

Les restrictions sémantiques des arguments verbaux : une question de fréquence d'usage

Beatriz Sánchez Cárdenas
Université de Strasbourg - LILPA EA 1339
beatriz.sanchez@unistra.fr



Synergies France n° 6 - 2010 pp. 41-50

Résumé : *La sémantique des arguments du verbe n'a pas fait l'objet d'études approfondies, exception faite de quelques travaux, comme ceux du laboratoire LADL de l'Université Paris VII sous la direction de Gaston Gross. Pourtant, apprenants et futurs enseignants de FLE trouveraient bénéfique une description des arguments du verbe d'un point de vue sémantique. Nous proposons une méthodologie pour permettre aux étudiants et futurs enseignants de FLE d'exploiter les corpus afin d'en tirer les informations pertinentes sur la formation correcte des énoncés en français. Notre étude est centrée sur le syntagme verbal et, plus précisément, sur l'exemple d'énumérer, qui servira à étayer nos propos.*

Mots-clés : *prédicat verbal ; classes sémantiques ; linguistique de corpus*

Abstract: *The semantic arguments of the verb have not been extensively studied, except for some research, such as those under the direction of Gaston Gross from the LADL laboratory, University of Paris VII. However, learners and future teachers of French as a Foreign Language (FFL) would benefit from a description of the verbal arguments from a semantic point of view. We propose a methodology to enable students and future teachers of FFL to exploit corpuses and thus extract relevant information on the correct enunciative formation in French. Our study focuses on the verb phrase and, more specifically, on the example énumérer, which will be used to support our remarks.*

Keywords: *verbal predicate; semantic classes; corpus based linguistics*

Resumen: *La semántica de los argumentos del verbo no ha suscitado apenas estudios exhaustivos, exceptuando trabajos como los del laboratorio LADL de la Universidad Paris VII bajo la dirección de Gaston Gross. Sin embargo, la descripción de los argumentos del verbo desde el punto de vista semántico resulta muy beneficiosa para los estudiantes de FLE así como para los futuros profesores de FLE. Exponemos una metodología destinada a permitir a los estudiantes de FLE y futuros profesores de FLE un análisis de los corpus que les permita extraer la información necesaria para la formación de enunciados en francés. Nuestro estudio se centra en el sintagma verbal y, más concretamente, en el caso del verbo énumérer que servirá para ilustrar dicha metodología.*

Palabras clave: *predicado verbal; clases semánticas; lingüística de corpus*

Linguistique et didactique sont deux disciplines autonomes (Galissou, 1989) qui peuvent se nourrir mutuellement sans entretenir pour autant des rapports de « vassalité » (Galissou, 1994). En l'occurrence, la didactique peut trouver des bénéfices dans la linguistique de corpus (Granger, 1997). Un cas de figure est la description sémantique des arguments verbaux basée sur une analyse de corpus. Nous proposons une méthodologie pour extraire ces informations du corpus. Elle s'avère utile pour l'apprenant de FLE et pour la formation de futurs enseignants de FLE dans le sens où elle contribue au processus d'« autonomisation formative » (Galissou, 1994) en ce qui concerne l'acquisition de la phraséologie propre de la langue étrangère. L'exemple du verbe *énumérer* illustrera nos propos.

1. Importance de la caractérisation sémantique des noms des arguments

La description du verbe vise traditionnellement à définir des phénomènes grammaticaux tels que la syntaxe, l'aspect ou le temps. Or, la description du syntagme verbal ne s'arrête pas à ce genre de phénomènes. Nous nous intéressons notamment aux types de noms (dorénavant N) des arguments verbaux. Cela est intéressant si nous considérons que des structures similaires supposent des interprétations sémantiques différentes :

- 1) Je compte sur ma sœur.
- 2) Je compte sur ma table.

Cette information est d'autant plus importante que ce n'est que lorsque le verbe est saturé en contexte par le substantif qu'il acquiert pleinement son sens. Par conséquent, une description sémantique des substantifs des arguments verbaux s'impose pour pouvoir rendre compte des constructions verbales. En effet, le sens des verbes est déterminé, en partie, par le choix lexical des noms admis en position de sujet ou de complément. Il est donc nécessaire d'établir des critères pouvant différencier les types de N se situant dans les arguments des verbes. Ce genre d'information s'avère profitable pour l'acquisition de la phraséologie, qui joue un rôle majeur dans l'enseignement/apprentissage d'une langue étrangère :

« La unidad fraseológica se convierte en una unidad excelente para la adquisición y retención tanto del léxico como de la sintaxis » (Molina García, 2006 : 6).

Nous explorons ici les applications pratiques de cette approche pour l'enseignement/apprentissage du FLE.

2. Les classes d'objets du LADL

La description sémantique des arguments du verbe n'est pas un sujet nouveau. L'analyse des classes d'objets de Gross (1994) en constitue un exemple remarquable : le laboratoire LADL décrit les arguments des verbes en fonction de traits syntactico-sémantiques. Par exemple « Luc part en voiture » est détaillé ainsi : « partir en/N0:hum/N1:mt-fc ». Cette description indique que le verbe « partir en » a un premier argument humain et un deuxième argument défini comme appartenant à la classe <moyen de transport> et à la sous-classe <transports ferroviaires> (Gross, 1994 : 20).

Malgré l'intérêt de cette approche, notre étude s'éloigne du modèle des classes d'objets pour plusieurs raisons :

- a) La classification des N dans les classes d'objets s'appuie sur leur distribution syntaxique. Le recours à la réalité référentielle des substantifs est écarté. Cela présente l'avantage d'une systématisation puisque les propriétés syntaxiques s'y prêtent mieux que les sémantiques. Or, la syntaxe peut ne pas être le seul critère retenu pour décrire les arguments verbaux sur le plan sémantique. Le recours à la référence est d'autant plus nécessaire dans le cadre d'une description visant à l'enseignement du FLE que les locuteurs n'emmagasinent pas les substantifs selon des critères linguistiques mais d'après leurs propriétés référentielles, tirées d'une connaissance sensorielle du monde.
- b) Les critères de description des arguments des verbes du LADL sont mixtes. Certains arguments sont décrits tantôt en fonction des opérateurs généraux, tantôt selon leur classe d'objet (cf. la description de « partir en »). Pour une meilleure cohérence, une séparation nette des critères formels (les opérateurs généraux) et des traits sémantiques (les classes d'objets) est souhaitable.
- c) La formalisation proposée par le laboratoire LADL a l'avantage d'être applicable à un modèle de traitement automatique de la langue. Son désavantage est que ce type de description n'est pas assez intuitif et, par conséquent, peu adapté à l'enseignement du FLE.
- d) Aucun inventaire des classes d'objet n'est, à notre connaissance, disponible, ce qui représente un désavantage majeur pour réutiliser ce modèle.

Toujours est-il que la description des arguments des verbes en fonction de l'appartenance des substantifs à leurs paradigmes sémantiques reste très utile pour l'apprentissage d'une langue étrangère. Nous proposons ici un modèle descriptif.

2. Pour une description du langage basée sur la fréquence d'usage dans les corpus

La description du syntagme verbal en fonction de la fréquence d'usage des arguments verbaux reflète l'usage réel de la langue telle qu'elle est pratiquée. Notre étude se situe donc dans une approche descriptive du langage. Elle s'intéresse aux réalisations effectives du langage plutôt qu'aux possibilités théoriques du système linguistique (Jacques, 2005 : 29). L'idée est que « plus une expression est fréquente, plus elle est usuelle et, vice-versa, moins elle apparaît dans le corpus, moins elle est usuelle [...]. La fréquence d'une expression dans le corpus est le reflet de sa fréquence réelle dans la langue » (Geyken, 2008 : 78). Ce type de description suppose la prise en compte des associations de mots les plus usuelles d'un point de vue statistique, c'est-à-dire des associations des mots les plus usuelles face à celles qui sont moins fréquentes :

« In order to account satisfactorily for language in use, a theory of language performance will be needed, a theory that is statistical and probabilistic, rather than certain and cut and dried. Of all the many words, uses, and structures that are possible in a language, it will show us how to pick out just those that are normal, and it will relate other uses to the norms by a theory of exploitations: a set of

exploitation rules that will say how a normal use may be exploited to form metaphors and other unusual uses, and what the constraints are. » (Hanks, 2000 : 12).

Si nous adoptons une telle approche, c'est parce que nous nous intéressons à ce que Hanks appelle le « linguistic behaviour » des verbes :

« When we look at a corpus and are astonished by the overwhelming and often unsuspected frequency of conventional phraseology, we are looking at traces of thousands of instances of fashionable linguistic behaviour. » (Hanks, 2000 : 11).

Le spécialiste Béjoint se réfère à ce phénomène en parlant de la description de l'« usage syntagmatique réel » (Béjoint, 2007 : 12).

En définitive, nous nous intéressons aux associations de verbes et substantifs que les usagers natifs ont tendance à produire avec un haut degré de répétition, autrement dit, aux « contextes par défaut » des verbes (Hanks, 2005 : 63). Nous pensons que l'apprentissage de ces régularités peut faciliter l'apprentissage d'une langue étrangère.

3. Description des N des arguments verbaux : paramètres de classification

Les données extraites du corpus restent information brute à moins de s'appuyer sur un modèle théorique solide. Il est donc nécessaire de définir clairement les critères de classification des N des arguments verbaux.

3.1. Traits linguistiques

Nous nous appuyons, premièrement, sur des notions linguistiques binaires classiques qui ont l'avantage de décrire n'importe quel N. Elles se fondent sur quatre couples de distinctions :

a) concret/abstrait : la différence entre ces deux traits se manifeste linguistiquement par le fait que les N abstraits ont « un certain type de complémentation » qui sont des arguments du N abstrait, contrairement aux N concrets qui n'ont pas « des places d'arguments vides à remplir » (Flaux et Van Velde, 2000 : 31).

Comparons par exemple *construction* (*la construction de la maison*) ou *mépris* (*le mépris de cet homme pour les richesses*) à un N comme *arbre*, qui « ne détermine pas autour de lui, comme le fait *construction*, des places d'arguments vides à remplir. Ses compléments, s'il en a, ne sont pas prédéterminés par le type de sens qu'il véhicule » (Flaux et Van Velde, 2000 : 31).

b) animé/inanimé : les N animés incluent les catégories des humains, animaux et végétaux. Le reste du lexique se range du côté inanimé. Vu la grande diversité de ce deuxième groupe, le trait inanimé pourrait être considéré comme étant peu opératoire (Flaux et Van Velde, 2000 : 35).

c) humain/non humain : cette division opère une distinction dans la classe des animés, pour mieux décrire les verbes qui exigent un argument humain pour leur réalisation, ce qui a une importance pour la description du lexique verbal.

d) naturel/fabriqué¹ : les N fabriqués dénotent « des objets concrets caractérisés comme "fabriqués", notion passive qui entraîne l'idée d'un agent. Ce faisant, ils constituent

une classe de N à la fois non prédicatifs et cependant pourvus d'arguments » (Flaux et Van Velde, 2000 : 37). Cette division distingue deux types de non humains. Sera considéré comme fabriqué tout N dont le référent a été produit par l'action humaine. Les référents des N naturels n'ont, au contraire, pas besoin d'une quelconque action humaine pour exister.

Mais ces traits ont un pouvoir explicatif restreint étant donné qu'ils fournissent des informations pauvres du point de vue lexical. Cette classification est à compléter, ensuite, par une seconde caractérisation de type sémantique.

3.2. Classes sémantiques des N

Les N des arguments verbaux sont décrits, deuxièmement, d'un point de vue sémantico-référentiel. Pour cela, des « classes sémantiques des N » sont définies. Pour donner un exemple, voici les catégories qui apparaissent le plus fréquemment dans les « verbes de comptage² » à partir de l'étude des corpus, à savoir : « N individu », « N institutionnel », « N profession », « N corps », « N flore », « N faune », « N objet naturel inanimé », « N artefact », « N événement », « N lieu construit », « N énonciatif », « N construction intellectuelle », « N propriété psychologique », « N classificateur », « N mesure », « N économique ». Chaque catégorie est définie, d'abord, d'un point de vue référentiel et puis à l'aide d'une série de tests distributionnels.

À titre d'exemple, et pour des questions d'espace, nous ne présentons que la catégorie « N institutionnel ». Elle regroupe les N faisant référence à des institutions, organisations, entités publiques ou entreprises comme *parlement*, *mairie*, *Peugeot*.

Pour ce qui est de sa distribution, la catégorie « N institutionnel » partage certaines caractéristiques avec celles des « N individu » ; ainsi, par un processus de métaphorisation ou de métonymie, ils peuvent apparaître en position de sujet avec des verbes comme *décider*, qui décrivent des actions propres aux entités cognitives (en principe réservées aux « N individu ») :

- Le « N institutionnel » a décidé de...

Ces N se trouvent par conséquent souvent en position de sujet des verbes de comptage :

3) La France dénombre près de 5 000 sanctuaires, églises, cathédrales, basiliques, chapelles, ermitages, oratoires ou calvaires³.

4) La préfecture de police a dénombré 800 manifestants dans ce cortège⁴.

Or, à la différence des « N individu », ils ne peuvent pas être qualifiés par un complément attribut animé :

*Le « N institutionnel » est un traître.

Ils peuvent être suivis d'un verbe décrivant la composition du sujet (*rassembler*, *regrouper*) et ayant comme complément d'objet un « N humain » :

- (Le/la) N institutionnel regroupe tant de « N profession⁵ ».

Ils admettent un prédicat décrivant leur localisation :

- N institutionnel se situe à Strasbourg.

Les N *parlement* et *mairie* réunissent ces propriétés :

- 5) Le parlement a décidé de...
- 6) *Le parlement est un traitre
- 7) Le parlement regroupe 76 députés.
- 8) Le parlement se situe à Strasbourg.
- 9) *La mairie a décidé de fermer le théâtre.*
- 10) *La mairie est un traitre.
- 11) *La mairie regroupe 13 conseillers municipaux.*
- 12) La mairie se situe sur la place.

La description des catégories sémantiques des arguments permet de mieux définir les structures des verbes. Elles ont pour but d'aider l'apprenant de FLE à construire des énoncés non seulement recevables mais aussi proches de ceux qu'un locuteur natif prononcerait. Nous l'illustrerons à l'aide de l'exemple d'*énumérer*.

4. L'exemple du verbe *énumérer*

Le *Trésor de la langue française informatisé* définit le verbe *énumérer* par : « énoncer un à un les éléments d'un ensemble » comme dans l'exemple « il a fort exactement énuméré toutes les circonstances » ou « il énumérait sur ses doigts les Ambertois qui avaient (...) fait fortune ». Le dictionnaire *Le Petit Robert* n'apporte pas plus d'éclaircissements. Dans le dictionnaire électronique *Les verbes du français* de Dubois et Dubois-Charlie⁶, *énumérer* est décrit comme « énumérer une suite ».

Ces dictionnaires permettent d'élucider la structure syntaxique d'*énumérer* (S V COD). Or, ces renseignements n'informent pas sur la nature du sujet ni sur les éléments de l'ensemble énuméré, ce qui est essentiel pour qu'un apprenant de FLE puisse produire des énoncés « naturels ». L'étude de corpus exposée ci-dessous peut être exploitée par les futurs enseignants de FLE pour accéder aux structures naturelles et, par ce biais, pour combler les lacunes des dictionnaires.

4.1. Description du corpus d'*énumérer*

Notre corpus se compose d'une série de textes journalistiques, littéraires et politiques qui nous ont semblé les plus appropriés pour rendre compte du type de langue que nous considérons, à savoir le registre formel du français contemporain (depuis 1900).

Nous avons pour cela utilisé la base de données Frantext⁷, où nous avons sélectionné les romans, essais et mémoires de 1900 à 2000. Nous avons aussi fait appel aux bases de données Wortschatz⁸ de l'Université de Leipzig. Pour finir, nous avons eu recours aux textes du Parlement européen par le biais de Copuseye⁹. Enfin, nous avons sélectionné une centaine d'occurrences de Google à partir de sources fiables. Voici un tableau récapitulatif des résultats auxquels nous avons abouti pour *énumérer* :

Tableau 1. Constitution du corpus d'*énumérer*

	<i>Frantext</i>	<i>Wortschatz</i>	<i>Parlement européen</i>	<i>Google</i>	<i>Total</i>
nombre d'occurrences	140	143	83	100	486

Les corpus sont exploités avec l'analyseur de corpus semi-automatique « WordSmith tools ».

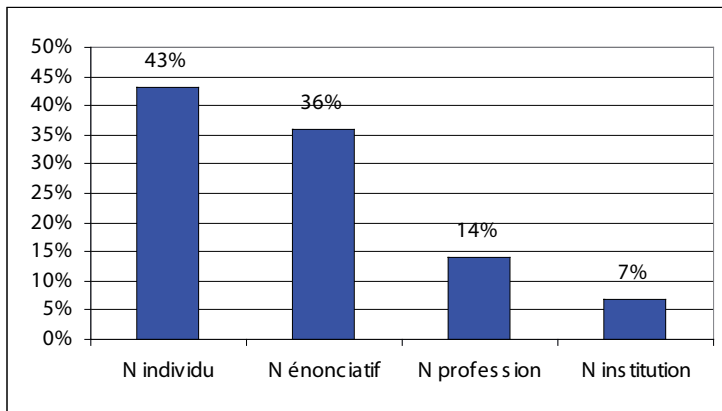
4.2. Le syntagme verbal d'*énumérer*

L'étude de corpus nous a permis d'observer la position sujet d'*énumérer* occupée par quatre types de N différents :

- i) Un être humain : *Rajsfus, président, M. Valverde López.*
- ii) Une institution : *Andra, Commission.*
- iii) Une discours (ou une partie d'un discours) : *proposition, communication.*
- iv) Un document (ou une partie de document) : *rapport, quotidien, lettre.*

D'après les tests distributifs, ces N correspondent aux catégories sémantiques de « N individu » (*femme, M. McCarthy*), « N profession » (*ministre, procureur*), « N institutionnel » (*Observatoire, Andra*) et « N énonciatif » (*communication, rapport*). Ils apparaissent dans notre corpus avec cette fréquence :

Figure 1. Types de N du sujet d'*énumérer*



Voici des exemples issus de nos corpus :

- 13) M. Foulon énumère les quatre facteurs de sa réussite.
- 14) Le **rapport** énumère de nombreuses violations flagrantes et massives des droits de l'homme.
- 15) Le **Président** énumère les principales décisions adoptées.
- 16) La **Commission** énumère 25 produits phytosanitaires dont l'utilisation est admise dans l'UE.

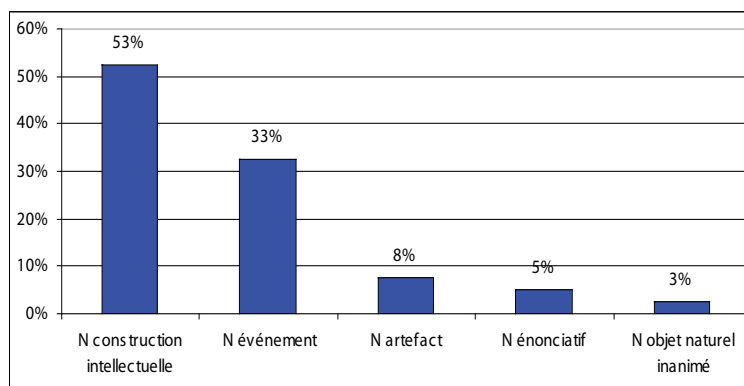
Les référents de ces agents sont des entités cognitivement capables d'accomplir une énumération¹⁰.

L'analyse du corpus révèle la position de complément d'objet direct occupée par les classes sémantiques suivantes :

- a) N construction intellectuelle : *causes, raisons, conséquences.*
- b) N évènement : *réformes, modifications, courses.*
- d) N artefact : *transports, produits phytosanitaires.*
- d) N énonciatif : *mots, discours.*
- e) N lieu construit : *sites, endroits.*
- f) N objet naturel inanimé : *matériaux, roches.*

Le graphique suivant permet de voir les pourcentages de la fréquence des catégories dans notre corpus :

Figure 2. Fréquences des types de N de l'UNDERGOER d'énomérer



Voici des exemples pour chacune de ces catégories :

- 17) Le rapport énumère les cinq **critères** du transport du bétail.
- 18) Le rapport énumère une douzaine de **rencontres** à caractère sexuel entre M. Clinton et l'ancienne stagiaire.
- 19) Le rapport énumère dans son avis 25 **médicaments** autorisés.
- 20) Le dictionnaire de Jacques Attali énumère plus de 400 **mots** clés du futur.
- 21) Une femme énumère les trois grandes familles de **roches** : les roches sédimentaires, magmatiques et métamorphiques.

Il est intéressant de noter qu'énomérer est souvent lié à des « N construction intellectuelle » qui servent à articuler le discours, tout comme les « N énonciatif », qui apparaissent dans 5% des cas. En outre, il y a très peu de N concrets dans le complément du verbe énumérer ; dans 89% des cas, il est abstrait.

Ce type de description du verbe a l'avantage de rendre compte des associations du verbe énumérer avec les substantifs dans l'énonciation ou, autrement dit, de l'« attraction » (en termes de Firth) que le verbe exerce sur ses arguments (Gledhill et Frath, 2007 : 65). Cette étude de la langue ne dissocie pas syntaxe

et lexique. Dans cette approche, le lexique et la grammaire forment un continuum. Cette analyse s'avère profitable pour les étudiants de FLE. Ainsi, l'introduction en cours de FLE de l'exploitation des corpus serait bénéfique.

Conclusion

Selon le protocole d'analyse établi, nous avons décrit les usages les plus courants du verbe *énumérer*, que l'on peut appeler aussi « normal use¹¹ » (Hanks, 2004 : 272) ou « conventions idiosyncrasiques » (Fontenelle, 2000 : 241). La méthodologie suivie ici pourrait être appliquée à d'autres groupes de verbes.

Les applications d'une telle approche sont à terme très fécondes. Suivant ces postulats, la création d'un dictionnaire des verbes, capable de renseigner les usagers sur la façon dont les mots se combinent, a une grande utilité pour les apprenants de langues étrangères. La voie des dictionnaires décrivant le contexte des mots a déjà été tracée, depuis longtemps, par des travaux comme celui de Descamps (1973). Il ne reste plus qu'à la suivre.

Bibliographie

Descamps J.L., 1973. « Présentation d'un dictionnaire contextuel de français pour les sciences de la terre ». *Langue française*, vol 17, n° 1, p. 81-111.

Flaux, N. & Van de Velde D., 2000. *Les noms en français : esquisse de classement*. Paris : Ophrys.

Fontenelle, T., 2000. « A Bilingual Lexical Database for Frame Semantics ». *International Journal of Lexicography*, n° 13/4, p. 232-248.

Galisson, R., 1978. *Recherches de lexicologie descriptive : La banalisation lexicale. Le vocabulaire du football dans la presse sportive*, Contribution aux recherches sur les langues techniques. Paris : Nathan.

Galisson, R., 1994. « Formation à la recherche en didactologie des langues-cultures ». *Études de Linguistique Appliquée*, n° 95, juillet-septembre 1994, p. 119-159.

Molina García, D., 2006. *Fraseología bilingüe. Un enfoque lexicográfico-pedagógico*. Granada : Editorial Comares.

Geyken, A., 2008. « Quelques problèmes observés dans l'élaboration de dictionnaires à partir de corpus ». *Langages*, n° 171/3, p. 77-94.

Granger S., 1997. « Automated retrieval of passives from native and learner corpora: precision and recall ». *Journal of English Linguistics*, n° 25/4, p. 365-374.

Gross, G., 1994. « Classes d'objets et description des verbes ». *Langages*, n° 115, p. 15-30.

Hanks, P., 2000. « Contributions of Lexicography and Corpus Linguistics to a Theory of Language Performance ». *Euralex Proceedings*, n° 1, p. 3-13.

Hanks, P., 2004. « The syntagmatic of metaphor and idiom ». *The International Journal of Lexicography*, n° 17/3, p. 245-274.

Jacques, M. P., 2005. « Pourquoi une linguistique de corpus ? ». In Williams G. (éd.), *La Linguistique de corpus, Journées de la Linguistique de corpus*. Rennes : Presses Universitaires de Rennes. p. 21-29.

Sánchez Cárdenas, B., à paraître. « Enhancing translation dictionaries through corpus analysis ». *International Journal of Translation*. Bahri publications.

Notes

¹ La notion de N « naturel » versus « fabriqué » est discutable, mais nous l'admettons provisoirement dans le cadre de cette recherche.

² Le groupe des « verbes de comptage » intègre les verbes *chiffrer, compter, dénombrer, comptabiliser, énumérer, numéroter* et *recenser*.

³ [www.lexpress.fr/informations/saintes-ecritures-la-france-des-pelerinages_614758.html]

⁴ [<http://www.lepoint.fr/actualites/2008-11-27/les-chercheurs-manifestent-l-agence-nationale-de-la-recherche/914/0/295442>]

⁵ Voir catégorie 3.

⁶ Ce dictionnaire a été élaboré par le laboratoire MoDyCo. Il est formé de 25 610 entrées verbales représentant 12 310 verbes différents.

⁷ La recherche de la forme conjuguée d'un verbe est « &#m infinitif du verbe », par exemple « &#mcompter ».

⁸ [http://corpora.informatik.uni-leipzig.de/?dict=fr_site]

⁹ [<http://corp.hum.sdu.dk>]

¹⁰ Bien entendu, ni les « N énonciatif » ni les « N institution » n'ont de propriétés cognitives *à priori* mais ils les acquièrent ; dans le premier des cas, par un phénomène de polyphonie (Ducrot 1972) et dans le deuxième cas par métonymie (Sánchez Cárdenas, à paraître).

¹¹ Fillmore (1975) avait même suggéré d'analyser le sens d'un mot selon sa ressemblance à l'usage prototypique, c'est-à-dire l'usage le plus courant. Hanks (2004) établit un parallélisme entre l'usage normal d'un verbe (celui qui est pratiqué) et le sens littéral (celui qui est possible). Ces distinctions reprennent la différence entre *performance* et *compétence*. Dans cette recherche, nous avons essayé de décrire cet « usage normal ».