



## **CogniTExtes**

Revue de l'Association française de linguistique cognitive

**Volume 19 | 2019**

**Corpora and Representativeness**

---

# **Jiyoung Yoon & Stefan Gries, *Corpus-based Approaches to Construction Grammar***

**Romain Delhem**



### **Édition électronique**

URL : <http://journals.openedition.org/cognitextes/1391>

ISSN : 1958-5322

### **Éditeur**

Association française de linguistique cognitive

### **Référence électronique**

Romain Delhem, « Jiyoung Yoon & Stefan Gries, *Corpus-based Approaches to Construction Grammar* », *CogniTExtes* [En ligne], Volume 19 | 2019, mis en ligne le 17 juin 2019, consulté le 19 juin 2019. URL : <http://journals.openedition.org/cognitextes/1391>

---

Ce document a été généré automatiquement le 19 juin 2019.



*CogniTExtes* est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

---

# Jiyoung Yoon & Stefan Gries, *Corpus-based Approaches to Construction Grammar*

Romain Delhem

---

## RÉFÉRENCE

Jiyoung Yoon & Stefan Gries (Eds.). 2016. *Corpus-based Approaches to Construction Grammar* [Constructional Approaches to Language, 19]. Amsterdam & Philadelphie : John Benjamins. vi, 268 pp. <https://doi.org/10.1075/cal.19>

- 1 *Corpus-based Approaches to Construction Grammar* est un ouvrage publié en 2016 et dirigé par Jiyoung Yoon et Stefan Gries. Avec l'avènement de d'outils numériques toujours plus performants, les corpus sont omniprésents et accessibles à tous, notamment sous forme électronique, et leur utilisation dans l'analyse linguistique est devenue incontournable. Elle l'est d'autant plus dans le cadre des approches constructionnistes, qui sont basées sur l'usage. Les jugements introspectifs de grammaticalité autrefois mis en avant laissent ainsi place à la recherche d'autres critères, tels que la fréquence statistique des constructions dans les corpus ainsi compilés.
- 2 Les grammaires de construction sont nées aux États-Unis (le nom apparaît la première fois chez Fillmore 1988) ; en conséquence, les travaux qui se situent dans cette approche se consacrent essentiellement à la langue anglaise, alors que les principes constructionnistes ont vocation à être universels. Yoon et Gries ont donc fait appel à des auteurs spécialistes de langues autres que l'anglais contemporain afin de décrire l'utilisation et l'utilité des corpus pour l'analyse, dans une approche constructionniste, de certains phénomènes grammaticaux de l'espagnol, de l'italien, du néerlandais ou du vieil-anglais.

- 3 Après une présentation des enjeux de l'ouvrage et de ses différents contributeurs dans le premier chapitre, le livre se divise en trois grandes parties. Celles-ci correspondent à des niveaux de complexité croissante dans l'utilisation des corpus :
  - l'étude de la fréquence statistique des constructions (chapitres 2, 3 et 4) ;
  - la mesure du degré d'attraction ou de répulsion entre deux mots, appelée *collocation*, ou entre un mot et une construction schématique complexe, appelée *colligation* ou *collostruction* (chapitres 5, 6 et 7) ;
  - l'analyse statistique multifactorielle ou multivariée d'un phénomène linguistique (chapitres 8 et 9).
- 4 L'ouvrage s'ouvre par une étude de Maaïke Beliën, qui porte sur les constructions à particule adpositionnelle en néerlandais (*hun fietspontje vaart her kanaal over* 'le ferry à vélos traverse le canal'), constructions où le verbe est suivi d'un complément de destination signalé par un élément adpositionnel postposé et qui garde un sens spatial. Selon l'auteure, ces constructions sont difficiles à analyser en termes de constituance et les précédentes recherches dans le domaine sont contradictoires. Après une comparaison sémantique, Beliën rapproche les constructions à particule des verbes complexes séparables, plutôt que des constructions prépositionnelles. Elle en déduit qu'à défaut de toujours former un constituant phonologique (un ensemble d'unités temporellement contiguës), le verbe et l'adposition de la construction étudiée forment un constituant conceptuel (des unités entretenant un lien de dépendance sémantique), selon la distinction proposée par Langacker (2008 : 205-214).
- 5 Dans le chapitre 3, Valeria Quochi aborde la question de l'acquisition de l'emploi léger de *fare* ('faire') dans le langage enfantin italien. En analysant la base de données CHILDES, qui rassemble des retranscriptions de discours oral spontané, l'auteure distingue trois macro-catégories d'emploi : ceux où le nom qui suit *fare* est morphologiquement lié à un verbe (*spesa* 'dépense'), ceux où il représente une entité physique (*balena* 'baleine') et ceux où il signifie une abstraction (*repertorio* 'répertoire'). Elle analyse les constructions à verbe léger (CVL) comme des réseaux de constructions organisés autour d'un prototype, en l'occurrence *< fare + nom dérivé d'un verbe intransitif >*. Par une étude de transcriptions longitudinales de sessions interactives avec 11 enfants, Quochi montre que les enfants commencent par produire les CVL lexicalisés conventionnels. Ils forment ensuite très tôt un schéma pivot en *fare* où le nom qui suit signifie d'abord un bruit, puis une action intransitive et enfin une action transitive plus complexe. En accord avec les théories basées sur l'usage, l'auteure montre que les CVL en *fare* sont un outil utilisé très tôt par les enfants (autour de 24 mois) pour signifier des situations pour lesquelles ils n'ont pas de lexème simple disponible.
- 6 Victoria Vázquez Rozas et Viola G. Miglio étudient les verbes de sentiment en espagnol et en italien dans le chapitre 4. Dans cette famille de verbes, il existe deux grands alignements entre fonction syntaxique et rôle sémantique : le rôle d'expérient peut être porté par le référent du sujet (ex. *amar/amare* 'aimer') ou de l'objet direct ou indirect (ex. *gustar/piacere* 'plaire'). Les auteures s'emploient à montrer que ces paires de verbes ne sont pas strictement équivalentes : chacune des constructions a une fonction pragmatique et discursive distincte. Ainsi les constructions à expérient objet sont davantage utilisées lorsque l'expérient est à la première personne, dans la mesure où seul le locuteur est à même de savoir ce qui l'affecte. Inversement, nous n'avons accès aux émotions des expérientes de troisième personne que par ce qu'ils nous donnent à voir, d'où l'idée qu'ils sont davantage agentifs et leur compatibilité plus grande avec les

constructions à expérience sujet. Les auteures démontrent ainsi qu'une analyse statistique de fréquence sur corpus selon des critères précis (personne et nombre, animéité, catégorie syntaxique) permet de mettre au jour des différences d'usage précises.

- 7 Le cinquième chapitre, écrit par Johan Pedersen, est une étude des verbes espagnols de mouvement du point de vue de la distinction entre langues à cadre verbal et langues à cadre satellitaire. L'espagnol figure généralement dans la première catégorie : le verbe signifie la trajectoire du mouvement et le mode de déplacement peut être facultativement codé à l'extérieur du verbe. L'auteur fait une analyse collocationnelle de la relation entre verbe de mouvement et construction téléique (*a + SN*) via un test exact de Fisher. Un étiquetage sémantique préalable des verbes montre que la construction téléique est typiquement associée à des verbes qui signifient une trajectoire, mais est également compatible avec des verbes de manière s'ils impliquent une certaine directionnalité (*correr, volar*). Après une comparaison entre l'anglais et l'espagnol dans les possibilités de complémentation des verbes, Pedersen raccroche la question des verbes directionnels à celle, plus large, des schémas de complémentation. À la distinction entre cadrage verbal et satellitaire, l'auteur préfère celle qui oppose les langues *verb-driven* ou *verb-framing*, comme l'espagnol, où le choix des compléments est surtout le fait du verbe, et celles qui sont *construction-driven* comme l'anglais, où la complémentation est plus libre.
- 8 Dans le chapitre 6, Jiyoung Yoon et Stefanie Wulff étudient les verbes espagnols pouvant avoir un complément infinitif ou un complément phrastique (en *que*). À l'aide d'une analyse des collexèmes distinctifs, qui étudie des constructions sémantiquement proches, les auteures montrent que les deux types de complémentation attirent des familles de verbes différentes, sans qu'il s'agisse de préférences absolues. Ainsi les compléments infinitifs sont davantage utilisés avec les verbes de désir et les compléments phrastiques avec les verbes de communication et d'activité mentale. Les auteures terminent sur l'hypothèse que les verbes qui sont davantage compatibles avec une complémentation phrastique sont ceux qui peuvent être réanalysés en marqueurs épistémiques (*yo creo, yo digo*).
- 9 Le chapitre 7 est une étude de Sarah Bernolet et Timothy Coleman portant sur l'alternance dite « dative » en néerlandais, c'est-à-dire la différence de sens entre une construction à deux objets et une construction avec un objet et complément prépositionnel en *aan* (l'équivalent de *à* en français et de *to* en anglais). Selon les auteurs, beaucoup d'études ont mis au jour les préférences de complémentation de certains verbes trivalents, mais ne prennent pas en compte la polysémie lexicale. Les auteurs distinguent donc, pour 15 verbes de transfert, deux acceptions (une concrète et une figurative) et proposent une analyse des collexèmes distinctifs pour voir quel schéma de complémentation est préféré par chacune des acceptions. Cette analyse montre que différentes acceptions d'un même verbe peuvent parfois privilégier une complémentation différente, ce que ne montraient pas les analyses par lexème.
- 10 Le chapitre 8 est consacré à une analyse multifactorielle de l'alternance *that/zéro* en anglais d'un point de vue diachronique. C. Shank, K. Plevoets et J. Van Bogaert commencent par donner les critères habituellement avancés dans la littérature pour expliquer la différence entre les deux constructions. Ils montrent ensuite que la croyance selon laquelle *that* aurait tendance à disparaître au profit du subordonnant zéro est fautive car ce n'est pas le cas pour les verbes *think* et *understand*. Dans la mesure où la pertinence des différents critères varie d'un verbe à l'autre, il est essentiel de faire non pas des analyses globales, mais détaillées, verbe par verbe et facteur par facteur. Les auteurs

estiment que le critère de choix le plus pertinent est la présence ou non d'éléments entre le verbe recteur et la subordonnée, et envisagent comme piste pour l'avenir la prise en compte de la prosodie et du rythme des énoncés.

- 11 Enfin, le chapitre 9 ferme la marche avec une étude par Natalia Levshina de la construction causative en *laten* du néerlandais (p. ex. *de general liet het leger de stad vernielen* 'le général ordonna à l'armée de détruire la ville'), équivalent de *faire faire* ou *laisser faire*. Elle montre que la question de la distinction des acceptions d'une même construction et l'existence d'un prototype doit être tranchée empiriquement par des analyses sémantiques sur corpus. À partir de 731 occurrences de la construction classées selon une trentaine de variables, l'auteure calcule la distance sémantique entre ces occurrences et la représente en deux dimensions via un outil de positionnement multidimensionnel. Il en résulte une structure elliptique sans centre. L'auteure en déduit que la construction n'a pas d'emploi prototypique, et que les différentes acceptions forment entre elles un continuum et diffèrent principalement selon deux dimensions sémantiques : le type d'évènement causé (mental, physique ou social) d'un côté, l'intentionnalité et le rôle sémantique du causataire de l'autre.
- 12 Le titre de la première partie (*Frequencies and probabilities*), qui regroupe les chapitres 2, 3 et 4, est relativement trompeur. Son but n'est pas vraiment d'exposer des analyses statistiques de certaines constructions, mais plutôt de montrer comment la prise en compte de faits linguistiques sur corpus peut faire avancer la recherche en interface syntaxe-sémantique et en acquisition du langage dans un cadre constructionniste. Il s'agit donc davantage d'un plaidoyer en faveur de l'utilisation de productions authentiques en grand nombre face aux jugements introspectifs de grammaticalité qui avaient cours autrefois en linguistique. Le chapitre 4 montre bien, cependant, comment des analyses statistiques sur corpus peuvent révéler que deux constructions généralement analysées comme synonymes ont en fait des fonctions pragmatiques et discursives différentes. Seules une approche et une méthode basées sur l'usage permettent donc de dépasser les intuitions parfois inexactes sur le sémantisme d'une construction.
- 13 La deuxième et la troisième partie, respectivement *Collostructional analysis* et *Multifactorial and multivariate analysis*, sont plus explicites dans leur utilisation des corpus et des outils statistiques et ouvrent davantage de pistes. Les auteurs de la deuxième partie (chapitres 5-7) montrent de façon convaincante que l'analyse collostructionnelle peut, en mesurant l'attraction entre différentes constructions, fournir des outils solides pour analyser le sémantisme de constructions que l'on croit bien souvent parfaitement synonymes. La contribution de Bernolet & Coleman montre ainsi que l'analyse sur corpus peut être mise au service d'une étude fine du sémantisme des verbes, tandis que celle de Yoon & Wulff explique les nuances sémantiques ténues qu'offre parfois la complémentation phrastique de certains verbes. La contribution de Pedersen, quant à elle, utilise les bases théoriques des grammaires de construction pour redéfinir la notion de cadrage chère à Talmy et comparer l'espagnol à d'autres langues d'un point de vue typologique.
- 14 Enfin, la troisième partie (chapitres 8 et 9) est extrêmement intéressante. Les auteurs y montrent (et il n'est pas vain de le rappeler) que la prise en compte d'un maximum de paramètres est essentielle pour l'analyse linguistique. Les résultats que les auteurs présentent font voler en éclats certaines idées communément admises : Shank *et al.* démontrent ainsi que le subordonnant *that* n'est pas en voie de disparition au profit du subordonnant zéro en anglais, tandis que la contribution de Levshina fait apparaître que

toutes les constructions n'ont pas d'emploi prototypique identifiable. Le fait de considérer un très grand nombre de critères grammaticaux relativement objectifs permet de faire apparaître des phénomènes qu'il aurait été impossible de mettre au jour sans analyse statistique. Le lecteur trouvera sans nul doute dans cette troisième partie une inspiration méthodologique qui viendra nourrir des études dans de nombreux domaines.

- 15 Cet ouvrage est original à deux titres. Il s'inscrit dans une approche clairement constructionniste, et il met l'accent sur la recherche empirique par corpus et sur les analyses statistiques pour analyser les faits de langue.
- 16 On pourra éventuellement reprocher à l'ouvrage (et non aux auteurs de façon individuelle) un manque de pédagogie. Les grammaires de constructions et leurs principes fondamentaux ne sont jamais présentés explicitement, y compris dans le premier chapitre d'introduction. Il n'y a pas non plus de passage consacré à l'explication des différents outils statistiques employés par les auteurs. Si les analyses de fréquence simples ne posent *a priori* aucun problème de compréhension, ce n'est pas le cas des analyses collostructionnelles ou multifactorielles ; or les auteurs partent en majorité du principe que ces outils sont connus et compris par le lecteur<sup>1</sup>.
- 17 En conséquence, l'ouvrage s'adresse principalement à des lecteurs qui ont non seulement un intérêt pour les approches constructionnistes et les analyses statistiques, mais également des connaissances établies dans ces deux domaines. On y trouvera en tout cas des applications statistiques stimulantes sur une grande variété de phénomènes linguistiques. Les différentes contributions constituent des apports importants à leur champ de recherche et ouvrent des perspectives de recherche et d'application intéressantes bien au-delà de leur domaine.

---

## BIBLIOGRAPHIE

Desagulier, Guillaume. 2017. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. New York: Springer

Fillmore, Charles. 1988. The Mechanisms of "Construction Grammar". *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*. 35–55.

Gries, Stefan. 2017. *Quantitative Corpus Linguistics with R: A Practical Introduction*, 2<sup>nd</sup> edn. London/New York: Routledge.

Langacker, Ronald. 2008. *Cognitive Grammar: An Basic Introduction*. Oxford: Oxford University Press.

## NOTES

1. Signalons qu'il existe désormais une littérature abondante sur ces questions et des outils statistique gratuits comme R assortis de manuels clairs. On pourra notamment se référer à des ouvrages tels que Gries (2017) ou Desagulier (2017).

---

AUTEUR

**ROMAIN DELHEM**

Sorbonne Université (CeLiSo – EA 7332) | Université Paris Nanterre (CREA – EA 370)