

Lecture bilingue augmentée par des alignements multi-niveaux

François Yvon¹ Yong Xu¹ Marianna Apidianaki¹

Clément Pillias² Pierre Cubaud²

(1) LIMSI, CNRS, Université Paris-Saclay, 91 403 Orsay, France

(2) CEDRIC, CNAM, 292 rue St Martin, 75003 PARIS, France

{yvon,yong,marianna}@limsi.fr, {clement.pillias,cubaud}@cnam.fr

RÉSUMÉ

Le travail qui a conduit à cette démonstration combine des outils de traitement des langues multilingues, en particulier l'alignement automatique, avec des techniques de visualisation et d'interaction. Il vise à proposer des pistes pour le développement d'outils permettant de lire *simultanément* les différentes versions d'un texte disponible en plusieurs langues, avec des applications en lecture de loisir ou en lecture professionnelle.

ABSTRACT

Augmenting bilingual reading with alignment information

We present in this work the result of the TRANSREAD project, aimed at exploring methods and tools for designing an augmented bilingual reading experience. Our proposal relies on bitext alignment models used for machine translation, as well as interaction and visualization techniques.

MOTS-CLÉS : Lecture électronique, Alignement automatique de bitextes.

KEYWORDS: Electronic Reading, Bitext Alignment.

1 Contexte

La lecture électronique est aujourd'hui une pratique bien établie pour les livres monolingues. Dans le cadre du projet TRANSREAD, nous avons conduit une réflexion sur les dispositifs de lecture électronique pour les textes existant en deux langues, en particulier les *livres bilingues*. Deux cadres applicatifs ont été initialement étudiés : la lecture sur terminal fixe, dans une optique de contrôle ou de validation de traductions humaines dans un contexte professionnel, d'une part ; la lecture sur terminal mobile pour une lecture de loisir, d'autre part. Ce second cadre applicatif vise une population de lecteurs désireuse de profiter des textes dans leur version originale, tout en ayant la possibilité de faire des allers-retours avec une langue plus familière, pour élucider des fragments difficiles ; l'utilisation de tels dispositifs dans un cadre d'apprentissage a également été évoquée.

Du point de vue technique, nous avons étudié comment la lecture bilingue pourrait être augmentée par l'exploitation d'alignements à différents niveaux de granularité entre les deux parties d'un bitexte. Les méthodes d'alignement de bitextes (Tiedemann, 2011) ont connu un fort développement ces années récentes, avec des applications notamment en traduction automatique. La plupart des outils existants de visualisation d'alignements ont été conçus dans un cadre de traduction automatique pour vérifier des annotations manuelles et offrent des possibilités de visualisation et d'interaction rudimentaires, très loin de ce qui est attendu pour une utilisation dans un cadre de lecture.

2 Réalisations

Les principales réalisations¹ de ce projet sont de trois ordres. Concernant les alignements automatiques, notre travail a principalement porté sur l'amélioration des techniques d'alignement de phrases, dont nous avons mis en évidence les faiblesses lorsqu'il s'agit de traiter des traductions d'œuvres littéraires, qui sont souvent moins littérales que les traductions techniques (Xu *et al.*, 2015); nous avons également proposé de nouvelles méthodes pour construire des ressources annotées pour l'alignement sous-phrastique, qui pose des difficultés encore plus grandes (Xu & Yvon, 2016).

Une seconde contribution de ce travail a été l'élaboration d'un format pour représenter les documents bilingues et leurs alignements : le défi, étant, par rapport aux propositions existantes, de représenter les documents électroniques en incluant toutes les informations typographiques qui sont indispensables pour un rendu graphique de qualité du texte. Notre proposition s'appuie sur le standard e-pub et repose sur une représentation explicite des alignements multi-niveaux dans un fichier XML qui permet de faire la correspondance entre les textes cible et source, tous les deux disponibles au format HTML.

Enfin, du point de vue de l'Interaction Humain-Machine, ce projet a permis de réaliser une étude fine de l'espace de design de la lecture bilingue (Pillias & Cubaud, 2015). La mise en œuvre qui fera l'objet de la démonstration permet de visualiser à la demande des alignements à plusieurs niveaux de granularité dans une représentation où les côtés source et cible sont en vis-à-vis. Dans cette visualisation, l'alignement est centré sur le mot situé au-dessus du doigt pour éviter l'occlusion. Il est également possible de faire apparaître/disparaître l'un des deux côtés du bitexte pour poursuivre la lecture dans une seule langue. Des expériences de lecture avec des sujets humains ont été réalisées.

Ce travail ouvre de nombreuses perspectives, tant dans l'optique d'augmenter la lecture bilingue avec d'autres informations (par exemple des informations de désambiguïsation), que dans la perspective de concevoir d'autres modes de visualisation et d'interaction avec des bitextes alignés.

Remerciements

Ce travail a été partiellement financé par l'ANR dans le cadre du projet ANR-12-CORD-0015. Nous remercions N. Kubler et M. Zimina (Univ. Paris Diderot) pour leurs contributions à ce projet.

Références

PILLIAS C. & CUBAUD P. (2015). Bilingual Reading Experiences : What They Could Be and How To Design for Them. In *Proc. IFIP Interact.*

TIEDEMANN J. (2011). *Bitext Alignment*. Number 14 in Synthesis Lectures on Human Language Technologies. Morgan and Claypool publishers.

XU Y., MAX A. & YVON F. (2015). Sentence alignment for literary texts. *Linguistic Issues in Language Technology*, **12**(6).

XU Y. & YVON F. (2016). Novel elicitation and annotation schemes for sentential and sub-sentential alignments of bitexts. In *Proc. LREC 2016*, Portorož, Slovenia.

1. Les publications, ressources et démonstrations produites par le projet sont sur le site : transread.limsi.fr.