

Extraction bilingue de termes médicaux dans un corpus parallèle anglais/français

Aurélie Névéol^{1,2}, Sylwia Ozdowska³

¹Laboratoire PSI – FRE CNRS 2645
INSA de Rouen - BP8 - Avenue de l'Université
76801 Saint Etienne du Rouvray Cedex , France
aneveol@insa-rouen.fr

²Equipe CISMef et L@STICS
Faculté de Médecine de Rouen, 1 rue de Germont
76031 Rouen – France.

³ Equipe de Recherche en Syntaxe et Sémantique
Université de Toulouse le Mirail, 5 allées Antonio Machado
31058 Toulouse Cedex 1 – France.
ozdowska@univ-tlse2.fr

Résumé. Le Catalogue et Index des Sites Médicaux Francophones (CISMef) recense les principales ressources institutionnelles de santé en français. La description de ces ressources, puis leur accès par les utilisateurs, se fait grâce à la terminologie CISMef, fondée sur le thésaurus américain Medical Subject Headings (MeSH). La version française du MeSH comprend tous les descripteurs MeSH, mais de nombreux synonymes américains restent à traduire. Afin d'enrichir la terminologie, nous proposons ici une méthode de traduction automatique de ces synonymes. Pour ce faire, nous avons constitué deux corpus parallèles anglais/français du domaine médical. Après alignement semi-automatique des corpus paragraphe à paragraphe, nous avons procédé automatiquement à l'appariement bilingue des termes. Pour cela, le lexique constitué des descripteurs MeSH américains et de leur traduction en français a fourni les couples amorces qui ont servi de point de départ à la propagation syntaxique des liens d'appariement. 217 synonymes ont pu être traduits, avec une précision de 70%.

1 Introduction

La recherche d'information, l'indexation, et la manipulation de ressources multimédia en général sont des domaines qui s'appuient sur l'utilisation d'une terminologie pour décrire les ressources disponibles et y accéder. Dans le domaine bio-médical, de nombreux travaux ont été réalisés en ce sens et plusieurs terminologies (par exemple, le MeSH¹ pour la gestion de connaissances, ou la SNOMED CT² pour les termes cliniques) ou ontologies (par exemple,

¹ Medical Subject headings. cf. <http://www.nlm.nih.gov/mesh/meshhome.html>

² SNOMED Clinical Terminology. cf. <http://www.nhs.uk/snomed/pages/default.asp>