

FRATH Pierre, & GLEDHILL, Christopher, 2005. Qu'est-ce qu'une unité phraséologique? in Bolly, C, Klein J-R & Lamirov, B. (éds) *La Phraséologie dans tous ses états*, Actes du colloque "Phraséologie 2005", Louvain-La-Neuve, 13-15 oct. 2005, Cahiers de l'Institut de Linguistique de Louvain 31/2-4:11-25. ISSN –0771-6524

Qu'est-ce qu'une unité phraséologique?

Frath, P. & Gledhill, C. (2005) in Bolly, C, Klein J-R & Lamirov, B. (éds) *La Phraséologie dans tous ses états*, Actes du colloque "Phraséologie 2005", Louvain-La-Neuve, 13-15 oct. 2005, Cahiers de l'Institut de Linguistique de Louvain 31/2-4:11-25.

Pierre Frath¹ et Christopher Gledhill²

¹Département d'anglais
Université de Reims Champagne-Ardenne
²Département d'anglais
Université Marc Bloch, Strasbourg

Abstract

What is a phraseological unit? Such ontological questions are usually either avoided altogether or answered away by resorting to an endless recess of ever smaller constituents organised by logical rules, thus unwittingly paving the way for the metaphysics of Universals and Laws of Nature. We offer a third way, based on reference. We argue that phraseological units do not have a separate existence between single words and open choice sentences. Rather, we consider them as variable referential denominators which are able to refer in their own specific way.

Mots-clés: Unité phraséologique, corpus, référence, dénomination, interprétant

Qu'est-ce qu'une unité phraséologique? La question ontologique, celle de l'être des choses, est assez angoissante et nous avons tendance à éviter de nous la poser. C'est ce que font par exemple les linguistiques descriptives, qui observent la langue, souvent dans de vastes corpus à l'aide d'outils d'exploration informatique. Elles constatent l'existence de très nombreux phénomènes, largement insoupçonnés, qu'elles décrivent alors, souvent sans essayer de les expliquer.

Ou alors, si nous essayons de répondre à la question de l'être, nous tendons assez naturellement à penser la complexité comme un assemblage d'éléments plus simples, jusqu'à faire reposer toute la pyramide explicative, *in fine*, sur des entités élémentaires. On postule alors volontiers des Primitives, des Universaux, des Relations, des Rôles, des Principes, des Règles, c'est-à-dire tout un Panthéon d'entités dont la multiplication n'est contrainte que par l'imagination des chercheurs. A défaut de se préoccuper de l'*existence* de ces entités au-delà des services rendus à la théorie, on risque de transformer la linguistique en une branche de la métaphysique. Il doit y avoir une autre voie, entre la non-réponse descriptiviste et l'explication par l'encodage/décodage, cette quête de la clé cachée dont Richard Rorty se moque lorsqu'il fait l'éloge du *Pendule de Foucault* de Umberto Eco, un livre qui selon lui est une satire de la quête essentialiste en sciences humaines (Rorty 1996). Nous examinons ces

deux attitudes dans la littérature sur les unités phraséologiques, avant de proposer une troisième voie, inspirée de C.S. Peirce, et fondée sur la référence.

Descriptivisme

L'attitude descriptiviste s'est considérablement développée avec l'avènement des grands corpus électroniques. Les outils d'exploration textuelle montrent clairement que le voisinage des mots n'est pas aléatoire. La présence de tel mot augmente la probabilité de la présence de tel(s) autre(s) mot(s) au sein de ce qui constitue alors une collocation, et dont voici quelques exemples en anglais¹, *blow your own trumpet*, *stark naked*, *piping hot*, *strong tea*, *powerful car*, *decay set in*, *spill the beans*, etc. Mais qu'est-ce qu'une collocation? Selon Van Roey (1990:46), "it is that linguistic phenomenon whereby a given vocabulary item **prefers**² the company of another item rather than its synonyms because of constraints which are not on the level of syntax or conceptual meaning but on that of usage". La raison de ces "préférences", comme le dit Van Roey, est ainsi mise sur le compte de l'usage. John Sinclair a formalisé ce point de vue dans sa dichotomie bien connue du "principe idiomatique" et du "choix ouvert". Dans certains cas, nous utiliserions les lois de la syntaxe pour construire des phrases "ouvertes", par exemple *she's blowing her nose* (elle s'est mouchée) ; dans d'autres, nous aurions recours à des expressions "idiomatiques" toutes faites, par exemple *she blew her own trumpet* (elle s'est fait mousser, elle a fait sa propre promotion). Notons qu'il ne s'agit pas là d'une véritable explication. Sinclair ne fait que nommer ces deux possibilités de création linguistique, sans qu'on voie clairement pourquoi il en est ainsi.

La linguistique de corpus a certainement fait faire un bon en avant considérable à la description des phénomènes linguistiques, et aussi à la prise de conscience de l'importance de l'usage de blocs lexicaux tout faits dans la langue. Elle a renforcé le camp des tenants de la lexico-grammaire, qui tiennent qu'une phrase est autre chose qu'un simple schéma général qui se "remplit" de vocabulaire au moment de l'énonciation.

Syntaxe et métaphysique

La deuxième attitude, qu'on peut qualifier de **syntaxique**, entend expliquer *pourquoi* les mots en préfèrent d'autres au sein d'entités polylexicales plus ou moins figées. Elle fait la même distinction que Sinclair entre l'"ouvert" et l'"idiomatique", et fait également la différence entre des *combinaisons libres* (par exemple *I want a baby*), les *collocations*, qui seraient soumises à certaines contraintes lexicales (on peut dire par exemple *strong tea* et *powerful car*, mais pas *powerful tea* ni *strong car*) et les *idiomes purs* (par exemple *kick the bucket* – mourir, qui est totalement lexicalisé) (Nesselhauf 2003:225-226). Mais elle s'intéresse surtout à la "colle" syntaxique plus ou moins forte qui relierait les mots les uns aux autres. Les assemblages collocationnels ou idiomatiques sont ainsi volontiers expliqués en termes d'opérateurs ou de fonctions, par exemple *stark* et *piping*, qui seraient des "intensificateurs", ou *a speck* dans *a speck of dust*, qui aurait une "fonction quantitative", etc. (Mel'cuk 1988). D'autres tentent d'établir des listes de relations entre les parties de l'expression phraséologique. On distingue alors volontiers des relations *sémantiques*, par exemple la partie et le tout (*duck foot*), la comparaison (*pumpkin bus*), la composition (*stone furniture*), le lieu (*Easter Oregon meal*), etc., ainsi que des relations *prédicatives* telles que la cause (*malarial mosquitoes*, *tidal wave*), la possession (*presidential power*), l'instrument (*manual signals*), etc. Ces distinctions sont

¹ Les deux auteurs de l'article sont des anglicistes.

² C'est nous qui soulignons

également faites pour des expressions verbales, par exemple *he heads the procession* (lieu), *he headed the ball* (instrument), *the salad is heading nicely* (production)³.

La mise en œuvre de ce type d'explication ne pose pas problème tant que l'on se trouve dans le cadre théorique prôné par le père fondateur de la sémantique, Michel Bréal, dans son *Essai de sémantique* :

"Nous appelons loi, prenant le mot dans le sens philosophique, le rapport constant qui se laisse découvrir dans une série de phénomènes. Un ou deux exemples rendront ceci plus clair. Si tous les changements qui se font dans le gouvernement et les habitudes d'un peuple, se font dans le sens de la centralisation, nous disons que la centralisation est la loi. [...] De même, si la grammaire d'une langue tend d'une façon constante à se simplifier, nous pouvons dire que la simplification est la loi de la grammaire [...] Il ne saurait être question d'une loi préalablement concertée, encore moins d'une loi imposée au nom d'une autorité supérieure." (p. 9)

On constate ou on établit des ressemblances entre des phénomènes, qu'on regroupe alors dans des catégories, qu'on nomme (*centralisation, simplification*). Tant que le linguiste est conscient qu'il agit et pense dans ce cadre, l'irruption de la métaphysique dans la théorie est contenue.

Il en va tout autrement lorsqu'il commence à attribuer une réalité platonicienne *en soi* aux catégories, lorsqu'il commence à entrer dans une logique de code, lorsqu'il considère l'énoncé comme le produit d'un calcul, comme l'émanation d'un fonctionnement logique immatériel sur un substrat biologique, le cerveau. Il entre alors dans une quête essentialiste, le plus souvent sans avoir conscience des tenants et des aboutissants métaphysiques de son propos. Les communautés de chercheurs développent alors des corps de croyances collectives, qui tirent leur évidence de leur règne sans partage sur les esprits, et rares sont les jeunes chercheurs qui osent alors mettre en doute ce qui relève désormais de la *doxa*. La croyance, par exemple, que le langage possède naturellement des caractéristiques logiques est profondément ancrée chez nombre d'auteurs, notamment cognitivistes. Voici par exemple un extrait d'un article de Thomas Hun-tak Lee, de l'université de Hong-Kong : "*The acquisition of [...] natural language logical structures*⁴ is of interest to linguistics and cognitive science for a number of reasons. They are to the best of our knowledge **a species-specific property** that has emerged in the course of human evolution. No other animal has a communication system that is anything close to **propositional logic** or **predicate calculus**. All human languages have them. **Operator-variable structures** are closely tied to the ability to handle **numbers**, which exhibits the property of 'discrete infinity', considered to be also the most elementary property of the **language faculty**." (Hun-tak Lee, 2002, p. 157).

La métaphysique cognitiviste tient tout cela pour acquis. Il n'en est rien. Faire reposer notre activité langagière sur notre génome ("*a species-specific property*"), cela peut sembler éminemment scientifique, beaucoup plus en tous les cas que de l'attribuer à Dieu, comme le faisait Descartes, mais en réalité, c'est faire preuve d'inculture philosophique. Le linguiste cognitiviste reste dans cette fuite devant l'être dont Heidegger dit qu'elle est typique de notre époque, dans l'illusion réductionniste selon laquelle on peut expliquer un phénomène par d'autres dont on ne sait rien. Il croit que les choses que nous avons nommées ("*logical structures, propositional logic, predicate calculus, numbers, language faculty*") possèdent une

³ Tous ces exemples de divers auteurs sont cités par H.U. Boas dans Boas (2003).

⁴ C'est nous qui soulignons.

existence délimitée en soi, et qu'à ces entités nommées correspondent des entités délimitées réelles. C'est une vue très répandue, mais aussi très naïve. Certes on ne place plus leur origine dans le Monde Idéal de Platon ; elles sont dorénavant codées dans le génome, où la sagacité des chercheurs finira bien par les dénicher. C'est ainsi que l'on voit périodiquement des scientifiques "découvrir" les gènes d'entités tels que l'intelligence, l'amour, l'homosexualité, sans qu'ils s'aperçoivent qu'il ne s'agit que de comportements divers et variés que nous regroupons sous des dénominations. Peut-être un jour trouvera-t-on les gènes de la *language faculty* avec son cortège d'opérateurs logiques et d'opérandes. En réalité, ces théories ne font rien d'autre que remplacer une cause première par une autre, Dieu par le Génome.

Le code linguistique n'est alors qu'une sorte de programme généré par le cerveau, qui produit ensuite les énoncés. La compréhension mutuelle est assurée par le caractère "logique" du programme, la logique étant universelle⁵. En conséquence, selon ce point de vue, aucun savoir linguistique n'est jamais acquis. Tout énoncé est éternellement recalculé grâce au code, comme le ferait un ordinateur. C'est ce point de vue qui est à la base des linguistiques cognitivistes telles celles de Jaekendoff, Langacker ou Fauconnier, qui produisent des théories où la langue est expliquée par autre chose, par des caractéristiques attribuées arbitrairement au cerveau, et dont on ne sait rien. La lexico-grammaire est une épine dans le pied de cette conception de la langue, puisqu'elle admet implicitement l'existence de blocs tout faits, dont l'assemblage n'a pas d'autre cause que l'usage, c'est-à-dire notre habitude de les utiliser ainsi.

Les théories syntaxiques tendent naturellement à expliquer les unités phraséologiques par une micro-syntaxe, valable également dans certains cas au niveau des mots monolexicaux composés, comme *psychanalyse* ou *danseur*. Le sens de tels mots peut être partiellement expliqué par les éléments qui les composent : la psychanalyse est l'analyse de la psyché, un danseur est quelqu'un qui danse. On achoppe toutefois assez rapidement sur des difficultés comme celles relevées par Gaston Gross (1996) : l'expression *cheveux noirs* subit assez facilement le test de la nominalisation de l'adjectif (*la noirceur des cheveux*), mais ce n'est le cas ni pour *colère noire* (**la noirceur de la colère*), ni pour *messe noire* (**la noirceur de la messe*). Pourquoi en est-il ainsi ? Rien dans l'approche syntaxique ne permet de le dire. Après avoir noté ce genre de phénomènes, les auteurs se lancent alors volontiers, comme G. Gross, dans des typologies gigantesques qui rendent le tableau global encore plus confus. Certains auteurs recensent ainsi plusieurs centaines types de phrasèmes nominaux différents, sans qu'on perçoive le moindre espoir de formulation de quelque *loi* que ce soit, dans le sens où Bréal emploie ce mot.

La raison en est le présupposé syntaxique de la théorie, qui ne permet pas de dire les limites entre le syntaxique et le lexical. Où s'arrête le code ? Où commence le sens lexical ? Y a-t-il d'ailleurs une limite, ou bien la sémantique est-elle le prolongement infraléxico de la syntaxe ? Faute d'un point de vue correct, la linguistique des unités phraséologiques est condamnée à errer entre les affirmations de la lexico-grammaire et les empiétements de la syntaxe.

La théorie en linguistique

Quel pourrait être un point de vue correct en linguistique ? Voici l'opinion de C.S Peirce (1984) sur les raisonnements, qui "*ne devraient pas prendre la forme d'une chaîne, jamais*

⁵ Pour une discussion sur le fond de la position cognitiviste, voir Frath (2005a, 2005b, 2005c)

plus forte que son maillon le plus faible, mais d'un câble dont les fibres peuvent être très fines pourvu qu'elles soient suffisamment nombreuses et torsadées serré" (p. 197). Il ajoute qu'il n'y a "*aucune connaissance absolument première d'un objet quelconque*" (p. 198). Une bonne théorie linguistique ne peut ainsi pas être une théorie du code, trop fragile (à chacun de ses chaînons) et trop métaphysique (il n'y a pas de connaissance première, le premier chaînon est ainsi de toute façon manquant).

L'alternative est la mise en relation de phénomènes les uns par rapport aux autres, selon un point de vue, avec formulation des différences. C'est ce que suggère Bréal. On parvient ainsi, *in fine*, à une théorie où chaque phénomène prend sa place au sein d'un ensemble général suffisamment souple pour intégrer sans s'effondrer d'autres observations. Nous avons appliqué cette méthode à la définition de l'unité phraséologique, prenant appui sur la notion de référence.

Référence

La référence a fait couler beaucoup d'encre. Elle a préoccupé de manière centrale la philosophie du langage, notamment sa lignée russellienne, qui attribue à la logique une position ontologique qui en fait la langue de l'univers. C'est d'ailleurs ce que Russell et Whitehead ont voulu démontrer dans leurs *Principia Mathematica*, une tentative pour réduire mathématiques à la logique. Dès lors, si cette entreprise réussit, tout ce qui peut être exprimé par la logique est vrai du monde. Le théorème d'incomplétude de Gödel a pourtant montré, dès les années 1930, que ce point de vue est erroné. En effet, il démontre qu'il est impossible de déterminer la valeur de vérité de toutes les propositions au sein d'une axiomatique. En conséquence, il ne saurait y avoir d'axiomatique du *tout*, fût-elle logique, car certaines propositions resteraient indécidables. Wittgenstein, de son côté, a montré, avec son Argument du Langage Privé⁶, que la parole ne peut pas être l'encodage de la pensée. Mais sans doute les préjugés ont-ils la vie dure, y compris chez certains philosophes du langage.

La linguistique cognitive, quant à elle, a hérité d'une forme dégradée de logicisme qui l'amène à ne pas s'intéresser à la référence, car elle croit, faussement, que cette question a été réglée une fois pour toute par la philosophie du langage⁷. En outre, la linguistique moderne, à la suite de Saussure et de Bloomfield, a fait l'impasse sur la référence en adoptant une conception dyadique du signe, constitué d'un signifiant et d'un signifié, qui ne fait pas de place à la référence. Également, les deux fondateurs du structuralisme ont mis explicitement de côté la question de la référence pour des raisons méthodologiques : la linguistique aurait suffisamment de pain sur la planche avec l'étude des signes pour ne pas se lancer dans des domaines qui sont plus, selon eux, du ressort des autres sciences, notamment la psychologie, la sociologie, la physique et la chimie⁸.

Pourtant, si nous parlons, c'est bien *de* quelque chose. Mais comment prendre en compte la référence dans la théorie linguistique ?

Dénomination, objet et interprétant

⁶ *Les investigations philosophiques*, paragraphes 242 à 315.

⁷ Pour plus de détails, voir Frath 2005a, 2005b, 2005c.

⁸ Pour plus de détails, voir Frath 2005b.

La sémiotique triadique de C.S Peirce (1978) permet de l'inclure avec élégance et simplicité⁹. Selon cet auteur, la **dénomination**, par exemple le mot *zinzolin* (choisi ici pour sa rareté dans l'espoir que le lecteur ne le connaît pas) et l'**objet** du signe, ici l'objet zinzolin, sont reliés par un lien constitutif et nécessaire sans lequel nous ne pouvons ni penser les choses, ni les percevoir. Pour tout objet que nous rencontrons, nous supposons l'existence d'un signe, et pour tout signe auquel nous sommes exposés, nous supposons l'existence d'un objet. C'est ce que certains auteurs (Martin 1976, Kleiber 1994, 2001) ont appelé la présupposition d'existence, une notion qui trouve son origine dans le couple *suppositio / denominatio* des philosophes du Moyen-Âge, pour qui la *denominatio* désigne la *suppositio*, c'est-à-dire ce qui est *supposé exister*. Ce lien ontologique ne présuppose pas de connaissances *a priori*. Toute connaissance est produite dans un second temps au sein de signes complexes, appelés **interprétants** par Peirce. Illustrons cela.

Si le lecteur vient d'apprendre l'existence du mot *zinzolin*, et s'il fait confiance à notre compétence linguistique, il sait maintenant qu'il existe pour notre communauté linguistique un **objet** qui correspond à *zinzolin*, même s'il ne le connaît pas lui-même. Une simple question pourra provoquer une explication au sein d'un signe interprétant définitoire : *zinzolin* est une couleur, un rouge violacé. Notons que sans le mot, la couleur zinzolin ne peut pas être perçue de manière délimitée : ce sera simplement une gradation de rouge. Notons également qu'il est difficile de parler de cette couleur sans le mot. La dénomination est donc un délimiteur de notre expérience commune. Sans le mot, pas de couleur zinzolin délimitée pour nous, même si elle existe physiquement. La dénomination participe ainsi de l'existence de l'objet : elle possède un caractère ontologique. Notons encore que la définition de *zinzolin* ci-dessus comprend d'autres dénominations : *couleur*, *rouge* et *violacé*. Un signe **interprétant** est ainsi constitué de **dénominations** qui permettent de voir un **objet** selon un certain angle, en l'occurrence *zinzolin* d'un point de vue définitoire.

La référence des unités monolexicales et polylexicales

Beaucoup de dénominations sont des unités *monolexicales*, par exemple *psychanalyse*, *danseur* ou *thé*. Elles réfèrent en vertu du lien ontologique qui les unit aux objets qu'elles désignent. Il ne saurait y avoir de thé pour nous sans le mot *thé*, et lorsque nous connaissons le mot *thé*, nous savons aussitôt que ce mot désigne une boisson d'un certain type. Une fois que le mot existe dans la langue, il vit une vie indépendante, riche de tous les contextes dans lesquels il apparaît. Il se peut que *psychanalyse* ait signifié la même chose qu'*analyse de la psyché* pendant quelque temps après sa création, mais très rapidement, ce mot a commencé à apparaître dans de multiples signes interprétants qui ont permis aux locuteurs d'en examiner l'objet sous tous ses aspects. Cela ne fut pas le cas d'*analyse de la psyché*, qui est un interprétant définitoire et une trace historique de la néologie.

Il existe beaucoup d'unités *polylexicales* qui fonctionnent de la même manière que les mots monolexicaux, en tant qu'elles réfèrent globalement, comme *psychanalyse*, et non analytiquement comme *analyse de la psyché*. Ainsi, *chemin de fer*, qui n'est pas qu'un chemin de fer, *carte bleue*, qui n'est pas forcément bleue, ou *carte verte*, qui n'est pas une carte. Ce qui compte, c'est la référence : toutes ces unités réfèrent globalement. La différence entre *cheveux noirs* et *colère noire* s'explique alors facilement : le premier est plutôt discursif (les

⁹ Notre point de vue n'est qu'*inspiré* de Peirce : il prend beaucoup de libertés avec cet auteur complexe. Par exemple, ce que nous appelons *dénomination*, pour nous placer dans une filiation continentale nominaliste qui remonte à Guillaume d'Ockham et à Pierre Abélard, Peirce l'a appelé le *representamen*. Il rejette d'ailleurs le nominalisme avec une certaine violence.

cheveux pourraient être d'une autre couleur) et réfère analytiquement, alors que le second réfère globalement à un type de colère, qui ne pourrait avoir d'autre couleur.

Ce qui compte, ce n'est donc pas l'aspect mono- ou polylexical d'une séquence de mots, c'est sa capacité à référer ensemble ou séparément. Il faut cependant noter qu'il n'y a pas de séparation radicale entre les deux types de phrasèmes, entre des expressions idiomatiques comme *messe noire* d'une part, et d'autre part des unités qui seraient entièrement discursives, comme *cheveux noirs*, ainsi qu'on le verra ci-après.

Collocations dénominatives nominales : *strong tea*

Reprenons l'exemple de J.R. Firth déjà cité plus haut : on peut dire *strong tea* et *powerful car*, mais pas *strong car* ni *powerful tea*, bien qu'on puisse dire *powerful argument* et *strong argument*. Une sémantique conceptuelle avancerait que *strong* et *powerful* posséderaient tous les deux le sème de force et que pour une raison ou pour une autre, *tea* et *car* sélectionneraient plutôt l'un que l'autre, et *argument* l'un ou l'autre. Mais on voit bien que l'explication est entièrement *ad hoc*. On peut aussi invoquer l'usage, comme le font les linguistes de corpus, mais cela ne constitue bien sûr pas une explication.

Cependant, *strong tea* n'est certainement pas une création linguistique *ex nihilo* produite par mon cerveau selon des règles pour exprimer la force d'un thé. Je ne pense pas d'abord à thé et à force, pour choisir ensuite *strong* plutôt que *powerful* pour quelque raison sémantique. Il s'agit d'une collocation dénominative toute faite, lexicalisée, disponible dans la langue, et que j'utilise pour référer au thé en relation avec sa concentration en tanin et en théine. Et effectivement, une requête dans le BNC¹⁰ montre que *strong* est l'adjectif le plus fréquent avec *tea* (28 occurrences). Il s'agit donc bien d'une collocation lexicalisée. Mais si je peux parler de la force d'un thé, je peux aussi admettre qu'il la possède moins ou autrement. Et en effet, le BNC révèle l'existence d'autres adjectifs qui peuvent exprimer divers degrés de force, tels que *weak*, *rich*, *ordinary*, *proper*, *robust*, etc. *Strong tea* est ainsi une collocation dénominative constituée de deux dénominations, *strong* et *tea*, capables de référer isolément, mais qui réfèrent ici conjointement. Elle ouvre sur sa gauche un paradigme commutatif scalaire dont *strong* est l'élément prototypique séminal. Nous avons donc ici une structure référentielle que nous représentons par la formule :

[*strong* : *weak*, *rich*, *ordinary*, *proper*, *robust*, etc.] + *tea*

Cheveux noirs est-il une collocation dénominative nominale ? Notre intuition linguistique nous dit que oui, car cet assemblage de mots semble certainement exister dans la langue. Et effectivement, une requête sur *cheveux* dans un corpus français de 25 MO révèle l'existence d'un paradigme de couleurs à droite de *cheveux*, dont *blancs* est l'élément séminal, car le plus fréquent (15 occurrences). La structure référentielle de *cheveux* est donc la suivante :

cheveux + [*blancs* : *noirs*, *roux*, *blonds*, *jaunes*, *châtain*, *gris*, *poivre et sel*, *verts*, ...]

Quant à *messe noire* et à *colère noire* notre corpus n'en a pas livré d'occurrences, ce qui n'empêche pas notre intuition linguistique de les accepter. Pourrait-on attribuer d'autres couleurs à la messe ou à la colère ? A la rigueur, mais uniquement dans un cadre de revitalisation du figement, si on affirmait par exemple que la messe était grise, pour signifier, mettons, qu'elle a été célébrée par des satanistes peu motivés, ou rouge, si la messe noire a été sanglante. Il faut noter que l'emploi d'une autre couleur se fait uniquement par rapport à la

¹⁰ *British National Corpus*, une archive électronique de 100 millions de mots.

couleur séminale de l'expression : nous ne comprendrions pas *messe grise* si nous ne savions pas que cette expression est construite sur *messe noire*. Mêmes remarques pour *colère noire*, pour laquelle des contextes de revitalisation du figement peuvent être imaginés facilement.

Collocations dénominales verbales : *rot set in*

Nous procédons de même avec l'exemple bien connu de Sinclair (1987: 150-159), qui a remarqué que le verbe *set in* prend généralement un sujet négatif (*rot, decay, panic, etc.*). Il ne donne pas d'exemples du **contraire**, mais le BNC 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*). Voici une concordance classée par ordre alphabétique du premier mot à gauche de *sets in*, réalisée avec Micro-Concord¹¹ à partir de résultats extraits du BNC à l'aide de son outil d'interrogation SARA. La 3^e personne (*sets in*) a été choisie pour des raisons de simplicité de traitement.

MicroConcord search SW: sets

60 characters per entry

Sort : 1L/SW shifted -1 characters.

1	inks and culinary amnesia	sets in	somewhere between the ti
2	te understandable anxiety	sets in	and the situation deteri
3	the old guard when apathy	sets in	. When mid-table apathy
4	. When mid-table apathy	sets in	, I force him to undertak
5	see. Even before autumn	sets in	, gardeners start thinkin
6	be before this behaviour	sets in	- indeed it was emphasis
7	sugar before the bloating	sets in	, but what might this be?
8	it is put down as boredom	sets in	and some other activity
9	s engaged in. As caries	sets in	at a weak point in the t
10	s of cumulative causation	sets in	. Instead of a see-saw
11	running, a vicious circle	sets in	- a process of cumulativ
12	ome so anxious that colic	sets in	. When times are bad an
13	actionary commodification	sets in	, the deleterious upshot
14	just in case complacency	sets in	! At the moment when co
15	nt when conscientiousness	sets in	, when a conscience is ac
16	sh in before conscription	sets in	. At best the Unix busi
17	efore irreversible damage	sets in	. Decay sets in if so
18	damage sets in. Decay	sets in	if soft tissues remain,
19	true that when a decline	sets in	in F1, it is very diffic
20	thereafter a slow decline	sets in	. It is also true that
21	After this, depression	sets in	and deep seated anger ca
22	ll stimulants, depression	sets in	when the initial stimula
23	owth stops, deterioration	sets in	. UNDERSTANDING COME da
24	dawns when disorientation	sets in	, and you begin to succum
25	butter, before dyspepsia	sets in	. of course, if the va
26	a kind of see-saw effect	sets in	with the recruitment tea
27	e on when serious erosion	sets in	, leaving behind eroded s
28	a long time until fatigue	sets in	. If a person continual
29	ales odorous air, fatigue	sets in	, due to the adaptation o
30	oils. If battle fatigue	sets in	, Mr Biedenkopf may have
31	e the spring thaw finally	sets in	, April can provide excel
32	long the chain the fixity	sets in	which determines a parti
33	a good-going winter gale	sets in	, a week or more may pass
34	stroys a kidney, gangrene	sets in	, now he's dead. Then t
35	's dead. Then the guilt	sets in	. If the body temperatu
36	w 35?C (95?C) hypothermia	sets in	. Thus it seems that fo
37	nce and control of the id	sets in	, associated with the dev
38	til sustained improvement	sets in	. However in practice d
39	ation of stock inevitably	sets in	, through a variety of fa
40	ed, and chronic infection	sets in	, there may be fibrous ti
41	professional insecurity'	sets in	. Institutionalism set
42	s in. Institutionalism	sets in	after a few months, and
43	case where one misfortune	sets in	train a whole sequence o
44	The north-east monsoon	sets in	during October. In all
45	s: the North-East monsoon	sets in	during October, the Sout
46	a sort of group mourning	sets in	which allows those peopl
47	in time, before necrosis	sets in	, he will be HIV-positive
48	Because when the night	sets in	, strange things happen o

¹¹ Micro-Concord, un concordancier de Mike Scott et Tim Johns, OUP.

49 the vital interview panic **sets in** . The emphasis on athle
50 f sharp passes, and panic **sets in** . Be on the look-out wh
51 hat I find is, once panic **sets in** w and that's usually eit
52 look-out when Pelagianism **sets in** : 'Thou shalt be justifie
53 bazaar! Procrastination **sets in** : jobs that ought to be d
54 ficult. Boredom quickly **sets in** . Umbrellarama... a for
55 igh winds - when the rain **sets in** , and especially if the w
56 cloak of grey as the rain **sets in** on finals day. This me
57 lexibility of the re-heat **sets in** meeting the more variabl
58 id. That's when reality **sets in** .' There is just time t
59 the winter weather really **sets in** . Managing director Gor
60 expansion as the recovery **sets in** . The secret of this st
61 e certainty of relegation **sets in** . WHEN THE worst thing
62 a certain kind of relief **sets in** . Despite my many and v
63 this is where the ritual **sets in** . Er so that gives you
64 pted, the statistical rot **sets in** . The danger area is at
65 l and air meet, where rot **sets in** and where a wind-blown s
66 years before the dry rot **sets in** . Here the smooth train
67 ops crack on as the storm **sets in** over So Kon Po. After
68 r pension, panic suddenly **sets in** as they consider the pro
69 r salary. Panic swiftly **sets in** when our boys twig they
70 masking tape, but tension **sets in** with the arrival of Clar
71 han most waters once thaw **sets in** . Head for tidal boatya
72 ykes and basins when thaw **sets in** , as excellent sport expe
73 a. West improve as thaw **sets in** Should revive quicker
74 sport expected. It then **sets in** train a labour process i
75 then the vertigo of time **sets in** , Seats and tables shou
76 re the bad winter weather **sets in** (hopefully!). We suspe
77 the warmer, drier weather **sets in** it is showing signs of d
78 ve before the bad weather **sets in** , especially if you have
79 r batteries before winter **sets in** once more. As winter
80 s the worst of the winter **sets in** , the goats move down to
81 n once more. As winter **sets in** , go to your local riding
82 settle down before winter **sets in** . No, I told myself, as
83 hem before a harsh winter **sets in** . The thing that I'm co
84 homeless people as winter **sets in** . Kurdish people are ha
85 efore the Armenian winter **sets in** with temperatures of 40
86 ur roof before the winter **sets in** . Finally, as the worst

On constate que l'élément séminal (le plus fréquent) est *winter*, 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 plus 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, ...*] + set in
=> [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 très 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*. Enfin, il peut y avoir attribution en discours de caractéristiques négatives au sujet. 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 type de processus est très fréquent, par exemple la création de classes en discours à l'aide de l'article défini (*we watch **the** morning swimmers do their lengths*, où l'article *the* crée une classe de nageurs matinaux en discours) ou l'allocation dynamique de caractéristiques sémantiques, comme dans *elle lui a révélé son origine*, où l'on comprend que l'origine était secrète¹².

¹² Il est à noter que les sémantiques qui reposent sur la résonance sémique entre les mots, comme la sémantique interprétative de F. Fastier, 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 (voir Frath 2004).

Les collocations dénominatives verbales fonctionnent ainsi de la même manière que les collocations dénominatives nominales, avec l'emploi d'un groupe lexical référentiel constitué dans la langue, dont l'un des mots est un élément séminal capable d'engendrer un paradigme lexical. Y a-t-il toujours un élément séminal ? Nos recherches semblent pour l'instant le montrer, mais il est encore trop tôt pour l'affirmer. Egalement, il semble que les éléments séminaux ne fonctionnent pas tous de la même manière, certains étant plus productifs que d'autres, avec la productivité liée à la force du figement (cf *cheveux noirs* v. *messe noire*).

Autres collocations dénominatives: *spill the bean*, *kick the bucket*, *kick the habit*

Les deux collocations dénominatives que nous venons d'analyser sont donc construites autour d'un pivot (*set in*, *tea*), avec un paradigme organisé à partir d'un élément séminal (*winter*, *strong*). Nous en examinons maintenant rapidement trois autres, *spill the beans* (révéler le secret, avouer), *kick the bucket* (mourir) et *kick the habit* (abandonner une habitude).

Le BNC révèle que l'expression *spill the beans* existe aussi sous la forme *spill it* et dans un autre corpus, nous avons trouvé *spill the details*. *The beans*, *it* et *the details* constituent ainsi une sorte de paradigme anaphorique qui renvoie au secret dont le dévoilement est exigé ou effectué, avec *the beans* comme élément séminal. *Spill the beans* se distingue ainsi de *kick the bucket*, une expression toute faite, qui n'admet aucun paradigme ni à droite, ni à gauche, dont *bucket* ou *kick* seraient les éléments séminaux. Cette expression réfère donc uniquement de manière globale, car ni *kick* ni *bucket* ne réfèrent à quoi que ce soit séparément. *Kick the habit* se construit selon le schéma séminal, puisque le BNC montre l'existence de tout un paradigme de mauvaises habitudes (*habits*) qui peuvent être abandonnées (*kicked*), par exemple *kick the drug habit* ou *kick the drug*. Dans ce dernier cas, le locuteur n'a pas précisé *habit*, sans doute parce qu'une des caractéristiques principales de la drogue est d'être une mauvaise habitude dont il est souhaitable de se débarrasser. Dans les exemples suivants, *kick the company-car habit* et *kick the mainframe habit*, le mot *habit* est quasiment obligatoire, sinon on risquerait de comprendre *kick* dans son sens de *donner un coup de pied*, respectivement à une voiture de service et à un ordinateur de grande taille. En revanche, dans la phrase *teaching her to kick the booze*, *the pick-ups*, *the self-destruction*, *the money*, le caractère addictif du premier mot (*booze* – boisson) implique que les mots suivants soient envisagés sous le même angle, celui de mauvaises habitudes dont il faut se débarrasser.

Typologie des unités phraséologiques

Certaines expressions très figées réfèrent de la même manière que des dénominations monolexicales, c'est-à-dire d'un seul tenant. C'est le cas de *kick the bucket*, dont la seule différence avec des dénominations comme *tea* ou *psychanalyse* est qu'elle accepte la conjugaison, là où les noms acceptent les désinences du pluriel. Il s'agit là de la première catégorie d'unité phraséologique (UP) de notre typologie, que nous appelons **UP lexicales**.

Une deuxième catégorie comprend des expressions comme *messe noire*, *colère noire*, ou *spill the beans*, qui acceptent un paradigme commutatif restreint construit avec parcimonie à partir de l'élément séminal. Dans certains cas, ce paradigme peut être anaphorique (*spill the beans*). Nous appelons cette catégorie les **UP semi-figées**

Une troisième catégorie comprend les UP que nous allons qualifier d'**UP ouvertes**, comme par exemple *kick the habit*, *winter set in*, *cheveux noirs* ou *strong tea*. Elles ouvrent un paradigme

commutatif parfois scalaire (*strong*) dans le voisinage du pivot. Ce paradigme est également construit à partir d'un élément séminal, mais de manière beaucoup moins parcimonieuse, et que l'on identifie en termes de fréquence d'usage. Les éléments du paradigme sont liés sémantiquement à l'élément séminal.

Une micro-syntaxe très liée à la sémantique est à l'œuvre dans les UP semi-figées et ouvertes. Selon le point de vue développé ici, un énoncé est constitué d'éléments référentiels à différents stades de figements, avec la syntaxe qui ne s'occupe que des détails de l'agencement et non du tout selon des critères très généraux. On peut pousser cette logique plus loin vers des énoncés encore plus ouverts, telles des phrases discursives, dont les structures sémantiques et syntaxiques comprennent certainement des éléments tout faits. Des analyses sur corpus de mots très fréquents comme *want* montrent qu'ils ont eux aussi des "*préférences*" lexicales, et spécifiquement, pour *want*, le Collins-Cobuild mentionne *want a baby*. Le "*choix*" dans les expressions non figées n'est peut-être pas si "*ouvert*" que cela. Il y aurait ainsi un continuum entre le mot simple et l'expression totalement figée d'une part, et la phrase d'autre part, construite par agencement et entrecroisement d'éléments référentiels plus ou moins lexicalisés, plus ou moins figés, ordonnés au niveau de la phrase par des régularités syntaxiques mémorisées¹³. La langue est faite d'habitudes acquises par l'apprentissage et par l'usage, et il serait certainement intéressant d'étudier la syntaxe dans ce cadre-là. Ce qu'elle perdrait en hégémonie, elle le gagnerait en termes de relations avec le sens lexical et le sens du texte. En outre, elle serait beaucoup moins métaphysique.

Définition de l'unité phraséologique

Qu'est-ce donc qu'une unité phraséologique, finalement ? Son mystère réside dans son figement variable, qui la distingue des mots simples et des phrases discursives. Elle intrigue en raison de l'habitude que nous avons de considérer les phrases et les syntagmes comme des constructions syntaxiques dans lesquelles des mots viennent s'insérer. Si on abandonne cette idée, l'unité phraséologique perd son mystère et devient ***une entité référentielle à géométrie variable*** au sein d'un continuum d'expressions référentielles qui vont de l'unité lexicale à la phrase en passant par l'unité phraséologique, puis au paragraphe et au texte en entier.

Références

- Boas Hans Ulrich (2003) : "Constructional productivity in English", in Proceedings of ESSE6, Ranam n°36/2003, P. Frath & M. Rissanen, eds.
- Bréal Michel (???) : *Essai de sémantique*. ???
- 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 National University, South Korea.
- Frath Pierre (2005a) : *Signe, référence et usage* (à paraître).
- Frath Pierre (2005b) : "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 (2005c) : "Post-cognitivism: a Plea for Reference in Linguistic Theory", *Proceedings of the Third International Workshop on Generative Approaches to the*

¹³ Plusieurs exemples pris dans la littérature française et britannique sont donnés dans Gledhill & Frath 2005 et Frath & Gledhill 2005.

- Lexicon*, Pierrette Bouillon and Kyoko Kanzaki, eds., University of Geneva, May 19-21, 2005.
- Frath Pierre & Gledhill Christopher (2005) : "Free-Range Clusters or Frozen Chunks? Reference as a defining criterion for linguistic units", in *RANAM (Recherches Anglaises et Nord-Américaines)*, n° 38, 2005, Strasbourg.
- Gledhill Christopher & Frath Pierre (2005) : "Une tournure peu en cache une autre : l'innovation phraséologique dans Trainspotting", *Les Langues Modernes*, n°3/2005.
- Gross Gaston (1996) : *Les expressions figées en français*, Ophrys.
- Hun-tak Lee, Thomas (2002) : "Two Types of Logical Structure in Child Language", in *Journal of Cognitive Science*, Vol. 3 N°2 December 2002 (published in 2004), Seoul, Korea
- Kleiber Georges (1994) : *Nominales. Essais de sémantique référentielle*, Armand Colin.
- Kleiber Georges (2001) : "Remarques sur la dénomination", in *Cahiers de praxématique*, 36-2001, Montpellier III, Bernard Bosredon, Irène Tamba et Gérard Petit, *dirs*, pp 21-41.
- Nesselhauf, Nadja (2003) 'The Use of Collocations by Advanced Learners of English and Some Implications for Teaching' in *Applied Linguistics* 42/2 : 223-42.
- Martin Robert (1976) : *Inférence, antonymie et paraphrase*, Klincksieck, Paris.
- Mel'cuk Igor (1988) *Dictionnaire Explicatif et Combinatoire du français contemporain*. Recherches lexicographiques II. Presses de l'Université de Montréal, Montréal.
- Peirce Charles Sanders (1978) : *Ecrits sur le signe*, rassemblés, traduits et commentés par Gérard Deledalle, Seuil.
- Peirce Charles Sanders (1984) : *Textes anticartésiens*. Aubier.
- Rorty Richard (1996) : "Le parcours du pragmatiste", in *Interprétation et surinterprétation*, U. Eco, coord., PUF 1996, CUP 1992.
- 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.
- Van Roey J. (1990) : *French-English Contrastive Lexicology: An Introduction*. Louvain-la - Neuve: Peeters.
- Wittgenstein Ludwig (1953, 1963) : *Philosophical Investigations*, translated by G.E.M. Anscombe, Basil Blackwell.