

Position et structure du groupe adjectival en serbe : une approche empirique
Aleksandra Miletic, CLLE UMR 5263, Université de Toulouse & CNRS, France

En serbe, l'adjectif est canoniquement antéposé à sa tête nominale (cf. *puna kuća* lit. 'pleine maison', 'maison pleine'), mais il peut également être postposé (cf. *kuća puna dece* lit. 'maison pleine enfants.GEN', 'maison pleine d'enfants'). Cette capacité adjectivale n'est cependant pas traitée dans les grammaires de référence (cf. Stanojčić et Popović, 2012 ; Mrazović, 2009), qui se contentent de décrire la structure canonique du GN, sans évoquer cet ordonnancement alternatif. Dans le but de proposer une modélisation plus complète de ce phénomène, nous nous servons d'un treebank pour identifier, extraire et analyser les occurrences des GA à l'intérieur d'un GN. À notre connaissance, il s'agit de la première étude empirique de ce sujet en serbe.

Cette étude est basée sur des critères syntaxiques liés à la structure du GA : présence ou absence de dépendants de l'adjectif, leur nature et position par rapport à l'adjectif. Nous examinons également les effets de deux principes syntaxiques globaux qui se sont montrés utiles à l'étude de la place de l'adjectif dans les langues romanes : le poids syntaxique et la minimisation de la longueur des dépendances (*dependency length minimization*, dorénavant DLM) (cf. Abeillé et Godard, 1999 ; Thuilier, 2012 ; Gulordava et Merlo, 2015 ; Gulordava *et al.*, 2015). Le premier correspond à la tendance des dépendants syntaxiquement complexes (lourds) à se trouver en postposition, alors que le deuxième exprime la tendance observée des langues à disposer les dépendants autour de leur gouverneur de sorte à minimiser la longueur des dépendances. Notre analyse est basée sur le treebank ParCoTrain-Synt (Miletic, 2018). Ce corpus de 101 000 tokens dispose d'annotations syntaxiques et morphosyntaxiques. En nous basant sur ces informations, nous avons effectué une extraction automatique des GA, suivie d'une analyse automatique de leur structure et position. Quant aux effets du poids syntaxique et de la DLM, ils ont été examinés manuellement.

Nous avons identifié 9666 occurrences du GA, dont 555 (5,74 %) postposés au nom. Différents patrons de comportement sont observés en fonction de la structure du GA. Les GA minimaux (constitués d'un adjectif sans dépendant) sont majoritairement antéposés (89,6 %), alors que les GA dotés d'un dépendant sous forme d'un nom fléchi ou d'un GP sont très majoritairement postposés (respectivement 92,44 % et 90,81 %). Cette distribution est en accord avec le principe du poids syntaxique et elle explique la différence entre les deux exemples donnés ci-dessus (*puna kuća* lit. 'pleine maison' vs *kuća puna dece* lit. 'maison pleine enfants.GEN'). Néanmoins, un GA lourd peut également se trouver en antéposition si le dépendant adjectival est antéposé à son tour (cf. *dece puna kuća* lit. 'enfants.GEN pleine maison'). Ce patron de linéarisation coïncide avec le principe de la DLM en permettant de minimiser la distance entre l'adjectif et son gouverneur. Il va également dans le sens des observations de Sadler & Arnold (1994), qui notent qu'en anglais le GA doit être à tête finale pour être antéposé. En revanche, les GA dotés d'un dépendant adverbial exhibent un comportement différent : ils préfèrent l'antéposition (55,23 %), et l'adverbe reste antéposé à l'adjectif quelle que soit la position de celui-ci par rapport au nom (cf. *sasvim puna kuća* lit. 'complètement pleine maison' vs *kuća sasvim puna* lit. 'maison complètement pleine'). Le fait qu'il s'agit d'un groupe à tête finale semble donc plus décisif que le poids syntaxique pour l'antéposition de ce type de GA. Quant à la DLM, elle ne semble pas affecter son positionnement. Le test χ^2 confirme que la position du GA et le type de dépendant sont statistiquement associés.

Les effets du poids syntaxique et de la DLM sont donc moins généralisés en serbe que dans les langues romanes : si le principe du poids permet d'articuler les tendances majoritaires, il n'explique pas les variations, et les effets de la DLM paraissent sporadiques. Une évaluation systématique de ces deux principes à travers la construction d'un modèle statistique prédictif permettra de mieux évaluer la présence et l'intérêt de ces phénomènes en serbe. Une autre piste possible consiste à confronter nos données à la *Dependency Locality Theory* (Gibson, 2000) dans le but d'évaluer sa capacité à expliquer les comportements observés. Néanmoins, ce travail nous a permis d'établir une première quantification de la position du GA en serbe et d'aboutir ainsi à une description plus complète du comportement du GA dans cette langue.

Références

- Abeillé, A., & Godard, D. (1999). La position de l'adjectif épithète en français: le poids des mots. *Recherches linguistiques de Vincennes*, (28), 9-32.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, 95-126.
- Gulordava, K., & Merlo, P. (2015). Structural and lexical factors in adjective placement in complex noun phrases across Romance languages. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning* (pp. 247-257).
- Gulordava, K., Merlo, P., & Crabbé, B. (2015). Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (Vol. 2, pp. 477-482).
- Miletic, A. (2018). *Un treebank pour le serbe : constitution et exploitations*. (Thèse de doctorat, Université Toulouse Jean Jaurès, Toulouse).
- Mrazović, P. (2009). *Gramatika srpskog jezika za strance*. Belgrade, Serbia: Izdavačka knjižarnica Zorana Stojanovića.
- Sadler, L., & Arnold, D. J. (1994). Prenominal adjectives and the phrasal/lexical distinction. *Journal of linguistics*, 30(1), 187-226.
- Stanojčić, Ž., & Popović, Lj. (2012). *Gramatika srpskog jezika*. Belgrade, Serbia: Zavod za udžbenike.
- Thuillier, J. (2012). *Contraintes préférentielles et ordre des mots en français* (Thèse de doctorat, Université Paris-Diderot-Paris VII, Paris).