contraintes liées à la notion de terminologie, aux critères de validation et à la finalité du travail d'analyse linguistique. Il s'agit pourtant d'une tâche essentielle : la validation repose sur l'interprétation de la valeur d'une séquence linguistique par rapport à un domaine donné<sup>5</sup>. La prise en charge du type d'utilisateur est un point difficile des systèmes de traitement des langues mais on commence à voir apparaître des interfaces conviviales, évoluées et adaptables, qui permettent de répondre en partie à ce problème [Boccon-Gibod, 1999].

Le champ d'activité de l'expert définit à son tour un type de langue particulier, dit *langue de spécialité*. Il s'agit d'une notion au cœur de l'activité terminologique, qui est directement associée à la notion de domaine. Une langue de spécialité est une langue propre à un domaine donné. La terminologie reflète l'organisation lexicale des éléments propres à la langue de spécialité en question. Comme nous l'avons vu dans la section précédente, la linguistique textuelle a permis de rénover le travail terminologique en prenant le texte comme source attestée de connaissances, au détriment de l'introspection d'expert. Cette façon de travailler permet d'obtenir un résultat beaucoup plus proche de la réalité linguistique que les terminologies classiques qui enregistrent souvent un savoir théorique éloigné des pratiques et des usages linguistiques en vigueur dans un domaine donné. Mais cette approche, fondée sur *l'attesté*, n'est pas sans susciter des questions nouvelles. Comme il n'existe pas deux textes (corpus) semblables, les termes et les relations entre les termes que l'on peut extraire à partir d'un corpus sont propres à ces textes (ou à ce corpus). Une terminologie est donc un objet intimement lié à sa source, au corpus et à – l'extrême – à l'application qui a permis de le constituer.

### 3.2 *Notion de sous-langage*

Le paragraphe précédent visait à montrer que la notion de domaine n'est pas suffisante pour définir de manière « naturelle » et rigoureuse une langue de spécialité. La notion de *sous-langage* est relativement proche de la notion de langue de spécialité et Harris [1991] a essayé d'en donner une définition formelle. Le travail de Dachelet [1994] sur le français s'inspire largement des éléments proposés par Harris.

Harris a montré, à partir d'une analyse essentiellement distributionnelle et syntaxique, que certains corpus ne font usage que d'une partie de la grammaire de l'anglais. Il serait sans doute plus exact de dire que ces corpus font un usage spécifique de la syntaxe de l'anglais (d'où la notion de sous-langage par rapport à la langue en soi, si tant est que ce concept de langue ait une réalité ; les auteurs prennent souvent des corpus journalistiques comme corpus

---

corpus qui est une « manifestation » de la notion de domaine mais qui ne contribue pas à éclairer les contours de ce dernier.

<sup>5</sup> Une importante contribution à cette question est apportée par Condamines [2003] qui propose même l'existence de « genres interprétatifs » (voir aussi [Rastier *et al.* 1994]). L'interprétation est un processus humain, les outils ne sont à ce stade que des aides à la décision (de validation ou de modélisation).

de référence<sup>6</sup> [Clément, 2001]). L'expérience sur de gros corpus a depuis montré qu'il ne s'agissait pas tant d'une grammaire différente que de différences dans la proportion relative des différentes structures rencontrées, confirmant ainsi l'idée de *continuum* aperçu dans la section précédente. Ainsi, le concept de langue de spécialité reçoit une justification *a posteriori* dans la mesure où l'on sait que le terme fait un usage particulier de certains traits linguistiques comme (en français) la détermination, les nombreux compléments prépositionnels, l'absence de relatives, etc. [Bourigault, 1994].

Les travaux de Biber sur des corpus variés (textes issus de journaux, de romans, de transcriptions de l'oral) montrent ainsi que les traits distinctifs sont davantage affaire de proportion que de présence ou absence [Biber, 1988]. Folch a montré le même phénomène pour la notion de genre, en analysant des corpus variés [Folch *et al.*, 2000]. Même au sein d'un journal comme le Monde, d'une section à l'autre, voire d'un genre à l'autre (article long, dépêche d'agence, commentaire, éditorial) la langue en jeu n'est pas la même. Les traits linguistiques à l'œuvre n'apparaissent pas dans les mêmes proportions. Pour autant, la finalité applicative de ces travaux n'est pas claire. La prise en compte du genre pour l'étiquetage morpho-syntaxique du corpus *le Monde* demande une mise en œuvre complexe pour des résultats qui semblent mitigés<sup>7</sup> [Illouz, 2000].

## 4 Modélisation conceptuelle

Outre ses aspects purement linguistiques, l'analyse terminologique comprend une étape de structuration des données et de mise en relation de ces données avec un modèle formel du domaine considéré.

### 4.1 De la langue au modèle formel

Dans la phase de modélisation, les entités manipulées sont des concepts, c'est-à-dire des abstractions *a priori* indépendantes de la langue. Il s'agit là d'un point de vue relativement classique, qui est bien représenté en sémantique, par exemple dans le formalisme des graphes conceptuels [Sowa, 1984]. Ce formalisme propose en effet un niveau de structuration où les objets sont reliés à des types, eux-mêmes structurés en treillis. On a donc bien ici un niveau linguistique, auquel on fait correspondre un niveau conceptuel, lui-même distinct du monde sensible. On retrouve la distinction classique entre le signifiant, le signifié, le concept (qui, en interaction avec d'autres, forme un modèle ou un espace mental [Johnson-Laird, 2001 Fauconnier 1984]) et l'objet physique<sup>8</sup>.

Cette description s'écarte de la triade sémiotique aristotélicienne en distinguant de manière explicite la notion de concept du signifié (voir par exemple [Depecker, 2000]), mais peut malgré tout se retrouver dans bon nombre de théories. Cette modélisation est toutefois régulièrement remise en cause dans la mesure où les différents niveaux de structuration ne peuvent être complètement indépendants les uns des autres. La façon de nommer les nœuds du treillis de concepts des graphes conceptuels est par exemple une première incursion du linguistique dans le niveau conceptuel. On pourrait certes nommer les types par de simples

---

<sup>6</sup> Cela est fortement contestable : un journal n'offre qu'un état de langue particulier, même s'il inclut souvent différents genres textuels (on n'écrit pas de la même façon un article de politique étrangère qu'un article de sport, un éditorial qu'un article de fond). La notion de corpus de référence est de plus en plus mise en cause : on ne sait pas bien définir ce qu'est une référence dans ce domaine. Il existe, au mieux, des corpus variés.

<sup>7</sup> La prise en compte des genres amène à constituer des corpus de taille moyenne permettant un moins bon apprentissage qu'un corpus de plus grande taille. L'apport de la notion de genre est alors gommé par l'insuffisance des données d'apprentissage.

<sup>8</sup> Sur tous ces points, la terminologie peut varier d'une étude à l'autre

identifiants mais le simple fait de leur donner des noms interprétables montre le lien entre les deux niveaux. Plus généralement, les relations analysables par analyse linguistique rendent partiellement compte de l'organisation conceptuelle d'un domaine.

#### **4.2 Du modèle formel à la langue**

La linguistique textuelle pose le problème de l'acquisition de connaissances à partir de corpus et de leur utilisation possible dans d'autres cadres applicatifs. Le constat est souvent fait que les ressources ainsi acquises sont difficilement réutilisables.

La modélisation d'un domaine donné, même d'un domaine *a priori* bien structuré comme la médecine, est une activité difficile, partielle et parfois partiale, parce qu'elle doit rendre compte de points de vue et de contraintes applicatives souvent divergents. Il n'en demeure pas moins, au-delà des petits décalages possibles dans la compréhension ou l'interprétation d'un individu à l'autre, que nous parvenons généralement à nous faire comprendre et à communiquer. De manière plus concrète, même sur un texte technique dont le contenu m'échappe, je peux le plus souvent dire des choses : que tel élément semble être une entité nommée, qu'elle est en relation avec tel autre élément, etc. Nous pouvons faire ce genre d'assertions parce que nous avons une connaissance générale, commune de la langue<sup>9</sup>.

Ceci n'implique pas l'existence d'une langue ou d'un corpus de référence au sens où on l'entend traditionnellement. Au contraire, ce niveau de compréhension, même limité, nous semble montrer l'existence d'une connaissance commune qui ressortit à la langue elle-même, les textes supportant quant à eux un certain niveau de variation et de spécialisation qui peut sans doute se mesurer de manière relative (ou différentielle, pour reprendre la terminologie employée par F. Rastier). Cela ne signifie pas qu'il y ait une langue générale par opposition à une langue de spécialité, mais cela signifie qu'il existe un certain état de langue, matérialisé par des textes, par opposition à un autre état de langue matérialisé par d'autres textes. Il s'agit donc d'un état relatif, il ne s'agit pas d'un absolu (la langue générale) par rapport auquel on mesurerait un degré d'éloignement (la langue de spécialité). Il existe un continuum qui peut se mesurer par l'analyse des textes.

### **5 Dynamique du sens**

Nous avons examiné dans la section précédente la notion de modèle, qui constitue en quelque sorte une image du monde physique. Nous revenons dans cette partie sur les liens entre les entités représentatives d'un domaine (termes, entités nommées) et le monde physique. Nous affirmons que le terme n'est pas référentiel en discours et non essayons de montrer que le cas des entités nommées n'est pas très différent.

#### **5.1 Référence et dénotation des entités en discours**

Comme nous l'avons vu en introduction, la terminologie classique, telle qu'elle est théorisée par Wüster [1981], est essentiellement un jeu d'étiquettes posées sur des entités extérieures. Chaque terme rend compte de manière non ambiguë d'un objet du monde ; c'est d'ailleurs un des points fondamentaux de la distinction entre lexique et terminologie. La théorie terminologique permet de modéliser un modèle à l'image du monde, car les termes sont censés référer directement à une réalité extérieure. Cette théorie a été depuis remise en cause à de multiples reprises. L'étude des termes en corpus a montré que ceux-ci se comportent pour une large part comme les autres syntagmes nominaux. Ils subissent des

---

<sup>9</sup> Nous nous inspirons ici d'un propos de B. Victorri.

variations linguistiques, possèdent des synonymes, intègrent un point de vue<sup>10</sup> [Condamines et Rebeyrolle, 1997].

Comme le rappelle Charolles [2002, p. 7-10], dans la littérature linguistique, les auteurs distinguent la dénotation de la référence. La dénotation « prend acte du fait que le signifié (des signes linguistiques) exprime un concept muni d'une certaine intension (i.e. un ensemble d'attributs caractéristiques en propre) et que cette intension délimite virtuellement son extension (c'est-à-dire une classe d'êtres satisfaisants ces attributs) ». L'acte de référence est quant à lui « un accord non entre deux pensées (...) mais entre deux pensées à propos de quelque chose et cela par le biais de la production, en contexte, d'une expression référentielle ». Nous reprenons à notre compte l'essentiel de ces définitions mais à l'inverse de Charolles, nous soutenons que l'on peut alors parler d'opération linguistique pour la dénotation, contrairement à la référence qui ressortit à une logique visant à apparier niveau linguistique et niveau extra-linguistique.

En discours, l'aspect dénotationnel est primordial, passant bien avant l'aspect référentiel. Cette distinction n'est pas seulement d'ordre philosophique : il s'agit d'une dimension essentielle de l'interprétation du texte dans la mesure où la dénotation, à l'inverse de la référence, peut supporter une sous-spécification par défaut<sup>11</sup>. Le texte n'a en effet pas besoin d'être ancré sur un environnement où chaque élément serait tangible ou « atteignable », comme muni d'une étiquette. Il est possible de tenir des discours abstraits, généraux, voire contraires à la réalité sans que cela nuise à la compréhension, tant que l'univers décrit est logique et cohérent. Le signe linguistique ouvre par ailleurs des « possibles », il possède une « vie sémasiologique » qui lui est propre, pour reprendre le terme de Depecker [2000]. Ce dernier souligne aussi la richesse du signifié en contexte, où des dimensions symboliques et sociales entrent en jeu<sup>12</sup>.

Cette variation du signe linguistique s'accorde mal avec l'analyse de la référence. On peut reprendre ici l'exemple bien connu de Frege, sur le fait que Vénus peut être désigné comme *l'étoile du soir* ou comme *l'étoile du matin*. Le fait que les deux dénominations désignent le même astre solaire n'est pas fondamental au niveau linguistique ; il est en revanche essentiel pour assigner une valeur de vérité aux assertions émises sur ces objets<sup>13</sup>. On rejoint ici le

---

<sup>10</sup> Peut-être faut-il parler ici de double point de vue : point de vue de celui qui choisit le terme, point de vue de celui qui l'interprète. On retrouve partiellement la proposition faite par Anne Condamines [2003] pour distinguer genre textuel et genre interprétatif. Nous développons dans la suite de cette article une vision mettant en avant la dimension interprétative qui est par nature présente lors de l'accès à la connaissances par les textes (nous nous plaçons là dans la lignée des travaux de Rastier, qui a bien mis en avant les aspects herméneutique de l'accès au texte, même technique).

<sup>11</sup> On retrouve ici la notion de profondeur variable de Kayser et Coulon [1981], que nous interprétons ainsi : l'analyse n'a pas à se préoccuper d'un certain nombre de traits sémantiques latents tant que ceux-ci ne sont pas activés de manière explicite.

<sup>12</sup> L'auteur souligne par exemple la différence entre un mot comme *eau* et son équivalent chimique H<sub>2</sub>O. Le mot *eau* prend son sens en contexte et se place au sein d'un réseau de notions, il n'en va pas de même de son équivalent chimique, au pouvoir évocateur bien moindre.

<sup>13</sup> Les cas où l'analyse de la référence est obligatoire sont somme toute relativement rares pour les applications de traitement automatique des langues. Le cadre privilégié est la commande de système, la communication avec des robots ou tout autre cadre où la parole entraîne une action très directe sur le monde. Les applications mettant en œuvre une analyse terminologique ont une visée opérationnelle que l'on ne peut nier, souvent sur des applications critiques. Quand EDF met au point une terminologie du domaine nucléaire, il s'agit bien évidemment de désigner des objets très concrets dans le cadre d'opérations parfois délicates. Mais, même dans ce cadre, la plupart des éléments sont décrits dans un cadre dénotationnel. Les éléments clés de l'opération, ceux pour lesquels l'analyse de la référence est cruciale, sont le plus souvent désignés par un identifiant unique

point de vue de Searle [1983] pour qui « l'acte de référence, le processus de référenciation », inclut une dimension externe à la langue « à savoir que l'entité visé par cet acte, la chose (au sens le plus général du terme) à laquelle il renvoie, existe au-delà de ce que le locuteur en dit, en dehors de son esprit et de celui de celles ou ceux à qui il s'adresse ». [Charolles, 2002, p. 38]

## 5.2 *Du Sens à la signification du terme*

A de rares exceptions près, nous sommes restés dans le cadre de cette étude à une réflexion touchant à l'analyse terminologique. Nous avons vu une évolution très importante entre les travaux de Wüster et ceux inspirés par une approche textuelle.

Un des éléments que nous avons relativement peu mis en cause est le statut privilégié du groupe nominal, en tant que support de la connaissance. Le verbe est parfois pris en compte pour la modélisation du domaine, notamment pour le repérage de liens entre termes. Il n'empêche, les modèles terminologiques sont essentiellement statiques, il s'agit essentiellement de taxinomies, c'est-à-dire de hiérarchies fondées sur le lien *est-un*, éventuellement assorties d'autres relations (cf. section 2.1. de cet article). Or, le texte offre une description dynamique du monde, il met en relation des entités et les fait évoluer au cours du temps (au cours de la narration). Du coup, il n'est pas étonnant de constater une dépendance étroite entre le résultat de la modélisation et le texte qui décrit des objets du monde en mouvement, dans un cadre dynamique.

Cette situation n'est pas propre à la terminologie. Tout texte, toute mise en relation d'items lexicaux offre ce cadre mouvant et permet de construire des descriptions de manière dynamique. On est là au cœur de la problématique de la sémantique lexicale : comment représenter de manière statique le sens d'une entité qui s'esquisse progressivement et évolue en fonction du contexte ? Il s'agit de la distinction bien connue entre sens (aspect statique) et signification (aspect dynamique) [Rastier, 1999]. Diverses solutions ont été proposées au niveau lexical : noyau de sens [Langacker, 1990], théorie du prototype [Kleiber, 1990], décomposition des mots en primitives sémantiques [Winograd, 1978] ou en sèmes [Greimas, 1966], processus génératif de construction du sens [Pustejovsky, 1991], etc. On retrouve des tentatives proches en terminologie, notamment pour limiter l'étendue des ontologies. Bachimont [2000] a ainsi développé la notion d'ontologie locale, pour limiter la portée de la modélisation fournie et régler de manière partielle le problème du glissement de sens. Peu de formalismes offrent cependant une vision réellement dynamique du terme au niveau linguistique. La dimension interprétative [Condamines, 2003] répond à une partie de la question, mais ne résout pas complètement le cas des relations qui évoluent entre termes, voire des contradictions qui peuvent apparaître au sein d'un texte si on en abstrait la dimension temporelle.

Sans revenir sur les problèmes pratiques de réutilisabilité des ressources terminologiques, on voit bien ici les enjeux fondamentaux du domaine. La linguistique textuelle opère sur des objets essentiellement dynamiques (les textes) pour élaborer des représentations par nature statique (les taxinomies). Le travail de modélisation replace les concepts dans un cadre dynamique, en offrant des modes de représentation appropriés (les actinomies, pour reprendre la terminologie employée par KOD [Vogel, 1988]). Cependant, les liens entre ce modèle formel opératoire et les modèles de l'analyse linguistique restent sans doute en grande partie à étudier.

---

(souvent une chaîne de caractères qui n'est pas à proprement parler linguistique) et des éléments non textuels (schémas, photos, ...).

## 6 Conclusion et perspectives

Cet article a permis de revenir sur plusieurs notions fondatrices de l'analyse terminologique, à savoir la notion de corpus, de domaine, de langue de spécialité. Nous avons essayé de montrer que l'analyse terminologique n'est pas très différente de l'analyse linguistique classique. Nous sommes également revenu sur la problématique classique de la distinction entre aspects statiques et aspects dynamiques de la représentation du sens. Cette distinction est visible dans les modèles formels instanciés à partir des textes mais nous avons essayé de montrer la question du sens reste en partie posée au niveau de l'analyse linguistique sous-jacente effectuée à partir de corpus.

Nous sommes là en quelque sorte au milieu du gué : une modélisation linguistique est-elle possible, sans être fondée sur un ordre ontologique pré-établi ? Cadiot et Visetti [2001] ont proposé des pistes pour dessiner « l'organisation et l'activité sémantiques » fondé sur des mécanismes d'instabilité et de stabilisation<sup>14</sup>. La mise en œuvre informatique d'un tel système est délicate mais la réflexion proposée rejoint nos préoccupations, visant à rendre le modèle sous-jacent davantage dynamique. Une modélisation de ce type permet de mieux prendre en compte la démarche interprétative à l'œuvre lors de l'accès au texte.

Ceci n'est pas entièrement nouveau : les techniques d'apprentissage essaient de rendre partiellement compte de ce caractère dynamique de l'analyse, comme l'ont montré plusieurs auteurs, y compris dans le cadre de l'analyse terminologique (voir par exemple [Bourigault, 1994]). Mais l'apprentissage seul n'est généralement pas suffisant car les corpus sont forcément limités : il est souvent nécessaire de recourir, en pratique, à des ressources existantes, y compris de type ontologique. Le continuum entre langue de spécialité et langue générale permet d'espérer que cette confrontation entre différents types de ressources ne soit pas vaine mais contribue à faire émerger des stratégies d'analyse originales quant au sens des unités en contexte.

## Références

- ATKINS S., CLEAR J. et OSTLER N. « Corpus design criteria ». *Literary and linguistic computing*. Vol 8(4). 1992. pp. 1-16.
- BACHIMONT B. Herméneutique matérielle et Artéfacture : des machines qui pensent aux machines qui donnent à penser ; Critique du formalisme en intelligence artificielle. Thèse de l'Ecole Polytechnique. 1996.
- BESSE B. « Le domaine ». In *Le sens en terminologie* (éd. H. Béjoint et P. Thoiron). Presses Universitaires de Lyon. 2000.
- BIBER D. *Variation across speech and writing*. Cambridge University Press. Cambridge. 1988.
- BIÉBOW B. et SZULMAN S. « TERMINAE : A linguistic-based tool for the building of a domain ontology ». *Actes 11<sup>th</sup> European Workshop, Knowledge Acquisition, Modeling and Management (EKAW 99)*. Dagstuhl Castle. Allemagne. 1999. p. 49-66
- BOCCON-GIBOD H., PICARD M.-L., BOURIGAULT D. et GARCIA D. « Enjeux de la maîtrise de la terminologie pour la production et la consultation de documents ». *Actes Conférence sur la coopération dans le domaine de la terminologie en Europe*. Paris. 1999.
- BOURIGAULT D. *Lexter, un logiciel d'extraction de terminologie. Application à l'acquisition des connaissances à partir de textes*. Thèse de l'Ecole des hautes Etudes en Sciences Sociales, Paris. 1994.

---

<sup>14</sup> Bien d'autres auteurs ont proposé des cadres allant dans ce sens. Pour ne prendre qu'un exemple : Siblot [1989] propose des mécanismes de réglage du sens d'après une connaissance partagée sur le monde. Le cadre est différent mais le but est le même, le contexte intégrant ici des éléments extra-linguistiques.

- BOUVERET M. « Approche de la dénomination en langue spécialisée ». In *Meta*. vol. 43, n° 3. 1998. pp. 393-410
- CABRÉ M.T., ESTOPÀ R., VIVALDI J. « Automatic Term Detection: a review of current systems ». In *Recent Advances in Computational Terminology*. John Benjamins. 2000.
- CADIOT P. et VISETTI Y.-M. *Pour une théorie des formes sémantiques*. Presses Universitaires de France. Paris. 2001.
- CARTIER E., DESCLES J.-P., JACKIEWICZ A. MINEL J.-L. « Filtrage automatique de textes : l'exemple des énoncés définitoires ». *Actes 1<sup>ères</sup> Journées FRANCIL*. Avignon. 1997.
- CHAROLLES M. *La référence et les expressions référentielles en français*. Ophrys. Paris. 2002.
- CLEMENT L. *Construction et exploitation d'un corpus syntaxiquement annoté pour le français*. Thèse de l'université Paris 7. Paris. 2001.
- COLLIER N., NOBATA C. et TSUJII J. "Comparison between Tagged Corpora for the Named Entity Task". In *Actes 18th International Conference on Computational Linguistics (COLING 2000)*. Saarbrücken, Allemagne. 2000. pp. 201-207.
- CONDAMINES A. *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*. Mémoire d'Habilitation à Diriger des Recherches de l'Université Toulouse 3. Toulouse. 2003.
- CONDAMINES A. et Rebeyrolle J. « Point de vue en langue spécialisée ». In *Meta*, Vol. 17(1). 1997. pp 174-184.
- DEPECKER L. « Le signe entre signifié et concept ». In *Le sens en terminologie* (éd. H. Béjoint et P. Thoiron). Presses Universitaires de Lyon. 2000.
- CULIOLI A. *Pour une linguistique de l'énonciation*. Ophrys. Paris. 2002.
- DACHELET R. *Sur la notion de sous-langage*. Thèse de l'Université Paris 8. Saint-Denis. 1994.
- DAILLE B. « Study and implementation of combined techniques for automatic extraction of terminology ». In *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. MIT Press. Cambridge. 1994. pp. 49-66.
- ENGUEHARD, C. « Acquisition de terminologie à partir de gros corpus ». *Actes Informatique & Langue Naturelle*, ILN'93, Nantes. 1993. pp.373-384.
- FAUCONNIER G. *Espaces mentaux*. Péditions de Minuit. Paris. 1984.
- FOLCH H., HEIDEN S., HABERT B., FLEURY S., ILLOUZ G., LAFON P., NIOCHE J. et PREVOST S. « TyPTex: Inductive typological text classification analysis for NLP systems tuning/evaluation ». In: *Second International Conference on Language Resources and Evaluation*. Athènes. Grèce. 2000, pp. 141-148.
- FREGE, G. « Über Sinn und Bedeutung ». In *Zeitschrift für Philosophie und philosophische Kritik*, 100. Trad. fr. « Sens et dénotation ». In *Ecrits logiques et philosophiques*. Seuil. Paris. 1971. pp. 102-126.
- GRABAR N. et HAMON T. « Les relations dans les terminologies structurées : de la théorie à la pratique ». In *Revue d'Intelligence Artificielle*. Vol. 18 n° 1. 2004. pp. 57-85.
- GREIMAS A.J. *Sémantique Structurale*. Larousse, Paris, 1966.
- GROSS G. « Degré de figement des noms composés ». *Langages*. n° 90. 1988. pp. 57-72.
- HABERT B., NAZARENKO A. et SALEM A. *Les linguistiques de corpus*. Armand Colin. Paris. 1997.
- HABERT B. et JACQUEMIN C. « Noms composés, termes, dénominations complexes: problématiques linguistiques et traitements automatiques ». *TAL*, vol. 34 n° 2. 1993. pp. 5-42.
- HARRIS Z. *A theory of language and information. A mathematical approach*. Oxford University Press. Oxford. 1991.
- Hofstadter D. *Gödel, Escher, Bach ; les Brins d'une Guirlande Eternelle*. InterÉditions. Paris. 1985.
- ILLOUZ G. « Vers un apprentissage en TALN dépendant du type de texte. » *Actes TALN 2000*. Lausanne. Suisse. 2000.

- JACKIEWICZ A. et MINEL J.-L. « L'identification des structures discursives engendrées par les cadres organisationnels ». *Actes TALN 2003*. Batz sur Mer. 2003. pp 155-164.
- JACQUEMIN, C. *Transformations des noms composés*. Thèse de l'Université Paris 7, Paris. 1991.
- JACQUEMIN, C. *Spotting and Discovering Terms through NLP*. MIT Press, Cambridge MA. 2001.
- JOHNSON-LAIRD P.N. « Mental models and deduction ». *Trends in Cognitive Science*. n° 5. 2001. pp. 434-442.
- KAYSER D. et COULON D. « Variable-Depth Natural Language Understanding ». *7<sup>th</sup> International Joint Conference on Artificial Intelligence*. 1981. Vancouver. pp.64-66.
- KLEIBER G. *La sémantique du prototype, Catégories et sens lexical*. PUF. Paris. 1990.
- KRIPKE S. *La logique des noms propres*. Éditions de Minuit. Paris. 1982.
- LANGACKER R. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Mouton de Gruyter. Berlin & New York. 1990.
- LEROY S. « Antonomase, métaphore et nom propre : identification ou catégorisation ? ». In *Travaux linguistiques du CerLiCO*. Vol 16 (*Morphosyntaxe du lexique – 2 : catégorisation et mise en discours*). Presses Universitaires de Rennes. Rennes. 2003. pp. 161-178
- NAZARENKO A. *Donner accès au contenu des documents textuels. Acquisition de connaissances et analyse de corpus spécialisés*. Mémoire d'Habilitation à Diriger des Recherches de l'Université Paris 13. Villetaneuse. 2004.
- NOY N. F., SINTEK M., DECKER S., CRUBEZY M., FERGERSON R.W. et MUSEN M.A. „Creating Semantic Web Contents with Protege-2000”. *IEEE Intelligent Systems*. Vol.16(2). 2001. pp. 60-71.
- PUSTEJOVSKY J. « The Generative Lexicon ». In *Computational Linguistics*. Vol 17(4). 1991. pp. 409-441.
- RASTIER F., CAVAZZA M. et ABEILLE A. *Sémantique pour l'analyse*. Masson. Paris. 1994.
- RASTIER F. « De la signification au sens : pour une sémiotique sans ontologie ». Paru en italien « Dalla significazione al senso : per una semiotica senza ontologia ». In *Eloquio del senso, a cura di Pierluigi Basso e Lucia Corrain, Costa & Nolan*. Milan. 1999. pp. 213-240 (disponible en français [www.revue-texto.net/Inedits/Rastier\\_Semiotique-ontologie.html](http://www.revue-texto.net/Inedits/Rastier_Semiotique-ontologie.html)).
- REBEYROLLE J. et TANGUY L. « Repérage automatique de structures linguistiques en corpus : le cas des énoncés définitoires ». In *Cahiers de grammaire*. N° 25. Université de Toulouse. 2000. pp. 79-88.
- SAGER J.C. *A practical course in terminology processing*. J. Benjamins. Amsterdam. 1990.
- SEARLE J. *L'intentionnalité. Essai de philosophie de l'esprit*. Éditions de Minuit. Paris. 1983.
- SEGUELA P. *Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques*. Thèse de l'Université Toulouse 3. Toulouse. 2001.
- SIBLOT P. « Isotopie et réglage du sens ». In *Cahiers de praxématique*, n° 12. 1989. pp. 91-110.
- SINCLAIR J. *Preliminary recommendations on Corpus Typology*. Technical report, EAGLES (Expert Advisory Group on Language Engineering Standards). 1996.
- SLODZIAN M. et BOURIGAULT D. « Pour une terminologie textuelle ». *Actes de la conférence TIA'1999*. Nantes. France. 1999.
- SOWA, J. F. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley. Londres. 1984.
- VOGEL C. *Génie cognitif*. Masson. Paris. 1988.
- WINOGRAD T. « On primitives, prototypes, and other semantic anomalies ». *Proceedings of the theoretical issues in natural language processing*. Urbana-Campaign. 1978. pp. 25 – 32.
- WÜSTER E. « L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses ». In *Textes choisis de terminologie I, Fondements théoriques de la terminologie*. Presses de l'université de Laval. 1981. pp. 55-114.