
Aide à la lecture en L2 : quel type d'analyse syntaxique automatique?

Neda Lestarevic^{*1}, François Stuck^{†1}, and Mathieu Valette^{‡1}

¹Equipe de Recherche Textes, Informatique, Multilinguisme (ERTIM) – INALCO – 2 Rue de Lille
75007 Paris, France

Résumé

Notre étude vise la conception et l'évaluation d'un dispositif en ligne d'aide à la lecture en L2, favorisant chez l'apprenant une pratique précoce et autonome de la lecture. La lecture est enrichie d'informations métalinguistiques générées automatiquement, destinées à aider la compréhension du texte et l'acquisition de compétences langagières. L'application intègre divers outils de Traitement Automatique des Langues, notamment un segmenteur en phrases et mots, un étiqueteur morphosyntaxique, un analyseur syntaxique. Dans l'étude présentée ici, nous cherchons à comparer l'apport didactique de deux analyseurs syntaxiques du français : TALISMANE, qui suit la norme d'annotation du French Treebank adaptée à la syntaxe du français et le Stanford Parser intégrant les dépendances universelles – une forme d'annotation universelle applicable à des langues éventuellement distantes syntaxiquement. L'enjeu consiste à découvrir quel type d'annotation est meilleur du point de vue de l'apprenant de FLE : celui proche de la langue cible ou celui, plus universel, se rapprochant potentiellement de sa langue maternelle ? Nous avons décidé d'évaluer cet apport didactique avec des apprenants locuteurs d'une langue syntaxiquement éloignée du français : une langue slave, le serbe, dont l'ordre des mots est libre et les cas expriment les fonctions syntaxiques. Nous avons choisi un groupe de 15 apprenants de FLE (niveaux A1-C2 du CERCL) de langue maternelle serbe. Chaque apprenant lira, via le dispositif, un jeu de textes adaptés à son niveau : textes sans aucune analyse syntaxique, juste annotés en parties du discours; textes analysés par Talismane; textes analysés par Stanford Parser. Il répondra ensuite à un questionnaire portant sur sa compréhension des textes et ses préférences vis-à-vis des trois types d'annotation proposée. Les résultats devraient permettre d'évaluer : l'utilité d'un analyseur syntaxique intégré à notre dispositif d'aide à la lecture ; le type d'analyse le mieux adapté pour des apprenants slaves de FLE.

Bibliographie

Antoniadis G. (2010). " De l'apport pertinent du TAL pour les systèmes d'ALAO : L'exemple du projet MIRTO ". *Actes du CMLF 2010*, Congrès Mondial de Linguistique Française.

Crinon J., Legros D., Marin B. et Avel P. (2007). Aides logicielles à la lecture de textes documentaires scientifiques. Alsics 10.1.

Lundquist L. (2008). *Navigating in Foreign Language Texts*, éditions Samfundslitteratur

*Intervenant

†Auteur correspondant: francois.stuck@inalco.fr

‡Auteur correspondant: mvalette@inalco.fr

(Danemark).

Nivre J., Kübler S. (2006). " Dependency Parsing ". Tutorial at *COLING-ACL*, Sydney 2006 <http://cl.indiana.edu/~md7/08/545/slides/dep-parsing.pdf>

Goodman, K.S. (1973). " Psycholinguistic universals of the reading process ". In F. Smith (Ed.), *Psycholinguistics and reading*, pp. 21-29. NY: Holt, Rinehart and Winston.

Urieli Assaf (2013). *Robust French syntax analysis: reconciling statistical methods and linguistic knowledge in the Talismane toolkit*. PhD thesis. Université de Toulouse II-Le Mirail.

Note biographique

Les auteurs sont membres de l'ERTIM de l'Inalco, une équipe au carrefour de l'ingénierie multilingue et du Traitement Automatique des Langues. L'ERTIM a pour objets principaux la recherche en sémantique des textes pour les applications ; le développement de méthodologie pour l'ingénierie des textes et des documents numériques multilingues et la production de ressources multilingues (lexicales, terminologiques, textuelles, didactiques).

Mots-Clés: aide à la lecture, langue seconde, FLE, traitement automatique des langues, analyse syntaxique