

iMAG : post-édition, évaluation de qualité de TA et production d'un corpus parallèle

Lingxiao WANG Ying ZHANG

GETALP – LIG, 41 rue des Mathématiques, BP 53, 38041 Grenoble Cedex 9

Lingxiao.Wang@imag.fr, Ying.Zhang@imag.fr

RÉSUMÉ

Une passerelle interactive d'accès multilingue (iMAG) dédiée à un site Web S (iMAG-S) est un bon outil pour rendre S accessible dans beaucoup de langues, immédiatement et sans responsabilité éditoriale. Les visiteurs de S ainsi que des post-éditeurs et des modérateurs payés ou non contribuent à l'amélioration continue et incrémentale des segments textuels les plus importants, et éventuellement de tous. Dans cette approche, les pré-traductions sont produites par un ou plusieurs systèmes de Traduction Automatique (TA) gratuits. Il y a deux effets de bord intéressants, obtenables sans coût additionnel : les iMAGs peuvent être utilisées pour produire des corpus parallèles de haute qualité, et pour mettre en place une évaluation permanente et finalisée de multiples systèmes de TA.

ABSTRACT

iMAG : MT-postediting, translation quality evaluation and parallel corpus production

An interactive Multilingual Access Gateway (iMAG) dedicated to a web site S (iMAG-S) is a good tool to make S accessible in many languages immediately and without editorial responsibility. Visitors of S as well as paid or unpaid post-editors and moderators contribute to the continuous and incremental improvement of the most important textual segments, and eventually of all. In this approach, pre-translations are produced by one or more free machine translation systems. There are two interesting side effects obtainable without any added cost: iMAGs can be used to produce high-quality parallel corpora and to set up a permanent task-based evaluation of multiple MT systems.

MOTS-CLÉS : post-édition, évaluation de systèmes de TA, production d'un corpus parallèle

KEYWORDS : post-edition, evaluation of MT systems, production of parallel corpora.

Nous proposons 3 démonstrations : (1) l'accès multilingue à un site Web, avec la post-édition de résultats de TA "à la Google"; (2) la post-édition en mode avancé; (3) la production d'un corpus parallèle.

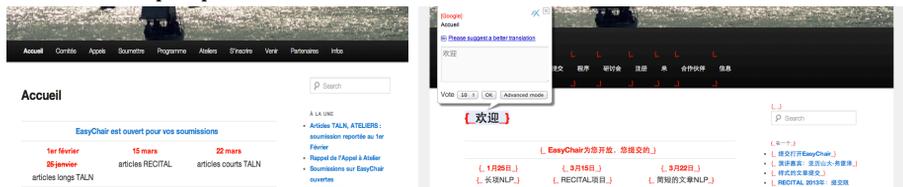


FIGURE 1 – Page originale en anglais et page accédée en chinois.

Voici un exemple d'accès au site Web de TALN 2013 en chinois. L'original est en anglais

