



HAL
open science

Introduction

Olivier Kraif, Agnès Tutin

► **To cite this version:**

Olivier Kraif, Agnès Tutin. Introduction. Cahiers de Lexicologie, 2016, Phraséologie et linguistique appliquée, 1 (108). <hal-01884939>

HAL Id: hal-01884939

<https://hal.science/hal-01884939v1>

Submitted on 9 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

INTRODUCTION

Olivier Kraif, Agnès Tutin
LIDILEM
Université de Grenoble Alpes

1. Phraséologie et linguistique appliquée

Ce volume des *Cahiers de Lexicologie* reprend un sous-ensemble de communications présentées lors du colloque *Rencontres Phraséologiques : des collocations aux séquences figées*, qui s'est tenu en novembre 2013 à Grenoble. Il résulte d'une sélection de plusieurs contributions plus spécifiquement consacrées aux aspects applicatifs des études phraséologiques, qu'il s'agisse de l'analyse des discours et des genres, de l'acquisition du langage, de la didactique des langues ou du traitement automatique des langues. Un autre volume issu des communications de ce colloque, intitulé *La phraséologie : clarifications et tendances actuelles*, à paraître chez Honoré Champion, coordonné par Francis Grossmann, Salah Mejri et Inès Sfar, aborde le sujet de la phraséologie française, de ses contributions dans le domaine de la phraséologie générale et de ses orientations et problématiques privilégiées.

Le thème de la phraséologie a le vent en poupe dans le domaine de la linguistique francophone, comme en témoigne la multiplicité de publications récentes, par exemple, pour n'en citer que quelques-uns : le numéro 189 de *Langages* de 2013, coordonné par D. Legallois et A. Tutin, intitulé « Vers une extension du domaine de la phraséologie », le numéro 159/160 de *Pratiques* de 2013 coordonné par L. Perrin, « Le figement en débat : figement linguistique et défigement discursif » ou l'ouvrage *Stéréotypie et figement : À l'origine du sens*, paru aux PUM en 2015 et coordonné par Vladimir Beliakov et Salah Mejri. La visée applicative semble également préoccuper les chercheurs, et ceci au-delà du domaine francophone : nous citerons la parution prochaine d'un ouvrage coordonné par Adriana Orlandi chez Peter Lang intitulé *Defining collocations for lexicographic purposes: from linguistic theory to lexicographic practice*, qui s'intéresse à la mise en œuvre lexicographique des concepts et aux définitions linguistiques de la collocation.

Depuis quelques années, le champ de la phraséologie se diversifie, comme le montrent les articles de ce volume, dans plusieurs directions. Les méthodes d'investigation sont désormais largement basées sur l'usage et exploitent de grands corpus traités par des méthodes numériques. Les objets d'étude s'étendent au-delà des expressions figées et s'intéressent également à des séquences préconstruites,

mais souvent compositionnelles, comme les collocations, les routines, les expressions conventionnelles pragmatiques ou des schémas lexico-syntaxiques. Les approches linguistiques sont également plurielles, puisqu'elles entretiennent des liens avec l'analyse de discours ou les grammaires de construction, au-delà de la sémantique et de la lexicologie. Le champ phraséologique apparaît dès lors comme un espace largement ouvert, et traversé par de multiples tensions dialectiques : zone grise où la syntaxe et le lexique se disputent des territoires partagés, zone de conflit entre principe de l'idiome et principe du libre choix, point de rencontre entre lexicologie et linguistique textuelle, espace de dialogue entre le contextualisme britannique et la phraséologie dite continentale, lieu de recouvrement entre critères linguistiques (restriction combinatoire, opacité sémantique) et critères purement statistiques et quantitatifs, espace continu ou fractionné en paliers discrets... Ce champ-là, qui n'est pas un champ de bataille, semble néanmoins voué à la dissolution des frontières, à la déconstruction des objets et à la remise en question des systèmes. Ce caractère problématique explique peut-être que des phénomènes initialement abordés comme périphériques et confinés dans les marges puissent aujourd'hui devenir centraux dans différents domaines d'application. Le traitement de la polylexicalité constitue en effet un véritable défi pour la structuration des dictionnaires, un obstacle à franchir pour la maîtrise d'une langue seconde, un ensemble de normes inconscientes à connaître pour s'intégrer dans une communauté de discours, un verrou technologique pour le traitement automatique des langues... Qu'il s'agisse de lexicographie, de didactique des langues, d'analyse du discours ou de TAL, on réalise aujourd'hui que la prise en compte des phénomènes phraséologiques constitue un problème *sui generis* qu'il est temps d'aborder de front. C'est dans cette perspective applicative et étendue que ce volume a pour ambition de rendre compte de travaux récents sur des thématiques émergentes dans ce domaine de la phraséologie, en abordant des thèmes et des objets encore peu étudiés dans ce domaine. Les objets originaux et les approches concrètes issues de ces terrains variés de la linguistique appliquée permettront peut-être, par la suite, de recomposer le champ théorique, en donnant enfin à la phraséologie la place centrale qui lui revient.

2. Présentation des contributions

Une première série d'article aborde la question de la phraséologie et du genre, dans les écrits littéraires romanesques, les corpus économiques ou le discours politique.

La phraséologie littéraire a souvent été étudiée chez des auteurs spécifiques. **Dirk Siepmann** a ici un objectif tout différent : en prenant en considération de gros corpus romanesques du français et de l'anglais (respectivement de 160 et 26 millions de mots), il a pour objectif de montrer que la phraséologie récurrente des romans

contemporains contribue à l'impression de « littérarité » du genre. À l'aide d'un corpus de contraste varié, il parvient à mettre en évidence la surreprésentation du lexique des mots simples spécifiques du genre en anglais et en français, par exemple les champs sémantiques des émotions et sentiments ou des verbes de parole, qui apparaissent d'ailleurs assez semblables dans les deux langues. Il y a donc bien un champ lexical propre au genre « où l'on sourit, rit et pleure avec une intensité accrue, où l'on perçoit plus de portails et de cheminées que dans la vraie vie [...] où l'on ne parle guère de commerce, industrie ou histoire et où les points cardinaux n'existent pas [...] ». L'examen des segments répétés et collocations révèle également clairement des « ingrédients de base » du langage romanesque, comme le style verbal ou la présence importante de comparaisons. Des « patrons narratifs », définis comme des cooccurrences récurrentes ayant une fonction narrative et/ou descriptive, peuvent être mis en évidence. Les collocations du genre romanesque, dont quelques exemples extraits à partir de la base *silence* sont présentés, présentent plus de proximité pour le français avec les autres genres écrits qu'en anglais. De façon intéressante, l'auteur montre aussi que, malgré les condamnations fréquentes des collocations stéréotypées, « les ressemblances entre littérature et paralittérature l'emportent sur les différences ». L'auteur aborde enfin la phraséologie autour des noms *soleil/sun* dont il relève des usages stéréotypés assez comparables dans les deux langues malgré des constructions grammaticales différentes, ainsi que des spécificités remarquables liés aux différents sous-genres (science fiction, par exemple).

C'est également dans le cadre d'un genre textuel spécifique, celui de la presse économique, que s'inscrit l'étude de **Lucia Rosenbaum Franková**. L'étude aborde la phraséologie des langues de spécialité de façon originale, en mettant en évidence les procédés syntaxiques et sémantiques réguliers. L'étude porte sur les « phrasèmes spécialisés » définis comme des séquences qui peuvent avoir une valeur terminologique mais aussi comme des syntagmes nominaux ou verbaux qui sont composés à la fois de mots de la langue générale et de termes. Une étude fine du phrasème spécialisé *plonger dans le rouge* montre bien qu'il s'agit d'une collocation qui se décompose en deux éléments : le verbe *plonger*, peut commuter avec une classe paradigmatique (Bolly 2011), alors que le syntagme *dans le rouge* apparaît, lui, fortement figé. L'examen est ensuite élargi au patron verbe + préposition + Nom de couleur. L'auteur montre l'existence d'une collocation productive dans le champ économique (*plonger/s'enfoncer dans le rouge ; passer/virer au vert*), et de métaphores fréquentes associées aux verbes de mouvement.

La contribution de **Wojciech Prażuch**, qui se situe dans la tradition française de l'analyse textométrique du discours politique, s'intéresse aux « mots de la crise ». Elle cherche à mettre en évidence, à travers l'entrée phraséologique, les changements du discours économique pré- et post-crise, en comparant les périodes 1981-2007 et 2008-2012. La méthode consiste principalement à observer les cooccurrences et les segments répétés d'un corpus diversifié de discours politiques d'1,3 M de mots pour

chaque tranche temporelle (discours de campagne, allocutions aux congrès des partis, discours du G20, ...). L'évolution fait apparaître quelques changements remarquables dans les paradigmes économiques. Certains termes de gauche comme *ouvrier* ou *populaire* tendent à disparaître alors que les termes *gouvernance*, *mobilité* et *flexibilité* apparaissent sur le devant de la scène. L'opposition entre *économie sociale/économie de marché* est supplantée par la dichotomie *économie réelle/virtuelle*. Une certaine critique du système semble transparaître à travers le paradigme de la nouveauté (*nouveau système monétaire, nouvel ordre, redéfinir* ...). Par ailleurs, l'étude phraséologique permet également de mettre en évidence un ensemble de métaphores productives, d'hyperboles fréquentes et une euphémisation croissante de la réalité économique. Bien que, comme le signale l'auteur « un discours ne pourrait se limiter à une série de formules et de séquences polylexicales relevant du "déjà dit" », l'entrée phraséologique permet d'approcher quelques faits « statistiquement saillants [qui] échappent au hasard » et qui permettent de mettre en évidence, au bout du compte, des modifications dans le discours économique qui ne sont pas radicales.

Le deuxième série de contributions aborde le traitement de la phraséologie en traitement automatique des langues.

C'est aux aspects lexicographiques et lexicaux que s'intéresse **Aude Grezka** à travers l'élaboration de la base FixISS, une base d'expressions adverbiales qui comporte toutes les formes possibles des expressions figées. Les expressions polylexicales présentent en effet des propriétés remarquables, dont il faut rendre compte dans le système de TAL, en particulier la variabilité syntaxique, morphologique, lexicale à côté de certaines formes de figement. Ces expressions comportent des difficultés importantes pour le TAL, en particulier pour la traduction automatique dont Aude Grezka montre la complexité à travers quelques exemples. Il n'est ainsi pas possible de traduire mot à mot des expressions comme *bête comme ses pieds* vers l'espagnol, et la plupart des systèmes de TA en ligne comme Reverso ou Systran ne donnent pas des résultats complètement satisfaisants. Il est ainsi essentiel de décomposer ces expressions, comme dans la base FixISS, qui inclut un nombre importants d'expressions adverbiales (6000) accompagnées d'informations syntaxiques et sémantiques structurées en XML, dans un codage inspiré de Odijk (2013). Des exploitations de cette base ont été proposées, en particulier pour la traduction et pour l'aide à la rédaction monolingue, à l'aide d'un prototype d'interface.

Une autre application importante du TAL dans le domaine de la phraséologie porte sur le repérage et l'extraction dans les textes des expressions polylexicales. C'est sur ce thème que porte la contribution d'**Olivier Kraif** qui présente un outil d'extraction des séquences phraséologiques basé sur des corpus arborés, le Lexicoscope. L'auteur montre d'abord l'intérêt des corpus arborés pour l'extraction de la phraséologie, ressources encore peu utilisées dans les outils en ligne. Le Lexicoscope permet d'extraire plusieurs types de séquences phraséologiques, tout

d'abord les collocations, avec des calculs de cooccurrences lexico-syntaxiques dans des *lexicogrammes*, mais il permet aussi d'avoir une vue générale des cooccurents à travers une analyse multivariée. Il permet également de comparer des cooccurents à travers l'analyse factorielle des correspondances, d'une classification hiérarchique ou d'échelonnement multidimensionnel. Un langage de requête a été développé, ainsi qu'un système de reformulation, qui facilite l'utilisation pour les non-experts. Enfin, une fonctionnalité assez nouvelle du système est la possibilité d'extraire des « arbres lexico-syntaxiques récurrents », permettant de repérer non seulement des cooccurrences lexico-syntaxiques binaires, mais également des expressions plus larges intégrant des éléments reliés syntaxiquement (et non nécessairement contigus), par exemple, à partir de *entrer dans colère*, l'expression *entrer dans une violente colère*. L'observation des propriétés syntaxiques sur les corpus permet d'en affiner les propriétés, qui sont à prendre en compte dans des bases d'expressions polylexicales. Ainsi, on pourra observer les alternances syntaxiques de collocations comme *réaliser -OBJ-> analyse*, dont la réalisation au passif est plus fréquente dans les écrits scientifiques (*analyse réalisée, l'analyse est réalisée*) que les réalisations à l'actif. L'outil permet ainsi d'établir pour une expression phraséologique donnée ses propriétés syntaxiques, morphologiques et combinatoires.

La dernière série de contributions porte sur l'acquisition des langues secondes et sur la didactique de la phraséologie.

Amanda Edmonds s'intéresse à la dimension psycholinguistique des unités phraséologiques dans l'acquisition des langues secondes, et notamment à la question de leur inscription mémorielle, comme unités à part entière. Partant du principe que les unités phraséologiques sont préconstruites, elle note que de nombreuses expériences ont mis en évidence un avantage psycholinguistique dans l'accès à ces unités, plus rapidement appréhendées par les locuteurs que des combinaisons libres de longueur équivalente. Cet avantage est interprété, chez de nombreux auteurs, comme une confirmation de l'hypothèse lexicaliste, selon laquelle les unités phraséologiques sont enregistrées en bloc dans le lexique mental. Par une nouvelle expérience qui vient compléter ses observations de 2013, Edmonds remet en cause cette hypothèse, du moins pour une certaine classe d'unités phraséologiques étroitement liées au contexte communicationnel, qu'elle nomme « expressions conventionnelles ». Ces expressions, telles que *avec plaisir, c'est normal, ne t'inquiète pas*, sont fonctionnellement liées à des facteurs pragmatiques, et utilisées dans des contextes d'énonciation bien déterminés. L'expérience de 2013 permettait de mettre en évidence un véritable avantage psycholinguistique de ces expressions présentées en contexte (dans des situations jugées typiques), pour un groupe de locuteur natifs par opposition aux locuteurs non natifs, qu'ils soient établis en France depuis longtemps ou non. Dans la nouvelle expérience présentée ici, la description des contextes d'énonciation n'est plus fournie, et l'avantage psycholinguistique disparaît : cette fois, les expressions conventionnelles ne sont pas

décodées plus vite que les substituts en combinaison libre, ce qui tend à infirmer l'hypothèse lexicaliste. Pour Edmonds, ces observations renforcent l'hypothèse concurrente de l'amorçage pragmatique, « selon laquelle un stock d'expressions serait activé ou déclenché par une situation en fonction des besoins communicatifs ». Par rapport aux collocations ou aux expressions figées, les expressions conventionnelles feraient donc exception, car le contexte communicationnel est déterminant dans leur reconnaissance en tant qu'unités.

La question de l'encodage des unités phraséologiques dans le lexique mental entre en résonance avec les préoccupations didactiques. Comme le notent de nombreux auteurs, la phraséologie constitue la part la plus difficile à maîtriser, pour un apprenant, du lexique de la langue – notamment en ce qui concerne les collocations du fait de leur caractère semi-figé Nesselhauf (2005). L'étude présentée par **Veronica Benigno, Olivier Kraif et Francis Grossmann** se situe dans ce champ applicatif, et propose d'élargir au domaine de la phraséologie la notion de *vocabulaire fondamental*, très largement balisée tout au long du siècle dernier avec Thorndike (1921), Ogden (1930), Vander Beke (1929), West (1953), puis avec les travaux fondateurs de Gougenheim sur le français qui remontent aux années 1950-1960. Ainsi, le concept de « collocation fondamentale » vient combler une lacune de ces études qui « apparaissent incomplètes pour avoir négligé l'analyse des associations typiques et fréquentes qui opèrent en tant qu'unité lexicogrammaticales ». Après une définition théorique des contours de son objet, selon une conception des collocations proches des définitions de Hausmann (1979) ou Tutin et Grossmann (2002), l'article se propose d'identifier la corrélation entre une évaluation subjective du caractère fondamental d'une série de collocations, effectuée auprès d'un échantillon de locuteurs, avec des critères objectifs tels que la fréquence et diverses mesures d'associations (information mutuelle spécifique, t-score ou rapport de vraisemblance). Ce croisement entre deux approches complémentaires, enquête intersubjective d'une part et observations sur corpus d'autre part, permet de tirer des conclusions à la fois sur le plan linguistique et méthodologique. D'une part, elle confirme l'intuition de Gougenheim que le caractère fondamental d'une unité, qu'il s'agisse d'un mot simple ou d'une collocation, n'est pas uniquement lié à sa fréquence – mais aussi à d'autres formes de saillance telles que « l'utilité communicative » – ce qui fait écho à l'étude d'Edmonds sur l'enracinement des unités dans le contexte communicatif. D'autre part, elle montre que les mesures objectives adaptées pour observer des collocations sur de grands corpus (ici de l'ordre du milliard de mots), présentent des profils particuliers : l'information mutuelle ou le *t-score*, très utilisés en phraséologie, paraissent inadaptés pour faire ressortir les phénomènes de haute fréquence, tandis que le rapport de vraisemblance semble biaisé par les récurrences thématiques liées à la constitution du corpus (trop resserré dans le temps). Le *z-score* et la fréquence brute apparaissant au final comme plus pertinentes, à condition de les associer à une mesure de dispersion, permettant d'éliminer les récurrences trop localisées dans quelques documents du corpus.

Outre le caractère fréquent ou fondamental des unités phraséologiques, deux autres aspects quelque peu paradoxaux sont de plus en plus souvent mis en évidence : leur productivité et leur régularité. On a trop longtemps enfermé les phrasèmes dans la gangue du figement et de l'idiosyncrasie, et contre cette conception **Maria Isabel Gonzáles-Rey** s'interroge : « comment faire [...] pour éviter la marginalisation des [unités phraséologiques] dans la section du vocabulaire comme des éléments pittoresques, alors qu'elles se trouvent au centre de la langue ? » Pour y répondre, elle nous propose de partir d'une réflexion théorique interrogeant les liens entre les grammaires de construction et la phraséologie. Citant Dobrovols'kij (2011), elle note que si toutes les constructions ne peuvent ressortir à la phraséologie, on peut toutefois considérer tous les phrasèmes comme des constructions, dotées de propriétés syntaxiques et lexicales spécifiques. Partant, on peut se demander ce que peut apporter le paradigme des grammaires de construction à la phraséologie : la réponse de Gonzáles-Rey se situe – comme l'article précédent – dans le domaine didactique. Les grammaires de construction permettent en effet d'appréhender les expressions polylexicales au sein d'une hiérarchisation, allant de schémas abstraits génériques (p.ex. N1 V N2 V) à des constructions spécifiques lexicalisées suivant des parcours différents (*N1 a plaisir à V, N1 prend plaisir à V, N1 a le plaisir de V*). Cette hiérarchisation peut déboucher sur un scénario didactique axé sur la confection de cartes heuristiques ou cartes mentales. Inspirées de Legallois (2014), ces cartes permettent de rattacher des expressions phraséologiques différentes à des schémas constructionnels sous-jacents, et de mettre en évidence le *continuum* lexico-syntaxique qui s'étend entre des expressions dites libres et des expressions plus figées, par exemple : *quelqu'un fuit, quelqu'un prend la fuite, quelqu'un prend ses jambes à son cou, quelqu'un prend ses cliques et ses claques*. De la sorte, on permet aux apprenants de saisir les deux faces lexicales et syntaxiques de l'objet phraséologique, autrement dit « d'affronter les constituants, d'une part, et leur combinaison, d'autre part » et de « joindre les deux aspects dans une didactique de la LE qui généralement sépare les deux ».

On ne saurait cependant réduire la problématique phraséologique à cette dialectique entre lexique et syntaxe : la perspective appliquée des travaux présentés dans ce numéro commande d'élargir le champ d'étude à la dimension discursive, notamment en se penchant sur les fonctions bien spécifiques jouées par les unités phraséologiques dans l'enchaînement textuel. Dans cette optique, **Thi Thu Hoai Tran, Agnès Tutin et Cristelle Cavalla** s'appuient sur une étude de corpus pour dresser une typologie détaillée des séquences lexicalisées à fonction discursive dans le genre des écrits scientifiques de sciences humaines. En se fondant sur le plan fonctionnel, elles distinguent deux grandes classes de marqueurs discursifs (MD), les MD à fonction métadiscursive (marqueurs d'enchaînement thématique, de topicalisation, de reformulation, d'exemplification, de comparaison) et les MD à fonction argumentative (marqueurs de causalité, de concession, d'opposition, de condition, de finalité). 171 MD ont été relevés dans ses 10 groupes, parfois subdivisés en sous-groupe (p.ex. addition, structuration, énumération, conclusion

pour l'enchaînement thématique). L'objectif de cette description est de déboucher sur une application didactique, à savoir l'apprentissage de l'écrit scientifique dans le contexte du Français sur objectif universitaire (FOU), la maîtrise de ces séquences ayant été identifiée comme une compétence clé pour ce type de production. À cette typologie, les auteurs ajoutent une grille d'analyse linguistique faisant intervenir les contextes gauche et droit des séquences, afin de montrer aux enseignants les trois paramètres linguistiques autour desquels ils peuvent centrer les activités pédagogiques : position syntaxique (p.ex. initiale ou médiane), portée syntaxique (p.ex. simple syntagme ou proposition) et valeurs sémantiques (p.ex. généralisation ou spécialisation, dans le cas d'une reformulation). Par cette description fine des phénomènes, les auteurs ouvrent la voie au développement de méthodes d'enseignement fondées sur une observation rigoureuse, l'analyse sur corpus permettant de remédier à des erreurs liées à l'intuition, ces séquences pouvant être trompeuses lorsqu'elles sont appréhendées hors contexte (cf. la distinction entre *c'est-à-dire*, *à savoir*, *en d'autres termes*). En s'inspirant des observations effectuées, les enseignants pourront commencer un travail de transposition didactique, par exemple en élaborant des activités de repérage ou de classification basées sur un métalangage simplifié.

Le dernier article de ce volume procède de la même posture méthodologique, consistant à partir d'une étude fine de corpus pour déboucher sur la confection d'activités didactiques. Dans cette démarche, **Laura Hartwell** s'appuie sur un vaste corpus de 33 millions de mots (issu de Scientext), comportant près de 8 900 publications d'articles de recherche en anglais dans les domaines de la biologie et de la médecine. Elle s'intéresse à des constructions verbales de la forme adverbe+verbe, qui apparaissent comme caractéristiques de l'anglais scientifique, où elles jouent un rôle de charnière entre deux propositions a et b pour exprimer la causalité ("a explains b"). Trois adverbes sont étudiés : *thus*, *therefore*, et *thereby*. L'objet étudié se rapproche, d'un point de vue fonctionnel, de la thématique des marqueurs discursifs abordée par Thi Thu Hoai Tran et al., avec les marqueurs d'explication et de conséquence. Mais ces trois constructions verbales sont abordées sous un autre éclairage : celui de la métaphore grammaticale, qui dénote la tendance, très fortement marquée dans l'anglais scientifique, à nominaliser les procès afin de stabiliser l'expression de concepts complexes. Cette notion de métaphore grammaticale, conceptualisée par Halliday (2004), indique un mode de construction spécialement adapté à l'activité scientifique, puisqu'il autorise la stabilisation de notions abstraites complexes. La composition nominale anglaise permet d'inclure de manière synthétique adjectifs, noms, adjectifs composés ("*intracellular reactive radicals*"), génitifs ("*blood's oxygen transport capacity*"), verbes en *-ed* ou *-ing* ("*marker assisted breeding*") en faisant abstraction de tous les relateurs (préposition, articles, conjonctions, connecteurs logiques) que Swales (2004) identifie comme les éléments les plus labiles de l'expression. Dans ce processus de stabilisation conceptuelle, la métaphore grammaticale permet de transformer un événement contingent tel que *cracks on the surface* en une entité complexe *surface cracks*.

L'usage des constructions *thus V-ing*, *therefore V-ed*, *thereby V-ing* peut s'interpréter dans ce cadre, celles-ci servant à articuler grâce à des verbes récurrents, généralement liés au lexique transdisciplinaire, des propositions a et b plutôt liées à des composés terminologiques stabilisés. L'étude de corpus permet d'identifier, pour chacune de ces expressions, les propriétés sémantiques des verbes les plus saillants (p.ex. verbes méthodologiques pour *therefore*, verbes exprimant des conséquences physiques après *thus*, verbes plus spécialisés et recherchés après *thereby*), les contextes rhétoriques typiques (p.ex. établir un lien entre les données et les hypothèses) et les traits grammaticaux (verbes en *-ing* après *thus* et *thereby*). Toutes ces observations sont ensuite réinvesties dans la confection de supports et d'activités, pour la plateforme Scientext Enigma, avec des exemples courts et représentatifs issus du corpus.

3. Conclusion

Les diverses contributions de volume, par la variété des terrains d'application dans lesquelles elles s'enracinent, montrent à quel point le territoire de la phraséologie est vaste, et encore largement inexploré. Cette difficulté à appréhender les phénomènes phraséologiques de façon panoramique trouve plusieurs explications. D'abord, contrairement aux domaines du lexique et de la grammaire, pris dans leurs acceptions traditionnelles, il semblerait qu'on ne puisse en cartographier que de petits territoires à la fois. L'observation des phénomènes dévoile de nombreuses régularités locales, des sous-systèmes lexico-syntaxiques, qu'on peut interpréter théoriquement en des termes différents mais analogues (métaphores grammaticales, constructions, collostructions, patterns, motifs...). Une autre difficulté inhérente à la description de ces sous-systèmes est liée au fait qu'ils ne peuvent être mis en évidence qu'à travers de vastes corpus, certaines régularités n'apparaissant que grâce à la récolte d'un très grand nombre d'occurrences (des centaines de millions de mots pouvant être requis, comme pour l'étude des sous-genre romanesques). Les instruments – et les corpus – permettant de saisir ces phénomènes de faible fréquence commencent seulement à se faire jour. Grâce à ces données et ces nouvelles méthodes d'observations, on peut espérer caractériser de mieux en mieux les propriétés combinatoires des unités phraséologiques, à la fois sur le plan de leur contexte lexico-syntaxique et sur leurs possibilités de variations internes. Une autre difficulté, enfin, est liée à l'intrication des niveaux descriptifs – relevée par Sinclair dans sa définition des *extended meaning units* : elles combinent les dimensions lexicales, syntaxiques, sémantiques, discursives (en lien avec le genre) voire pragmatique, le contexte communicationnel étant parfois intrinsèquement lié à la définition de l'unité phraséologique en tant que telle (par exemple, pour ce qui concerne les expressions conventionnelles). Ce sont donc des sous-systèmes multidimensionnels qui s'offrent à la description, où aucune composante ne peut-être escamotée : ni les propriétés combinatoires, ni les fonctions discursives, ni l'inscription dans un genre textuel ou une situation de communication

donnée. Pour la description et l'exploration de ces territoires au relief complexe et accidenté, les terrains applicatifs constituent un point d'entrée intéressant : qu'il s'agisse de lexicographie, de traduction, d'analyse du discours, de didactique des langues, d'acquisition ou de traitement automatique, ils ont besoin de s'appuyer sur des descriptions minutieuses pour résoudre des problèmes concrets. Ces travaux en linguistique appliquée, fondés sur de vaste corpus et des outils toujours plus performants, nous permettront sans doute d'avancer dans la compréhension du phénomène phraséologique pris dans sa généralité.

Bibliographie

BELIAKOV Vladimir et MEJRI Salah (2015) : *Stéréotypie et figement : À l'origine du sens*, Toulouse, PUM.

BIBER Douglas, CONRAD Susan et CORTES Viviana (2004) : « If you look at...: Lexical bundles in university teaching and textbooks », *English for Specific Purposes*, 23, p. 397-423.

BOLLY Catherine (2011) : *Phraséologie et collocations. Approche sur corpus en français L1 et L2*, Bruxelles, Peter Lang.

DOBROVOL'SKIJ Dimitrii (2011) : « Phraseologie und Konstruktionsgrammatik », in A. Lasch et A. Ziem (Hrsg.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, Tübingen, Stauffenburg, p. 111-130.

GOUGENHEIM Georges, MICHÉA René, RIVENC Paul P. et SAUVAGEOT Aurélien (1964) : *L'élaboration du français fondamental (1er degré)*, Nouvelle édition refondue et augmentée, Paris, Didier.

GROSSMANN Francis, MEJRI Salah et SFAR Inès (Ed)(à paraître) : *La phraséologie : clarifications et tendances actuelles*, Paris, Honoré Champion.

HALLIDAY M. A. K. (2004) : "Lexicology", in Halliday, M. A. K., Teubert, W., Yallop, C. & Čermáková, A. (Eds.) *Lexicology and Corpus Linguistics*, London, Continuum, p. 1-22.

HAUSMANN Franz, Joseph (1979) : « Un dictionnaire des collocations est-il possible ? ». *Travaux de linguistique et de littérature XVII* (1), Strasbourg, p. 187-195

LEGALLOIS Dominique (2014) : « La grammaire d'une langue peut-elle s'apprendre à partir de ses unités phraséologiques? », in M. I. González-Rey (dir.), *Outils et méthodes d'apprentissage en Phraséodidactique*, Cortil-Wodon, InterCommunications & E.M.E.

LEGALLOIS Dominique et TUTIN Agnès (Ed)(2013) : « Vers une extension du domaine de la phraséologie », *Langages 189*, mars 2013.

NESSELHAUF Nadja (2005) : *Collocations in a learner corpus*, Amsterdam/Philadelphia, John Benjamins.

ODIJK Jan (2013) : *Identification and Lexical Representation of Multiword Expressions*, in P. Spyns et J.E.J.M Odijk (Eds.), *Essential Speech and Language Technology for Dutch. Results by the STEVIN-programme* Berlin/Heidelberg: Springer.

OGDEN Charles Kay (1930) : *Basic English, a general introduction, with rules and grammar*, London, K. Paul.

ORLANDI Adriana (Ed) (à paraître) *Defining collocations for lexicographic purposes: from linguistic theory to lexicographic practice*. Berne, Peter Lang.

PERRIN Laurent (Ed)(2013) : « Le figement en débat : figement linguistique et défigement discursif », *Pratiques* , 159/160.

SALEM André (1987) : *Pratique des segments répétés. Essai de statistique textuelle*, Paris, Klincksieck.

SWALES John (2004) : *Genre analysis: English in academic and research settings*, Cambridge, Cambridge University Press.

THORNDIKE Edward L. (1921) : *The Teacher's Word Book*, New York Teachers College, Columbia University.

TUTIN Agnès et GROSSMANN Francis (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif, *Revue française de linguistique appliquée*, VII-1, p. 7-25.

VANDER BEKE G. E., (1929) : « French Word Book », dans *Publications of the American and Canadian Committees on Modern Languages*, vol. XV, New-York, The Macmillan Co.

WEST Michael (1953) : *A General Service List of English Words*, London, Longman.