

Les outils libres du traducteur

Didier Briel

didier@didierbriel.fr

Rencontres informelles SFT

Lyon

mars 2011

Sommaire

- Concepts
- Écosystème des outils libres
- Rôle des différents logiciels
- OmegaT
- Alignement
- Logiciels libres et investissement



Concepts

Logiciels libres

Traduction Assistée par Ordinateur

Concepts Logiciels libres

- Un logiciel libre n'est pas un *freeware*
- Un logiciel libre est caractérisé par quatre libertés fondamentales
 - Pouvoir exécuter le programme, pour n'importe quel usage
 - Pouvoir étudier le fonctionnement du programme, et l'adapter à ses besoins
 - Pouvoir redistribuer des copies du logiciel (y compris de façon payante)
 - Pouvoir distribuer à d'autres des copies des versions modifiées

Concepts

Traduction Assistée par Ordinateur

- Parfois appelée « Mémoire de traduction »
- Il s'agit d'une aide pour le traducteur, pas d'une traduction automatique (*Machine Translation*)
- L'ordinateur permet de réutiliser des phrases ou des segments de phrases déjà traduits
- De façon étendue, la TAO est l'ensemble des aides apportées au traducteur par le ou les logiciels (glossaires, dictionnaires, etc.)

Rôle des différents logiciels

OmegaT
Okapi
bitext2tmx
Bligner
LF Aligner

Rôle des différents logiciels

- OmegaT
 - Traduction assistée par ordinateur
 - Traduction
 - Mémoires de traduction
 - Glossaires
- Okapi
 - Rainbow
 - Conversions de fichiers
 - Utilitaires
 - Checkmate
 - Assurance qualité
 - Olifant
 - Éditeur de mémoires de traduction
- bitext2tmx, bligner, LF Aligner
 - Alignement

Écosystème des outils libres

Multiplateforme

Java ou interpréteur

Conception modulaire

Standards et transparence des données

Écosystème des outils libres

Multiplateforme

- Les logiciels libres les plus connus
 - Firefox (navigateur)
 - OpenOffice.org/LibreOffice (suites bureautiques)
- fonctionnent sur les trois systèmes principaux
 - Windows
 - Linux (système d'exploitation libre)
 - Mac OS X
- Être multiplateforme
 - Alimente la logithèque des systèmes « pauvres »
 - Augmente la part de marché lorsque le logiciel n'est pas disponible initialement sous Windows
- Méthode pour être multiplateforme
 - Utiliser des méthodes « standard » quand cela est possible
 - Adapter une partie du logiciel pour chaque système
 - Compiler (produire un exécutable) pour chaque système

Écosystème des outils libres

Java ou interpréteur

- Java est une machine virtuelle
 - OmegaT
 - Okapi
 - bitext2tmx
- Un interpréteur isole l'exécution du système réel
 - BASIC
 - Perl
 - bligner
 - LF Aligner
 - Python
 - bligner
- Ils permettent d'être multiplateforme sans aucun développement

Écosystème des outils libres

Conception modulaire

- Pas d'outils tout en un
- Pourquoi créer un assistant lorsque le système d'exploitation hôte fournit les outils ?
- Pas d'impératif commercial
 - Pas besoin de fournir « autant de fonctions que le concurrent »
 - Si un autre logiciel libre fournit la même fonction, pas besoin de redévelopper
- Un programme s'appuie souvent en partie sur des bibliothèques
 - OmegaT
 - Vérificateur orthographique Hunspell
 - Firefox
 - OpenOffice.org/LibreOffice
 - Vérificateur linguistique LangageTool
 - OpenOffice.org/LibreOffice
 - Lemmatiseurs de Lucene

Écosystème des outils libres

Conception modulaire

- Relations symbiotiques
 - Okapi dispose de fonctions dédiées à OmegaT
 - OmegaT a été modifié pour mieux collaborer avec Okapi
- Réutilisation d'autres logiciels
 - LF Aligner
 - Pratiquement aucune tâche propre
 - Agrégation de sept logiciels différents, y compris un traitement de texte
 - Boltran
 - OmegaT « anywhere » (version web d'OmegaT)
 - Autshumato ITE
 - Dédié aux onze langues d'Afrique du Sud
 - Utilise OmegaT et OpenOffice.org
 - La relation n'est pas à sens unique
 - Développements de Boltran et Autshumato ITE dans OmegaT

Écosystème des outils libres

Standards et transparence

- Avec un code source disponible, le « secret de fabrication » n'a aucun intérêt
- Utilisation de standards
 - Réduction des coûts
 - Augmentation de l'interopérabilité
 - On peut commencer une tâche dans un logiciel, et la finir dans un autre
 - Données transparentes
 - Peuvent généralement être examinées et modifiées avec un simple éditeur de texte
- Standards utilisées par les outils libres de TAO
 - TMX, mémoires de traduction
 - Fichiers textes (CSV et TSV), glossaires
 - XLIFF, échange de fichiers multilingues
 - TBX, base terminologique
 - SRX, description des règles de segmentation
 - ITS, description des éléments à traduire



OmegaT

Caractéristiques principales

Fonctions principales

Modules d'extension

Échange avec les autres logiciels de TAO

OmegaT

Caractéristiques principales

- OmegaT est un logiciel totalement autonome
 - Aucune de ses fonctions ne dépend de l'installation d'autres logiciels (ex. : Microsoft Office)
- Son interface graphique dépend du système d'exploitation
- *Segments* source et cible affichés en ligne (comme Trados TagEditor), plutôt que côte à côte (comme Trados Studio, Wordfast Pro et MemoQ)

OmegaT

Caractéristiques principales

- Pas de format intermédiaire
 - Pas de préparation
 - Importation ou conversion
 - Pas de « nettoyage »
 - Modification dynamique des projets instantanée (ajout/modification/suppression de documents)
 - Très peu d'impact sur le formatage
- Pas de base de données
 - Toutes les données sont traitées en mémoire
 - Très grande rapidité
 - Limite à la taille des données

OmegaT

Fonctions principales

- Correspondances partielles (*fuzzy matching*)
- Propagation automatique des traductions
- Glossaires (vérifications terminologiques)
- Recherche de termes dans le projet, dans les mémoires de référence et dans des documents de référence
- Projets comportant un nombre quelconque de dossiers et de fichiers, dans tous les formats acceptés
- Gestion des écritures de droite à gauche et mixtes
- Validation des balises (vérification du formatage)
- Vérification orthographique (Hunspell ou JMySpell)
- Dictionnaires (mono et multilingue) au format StarDict et Lingvo DSL
- Traduction automatique (Google Translate, Apertium, Belazar)

OmegaT

Modules d'extension (*plugins*)

- Distribués séparément, principalement pour des raisons juridiques (compatibilité des licences)
 - OmegaT : GPL v2
 - Modules d'extension : GPL v3 et Apache
- Vérificateur linguistique (LanguageTool)
 - Règles de base applicables à la plupart des langues
 - Règles spécifiques dans 20 langues
- Lemmatiseurs
 - Permettent de calculer les racines des mots (*lemmes*)
 - Correspondances plus nombreuses et plus intelligentes, à la fois pour les correspondances et les glossaires
 - Calcul des mots pour les langues asiatiques (CJK)
 - Éliminent certains mots courants (*mots non significatifs*) des correspondances, améliorant ainsi leur pertinence
 - 24 langues disponibles

OmegaT

Modules d'extension (*plugins*)

- Formats de fichier supplémentaires (Okapi)
 - InDesign IDML
 - ...
 - Trados TagEditor TTX
- Lancement de scripts dans OmegaT
 - Permet de lancer des traitements sur les données en mémoire dans OmegaT
 - Permet de créer des interfaces graphiques et des fonctions interactives
 - Nombreux langages pris en charge
 - Groovy
 - ...
 - JavaScript

OmegaT

Échange avec les autres logiciels de TAO

- Mémoires de traduction
 - TMX version 1.4b niveaux 1 et 2
 - Lecture
 - Création
 - Informations textuelles préservées, ainsi que les repères de formatage
 - Nombre de mémoires de traduction illimité
- Glossaires
 - Fichiers textes
 - Colonnes séparées par des tabulations ou des virgules (CSV)
 - Colonnes Source, Cible et Commentaire
 - Lecture
 - Création
 - TBX
 - Lecture
 - Nombre de glossaires illimité
- Documents bilingues
 - PO
 - XLIFF
 - TTX
 - TXML



Alignement

Stratégies d'alignement

bitext2tmx

bligner

LF Aligner

Alignement

Stratégies d'alignement

- L'alignement consiste à récupérer des traductions existantes à partir de documents monolingues déjà traduits
- Alignement visuel
 - Le logiciel effectue un premier alignement
 - L'ajustement s'effectue ensuite manuellement
 - Avantage
 - On peut obtenir un alignement dont on est sûr
 - Inconvénient
 - L'alignement automatique initial est en général sommaire
 - Demande énormément de temps
- Alignement automatique
 - Le logiciel effectue l'alignement sans intervention de l'utilisateur
 - Avantage
 - On peut traiter de grands volumes de données
 - La qualité de l'alignement est souvent « suffisante »
 - Inconvénient
 - L'alignement automatique n'est pas parfait

Alignement bitext2tmx

- Alignement visuel
- Programme Java
- Fonctionne avec des fichiers texte
- Prise en main simple
- Règles de segmentation simplifiées et non modifiables

Alignement bligner

- Alignement automatique
- Programme Perl ou Python
- Fonctionne avec des fichiers texte
- Application spartiate
 - Modification des options directement dans le fichier source
- Règles de segmentation de type SRX
- Mode paragraphe
- Mode « interactif »
 - Choix des fichiers
 - Choix des langues
- Mode ligne de commande
- Algorithme simple pour gérer un nombre différent de phrases au sein d'un paragraphe

Alignement

LF Aligner

- Alignement automatique
- Programme principal Perl
 - Exécutable et autres composantes livrés sous forme binaire sous Windows
- Fonctionne avec de nombreux formats
 - Texte, HTML (y compris téléchargement dynamique), .doc, .docx, .rtf, .pdf
 - Alignement direct de la législation européenne
 - Conversion interne en fichiers texte
- Algorithme d'alignement sophistiqué d'Hunalign
 - Dictionnaires pour 32 langues
 - Longueur des phrases
- Alignement semi-visuel
 - Corrections dans Excel

Logiciels libres et investissement

Coût d'utilisation

Développement

Assistance

Logiciels libres et investissement

Coût d'utilisation

- Logiciel propriétaire
 - L'achat initial d'un logiciel propriétaire n'est pas le seul coût
 - Coûts de maintenance ou de mise à jour
 - Coût de gestion des licences
 - Utilisation de la ou des licences sur plusieurs machines
 - Risque en cas de perte de disque dur
- Logiciel libre
 - Immédiatement disponible
 - Sur toutes les machines
 - En version « portable » (clé USB)
 - Sans installation (Java Web Start)
 - Investissement intellectuel différent
 - Compréhension plus profonde des processus
 - Pas de clickodromes et de « certification » à base de questionnaires à choix multiples

Logiciels libres et investissement

Développement

- Logiciels propriétaires
 - Il faut les prendre « tels quels »
 - Pouvoir obtenir une adaptation est rare
- Logiciels libres
 - Les sources étant disponibles, il est possible de faire effectuer des développements spécifiques
 - Coût parfois inférieur au prix d'une seule licence propriétaire
 - Développements pérennes et maintenus gratuitement par la communauté

Logiciels libres et investissement

Assistance

- Il ne s'agit pas de tout ou rien
 - Logiciels propriétaires
 - Assistance payante et professionnelle
 - Logiciels libres
 - Se débrouiller avec le manuel et la communauté
- Les logiciels libres peuvent aussi bénéficier d'une assistance

Logiciels libres et investissement

Assistance OmegaT

- Assistance bénévole
 - Développeurs sur Sourceforge
 - Soumissions d'incidents
 - Demandes d'améliorations
 - Autres utilisateurs sur le groupe Yahoo!
 - Assistance
 - Trucs et astuces
 - Discussions sur les fonctionnalités actuelles et futures
 - 1700 membres
 - 24 h/24 (présence dans la plupart des fuseaux horaires)
 - Langue principale anglais, mais assistance possible dans de nombreuses langues (français, russe, japonais, etc.)
 - Écrire dans sa langue maternelle est encouragé

Logiciels libres et investissement

Assistance

- Assistance professionnelle abordable
 - Trados coûte 845 € (hors remise)
 - SDL Trados Studio 2011 Freelance à 676 € (sans « 2007 »)
 - Ce prix permet d'acheter de nombreuses heures d'assistance
- Le code étant disponible, on peut choisir son prestataire librement
- Il est parfois possible d'utiliser des compétences internes
 - Salarié formé au dépannage et au développement OmegaT

Logiciels libres et investissement

Assistance

- Formation OmegaT en présentiel (SFT)
 - Lyon
 - 25 mai 2012