



Les outils libres du traducteur

Didier Briel
didier@didierbriel.fr
Matinales IDF SFT
Paris
juin 2017

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

Heartsome TMX Editor

LF Aligner

Rôle des différents logiciels

- OmegaT
 - Traduction assistée par ordinateur
 - Traduction
 - Mémoires de traduction
 - Glossaires
 - Alignement visuel
 - Initialement, création individuelle
 - Communauté de traducteurs
 - Développé activement
- Okapi
 - Rainbow
 - Conversions de fichiers
 - Utilitaires
 - Checkmate
 - Assurance qualité
 - Olifant
 - Éditeur de mémoires de traduction
 - Initialement, création professionnelle
 - Communauté d'informaticiens
 - Développé activement

Rôle des différents logiciels

- Heartsome TMX Editor
 - Éditeur de mémoires de traduction
 - Conversion de glossaires et mémoires de traduction
 - Ancien logiciel commercial
 - Aucune communauté
 - Aucun développement
- LF Aligner
 - Alignement automatique
 - Création individuelle d'un traducteur
 - Aucune communauté
 - Maintenu



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/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
 - Heartsome TMX Editor
- Un interpréteur isole l'exécution du système réel
 - BASIC
 - Perl
 - LF Aligner
- 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/LibreOffice
 - Vérificateur linguistique LangageTool
 - OpenOffice/LibreOffice
 - Lemmatiseurs de Lucene et Hunspell

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
 - Il s'agissait d'un OmegaT « anywhere » (version web d'OmegaT)
 - Autshumato ITE
 - Dédié aux onze langues officielles d'Afrique du Sud
 - Utilise OmegaT et OpenOffice
 - 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

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
- Système de recherche sur Internet
- Projets comportant un nombre quelconque de dossiers et de fichiers, dans tous les formats acceptés
- Projets en équipe avec un nombre illimité de traducteurs
- Gestion des écritures de droite à gauche et mixtes
- Protection et validation des balises (vérification du formatage)
- Reconnaissance morphologique
 - Lemmatisation (correspondances partielles, glossaires)
 - Mots non significatifs (correspondances partielles)
 - Calcul des mots (particulièrement pour les langues asiatiques)
 - Plus de 100 langues disponibles (36 Lucene, plus de 100 Hunspell)

OmegaT

Fonctions principales

- Vérification orthographique (Hunspell)
- Vérification linguistique (LanguageTool)
 - Règles de base applicables dans la plupart des langues
 - Règles spécifiques dans 29 langues
- Dictionnaires (mono et multilingue) au format StarDict et Lingvo DSL
- Saisie automatique
 - Entrées de glossaire
 - Abréviations
 - Saisie historique et prédictive
- Traduction automatique (Google Translate, Microsoft Translator, Apertium, Yandex, MyMemory)
- Exécution de scripts
 - Équivalent des macros Word ou Excel
- API publique pour modules d'extension
 - Formats de fichier supplémentaires (Okapi)
 - Traduction automatique locale (Apertium)
 - Support des dictionnaires japonais EPWING (Hiroshi Miura)
 - Plus de 20 modules disponibles

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
 - Connexion au site terminologique TaaS (Terminology as a Service)
 - 24 langues officielles européennes plus russe et turc
 - Collaboration avec utilisateurs memoQ et Trados Studio

OmegaT

Échange avec les autres logiciels de TAO

- Documents bilingues
 - PO (format de localisation libre)
 - XLIFF (format d'échange standard)
 - TTX (ancien format d'échange Trados)
 - TXML (format d'échange Wordfast Pro)



Alignement

Stratégies d'alignement

OmegaT 4

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 OmegaT 4

- Alignement visuel
- Fonctionne avec tous les formats reconnus par OmegaT
- Le formatage est préservé
- Choix d'algorithmes et de paramètres pour l'alignement initial
- L'alignement visuel permet de segmenter, de fusionner et de modifier les segments
- Statut visuel disponible (accepté, à réviser)

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
 - Le formatage est perdu
- Algorithme d'alignement sophistiqué d'Hunalign
 - Dictionnaires pour 32 langues
 - Longueur des phrases
- Alignement semi-visuel
 - Corrections dans un éditeur dédié ou 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
 - Le code source étant disponible, 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
 - 2577 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
 - SDL Trados Studio 2017 coûte 695 € (hors remise)
 - Ce prix permet d'acheter de nombreuses heures d'assistance
- Le code source 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

- Téléformation Initiation à OmegaT (SFT)
 - 12 octobre 2017