

État des lieux de TRADOH

Projet Traouiéro, document L3.3.a

Ce document est un rapport d'avancement partiel de la sous-tâche 3.3 (phase a) et est joint au rapport d'avancement pour les tâches 2, 3 et 4 (livrable L234.1).

Contenu

1. INTRODUCTION	2
1.1 Objet de la sous-tâche ST3.3	2
1.2 Objet de ce document.....	2
2. PHASE A : TRADOH_v1 VERS TRADOH_v2	2
2.1 Spécifications	2
2.2 Particularités et réutilisabilité	3
2.3 Dépôt	3
2.4 Intégration à SECTRa_w/iMAG_v2	3
3. PHASE B : TRADOH++	3
3.1 Objectifs	3
3.2 Services de traduction automatique.....	3
4. RÉFÉRENCES	4
5. ANNEXES	4
5.1 Appel de TRADOH dans le cas le plus simple d'un texte monolingue	4
5.2 Appel de TRADOH dans le cas « multilingue »	5
5.3 Appel de TRADOH dans le cas « multilingue & multitructeur »	6
5.4 Organisation de TRADOH	6
5.5 Projet TRADOH++ dans la forge du LIG.....	6

État des lieux de TRADOH

Achille FALAISE, Valérie BELYNCK, Christian BOITET

1. INTRODUCTION

1.1 OBJET DE LA SOUS-TÂCHE ST3.3

La sous-tâche ST3.3 participe à la tâche 3, dédiée à l'opérationnalisation du service d'accès multilingue à des sites Web par des iMAG¹ dédiées, permettant la post-édition contributive avec ou sans modération.

TRADOH++ sera un serveur paramétrable d'appel de services de traduction automatique, et constituera à ce titre un service essentiel pour l'opérationnalisation du logiciel iMAG. TRADOH++ sera basé sur le logiciel TRADOH, développé par VO-TRUNG Hung dans sa thèse et mis en service fin 2004 sur le site du GETA (devenu GETALP en 2007).

TRADOH est un service Web auquel on envoie un texte brut, a priori multilingue et multiformat, une langue cible, et une liste de 1 à 4 serveurs de TA accessibles par le Web. Il fait appel à SANDOH, autre logiciel du même auteur, pour segmenter le texte en zones homogènes du point de vue de la langue et du codage (ex : <japonais, UTF-8>, <japonais, EUC-jp>), ce qui donne un fichier XML d'où l'on peut recréer exactement le fichier de départ.

En fonction du graphe des couples de langues offerts par chacun des systèmes de TA demandés, TRADOH transforme chaque zone homogène dans un codage accepté par le système de TA, et appelle ce système, éventuellement 2 fois (en passant par l'anglais) s'il n'offre pas de traduction directe pour le couple <langue source de la zone, langue cible demandée>. Enfin, il affiche les résultats dans 1 à 4 fenêtres parallèles. Comme le service GoogleTranslate n'existait pas en 2004, l'instance déployée à l'époque appelait seulement les serveurs de Systran, Softissimo (Reverso), GistinTime et FreeTranslation, qui ne sont plus disponibles sous la même forme actuellement (les API et les conditions d'utilisation ont changé).

Il s'agit dans un premier temps (phase ST3.3.a) de reprendre le code de TRADOH, de le déposer sur une forge et à l'APP (Agence pour Protection des Programmes). Dans un second temps (phase ST3.3.b), il s'agira de l'étendre en un service encore plus paramétrable (dit TRADOH++), et également appelable par programme. Cette sous-tâche est ainsi décomposée en 2 phases :

phase ST3.3.a :	TRADOH : récupération du code, actualisation, et dépôt
phase ST3.3.b :	TRADOH++ : extension du service

1.2 OBJET DE CE DOCUMENT

Ce document est donc un rapport d'avancement partiel de la phase a de la sous-tâche 3.3. Il est joint au rapport d'avancement L234.1 des tâches 2 à 4 à T0+6 (15/7/11).

L'objectif principal de ce livrable est de préparer le dépôt des logiciels TRADOH et SANDOH actualisés, qui vont être mis en service dans la version 2 du logiciel SECTra_w/iMAG.

2. PHASE A : TRADOH_v1 VERS TRADOH_v2

2.1 SPÉCIFICATIONS

Les spécifications générales de TRADOH ont été données plus haut.

Ses spécifications détaillées sont contenues dans la thèse de son auteur, VO TRUNG Hung (2004), mais il manque les documents de génie logiciel nécessaires pour pouvoir parler

¹ Passrelle interactive d'accès multilingue (iMAG = interactive Multilingual Access Gateway).

d'opérationnalisation. Il faudra les créer pour la version « opérationnalisée » TRADOH_v2 (rétro-ingénierie), puis pour le futur TRADOH++.

2.2 PARTICULARITÉS ET RÉUTILISABILITÉ

TRADOH est un script CGI codé en Perl. Il peut être associé à un formulaire Web pour effectuer des requêtes (cf. annexe).

2.3 DÉPÔT

Le code de TRADOH a été déposé sur la nouvelle forge du laboratoire LIG, à l'adresse suivante : <https://forge.imag.fr/projects/tradoh/> (accès restreint aux membres de la forge inscrits sur le projet).

Le dépôt de ce code auprès de l'APP (Agence pour Protection des Programmes) est en cours.

2.4 INTÉGRATION À SECTRA_w/IMAG_v2

TRADOH_v2 sera intégré à la version 2 de SECTra_w/IMAG, pour remplacer les appels directs actuels bien plus limités.

3. PHASE B : TRADOH++

3.1 OBJECTIFS

Le logiciel TRADOH++ doit permettre l'appel d'autres services de traduction automatique. Il sera directement basé sur TRADOH_v2.

Cette phase comprend deux étapes : l'identification de services de traduction automatique appelables en ligne, et l'extension de TRADOH pour permettre l'appel de ces services. Actuellement, seule cette première étape est réalisée.

3.2 SERVICES DE TRADUCTION AUTOMATIQUE

Certains services sont actuellement disponibles, et seront bientôt utilisables via TRADOH_v2. Nous avons identifié les services suivants :

- Une adaptation de Systran 6, qui fonctionne sur l'un de nos serveurs et est actuellement appelée directement par SECTra_w/IMAG.
- Le service de traduction gratuit de Google, qui est actuellement appelé directement par SECTra_w/IMAG. Toutefois, ce service ne sera plus fourni par Google après le 31 décembre 2012.
- La version serveur de Systran 7.
- Le nouveau service (payant) de traduction en ligne de Systran (SystranLink).
- Le service de traduction automatique statistique développé dans l'équipe, basé sur le logiciel libre MOSES. On construira un système système pour chaque couple de langue et chaque mémoire de traductions en extrayant (grâce aux annotations de qualité présentes dans SECTra_w) une sous-mémoire représentant le meilleur compromis entre quantité et qualité.
- L'appel aux systèmes « experts » de traduction automatique écrits en Ariane-G5/Héloïse/Ariane-Y qui seront opérationnalisés dans le cadre de la tâche LING++ (ST4.1) du projet, entre autres pour le français↔russe et le français↔anglais.
- Le service de traduction automatique anglais↔malais SiSTeC-ebmt de l'USM (Universiti Sains Malaysia, Penang) et de la MMU (Multimedia University, Cyberjaya, Malaisie), avec lequel un effort d'intégration est en cours dans le cadre d'une coopération entre le projet Traouiero et le projet malaisien Brain Gain.

4. RÉFÉRENCES

- [1] **Vo-Trung H. (2004)** *Construction d'un outil pour analyser un document multilingue hétérogène en zones homogènes*. Proc. Research Informatics Vietnam & Francophone (RIVF), Hanoï, Vietnam, février, pp. 175-178, 4 p. <http://www-clips.imag.fr/~genthial/Publis/Hung-RIVF-04.pdf>
- [2] **Vo-Trung H. (2004)** *Système pour l'analyse des documents hétérogènes*. Proc. JADT'2004, Louvain la neuve, Belgique, mars, vol. 2/2, pp. 1177-1184, 8 p. <http://www-clips.imag.fr/~genthial/Publis/Hung-JADT-04.pdf>
- [3] **Vo-Trung H. (2004)** *Réutilisation de traducteurs gratuits pour développer des systèmes multilingues*. Proc. RECITAL-2004, Fès, Maroc, 19-22 avril, 8 p. <http://www-clips.imag.fr/~genthial/Publis/Hung-RECITAL-03.pdf>
- [4] **Vo-Trung H. (2004)** *Méthodes et outils pour utilisateurs, développeurs et traducteurs de logiciels en contexte multilingue*. Thèse de doctorat, INPG, 22/12/2004, 229 p.
- [5] **Vo-Trung H. & Boitet C. (2006)** *TRADOH, un méta-système de traduction automatique*. Proc. RIVF-06, Beyrouth, mai 2006, 12 p.

5. ANNEXES

5.1 APPEL DE TRADOH DANS LE CAS LE PLUS SIMPLE D'UN TEXTE MONOLINGUE

Cet écran correspond à l'appel le plus simple de TRADOH, sur un texte supposé monolingue. Dans ce cas, TRADOH n'appelle pas SANDOH (segmenteur et normaliseur des zones homogènes <langue, codage>), mais suppose que le texte d'entrée est dans la langue source indiquée par l'utilisateur, et en UTF-8.

On voit dans le panneau de gauche les 2 autres possibilités, « Texte multilingue » et « Multitraducteur ».

TRADUCTION AUTOMATIQUE MULTILINGUE

[English](#)
[Français](#)
[Tiếng Việt](#)

[Introduction](#)

[Texte monolingue](#)

[Texte multilingue](#)

[Multitraducteur](#)

TRADUCTION D'UN TEXTE MONOLINGUE

Choisissez la langue source

Choisissez la langue cible

Texte entrez un texte à traduire

Fichier choisissez un fichier disponible

pour traduire en langue souhaitée

Vous pouvez m'envoyer des courriers à Hung.Vo-Trung@imag.fr

Figure 1 : Formulaire Web de TRADOH pour un texte monolingue (anglais-français)

Figure 2 : Formulaire Web de TRADOH pour un texte monolingue (français-coréen)

5.2 APPEL DE TRADOH DANS LE CAS « MULTILINGUE »

Figure 3 : Formulaire Web de TRADOH pour le cas « multilingue »

Ici, le texte d'entrée est en français et en chinois (simplifié), et l'utilisateur voudrait le voir en japonais.

5.3 APPEL DE TRADOH DANS LE CAS « MULTILINGUE & MULTITRADUCTEUR »

TRADUCTION AUTOMATIQUE

Choisissez des traducteurs :

Systran
 Reverso
 Gist-in-Time
 FreeTranslation

Choisissez la langue source : Français

Choisissez la langue cible : Chinois Simplifié

Texte entrez un texte à traduire

TRAduction : Outils Unifiés, Intégrables, Embarquables, et Ressources Opérationnelles

Fichier choisissez un fichier disponible

pour traduire en langue souhaitée

Figure 4 : Formulaire Web de TRADOH pour le cas « monolingue & multitracteur »

5.4 ORGANISATION DE TRADOH

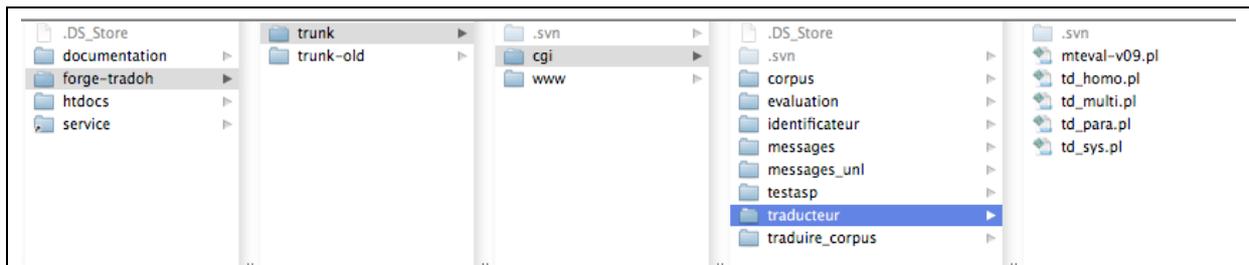


Figure 5 : Organisation de l'implémentation actuelle de TRADOH

5.5 PROJET TRADOH++ DANS LA FORGE DU LIG

Figure 6 : Page d'accueil du projet TRADOH++ sur la forge du LIG