

## Acquisition de structures grammaticales en corpus

Rompant avec la thèse générativiste, des travaux récents en linguistique cognitive ont mis en évidence l'influence de la fréquence d'utilisation des structures linguistiques sur l'évolution de la langue et tendent à montrer que les locuteurs sont sensibles à ces informations d'usage : l'usage modèle la grammaire (cf. Langacker 1987, Bybee 2006). Ce changement de paradigme vers des modèles basés sur l'usage (*usage-based models*) a inévitablement conduit ses partisans à donner une place plus importante aux données empiriques, telles que celles fournies par un corpus. Dans la tradition générativiste, de telles données relevant de la performance n'ont pas leur place dans l'analyse linguistique dédiée à l'étude de la compétence, ce qui a conduit au rejet du corpus ; dans un modèle basé sur l'usage, la distinction entre compétence et performance étant abolie, cette critique est nulle et non avenue. L'augmentation considérable de la taille des corpus disponibles et l'automatisation informatique des traitements ouvrent de nouvelles perspectives aux linguistes en leur permettant d'accéder à des données naturelles nombreuses et diversifiées. Ainsi, de nombreuses études cognitives sont basées sur une grande quantité de données de corpus (Gries et Stefanowitsch 2003, 2005, Gries et al. 2005).

Cependant, comme Gries et al. (2005) le rapportent, certaines études exploitant les fréquences d'usage aboutissent à des résultats incohérents. En effet, l'utilisation des corpus dans un tel cadre théorique n'échappe pas aux autres écueils méthodologiques dénoncés par Chomsky et les générativistes, notamment que (i) en tant que fraction de l'ensemble infini des phrases d'une langue, un corpus est nécessairement déséquilibré et non représentatif et (ii) un corpus est incapable de présenter des preuves négatives (*i.e.* des phrases non grammaticales). Stefanowitsch et Gries (2003) soutiennent que ces limites sont uniquement dues à l'utilisation de fréquences brutes et donc biaisées, et pour remédier à ces problèmes, présentent l'analyse collostructionnelle, un ensemble de procédures statistiques permettant d'utiliser les fréquences brutes de manière appropriée. Grâce à cette méthode, les corpus peuvent être utilisés avec toute la rigueur scientifique nécessaire.

Dans cette présentation, nous partons de l'idée que les corpus devraient jouer un plus grand rôle dans l'analyse grammaticale. Suivant Stefanowitsch (2006), notre hypothèse de départ est qu'il est possible de dériver une grammaire à partir d'un corpus grâce à l'analyse collostructionnelle. Nous avons testé cette hypothèse dans le cas des constructions argumentales, *i.e.* une approche de la structure argumentale des verbes dans le cadre des grammaires de constructions. A partir de l'analyse de Goldberg (1995), nous avons conçu et implémenté trois indices statistiques de la forme de l'analyse collostructionnelle basés sur certaines caractéristiques essentielles des constructions argumentales. Nous avons testé ces indices sur le corpus ICE-GB (1 million de segments) afin de vérifier à quel point ils permettent d'identifier trois constructions de l'anglais préalablement annotées. Bien que les résultats affichent une tendance positive, ils montrent que l'analyse collostructionnelle requiert des ajustements supplémentaires ; de plus, l'étude a mis en évidence certaines incohérences dans la définition des concepts théoriques des grammaires de construction. Nous terminerons la présentation en offrant des perspectives de développements futurs vers une nouvelle technique d'exploration de corpus.

## Bibliographie

- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language* 82 (4), 711-733.
- Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gries, S. T., B. Hampe, and D. Schönefeld (2005). Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16 (4), 635-676.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Volume 1: Theoretical Prerequisites*.

Stanford University Press.

Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2.

Stefanowitsch, A. and S. T. Gries (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8 (2), 209-243.

Stefanowitsch, A. and S. T. Gries (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1 (1), 1-43.