HAL
open science

# Multilinguïsation de services de gestion de contenu.

Najeh Hajlaoui

## To cite this version:

Najeh Hajlaoui. Multilinguïsation de services de gestion de contenu.. IC "Ingénierie des Connaissances", plate-forme AFIA "Association Française pour l'Intelligence Artificielle", Jul 2007, Grenoble, France. 2 p. hal-00390864

**HAL Id: hal-00390864**

**https://hal.science/hal-00390864v1**

Submitted on 2 Jun 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multilinguïsation de services de gestion de contenu

Najeh Hajlaoui

Laboratoire LIG, GETALP, Université Joseph Fourier
385 rue de la Bibliothèque, BP n° 53
38041 Grenoble, cedex 9, France
Najeh.Hajlaoui@imag.fr

**Résumé** : La multilinguïsation des services de gestion de contenu est peu fréquente. Elle constitue un problème important et difficile. En effet, cela dépend de la situation traductionnelle qui représente un ensemble de facteurs prépondérants : types et niveau d'accès possibles, ressources disponibles, compétences langagières et linguistiques des intervenants pour la multilinguisation des applications… Plusieurs stratégies de multilinguisation sont alors possibles (par adaptation interne ou externe, par traduction …). Nous illustrons cette étude par un cas réel de portage linguistique (arabe vers français) d'application d'e-commerce déployée et traitant des textes de SMS spontanés concernant l'achat et la vente de voitures d'occasion. Malgré la grande distance entre l'arabe et le français, les méthodes de localisation utilisées marchent bien à cause de la proximité des deux sous-langages.

**Mots-clés** : Localisation, Enoncés Spontanés en Langue Naturelle (ESLN), Extraction de contenu, SMS, CRL (Content Representation Language).

## 1 Introduction

La multilinguïsation des services de e-commerce traitant des énoncés spontanés en langue naturelle est un problème important mais difficile et très peu de services le font. En effet, cela dépend de deux facteurs par rapport à la situation traductionnelle :

- Le niveau d'accès aux ressources des applications, dans lequel quatre cas se présentent : accès complet au code source, accès limité à la représentation interne, accès limité au dictionnaire, et aucun accès.
- Le niveau de compétences langagières et linguistiques des intervenants pour la multilinguisation des applications qui peut-être défini par rapport à la langue source ou par rapport aux compétences linguistiques de l'équipe.

Nous présentons d'abord CATS, une application déployée pour laquelle les deux facteurs présentés ci-dessus sont assurés et plusieurs stratégies de multilinguisation sont alors possibles. Nous présentons, ensuite deux stratégies déjà expérimentées.

## 2 Présentation de CATS

CATS est un système d'achat et de vente basé sur l'utilisation des SMS en arabe (Daoud 2006). Il est déployé en Jordanie par la société FastLink. Bien qu'il n'y ait pas

de transaction directe, CATS aide les utilisateurs à vendre et acheter en les mettant en relation (sans qu'ils aient à se déplacer). Ces SMS sont envoyés à un numéro spécial unique et sont enregistrés automatiquement dans un corpus de données. Leurs textes rédigés d'une façon spontanée, sont analysés et une réponse est envoyée automatiquement à l'expéditeur du SMS en cas de correspondance de la demande avec l'une des propositions.

## 3 Illustration : portage interne et externe

Une première méthode de portage de CATS de l'arabe vers le français et qui nécessite un accès total aux ressources de l'application consiste à adapter son extracteur de contenu prévu initialement pour l'arabe. Cela pose bien sûr le problème de trouver un corpus et un dictionnaire français fonctionnellement équivalents à ceux de l'arabe (Hajlaoui 2006).

Notons que la partie grammaticale a été très faiblement modifiée, ce qui prouve que, malgré la grande distance entre l'arabe et le français, ces deux sous-langages sont très proches l'un de l'autre, une nouvelle illustration de l'analyse de R. Kittredge (Kittredge and Lehrberger 1982).

Une deuxième méthode consiste à adapter un extracteur de contenu pour la même langue (ici, français) mais pour un autre domaine celui du tourisme. Cette méthode nécessite un simple accès à la représentation interne de l'application CATS mais elle demande un compilateur de format qui permet de traduire de la représentation IF (Interchange Format) (Blanchon 2004) à la représentation CRL-CATS (Content Representation Langage) (Daoud 2006).

Nous avons expérimenté les deux méthode de portage, interne et externe. Elles marchent bien et donnent respectivement un pourcentage de portage par rapport à version originale de 98% et de 77 %.

## 4 Conclusion

Nous avons choisi CATS, comme application à localiser parce que c'est une plate-forme qui traite des ESLN et qu'on a accès à toutes ses ressources. Nous avons présenté deux méthodes de portage « interne » et « externe » qui nécessitent respectivement un accès total et un accès limité à la représentation interne de CATS. Malgré la grande distance qui existe entre le français et l'arabe, les deux méthodes donnent de bons résultats à cause de la proximité des sous-langages.

# Références

BLANCHON H. (2004). Comment définir, mesurer et améliorer la qualité, l'utilisabilité et l'utilité des systèmes de TAO de l'écrit et de l'oral. Une bataille contre le bruit, l'ambiguïté, et le manque de contexte. Grenoble, Université Joseph Fourier**:** 380 p.

DAOUD D. M. (2006). It is necessary and possible to build (multilingual) NL-based restricted e-commerce systems with mixed sublanguage and contend-oriented methods. *GETA, CLIPS*. 296 p. Grenoble.

HAJLAOUI N. (2006). Recherche et production de corpus de messages pour la multilinguïsation de sites. In Business Information Association Conference. Bonn, Allemagne.

KITTREDGE R. & LEHEBERGER J. (1982). Sublanguage : study of language in restricted semantic domain.