

Qu'est-ce qu'une linguistique de la dénomination, de la référence et de l'usage?

PIERRE FRATH

Université de Reims Champagne-Ardenne

pierre.frath@univ-reims.fr

Résumé

Il est remarquable que la plupart des doctrines linguistique actuelles sont très platoniciennes, le plus souvent à l'insu de leurs protagonistes. Elles font un usage massif de principes, de règles, de fonctions, de relations, de primitives, d'universaux, de concepts, de sèmes, et autres entités de ce type, qui n'ont d'autre existence que d'être les causes « profondes » supposées de tel ou tel phénomène « de surface ». Nous argumentons ici que la linguistique gagnerait beaucoup à abandonner la métaphysique au profit de l'observation de l'usage, notamment à travers les corpus électroniques, dans une optique dénominative et référentielle.

Trois approches de la théorie en linguistique

Qu'est-ce qu'une linguistique de la dénomination, de la référence et de l'usage ? Avant de répondre à cette question, il vaut mieux peut-être commencer par dire ce qu'elle *n'est pas*. Cette partie sera suivie d'une autre plus constructive, dans laquelle on tentera de présenter les avantages de ce point de vue pour notre discipline.

Le localisme

On peut sans doute caractériser les travaux de linguistique de trois manières, selon leur usage de la théorie, à savoir l'**écléctisme**, la **modélisation**, et l'**indifférence**. La première mène à ce qu'on peut appeler une **linguistique localiste de l'explication**. Le linguiste ne se revendique d'aucune doctrine en particulier et il s'arroge le droit de piocher où bon lui semble parmi les nombreux points de vue à sa disposition. De tels travaux sont dès lors libérés de toute contrainte doctrinale et ils peuvent se développer dans toutes les directions, au gré des problèmes qui se posent ou des nouvelles idées qui peuvent surgir. La compréhension du langage se fait par l'accumulation de résultats divers et variés, mais sans qu'une vision d'ensemble ne s'en dégage explicitement. La qualité de l'explication linguistique locale est alors payée par le flou philosophique de la vue d'ensemble. Mais cela ne gêne pas les localistes, car une vision globale unique de la langue et de la pensée leur semble de toute façon inaccessible, ou alors sujette à caution.

Une linguistique des modèles

L'attitude localiste trouve en effet sa source dans les excès de la théorisation qui a dominé la linguistique du 20^e siècle, et qui la domine encore dans une certaine mesure, notamment en syntaxe et en sémantique. Appelons cette linguistique une **linguistique des modèles**. Elle n'est pas sans liens avec la pensée de Descartes. En effet, cet auteur a voulu construire une compréhension du monde plus cohérente que celle de la physique médiévale, pour laquelle chaque phénomène était explicable par une cause *ad hoc* particulière, de nature platonicienne. Il a ainsi montré que la logique, les mathématiques et la géométrie étaient en mesure de rendre compte de la réalité de manière tout à fait satisfaisante. Dès lors, plus besoin d'une multiplicité de causes isolées : il suffit qu'on puisse décrire les phénomènes examinés dans un langage logique ou mathématique. Ce point de vue novateur a permis l'explosion de la recherche scientifique à partir de la Renaissance. Les linguistiques des modèles se réfèrent à cet héritage cartésien plus ou moins explicitement. Elles construisent des ontologies exprimées dans des formalismes mathématiques et logiques, auxquelles elles tentent d'accrocher les observations linguistiques. Si elles y parviennent, ces observations sont réputées expliquées; sinon, on peut soit modifier le modèle, soit l'abandonner complètement pour en construire un autre.

Ce point de vue présente cependant un certain nombre d'inconvénients. D'abord, une grande partie de l'énergie des linguistes est dépensée à résoudre des problèmes liés à la théorie. Ainsi, on peut dire que le Lexique Génératif (LG) de James Pustejovsky¹ est une tentative pour résoudre les problèmes posés par les théories syntaxiques issues de la grammaire générative. Ces théories se passeraient volontiers du lexique et préféreraient n'avoir à traiter que des universaux et des principes d'association. D'ailleurs, certaines d'entre elles² installent une atomisation du lexique, où chaque mot est subdivisé en autant d'entités référentielles univoques qu'il comprend de sens. Le LG s'insurge contre cette attitude et redonne une existence théorique au lexique en l'imposant entre le niveau des règles syntaxiques et celui des *qualia*, une structure en quatre éléments (formel, constitutif, télique et agentif) qui regroupent une liste ouverte de concepts élémentaires. Il donne ainsi un coup d'arrêt à l'impérialisme de la syntaxe, dont les analyses s'arrêtent alors au niveau des mots, et il redonne toute son importance au lexique. Mais le LG génère lui aussi des difficultés, ce qui fait qu'il est en constante mutation. On finit alors par s'intéresser plus au modèle qu'à la langue. Par surcroît, l'adoption d'une linguistique des modèles n'est pas sans conséquences sur l'idée qu'on est alors amené à se faire de l'activité langagière, à savoir celle d'une parole qui serait le produit de calculs toujours renouvelés à chaque nouvelle énonciation, à partir de règles qui combinent des informations lexicales. On verra dans la suite du texte que ce point de vue mène à des impasses.

1. Voir par exemple Pustejovsky (1993) et (1995).

2. Par exemple Fodor et Lepore (1998).

Ensuite, une telle linguistique formelle est malgré tout très métaphysique. Elle repose sur l'idée cartésienne que la logique et les mathématiques sont les langages de la nature, et donc aussi ceux de la pensée et du langage. Une description formelle, mathématique ou logique, est ainsi *ipso facto* vraie, et fiable en principe à toutes les autres qu'on a pu faire de cette manière. C'est manifestement un *a priori* ontologique extrêmement lourd, dont on aimerait qu'il soit revendiqué explicitement, et non abandonné implicitement au sens commun.

Egalement, ces linguistiques formelles se considèrent comme profondément naturalistes et rejettent avec force le dualisme cartésien du corps et de l'esprit. Cependant, elles sont malgré tout obligées d'attribuer à l'esprit une existence séparée, car si notre fonctionnement mental et linguistique n'est qu'un déroulement de procédures formelles, une sorte de mécanisme donc, alors qui appuie sur le bouton? C'est la problématique bien connue de l'homoncule cartésien aux commandes du cerveau, celle du fantôme dans la machine de Gilbert Ryle (1947). C'est aussi l'argument de Ronald Searle (1980) dans sa parabole de la *Chambre chinoise*. Ces théories formelles sont en fait profondément métaphysiques, le plus souvent à l'insu de leurs protagonistes³.

Une linguistique de l'accumulation

Il existe une troisième grande famille de pratiques, celles qu'on peut appeler les **linguistiques de l'accumulation**. Ici l'attitude face à la théorie est essentiellement **l'indifférence**. Ce point de vue est à l'œuvre dans la compilation de dictionnaires, de listes d'expressions idiomatiques, de terminologies, de descriptions syntagmatiques exhaustives, etc. La linguistique de corpus est un bon exemple de linguistique de l'accumulation. Tout praticien remarque vite que les corpus sont d'immenses réservoirs de faits de langue qui ont jusqu'ici été en grande partie négligés, car ils sont peu accessibles à l'intuition seule. Il y a là un gisement de faits qui débouche naturellement sur la compilation. La réalisation par l'INaLF du *Trésor de la Langue Française* à partir de la base de données textuelles *Frantext* en est un bon exemple, ainsi que celle du dictionnaire *Collins-Cobuild*, entamée par John Sinclair et son équipe de l'Université de Birmingham deux décennies plus tard. Il y a là une attitude face à l'objet langue assez proche de celle des sciences exactes à leur début, à savoir observation, description et compilation.

Les corpus fournissent des données textuelles réelles, qui peuvent parfois être trompeuses ou incomplètes, mais qui permettent de résoudre certains problèmes. Pour ne donner qu'un exemple, une étude en corpus du verbe *commencer*⁴ a révélé que les structures en *commencer + N non procédural*, comme *elle a commencé un roman*, sont presque inexistantes en corpus, et qu'en conséquence les théories qui les ont mises sur un pied d'égalité avec la structure *commencer + infinitif*, comme *elle a commencé à lire un livre*, analysent une réalité tronquée.

3. Pour une critique en règle du cognitivisme, voir Frath 2007

4. Voir Frath (2002)

Cependant, les linguistiques de l'accumulation ne fournissent pas de vision globale de la langue. Observer est une chose, comprendre et interpréter en est une autre. Pour cela, il faut malgré tout une théorie, c'est-à-dire une vision d'ensemble qui produise un lien entre les observations, et qui leur donne sens.

Définition

Qu'est-ce donc, finalement, qu'une linguistique de la dénomination, de la référence et de l'usage, du moins telle que nous la concevons ? Elle se caractérise par *une ambition explicative générale*, fondée sur *l'adoption d'un point de vue qui soit aussi peu métaphysique que possible*, et par *une méthodologie descriptive qui utilise les corpus*. Elle considère que la parole n'est pas produite par un calcul cent fois, mille fois renouvelé, mais qu'au contraire le langage est une habitude caractéristique de notre espèce, qui fonctionne par construction et extrapolation à partir de l'usage. **L'usage** doit ainsi constituer le sujet d'étude principal d'une telle linguistique, d'où l'accent mis sur les corpus. Ensuite, elle considère que si nous parlons, c'est essentiellement pour parler *de* quelque chose, c'est-à-dire de notre expérience commune et individuelle. **La référence** est ainsi un point de vue central pour l'étude du langage⁵. Il ne s'agit pas de la référence logiciste qui prévaut dans la philosophie analytique, mais d'une conception sémiotique du signe qui trouve son origine chez les nominalistes médiévaux et dans la conception triadique du signe de C. S. Peirce⁶. Cet auteur distingue la **dénomination** (qu'il appelle *representamen*), opaque et arbitraire, qui dénomme des objets de notre expérience de manière globale, par exemple *chat*, *psychanalyse*, ou *verre à vin*. Il distingue aussi l'**objet**, par exemple les référents chat, psychanalyse ou verre à vin, ainsi que les **interprétants**, c'est-à-dire les signes discursifs complexes à propos d'un ou plusieurs objets ou dénominations. La dénomination étant un objet social, le point de vue est d'emblée *communautaire* et *externaliste* là où les linguistiques cartésiennes sont *individuelles* et *internalistes*.

Un exemple dans le domaine de la phraséologie : le fusil de chasse et l'outil à travailler

Pour montrer la validité du point de vue défendu ici, la suite de l'article sera consacrée à quelques exemples dans le domaine de la phraséologie, un domaine où règne une pléthore déroutante de termes. Cela est dû notamment à la diversité doctrinale de leurs origines. Par exemple, le **synthème** provient de la tradition fonctionnaliste, la **collocation** de la linguistique de corpus, la **cooccurrence** de la linguistique quantitative, la **locution** de la tradition grammaticale scolaire, etc. Il y

5. Pour une analyse de la notion de référence en linguistique, voir Frath (2005a) et Frath (2005b)

6. Peirce (1978)

a aussi la **synapsie**⁷, l'**expression idiomatique**, le **nom composé**, l'**expression toute faite**, et ainsi de suite. Les objets auxquels ces entités réfèrent peuvent se recouper, par exemple le syntème désigne aussi des mots dérivés comme *désirable* ou *refaire*. Certains termes couvrent des champs très étroitement définis, par exemple le gallicisme. D'autres, comme la collocation ou la cooccurrence, sont très généraux (trop sans doute) et recouvrent *tous* les phénomènes.

Les études sur les unités phraséologiques (UP) peuvent adopter l'un ou l'autre des points de vue mentionnés ci-dessus, localiste, formel ou accumulatif. Par exemple, Catherine Paulin (2002) décrit comment les contacts de langue dans un environnement multi-ethnique tel qu'on peut en trouver en Grande-Bretagne peuvent générer des unités phraséologiques diverses et variées. Anita Nascicione (2002) décrit la formation d'UP en anglais, et son travail a débouché sur un ouvrage qui en recense de très nombreuses, ce qui montre que le localisme peut générer l'exhaustivité et produire un travail accumulatif. Citons aussi Gaston Gross, qui liste de grands nombres d'UP classées selon divers critères, notamment dans son ouvrage de 1996.

Limites de l'explication cognitive

Pour illustrer le point de vue formel sur les UP, nous reprenons des exemples développés dans Christian Bassac (2006 : 56-57). Selon cet auteur, qui se place dans le cadre du Lexique Génératif, le N1 dans *N1 de N2* posséderait une variable télique de type *activité* que certains N2 peuvent saturer, ce qui autorise des constructions comme *fusil de chasse*, *chaussures de marche*, *poignée de maintien*. Si *fer de repassage* n'est pas bon, c'est que « l'événement représenté par le nom ne renvoie pas à un événement de type activité », d'après Bassac. Dans *N à V_{inf}*, « le verbe à l'infinitif est une transition » entre « les deux sous-événements dont elle est constituée », ce qui donne des composés comme *fer à repasser*, *fer à friser*, *poudre à lessiver*, etc. Si *fusil à chasser* ou *outil à travailler* ne sont pas viables, c'est que le verbe ne renvoie pas à une transition.

On pourrait éventuellement discuter la pertinence de ces critères, et se demander par exemple en quoi le déverbal *repassage* ne désigne pas une activité. Mais peu importe au fond, car ce qui est en cause ce ne sont pas les critères en eux-mêmes, mais toute la démarche. Admettons donc que notre cerveau dispose de critères, quels qu'ils soient, qui lui permettent effectivement de séparer le bon grain de l'ivraie, *fusil de chasse* de *fusil à chasser*. Pour parler de l'objet *fusil de chasse*, mettons, quelque chose dans mon cerveau sélectionne *fusil* et *chasser* dans notre « lexique mental », ainsi que la structure *à + infinitif* dans notre répertoire grammatical. Il instancie alors *chasser* à sa place après *fusil à*, et là il remarque que ça ne marche pas (*chasser* n'est pas une transition - ou tout autre critère). Il choisit alors une autre structure, celle avec *de*, et là ça marche (*chasse* est une activité). Il y aurait donc une sorte de surveillant qui vérifierait en permanence que les règles sont bien appliquées. On le voit : la nécessité d'un fantôme dans la

7. Chez Benveniste (1974)

machine est bel et bien consubstantielle à l'explication cartésienne cognitive. Ou alors y aurait-il des règles qui régulent les règles ? Mais jusqu'où ira-t-on alors ? Non, il faut bien poser un moment donné une entité qui décide, une intentionnalité. Et si on la pose, on est obligé de supposer qu'elle possède toutes les connaissances contenues dans les règles, sinon elle ne pourrait pas accepter ou refuser telle ou telle combinaison. Et dans ce cas, pourquoi cette cérémonie des règles ?⁸ Il lui suffirait de conjoindre les mots selon son intention, puisqu'elle sait ce qui est correct.

Figement et colle

Il y a aussi l'usage magique qui est fait de certaines notions, comme par exemple celle de *figement* pour expliquer la formation des UP. Les auteurs sont d'accord pour remarquer que les degrés de figement des UP sont variables, qu'il y a des expressions très figées, comme par exemple *col vert*, *cordon bleu*, *pomme de terre*, *casser sa pipe*, qui n'acceptent aucune transformation ou insertion (**un col très vert*, **un cordon tout à fait bleu*, **une pomme de terre tourbeuse*, **casser sa bouffarde*), et d'autres moins figées, comme *verre de vin (rouge, blanc, rosé, ...)*, *cheveux noirs (blancs, roux, ...)*. Le figement sert souvent de critère « objectif » pour fournir une séparation discrète entre les catégories. Nesselhauf (2003) par exemple, distingue entre les **combinaisons libres**, comme celle du verbe *to want* avec une multitude d'objets (*toys, a child, a car, the truth*, etc.), les **collocations**, comme *perform* qui accepte *a survey* comme objet, mais pas *a task*, selon elle, et les **idiomes**, comme *kick the bucket*, qui seraient totalement figés.

Le figement est ainsi pensé comme produit par une sorte de « colle » syntaxique ou sémantique, plus ou moins forte, qui rend compte à la fois des affinités entre les composants de l'UP et de la force qui les lie. Naturellement, toutes ces explications devraient s'effondrer si des occurrences d'expressions réputées agrammaticales sont effectivement produites, par exemple *c'est un cordon tout à fait bleu*, si on veut remarquer que la cuisinière, en plus d'être bonne, est habillée en bleu de pied en cape. La « colle » syntactico-sémantique, pourtant posée comme constitutive de l'UP, est alors dissoute dans le solvant des exceptions. Que reste-t-il alors de l'explication ?

Résolution du problème

L'observation des corpus indique une voie vers la résolution du problème, qui se fonde sur la nature dénominate et référentielle du signe. Prenons l'exemple d'*outil de travail* v. *outil à travailler* dans Bassac (2006 : 56-57) La consultation d'un corpus d'articles du *Monde* donne vingt occurrences d'*outil de travail*, aucune d'*outil à travailler*, ni même d'*outil à V_{inf}*, bien que par exemple *outil à travailler le bois* soit possible. Un autre corpus, de textes littéraires celui-ci, révèle

8. Voir l'argumentation sur les règles que Wittgenstein développe dans ses *Investigations Philosophiques* §202-242

l'existence d'une seule occurrence d'*outil* à V_{inf} , *outil à couper les oreilles*. En revanche *outil de* est très productif, puisqu'on trouve aussi *outil(s) de production, de promotion, de dissuasion, de transformation, de communication, de recherche, de défense*.

Cette disparité s'explique par une différence essentielle entre *outil de travail* et *outil à travailler*, que le LG ne fait pas : le premier existe dans la langue, et le second non. Mon cerveau n'a ainsi pas besoin de faire de longues vérifications lorsqu'il entend *fusil de chasse* ou *outil de travail* : il les reconnaît immédiatement. S'il entendait *fusil à chasser* ou *outil à travailler*, il ne les reconnaîtrait pas et chercherait une autre interprétation, discursive. Le LG traite ainsi de la même manière un signe qui existe dans la langue, *outil de travail*, c'est-à-dire une dénomination, et un signe qui n'existe pas, *outil à travailler*, mais qui pourrait apparaître sous la forme d'*outil à travailler le bois*, par exemple, c'est-à-dire un interprétant, un signe discursif.

Référence et dénomination

Deux questions se posent à propos de la dénomination et de la référence. La première concerne l'*existence* du signe dans la langue : s'agit-il ou non d'une *dénomination* existante ? Ainsi *oculiste* est une dénomination et *médecin des yeux* une désignation discursive, éventuellement définitoire. La seconde question, qui permet d'ailleurs de se prononcer sur la nature dénominative du signe, est celle du *type* de la référence : le signe réfère-t-il synthétiquement ou analytiquement ? *Oculiste* réfère synthétiquement, et *médecin des yeux* analytiquement, car *médecin* et *yeux* réfèrent ici individuellement. Mais, se dira peut-être le lecteur, il est normal qu'*oculiste* réfère synthétiquement, puisqu'il s'agit d'un mot monolexical. Notons d'abord que *pomme de terre* réfère de la même manière qu'*oculiste*, en un seul morceau (ni *pomme* ni *terre* ne réfèrent séparément), et pas différemment de *gendarme*, un monolexème pourtant proche par la forme du polylexème *pomme de terre*. Notons ensuite qu'*oculiste* est dans le même cas que *gendarme* : sa nature compositionnelle se repère assez aisément grâce à un radical qu'on perçoit aussi dans *oculaire*. Autrement dit, le caractère synthétique ou analytique de la référence n'est pas dans la dépendance de la mono- ou polylexicalité de la dénomination. Il existe d'ailleurs des entités dénominatives bien plus longues que les mots composés : les proverbes⁹. Ainsi, *Pierre qui roule n'amasse pas mousse* réfère bien synthétiquement à un type de comportement, sur lequel on porte un jugement. Notons que cette référence n'est pas univoque. Le proverbe n'est pas juste le nom du comportement : sans le proverbe, ce comportement n'aurait pas de dénomination, donc pas d'existence sociale, et il faudrait, si on voulait en parler, procéder à chaque fois à la construction d'un signe discursif, d'un interprétant, par exemple : « S'il n'est pas devenu riche, c'est qu'il n'a jamais réussi à se stabiliser

9. Voir l'argumentation de Kleiber (1994)

professionnellement », ou quelque chose de ce genre. La langue construit ainsi notre univers, et l'objet *pour nous* est une fonction du langage.

Une dénomination est ainsi caractérisée par sa référence globale. Référence synthétique et dénomination vont de pair. Une dénomination nomme un objet socialement existant, et d'ailleurs, elle est le signe de cette existence. Un objet sans nom est inexistant pour nous. Nous supposons que chaque objet de notre expérience, mettons tel arbre à fleurs mauves, possède une dénomination, en l'occurrence *jacaranda*, même si elle est inconnue de nous. Inversement, à chaque dénomination, par exemple *zinzolin*, nous supposons que correspond un objet, en l'occurrence une couleur, même inconnu de nous. Ce qui compte, c'est la *reconnaissance* du signe. Ainsi les constituants de la dénomination ne sont-ils que des indices qui nous permettent de la reconnaître, et non des éléments qui nous permettent de la « calculer » à l'aide de règles. Nous reconnaissons *outil de travail* ; nous ne reconnaissons pas *outil à travailler*. Nous reconnaissons *Pierre qui roule n'amasse pas mousse* comme une dénomination existante, mais pas *la pierre a roulé en bas de la pente et elle n'a ainsi pas pu se recouvrir de mousse*. La dénomination est un délimiteur d'expérience collective, et elle est reconnue d'un seul coup par sa forme. Si la forme est modifiée de manière trop drastique, nous ne la reconnaissons pas, et nous commençons à l'interpréter discursivement. Ainsi, Gaston Gross remarque qu'on ne peut pas remplacer *pipe* par *bouffarde* dans *casser sa pipe*, ni *anglaise* par *britannique* dans *clé anglaise*. Mais cette impossibilité n'est pas liée à quelque colle syntactico-sémantique qui agirait sur les sèmes de *pipe* et pas sur ceux de *bouffarde*. Si nous entendions *psycho-étude* ou *esprit-analyse* à la place de *psychanalyse*, nous ne les comprendrions pas non plus, parce que nous ne les **reconnâtrions** pas.

Une théorie générale du fonctionnement des unités phraséologiques

Un énoncé quelconque est ainsi composé d'éléments que nous **reconnaissons**, et d'autres que nous **interprétons**. Mais il n'y a pas discontinuité entre la référence synthétique et la référence analytique, entre celle, opaque, de la dénomination, et celle, discursive du signe interprétant. Ce ne sont pas des réalités discrètes : il existe toute une gradation référentielle, que nous allons essayer de cerner, des mots monolexicaux au texte tout entier.

Parmi les **mots monolexicaux** on peut distinguer tout d'abord les mots entièrement opaques, comme *thé* ou *voiture*, que nous percevons en un seul morceau par la forme, même s'ils sont historiquement composés ou dérivés (*voiture* vient du latin *vectum*, supin de *vehere* : porter, transporter). Ces mots réfèrent globalement. C'est aussi le cas des dérivés, comme *danseur*, ou des composés comme *psychanalyse* ou *gendarme*. Ces derniers sont moins opaques, puisque nous percevons l'effort néologique fait à leur création, mais aucune de leurs parties ne réfère séparément : *gens* et *arme* de désignent rien d'individuel dans *gendarme*.

C'est aussi le cas d'UP comme *messe noire*, *carte bleue*, *casser sa pipe*, *chemin de fer*, *prendre une veste*, *fusil de chasse*, *machine à laver*, etc., qui réfèrent bien synthétiquement, même si, comme pour *gendarme* et *psychanalyse*, nous sommes en mesure de percevoir leurs composants d'origine. Appelons ces UP des **UP lexicales**. Certains des éléments qui les composent peuvent être remplacés par d'autres, à la condition qu'on puisse percevoir un lien avec l'UP d'origine. Ainsi, on pourrait dire que la messe était *grise*, si, mettons, les satanistes étaient peu motivés, ou *rouge*, si elle était sanglante. Mais on ne comprend ces UP modifiées que par rapport à l'UP d'origine, *messe noire*. *Tel politicien a pris un smoking* se comprend en référence à *prendre une veste* s'il est clair qu'il a perdu les élections malgré ou en raison de son goût notoire pour ce vêtement, par exemple. Même remarque par exemple pour *prendre la main dans le Vuitton*, *faire d'une horloge douze coups*, *fusil de chiasse*, *machine à délayer*. Il y a suffisamment d'indices pour que nous puissions **reconnaître** les dénominations, et aussi **interpréter** les modifications en fonction de la situation.

Les UP lexicales diffèrent des monolexèmes par les modifications morphologiques qu'elles acceptent (par exemple la conjugaison pour les verbes), et aussi par la variabilité plus grande de leurs composants : il est plus facile de modifier *messe noire* que *psychanalyse*. Notons que leur capacité référentielle reste intacte malgré les éventuelles modifications, dont l'ampleur est limitée par la nécessité de la reconnaissance de la dénomination. Des insertions sont possibles dans les UP lexicales (*messe très noire*), mais pas au sein des monolexèmes, du moins en français. En anglais, on a des composés du type *the Soviet **fucking** Union*, mais aussi, des insertions dans des monolexèmes comme *extra-**fucking**-ordinary* ou *psycho-**bloody**-analysis*. A noter que ces insertions sont toujours dans le registre du juron.

Examinons maintenant les UP moins figées, comme par exemple *cheveux noirs*, *outil de travail* ou *fait de société*. Il est clair que lorsque nous les prononçons, nous ne faisons pas acte de création linguistique. Nous ne conjoignons pas *cheveu* avec *noir* après un examen de la couleur et un inventaire des structures grammaticales autorisées, en l'occurrence *N + Adj*. Ceci ne serait qu'un avatar de l'homoncule cartésien aux commandes du cerveau, une solution que nous avons rejetée. En réalité, nous utilisons un polylexème existant, *cheveux noirs*, que nous avons déjà entendu et utilisé d'innombrables fois. C'est une dénomination, qui évoque donc autre chose que juste une description de la couleur. Dire de quelqu'un qu'il a les cheveux *noirs*, *blancs*, *poivre et sel* ou *blonds* n'est pas juste une référence à une couleur : il y a également des indications concernant l'âge, peut-être l'origine ethnique, ou le degré d'intelligence (pour les blondes). Les UP de ce type réfèrent donc bien synthétiquement, mais aussi analytiquement. Ainsi *noir* dans *cheveux noirs* réfère à une couleur, mais pas dans les UP lexicales *caisse noire*, *boîte noire* ou *messe noire*.

Appelons ces UP des **UP semi-figées**. Elles ont d'autres caractéristiques, perceptibles à l'examen des corpus. Ainsi une requête sur *fait*¹⁰ de donne trois types d'usages : *fait de + inf* (par exemple, *le fait de choisir*), *fait de + agent* (par

10. La requête concerne seulement la forme nominale de *fait*.

exemple, c'est *le fait des catholiques*) et *fait de + type de fait* (par exemple *fait de société*). Pour le premier, nous n'avons pas noté de type de verbe particulier: tous les infinitifs semblent être en mesure d'occuper cette position, et il s'agit donc d'une structure discursive. Pour le deuxième, le N2 réfère à des groupes humains (catholiques, juifs, communistes, ...) ou à des personnes. Pour le troisième type, voici des N2 collectés dans deux corpus, par ordre de fréquence : *société, résistance, guerre, grève*, qui forment les dénominations *fait de société, fait de résistance, fait de guerre, fait de grève*. On peut déceler une parenté sémantique entre ces N2 : ils réfèrent à des entités qui concernent la vie en société. Effectivement, il ne semble pas possible pour les N2 de référer à des objets matériels (*un fait de couteau, de voiture, ..*) ou des entités abstraites (*un fait d'intelligence, de beauté, de gentillesse, ...*). Nous avons aussi relevé *fait de science* et *fait de volupté*, et nous ajoutons *fait de langue* à la liste, bien que nous ne l'ayons pas trouvé en corpus. La langue et la science concernent bien des entités de la vie sociale, et leur usage avec *fait de* ne contredit pas nos observations. Mais qu'en est-il de *fait de volupté* ? Voici la phrase qui contient l'expression, extraite de *Volupté*, de Sainte-Beuve :

« *Les matérialistes (et de nos jours la plupart des hommes le sont, du moins en pratique) envisagent le fait de volupté comme indépendant presque du reste de la conduite ... Les pères, frères aînés et tuteurs, dans les conseils qu'ils donnent à ce propos en font communément une affaire d'hygiène, d'économie, de régularité.* »

La volupté est ici clairement considérée en relation avec son importance pour la santé publique et l'économie, des faits de société, donc.

Ces UP peuvent être décrites comme composées d'un **pivot** (*fait de*) suivi d'un **paradigme** de N2 (*société, résistance, guerre, grève, ...*) qui partagent une caractéristique sémantique, laquelle s'impose aux éventuels N2 discursifs, comme *volupté*, vus alors sous l'angle social. Nous schématisons cela ainsi :

pivot + [**élément séminal** : *paradigme dénominatif => paradigme discursif*].

Soit, pour *fait de + N2* :

fait de + [**société** : *résistance, guerre, grève, science, langue, ... => volupté, ...*]

Une analyse de John Sinclair revisitée : la locution verbale *set in*

Est-ce que cette structure possède une validité générale ? Il semble que oui. En tous les cas, nous en avons trouvé de très nombreuses aussi bien en anglais qu'en

français. Pour l'anglais, nous avons notamment repris¹¹ l'exemple bien connu de Sinclair (1987: 150-159), qui a remarqué dans des corpus électroniques que le verbe *set in* prend généralement un sujet négatif (*rot, decay, panic set in*, etc.). Il ne donne pas d'exemples du contraire, mais une requête sur *sets in* dans le *British National Corpus* révèle l'existence de sujets neutres (*ritual, fixity, a process of cumulative causation, reality, ...*), et même de sujets positifs (*sustained improvement, recovery, a kind of relief*). Comment l'expliquer ?

On constate, dans une requête sur *sets in*, que l'élément le plus fréquent est *winter* (*winter sets in*) avec 8 occurrences sur 86. Si on y ajoute les autres mots qui ont un rapport avec le mauvais temps, on en compte 16 de plus (*autumn, thaw, gale, monsoon, rain, re-heat, winter weather, bad weather, dry weather*) soit près du tiers de l'ensemble des occurrences. Du mauvais temps à des choses négatives, il n'y a qu'un pas, d'où sans doute des mots comme *panic, rot, decay, tension, boredom*, etc. On peut donc formuler le schéma référentiel suivant :

[*winter* : *bad weather, gale, rain, ... => bad things* : *rot, panic, fatigue, ...*] + **set in**

Comment justifier les sujets positifs de *set in* ? Peut-être l'élément séminal en est-il *thaw*, le dégel, un événement positif qui a lieu après un élément négatif (le gel). Il s'agit d'une amélioration, ce qui est aussi le sens des trois éléments positifs relevés dans la concordance : *sustained improvement, recovery, a kind of relief*. Ce qui donne le schéma suivant :

[*thaw* : => *improvement, recovery, relief, ...*] + **set in**.

Enfin, il y a aussi, comme pour *volupté*, attribution en discours de caractéristiques sémantiques. C'est le cas de *Pelagianism sets in*. La présence de *sets in* transforme *Pelagianism*, pourtant une doctrine optimiste qui nie le péché originel et affirme l'existence du libre-arbitre, en une entité négative. Ce genre d'allocation dynamique de sens est très fréquent, par exemple, dans *elle lui a révélé son origine*, on comprend que l'origine était secrète¹².

Si nous retournons à notre problème de départ, à savoir la distinction entre *outil de travail* et *outil à travailler*, nous avons établi que le premier est une dénomination est que le second ne peut être, à la rigueur, qu'une séquence discursive. On peut aussi comparer *outil* + *à* + *V_{inf}* avec d'autres occurrences de *N* + *à* + *V_{inf}* et constater qu'il y a peu d'UP de ce type dans les corpus. Il y a bien *pince à épiler, machine à laver, salle à manger, chambre à louer*, mais en nombre assez restreint. Et peut-être le lecteur ressentira-t-il comme nous une différence entre ces UP et *outil de travail* : les premières semblent plus figées. La raison en est maintenant claire : il s'agit d'UP lexicales, qui n'acceptent pas facilement de modifications,

11. Frath 2008

12. Il est à noter que les sémantiques qui reposent sur la résonance sémique entre les mots, comme la sémantique interprétative de F. Rastier, ne peuvent pas rendre compte de l'allocation dynamique de sens, sauf à postuler des sèmes afférents de tous les types possibles au sein de tous les mots. La sémantique générative de Pustejovsky ne le peut pas non plus, pour la même raison (voir Frath 2004).

alors qu'*outil de travail* est une UP semi-figée, où *travail* est un générateur de N2 déverbaux, lesquels, placés après *outil* les font voir au sein de l'UP (*outil de communication, de promotion, ...*) en relation avec l'idée de travail.

Conclusion

On voit que ce type d'analyse est loin des théories syntactico-sémantiques qui formulent des principes généraux *a priori* (et souvent *ad hoc*) s'appliquant au lexique. Ici, les conclusions sont le fruit de l'observation et de la généralisation *a posteriori* selon un point de vue, en l'occurrence la référence et la dénomination. Il est possible que d'autres points de vue puissent être productifs également, mais pour l'instant, celui-ci semble relativement inexploré et il livre un grand nombre de conclusions. La référence se fait par des **monolexèmes dénominatifs** (*thé, psychanalyse*), des **UP lexicales** (*prendre une veste, une boîte noire*), et des **UP semi-figées** (*fusil de chasse, d'assaut, de compétition, ...*) dans lesquelles un **pivot** lexical est accompagné d'un **paradigme** généré par similitude à partir d'un **élément séminal**. L'observation des corpus permet de donner des descriptions ancrées dans les données réelles, même si elles ne couvrent pas *toute* la langue. Il reste maintenant à appliquer la même méthode à la construction des phrases et des textes.

Bibliographie

- Bassac Christian (2006) : *Morphologie et information lexicale*, Mémoire d'habilitation à diriger les recherches, Université de Bordeaux 3.
- Benveniste Emile (1974) : *Problèmes de linguistique générale*, 2. Gallimard.
- Fodor, J. & Lepore, E., 1998 : « The emptiness of the lexicon », in *Linguistic Inquiry*, 29.2 p. 269-288.
- Frath Pierre (2002) « Etude du verbe 'commencer' en contexte », *Journal of French Language Studies*, 12.2 (2002), pp 169-180, Cambridge University Press
- Frath Pierre (2004) : « Rules and Intentionality in the Generative Lexicon », in *Journal of Cognitive Science*, Vol. 3, No. 2 (2002), Institute for Cognitive Science, Seoul
- Frath Pierre (2005a) : « Pour une sémantique de la dénomination et de la référence », in *Sens et références, Sinn und Referenz*. Adolfo Murguía (dir.), Günther Narr Verlag, Tübingen
- Frath Pierre (2005b) : « Post-cognitivism: a Plea for Reference in Linguistic Theory », *Proceedings of the Third International Workshop on Generative Approaches to the Lexicon*, Pierrette Bouillon and Kyoko Kanzaki, eds., University of Geneva, May 19-21, 2005.
- Frath Pierre (2007) : *Signe, référence et usage*, Editions Le Manuscrit, Paris.
- Frath Pierre (2008) : « Pour commencer, il faut arrêter de décoder. Plaidoyer pour une linguistique sans métaphysique », *Journal of French Language Studies* 18/2 (2008) : 147-173. Cambridge University Press.
- Gross Gaston (1996) : *Les expressions figées en français*, Ophrys.
- Kleiber Georges (1994) : *Nominales. Essais de sémantique référentielle*, Armand Colin.
- Naciscione Anita (2001) : *Phraseological Units in discourse: towards applied stylistics*, Latvian Academy of Culture, Riga
- Naciscione Anita (2002) : « Phraseological metaphor: dead or alive? », in *Ranam n°36/2002*, pp.23-30, Frath P. & Rissanen M., eds, Strasbourg.

- Nesselhauf, Nadja (2003) 'The Use of Collocations by Advanced Learners of English and Some Implications for Teaching' in *Applied Linguistics* 42/2 : 223-42.
- Paulin Catherine (2002) : « Language crossing: lexical and phraseological specificities », in *Ranam* n°36/2002, pp.57-66, Frath P. & Rissanen M., eds, Strasbourg.
- Peirce Charles Sanders (1978, rééd.) : *Ecrits sur le signe*, rassemblés, traduits et commentés par Gérard Deledalle, Seuil.
- Pustejovsky James (1993): « Type Coercion and Lexical Selection ». In *Semantics and the Lexicon*. J. Pustejovsky, ed., Kluwer Academic Publishers. Dordrecht.
- Pustejovsky James (1995) : « Linguistic Constraints on Type Coercion ». In *Computational Lexical Semantics*, Patrick Saint-Dizier & Evelyne Viegas, eds. Studies in NLP. Cambridge University Press.
- Ryle G. (1949, 1978) : *La notion d'esprit*. Traduit par Suzanne Stern-Gillet. Payot.
- Searle John R. (1980): Minds, Brains and Programs. *Behavioral and Brain Sciences* 3 (3): 417-457.
- Sinclair John (1987) : *Looking up : An Account of the COBUILD Project in Lexical Computing*. John Sinclair, ed. Collins Cobuild.
- Wittgenstein Ludwig (1953, 1963) : *Philosophical Investigations*, translated by G.E.M. Anscombe, Basil Blackwell.