



HAL
open science

Pour une Analyse Cognitive de Discours.

Matthieu Bach

► **To cite this version:**

Matthieu Bach. Pour une Analyse Cognitive de Discours.: Twitter comme corpus, la grammaire de construction comme outil.. Repères-Dorif, 2020, 22. hal-02966463

HAL Id: hal-02966463

<https://hal.science/hal-02966463>

Submitted on 14 Oct 2020

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

Pour une Analyse Cognitive de Discours.

Twitter comme corpus, la grammaire de construction comme outil.

The present contribution proposes to lay down the methodological foundations of a Cognitive Discourse Analysis by matching the so-called usage-based Construction Grammar with a systematic approach to digital corpora in order to allow a highly precise observation of discourse, understood here as a cognitive structure in the form of a discourse frame. The proposed argumentation will be illustrated by a micro-qualitative analysis of the construction <[[X] + [sur] + [Y]]>, such as "je suis sur Paris".

La présente contribution se propose de poser les jalons méthodologiques d'une Analyse Cognitive de Discours en appareillant la Grammaire de Construction dite *usage-based* à une approche systématique de corpus numériques pour permettre une observation à haut degré de granularité du discours, compris ici comme une structure cognitive ayant la forme d'un frame discursif. L'argumentation proposée sera illustrée par l'analyse microqualitative de la construction <[[X] + [sur] + [Y]]>, telle que « je suis sur Paris ».

Cognitive Linguistics – Construction Grammar – Cognitive Discourse Analysis – Linguistics of Corpus

Linguistique Cognitive – Grammaire de Construction – Analyse Cognitive de Discours – Linguistique du Corpus

1. Éléments de Contexte¹

La présente contribution entend démontrer comment l'analyse de discours peut se muer en une Analyse *Cognitive* de Discours en s'appuyant sur l'un des développements majeurs de la linguistique cognitive qu'est la grammaire de construction (abrégé ci-après en GxC). Les constructions (Goldberg, 2019 ; Lasch, 2016, 2018+, abrégé ci-après en Cx) sont des structures linguistiques et cognitives holistiques encapsulant l'ensemble des connaissances individuelles et supraindividuelles en un réseau nommé le constructicon (cf. Lyngfelt *et al.*, 2018).

Dans la perspective de la GxC *usage-based* (Goldberg, 2019), les Cx sont des structures strictement sémantiques et syntaxiques qui n'intègrent *de facto* pas de paramètres pragmatiques et discursifs. Néanmoins, la recherche récente (Bach, e.p.a, b) a démontré le lien

¹ Nous tenons à remercier ici les deux relecteurs anonymes qui, par leurs critiques constructives, ont permis d'améliorer grandement le présent travail.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

entre des discours et des Cx plus ou moins spécialisées. On postulera donc à la suite de ces travaux que c'est à partir du discours (*sensu* LONGHI 2018, cf. Bach, 2019a) qu'est mobilisée une construction au profit d'une autre et que les lexèmes ou les termes sont actualisés ; ainsi, l'analyse des Cx peut servir une analyse (cognitive) du discours.

2. Introduction à la grammaire de construction *usage-based*

La notion centrale des GxC est la Cx. Une Cx est un schéma holistique possédant une force conceptuelle reposant sur une structure sémantique stabilisée couplée, par le biais d'une matrice cognitive de surface, à un schéma formel grammatical. La Cx possède une fonction intentionnelle particulière répondant à un besoin épistémolinguistique dans une situation de communication particulière, dans une culture et à un temps et un moment donnés. Une Cx n'est jamais isolée, elle est toujours prise dans un réseau de constructions (= le constructicon, Lasch, 2016 ; Goldberg, 2019) allant de Cx abstraites vers des Cx concrètes, et des Cx concrètes aux réalisations constructionnelles matérielles.

Une Cx est classiquement définie en reprenant la proposition de Goldberg (2019 : 7 [ma trad., MB]) :

Tout schéma linguistique est reconnu comme une construction tant qu'un aspect de sa forme ou de sa fonction n'est pas strictement prévisible à partir de ses composants ou d'autres constructions reconnues comme existantes. De surcroît, les schémas sont stockés en tant que constructions même s'ils sont entièrement prévisibles à condition qu'ils apparaissent avec une fréquence suffisante.²

Il s'agit donc d'un couple de sens et de forme ayant une fréquence élevée et qui est non strictement prédictible (cf. également Lasch, 2016, 2018+) comme le sont les figements (cf. p. ex. Bach, 2019b). On notera de plus que les constituants de la Cx possèdent des forces sémantique et syntaxique indépendantes majorées par la construction elle-même.

Un exemple classique de Cx est la Cx de transfert <X TRANSFERT Y à Z> généralement générée par le verbe ditransitif « donner », comme en (i) :

(i) Julianex donne_{TRANSFERT} le livre_Y à Mathilde_Z.

En (i), l'AGENT (Juliane), le prédicat d'ACTION actualisé par le verbe donner, l'OBJET EFFICIENT (le livre) et le BENEFACTIF (Mathilde) sont respectivement couplés aux rôles grammaticaux sujet, verbe, COD, COI de telle manière que l'on peut représenter la construction de la façon suivante :

² Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

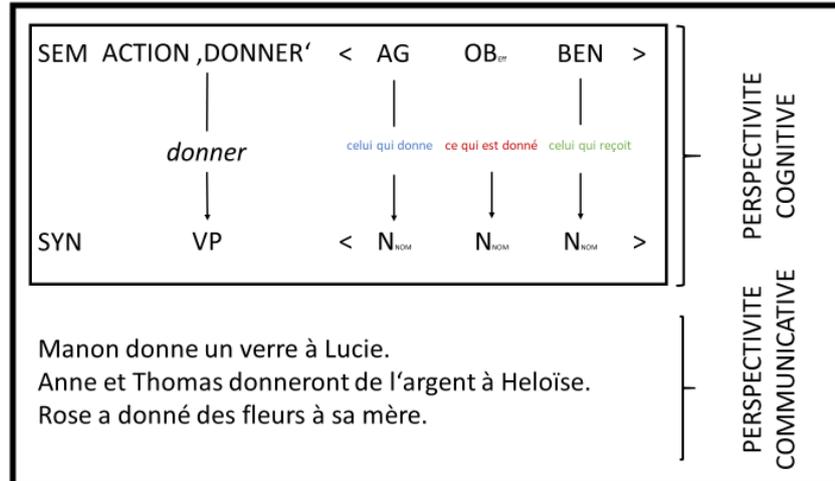


Fig. 1 : Construction de TRANSFERT instanciée au verbe 'donner'

Le caractère prototypique des Cx est à souligner ; pour la Cx de transfert, la figure 1 est son prototype, mais une telle structure de sens et de forme autorise un certain nombre de variations comme en (ii) et (iii).

- (ii) Anna-Lena a refusé l'offre d'emploi de son conseiller Pôle-Emploi.
- (iii) Simone a transmis le virus à toute sa famille.

En (ii), on constate que le BENEFACTIF (Anna-Lena) n'accepte pas l'objet transmis par l'AGENT ; or, ceci est un des paramètres de base de la Cx de TRANSFERT. Dans une veine similaire, l'exemple (iii) reprend cet élément et montre également que l'OBJET transmis n'est pas forcément physique et matériel. La Cx s'active et déploie son sémantisme même pour les exemples suivants :

- (iv) « vous vous approchez de votre partenaire pendant qu'il(elle) est assis(e), *donnez-lui un baiser* sur le front. »³
- (v) « Le suspense était mince, mais *la confirmation a dû être accueillie* avec soulagement du côté *des dirigeants* de la Mitchelton-Scott. »⁴

On reconnaît une structure relativement stable « donner un baiser » où ce qui est donné n'est pas immuable, et une structure où l'AGENT est un organisme international (l'Union Cycliste Internationale), les BENEFACTIFS (les dirigeants de l'équipe cycliste) sont multiples et l'OBJET (une décision) est un élément volatil. Toutefois, les deux énoncés sont compréhensibles grâce à la *puissance* sémantique apportée par la Cx.

Ce qui nous intéresse dans la présente contribution est la sélection d'une Cx au profit d'une autre et/ou l'actualisation d'un lexème par rapport à un autre (pourquoi l'on « donne » et non

³ <https://laviedesreines.com/amour/comment-embrasser-les-secrets-du-baiser-parfait/> (28 décembre 2019).

⁴ https://www.eurosport.fr/cyclisme/mitchelton-scott-confirnee-en-world-tour-pour-2020_sto7573792/story.shtml (28 décembre 2019).

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

pas l'on « jette » un baiser ?). Lasch (2018+) propose d'intégrer dans la configuration de la Cx les *Frame Elements* (abrégés ci-après en FE), qui sont des composants sémantiques du frame mobilisé par la situation d'énonciation.

La notion de frame repose sur l'idée basique qu'il est impossible de comprendre le sens d'un mot, d'une phrase, d'un texte sans avoir accès à différentes connaissances minimales (linguistiques, épistémiques, etc.). Ces connaissances sont continues dans un frame en ce sens qu'elles sont à disposition du frame à tout moment et qu'elles sont liées les unes aux autres. Il s'agit alors d'une structure cognitive cohérente composée de sous-concepts qui sans eux ne permettraient pas d'accéder à la totalité sémantique d'une unité linguistique. Le frame est composé de valeurs figées immuables et de valeurs variables, les deux évoluant avec l'expérience de communication. Les comparaisons entre les variables, entre le prototype de la variable et la réalisation, etc. sont hautement pertinentes en analyse de discours (cf. Ziem, 2008).

On peut définir les FE comme des actualisations matérielles d'un frame plus profond sémantiquement, pour nous il s'agit du frame discursif (cf. *infra*) ; concrètement, il s'agit en (i) des rôles communicationnels d'un individu qui donne (le DONNEUR), d'un individu qui reçoit (le RECEVEUR) et d'un objet (OBJET DONNÉ) qui est transféré volontairement du DONNEUR au RECEVEUR. Ces rôles communicationnels sont stockés dans le frame discursif et injectés dans la Cx pour générer le construct. Sans cette injection, il ne serait pas possible de passer d'une entité conceptuelle à un énoncé linguistique comme attendu par la communauté langagière dans une situation de communication précise.

Bien sûr, dans ces exemples simples cette intégration n'a pas grand intérêt ; en revanche, dans une approche plus globale telle que celle exposée *infra* la plus-value d'un tel ajout est considérable.

3. Du constructicon au discours

Dans la perspective fonctionnelle-cognitive qui est la nôtre, il est unanimement reconnu que les Cx apparaissent sous la forme d'un vaste réseau, qui est appelé le constructicon (cf. Goldberg, 2019, Lasch, 2016 pour un exemple avec les Cx non-agentives, et la Fig. 6 du présent travail). Ce réseau est une hiérarchie organisant les connaissances du plus abstrait au plus concret de telle manière qu'une Cx concrète hérite des spécificités des Cx abstraites qui l'ont générée.

À cet égard, il est fondamental de s'interroger sur l'organisation du réseau et sur son utilité cognitive. À la suite des travaux cités, nous postulons que le réseau dispose d'une organisation verticale allant du plus abstrait au plus concret :

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

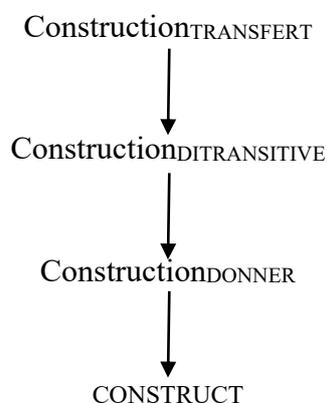


Fig. 2 : Déploiement d'un segment simplifié du constructicon

La question essentielle est dès lors la suivante : comment le système conceptuel est-il capable de sélectionner ces différentes Cx^5 et comment le construct est-il correctement réalisé, c'est-à-dire comment la combinaison lexicale adéquate est-elle sélectionnée parmi le vaste répertoire stocké chez le locuteur natif ?

La solution que nous proposons implique l'utilisation de l'entité supérieure qu'est le discours. Nous entendons le discours à la suite de Longhi (2018) et développons son approche phénoménologique pour notre objet d'étude (cf. Bach, 2019a, b) de telle manière que l'on peut définir le discours comme une structure conceptuelle hautement organisée autour de paramètres textuels, sociaux, culturels, pragmatiques et épistémiques.⁶

Longhi (2008, 2011, 2018) développe l'école française d'analyse de discours autour de sa Théorie des Objets Discursifs. Pour résumer cette théorie, il convient de distinguer quatre notions principales : les *topoi*, le sens commun, la doxa (et l'idéologie, que nous laissons de côté ici).

Les *topoi* sont des atomes de sens socialement structurés par l'expérience sociolinguistique. Réunis en un ensemble à l'occasion de situations de communication spécifique, ils forment un ensemble dont le sens commun est le concept structurant. Le sens commun renvoie à la totalité des structures normatives invoquées par les locuteurs d'une communauté d'individus dans la réalité langagière. La doxa est chez Longhi une région topique délimitée du sens commun.

Le sens commun est la porte d'entrée privilégiée au discours, dans la mesure où il réunit des événements discursifs matériels. Il est alors l'écho sociolinguistique de la dimension

⁵ Le développement conceptuel aurait pu sélectionner la construction transitive avec marquage spécifique du bénéfactif (p. ex. (i') : Juliane donne à Mathilde le livre, ou encore (i'') : C'est à Mathilde que Juliane donne le livre).

⁶ Il convient de souligner dès à présent que notre approche se distingue en tout point de la proposition de Ghigliione *et al.* (1995), et ce autant au niveau théorique, que méthodologique.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

épistémique du discours. Selon la perspective choisie, le sens commun permet (i) aux locuteurs de comprendre le sens transféré lors de l'échange communicationnel, et (ii), en production, de déployer le sens transféré de manière adéquate.

Dans cette théorie, le sens commun et le discours actualisent une position hautement sociale et normative, mais sont également le reflet d'un ensemble de positions individuelles, et ce en synchronie comme en diachronie. Ce faisant, la reconstruction du sens commun nécessite le recours à un corpus d'énoncés authentiques, qui actualise un discours dans une situation de communication. À ce stade, Longhi renvoie vers la phénoménologie afin de prendre en compte l'ensemble de la structure pragma-sémantique déployée, par le discours en tant que sens commun, dans la situation d'énonciation.

La philosophie phénoménologique repousse (Merleau-Ponty *i. a.* 1945) la traditionnelle dichotomie entre esprit et corps au profit d'une structure phénoménale unifiée dont les constituants sont interdépendants et significatifs. Le sens est autant le résultat de mécanismes cognitifs que de l'accumulation de perceptions sensorielles et corporelles. Plus encore, le sens transféré est la structure conceptuelle émergente de la réunion de ces différents segments de sens. Ainsi, l'appréhension du monde et des objets du monde est le résultat d'une perspective individuelle située, corporelle et épistémique, contrainte par l'expérience de l'environnement local de l'individu et normée par les superstructures discursives et sociales (et culturelles).

L'analyse de discours dans la perspective de Longhi permet précisément de faire ressortir ces ressorts conceptuels transcendants et surtout interindividuels, c'est-à-dire sociaux (voire culturels).

En repartant de cette base, nous poursuivons l'idée phénoménologique d'un point de vue cognitif : cela revêt alors l'idée d'une cognition incarnée (au sens de *embodied*, cf. Lakoff, Johnson, 1999). La cognition humaine est le reflet de l'expérience humaine de la vie sur Terre, et en parallèle, elle contraint *ipso facto* l'expérience de la vie. Cela est à coupler à l'axiome désormais largement acquis que le savoir, c'est la langue et réciproquement (Ziem, 2008) ; or, comme la cognition repose essentiellement sur la manipulation, la combinaison et la hiérarchisation des connaissances, on peut raisonnablement postuler que la langue façonne la cognition qui façonne l'expérience de vie et réciproquement. Ainsi, l'étude de la langue permet une observation de la cognition des locuteurs et des savoirs portés par ces derniers.

Cette argumentation strictement théorique est vérifiable selon la perspective apportée par diverses disciplines ; pour ce qui est de la linguistique, il semble nécessaire de faire intervenir la notion de discours afin de rester cohérent avec la tradition sémantique cognitive contemporaine et de permettre une représentation rigoureuse de la réalité cognitive. En effet, il est peu probable que les connaissances ne soient pas organisées.

Ainsi, nous revendiquons la thèse que les connaissances sont déployées et organisées en discours, que chaque discours est indexé à un domaine socio-économique correspondant à des situations de communication similaires. Ainsi, un segment d'une langue naturelle (*sensu* potentialité cognitive prototypique actualisée prototypiquement en énoncés) n'est analysable que par la réalité langagière exprimée lors de situations de communication relativement similaire, compilée en un corpus cohérent et homogène ; chaque corpus est un outil permettant l'observation d'un discours spécifique (Longhi, 2018) ; chaque discours est une structure conceptuelle organisée autour de différents paramètres linguistiques.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

Ainsi, l'analyse de constructs et l'analyse des Cx permettent d'observer le discours systématiquement et rigoureusement en remontant du texte au discours, avec la nécessité d'accéder à la réalité langagière et communicationnelle par les corpus ; ce qui implique une réflexion méthodologique.

4. Étude de cas microqualitative d'une construction du français moderne général réalisé sur Twitter

Les sciences humaines et sociales deviennent toujours plus des sciences fondées sur l'empirisme afin de gagner en rigueur et exactitude, et se rapprochent des sciences dites dures. Dans cette quête de la scientificité, les bases de données authentiques, organisées et représentatives de la réalité prennent une importance fondamentale. C'est ainsi que la discipline linguistique a pleinement exploité les évolutions d'une de ces branches : la linguistique de corpus (traduction aussi neutre linguistiquement que théoriquement de *corpus linguistics*).

Se sont alors développés des super-corpus de plusieurs millions voire milliard de mots pour la plupart des principales langues naturelles (cf. Renouf, 2007 pour un aperçu de l'évolution des corpus). Dans le même temps, des corpus ad hoc plus ou moins spécialisés ont commencé à émerger au fur et à mesure que la technologie numérique s'est développée (Lobin, 2018) et cela a permis de simplifier et d'individualiser les processus de requêtes par le truchement de logiciels simples d'utilisation et gratuits (cf. pour une présentation Look, 2016).

Ce faisant, la linguistique s'est transformée d'une science du langage introspective en une science des données (quand ce n'est pas de la *data* !) hors-sol, ce qui n'a pas résolu le problème de représentativité des analyses et des interprétations linguistiques, mais seulement déplacé le problème sous couvert de la scientificité et de la numérisation / automatisation des procédés analytiques. Cette linguistique des données a perdu de vue l'idée même que l'énoncé ne saurait être un produit hors-sol, simple suite de lettres ou de sons dans une analogie à une ligne de code informatique faite de zéro et de un.

Mukherjee (2010 : 371) fait également le constat d'une dérive de la linguistique de corpus et souligne que bien que l'on ne puisse que se satisfaire de l'intégration croissante des corpus en linguistique pour toutes les raisons détaillées jusqu'à présent, il n'en reste pas moins indéniable que certaines approches sont moins pertinentes que d'autres, et que le risque de s'éloigner du langage au profit des chiffres est certain :

Pour moi, l'implication est évidente : les linguistes qui effectuent des analyses statistiques quantitatives de données de corpus (en utilisant, par exemple, le logiciel R, mais on pourrait aussi ajouter des scripts Perl ou tout autre logiciel standard pour l'analyse de corpus) ne s'intéressent pas vraiment à la langue (ou au sens de la langue), mais ce sont des statisticiens qui se trouvent à travailler avec des données linguistiques.⁷ (MUKHERJEE 2010 : 374 [ma trad., MB])

⁷ To me, the implication is obvious: linguists who engage in quantitative-statistical analyses of corpus data (using, for example, R software, but one could also add Perl scripts or any standard software package for

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

Il y a donc un développement en deux parties pour les corpus qui se profilent pour les années à venir (MUKHERJEE 2010 : 376).

- (i) Les corpus vont prendre, d'un côté, une coloration transversale (comme nous l'avons souligné *supra*) et irriguer tous les champs des sciences du langage. Il s'agira davantage d'une méthode : création d'un protocole de collecte, collecte de données, valorisation des données à partir de théories, etc.
- (ii) D'un autre côté, la linguistique de corpus va se constituer en discipline et développer son propre corpus théorique. L'enjeu consiste alors en l'articulation de la méthode ou la méthodologie de corpus avec l'analyse linguistique, qui tend de plus en plus à être orientée vers des questions cognitives.

L'énoncé est, on l'a vu, un produit phénoménologique, incarné et situé (Lakoff, Johnson, 1999), devant être examiné en tant que tel et par conséquent être remis dans son contexte *lato sensu*. Il convient donc de manipuler avec prudence les bases de données et de conférer au critère de représentativité des données vis-à-vis de l'objet d'étude (Biber, 1993) un rôle primordial dans la constitution des corpus. Cette nécessité impose également de ne pas s'attaquer à des corpus de plusieurs millions de mots, illisibles aux yeux humains, et que seul un logiciel puisse traiter, au risque de mettre en péril la qualité de l'analyse (cf. l'éloquent argumentaire de Adamzik (2016 : 317) qui montre par l'exemple les variations observables entre différents logiciels ne serait-ce que pour le comptage des mots d'un texte simple).

Ce faisant il convient de marquer la distinction entre la linguistique *de* corpus et la linguistique *du* corpus à la suite de Mukherjee (2010). Cette dernière est une approche hypothético-déductive visant à la collecte de données en fonction d'une hypothèse générale de recherche. Une première observation du corpus permet de raffiner la problématique et éventuellement de proposer des questions de recherche plus précises, afin de s'orienter vers l'une ou l'autre des théories linguistiques. Cette première phase inductive est suivie d'une étape déductive, où les analyses sont menées en accord avec les théories choisies au préalable. Une telle façon de faire permet, de plus, d'évaluer les théories au regard de la réalité langagière. Cette approche peut alors être qualifiée de *human-based* et ne laisse pas la clé de l'analyse aux ordinateurs et aux programmes informatiques, même si elle s'en sert à certains égards.

Le corpus est dès lors compris comme une collection d'énoncés similaires traités pour être analysables et stockés numériquement. Le corpus, dans notre perspective, s'enrichit de données extralinguistiques, communicationnelles, environnementales, de telle manière que la finitude de la collection ne correspond pas à la finitude conceptuelle du corpus. Celui-ci devient « sans frontière » (LONGHI 2018 : 38–68) et permet un passage permanent entre l'énoncé et son co(n)texte.

corpus analysis) are not really interested in language (or meaning in language), but they are statisticians who happen to work with linguistic input.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

En d'autres termes, la linguistique du corpus est une linguistique outillée qui, en remplaçant le Stabulo d'autrefois, rend l'analyse plus rapide, la détection de récurrences plus efficace et plus rigoureuse ; bref, l'outil informatique, quel que soit le logiciel employé, est une aide à l'analyse dans ce qu'elle a de fastidieux. Cela fait, l'interprétation revient au linguiste, qui seul peut atteindre le degré de granularité le plus élevé dans l'observation de phénomènes langagiers *a fortiori* sémantiques. Et comme la langue est davantage un vecteur de sens – nous transférons bien des informations pour atteindre un but sociocommunicatif – qu'un vecteur de forme, c'est l'analyse du sens qui doit primer sur la forme, de telle manière que la forme reflète le sens et non l'inverse). Cette finesse d'analyse ne peut pas être atteinte par l'outil informatique sauf s'il a déjà été programmé pour identifier le phénomène langagier, auquel cas l'intérêt de l'analyse linguistique s'amointrit.

La linguistique du corpus se veut donc être une linguistique raisonnée se fondant sur des corpus plus ou moins spécialisés à taille humaine, où l'analyste reste le linguiste en tant qu'expert du matériel langagier (cf. notre argumentaire *in* Bach, 2019a, e.p.a et Gautier *et al.* 2019) pouvant être accompagné d'un expert de la spécialité, *modulo* le degré de spécialisation du corpus étudié.

La figure 4 résume le protocole d'étude esquissé dans cette introduction à la linguistique du corpus :

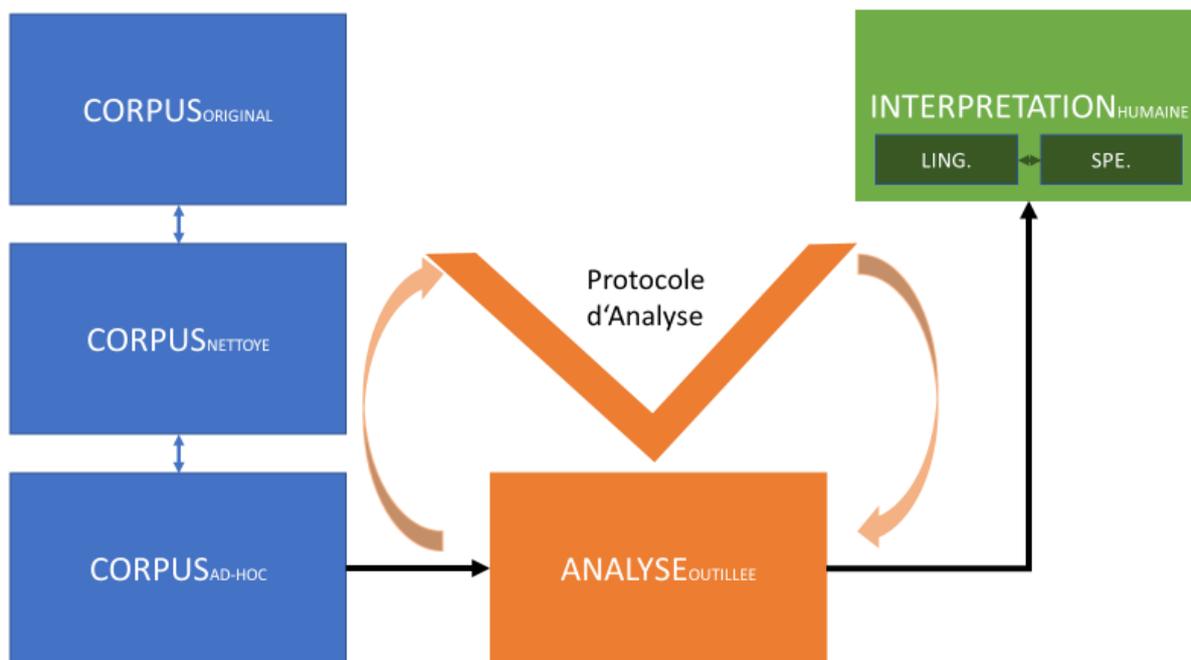


Fig. 4 : La linguistique du corpus à l'épreuve de l'analyse langagière

4.1 Twitter comme Corpus

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

Ainsi, nous employons ici Twitter⁸ non pas comme un corpus de référence, ni comme un corpus possédant une quelconque légitimité sur le français, mais comme un réservoir d'exemples réels et comme terreau objectif et connecté à la réalité langagière, d'où peuvent germer des constatations linguistiques menant des théories qui se devront d'être validées empiriquement sur des corpus plus homogènes représentatifs de la langue telle qu'elle est effectivement utilisée par les locuteurs.⁹ Il nous faut donc convenir que les énoncés collectés sur Twitter ne correspondent qu'à une partie de la potentialité socio-linguistico-épistémique de la langue française ; dès lors, ces énoncés représentent un échantillon de la langue française et ne sont donc pas représentatifs de cette dernière dans sa totalité.

4.2 Protocole d'analyse

Dans ce contexte, nous proposons le protocole d'analyse suivant. À partir d'un corpus général de tweets, il convient de faire un premier tri afin de retirer le bruit, c'est-à-dire les éléments ne répondant pas aux critères de qualité du corpus. Dans notre cas, il s'agit de retirer les tweets en langue étrangère ou qui emploient la proposition « sur » dans son sens traditionnel :

- (1) Los casos nuevos de coronavirus en Corea del Sur caen por debajo de 30 por tercer día consecutivo
- (2) Salut tu n'aurais pas des infos sur Dijon ?
- (3) Petite vanne sur Marseille mais largement sa place dans club légendaire

Une fois ce nettoyage réalisé, il convient de procéder à une analyse en « première lecture », il s'agit de faire un nouveau tri, cette fois qualitatif, afin de sélectionner les éléments qui correspondront au corpus d'étude. Le degré de granularité de ce tri correspond au degré de qualité du corpus et assure *de facto* le degré de qualité de l'analyse et du travail final.

On remarquera que le corpus *ad hoc* peut revêtir un caractère plus ou moins spécialisé selon les critères de sélection des composants du corpus : à partir d'un corpus tout à fait général, il est possible de ne sélectionner que les tweets relevant de la branche du vin. Le corpus d'étude sera *ipso facto* relativement spécialisé. Il est ensuite possible de ne sélectionner que les tweets rédigés par des « experts », ce qui permet de déplacer le curseur du continuum du spécialisé vers le « plutôt très spécialisé ». Le corpus *ad hoc* représente dès lors l'échantillon de la réalité langagière que le chercheur souhaite analyser.

Cette étape fastidieuse, mais fondamentale réalisée, l'étude peut entrer dans sa phrase analytique.

⁸ Pour une étude sur les pratiques sociolinguistiques sur Twitter, nous renvoyons à la monographie de Dang-Ahn (2019) et pour des réflexions générales sur les formes et pratiques numériques, cf. la monographie de Lobin (2018).

⁹ Sauf, bien entendu, si l'objectif est d'analyser le français sur un support numérique, tel que Twitter.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

L'analyse du corpus – quel que soit l'objectif final de l'étude : morphologique, textuel, discursif, etc. – nécessite également un protocole d'analyse spécifique à l'objectif d'étude ce qui permet de répliquer l'analyse et lui assure un haut degré de scientificité (cf. Popper, 1945, Kuhn, 1962). On pourra ainsi partager le protocole d'analyse sur des plateformes de partage telle que OSF (*Open Science Network*)¹⁰.

Il convient ensuite d'améliorer ledit protocole d'analyse voire l'appareil théorique afin d'améliorer l'ensemble de l'étude. Pour un exemple d'évolution du protocole d'analyse en sémantique propositionnelle, car la première version ne permettait pas d'accéder à la totalité de la structure sémantique, cf. les études de Bach (2017, soumis).

Cette étape théorique intermédiaire améliore le matériel analytique et permet de réaliser une seconde étude, qui sera cette fois finale. Cette étude se doit de prendre en compte le contexte d'énonciation (c'est-à-dire l'environnement¹¹ *lato sensu*, la situation de communication en elle-même, l'architecture socioconceptuelle déployée, etc.) et le cotexte, c'est-à-dire l'environnement phrastique immédiat. Cela est en effet nécessaire dans la mesure où nous nous positionnons dans la perspective discursive de Longhi (2018) et, donc, de la philosophie phénoménologique de Merleau-Ponty (1945) et cognitive de Lakoff et Johnson (1999).

Le produit de cette analyse est le produit final sur lequel se base l'interprétation linguistique et, *modulo* le degré de spécialisation du corpus, l'interprétation de la perspective du spécialisé.

La conséquence immédiate et évidente est la prise en compte de ces données extralinguistiques dans l'analyse. Pour dresser un parallèle avec la *Grounded Theory* (GLASER, STRAUSS 1967), les données extralinguistiques et les données linguistiques sont dans un même rapport que les données et la théorie : les premières enrichissent les secondes. Dans cette perspective, les modèles théoriques ne sont réalisés qu'à partir de la réalité langagière ; ce sont les données expérimentales qui façonnent l'outil théorique. De ce dernier découle un outil analytique, qui sert à l'analyse des données. Il s'agit alors d'une phase test. Une fois la théorie et l'outil validés, il est possible d'abord d'améliorer l'un ou l'autre, puis d'élargir le corpus et/ou de le diversifier pour systématiser la théorie ou l'analyse. Le corpus dispose dans cette perspective d'un rôle prépondérant permettant d'aboutir à une augmentation significative de la scientificité et de la rigueur des travaux.

Ayant pris ces précautions et en ayant en tête ces paramètres méthodologiques, nous nous proposons de mettre en évidence une Cx du français : <[[X] + [sur] + [Y]]>.

¹⁰ https://osf.io/?view_only= (17 février 2020).

¹¹ Y compris au sens technique de Paveau (2013), qui montre l'importance particulière de la prise en compte de l'environnement (et de la « contextualisation technorelationnelle » de l'énoncé) dans l'analyse d'énoncés sur support numérique (cf. également Longhi, 2020).

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

4.3 Exemple d'analyse¹²

Dans cette section, nous étudions la Cx <[[X] + [sur] + [Y]]>, où X est un verbe (de position ou de mouvement) associée à la préposition « sur » suivi de Y qui est un lieu. L'exemple prototypique pour l'expression de la position est « je suis sur Paris », pour l'expression du mouvement « je vais sur Paris ». S'il est difficile de dater l'émergence de cette Cx, il convient de constater que son emploi n'est pas récent ; par exemple, Olivier de Kersauson employait ponctuellement la Cx « [[X] + [sur Paris]] » dans les émissions des années 1990 des *Grosses Têtes* sur RTL.

4.3.1. De l'usage de la construction <[[X] + [sur] + [Y]]>

L'Académie Française rejette néanmoins l'emploi de la Cx et s'en explique dans un article reproduit *verbatim* :

La préposition sur ne peut traduire qu'une idée de position, de supériorité, de domination, et ne doit en aucun cas être employée à la place de à ou de en pour introduire un complément de lieu désignant une région, une ville et, plus généralement, le lieu où l'on se rend, où l'on se trouve. (ACADEMIE FRANÇAISE 2011)

Le premier exemple contrastif donné étant « je suis à Paris » (emploi accepté) et « je suis sur Paris » (mésemploi).

Entre la règle de l'Académie et l'emploi, il est fondamental de se tourner vers un corpus pour apprécier l'usage réel de cette Cx. Ainsi, Twitter comme corpus dynamique offre un accès à la réalité langagière dans un état d'instantanéité que l'on pourrait qualifier d'hypersynchrone.

Une première étude (Bach 2019c) a mis en évidence la Cx « [[X] + [sur Paris]] », puis a montré dans un second temps que cette Cx n'est pas exclusivement réservée à la capitale, mais qu'elle est également employée pour d'autres villes françaises :

- (1) Après un passage *sur Nantes* puis Angers, le FC Bicyclette débarque les 7 et 8 février *sur Le Mans* !¹³
- (2) C'est bon sa : La semaine prochaine je descend *sur Montpellier* voir mon frere , Des bons podcast en perspective
- (3) QUEL PLAISIR DE REVENIR *SUR LILLE*
- (4) « dans le Centre (-43%), à Strasbourg (-35%), en Bretagne (-34%), à Toulouse (-30%), en Normandie (-29%) et *sur Lille* (-25%). » (frTenTen17 in Bach 2019c)

¹² En plus du corpus Twitter, nous utilisons comme corpus témoin le corpus FrTenTen17 de 6,8 milliards de mots hébergé sur Sketch Engine.

¹³ Tous les exemples sont issus de tweets authentiques non corrigés.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

On peut se demander pourquoi les locuteurs ont eu besoin d'inventer une nouvelle Cx alors que la Cx canonique « verbe + à + ville » est productive, efficace et plus économique, car plus courte. (4) montre qu'au sein d'une même proposition il existe une différence entre les prépositions « à » et « sur », car l'action est effectuée « à Strasbourg », mais « sur Lille ». On peut donc postuler une différence liée au niveau de précision sémantique voulant être transférée par le locuteur au regard de son rapport phénoménal avec l'entité géographique concernée. La Cx avec « à » indique une localisation précise, alors que la Cx avec « sur » semble être moins précise.

L'analyse de ces exemples nous pousse ainsi à proposer deux hypothèses discursives pour la sélection exclusive d'une Cx plutôt qu'une autre (c'est-à-dire l'introduction du groupe prépositionnel par la préposition « sur » au profit de la préposition « à »).

La première évoquée ci-dessus est le mode de saisie de la localisation – c'est-à-dire la mise en perspective explicite de l'appréhension cognitive du sujet sur son rapport au monde (limité ici à une ville) – que le locuteur souhaite transférer :

- est-ce un point fixe ? Alors, on aura tendance à employer la Cx avec « à » ;
- est-ce un point inconnu ou possiblement mobile dans le temps ? Alors, on préférera l'emploi de la Cx avec « sur ».

D'autres exemples attestent ce raisonnement :

- (5) Nous voyons que sur Marseille des actions de solidarités sont mis en place un peu partout. Bravo à tous.
- (6) Ben oui si ils y en a qui veulent venir sur Marseille et qui galèrent, si je peux vous aider dans mes capacités « of course » n'hésitez pas

La Cx est limitée aux villes et pas à d'autres entités géographiques, on peut ainsi faire l'hypothèse, d'un point de vue diachronique, que cette innovation linguistique suivrait la logique d'expansion des grandes agglomérations françaises (cf. Biarez, 1989 d'un point de vue politique ; Le Jeannic, 1997 d'un point de vue géographique ; INSEE 2018 d'un point de vue démographique).

Une deuxième hypothèse est le degré de précision connu par le locuteur :

- le locuteur sait-il où il va *précisément* ? Si oui, il aura tendance à choisir la Cx avec « à » ; si non, c'est la Cx avec « sur » qui tendra à être sélectionnée ;
- le locuteur a-t-il *vraiment* l'intention de ne rester qu'à un point précis de la ville dans laquelle il se rend ? Si oui, la Cx avec « à » sera plutôt sélectionnée ; si non, c'est la Cx avec « sur » qui prendra le dessus.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezone/ezone_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

Au-delà de la manipulation, de la transformation et du transfert d'informations conceptuelles, il nous faut ajouter un éclairage pragma-situationnel (autrement dit, l'intention du locuteur) pour l'analyse de la génération d'une Cx.

Cette analyse montre bien, à notre sens, tout l'intérêt d'une analyse constructionnelle et de l'apport de cette dernière dans un appareil analytique éventuellement plus complexe, en vue d'une analyse de discours empirique.

L'analyse fine du corpus a, néanmoins, fait ressortir un phénomène qu'il nous faut remarquer : la Cx avec « sur » est activée pour signifier non seulement le déplacement, mais également pour la localisation sans déplacement, tel que le met en évidence la figure suivante.



Fig. 5 : Exemples de la construction <[[X] + [sur] + [Y]]> pour exprimer la localisation

Pour la recherche d'appartements ou d'emplois, la différence entre « sur » et « à » est tout à fait notable, car elle implique une variation discursive fondamentale. En effet, une future locataire ou propriétaire ne recherche pas la même catégorie de biens si elle cherche « à + VILLE », c'est-à-dire dans une région plutôt centrale, en tous les cas à l'intérieur de la ville, ou si elle cherche « sur + VILLE », autrement dit la région de la ville. Il en va de même pour la recherche d'emploi par exemple.

Pour d'autres énoncés, comme en (7), l'utilisation de « sur » plutôt que de « à » reflète plus précisément la réalité extralinguistique de la proposition sémantique :

- (7) mon bb tu me manques trop grave hâte qu'on se revoit sur metz .
- (8) On organise une distribution d'habits chauds pour les sans abris sur Dijon, Dole et Besançon !

Les maraudes ne restent jamais à un point fixe dans une ville, elles sont dynamiques dans le temps et l'espace ; l'emploi de « à » est, semble-t-il, contre-intuitif, et c'est bien la préposition « sur » qui semble la plus adaptée.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

Il convient de souligner ici une faille dans l'outil d'analyse qui n'était pas prévue par l'objectif initial de l'étude : la Cx a en première instance été observée pour exprimer le déplacement d'un lieu à un autre, avec la contrainte que l'endroit d'arrivée soit une ville relativement importante.

L'analyse des corpus a mis en évidence que la Cx est actualisée pour exprimer les déplacements dans la ville et également la localisation sans mouvement dans la ville.

4.3.2. De la construction au discours et réciproquement

L'analyse constructionnelle nous permet d'accéder au discours comme lieu de stockage dynamique de connaissances multimodales. La Cx licenciant la préposition « sur » permet de renvoyer dans la situation de communication une appréhension phénoménale polarisée situationnellement de l'individu :

- (9) Ça serait le comble pour moi de commencer une love story sur Montpellier alors que je vais sûrement déménager
- (10) Je recherche avec mon conjoint un appartement de 40 à 50m² sur Caen dans le secteur Calvaire st Pierre/péricentre/Campus1 à louer.
- (11) je viens de recevoir mon jogging blanc il est transparent, super faut que je le renvoie, y'a t'il des postes ouvertes sur Besançon?

En (9–11), les locuteurs ne renvoient pas à un point géographique précis, mais à une référence épistémique dont la dénomination est une ville. Autrement dit, ils montrent aux interlocuteurs que leur énoncé n'est pas borné (conceptuellement et par conséquent géographiquement), mais qu'il est large, mobile, voire dynamique.

Le traitement discursif opéré avant la sélection constructionnelle – actualisé dans l'énoncé – permet une prise de position de l'individu dans le monde et sur les objets du monde qui le touchent :

- (12) Il n'y a pas *qu'à Paris, sur Montpellier* c terrible

En (12), l'individu vise une entité géographique précise, un point très concret, c'est-à-dire Paris, et l'associe à un point plus vague géographiquement articulé autour d'une entité connue, celle de Montpellier. On obtient donc une différence de traitement cognitif qui s'explique pragmatiquement et conceptuellement par l'emploi d'un signe linguistique spécifique reflétant cette divergence conceptuelle dans la matérialité énonciative du texte.

L'analyse de discours est dans sa forme ultime une analyse des connaissances (Busse 2012, 2018, Ziem, 2008) : l'ACD telle qu'elle a été théorisée ici permet une analyse textuelle des énoncés pour reconstruire la structure discursive permettant le transfert de connaissances entre

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

individus. Les connaissances sont sélectionnées en discours (dans le frame discursif) selon les nécessités de la situation de communication ; elles sont ensuite injectées dans la réalité langagière et communicationnelle par la matérialité d'un artefact langagier, ici un texte numérique, qui s'effectue par la sélection spécifique et le déploiement de constructions.

4.3.3. Deux constructions coordonnées par [sur] ?

Outre, les emplois particulièrement éloignés du prototype de la réalisation constructionnelle, il semblerait qu'il n'existe pas une Cx mais bien deux Cx générant un sens non-compositionnel relativement peu prédictible (sans contexte).¹⁴

Une première Cx est déployée selon l'intention du locuteur de transférer l'idée d'un mouvement ou d'une position. Il lui faut alors spécifier si le mouvement est dynamique (d'un point A à un point B) ou s'il est local (p. ex. dans un arrondissement d'une ville). La perspective épistémique doit être explicitée : la focale était fixée sur le point de départ ou le point d'arrivée.

À ce moment du développement conceptuel, la deuxième construction, à savoir le groupe propositionnel signifiant la localisation, est mobilisée. Or, il est fondamental de sélectionner le premier composant du groupe qui est la préposition. Les deux constructions sont alors couplées pour accorder le prédicat, ici le verbe prédicatif, à la préposition, selon les contraintes syntactico-sémantiques des deux entités.

Les deux Cx sont à concevoir avec une forme fixe et forme variable (un slot : α) :

- $\langle [[X] + [\alpha]] \rangle$, qui est la construction verbale, et
- $\langle [[\alpha] + [\text{prep.}] + [Y]] \rangle$ qui est la construction propositionnelle locative.

Le discours entre en jeu à cet endroit du développement conceptuel pour injecter dans ce qui n'est pour l'heure qu'une structure relativement abstraite les atomes conceptuels contenant des informations conceptuelles générées par la situation de communication et le genre textuel. α est spécifié et articule les deux Cx.

Le développement conceptuel arrive à son terme en stabilisant une Cx, somme des deux Cx, générant un construct : la Cx prend sa forme matérielle déployée dans la réalité langagière.¹⁵

¹⁴ Nous tenons ici à remercier Alexander Lasch de nous avoir mis *sur* la bonne voie !

¹⁵ Il est à noter que cet argumentaire est encore peu fondé sur des faits empiriques ; il convient alors de multiplier les contributions de ce genre pour améliorer la réalité conceptuelle en se fondant sur l'observation de corpus variés.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

On peut alors représenter le segment du constructicon activé par la Cx <<[X] + <[sur] + [Y]>> comme suit :



Fig. 6 : Segment du constructicon

Dans tous les cas, il est désormais impensable de considérer que la Cx actualisée par la préposition « sur » soit une réalisation fautive de la langue française. Elle est bien davantage une innovation diachronique reflétant les besoins interactionnels, discursifs et par conséquent conceptuels des locuteurs francophones.

5. Conclusion

L'objectif de la présente contribution était de développer une approche rigoureuse d'analyse de discours en s'appuyant sur l'emploi de corpus numériques. Nous avons ainsi proposé un protocole d'étude reposant sur l'analyse systématique de Cx afin d'observer le discours que l'on comprend ici comme un frame discursif. Nous avons nommé cette approche Analyse Cognitive de Discours dans la mesure où elle allie le socle théorique de l'AD française de Longhi (2018) aux outils d'analyse et de représentations cognitifs de la sémantique cognitive.

Ce cadre méthodologique a été opérationnalisé pour une analyse systématique de discours numériques tels qu'ils existent sur les réseaux sociaux comme Twitter. On a pu démontrer que cette opérationnalisation nécessite une réflexion spécifique sur l'emploi des corpus et le développement d'un protocole d'analyse particulier.

L'exemple de la Cx <[[X] + [sur] + [Y]]> a mis en évidence l'intérêt d'une telle approche en démontrant la fiabilité du protocole, nous permettant ainsi d'observer que la Cx à l'œuvre est

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

en réalité la coactivation de deux Cx fusionnées : <<[X] + <[sur] + [Y]>>. Du point de vue discursif l'appareil théorique et analytique permet, à partir de l'étude d'énoncés pris dans leur contexte d'énonciation (i) d'expliquer le phénomène, (ii) de retracer la génération cognitive du phénomène, (iii) et de représenter le tout. Ainsi, la génération de la Cx s'explique par la mobilisation d'un segment du constructicon qui s'est construit diachroniquement en devenant une norme sociale, par l'injection de connaissances spécifiques (en particulier la nature du lieu – ce doit être une ville –, la taille de la ville – ce doit être une *grande* ville –, etc.) et par la contrainte discursive héritée de la double pression normative de la situation de communication et du genre de texte mobilisé dans cette dernière.

En utilisant un corpus compilé sur Twitter, l'étude de cette Cx repose sur une observation fine de la réalité langagière en hypersynchronie et le traitement d'un phénomène langagier en le caractérisant par des données concrètes, réelles, et actuelles. Le point négatif d'une telle étude est l'aspect biaisé du corpus dans la mesure où il n'est pas représentatif de l'entièreté de la production langagière française. D'autres études complémentaires sont donc nécessaires, et permettront de développer (i) le socle méthodologique de l'Analyse Cognitive de Discours, (ii) le constructicon pour le français moderne.

Ces avancées en linguistique fondamentale pourraient bénéficier concrètement, par exemple, à l'enseignement du français sur objectif spécifique (Bach, Wagener, 2019). En effet, l'enseignement aurait tout à gagner en se tournant vers les corpus, comme reflets de la réalité langagière d'une langue naturelle, et vers les Cx, comme structures de sens et de forme stabilisées et récurrentes dans cette réalité langagière.

Bibliographie

- ACADEMIE FRANÇAISE, « sur » in « Dire, ne pas dire », <http://www.academie-francaise.fr/sur>, 6 octobre 2011, (date de dernière consultation : 06 février 2020).
- ACHARD-BAYLE, Guy, « Texte, discours, cognition », *Semiotica*, n° 223, 2018, p. 71–86.
- ADAMZIK, Kirsten, *Textlinguistik. Grundlagen, Kontroversen, Perspektiven. 2. Auflage*, Berlin, Boston, De Gruyter, de Gruyter Studium, 2016.
- BACH, Matthieu, *Prototypicité discursive dans les discours de vente du vin. Étude contrastive français-allemand en Bourgogne et en Rhénanie-Palatinat* [Mémoire M2], Dijon, Université de Bourgogne, 2017.
- BACH, Matthieu, « Sensorial discourse and corpus in the digital humanities era: The example of the wine language », *Digital Scholarship in the Humanities*, 2019a, <https://doi.org/10.1093/llc/fqz006> [consulté le 12/07/2020].
- BACH, Matthieu, « Fixités thématiques et figements lexico-grammaticaux dans les discours numériques du vin en France et en Allemagne », *Interfaces Numériques*, n° 8, 2, p. 305–322, 2019b.
- BACH, Matthieu, « 'monter sur Paris' : analyses sur corpus d'une nouvelle construction », Site Internet Personnel, 2019c, en ligne : <https://matthieubach.wordpress.com/2019/09/21/%e2%80%89monter-sur-paris%e2%80%89analyses-sur-corpus-dune-construction-du-francais-moderne/> [consulté le 12/07/2020].

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

- BACH, Matthieu, « Fachtextmuster im Kontrast am Beispiel französisch-deutscher Online-Weinbeschreibungen », in GAUTIER, Laurent, VARGA, Simon, SCHREIBER, Michael (éds.), *Fachsprachen Kontrastiv*. Frankfurt/Main, Peter Lang, Kontraste/Contrastes, soumis.
- BACH, Matthieu, « 0. Discours spécialisés du sensoriel, une introduction », in BACH, Matthieu (éd.), *Discours du Sensoriel. Cognition, Expérience, Texte*, Berlin, Frank & Timme, *Forum für FachsprachenForschung*, e. p.a.
- BACH, Matthieu, « 8. Evoquer l'idée de luxe en mobilisant des constructions de spécialité. Analyses microqualitatives de descriptions de chambres de palaces parisiens », in BACH, Matthieu (éd.), *Discours du Sensoriel. Cognition, Expérience, Texte*, Berlin, Frank & Timme, *Forum für FachsprachenForschung*, e. p.b.
- BACH, Matthieu, WAGENER, Albin, « L'apprentissage du français sur objectif spécifique et le rôle de la terminologie à l'épreuve des discours de spécialité du vin », *Le Langage et l'Homme*, n° 54, vol. 2, 2019, p. 103–118.
- BIAREZ, Sylvie, « L'évolution du développement dans les grandes villes françaises », *Revue internationale d'action communautaire*, n° 22, 1989, p. 105–115.
- BUSSE, Dietrich, *Frame Semantik. Ein Kompendium*, Berlin, Boston, de Gruyter, 2012.
- BUSSE, Dietrich, « 1. Diskurs und Wissensrahmen », WARNKE, Ingo H. (éd.), *Handbuch Diskurs*, Berlin, Boston, de Gruyter, *Handbücher Sprachwissen*, n° 6, 2018, p. 3–29.
- DANG-AHN, Mark, *Protest twittern. Eine medienlinguistische Untersuchung von Straßenprotesten*, Bielefeld, transcript Verlag, *Locating Media/Situierte Medien*, n. 22, 2019.
- BIBER, Douglas, « Representativeness in Corpus Design », *Literary and Linguistic Computing*, vol. 8, n° 4, 1993, p. 243–257, en ligne : <https://doi.org/10.1093/lle/8.4.243> [consulté le 12/07/2020].
- GAUTIER, Laurent, BACH, Matthieu, MERIC, Olivier, « Données textuelles, données expérimentales : quels apports pour une sémantique de la sensorialité ? », Colloque PRELA 2019, Lyon, Juin, <hal-02165279>, 2019.
- GHIGHLIONE, Rodolphe, KEKENBOSCH, Christiane, LANDRE, Agnès, *L'analyse cognitivo-discursive*, Grenoble, Presses Universitaires de Grenoble, *La psychologie en plus*, 1995.
- GLASER, Barney G., STRAUSS d'Anselm L., *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Hawthorne, Aldine de Gruyter, 1967.
- GOLDBERG, Adele E., *Explain me this. Creativity, Competition, and the Partial Productivity of Constructions*, Princeton, Oxford, Princeton University Press, 2019.
- INSEE, « Entre 2011 et 2016, les grandes aires urbaines portent la croissance démographique française », *INSEE FOCUS*, n° 138, 2018, en ligne : <https://www.insee.fr/fr/statistiques/3682672> [consulté le 12/07/2020].
- KUHN, Thomas S., *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press, 1962.
- LAKOFF, Georges, JOHNSON, Mark, *Philosophy in the Flesh*, New York, Basic Books, 1999.

Bach, M. (2020) : « Pour une Analyse Cognitive de Discours. Twitter comme corpus, la grammaire de construction comme outil. » In : *Repères DoRiF* 22 [En ligne]. http://www.dorif.it/ezine/ezine_articles.php?id=494. (Creative Commons 4.0 (BY NC ND)).

- LASCH, Alexander, *Nonagentive Konstruktionen des Deutschen*, Berlin, Boston, de Gruyter, *Sprache und Wissen*, n° 25, 2016.
- LASCH, Alexander, *Überlegungen zu Konstruktionen als Bedeutungsformpaaren*, <https://www.gls-dresden.de/tag/bedeutungsformpaar>, 2018+.
- LE JEANNIC, Thomas, « Trente ans de periurbanisation : extension et dilution des villes », *Economie et statistique*, n. 307, p. 21–41, DOI : 10.3406/estat.1997.2578, 1997.
- LOBIN, Henning, *Digital und Vernetzt. Das neue Bild der Sprache*, Stuttgart, J.B. Metzler, 2018.
- LONGHI, Julien, *Objets discursifs et doxa. Essai de sémantique discursive*, Paris, L'Harmattan, *Sémantiques*, 2008.
- LONGHI, Julien, *Visées discursives et dynamiques du sens commun*, Paris, L'Harmattan, *Sémantiques*, 2011.
- LONGHI, Julien, *Du discours comme champ au corpus comme terrain. Contribution méthodologique à l'analyse sémantique du discours*, Paris, L'Harmattan, *Humanités Numériques*, 2018.
- LONGHI, Julien, « Explorer des corpus de tweets : du traitement informatique à l'analyse discursive complexe », *Corpus*, n° 20, 2020, en ligne : <https://journals.openedition.org/corpus/4567> [consulté le 12/07/2020].
- LOOCK, Rudy, *La traductologie de corpus*, Lille, Presses du Septentrion, *Traductologie*, 2016.
- LYNGFELT, Benjamin, BORIN, Lars, OHARA, Kyoko, TIMPONI TORRENT, Tiago (éds.), *Constructicography. Constructicon development across languages*, Amsterdam, Philadelphie, John Benjamins, *Constructional Approaches to Language*, n° 22, 2018.
- MERLEAU-PONTY, Maurice, *Phénoménologie de la perception*, Paris, Gallimard, Bibliothèque des idées, 1945.
- MUKHERJEE, Joybrato, « Corpus linguistics versus corpus dogmatism – *pace* Wolfgang Teubert », *International Journal of Corpus Linguistics*, n° 15, vol. 3, 2010, p. 370–378.
- PAVEAU, Marie-Anne, « Analyse discursive des réseaux sociaux numériques », in Dictionnaire d'analyse du discours numérique, Technologies discursives [Carnet de recherche], 2013, en ligne : <http://technodiscours.hypotheses.org/?p=431> [consulté le 12/07/2020].
- VON POLENZ, Peter, *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens*, 3. Auflage, Berlin, New York, de Gruyter, de Gruyter Studienbuch, 2008.
- POPPER, Karl, *The Open Society and Its Enemies*, London, Routledge, 1945.
- RENOUF, Antoinette, « Corpus development 25 years on: from super-corpus to cybercorpus », in FACCHINETTI, Roberta (éd.), *Corpus Linguistics 25 Years on*, Amsterdam, New York, Rodopi, 2007, p. 27–50.
- ZIEM, Alexander, *Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz*, Berlin, New York, de Gruyter, *Sprache und Wissen*, n° 2, 2008.