

PRÉSENTATION

Aude GREZKA

Françoise MARTIN-BERTHET

Lexiques Dictionnaires Informatique (LDI)
CNRS - Université Paris 13, UMR 7187

INTRODUCTION

L'organisation du lexique en sous-ensembles sémantiques est une pratique ancienne, destinée à donner un accès naturel et lisible à la nomenclature. On cherche aujourd'hui à construire des classes sémantiques appuyées sur des critères linguistiques, de manière inductive et interne à chaque langue.

Les entrées sont des emplois. Les dégroupements sont un préalable ; on constate cependant, dans la pratique de la recherche, que les regroupements peuvent agir en retour sur les dégroupements, notamment en les affinant : les deux démarches sont interdépendantes et se soutiennent l'une l'autre.

Concernant les verbes, plusieurs travaux récents s'attachent à établir des classes sémantiques et syntaxiques ; notamment, pour l'anglais, Levin, 1993 ; pour le français, Dubois et Dubois-Charlier, 1997 ; François, Le Pesant & Leeman (éds), 2007. Une partie des contributions présentées ici appliquent une grille de description en cours d'élaboration au LDI dans le cadre d'un groupe de travail dirigé par Michel Mathieu-Colas. Cette grille consiste en un faisceau de propriétés conçu pour atteindre un niveau de regroupement linguistiquement pertinent¹. Nous en indiquons les principaux aspects, avant de présenter les articles de ce numéro.

1 Cf. Grezka, 2006.

GRILLE DE DESCRIPTION : PRINCIPAUX PARAMÈTRES

Nom de la classe

Les classes générales sont désignées par un nom (grammatical) prédictif : “changement de couleur”, “cognition”, “coups”, “opérations numériques”, etc. On ne dispose généralement pas de verbe hyperonyme simple susceptible de servir de dénomination. Les verbes les plus généraux (comme *penser* ou *savoir* pour la “cognition”, *parler* pour la “parole”, *compter* ou *calculer* pour les “opérations numériques”) sont eux-mêmes engagés dans le jeu des différences sémantiques et syntaxiques qui structurent l’ensemble et ne les dominent pas ; ou encore, le verbe le plus général non marqué, effectivement substituable aux autres, et qui peut leur servir de genre prochain (par ex. *frapper* pour les verbes de “coups”) a d’autres emplois classés ailleurs : on risquerait de confondre la désignation de la forme (le verbe *frapper*), de l’emploi (*frapper* = coups physiques entre humains), de la classe (“frapper”). On peut trouver, notamment pour les classes plus petites, un déverbal morphologique (quel que soit le sens de la dérivation) correspondant à l’un des verbes de la classe et capable de la coiffer : Verbes d’“opération numérique#2 : addition” : *additionner*, *ajouter*, *sommer*, *totaliser* ; ou un infinitif se nominalise, comme pour le “savoir”, dont les six sous-classes comportent ou non le verbe *savoir*. Dans d’autres cas, c’est un supplétif qui s’impose : “coups”, à défaut du déverbal *frappe* dont les emplois sont restreints. On peut jouer sur la richesse de la morphologie nominale déverbale pour proposer un dérivé non usuel (mais attesté) comme “ayance”, capable de rassembler : “possession”, “acquisition#1 : obtention”, “acquisition#2 : approvisionnement”, “acquisition#3 : appropriation”, “acquisition#4 : capture”, “dépossession d’autrui”, “dépossession volontaire”, “manque”. Ces noms de catégories sémantiques sont destinés à être indépendants des classes grammaticales.

On peut en revanche alléger la métalangue en utilisant des verbes prototypes pour symboliser des classes terminales : par exemple “coups#1 : type **frapper**” ; “apprentissage#1 : type **étudier**”. L’expression du sens de la classe est alors renvoyée à la caractérisation sémantique.

Caractérisation sémantique

La caractérisation sémantique a la forme d’une expression verbale ; par exemple, “addition” : “ajouter plusieurs valeurs numériques” ; “apprentissage#1 : type **étudier**” : “acquérir par un effort conscient un ensemble de connaissances appartenant à une discipline” ; “coups#2 : type **poignarder**” : “donner un ou plusieurs coups avec un instrument spécifique”. Cela correspond à une phrase, dans la mesure où le sujet est spécifié (humain dans ces exemples). Elle est formulée intuitivement de manière à exprimer autant que possible la partie commune du sens des entrées, c’est-à-dire le sens de la classe. Il ne s’agit pas d’une définition, mais d’un guide pour la recherche et la consultation.

Liste de verbes

On rassemble dans la liste des verbes de chaque classe les unités simples et les unités polylexicales (Gross, 1996 ; Mejri, 1997), par exemple

verbes de “délire” : *dérailler, déraisonner, battre la campagne, perdre les pédales...* ; verbes de “coups” : *dérouiller, tabasser, casser la figure, passer à tabac...* On sait comme la reconnaissance théorique de ces unités composées reste le plus souvent lettre morte dans les applications et les traitements lexicographiques. On ne méconnaît pas les problèmes posés par la délimitation de l’unité lexicale en général et par l’identification des différents types d’expressions verbales complexes en particulier. Mais on avancera en accumulant les données ; l’affinement de la typologie est l’un des chantiers transversaux en cours. Par sécurité, les expressions pourront aussi être indiquées sous leur verbe simple (comme il est traditionnel en lexicographie, quand elles sont traitées) : *mettre la main sur* (un bien), *mettre* Poss *main sur la figure de* (quelqu’un), sont des constructions de *mettre* ou des expressions à base *mettre*, et des verbes respectivement d’“ayance” et de “coups”, substitués de *s’emparer de, cogner sur*². La numérisation autorise l’excès et la redondance, défauts plus facilement corrigibles que le manque.

Propriétés structurelles

Elles sont représentées sous la forme de schémas d’arguments. Les verbes d’une classe peuvent entrer dans plusieurs schémas, avec des répartitions variables.

Les verbes de “coups#1” (verbes généraux : *cogner, frapper, taper*) entrent tous dans les deux schémas suivants :

N0<hum> V N1<hum> [Prép N2<npc>]

Aude frappe Marc [à l’épaule]

N0<hum> V N1<npc> de N2<hum>

Aude frappe l’épaule de Marc

La question se pose dans ce cas de la réduction à une construction de base, dont l’autre serait une reconstruction.

Les verbes de “transfert d’objet” entrent dans deux schémas à trois arguments, distingués par l’ordre des objets, avec une distribution complémentaire :

N0<hum> V N1<objet> à N2<hum>

Verbes : *céder, donner, fournir, livrer, offrir, procurer ...*

N0<hum> V N1<hum> de N2<objet>

Verbes : *doter, gratifier, munir, pourvoir ...*

Les verbes d’“opération numérique#3 : soustraction” entrent dans deux schémas à trois arguments distingués par la préposition introduisant N2, avec une intersection (il y a des verbes communs aux deux constructions) :

² Le registre familier est particulièrement riche en expressions de ce type. Elles présentent souvent une variation interne qu’aucune norme ne vient fixer : par exemple *casser la figure, gueule, tronche...* On devra établir des conventions de lemmatisation (cf. Petit, 2003).

N0<hum> V N1<val.numériques> [de N2<val.numériques>]

Verbes : *décompter, déduire, défalquer, ôter, retrancher, retirer, Soustraire*

N0<hum> V N1<val.numériques> [sur N2<val.numériques>]

Verbes : *défalquer, précompter, prélever, retenir, retrancher ...*,

Les arguments sont spécifiés par leur “classe d’objets” (Gross, 1994 ; Le Pesant & Mathieu-Colas, 1998), comme indiqué dans l’exemple ci-dessus. La distinction des emplois opérée par les classes d’arguments retentit évidemment sur les classes de prédicats. Par exemple, *apprendre* (+ un événement) se range avec *connaître, savoir*, dans une classe de verbes de “savoir” caractérisée par cet objet ; *apprendre* (+ une discipline) se range avec *étudier, potasser, se mettre à*, etc., dans une classe de verbes d’“apprentissage”. Les classes de lieux ou d’instruments sont également prises en compte selon leur relation à la constitution de la classe et à l’emploi de chacune de ses unités : *apprendre* (+ une discipline) *à l’école, dans les livres* ; *apprendre* (+ un événement) *par la presse*. Les lieux ou instruments spécifiques apparaissent aussi dans la rubrique des **Adverbiaux appropriés**, où figurent encore les intensifs de la classe ou d’une partie de la classe (type **frapper** : *à bras raccourcis, sauvagement*).

Reconstructions

On indique les principales reconstructions des schémas de base (passif, constructions pronominales, constructions causatives, pronominalisation des arguments), notamment lorsqu’elles sont caractéristiques. Par exemple, l’objet de *tomber sur* (quelqu’un) “agresser” se pronominalise en *lui / leur – dessus*, comme *taper sur, cogner sur* ; celui de *tomber sur* (quelqu’un) “rencontrer” se pronominalise en *sur lui / elle / eux* :

Aude est tombée sur Marc [à coups de parapluie / à bras raccourcis]

> *Aude lui est tombée dessus*

Aude est tombée sur Marc [à la cafeteria]

> *Aude est tombée sur lui*

La périphrase passive en *se faire* est typique d’un certain nombre de classes dont la typologie reste à faire (*se faire agresser, se faire coiffer*).

Une difficulté se présente lorsque l’orientation entre deux constructions liées n’a pas d’évidence ; par exemple entre constructions transitives et intransitives pour les verbes de “changement de couleur” : *blanchir* tr. et intr. On notera le chassé-croisé des verbes *blanchir* et *rougir* dans le *Petit Robert* : le premier est d’abord décrit comme transitif, le second comme intransitif. Il peut donc y avoir hésitation entre deux orientations : du transitif à l’intransitif “neutre” ; de l’intransitif au transitif “causatif”. De plus, il y a des intransitifs isolés : par ex. *grisonner*. Il semble donc préférable, dans tous les cas semblables, de traiter les constructions transitives et intransitives dans des classes autonomes. Simplement, pour ne pas perdre le lien entre constructions associées, on opérera des renvois d’une classe à l’autre sous la rubrique “Reconstructions”³.

3 Ce point est développé dans Grezka, à par. en 2008 ; Mathieu-Colas, à par. en 2008.

De même, les constructions pronominales réciproques de verbes de “coups” (*se battre*, *se taper dessus*) seront traitées comme reconstructions des verbes concernés, mais aussi comme entrées dans une classe de verbes signifiant l’“échange de coups”, où ils sont rassemblés avec des verbes non pronominaux (*en découdre*) ou “essentiellement pronominaux” (*se colleter*, *se frriter*).

Propriétés morphologiques

Cette rubrique doit préparer la correspondance avec les classes de noms et d’adjectifs.

On rassemble ici les noms et adjectifs prédicatifs morphologiquement et sémantiquement associés. Un déverbal morphologique peut ne pas figurer dans la classe du verbe correspondant (*frappe* est séparé de *frapper* “coups physiques humains”). Inversement, on trouvera *rupture* (*de stock*) dans une classe “manque”, où ne figure pas le verbe *rompre*. Les supplétismes du type *frapper* / *coup*, *tomber* / *chute*, dont le principe est reconnu depuis longtemps, demandent un traitement systématique. On recueillera donc les noms prédicatifs autonomes : par exemple, avec *affrontement(s)*, *bagarre(s)*, *heurts* : *échauffourée(s)*, *émeute(s)*, *violences urbaines* ; avec *maltraitance* : *séances*. Les formes savantes de dérivation, à partir d’un certain degré d’éloignement de la forme populaire, tendent d’ailleurs à perdre la relation morphologique au profit de la relation sémantique et syntaxique, qui devient le critère prédominant⁴.

Les participes passés adjectivés sont mentionnés dans cette rubrique ; certains d’entre eux sont également indiqués dans la rubrique précédente par la construction résultative.

Classes voisines

La classification terminale, étant linguistique et non ontologique, se dégage du lexique par induction. Elle est en constant réexamen jusqu’à couverture aussi exhaustive que possible de la nomenclature. Pour cette raison, et pour guider la consultation de proche en proche, il est utile de prévoir que la fiche signalétique d’une classe comporte systématiquement l’indication des classes sémantiquement voisines ; par ex. “soustraction” : “diminution”, “prélèvement”.

ARTICLES

Les huit premiers articles résultent du travail mené au LDI⁵.

Les contributions de Michel Mathieu-Colas et Aude Grezka abordent des questions générales. Le premier, après un rappel théorique et terminolo-

4 “Personne ne peut déduire *rupture* de *rompre* en français, si ce n’est par l’usage qui en est fait ; la différence formelle est presque aussi grande qu’entre *chute* et *tomber*” (Rey-Debove, 2004).

5 Nous remercions John Humbley pour les résumés en anglais qu’il a écrits ou corrigés.

gique concernant les notions de *domaine* et de *classe*, montre l'intersection entre les deux ensembles et la nécessité de cette double catégorisation notamment pour le traitement automatique des langues. La seconde, après une synthèse historique, distingue *type de procès* (action, état, événement) et *aspect*, puis *aspects inhérents* (véhiculés par le sens lexical) et *aspects contextuels* (véhiculés par le contexte) ; l'auteur montre les compatibilités entre ces trois dimensions, et développe plus particulièrement le cas du verbe *sentir*.

Les quatre articles suivants présentent des classes de verbes et les questions soulevées par leur description. Robert Vivès a classé les verbes de "cognition" ; il souligne les difficultés propres à cet ensemble (grande polysémie, différences sémantiques fines), le rôle crucial des classes d'arguments et des adverbiaux appropriés, et détaille le cas du verbe *penser*. Monia Bouali a classé les verbes qui marquent le changement d'une couleur à une autre, selon les constructions (intransitives et transitives) et les classes d'arguments concernées (objets, éléments naturels, humains) ; elle montre les correspondances entre verbes et parties du corps ainsi que les métonymies autorisées (*les cheveux de Marc blanchissent* > *Marc blanchit*), qu'elle présente dans un tableau synoptique. Itab Ben-Henia, à partir de son travail sur les classes de verbes d'"ayance" (possession, dépossession, acquisition, manque), insiste sur l'intégration nécessaire des *séquences verbales figées* ; avec deux exemples (*faire main basse sur*, *mettre la main sur*), elle soulève le problème de leur délimitation, en l'occurrence du statut de la préposition, intégrée au verbe ou construction du verbe. Le figement et la métonymie sont également abordés dans l'article de Françoise Martin-Berthet, avec le cas des verbes de "coups" et de leurs relations syntaxique, sémantique et morphologique aux noms de partie du corps : liens entre nom d'humain et nom de partie du corps dans les constructions, verbes dérivés de noms de partie du corps, expressions verbales à partie du corps intégrée.

Les deux dernières contributions émanant du LDI ouvrent des chantiers transversaux. L'article de Jean-François Sablayrolles et Soundous Ben Hariz Ouenniche concerne celui de la veille néologique, qui doit alimenter en permanence les entrées et faire apparaître les causes et les matrices de la néologie verbale, matrices non seulement morphologiques mais aussi sémantico-syntaxiques, avec les changements de classe grammaticale (conversion) et de schéma d'arguments (transitivité, classes d'objets). Le travail d'Emmanuel Cartier porte sur l'informatisation de la description linguistique avec un analyseur développé par lui-même, TextBox, où toutes les informations linguistiques sont librement définies dans des fichiers distincts, ce qui permet l'implémentation de descriptions fines ; il l'applique ici au repérage automatique des expressions verbales de "cognition", à partir des données de Robert Vivès.

Les trois articles suivants ouvrent des perspectives sur d'autres langues et sur des approches voisines et complémentaires.

Xavier Blanco (Université Autonome de Barcelone, Laboratoire fLex-Sem - SGR 01029) s'inspire du modèle d'Igor Mel'čuk et Alain Polguère (2007), pour présenter une *hiérarchie d'étiquettes sémantiques* en cours

d'élaboration pour les verbes de l'espagnol ; les "étiquettes" sont des expressions verbales capables d'être le genre prochain de la définition des entrées ; les entrées sont des emplois décrits par une *formule actancielle*, qui précise les *actants sémantiques* liés au verbe, c'est-à-dire les participants à la situation à laquelle il réfère⁶.

Les deux derniers articles élargissent le champ en direction des expressions prédicatives à support verbal et nominal. Michele Prandi et Elisa Caligiana (Université de Bologne - SSLiMIT, Forlì) analysent les métaphores à l'œuvre dans les verbes supports et appropriés de noms prédicatifs de sentiment ; entre métaphores vives et catachrèses, elles dessinent des réseaux lexicalisés mais productifs de concepts métaphoriques (cf. Lakoff & Johnson, 1981) souvent communs à plusieurs langues (exemples en français, en italien et en anglais). Raffaele Simone et Francesca Masini (Université Roma Tre) s'attachent aux *noms supports* véhiculant des traits de nature verbale, tels que l'aspect (*coup de colère*, *éclat de rire*), là encore souvent identiques ou voisins dans plusieurs langues.

CONCLUSION

Si la classification morphologique et syntaxique des verbes a pu se faire dans un relatif consensus, il n'en va pas de même avec la classification sémantique (si liée soit-elle à la syntaxe), sans cesse remaniable, liée à la fois à l'intuition partagée et à des décisions théoriques et méthodologiques. Le rapprochement des travaux et la discussion devraient conduire à des résultats susceptibles de recevoir des applications tant dans le traitement automatique que dans la didactique des langues.

BIBLIOGRAPHIE

- DUBOIS J. & DUBOIS-CHARLIER F. (1997), *Les Verbes français*, Larousse.
- FRANÇOIS J., LE PESANT D. & LEEMAN D. (éds) (2007), *Le classement syntactico-sémantique des verbes français*, *Langue française* 153.
- GREZKA A. (2006), *Les prédicats de perception. Traitement de la polysémie (Les sens des sens)*, thèse de doctorat en Sciences du Langage, Université Paris XIII Villetaneuse.
- GREZKA A. (à paraître en 2008), "Des domaines aux classes sémantiques : l'exemple des verbes de <coiffure>", *Le traitement du lexique. Catégorisation et actualisation*, Actes du Colloque international des doctorants et jeunes chercheurs en Linguistique, 17-18 novembre 2006, Faculté des Lettres et des Sciences Humaines de l'Université de Sousse, Tunisie.
- GROSS G. (1994), "Classes d'objets et description des verbes", *Langages* 115, 15-30.
- GROSS G. (1996), *Les expressions figées en français*, Ophrys.

6 Le *Dictionnaire du français usuel* de J. Picoche et J.-C. Rolland (De Boeck-Duculot, 2002), à visée didactique, décrit également les entrées par une formule actancielle dont la présentation est voisine.

- LAKOFF G. & JOHNSON M. (1981), *Metaphors we Live by*, The University of Chicago Press, Chicago - Londres.
- LE PESANT D. & MATHIEU-COLAS M. (éds) (1998), *Les classes d'objets, Langages*, 131, Paris, Larousse.
- LEVIN B. (1993), *English Verb Classes and Alternations, A preliminary investigation*, The University of Chicago Press.
- MATHIEU-COLAS M. (à paraître en 2008), "Les classes de verbes : syntaxe et sémantique", *Le traitement du lexique. Catégorisation et actualisation*, Actes du Colloque international des doctorants et jeunes chercheurs en Linguistique, 17-18 novembre 2006, Faculté des Lettres et des Sciences Humaines de l'Université de Sousse, Tunisie.
- MEJRI S. (1997), *Le figement lexical. Description linguistique et structuration sémantique*, Publications de la Faculté des Lettres de la Manoubia.
- MEL'ČUK I. & POLGUÈRE A. (2007), *Lexique actif du français*, De Boeck.
- PETIT G. (2003), "Figement lexical et lemmatisation", *Cahiers de lexicologie* 82 (2003-1).
- PICOCHÉ J. & ROLLAND J.-C. (2002), *Dictionnaire du français usuel*, De Boeck-Duculot.
- REY-DEBOVE J. (2004), *Le Robert Brio, Analyse des mots et régularités du lexique*, Préface.

Abréviations pour les traits syntactico-sémantiques utilisés dans les articles :

hum	humain
humcol	humain collectif
ani	animal
vég	végétal
in	inanimé
inc	inanimé concret
ina	inanimé abstrait
npc	nom de partie du corps
loc	locatif
tps	temps