

Ellipse et catégorisation des unités lexicales

François Trouilleux (UBP, LRL)

Le concept de catégorie syntaxique se situe à l'interface entre deux types de données : des données du lexique, c'est-à-dire des étiquettes qu'on associe à des unités lexicales hors énoncés, et des données de la syntaxe, c'est-à-dire des étiquettes qu'on associe à des nœuds dans un arbre syntaxique, donc à un énoncé.

L'architecture classique des outils de TAL illustre bien cette caractéristique. On définit d'une part des dictionnaires sous la forme de listes de mots ou de lemmes auxquels on associe des catégories (A, N, DET, PRO, etc.). On le fait en ayant d'autre part en tête une certaine idée des arbres syntaxiques qu'on doit obtenir, mais en général sans avoir une grammaire formalisée qui permette d'obtenir effectivement ces arbres. Dans la perspective d'une définition d'une grammaire des groupes nominaux et prépositionnels noyau avec le logiciel NooJ, l'exposé tente de montrer comment des choix faits au niveau de la description syntaxique (la grammaire) peuvent rendre inutile l'assignation de certaines catégories dans le lexique.

On s'intéresse en particulier à l'assignation d'une lecture adverbiale pour certaines prépositions et aux catégories que certains auteurs assignent aux numéraux cardinaux. Dans les deux cas, la réduction des ambiguïtés catégorielles fait suite à un traitement syntaxique mettant en [jeu](#) une sorte d'ellipse.