

Intérêt et pièges de la Toile en tant que corpus pour la recherche en linguistique (sous l'angle de recherches sur la valence verbale)

Emmanuel DERONNE

ATILF UMR 7118 CNRS – Nancy-Université

It may seem a crazy bet or at least a challenge to exploit such a corpus as the Web for linguistic studies. We recognize the strange, nearly mysterious and indeed imperfect status of the Web corpus. This paper aims to expose, nevertheless, how it is possible to develop a not so complex questioning strategy to achieve new sophisticated syntactic and semantic descriptions of French verbs. The subject here especially treated is the verb alternations, including the lexical means. Even if we cannot easily appeal to statistical analysis to exploit the Web corpus, the Web is very useful to study the frequent variations, but gives also easier access to rare or new phrases, which reflect the advantages of such a mechanism of alternation: among others, it offers easy topicalisation, by the means of subjectivization, of rather many elements of a verbal sentence. This progress can also be precious for translation.

1. Introduction

Le développement du réseau Internet a d'ores et déjà un nombre important de conséquences pour les chercheurs en linguistique. La mise à disposition d'un nombre incommensurable de documents écrits ou même sonores dans de très nombreuses langues ne pouvait manquer de révolutionner la notion même de corpus.

Il s'agit donc, à cette étape encore relativement récente d'un processus important et en plein développement, de s'interroger sur les avantages et les inconvénients de ces documents en tant que supports de recherches linguistiques.

Personnellement, j'ai choisi de recourir à la Toile pour mes recherches linguistiques depuis 2006. C'est donc en tant qu'utilisateur, en tant que praticien (n'étant ni théoricien de la communication ni spécialiste d'informatique) que j'ai envisagé d'exposer le grand avantage que représente pour moi le fait de disposer d'un tel outil, mais aussi les nombreuses difficultés que l'on rencontre dans ces travaux encore laborieux et artisanaux. Il s'agira donc ici d'un témoignage élargi à une réflexion méthodologique.

Après avoir exposé sommairement en quoi le type de recherches que je mène impose le recours à la Toile, j'examinerai le corpus d'un nouveau genre constitué par la multitude de documents consultables grâce à Inter-

net sous ses deux dimensions quantitative et qualitative, en illustrant cette présentation par des exemples commentés de recherches sur différents verbes. Je terminerai par des considérations plus générales.

2. Les besoins d'une recherche sur la valence verbale et plus précisément sur les conversifs

Je rappellerai brièvement et caricaturalement que depuis les années 70, au-delà de l'analyse des *voix*, outil morphologique, les chercheurs de différentes obédiences se sont intéressés aux autres outils linguistiques permettant de faire varier la valence verbale et particulièrement le sujet du verbe.

On a exploré à ce titre la possibilité pour certains verbes de présenter des constructions syntaxiques inverses du type "La branche casse" vs. "Pierre casse la branche". La notion alors proposée de *verbes symétriques* est néanmoins à la fois trop restrictive parce que limitée à la description de duos de constructions et trop lâche dans ses exigences de relations sémantiques entre les verbes ainsi étiquetés (les deux phrases citées ne sont pas équivalentes sémantiquement).

On a étendu cette recherche aux verbes supports entrant dans des *constructions converses*, comme *donner et recevoir* (G. Gross, 1989), en opposant par exemple *donner des coups* à *recevoir des coups*.

Au-delà de cette opposition qui équivaut à celle de l'actif et du passif, on peut étendre ces recherches à tous les cas de variation sémantique du sujet. Ainsi, le verbe *voir* "peut perdre son sens de verbe de perception et permettre des relations de diathèse particulières, en particulier pour les circonstanciels de temps ou de lieu", comme dans "Saint-Malo a vu naître Chateaubriand" (Muller, 2005: 12). On peut dire que le verbe *voir* sert à former en français la *diathèse temporelle* ou *locale*, de même que la voix active et la voix passive des verbes d'action servent respectivement à former les *diathèses de l'agent* et *du patient*. On peut dire aussi que par ces moyens on change l'*orientation* de la phrase verbale. Le plus souvent, ce changement a pour fonction de permettre à un élément sémantique particulier d'accéder, à travers la fonction de sujet syntaxique, au statut de *thème*.

Il existe un certain consensus minimal autour de cette notion. Alain Lemaire (par exemple, 1989) et Claude Muller parlent de *diathèses* du verbe. Igor Mel'cuk (par exemple dans Mel'cuk, Clas et Polguère, 1995: 130-131) préfère, dans la lignée de Gaston Gross, parler de *conversifs*.

André Martinet souligne dans sa *Syntaxe générale* (1985: 216) que la langue française "fait feu de tout bois" dans ce domaine, expression très révélatrice de la variété des outils mis en œuvre par la langue française. Mon

objet d'étude est donc la recherche des moyens morphologiques ou lexicaux permettant au français d'exprimer différentes diathèses, c'est-à-dire de faire varier le thème d'une phrase verbale en lui accordant la fonction syntaxique de sujet.

Pour ces recherches, je m'appuie massivement sur le corpus de la Toile interrogé à partir du moteur de recherche *Google* (après une brève période de confrontation de ses résultats avec ceux du moteur *Yahoo*). On trouvera sur les archives Hal les deux longs articles (Deronne, 2008 a et b) qui exposent les fondements théoriques et décrivent les problèmes méthodologiques de ces recherches personnelles, qui ont par ailleurs fait l'objet d'une communication (Deronne, 2010) au dernier Congrès Mondial de Linguistique Française, à La Nouvelle Orléans.

Ainsi, le verbe *afficher*, dont j'ai décrit la variation (Deronne, 2010) dans le contexte de l'affichage par un thermomètre de la température ambiante ("afficher telle température, tant de degrés"), accepte comme sujets, outre le thermomètre, le lieu, la saison, etc.

Pour évoquer une de mes études en cours, outre *trouver* et *découvrir* qui ont pour sujet l'agent de la recherche (par exemple un archéologue menant des fouilles sur un site), on dispose en français (notamment dans ce domaine de l'archéologie) de verbes comme *livrer* qui permettent de thématiser d'autres éléments sémantiques que l'agent. On peut dire par exemple que "des fouilles ont *livré* trois ossements humains", que "tel site a *livré* deux statues d'Auguste" ou même que "la faune a *livré* les ossements de deux rhinocéros laineux". *Livrer* est le conversif (lexical) de *trouver* dans ces constructions.

Le verbe *embarquer*, phénomène naturel ou accidentel lors duquel de l'eau entre dans un navire au-dessus de la ligne de flottaison (cf. Deronne, 2008a, 2008b et 2010) se prête à lui seul, sans parler d'un jeu limité (mais néanmoins varié) de la voix passive, à plusieurs variations d'orientation. Il se dit en effet notamment du bateau qui subit ce phénomène ("le bateau a embarqué de l'eau par l'avant"), de l'eau qui pénètre dans le bateau, des hommes impliqués par ce phénomène en tant que présents sur cette embarcation (passagers ou plus souvent pilotes).

Je reviendrai sur le contexte archéologique du verbe *livrer* pour illustrer la méthodologie de ces recherches en cours, puis plus brièvement sur le verbe *embarquer* pour la variété des textes mobilisés et les types de progrès réalisables.

De telles précisions, qui devraient faire partie de la description syntaxique et sémantique des verbes du français et rendraient de précieux services aussi bien à la didactique du français qu'aux traducteurs, ne sont pas encore prises en compte par les dictionnaires (même si ici ou là certains de ces emplois sont décrits parmi d'autres, de façon non systématique) et

constituent donc un champ de recherche important, qui relève du cadre plus général de la description de la valence verbale.

Or de telles recherches nécessitent un corpus important. Les emplois verbaux très répandus (pensons au verbe *sembler*, conversif de *considérer comme* ou de *croire*, cf. latin *videri*, de forme passive) pourront s'enrichir d'une multitude de contextes. Les emplois spécialisés (pour avoir besoin de parler d'un procès sous plusieurs angles sont le plus souvent des spécialistes, amateurs ou professionnels) bénéficieront de la présence sur la Toile de textes nombreux et autrement très difficiles d'accès.

Pour repérer ces domaines spécialisés, il est d'ailleurs indispensable dans un premier temps de procéder à recherche très large, sans préjugés. Ainsi, le verbe *livrer*, dans le sens évoqué ci-dessus, aurait pu relever de l'enquête policière ou de l'exploitation minière, mais ce sont les fouilles archéologiques qui constituent nettement son domaine majoritaire, dans lequel on rencontre en outre de nombreux verbes voisins aux emplois également variés.

Soulignons également l'utilité des données diachroniques pour remonter l'écheveau de ces emplois multiples, ce qui ajoute encore à la variété du corpus nécessaire (et malheureusement très souvent insuffisants à ce jour) pour de telles recherches.

Par ailleurs, des calculs de fréquence pourraient permettre de confronter la configuration étudiée à des formulations analytiques ou synthétiques concurrentes.

Enfin, ces recherches ont besoin de textes intégraux (ou au moins de contextes très larges) pour évaluer le caractère thématique ou rhématique du sujet du verbe, disons plus exactement pour évaluer la dynamique des informations associée aux éléments de la construction étudiée.

La systématisation de ces recherches fines nécessite, on le voit, un corpus aux multiples qualités: étendu, varié, évaluable numériquement et pratique (fiable et facile à exploiter).

3. La Toile, un corpus très particulier: des limites à évaluer et des difficultés à surmonter

Pour répondre à ces besoins, la Toile se présente, en apparence, comme un outil idéal en permettant d'accéder *en une seule fois* et en un temps record à une quantité de données extrêmement importante, infiniment supérieure aux possibilités antérieures. Mais c'est tout de même un corpus très particulier.

3.1 *Le plus grand corpus de textes du monde est un ensemble flou*

La première caractéristique du *corpus* particulier représenté par les textes et documents présents sur la Toile est assurément de ne pas être quantifiable. Intrinsèquement, parce que ce corpus est vivant, se fait et se défait en partie à chaque minute. Pour des raisons techniques, également, parce que les outils permettant d'accéder à son contenu (les moteurs de recherche) ne peuvent pas dépouiller ces textes de façon exhaustive.

Quelle est donc la capacité de dépouillement de ces moteurs de recherche? En 2004 et en 2005, Google et Yahoo se sont affrontés quant au nombre de milliards de *pages* qu'ils revendiquaient avoir traitées. Des spécialistes comme Jean Véronis, professeur d'Informatique et de Linguistique à l'Université de Provence¹, ont jeté le doute sur ces revendications². Pour ce qui nous concerne, il faudrait également savoir quelle proportion de textes est rédigée en français, la langue anglaise étant nettement majoritaire.

Par ailleurs, ce *corpus* est, grossièrement, composé de deux sous-ensembles dont la nature respective fait obstacle partiellement à leur évaluation numérique. La Toile est en effet constituée en partie d'un ensemble de textes originaux (blogs, forums de discussion, pages personnelles sur des réseaux sociaux, rapports, informations, sur des sites personnels ou professionnels), particulièrement marqué par une grande expansion et une grande volatilité. Elle archive par ailleurs un très grand nombre de publications imprimées anciennes que l'on essaie de préserver et de rendre accessibles à tous, ensemble que divers projets nationaux, européens ou mondiaux complémentaires ou concurrents essaient d'accroître jusqu'à une (utopique?) exhaustivité. Dans le cadre de son projet universel, Google estime cette fois à 130 millions le nombre total de livres imprimés à numériser³. Il revendiquait fin 2010 la numérisation de 500 milliards de mots (soit 4% du total des textes imprimés, selon certains), dont 40 milliards en français⁴. Ce domaine est en expansion

¹ Voir sa présentation sur son site (qui mentionne également son blog intitulé "Technologies du Langage"): <http://sites.univ-provence.fr/veronis>.

² La notice sur Jean Véronis dans Wikipédia rappelle l'impact de ces polémiques en 2005. L'article en jeu s'intitule "Yahoo: pages manquantes?" et il est composé de quatre billets successifs, suivis de discussions souvent intéressantes de lecteurs. Le premier a pour adresse: <http://blog.veronis.fr/2005/08/yahoo-pages-manquantes-1.html>.

³ Voir par exemple l'article de Joffrey Bollée dans *L'Express* du 09/08/2010, disponible à l'adresse http://www.lexpress.fr/culture/livre/google-estime-a-130-millions-le-nombre-de-livres-dans-le-monde_911393.html, qui renvoie en outre à l'histoire de ce projet. On trouvera l'article de Leonid Taycher (5 août 2010), qui explique le mode de calcul de cette estimation, à l'adresse: <http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html>.

⁴ Les articles de Jean Veronis, "Google: Le plus grand corpus linguistique de tous les temps", <http://blog.veronis.fr/2010/12/google-le-plus-grand-corpus.html>, et d'Yves Miserey sur le site du journal *Le Figaro*, publié le 16/12/2010,

constante lui aussi, avec théoriquement peu de volatilité, l'objectif étant cumulatif.

Mais ces deux sous-ensembles ne sont pas intégralement consultables. De nombreux résultats répertoriés par Google renvoient à des sites payants réclamant inscription, achat ou abonnement et ne sont donc pas directement accessibles aux chercheurs, sauf convention collective. De même, une partie seulement des *Google Books* (20% selon certaines sources, mais combien en français?) est accessible en intégralité⁵ (et téléchargeable), Google cherchant laborieusement à établir des accords avec les éditeurs pour les livres soumis à des droits d'auteur. Pour ces derniers, dont le site ne donne qu'un aperçu de taille très variable, *Google Books* n'est qu'un catalogue bibliographique augmenté de quelques extraits. Il s'ensuit par ailleurs, pour ce qui est du degré (virtuel et idéal) de la *représentativité* des textes de la Toile, que, contrairement aux textes du réseau, qui sont fort marqués par la culture médiatique, la mode et même l'actualité immédiate, les archives sont par définition déséquilibrées en faveur des textes anciens, ce qui accentue la différence avec les autres textes de la Toile.

3.2 *L'efficacité des moteurs de recherche est relative*

Cet ensemble monstrueux n'est pas consultable directement. Or on ne peut être sûr que les moteurs de recherche ramènent de façon objective et fiable les informations que nous avons demandées, car leur fonctionnement, qui n'est pas standardisé, est opaque. Il faut l'aide de spécialistes pour essayer de percevoir leurs caractéristiques.

D'abord destinés aux seules pages *web*, les moteurs de recherche ne dépouillent pas dans les mêmes proportions les textes de différents formats (pdf...). Ils ne recensent pas non plus la totalité (ni la même proportion) des documents très longs: Jean Véronis a prouvé en 2005 (Billet 1) que Yahoo recensait une plus petite partie des textes longs que Google, qui est pourtant très loin de leur appliquer un traitement exhaustif.

Les moteurs de recherche opèrent en outre, de fait, des choix dans le corpus, en recourant à un algorithme évolutif pour hiérarchiser les pages qu'ils recensent. Dans un article d'un *Dossier Pour la Science*, Jean-Paul Delahaye (2010), de l'Université de Lille, tout en expliquant l'importance du

<http://www.lefigaro.fr/sciences/2010/12/16/01008-20101216ARTFIG00752-google-numerise-des-milliards-de-mots.php>, commentent tous deux l'article "Google Opens Books to New Cultural Studies", de John Bohannon (*Science*, 17 December 2010: 1600).

⁵ L'article intitulé "Pourquoi est-il impossible de lire le livre dans son intégralité?", sur la page <http://books.google.fr/support/bin/answer.py?hl=fr&answer=43729&ctx=cb&src=cb&cbid=11vesah7j4dzu&cbrank=1>, a été rédigé par le centre d'aide destiné aux utilisateurs des *Google Books*.

critère de la *popularité* d'une page dans ce *Pagerank*, proteste contre l'opacité globale de ce traitement, qu'il attribue prioritairement au souci d'échapper à des contestations juridiques des entreprises qui pourraient s'estimer lésées par un mauvais classement aux fondements trop empiriques, voire arbitraires.

Le souci de lutter contre les sites manipulant leur contenu pour attirer les internautes vers leurs publicités amène par ailleurs à l'élimination de nombreuses réponses, sur la base de l'établissement d'une liste noire constamment remise à jour⁶. La série de billets de Jean Véronis⁷ traite précisément de ce problème des "pages manquantes", qui est à l'origine de la grande fluctuation dans le nombre de résultats affiché à l'occasion d'une requête.

Google indique en effet sur la première page de réponses à une requête large un nombre total de réponses très important, qui peut atteindre plusieurs millions. Par exemple, la requête "lire des livres" se voit attribuer d'abord 3.910.000 résultats (le 14/1/2012 à 17h20). Ce nombre de réponses est très souvent provisoire. Il ne se stabilise souvent qu'à la page 4 (qui porte cette fois la mention "page 4 sur environ 8.850.000 résultats"). Ce pic reste stable jusqu'à la dernière page consultable, en l'occurrence la page 53. Et c'est seulement en atteignant cette dernière page que l'on apprend que, "pour limiter les résultats aux pages les plus pertinentes (total: 530) Google a ignoré certaines pages à contenu similaire". On constate à cette occasion que Google n'a pas fourni le nombre maximal d'occurrences (qu'il a fixé à 1000), sans aucune explication supplémentaire. La récupération (proposée par un lien) des "pages ignorées" aboutit au résultat légèrement supérieur de 680 résultats, toujours en deçà des 1000 exemples possibles et surtout des millions d'exemples revendiqués.

Les processus d'élimination (censés éliminer ce que Google nomme les résultats "non pertinents"), manquant partiellement d'efficacité dans leur recherche des sites-parasites, ne parviennent pas toujours à privilégier le texte "original" par rapport à ses copies⁸ et ont tendance à évacuer automatiquement certaines pages pour des raisons formelles et techniques alors même que leur *contenu* n'est pas semblable ou au contraire à engranger plusieurs états différents des pages dynamiques⁹. En

⁶ Voir sur ce point la présentation par Google de la version *Panda* de son système de "nettoyage": <http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html>, et un commentaire critique, dans un article du 25 février 2011: (<http://www.blogdumoderateur.com/index.php/post/google-ameliore-ses-resultats-de-recherche>). Sur le principe, cf. déjà Véronis en 2005 (Billet 3).

⁷ Et particulièrement Véronis, 2005, Billet 3.

⁸ Voir par exemple les craintes récentes de certaines entreprises, dans un article du 23 septembre 2011, (<http://actu.abondance.com/2011/09/google-panda-premiere-jurisprudence-en.html>).

⁹ Véronis, 2005, Billet 4, dans les discussions.

outre, la réduction volontaire des résultats réellement consultables au nombre de 1000 les pousse à limiter fortement le nombre d'occurrences issues d'un même document. Le moteur de recherche intégré à Persée (<http://www.persee.fr>), portail de revues francophones en sciences humaines et sociales, fait mieux sur ce point, en citant toutes les occurrences demandées.

L'intérêt d'études statistiques est clairement anéanti par ces mystères et ces variations aux logiques peu explicites: le nombre de pages "est une chimère"¹⁰. On peut tout au plus faire état du nombre définitif de réponses obtenues tel jour grâce à tel moteur de recherche (en citant le nombre d'exemples consultables en allant à la dernière page accessible des résultats). Cela permet de prendre connaissance de la rareté ou de la fréquence *relative* d'un mot ou d'une séquence de mots, sans aucune garantie scientifique.

3.3 *La formulation de requêtes adressées à ces moteurs de recherche relève d'un apprentissage assez complexe et présente des aspects déroutants*

En attendant de futures améliorations, on peut essayer, malgré ses défauts, de tirer parti de façon intéressante de ce corpus infini et en grande partie accessible. On peut même, à certains points de vue, le considérer comme incontournable et pour longtemps irremplaçable.

Les indications qui suivent font état d'une démarche expérimentale acquise empiriquement. L'analyse sommaire des exemples de recherches choisis reviendra sur le caractère déroutant des indications chiffrées.

Les modes de "recherche avancée" de Google (ou d'autres moteurs de recherche) peuvent être précieux, malgré leurs graves défauts¹¹, à condition justement qu'on en connaisse les limites.

Ainsi, dans les requêtes, les majuscules ne sont pas distinguées des minuscules, et les accents ne sont pas pris en compte a priori.

Les opérateurs booléens ET (mobilisé par la formule *AND* ou implicitement par une requête multiple quelconque) et OU (*OR*) n'ont pas le même degré de fiabilité. Le premier est assez fiable, malgré une sensibilité anormale et variable à l'ordre des mots de la requête. Il est cependant nettement inférieur à la mise entre guillemets d'un mot ou d'une séquence de mots,

¹⁰ Véronis, 2005, Billet 4.

¹¹ Jean Veronis, <http://blog.veronis.fr/2005/01/web-comptes-bidons-chez-google.html> (qui renvoie également à un article de Danny Sullivan du 28/01/2005, "Questioning Google's Counts", <http://searchenginewatch.com/article/2062882/Questioning-Google-Counts>) dénonce les imperfections de cette recherche "avancée" en en donnant de nombreux exemples. J'essaierai pour ma part de préciser dans quelles limites elle peut néanmoins rendre certains services.

qui, quand cette requête est possible, amène des résultats très précis et force également le moteur de recherche à tenir compte des accents. L'utilisation de *OR* n'est pas systématiquement fiable, sans parler du fait qu'elle risque de nous priver de nombreux exemples en raison du plafonnement à 1000 réponses.

Premier exemple de recherche¹²: le parasitage par les homographes vrais ou approximatifs

Sans les guillemets, la requête /livré/ est assimilée à la requête /livre/ et fournit presque exclusivement des occurrences de ce nom. La requête simple /"livre"/ en tant que forme du présent du verbe *livrer* ne peut aboutir, pour la même raison.

Si on s'intéresse au participe passé passif (et aux formes verbales composées) du verbe *livrer*, on formulera donc précisément la requête /"livré"/, puis successivement /"livrée"/, /"livrés"/ et /"livrées"/.

Deuxième exemple de recherche: essai d'élimination des concurrents

Le recours à l'opérateur SAUF (représenté par le signe de la soustraction) permet de résoudre ce problème dans des cas particuliers.

Ainsi, si l'on s'intéresse au verbe "lancer", la requête /"lance"/ sur le corpus des "pages en français" obtient un nombre provisoire affiché de 55 millions (puis un pic de 55,5 millions) de réponses, qui se réduit finalement à 786 occurrences consultables. Mais on constate très facilement, en tête de liste, la présence parasite dans ces résultats du nom du contesté champion cycliste américain *Lance Armstrong*. On peut l'éliminer classiquement de ces occurrences en formulant la requête complexe /"lance" -Armstrong/. On obtient alors le nombre provisoire incohérent de 129 (puis 128) millions de réponses (supérieur à celui de la requête précédente!), qui se réduit à 983 occurrences consultables. On est ainsi passé mystérieusement de 786 réponses hétéroclites à 983 occurrences mieux ciblées. L'élimination des réponses parasites issues du monde médiatique et surreprésentées sur la Toile est donc possible. On peut par ailleurs procéder à un grand nombre d'éliminations successives, ce qui permet d'affiner progressivement la requête (avec, toutefois, l'inconvénient, non évalué précisément, d'éliminer alors conjointement des occurrences pertinentes).

¹² Tous les exemples issus de la Toile ont été vérifiés en janvier 2012, date de la remise de cet article. Etant donné la présence fréquente de guillemets dans leur formulation, les requêtes ont été présentées conventionnellement entre barres obliques. Le nombre des exemples a été réduit de façon drastique dans le cadre de cet article à visée méthodologique, d'autant que chacun, et c'est là l'objet même de cette communication, peut retrouver ces données très facilement, s'il le souhaite, en ayant recours à la Toile.

Il ne faut pas croire par ailleurs que l'élimination (revendiquée par Google en ces termes) des pages moins "pertinentes", la mise en réserve des "pages ignorées" (qui sont, elles, récupérables) et l'évacuation des résultats des sites jugés non fiables (les pages *manquantes* des décomptes précédents) aient résolu le problème des doublons. Internet est le monde des échanges, du partage et de l'imitation.

Troisième exemple de recherche: les doublons

La voix passive n'est pas négligée dans la recherche de conversifs. Pour vérifier l'existence du pendant passif de l'actif /"une grue a tué"/ (dans le contexte d'accidents dans le domaine des travaux publics: j'ai pris ici un exemple de fréquence réduite, plus parlant)¹³, on peut formuler une requête /"tué par une grue"/, qui obtient un nombre provisoire de 49700 réponses, qui se réduisent très vite à 27 réponses effectives (théoriques)¹⁴.

Si l'on examine ensuite l'ensemble des 26 occurrences (après avoir supprimé un exemple avec l'homonyme la *grue* oiseau), on constate des répétitions qui ramènent ce nombre à 11 seulement. Les doublons les plus nombreux (respectivement 4 et 3 mentions) relèvent de nouveau du domaine médiatique. Ils figurent dans le résumé d'un film et dans celui d'un épisode d'une série américaine.

Le tri des doublons, tout à fait nécessaire en théorie, est évidemment très difficile et très long à réaliser quand les résultats sont très nombreux. Dans la mesure où le nombre de doublons ne peut être estimé en moyenne ni anticipé (et dans la mesure où il ne serait pas sérieux de présupposer, dans le cas de comparaisons entre formulations concurrentes, que chaque requête ramène en moyenne le même pourcentage de doublons), toute étude statistique fondée sur les nombres bruts de résultats effectifs manque donc totalement de fiabilité. C'est seulement à titre indicatif (et sous réserve d'un dépouillement minutieux ultérieur, s'il s'avère important pour la recherche menée) que l'on peut mentionner ces résultats approximatifs, comme j'ai pris la liberté de le faire moi-même à l'occasion parce qu'il s'agit d'une étude en cours.

¹³ *Tuer* est un verbe très polyvalent qui a été étudié en tant que causatif, y compris dans sa capacité à adopter divers sujets, par Sylvain Kahane et Igor Mel'cuk (2006).

¹⁴ On peut accéder à 11 réponses de plus en réclamant les pages "ignorées" par Google (disponibles d'un simple clic), mais ce ne sont (comme presque toujours) que des doublons issus des mêmes sites. On en reste donc à 27.

Quatrième exemple de recherche: perte de données avec les requêtes complexes

Par ailleurs, la requête /"(tué OR tuée OR tués OR tuées) par une grue"/ ne fournit curieusement pas de réponses incluant le participe au pluriel. D'autres requêtes comportant un nombre inférieur de formes du verbe *tuer* font également clairement apparaître à chaque fois un déficit d'occurrences pertinentes. Cela signifie que les requêtes complexes laissent échapper une partie des résultats que l'on peut obtenir par la série de requêtes simples correspondante.

Ajoutons que la vérification de ces occurrences ne peut pas se pratiquer directement dans tous les cas. Une partie des extraits fournis dans les résultats ne comporte pas le ou les mots demandés: il faut examiner le contexte large (et quelquefois très large) en consultant la page-source pour retrouver l'occurrence recherchée. Ce phénomène est amplifié quand la requête recourt au *joker* (l'astérisque peut remplacer tout mot dans une séquence), lequel semble dérouter le moteur de recherche. Autrement dit, l'indexation n'est pas précise pour tous les "résultats" proposés par Google (Yahoo est plus honnête sur ce point).

En outre, évidemment, les requêtes font apparaître dans les résultats des constructions non souhaitées, à éliminer également, et, en l'absence de lemmatisation, il faut formuler une à une les requêtes concernant les formes qui nous intéressent. Les documents en format *PDF* figurant dans les résultats se prêtent en revanche le plus souvent aux recherches par troncation (il est souvent nécessaire de les télécharger pour travailler plus aisément sur chacun d'eux).

3.4 Statut, dans ce contexte, de la diversité linguistique des textes

La langue des textes du réseau est-elle un obstacle supplémentaire à l'efficacité de ce type de recherches? De fait, tout le monde écrit maintenant sur la Toile. Les internautes deviennent rédacteurs de l'encyclopédie *Wikipédia*, de sites personnels, de pages dans les sites de réseaux sociaux, interviennent dans les forums de discussion pour exprimer leur opinion, certains en affichant un renoncement total à la norme.

En fait, si cette présence sur la Toile de français non standard possède un grand intérêt, elle rend nécessaire de multiplier encore les requêtes ne serait-ce que pour tenir compte des variantes orthographiques non normées. C'est indispensable quand on vise des supports de sources très diverses. Le participe passé des verbes du premier groupe *conjugués avec l'auxiliaire avoir*, notamment, est très souvent, dans les textes créés pour le Web, soit neutralisé (à la forme masculine), soit systématiquement accordé avec le sujet (!), voire remplacé par son homophone l'infinitif présent.

Beaucoup de manipulations, de tris, de recherches voisines et fastidieuses, pas de garantie scientifique de représentativité des résultats, dont le dénombrement est peu fiable, tels sont donc les inconvénients notables de la Toile en tant que corpus pour la recherche linguistique.

4. Un support néanmoins incontournable

Si l'on en revient, cependant, à une comparaison triviale avec la recherche *à l'ancienne*, comment contester la supériorité de la consultation des textes sur la Toile?

Si l'on exclut la partie des textes d'accès limité ou payant, comparable après tout en cela aux textes imprimés classiques, pour le reste, il est inutile d'insister sur la facilité de consultation de textes innombrables de différentes époques, consultation sans limites de lieu, de temps, presque de durée (sauf volatilité du support, mais les doublons les remplacent assez souvent), sans formalités. Le recours aux moteurs de recherche est, quant à lui, autorisé sans limites¹⁵.

Ce changement total de paradigme est encore somme toute très récent.

Il est toutefois indispensable, afin de tirer parti de ces richesses, de chercher un *modus operandi* adapté aux recherches menées. Je présenterai cette démarche empirique de deux façons. Tout d'abord, je ferai état de séries de requêtes menant, pas à pas, à l'étude des conversifs (la première étape consiste dans la recherche des parasyonymes venant "concurrencer" ou compléter l'emploi du verbe visé¹⁶). Ensuite (p.41-42), je présenterai très succinctement, dans leurs différentes dimensions, les résultats obtenus sur le verbe *embarquer*, déjà évoqué.

4.1 *Un bombardement de requêtes*

Pour compenser en partie les lacunes ou défauts du corpus et des moteurs de recherche, il convient d'adopter une stratégie de questionnement (de formulation de requête), en jouant, évidemment, sur la possibilité de multiplier ces requêtes.

La recherche de conversifs, je le rappelle, doit d'abord dégrouper les sens d'un verbe avant de chercher les modalités de variations de l'orientation de la phrase: autrement dit, elle considère les autres sens (et même les

¹⁵ La seule contrainte pour l'internaute, en cas d'utilisation intensive, est de prouver en recopiant un code de sécurité, que son ordinateur ne procède pas par malveillance à des recherches automatiques.

¹⁶ Il convient en effet, étant donné la complexité des données, de valoriser un verbe lors de chaque étude détaillée, mais de fait, avec plus de moyens, il serait plus efficace de rendre compte conjointement de la variation diathétique et de la complémentarité de tous les verbes d'un même réseau.

nouveaux sens, notamment "causatifs", qui apparaissent à cette occasion) comme des homonymes. Pour diminuer le nombre de résultats faisant apparaître ces homonymes, pas d'autre solution que de repérer initialement dans la masse des réponses les domaines les plus pertinents et de formuler ensuite des requêtes complexes, plus précises. J'illustrerai sommairement les étapes de ce processus méthodique.

On peut restreindre le champ dans un premier temps grâce à des mots clés représentatifs d'un domaine. Je reprendrai ici l'exemple de recherches autour du verbe *livrer* au sens *de fournir comme résultat concret, donner accès à, mettre à la disposition de* dans le domaine de la recherche archéologique.

Cinquième exemple de recherche: une approche très lâche

Par exemple, on demandera /"livré" archéologie/ et le mot clé étendra automatiquement la recherche à des textes comportant les mots *archéologue(s), archéologie* et *archéologique(s)*.

Sixième exemple de recherche: un joker pour pallier l'absence de catégorisation grammaticale

Une requête comportant un *joker* sera plus précise: /"archéologues ont * des vestiges"/ (*vestiges* ayant été repéré comme très fréquent dans cette position) fournira une série très riche de verbes "concurrents". On y rencontre les verbes attendus (notre première liste intuitive augmentée des recherches de paronymes dans des dictionnaires spécialisés: *découvrir, trouver, exhumer, dégager, déterrer, mettre au jour*), mais on peut également ainsi leur ajouter *mettre à jour* et *retrouver*¹⁷.

Septième exemple de recherche: approfondissement de ces premiers résultats

Il s'agit ensuite de vérifier, d'étudier et d'évaluer chacun de ces verbes grâce à une requête explicite (/ "archéologues ont découvert"/, / "archéologues ont trouvé"/, etc.).

Comme annoncé, la première surprise est donc venue du verbe *retrouver*, que je considérais a priori comme pléonastique (*trouver qqch d'enfoui, datant d'une époque passée, c'est toujours le re-trouver*), mais qui est finalement bien attesté, et même dans des proportions proches de celles

¹⁷ De fil en aiguille, ultérieurement, il faudra s'intéresser aux verbes en *re-* sur le modèle de ce *retrouver*: *redécouvrir, réexhumer, remettre au jour...* afin d'examiner si, comme c'est le cas pour *redécouvrir*, leur sens n'est pas nécessairement itératif.

du simple *trouver* et dans des contextes apparemment non spécifiques et non-itératifs

La seconde surprise est venue de *mettre à jour*, qui est généralement condamné. Le *TLFi* ne le valide d'ailleurs pas clairement¹⁸. Le dictionnaire en ligne canadien *Termium Plus*, qui le dénonce comme résultant dans ce sens d'une confusion avec *mettre au jour*¹⁹, formule toutefois une remarque intéressante. Il signale le caractère archaïque et déclinant de *mettre au jour* (remplacé aujourd'hui par *donner le jour*, *donner naissance*, *mettre au monde*), déclin qui pourrait être à l'origine de la forte concurrence constatée (*mettre à jour* obtient (avant tri) dans cette recherche 460 résultats effectifs, contre 476 pour *mettre au jour*). Dans l'attente d'une analyse systématique de ces premiers résultats, il est impossible de conclure. Les deux expressions se rencontrent en tout cas dans des contextes équivalents et cela à des dates variées. La Toile fournit ici des matériaux suffisants pour qu'on se pose la question de la validation de *mettre à jour* en rompant avec sa condamnation traditionnelle.

Huitième exemple de recherche: vers les emplois conversifs!

Cet inventaire constituait un premier objectif: le recensement des verbes qui de façon concurrente ont vocation, dans le contexte des fouilles archéologiques, à en exprimer le résultat, avec pour sujet l'agent de l'action préalable de *fouiller*, *chercher* (les archéologues). La recherche des conversifs consiste ensuite à traquer les autres constructions syntaxiques auxquelles se prêtent ces mêmes verbes, et notamment la série de sujets différents auxquels ils parviennent à s'adapter. La souplesse de nombreux verbes de ce point de vue est en effet fortement sous-estimée en français et il est très difficile actuellement de trouver de telles constructions, qui ne sont pas encore répertoriées²⁰.

Il s'agit maintenant, afin de trouver de telles occurrences de constructions originales, de ne plus faire figurer de sujet dans la requête. En revanche, pour éviter un éparpillement extrême des réponses, il est nécessaire d'ajouter au verbe un complément spécifique, tel que "vestiges".

¹⁸ *Mettre à jour* ne dispose pas d'une entrée dans le *TLFi*, mais il sert à définir *fouiller* et apparaît dans les articles *squelette* et *trésor*. Par ailleurs, l'article *jour* mentionne l'équivalence *mettre au jour* / *mettre à jour*, seulement, il est vrai, avec le sens de "Exposer, (s')étaler au (grand) jour, mettre à/au jour. Mettre à découvert; au fig. découvrir, faire apparaître, divulguer, publier".

¹⁹ Sous l'entrée *jour* (http://www.btb.termiumplus.gc.ca/tpv2guides/guides/clefsfp/index-fra.html?lang=fra&lettr=indx_catlog_z&page=1471).

²⁰ Il ne s'agit ici, je le rappelle, que d'une partie de la recherche de conversifs, celle qui vise les divers emplois d'un même verbe. La recherche des conversifs est de fait beaucoup plus étendue : le conversif peut être notamment un autre verbe (*recevoir* vs. *donner*) ou un des verbes recensés précédé d'un "auxiliaire" (le verbe *permettre* se rencontre dans ces contextes: *permettre de découvrir*, *de trouver*...).

Par exemple, la seule requête /"exhumé des vestiges"/ donne accès (outre *archéologues* et ses équivalents, fortement majoritaires mais évidemment déjà connus), à une série de sujets non agentifs, à savoir *les fouilles* (6 fois), comme dans :

- (1) Les premières fouilles de Xanthos (1950-1961) avaient exhumé des vestiges de constructions extrêmement mal conservées (...).
http://ausonius.u-bordeaux3.fr/axes/Kelainai/Kelainai_I_Courtils.html.

Egalement un *labour* (une fois) et *des travaux* (4 fois): il s'agit de travaux distincts de travaux de fouilles concertés, comme dans:

- (2) Bugeal. — Près du hameau de Massoulre, des travaux agricoles ont exhumé des vestiges de constructions (murs, dalles creusées de rigoles pour l'évacuation des eaux, débris de colonnes, briques, tuiles) et des tessons gallo-romains. Communication de M. Marius Vazeilles (P.F. Fournier, *Gallia*. Tome 23 fascicule 2, 1965. p. 389.
 url: http://www.persee.fr/web/revues/home/prescript/article/galia_0016-4119_1965_num_23_2_2428)

Egalement, une fois, des *travailleurs*.

Enfin, *les crues* (associées ailleurs aux verbes *recouvrir* ou *dégager*):

- (3) Les crues qui ont balayé la Tunisie centrale à l'automne 1971 ont exhumé des vestiges romains enfouis depuis plus de dix siècles!²¹

Si les réponses sont trop hétéroclites (avec des compléments moins spécifiques), il redevient utile de préciser la requête au moyen d'un mot-clé supplémentaire (tel qu' "archéologie", comme précédemment).

La suite du travail consiste à élaborer un classement sémantique provisoire des sujets recensés afin de distinguer les emplois dits causatifs (ou d'autres encore) des véritables emplois de conversifs²².

L'énumération précédente correspond à 4 catégories : (1) les *fouilles* (action visant la découverte) // (2) les *travaux* autres / (3) les *travailleurs* (action autre ne visant pas la découverte / agent de cette action) // (4) les *crues* (phénomène naturel). Les cas (2) et (4) correspondent à des emplois dits causatifs du verbe *exhumer* ("avoir pour conséquence la mise au jour de"). Le cas (3) repose sur un emploi d'*exhumer* qui mérite d'être envisagé comme un conversif du cas (2), même si ce type de variation n'a pas encore été étudié systématiquement. Le cas (1), enfin, donne à *exhumer* le statut de "verbe de réalisation" (Mel'cuk, Clas et Polguère, 1995: 141-2). Chacun de ces emplois devra ensuite à son tour être examiné, vérifié, validé par d'autres occurrences. Il faudra également étudier si le passif de ces constructions existe.

²¹ Cette phrase figure dans un dossier sur l'érosion et son rôle dans le vieillissement des montagnes : www.sciencesettechniques.com/dossier.php?dos=39.

²² Iordanskaja & Mel'cuk (2002) aident à trancher entre emplois causatifs et emplois conversifs qui, tous deux, s'accompagnent d'une variabilité du sujet syntaxique.

Neuvième exemple de recherche: resserrer les mailles du filet

Cette première liste (pour le seul *exhumer*) a été obtenue grâce à la présence dans la requête d'un complément assez général (les "vestiges"). En substituant à ce nom d'autres mots relevant du domaine spécialisé de l'archéologie, on peut, par tâtonnement, à l'aide de nouvelles séries de requêtes /"exhumé une statue"/, "exhumé des ossements"/ (précédées du verbe *avoir* au singulier puis au pluriel à différents temps), dégager un nouvel emploi:

- (4) Au lieu-dit Les Grandes Lignières (...) lors de la construction du canal de Saint-Julien, un bulldozer a exhumé un 'fond de cabane' du premier âge du Fer"
(Laurent Denajar, *Carte archéologique de la Gaule: 10. L'Aube* (2005), 471:
<http://books.google.fr/books?isbn=2877540936>, consultation en mode aperçu limité.)

Cet exemple est un conversif du verbe d'action *exhumer* qui complète les cas précédents (2) et (3): *exhumer* supporte ici, classiquement, d'avoir pour sujet l'instrument des travaux (le *bulldozer*).

L'accès à des données de fréquence plus limitée grâce à ce maillage très fin pose évidemment le problème de la norme et celui de la limite de pertinence de telle ou telle construction. Dans cette dernière recherche, bien que le nom *crue* soit, d'après la Toile, rarement sujet du verbe *exhumer*, une série d'autres contextes confirme son intérêt. En revanche, un exemple comportant le nom *trouvailles* comme sujet, dans un article pourtant soutenu, voire lyrique²³, ne peut être pris en compte.

Il m'est impossible d'aller plus loin dans l'illustration concrète de cette recherche en cours dont, on le voit, les ramifications sont nombreuses. La question de la valorisation et du regroupement de ces faisceaux de données dans les dictionnaires se pose aujourd'hui nettement. Elle est la conséquence de l'enrichissement que permet le recours à la Toile.

4.2 Des progrès liés au grand nombre, même imparfaitement maîtrisé

On appréciera par ailleurs la variété des sources consultables et la facilité d'accès au contexte large, voire intégral si on le souhaite. La Toile permet d'accéder à des séries de textes ou à des œuvres intégrales très denses, qui permettent à leur tour, si elles se distinguent des autres par leur richesse, de mener des recherches complémentaires au-delà de cette cueillette, de cette collection d'extraits, on l'a vu un peu aléatoire (ou en tout cas non maîtrisable).

²³

Ce texte du 9 mars 2011, assez bref, traite des thermes romains de Banassa au Maroc:
<http://www.archimedia.ma/avis-paroles-dexperts/architecture-et-patrimoine-/2027-les-thermes-antiques-du-maroc>.

Ainsi, une recherche sur le verbe *livrer* pourra bénéficier de la présence sur la Toile d'un ensemble de comptes rendus de missions archéologiques parmi lesquels un des rapports concernant Feurs dans la Loire (publié sur le site [archeodunum.ch](http://www.archeodunum.ch))²⁴ se révèle d'une très grande richesse en livrant plus de 140 occurrences variées en 144 pages! La *recherche avancée* "par site" proposée par Google n'est pas efficace sur ce point. Si elle augmente le nombre de documents cités issus du même site, elle continue cependant à restreindre drastiquement le nombre de résultats issus de chaque document, la requête précise */"livré"/* limitée au site *archeodunum.ch* ne produisant que 16 résultats au lieu des centaines attendues, nouvelle preuve du dépouillement très partiel des documents longs. La consultation des documents intégraux les plus riches reste donc indispensable. Mais comment deviner leur richesse?

La recherche déjà mentionnée sur le verbe *embarquer* (Deronne, 2008 a), avec laquelle je vais terminer cet aperçu, a pu profiter plus facilement de la présence sur la Toile de *séries* de rapports d'accidents rédigés par les sites français du BEAmer (Bureau d'enquêtes sur les événements de mer, <http://www.beamer-france.org/index.php>) et surtout canadien du BST, Bureau de la sécurité des transports du Canada, en anglais TSB (<http://www.tsb.gc.ca/fra/avis-notices.asp>). Elle a bénéficié aussi, par ailleurs, de la présence de textes (normés ou non normés) aussi divers que des sites ou blogs (de navigateurs professionnels ou amateurs, de constructeurs de bateaux), de journaux de bord ou de récits d'explorateurs et de voyageurs du 18^{ème} et du 19^{ème} siècles, ainsi que de romans d'aventure (notamment de Dumas et de Jules Verne, que Frantext ne proposait pas exhaustivement mais que j'ai pu interroger sur les sites d'amateurs éclairés).

Ce sens d'*embarquer*, peu valorisé par les dictionnaires, a été amplement conforté par de nombreux exemples qui ont permis d'approfondir son fonctionnement et légitiment son dégroupement d'avec les autres sens d'*embarquer*. Une construction nouvelle a été dégagée ("les hublots embarquent de l'eau"), une autre étendue, et le nom *embarquement (d'eau)* a reçu une confirmation éclatante²⁵: il avait été récusé dans ce sens par Alain Rey (1992), et était absent des dictionnaires, sauf, à ma connaissance, du *Petit dictionnaire de marine* de Robert Gruss (1952), que j'ai consulté a posteriori, après la redécouverte du nom sur la Toile.

²⁴ Ce rapport de Sébastien Freudiger concernant le 45-47, rue de Verdun à Feurs, dans la Loire figure à l'adresse http://www.archeodunum.ch/FILES/mc7/52_tmp_218.pdf. Il est recensé par la base Frantiq, <http://koha.mom.fr/> Fédération et Ressources sur l'Antiquité.

²⁵ Par exemple, dans "Lors de l'élaboration du plan de la porte latérale, on n'a pas tenu compte des risques potentiels associés à l'embarquement d'eau pendant une opération de récupération de personnes à l'eau", www.tsb.gc.ca/fra/rapports-reports/marine/2009/m09l0068/m09l0068.asp (modifié 19/11/2010).

Du point de vue diachronique, une antédation du verbe a été possible et un texte du début du 19^{ème} siècle a montré que dès cette date il était déjà utilisé dans la plupart des constructions recensées. Seule l'insuffisance du corpus des textes anciens de navigateurs et de textes techniques antérieurs à 1950 freine de nouvelles antédations.

Des emplois causatifs non encore recensés d'*embarquer* et *de faire embarquer* (Deronne, 2008 b) ont, enfin, pu m'être confirmés par le témoignage d'un auteur contacté par l'intermédiaire de son site personnel consacré à la marine marchande. La Toile peut mettre les chercheurs *très facilement* en relation avec les auteurs.

Dans le même ordre d'idées, la richesse de la Toile permet, dans sa profusion d'opinions diverses, d'accéder à des réflexions métalinguistiques intéressantes, allant du commentaire explicite à la mise entre guillemets, au jeu de mots ou encore à la contestation d'une de ces formulations²⁶.

5. Bilan

Le recours à ce corpus illimité et mouvant suggère de la part du chercheur un comportement adéquat, dans le respect de certaines précautions.

Plus encore que dans d'autres domaines, dans la mesure où les données de la Toile sont très hétéroclites, le statut, le genre de discours et le registre des documents exploités devrait être suggéré avec précision et les textes cités fidèlement sans toilette.

De même, la confusion dans les données chiffrées, outre un avertissement général, suggère de distinguer le nombre brut d'occurrences effectivement citées ("avant tri") du nombre final d'occurrences "pertinentes" vérifié par le chercheur.

J'ai essayé de montrer comment, pour surmonter les lacunes des moteurs de recherche, on pouvait procéder à des recherches tâtonnantes, concentriques, nécessitant une stratégie patiente d'interrogation par approches successives. Mais, dans l'état actuel de la Toile et malgré sa croissance, il n'est pas (encore?) possible de se passer des supports papier: la consultation de dictionnaires et d'autres ouvrages spécialisés anciens non numérisés ou de textes plus récents soumis au droit d'auteur est encore indispensable. Par exemple, à propos du verbe *livrer*, le volumineux ouvrage de Vialou (2004) n'est pas exploitable²⁷ sur le site des

²⁶ Ainsi, des internautes contestent la formulation "La route a tué tant de personnes ce week-end" parce qu'elle voile, selon eux, les vraies responsabilités (des conducteurs ou de l'Etat). Bonne mise en garde quant à la valeur éventuellement argumentative de certaines de ces variations de constructions.

²⁷ "Aucun aperçu disponible", déclare sobrement la page qui lui est consacrée: <http://books.google.fr/books?id=y24qHAAACAAJ&dq=denis+vialou&hl=fr>.

Google Books, qui est bien à ce jour avant tout un *catalogue* des livres de la bibliothèque mondiale.

On notera enfin que le recours à des expérimentations plus classiques pour vérifier la validité d'une hypothèse ou le degré d'acceptabilité d'une structure ne s'en trouve pas condamné. La recherche sur la Toile peut être intégrée à une stratégie de recherche plurielle.

Finalement, la Toile est incontournable pour certaines recherches. Son hétérogénéité, quoiqu'elle complique les recherches, est un atout irremplaçable et son immensité également. Ajoutées à sa facilité (relative) de consultation, elles lui confèrent une *valeur heuristique* extraordinaire, en permettant d'accéder de façon ciblée à un nombre très important de données voisines. La Toile peut aussi mettre à la disposition du chercheur, pour inaugurer une recherche ou la relancer, des données moins ciblées dans lesquelles il peut trouver des faits stimulants dignes d'intérêt. Enfin, elle permet de multiplier les hypothèses *à peu de frais*, de tester des intuitions sans véritables limites.

Quels que soient l'évolution future des corpus exploitables et leur degré de préparation à la recherche linguistique, l'étude de la valence verbale, visant à distinguer des variations fines de constructions syntaxiques voisines ou complémentaires, parmi lesquelles les constructions converses, n'en restera pas moins difficile à généraliser et la démarche empirique ici illustrée sera difficilement opérée à l'aide de logiciels. On pourra tout au plus cataloguer dans la description de la langue les régularités absolues, telle la souplesse du choix du sujet des verbes causatifs. Ce travail de fourmi ne pourra donc pas être remplacé de sitôt.

Bibliographie

- Delahaye, J.-P. (2010): Le secret de Google. In: Dossier Pour la Science n°66, janvier-mars 2010, 64-69.
- Deronne, E. (2008a): Embarquement immédiat, Etude des diathèses renvoyant au verbe embarquer (de l'eau), à l'exclusion des relations de causalité. Disponible: http://halshs.archives-ouvertes.fr/halshs-00276585_v1/.
- (2008b): Embarquement différé, Etude des formulations renvoyant aux relations de causalité. Disponible: http://halshs.archives-ouvertes.fr/halshs-00311980_v1/.
- (2010): Analyse de verbes qui sont plusieurs fois leur propre conversif. In: Neveu, F., Muni Toke, V., Durand, J., Klingler, T., Mondada L. et Prévost S. (éds): Actes de la section "Lexique et morphologie" du 2e Congrès Mondial de Linguistique Française (CMLF'10), 909-925.
- (2012): Livraisons en tous genres: le verbe 'livrer' comme conversif (dans le domaine de l'archéologie), (CMLF'12), Lyon.
- Gross, G. (1989): Les constructions converses du français. Genève (Droz).

- Gruss, R. (3e éd., 1952): Petit dictionnaire de marine. Paris (Société d'éditions géographiques maritimes et coloniales).
- Iordanskaja, L. & Mel'cuk I. (2002): Conversif ou causatif? In: Cahiers de lexicologie 80-2002-1, 105-119.
- Kahane, S. & Mel'cuk, I. (2006): Les sémantèmes de causation en français. In: Hamon, S. et Amy, M. (éds): La cause: approche pluridisciplinaire. Linx 54, 247-292.
- Lemaréchal, A. (1989): Les parties du discours, sémantique et syntaxe. Paris (PUF).
- Martinet, A. (1985): Syntaxe générale. Paris (A. Colin).
- Mel'cuk, I., Clas, A. & Polguère, A. (1995): Introduction à la lexicologie explicative et combinatoire. Louvain-la-Neuve (Duculot).
- Muller, C. (2005): Diathèses et voix en français. In: Interaction entre sémantique et pragmatique, Actes du XIe Séminaire de Didactique Universitaire (Association des Chercheurs en Linguistique Française, Université Ovidius, Constanta 2004). Bucarest (Editura ASE), 73-95. Disponible: <http://erssab.u-bordeaux3.fr/spip.php?article15>.
- Rey A. (dir.) (1992): Dictionnaire historique de la langue française. Paris (Dictionnaires Le Robert).
- Véronis, J. : <http://blog.veronis.fr/>
- Vialou D. (éd.) (2004): La Préhistoire, histoire et dictionnaire. Paris (R. Laffont).

Remarque

Tous les exemples issus de la Toile ont été vérifiés en janvier 2012. Les exemples de recherches sont fondés sur des séries de requêtes formulées, sauf mention contraire, entre juin et septembre 2011.