

Système d'analyse catégorielle ACCG : adéquation au traitement de problèmes syntaxiques complexes

Juyeon Kang, Jean-Pierre Desclés

LaLIC, 28, Rue Serpente, 75006 Paris, France

kjuyeon79@yahoo.fr, jean-pierre.desclés@paris-sorbonne.fr

Le système catégoriel, ACCG (*Applicative Combinatory Cateogrial Grammar*), est conçu pour objectif de produire des calculs catégoriels et d'engendrer des structures opérateur/opérande des phrases du français et du coréen. Il s'agit d'un système motivé pour l'analyse de bonne qualité linguistique pouvant traiter de phénomènes langagiers précis, particulièrement, le système casuel, le double cas, la flexibilité de l'ordre des mots dans la langue coréenne, et la coordination, la subordination ainsi que la thématization dans les langues française et coréenne. Pour cela, nous nous sommes basés sur la Grammaire Catégorielle Combinatoire Applicative (GCCA) (Desclés et Biskri, 2005 ; Kang et Desclés, 2008) qui est sous-jacente à un modèle mettant en œuvre le calcul fonctionnel des types de Church, les combinateurs de la Logique Combinatoire (LC) (Curry et Feys, 1958) et des méta-règles (deux types de méta-règles : a/ méta-règles pour contrôler l'intervention des combinateurs ; b/ méta-règles « contextuelles » pour lever l'ambiguïté dans l'assignation des types à certaines unités polysémiques). Ces méta-règles contextuelles sont des sortes de règles contextuelles au sens de J.-P. Desclés (Desclés, 2006).

Ce système ACCG (cf. Figure 1) est composé d'une interface, d'un lemmatiseur, de deux dictionnaires des mots typés et d'un analyseur catégoriel : 1) le lemmatiseur que nous avons développé permet essentiellement de segmenter des phrases en entrée en morphèmes ou en mots, ce qui est une étape pour une analyse morpho-syntaxique des langues agglutinantes comme le coréen ; 2) les deux dictionnaires préalablement définis contiennent environ milles unités linguistiques typées pour le coréen et le français¹ ; 3) l'analyseur catégoriel implémenté en OCaml.



Figure 1. Interface de l'analyseur ACCG

Le module « Demo » de l'ACCG propose des analyses syntaxiques de phénomènes précis tels que le double cas, l'ordre des mots, la coordination, la subordination et la topicalisation.

Le module « Parser » est une composante essentielle de l'ACCG en effectuant des analyses syntaxiques. Le lemmatiseur et les dictionnaires prédéfinis interviennent dans ce module. Le processus d'analyse est le suivant : 1) choisir une langue spécifique ; 2) entrer une phrase à analyser ; 3) cliquer sur le bouton '1. Lemmatiser' si on veut vérifier le résultat de lemmatisation ; 4) cliquer sur le bouton '2. French Parser' ou '3. Korean Parser' pour l'analyse catégorielle de la phrase segmentée.

BISKRI I., DESCLÉS J.-P. (2005). Applicative and Combinatory Categorical Grammar and Subordinate Constructions in French, *International Journal on Artificial Intelligence Tools*, Vol.14, N°1&2, 125-136.

DESCLÉS J.-P. (2006). Contextual Exploration Processing for Discourse Automatic Annotations of Texts. *FLAIRS-19 (International Florida Artificial Intelligence Research Society Conference)*, 281-284.

KANG J.Y., DESCLÉS J.-P. (2008). Korean Parsing based on the Applicative Combinatory Categorical Grammar. *The 22nd Pacific Asia Conference on Language, Information and Computation*, 215-224.

¹ Ces deux dictionnaires sont chacun construits pour le français et le coréen. Ils sont accompagnés de dictionnaires restreints pour l'anglais et le japonais : le dictionnaire pour le français contient une centaine d'entrées pour l'anglais et celui pour le coréen une centaine d'entrées pour le japonais, de façon à tester la pertinence des analyses linguistiques de problèmes syntaxiques particulièrement complexes.