

Détection et correction d'erreurs grammaticales

Jean-Philippe Prost et Mathieu Roche

9 novembre 2012

Résumé

L'objet du stage est de développer une application de correction grammaticale, qui réponde aux critères de participation au challenge international HOO.

le challenge international *Help Our Own*¹ (HOO) vise à fournir un cadre d'évaluation pour des systèmes de correction grammaticale (ou *assistance à la rédaction*), sur des tâches de détection et de correction très spécifiques. En 2012, l'objectif fixé était de détecter et corriger des erreurs de pronoms et de prépositions en anglais, réparties en six catégories d'erreurs différentes. Ces catégories incluent, par exemple :

- le remplacement de préposition (ex. : *When I arrived at London*)
- les prépositions absentes (ex. : *I gave it John*)
- les prépositions inutiles (ex. : *I told to John that ...*)
- etc.

L'objectif scientifique est de proposer, tester, comparer et évaluer différentes stratégies de détection et de correction pour la même couverture de phénomènes linguistiques que celle de l'édition HOO 2012. Ces stratégies seront basées sur l'exploitation des données statistiques fournies par un corpus de n-grammes — en l'occurrence le corpus de Google, établi sur la littérature recensée dans Google Books.

Ces stratégies devront, bien sûr, faire l'objet d'une implantation logicielle.

Parmi les questions à aborder, on pourra trouver les suivantes :

- Peut-on généraliser les n-grammes, en intégrant, par exemple, des connaissances sémantiques, et/ou syntaxiques ?
- Comment intégrer des mesures d'association entre mots (*dice, information mutuelle, ...*) ?
- Quelle(s) taille(s) de n-grammes choisir, et quelles stratégies de choix associées ? Peut-on construire des mesures qui rendent compte de ces différentes stratégies ?

Références

- [1] Jill Burstein, Martin Chodorow, and Claudia Leacock. Automated essay evaluation : The criterion online writing service. *AI Magazine*, 25 :27–36, 2004.
- [2] Robert Dale, Ilya Anisimoff, and George Narroway. Hoo 2012 : A report on the preposition and determiner error correction shared task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 54–62, Montréal, Canada, June 2012. Association for Computational Linguistics.
- [3] Rogelio Nazar and Irene Renau. Google books n-gram corpus used as a grammar checker. In *Proceedings of the Second Workshop on Computational Linguistics and Writing (CL&W 2012) : Linguistic and Cognitive Aspects of Document Creation and Document Engineering*, pages 27–34, Avignon, France, April 2012. Association for Computational Linguistics.

1. <http://clt.mq.edu.au/research/projects/hoo/>