

BULLETIN D'INFORMATIQUE APPROFONDIE ET APPLICATIONS

COMPUTATION - INFORMATION

N° 61 - MARS 2002

DIRECTEUR :

Jean - Michel Knippel

REDACTEUR EN CHEF :

Edmond Bianco

REDACTEUR ADJOINT :

Sami Hilala

SECRETARIAT :

Kalassoumi Adjilani

Université de Provence
Equipe Hermès. Case 33
3, place Victor Hugo
F - 13331 Marseille Cedex 3
Téléphone: (0)4 91 10 62 30
Télécopie : (0)4 91 50 91 10

1 EDITORIAL
Informatique et complexité

par Edmond Bianco

DEPOSITAIRE :

Université de Provence
Bibliothèque Vniversitaire
1, place Victor Hugo
F - 13331 Marseille Cedex 3
Téléphone: (0)4 91 10 85 29
Télécopie : (0)4 91 95 75 57

**5 BASES DE DONNEES ET LEXIQUES-
GRAMMAIRES: UN EXEMPLE
D'EXPLOITATION LINGUISTIQUE
DE L'OUTIL INFORMATIQUE**

par Nadia Mesli

IMPRIMEUR :

Université de Provence
Service Reprographie
3, place Victor Hugo
F - 13331 Marseille Cedex 3
Téléphone: (0)4 91 10 60 48

23 VOZZAVEDIBISAR
Petit lexique à usage de littérature politique

par Edmond Bianco

<http://scamup.univ-mrs.fr/biaa>

Publication triannuelle, gratuite, de l'Université de Provence

Impression : Février 2003

ISSN 0291 - 5413

ÉDITORIAL

Informatique et complexité

Edmond Bianco

Le premier abord du Monde est complexe.

L'interprétation de cette première phrase est elle-même complexe, en ce sens que le sens du mot 'complexe' n'est pas spécialement clair. L'univers que nous percevons dès notre débarquement sur cette planète est un effroyable fouillis dont on a immédiatement le plus grand mal à distinguer les détails. Ce n'est qu'au cours du temps que peu à peu on parvient à séparer des éléments remarquables qui surgissent lentement du magma, dont ils finissent par s'extraire à force qu'on les distingue. (Voyez combien cette phrase, pourtant simple, paraît complexe ! Mais peut être n'est-elle que compliquée ?). Une certaine constance, dans l'apparition d'un détail particulier, finit par faire en sorte, qu'on se l'approprie tel quel en le casant dans l'ambiance de son apparition. L'éducation est une forme d'organisation destinée à favoriser par acquisitions successives, l'appréhension de l'univers, et de le rendre sinon convivial, du moins vivable.

Les phénomènes de l'acquisition sont eux-mêmes complexes, mais ce qu'on peut dire simplement c'est qu'à partir du moment où un phénomène est devenu familier, il perd son apparence complexe. A titre d'exemple, on prend quelques objets courants : 'le mot', 'le crayon', 'le nuage'. En se limitant à un simple point de vue utilitaire, la compréhension de l'objet 'crayon' implique son utilité, la manière de s'en servir, la manière de l'entretenir. L'acquisition de ces quelques notions est en général rapide, après quoi le crayon perd tout mystère. Pour ce qui est du nuage, les niveaux d'appréhension croissent rapidement. Du point de vue du contemplatif, le nuage devient un objet décoratif, un ornement de paysage, il peut aussi devenir oppressant dans sa version pluvieuse et sombre, voire menaçant dans sa forme orageuse. Il n'est pas impossible d'ailleurs, que ce dernier avatar ait puissamment contribué à la naissance des religions par l'intermédiaire de superstitions, qui se seraient peu à peu organisées. Par sa nature même, la notion de religion devient d'une complexité insondable, car, pratiquant l'inexplicable, elle fuit toute tentative d'appropriation, qui se retrouve perpétuellement rejetée en l'être suprême.

Pour en revenir à notre nuage, la difficulté de son interprétation s'accroît d'autant que l'apparence de ce phénomène dépend fortement de la distance à laquelle on l'observe. Quant à l'objet 'mot', l'appropriation en est encore plus touffue, variable selon les individus, variable au cours du temps. Des mots comme 'table', 'bureau', ont des vies riches, partant des tables de la loi de Moïse, à la table d'hôte, en passant par tout ce qui est plat et où peuvent se poser des choses dessus, et du bureau, le meuble, jusqu'aux immenses bureaux embusqués dans les deux tours de Manhattan.

Comme en informatique, la complexité se scinde en deux parties 'conceptuellement' disjointes, il y a la complexité objet, celle que la complexité de la compréhension est chargée d'analyser, et, en face il y a la complexité des phénomènes de compréhension. A une échelle de complexité bien plus faible l'informatique est structurée de la même manière, il existe un ensemble objet constitué du code et de la donnée, et il existe un algorithme capable de traiter cette donnée par interprétation du code, c'est la machine universelle. La complexité de la machine universelle est une sorte de reflet figé de la complexité de structure du code qui, lui, peut varier à l'infini dans son expression.

La complexité du phénomène de compréhension ne simplifie pas l'approche des phénomènes complexes. En effet, on ne comprend bien que des choses qu'on puisse ramener facilement à ce qu'on connaît déjà. Or, il existe des objets, des manifestations qui ne se laissent pas ramener facilement à des notions connues. Pour aborder l'inconnu, l'esprit a besoin de se forger des hypothèses. Et il est assez facile d'imaginer, par exemple, que la foudre, avec son éclair aveuglant et son tonnerre assourdissant, faute de mieux, ait induit l'hypothèse d'une colère éruptée par une puissance supérieure. Simple transposition d'un phénomène bien humain et bien connu. De la même manière, l'apparition dans le ciel d'une étoile très brillante ne peut être interprétée que comme le signal d'un bouleversement important de la vie des hommes. Il faudra de très longues études pour 'décomplexifier' la nature des nova et supernova.

Des quelques remarques qui précèdent, on pourrait déduire que la complexité d'un phénomène ou d'un objet est très fortement liée au mystère qui l'entoure. D'où l'idée de décomposer l'objet de l'étude dans la limite du possible, en des éléments plus simples jusqu'à ce qu'on atteigne des parties déjà connues. L'élément simple est, en fait, un élément qu'on a pu déjà s'approprier, qu'on sait isoler, et à qui on attache un jeu bien défini de propriétés. Mais est-il pour autant moins complexe ? Une telle décomposition n'est pas toujours possible de manière directe. Dans le cas du soleil, comment imaginer le mécanisme de sa fabuleuse production d'énergie sans connaître au préalable un peu de physique nucléaire.

Je distinguerais volontiers dans un même domaine, deux aspects de la complexité. La complexité structurelle, celle qui caractérise l'organisation du milieu, et par opposition la complexité dynamique, qui découle des déformations possibles du milieu. Par exemple, en informatique un langage se définit avec des règles de construction, et toute phrase qui respecte ces règles est une phrase du langage. Bien entendu ces règles se partagent la syntaxe et la sémantique du langage. La complexité structurelle découle de l'organisation du langage, elle est la conséquence de l'application des règles. La variété illimitée des phrases, qu'on peut construire, constitue alors la complexité dynamique. La complexité structurelle est du ressort de la machine universelle. La complexité dynamique est formellement du ressort du système de gestion.

Il est bien connu que, débutant dans une branche quelconque de la connaissance, on éprouve un sentiment puissant de mystère et d'inconnu, et au fur et à mesure qu'on pénètre dans le sujet les ombres s'éclairent, les difficultés s'effacent lentement, et après une bonne pratique du sujet, tout paraît tout à coup évident (Il y a bien entendu des cas où rien ne se passe d'une manière aussi aisée !). On a ainsi construit des concepts complets, autonomes dont on sait parfaitement comment ils s'articulent à d'autres concepts tout aussi bien connus, et la complexité a disparu, tout au moins à ce niveau. Il suffit alors de nommer précisément le nouveau concept pour désigner clairement un bloc de complexité dont les propriétés essentielles sautent instantanément à l'esprit, sans qu'on ait à en refaire chaque fois, toute l'analyse.

C'est ainsi que parler de 'Machine Universelle' définit parfaitement une classe d'algorithmes dont chacun peut être très complexe, et dont la complexité sera fonction de la nature du langage pris en compte.

Le sujet est complexe et d'une aridité qui ne laisse aucune chance de faire surgir la moindre idée amusante, la moindre trace d'humour, mais peut être faudrait il rechercher un domaine où complexité et humour feraient bon ménage. En cherchant un peu... Par exemple la politique, sujet éminemment complexe, épaulée par son ombre, l'économie. Il existe des lois simples qui représentent des phénomènes extrêmement complexes, par exemple on sait que les promesses électorales donneront des réalisations exactement contraires. Par exemple, une promesse de réduction d'impôt aboutira forcément à un accroissement substantiel de ce que vont payer les contribuables. Pareil pour l'immense complexité du code pénal, le cas n'est pas rare, non plus, de voir d'énormes délinquants invulnérables de par leur situation sociale, tonner contre de minuscules délinquants à leur corps défendant.

Peut être n'est il pas mauvais de relire DESCARTES, Le discours de la méthode?

Mais il n'est pas pour autant nécessaire de lire les œuvres d'Edgard Morin sur la complexité, bien que ses jeux de mots du style « Intelligence de la Bêtise et Bêtise de l'Intelligence » soient déjà une amorce plus ou moins voulue du rapprochement humour-complexité.

Edmond Bianco

Bases de données et lexiques-grammaires : un exemple d'exploitation linguistique de l'outil informatique

Nadia MESLI

mesli@up.univ-mrs.fr

1. Introduction

Nous présenterons ici comme exemple d'exploitation linguistique de l'outil informatique un projet d'élaboration de dictionnaires électroniques bilingues (allemand-français) mené depuis plusieurs années dans le cadre de maîtrises en linguistique allemande à l'université de Provence (Aix-en-Provence). Cette étude linguistique (Mesli 1998) a son origine dans des travaux sur les propriétés syntaxiques et sémantiques d'expressions relevant du domaine des émotions et sentiments en allemand (Bresson 1987, Mesli & Bresson 1992, Bresson & Dobrovol'skij 1998), et s'apparente aux tentatives d'élaboration de dictionnaires syntaxiques et sémantiques systématiques spécialisés, ou *grammaires locales*, constituant une application directe de la théorie du lexique-grammaire (Gross 1995, Balibar-Mrabti 1995). L'objectif visé par un tel projet est d'une part de recenser le vocabulaire spécifique à un domaine de connaissance extralinguistique et encyclopédique organisé par concepts sous la forme d'un thesaurus, et d'autre part d'en établir les propriétés combinatoires. Il s'agira de décrire la démarche linguistique et l'application informatique nécessaires à l'établissement de dictionnaires électroniques spécialisés, en partant des critères de sélection du domaine d'étude pour aboutir aux lexiques-grammaires sur support informatique.

2. Linguistique et informatique : une collaboration fructueuse

La collaboration linguistique-informatique n'est pas nouvelle. Les techniques formelles développées en informatique théorique, en logique mathématique ou en intelligence artificielle ont très tôt suscité l'intérêt de linguistes et informaticiens soucieux de développer des outils informatiques pour le traitement automatique du langage naturel ayant des buts aussi divers que la traduction automatique, la vérification et l'explicitation de théories linguistiques, ou la simulation des capacités humaines de compréhension et de production du langage. De cette collaboration est née la linguistique informatique (*computational linguistics*) qui a entrepris de définir de véritables langages ou formalismes adaptés aux applications linguistiques, et qui s'est également donné pour tâche d'implémenter dans ces logiciels les théories et les descriptions linguistiques particulières. C'est ainsi que sont apparus deux types de formalismes, les formalismes outils développés dès le départ dans une perspective de traitement automatique du langage naturel, et les formalismes linguistiques, syntaxiques ou grammaticaux, élaborés en tant que théorie du savoir linguistique. La linguistique informatique a ainsi, en retour, influencé la linguistique théorique, censée désormais non plus fournir des modèles de représentation linguistiques abstraits et non vérifiables, mais des descriptions entièrement explicites et cohérentes du savoir linguistique. L'exemple le plus typique d'influences mutuelles entre recherches linguistiques, linguistiques-informatiques ou informatiques semble être reflété par le traitement des données lexicales.

En effet, tous les modèles linguistiques modernes accordent une place centrale au lexique. Délaissé en un premier temps par les modèles génératifs qui privilégiaient un système d'analyse stratificationnel basé sur la séparation des différents niveaux d'analyse morphologique, syntaxique et sémantique, il joue un rôle croissant dans les théories linguistiques plus récentes intégrant des données syntaxiques et sémantiques dans le lexique. D'une manière générale, le module d'analyse syntaxique se réduit et la composante lexicale s'alourdit. Le lexique, qui à l'origine présentait une structure minimale réduite à une simple liste d'entrées morphologiques, devient un objet d'analyse complexe, avec des relations internes relevant de différents niveaux d'analyse.

Cette évolution est très marquée dans le cadre des formalismes grammaticaux basés sur l'unification qui traitent du problème de l'encodage de l'information linguistique, syntaxique ou sémantique. Issus de la grammaire générative chomskienne, tous ces formalismes ont été à l'origine d'un mouvement de réhabilitation du lexique qui s'est étendu à de nombreuses théories linguistiques, dont la théorie chomskienne elle-même, qu'ils ont par la suite refusé de suivre dans ses récents développements. Si les premiers formalismes linguistiques tels que la Grammaire fonctionnelle lexicale (*Lexical Functional Grammar* ; abréviation : *LFG*) de Bresnan (1982) ou la Grammaire syntagmatique généralisée (*Generalized Phrase Structure Grammar* ; abréviation : *GPSG*) de Gazdar *et al.* (1985) attribuent une place importante au lexique et maintiennent le principe d'une syntaxe séparée de la sémantique, il n'en va pas de même pour des formalismes plus récents tels que la Grammaire syntagmatique guidée par la tête (*Head-driven Phrase Structure Grammar* ; abréviation : *HPSG*) de Pollard & Sag (1987, 1993), qui exploitent pleinement le pouvoir de l'unification et enrichissent les structures lexicales d'informations plutôt complexes relevant à la fois du niveau d'analyse phonologique, morphologique, syntaxique et sémantique.

On constate une évolution semblable dans le cadre des théories linguistiques visant à la fois une description exhaustive des possibilités du système linguistique et par là même, une amélioration des traitements automatisés du langage. L'ensemble des travaux du Laboratoire d'Automatique Documentaire et Linguistique (L.A.D.L.) de Paris entrepris pour l'élaboration d'un grand lexique-grammaire du français (Vivès 1988, Gross 1993) de même que la théorie Sens-Texte de Mel'chuk, ainsi que son application directe, le Dictionnaire explicatif et combinatoire (Mel'chuk 1984, 1988), peuvent être cités en référence pour illustrer la place centrale accordée aux données lexico-grammaticales dans des approches syntaxiques et sémantiques du langage.

Le projet d'élaboration de dictionnaires électroniques bilingues que nous décrivons ici s'insère dans cette tradition des lexiques-grammaires développés par le L.A.D.L. de par la place centrale accordée aux données lexico-grammaticales, de même que par son approche à la fois syntaxique, sémantique et pragmatique. Il s'en distingue cependant de par sa démarche particulière d'exploitation informatique de ses données linguistiques dans un logiciel de gestion de bases de données existant. Cette double démarche linguistique et informatique sera donc présentée ici pour illustrer toutes les phases d'analyse nécessaires à l'établissement de dictionnaires électroniques spécialisés, et orienter ce travail linguistique parmi les contributions modestes à la linguistique informatique.

3. Elaboration d'un lexique-grammaire du point de vue linguistique

3.1. Choix d'un domaine et constitution d'un corpus

Le choix du domaine d'étude ainsi que la constitution d'un corpus d'exemples-types d'emplois d'un champ lexical et sémantique sélectionné représentent l'étape préliminaire et indispensable au projet d'établissement d'un lexique-grammaire. Il n'y a en principe aucune limitation en ce qui concerne le choix d'un domaine. L'important est de choisir un ensemble qui présente une certaine cohésion conceptuelle. Il peut s'agir par exemple de domaines tels que les médias (presse, télévision, radio, édition...), les catastrophes naturelles (tremblements de terre, inondations, activité volcanique...), les spectacles (cinéma, production de films, concerts, théâtre...), les sports (cyclisme, tennis, athlétisme, sports d'équipe avec ballon...), la vie politique (élections, débats parlementaires, législation...), les conflits (guerres, réfugiés, interventions, négociations, racisme...), le social (pauvreté, soins médicaux, émigration, droit d'asile...), ou la santé (maladies infectieuses, drogues, médicaments, médecine...).

La constitution du corpus d'exemples-types d'emplois, indispensable à l'établissement de dictionnaires électroniques monolingues ou bilingues, se fait en rassemblant un ensemble de textes en allemand et éventuellement en français traitant du sujet choisi : ouvrages de caractère encyclopédique (manuels, dictionnaires encyclopédiques), presse générale ou spécialisée, CD-ROM, banques de données sur INTERNET, etc. Le choix de la forme du support du corpus dépend évidemment des conditions particulières de chaque étudiant et de sa localisation géographique pendant la réalisation du mémoire de maîtrise.

Nous illustrerons notre exposé par des exemples relevant du domaine des maladies infectieuses en général et du domaine des virus en particulier. Ces derniers exemples sont tirés du corpus de Bussienne (1996), constitué principalement à partir d'articles du journal *Tageszeitung (TAZ)* sur CD-ROM parus de 1988 à 1994.

3.2. Cadre théorique

Tout domaine de la réalité extralinguistique est organisé conceptuellement en fonction d'un ensemble de connaissances. Selon la nature du domaine, les concepts de base peuvent être exprimés par des prédicats d'action, d'état, de processus, de propriété. Ces prédicats s'associent à des arguments qui réfèrent plus ou moins directement aux participants de la réalité extralinguistique (von Polenz 1988 qualifie ces arguments de *Referenzstellen* ou *Bezugsstellen*), ou du moins à la représentation que nous nous en faisons : un prédicat de maladie (état ou processus) se combine par exemple avec des arguments qui représentent le patient (celui qui est affecté par la maladie), la spécification de l'affection (le type de maladie) et l'indication de la cause (comment, à la suite de quoi provient la maladie). Les prédicats liés au concept de maladie peuvent d'autre part, soit constituer le noyau sémantique des phrases-types d'emplois en tant que prédicats primaires ($PRED_{\text{primaire}}$), représentatifs du domaine de base et imposant une structure d'arguments ($ARG_{0..n}$, ARG_0 représentant une structure d'arguments non réalisée dans le cadre d'ellipses contextuelles) :

$PRED_{\text{primaire}}(ARG_{0..n}) : PRED_{\text{infection, infecter, s'infecter, infectieux, contaminer, contamination, contagieux...}}(ARG_{0..n})$

soit compter comme argument d'un autre prédicat, considéré comme secondaire car classifieur d'un sous-domaine du concept de base (PRED_{secondaire}), et pouvant hériter de la totalité ou d'une partie de la structure d'arguments existante (ARG_{0..n}PRED_{primaire}(ARG_{0..n}), ARG₀PRED_{primaire}(ARG_{0..n}) représentant un prédicat primaire élidé dans le cadre de la structure d'arguments du prédicat secondaire) :

PRED_{secondaire}(ARG_{0..n}PRED_{primaire}(ARG_{0..n})) : PRED_{secondaire}(PRED_{primaire}(ARG_{0..n})) | PRED_{secondaire / primaire}(ARG_{0..n})
 PRED_{secondaire}(PRED_{primaire}(ARG_{0..n})) : PRED_{diffusion de, transmission de, risque de, cas de...}(PRED_{infection}(ARG_{0..n}))
 PRED_{secondaire / primaire}(ARG_{0..n}) : PRED_{diffusion de, transmission de, risque de, cas de... / infection}(ARG_{0..n})

Elaborer un dictionnaire électronique de champs lexicaux et sémantiques sélectionnés, représentatifs d'une situation extralinguistique donnée, reviendra donc à recenser le vocabulaire spécifique à ce domaine et à décrire les propriétés combinatoires des expressions linguistiques désignant les différents objets extralinguistiques participant à cette situation. Si l'on reprend le concept de base des maladies infectieuses noté INFECT, représenté par un prédicat primaire ou une combinaison de prédicats primaires et secondaires (PRED_{INFECT} = PRED_{primaire} | PRED_{secondaire / primaire}) reliés aux trois arguments sémantiques cités précédemment :

ARG1 : **qui** est infecté ou contaminé
patients / personnes / malades
 ARG2 : **nature** de l'infection
le virus du sida / le virus de l'hépatite C / le VIH / le sous-type 0
 ARG3 : **origine** de l'infection ou de la contamination
à la suite de l'administration de Gammagard / d'origine vaccinale / via le lait maternel / après avoir utilisé des produits sanguins

les différentes combinaisons observées peuvent être maximales (saturation des positions d'arguments autour des prédicats ; exemples a.), insaturées (absence de certains arguments autour des prédicats ; exemples b. / c. / d. / e. / f. / g.), ou minimales (prédicats sans arguments : exemples h., ou réalisation de certains arguments en l'absence des prédicats : exemples i. / j.) :

- a. PRED_{INFECT}(ARG1, ARG2, ARG3) :
les malades infectés par le virus de l'hépatite C à la suite de l'administration de Gammagard / les personnes contaminées par le virus du sida à la suite de l'utilisation de produits sanguins / la transmission mère-enfant du virus du sida par le lait maternel
- b. PRED_{INFECT}(ARG1) :
contamination zairoise / infection des malades
- c. PRED_{INFECT}(ARG2) :
infection virale / infection par le sous-type 0
- d. PRED_{INFECT}(ARG3) :
cas de contamination à la suite de l'utilisation de produits sanguins
- e. PRED_{INFECT}(ARG1, ARG2) :
les malades infectés par le virus de l'hépatite C
- f. PRED_{INFECT}(ARG2, ARG3) :
le risque de contamination virale via le lait maternel
- g. PRED_{INFECT}(ARG1, ARG3) :
les malades infectés à la suite de l'administration de Gammagard
- h. PRED_{INFECT}() :
il y a des risques d'infection / des cas de contamination / deux cas suspects
- i. (ARG1, ARG2, ARG3) / (ARG1) / (ARG1, ARG2) / (ARG1, ARG3) :
malades (séropositifs) (à la suite de l'utilisation de produits sanguins)
- j. (ARG2) / (ARG2, ARG3) :
maladies présentant des caractéristiques virales / maladies d'origine vaccinale

La combinatoire extralinguistique globale du concept de base *maladies infectieuses* peut être schématisée selon des graphes orientés ou automates finis du type de ceux de la figure 1 et donnera lieu à l'établissement de grammaires locales morphosyntaxiques recensant toutes les variations lexico-syntaxiques du domaine :

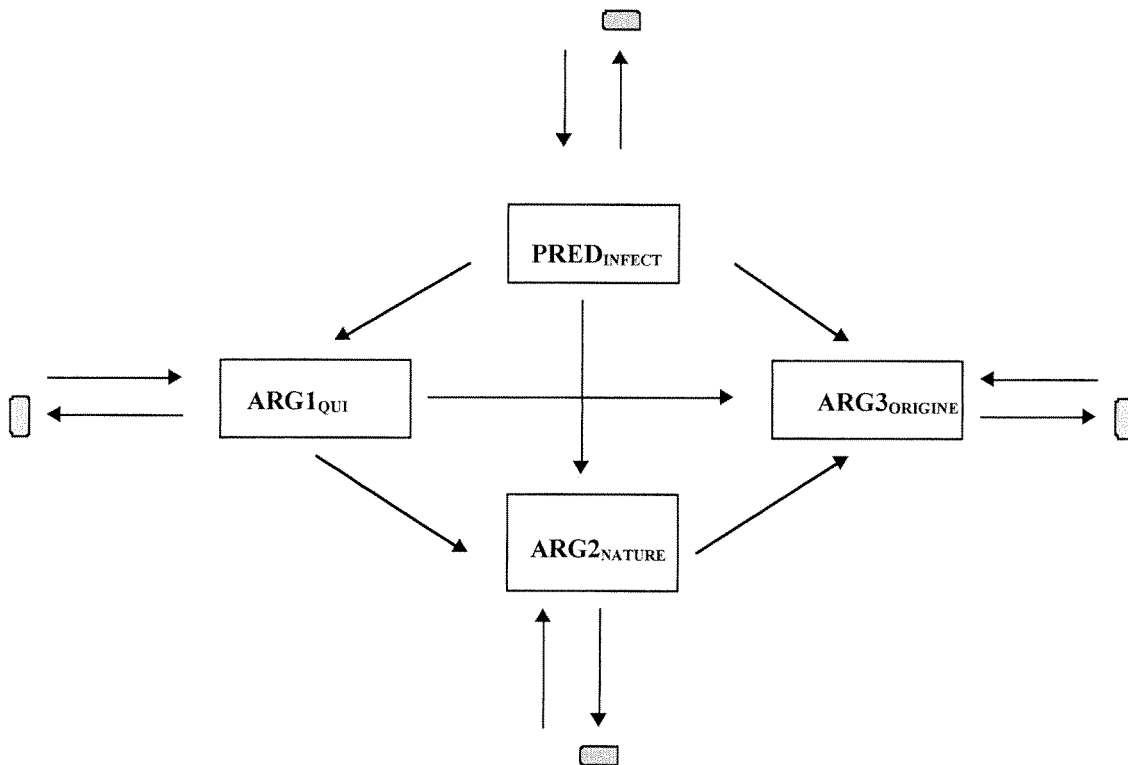


Figure 1. Schématisation de la combinatoire extralinguistique globale : concept maladies infectieuses

La figure 2 illustre, sous la forme d'un tableau organisé autour de la combinatoire extralinguistique $PRED_{INFECT}(ARG1_{QUI}, ARG2_{NATURE}, ARG3_{ORIGINE})$, un certain nombre de constructions toutes dépendantes du type sémantique et morphosyntaxique des prédicats représentatifs du concept de base.

En effet, l'ordre de surface et la forme syntaxique des arguments (adjectifs : A, noms / pronoms : N, compléments de noms : *deN*, compléments prépositionnels nominaux ou verbaux : *pN* / *pV*, indicés conformément à leur statut extralinguistique) pourra varier selon que le prédicat se présente sous la forme d'un participe passé ($VP2_{pred}$), d'une construction verbale active (V_{pred}) ou passive ($V-VP2_{pred}$) : exemples a., d'un adjectif (A_{pred} : exemple b.), d'une construction verbo-nominale ($V-N-N_{pred}$ / $V-N_{pred}$: exemples c.), d'un nom seul ou combiné avec un autre nom ou un adjectif (N_{pred} / $N-N_{pred}$ / $N-A_{pred}$: exemples d.), ou en l'absence de prédicat (exemples e.) :

- a. $N_1 VP2_{pred} (pN_2) (pN_3) | N_1 V-VP2_{pred} pN_2 pN_3 | N_1 / N_2 V_{pred} pV_3$
- b. $N_2 A_{pred}$
- c. $V-N-N_{pred} | V-N_{pred} pN_3$
- d. $N-A_{pred} | N_{pred} A_1 / A_2 | N_{pred} pN_2 | N_{pred} (N_1) (deN_2) (pN_3) | N-N_{pred} (A_2) (pN_3) | N-N_{pred} deN_2 pN_3$
- e. $N_2 deN_1 | N_1 (A_2) (pN_3) | N_2 (deN_3)$

PRED _{INFECT}	ARG1 _{QUI}	ARG2 _{NATURE}	ARG3 _{ORIGINE}
VP2	N	pN	pN
infecté _{VP2}	toutes les personnes _N		
infecté _{VP2}	certains patients / des malades _N	par le virus du sida _{pN}	
infecté _{VP2}	des patients _N	par le VIH _{pN}	
infecté _{VP2}	la région de Kikwit _N	par le virus Ebola _{pN}	
contaminé _{VP2}	un singe / plusieurs personnes _N		
contaminé _{VP2}	17 personnes _N	par la bactérie listeria monocytogenes _{pN}	
contaminé _{VP2}	des personnes _N	par le virus du sida _{pN}	à la suite de l'utilisation de produits sanguins _{pN}
V-VP2	N	pN	pN
être infecté _{V-VP2}	quelques patients _N	par le virus de l'hépatite C _{pN}	à la suite de l'administration de Gammagard _{pN}
être contaminé _{V-VP2}	plusieurs dizaines de personnes _N	par le virus de l'hépatite C _{pN}	à la suite d'injections d'immoglobulines Gammagard _{pN}
V	N	N	pV / pN
s'infecter _V	cette femme _N		après avoir manipulé du sang de cet animal _{pV}
se transmettre _V		le virus _N	par les fluides corporels _{pN}
A		N	
infectieux _A		maladies _N	
contagieux _A		germe _N	
V-N			pN
il y a contamination _{V-N}			par voie sexuelle _{pN}
V-N-N			
il y a des risques de contamination / d'infection _{V-N-N}			
N	A / N	A / pN / deN	pN
infection _N		virale _A	
infection _N		par le virus du sida / le sous-type O / les VIH 1 et 2 _{pN}	
contamination _N	zaïroise _A		
transmission _N	mère-enfant _N	du virus du sida _{deN}	
transmission _N		du virus du sida _{deN}	par l'allaitement maternel _{pN}
deux cas _N			
cas _N		d'hépatite C _{deN}	à la suite de l'utilisation de lots d'immoglobuline _{pN}
N-N		A / deN	pN
cas d'infection _{N-N}			
risque d'infection _{N-N}			
risque de contamination _{N-N}		virale _A	via le lait maternel _{pN}
risque de transmission _{N-N}		du virus du sida _{deN}	par le lait maternel _{pN}
N-A			
risque infectieux _{N-A}			
	deN / N	N / A	pN / deN
	des malades _{deN}	statut séropositif _N	
	patients / malades _N	(séropositifs) _A	(à la suite de l'utilisation de produits sanguins) _{pN}
		maladies présentant des caractéristiques virales _N	
		maladies _N	d'origine vaccinale _{deN}

Figure 2. Exemple de construction autour de la combinatoire extralinguistique : concept maladies infectieuses

3.3. Traitement du corpus

Le choix des textes représentatifs du domaine sélectionné étant fixé, de même que le cadre théorique de l'étude, il s'agira de constituer des fiches comportant les expressions récurrentes du domaine. Ce recueil de phrases-types d'emplois servira ensuite de support d'étude pour l'analyse syntaxique et sémantique des diverses unités lexicales, l'objectif visé étant de montrer comment fonctionne, pour un champ lexical et sémantique donné, le passage du niveau extralinguistique aux constructions syntaxiques. Pour ce faire, toutes les expressions doivent être étudiées selon leur combinatoire, en privilégiant les associations à deux ou plusieurs termes. Le concept de base *maladies infectieuses* présente par exemple des associations telles que :

<i>Epidemie, ausbrechen</i>	<i>die in Kikwit ausgebrochene Epidemie</i>
<i>Epidemie, Ausbruch</i>	<i>der Ausbruch einer tödlichen Epidemie</i>
<i>Epidemie, Ausmaß</i>	<i>das Ausmaß der Epidemie</i>
<i>Epidemie, Übertragung</i>	<i>die Übertragungswege der Epidemie</i>
<i>Infektion, auftauchen</i>	<i>plötzlich auftauchende Infektionen</i>
<i>Infektion, möglich sein</i>	<i>eine Infektion ist nur durch Kontakt mit Blut möglich</i>
<i>Krankheit, Endstadium</i>	<i>im Endstadium der Krankheit</i>
<i>Krankheit, kommen</i>	<i>wie war die Krankheit über die Menschen gekommen?</i>
<i>Krankheit, Verbreitung</i>	<i>die Verbreitung der Krankheit stoppen</i>
<i>Virus, Inkubationszeit</i>	<i>die Inkubationszeit des Virus</i>
<i>Virus, sich übertragen</i>	<i>das Virus überträgt sich durch Kontakt mit Körpersäften</i>
<i>Virus, tödlich wirken</i>	<i>das Virus wirkt in 90 Prozent aller Fälle tödlich</i>

Dans ces exemples, le travail de classification syntaxique n'a pas encore eu lieu ; il s'agit simplement d'un relevé des associations entre le concept de base choisi (*Epidemie, Infektion, Krankheit, Virus*) et d'autres termes (*ausbrechen, Ausbruch, Ausmaß, Übertragung, auftauchen, möglich sein, Endstadium, kommen, Verbreitung, Inkubationszeit, sich übertragen, tödlich wirken*). Cette étape est essentielle car elle sert à constituer le matériau qui permettra ensuite l'analyse linguistique et l'établissement d'une syntaxe de chacune des expressions de base. A partir de ces recueils d'expressions constitués empiriquement, il faudra ensuite établir le système de description syntaxique et sémantique rendant compte de leur construction et permettant de prévoir ou d'analyser d'autres expressions de ce domaine à partir des éléments de base. Ces règles feront intervenir la classe lexicale des unités linguistiques (verbes, noms, adjectifs) et la relation syntaxique entre les éléments.

4. Mise en place d'un lexique-grammaire sur support informatique

Nous présenterons ici les solutions adoptées pour le traitement électronique des données de nos corpus en vue de l'élaboration de lexiques-grammaires de champs lexicaux et sémantiques sélectionnés. Le choix d'un logiciel de gestion de bases de données tel que DBASE 5.0 pour Windows nous semblait le plus approprié pour l'exploitation de nos données rassemblées sous forme de fiches d'exemples-types d'emplois présentant un certain nombre de rubriques ou champs suivis de toute l'information linguistique et extralinguistique relative à ces emplois. De cette façon, nous avons toutes les garanties d'atteindre les objectifs visés, à savoir une liste indexée de toutes les descriptions sémantiques des structures argumentales autour des prédicats représentatifs des concepts de base ainsi que l'ensemble des réalisations syntaxiques et lexicales des différentes combinaisons extralinguistiques.

4.1. Structure de la base de données

Quel que soit le domaine d'analyse choisi, la base de données doit permettre de représenter les propriétés combinatoires de toute unité lexicale. Pour cette raison, toute base de données présentera une même structure de fiche comportant cinq parties consacrées respectivement au concept de base (CB), aux arguments de ce concept dont le nombre varie selon les domaines traités (ARG), aux modificateurs (MODIF), à la phrase représentative du domaine (EXPRESSION) ainsi qu'à sa source (REFERENCE).

Le libellé sémantico-pragmatique du concept de base (noté en majuscules) doit être considéré comme une métalangue représentative du domaine choisi. Lorsque le champ lexical et sémantique sélectionné est directement représenté par des prédicats constituant le noyau sémantique des phrases-types d'emplois, le libellé est alors un terme classifieur de ces prédicats et le concept de base est considéré comme primaire :

CB_{primaire} : INFECT
eine Infektion / Krankheit haben, erkranken an, infiziert sein / werden

CB_{primaire} : VIRUS
Viren sein

Dans le cas contraire, le libellé fait apparaître un deuxième terme classifieur d'un sous-domaine d'étude et de ce fait, le concept de base est considéré comme secondaire :

CB_{secondaire} : APPARITION/INFECT
auftauchende Infektionen, der Ausbruch einer Epidemie, eine Krankheit ist aufgetreten

CB_{secondaire} : APPARITION/VIRUS
das Hereinbrechen des Virus, das Virus taucht auf / hat sich eingeknistet / wird produziert

Selon la nature du domaine d'analyse, les concepts de base secondaires peuvent avoir un emploi plus large que les concepts de base primaires ; c'est le cas notamment du domaine des virus dont le terme classifieur VIRUS regroupe des unités linguistiques qui n'ont pas de contenu prédicatif et qui apparaissent majoritairement en tant qu'arguments de prédicats secondaires, les emplois prédicatifs de ce concept de base primaire étant réservés aux seules associations du type *Viren sein*.

Les réalisations lexico-syntaxiques d'un concept de base primaire ou secondaire peuvent alors se présenter sous la forme d'un prédicat actualisé, c'est-à-dire en emploi verbal ou en position prédicative : un verbe plein (CB_V), une association d'un nom prédicatif, précédé ou non d'une préposition, et d'un verbe support (CB_N + CB_VSUP), dans certains cas d'un nom et d'un verbe approprié (CB_N + CB_V), d'un nom, d'un adjectif ou d'un verbe en position attributive et d'un verbe support (CB_N / CB_A / CB_V + CB_VSUP) ; ou d'un prédicat non actualisé, c'est-à-dire en emploi non verbal ou en position non prédicative : un prédicat nominal ou adjectival seuls ou combinés (CB_N, CB_A, CB_N + CB_A). La notion de verbe support est ici une notion large qui regroupe à la fois les verbes supports de nominalisation, les verbes copules, les verbes factitifs / causatifs ou les semi-modaux (constructions CB_V + CB_VSUP du type : *sein / haben* + infinitif en *zu*). La notion d'adjectif est également une notion large qui regroupe les adjectifs primaires ou dérivés (participes passés/ présents) :

CB	CONSEQUENCE/VIRUS
CB V	<i>auslösen, angreifen, lahmlegen, verursachen, bewirken, entvölkern, erkranken...</i>
CB N+ CB VSUP	<i>zum Opfer gefallen sein, Träger sein, Erbrechen / Fieber und Schnupfen bescheren</i>
CB A+ CB VSUP	<i>infiziert sein, löchrig machen, tot umfallen</i>
CB N	<i>Infektion, Virusinfektion, Viruskrankheit, Auswirkungen...</i>
CB A	<i>verseucht, infiziert, betroffen, befallen, positiv, haltig...</i>

CB	LUTTE/VIRUS
CB V + CB VSUP	<i>zu kämpfen haben</i>

CB	MUTATION/VIRUS
CB N + CB V	<i>ihre Oberflächenstruktur ändern</i>

CB	TEST/VIRUS
CB N + CB A	<i>HIV-Test bei Spenderblut durchgeführt, alle vorgenommenen HIV-Tests</i>

CB	VIRUS
CB N + CB VSUP	<i>Viren sein</i>

Chaque argument ou modifieur associé à ce concept de base est analysé selon sa nature ou *définition* extralinguistique (ARG_DEF / MODIF_DEF), sa représentation *lexicale* (ARG_LEX / MODIF_LEX), *syntactique* (ARG_SYN / MODIF_SYN) et éventuellement sa particularité *sémantique* (ARG_SEM / MODIF_SEM). L'entrée lexicale suivante illustre une analyse globale d'un exemple-type d'emploi des conséquences liées au domaine d'étude *maladies infectieuses / virus* :

CB	CONSEQUENCE/VIRUS
CB A	<i>hingerafft</i>
ARG1 DEF	<i>cause</i>
ARG1 LEX	<i>durch ein angeblich mysteriöses Virus</i>
ARG1 SYN	<i>pN durch</i>
ARG2 DEF	<i>conséquence</i>
ARG2 LEX	<i>Robben</i>
ARG2 SYN	<i>Nnominatif</i>
ARG2 SEM	<i>victime animale</i>
MODIF DEF	<i>temps</i>
MODIF LEX	<i>vor 4 Jahren</i>
MODIF SYN	<i>pN vor</i>
MODIF SEM	<i>date</i>
EXPRESSION	<i>auch die vor 4 Jahren durch ein angeblich mysteriöses Virus hingerafften Robben</i>
REFERENCE	<i>TAZ 1994 N°4323 S.3</i>

et correspond à la structure de table suivante (VIRUS.DBF) :

NOM	TYPE	LARGEUR	DECIMALES	INDEX
CB	Caractère	22	0	Croissant
CB_N	Caractère	40	0	Aucun
CB_V	Caractère	22	0	Aucun
CB_A	Caractère	22	0	Aucun
CB_VSUP	Caractère	22	0	Aucun
ARG1_DEF	Caractère	25	0	Aucun
ARG1_LEX	Caractère	75	0	Aucun
ARG1_SYN	Caractère	16	0	Aucun
ARG1_SEM	Caractère	16	0	Aucun
ARG2_DEF	Caractère	25	0	Aucun
ARG2_LEX	Caractère	50	0	Aucun
ARG2_SYN	Caractère	16	0	Aucun
ARG2_SEM	Caractère	16	0	Aucun
ARG3_DEF	Caractère	25	0	Aucun
ARG3_LEX	Caractère	50	0	Aucun
ARG3_SYN	Caractère	16	0	Aucun
ARG3_SEM	Caractère	16	0	Aucun
MODIF_DEF	Caractère	25	0	Aucun
MODIF_LEX	Caractère	50	0	Aucun
MODIF_SYN	Caractère	16	0	Aucun
MODIF_SEM	Caractère	16	0	Aucun
EXPRESSION	Caractère	150	0	Aucun
REFERENCE	Caractère	36	0	Aucun

Un concept de base primaire ou secondaire peut être représenté par une unité linguistique de nature non prédicative, mais comportant une prédication à un second niveau. Ce cas se produit par exemple avec des noms d'agent dérivés d'une structure prédicative (*der Tote = derjenige, der tot ist ; der Infizierte = derjenige, der infiziert wurde / war / ist*), qui ne constituent pas le prédicat de la phrase-type d'emploi, mais qui permettent d'identifier le concept de base :

CB	CONSEQUENCE/VIRUS
CB N	<i>(Tote)</i>
ARG1 DEF	cause
ARG1 LEX	<i>durch neuen Virus</i>
ARG1 SYN	pN <i>durch</i>
ARG2 DEF	conséquence
ARG2 LEX	<i>drei Tote</i>
ARG2 SYN	Nnominatif
ARG2 SEM	victime humaine
EXPRESSION	<i>Drei Tote durch neuen Virus</i>

CB	CONSEQUENCE/VIRUS
CB N	<i>(Infizierte)</i>
ARG1 DEF	cause
ARG1 LEX	<i>HIV-</i>
ARG1 SYN	N-
ARG2 DEF	conséquence
ARG2 LEX	<i>350 Infizierte</i>
ARG2 SYN	Nnominatif
ARG2 SEM	victime humaine
EXPRESSION	<i>350 HIV-Infizierte</i>

Enfin, un concept de base peut être étudié de façon complémentaire selon ses relations privilégiées avec des modificateurs spécialisés, concrets ou abstraits, voire métaphoriques, spécifiques du domaine d'emploi du champ lexical et sémantique sélectionné (il s'agit dans ces derniers exemples d'emplois non prédicatifs du concept de base VIRUS) :

CB	VIRUS
CB N	<i>Iwans Virus, das böse Virus, ein heimtückisches Virus, Virus des Nationalismus...</i>
MODIF DEF	caractéristique concrète / abstraite, identification concrète / abstraite

4.2. Analyses combinatoires

Les fiches d'analyse présentées ici sous forme condensée constituent des grammaires locales en application directe de la théorie du lexique-grammaire dont le postulat fondamental localise les éléments de sens dans des phrases élémentaires et non pas dans les mots (GROSS 1993, 1995). Elles sont issues d'une base de données indexée sur le concept d'étude *maladies infectieuses / virus* et illustrent toutes les combinatoires extralinguistiques recensées pour ce concept. Elles présentent les libellés des concepts de base primaires ou secondaires, la réalisation lexicale et syntaxique des prédicats sous-jacents, ainsi que la combinatoire d'arguments sémantico-pragmatiques, voire de modificateurs, représentative du domaine ou sous-domaine d'étude. Le lexique-grammaire dont sont issues ces fiches d'analyse pourra fournir la liste des expressions du corpus correspondant à chacun de ces domaines ou sous-domaines d'étude, et donnera lieu à l'établissement de grammaires locales morphosyntaxiques recensant toutes les variations lexico-syntaxiques du domaine d'étude.

CB	ANEANTISSEMENT /VIRUS
CB V	<i>vernichtet werden, inaktiviert werden, inaktivieren, abgetötet werden, einkapseln, absterben</i>
CB V + CB VSUP	<i>beizukommen sein</i>
CB A + CB VSUP	<i>unschädlich gemacht werden, fertig werden, resistent gemacht werden</i>
ARG1	virus tué
ARG2	moyen
ARG3	qui tue
MODIFIEUR	lieu

CB	APPARITION /VIRUS
CB V	<i>festgestellt werden, vorkommen, sich einnisten, passieren, auftauchen, existieren, erzeugt werden, produziert werden, hergestellt werden, auftreten</i>
CB N	<i>Einpflanzen, Hereinbrechen</i>
ARG1	virus
ARG2	lieu
ARG3	date

CB	CONSEQUENCE /VIRUS
CB V	<i>sterben, auslösen, schmälern, angreifen, ausgelöst werden, fordern, verursachen, niederstrecken, bewirken, verschlimmern, provozieren, lahmliegen, nehmen, reißen, entvölkern, verursacht werden, begünstigen, beruhen, befallen sein, erkranken</i>
CB N	<i>Tote, Infizierte, Infekte, Infektion, Virusinfektion, Viruskrankheit, Auswirkungen</i>
CB A	<i>hingerafft, verseucht, infiziert, virusinfiziert, betroffen, befallen, positiv, haltig</i>
CB N + CB VSUP	<i>zum Opfer gefallen sein, Träger sein, Erbrechen / Fieber und Schnupfen bescheren</i>
CB A + CB VSUP	<i>infiziert sein, löchrig machen, tot umfallen</i>
ARG1	cause
ARG2	conséquence
MODIFIEURS	temps, lieu, particularité, intensité

CB	CONTAMINATION/ VIRUS
CB V	<i>infiziert werden, bekommen, sich zuziehen, weitergeben, angesteckt werden, sich infizieren, übertragen werden, sich anstecken, infizieren, übertragen, andocken, weitertragen, sich einnisten, sich holen</i>
CB N	<i>Infektionen, Infizierung, Übertragung, Virusübertragung, Ansteckung, Andocken, Kontakt, Transmissionswege</i>
CB A	<i>angesteckt, ansteckend</i>
CB N + CB VSUP	<i>seine Erbanlagen übertragen, in Kontakt geraten</i>
ARG1	patient infecté
ARG2	agent contaminant
ARG3	moyen
MODIFIEURS	temps, lieu, cause

CB	LOCALISATION/ VIRUS
CB V	<i>sitzen, herumschwimmen, orten, festmachen</i>
CB A + CB VSUP	<i>konzentriert sein</i>
ARG1	virus
ARG2	lieu
MODIFIEURS	manière, intensité, moyen

CB	LUTTE /VIRUS
CB V	<i>überleben, wirken, bekämpfen, bremsen, angreifen, sich wehren</i>
CB N	<i>Kampf, Aktion, Bekämpfung, Mittel, Heilmittel</i>
CB N + CB V	<i>Antikörper enthalten, Antikörper tragen</i>
CB V + CB VSUP	<i>zu kämpfen haben</i>
CB N + CB VSUP	<i>Gegenmittel geben, in Sicht sein</i>
ARG1	contre quoi
ARG2	qui combat
MODIFIEURS	manière, temps, moyen

CB	MUTATION / VIRUS
CB V	<i>gewinnen, sich verändern</i>
CB N	<i>Mutanten, Neubildung, Mutationen, Entwicklung</i>
CB A	<i>sich verändernd</i>
CB N + CB V	<i>ihre Oberflächenstruktur ändern</i>
CB N + CB VSUP	<i>Veränderungen durchmachen</i>
ARG1	virus
ARG2	but de l'évolution
MODIFIEURS	temps, caractéristique, intensité

CB	PREVENTION / VIRUS
CB V	<i>schützen</i>
CB N	<i>Inspektion, Super-Impfstoff, Hemmer AZT, Schutz, Schutzmaßnahmen, Impfung, Impfstoff</i>
ARG1	ce qui est protégé
ARG2	contre quoi
ARG3	agent protecteur
MODIFIEUR	temps

CB	PROPAGATION /VIRUS
CB V	<i>sich ausbreiten, eindringen, verbreitet werden, sich vermehren, sich verbreiten, sich verteilen, wuchern</i>
CB N	<i>Verbreitung, Ausbreitung, Verbreitungsgrad, Virusepidemie, Vermehrung</i>
CB A	<i>verbreitet</i>
CB A + CB VSUP	<i>verbreitet sein</i>
ARG1	virus
ARG2	lieu
MODIFIEURS	intensité, temps, recherche, moyen

CB	PROVENANCE/ VIRUS
CB V	<i>herkommen, stammen, kommen, entstehen</i>
CB N	<i>Entstehung</i>
ARG1	virus
ARG2	provenance
MODIFIEUR	temps

CB	RECHERCHE / VIRUS
CB V	<i>entdecken, nachweisen, analysiert sein, gesehen werden, isoliert werden, sich aufspüren lassen, sehen</i>
CB N	<i>Forschung, Virusforschung, Erforschung, Entdeckung, Nachweis, Arbeit</i>
CB V + CB VSUP	<i>nachzuweisen sein</i>
ARG1	agent chercheur
ARG2	objet de la recherche
ARG3	lieu
MODIFIEUR	manière

CB	RISQUE /VIRUS
CB V	<i>bedrohen</i>
CB N	<i>Anfälligkeit, Gefahr, Risiko</i>
CB A	<i>gefährdet</i>
CB N + CB VSUP	<i>ein Risiko darstellen</i>
CB A + CB VSUP	<i>ausgesetzt sein, gefährdet sich fühlen</i>
ARG1	virus dangereux
ARG2	ce qui est en danger
MODIFIEUR	intensité

CB	TEST /VIRUS
CB V	<i>untersucht werden, untersucht sein, überprüft werden, getestet werden, testen, sich testen lassen, testen lassen, untersuchen</i>
CB N	<i>Test, Aidstest, Antikörpertest</i>
CB A	<i>getestet, untersucht</i>
CB N + CB A	<i>Tests vorgenommen, Test durchgeführt</i>
CB N + CB VSUP	<i>einen Test machen lassen, einen Antikörpertest machen, zu einem Test gebeten werden, den Test durchführen, Tests durchgeführt werden, mit Tests untersucht werden, zum Antikörpertest drängen, Untersuchungen vornehmen</i>
ARG1	ce qui est testé
ARG2	qui teste
ARG3	sur quoi
MODIFIEURS	manière, temps, date, lieu

CB	VIRUS
CB N	<i>Virus</i>
CB N + CB VSUP	<i>Viren sein</i>
ARG1	virus identifizé
MODIFIEURS	caractéristique concrète / abstraite, identification concrète / abstraite

A titre d'exemple, nous présenterons dans la figure 3 une grammaire locale du concept de base CONTAMINATION / VIRUS organisée autour de la combinatoire extralinguistique $PRED_{CONTAMINATION/VIRUS}$ ($ARG1_{PATIENT\ INFECTE}$, $ARG2_{AGENT\ CONTAMINANT}$, $ARG3_{MOYEN}$) :

$PRED_{CONTAMINATION/VIRUS}$	$ARG1_{PATIENT\ INFECTE}$	$ARG2_{AGENT\ CONTAMINANT}$	$ARG3_{MOYEN}$
N		N-	
HIV-Ansteckung		HIV-	
HIV-Infizierung		HIV-	
HIV-Übertragung		HIV-	
Virusübertragung		Virus	
N		N-	pN durch
HIV-Infektionen		HIV-	durch Blut und Blutprodukte
Infizierung		HIV-	durch HIV-verseuchtes Blut
HIV-Übertragungen		HIV-	durch Blutprodukte
N		Ngénitif	
Andocken		des HI-Virus	
Übertragung		des Humanen Immundefekt Virus	
Transmissionswege		des HIV-Virus	
N		pN mit	
Ansteckung		mit dem Virus	
Infizierung		mit dem HIV-Virus	
Kontakt		mit dem Virus	
A		Nnominatif	
ansteckend		der Virus	
A		pN mit	
angesteckt		mit dem Aids-Virus	
V	Nnominatif	pN an, mit	
sich anstecken	so viele junge Menschen	am Aids-Virus	
sich anstecken	mehr als 2000 Empfänger von Blut und Blutprodukten	mit dem Aids-Virus HIV	
sich infizieren	er	mit dem Aids-Virus	
sich infizieren	der Junkie	mit dem HIV-Virus	
infiziert werden	drei Patienten	mit dem HIV-Virus	
V	Nnominatif	pN mit	pN bei, durch, über
sich infizieren	immer mehr Menschen	mit dem tödlichen Aids-Virus	bei heterosexuellen Kontakten
sich anstecken	4000 bis 5000 Bundesbürgerinnen	mit dem HIV-Virus	bei ungeschütztem Geschlechtsverkehr
angesteckt werden	Aids-Infizierte	mit dem HIV-Virus	durch Bluttransfusionen
angesteckt werden	rund 1500 Bluter	mit dem Immunschwächevirus	durch verseuchte Blutgerinnungspräparate
infiziert werden	sie	mit dem HI-Virus	über ein Blutpräparat oder eine Transfusion
V	Naccusatif	Nnominatif	
infizieren	mehr als 40 verschiedene Gattungen von Raupen	der Virus Autographa Californica	

V	pN an, in, von..zu	Nnominatif	
andocken	an gesunde Zellen	das Virus	
sich einnisten	ins Gehirn	der Virus	
übertragen werden	von Tier zu Tier	der Virus	
V		Nnominatif	pN durch, über, via
übertragen werden		der Virus	durch heterosexuellen Geschlechtsverkehr
übertragen werden		das Virus	über Geschlechtsverkehr
übertragen werden		der Virus	über das Trinkwasser
übertragen werden		der Virus	über Betreuungspersonen
übertragen werden		das HIV-Virus	via Serum
V		Naccusatif	Nnominatif
übertragen		einen bislang unerforschten Virus	die weiße Treibhausfliege
weitertragen		den Virus	Transportfahrzeuge
V		pN mit	pN bei
sich infizieren		mit dem HI-Virus	bei einer Bluttransfusion
V	Nnominatif	N-	pN durch
infiziert werden	rund 80 Prozent der Bluter-Kranken	HIV-	durch HIV-verseuchtes Blut
V	Nnominatif	Naccusatif	pN bei, durch
sich holen	sie	das Virus	bei vier Bluttransfusionen
bekommen	meine Frau	den HIV-Virus	durch mich
sich zuziehen	Brasilianerinnen	den Virus	durch heterosexuellen Geschlechtsverkehr
sich zuziehen	25 Babys	den HIV-Virus	durch eine Bluttransfusion
V	Naccusatif	pN mit	Nnominatif
infizieren	die Testpersonen	mit dem HI-Virus	die Wissenschaftler
V	Ndatif	Naccusatif	Nnominatif
weitergeben	ihr	den Virus	ein Geliebter
N-V	pN auf	Nnominatif	
seine Erbanlagen übertragen	auf die Bakterie	dieser Virus	
pN-V	Nnominatif	pN mit	
in Kontakt geraten	sie	mit dem Aids-Virus	

Figure 3. Une grammaire locale du concept de base CONTAMINATION / VIRUS

5. Conclusion

L'objectif de notre étude était de présenter un exemple d'exploitation linguistique de l'outil informatique illustré par un projet d'élaboration de dictionnaires électroniques bilingues spécialisés mené dans le cadre de maîtrises en linguistique allemande à l'université de Provence. Nous espérons avoir démontré l'intérêt de l'utilisation d'un logiciel de gestion de bases de données tel que DBASE 5.0 pour nos descriptions syntaxiques et sémantiques basées sur la notion de prédicat sémantico-pragmatique. La structure des bases de données semble en effet être parfaitement adaptée à l'encodage de l'information linguistique associée aux lexiques-grammaires et rend possible la mise en relation de phrases d'une même langue, de même que l'établissement de réseaux sémantiques relatifs à un domaine extralinguistique et encyclopédique. Cette dimension monolingue de notre exploitation linguistique de l'outil informatique ne doit pas être considérée comme une vue exclusive d'une base de données, mais plutôt comme une étape préliminaire à notre projet d'élaboration de dictionnaires bilingues, voire même multilingues, pour des domaines délimités. En effet, la comparaison de phrases parfaitement définies syntaxiquement et sémantiquement ne doit pas s'arrêter aux limites d'une seule langue. De même que certains modèles de traduction automatique mettent en jeu une interface sémantique ou langue pivot compatible avec toutes les langues traitées lors du passage de la langue source à la langue cible, nous nous fixons pour objectif futur d'exploiter à nouveau l'outil informatique dans le but de mettre en relation les phrases de différentes langues au moyen d'une langue pivot située au niveau de la combinatoire sémantico-pragmatique.

6. Bibliographie

BALIBAR-MRABTI Antoinette, 1995. « Une étude de la combinatoire des noms de sentiment dans une grammaire locale », in *Langue Française* 105, Larousse, Paris, pp. 88-97.

BRESNAN Joan (éd.), 1982. *The Mental Representation of Grammatical Relations*, Cambridge, MIT Press, Massachusetts.

BRESSON Daniel, 1987. « Etude des propriétés syntaxiques d'expressions verbales psychologiques de l'allemand », in *Cahier d'Etudes Germaniques* 12, Aix-en-Provence, pp. 123-143.

BRESSON Daniel & DOBROVOL'SKIJ Dmitrij, 1998. « Semantik und Syntax der ANGST-Ausdrücke. Versuch einer integrativen mehrsprachlichen Analyse », in *Abstrakte Nomina, Studien zur deutschen Sprache* 10, Bresson / Kubczak (éd.), Gunter Narr Verlag Tübingen, pp. 163-186.

BUSSIENNE Armelle, 1996. *Elaboration d'un dictionnaire syntaxique allemand dans le domaine : virus*, Mémoire de Maîtrise d'allemand, Université de Provence, Aix-en-Provence.

GAZDAR Gerald & al., 1985. *Generalized Phrase Structure Grammar*, Basil Blackwell, Oxford.

GROSS Maurice, 1993. « Lexicon Based Algorithms for the Automatic Analysis of Natural Language », in *Theorie und Praxis des Lexikons*, Beckmann / Heyer (éd.), Walter de Gruyter, Berlin / New York, pp. 218-236.

GROSS Maurice, 1995. « Une grammaire locale de l'expression des sentiments », in *Langue Française* 105, Larousse, Paris, pp. 70-87.

MEL'CHUK Igor & al., 1984. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I*, Les Presses de l'Université de Montréal, Montréal.

MEL'CHUK Igor, 1988. « Paraphrase et lexique dans la théorie linguistique Sens-Texte », in *Lexique* 6, Presses Universitaires de Lille, pp. 13-54.

MESLI Nadia & BRESSON Daniel, 1992. « Nominale Prädikate und Funktionsverben bei Martin Luther am Beispiel des lexikalischen Bereichs der Gefühle und der zwischenmenschlichen Beziehungen », in *Cahier d'Etudes Germaniques* 23, Aix-en-Provence, pp. 75-99.

MESLI Nadia, 1998. « Dictionnaires électroniques bilingues de champs lexicaux et sémantiques sélectionnés », in *Cahier d'Etudes Germaniques* 35, Aix-en-Provence, pp. 29-50.

von POLENZ Peter, 1988. *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens*, Walter de Gruyter, Berlin / New York.

POLLARD Carl & SAG Ivan, 1987. *An Information-Based Theory of Syntax and Semantics. Volume One : Fundamentals*, University of Chicago Press, Chicago.

POLLARD Carl & SAG Ivan, 1993. *Head-Driven Phrase Structure Grammar*, University of Chicago Press, Chicago.

VIVÈS Robert, 1988. « Lexique-grammaire, nominalisations et paraphrases », in *Lexique* 6, Presses Universitaires de Lille, pp. 139-156.

VOUZZAVEDIBIZAR

Petit lexique à usage de littérature politique

Edmond Bianco

En quelques dizaines d'années, si la langue Française n'a pas sensiblement changé, l'usage des mots, lui, a subi une forte évolution. Si je m'étais permis, alors modeste élève de sixième dans les années quarante, de rendre une rédaction, comme le font les collégiens à l'heure actuelle, le jour des résultats de la composition, le professeur arborant un air dégoûté aurait présenté ma copie tenue dédaigneusement par un angle entre le pouce et l'index, comme un objet très sale, en éructant une formule vexante pour en exprimer une violente aversion.

Des mots du langage courant, considérés voila encore fort peu comme des termes très grossiers, sont devenus très usités dans le domaine de l'écriture demeuré longtemps rigoureusement expurgé. Peut être doit-on cette évolution à Céline qui fut le premier à oser transcrire une sorte de langage parlé et à lui attribuer un titre de noblesse. Les "gros mots" ne tardèrent pas à suivre. De telle manière que non seulement en littérature, mais encore dans la presse, à la radio, à la télévision, des mots considérés jadis, c'est-à-dire il y a seulement quelques années, comme parfaitement orduriers et interdits de séjour dans les lieux où coule la culture, ont non seulement conquis droit de cité, mais encore deviennent la marque d'un esprit fort.

Il m'apparaît en conséquence à la fois utile et indispensable, pour la commodité de l'expression, au moins dans un domaine restreint, de se donner quelques définitions précises destinées à argumenter solidement sans se perdre à chaque fois dans des détails alourdissants. En effet et par exemple, on se préoccupe de sujets qui touchent à la politique, l'actualité met en scène des personnages quelque fois hauts en couleur, quelque fois peu brillants, mais qui toujours ont des attitudes que des expressions fortes du langage courant caractérisent en toute clarté. Pourquoi se priver d'une telle efficacité? C'est bien la raison pour laquelle je propose le petit lexique suivant dont on pourra à titre de commodité ne retenir que les clefs d'accès.

C1) Connerie.

Désigne une attitude généralement choquante, soit par excès de malice soit par défaut de malignité.

C2) Gros con.

Personnage qui, dans l'expression de sa connerie, prend une attitude largement éléphanterque, mais en général pas par excès de malice.

C3) Petit con.

Se dit d'un personnage dont la connerie est nettement agressive. Un excès de malice peut en affiner la qualité.

C4) Petit con de la dernière pluie.

Parfaitement défini par la chanson de Brassens. Il s'agit d'un débutant qui est convaincu d'avoir inventé cette précieuse compétence.

C5) Vieux con.

Ce n'est pas une question d'âge, mais plutôt de durée dans la pratique. Le vieux con a pris le temps qu'il faut pour se sentir parfaitement à l'aise dans son état. Certains cas particulièrement graves peuvent même en concevoir une certaine fierté.

C6) Sale con.

Aucune lessive, si blanchissante soit-elle ne saurait le récurer. Non seulement ce personnage déborde de malice, mais encore plus de malignité agressive, il est dangereux.

C7) Vrai con.

L'adjectif marque surtout l'effet de surprise du partenaire qui découvre la profondeur de l'atteinte de son vis-à-vis.

C8) Con de première.

Le talent du sujet est visiblement réjouissant, le personnage montre une aptitude particulièrement brillante.

C9) Con fini.

N'a pas forcément de rapport avec le chef du parti fasciste italien, le participe passé marque plutôt la complétude de la propriété, il marque également le côté irrécupérable de l'individu.

C10) Le con de service.

Ne désigne plus un personnage, mais une fonction, beaucoup de gens peuvent à un moment ou à un autre endosser l'uniforme de cette digne fonction. Même quelque fois à leur insu.

C11) Complètement con.

Marque l'avortement d'une vaine tentative de récupération désespérée.

C12) Pauvre con.

Caractérise l'amateur éclairé qui en rajoute. Contient une nette touche de mépris envers une victime accablée d'un tel fardeau.

C13) Fieffé con.

Une nuance de noblesse souligne l'ardeur que porte le sujet en son sein.

C14) Con féal.

N'existe pas dans l'usage courant et c'est bien dommage.

Jusque là ce sont des désignations d'acteurs, il existe aussi des actes:

C15) Jouer au con.

Attitude maligne destinée à mettre l'adversaire en porte à faux en jouant une partition encore plus stupide ou ignoble que la sienne.

C16) Passer pour un con.

Avoir une réaction malheureuse qui peut laisser à penser qu'on peut être caractérisé par l'un des cas C1 à C14. On peut également être la victime d'une entreprise particulièrement vicieuse qui organise un piège.

Bien entendu, ce modeste petit lexique ne saurait prétendre à l'exhaustivité, et je fais aussi confiance à l'inventivité de mes contemporains pour en enrichir le vocabulaire au cours du temps. Un petit exemple pour en roder l'usage ne peut pas nuire.

Lors de son allocution du 14 juillet 2001, notre "C2" de Savate Premier a dû employer quelques contorsions de vocabulaire pour essayer de faire avaler à ces "C12"s de Français, toutes ces histoires malodorantes de paiements en argent, bien que l'on ait cru depuis longtemps que celui-ci n'avait pas d'odeur. Se comportant comme un "C13", Savate Premier a dû même souffler un "pschitt" sonore pour en écarter le remugle, ce qui confirmait le nauséabond de l'affaire. Le Grand Vizir Isnosmile, qui brigue la place de Calife à la place du Calife, a une forte tendance à "C15", faisant ce que le Calife n'oserait pas faire, en fait pour essayer de le faire "C16". Heureusement il n'est pas nécessaire de pousser beaucoup le Calife.

Edmond Bianco

