

# L'apport d'une approche symbolique pour le repérage des entités nommées en langue amazighe

Meryem Talha\*, Siham Boulaknadel\*\*, Driss Aboutajdine\*

\*LRIT, Unité Associée au CNRST  
Faculté Des Sciences, Université Mohammed V - Agdal Rabat, Maroc  
meriem.talha@gmail.com  
aboutaj@fsr.ac.ma

\*\*Institut Royal de la Culture Amazighe  
Avenue Allal El Fassi, Madinat Al Irfane, Rabat-Instituts, Maroc  
boulaknadel@ircam.ma

**Résumé.** Le repérage des Entités Nommées (REN) en langue amazighe est un prétraitement éventuellement essentiel pour de nombreuses applications du traitement automatique des langues (TAL), en particulier pour la traduction automatique. Dans cet article, nous présentons une chaîne de repérage des entités nommées en amazighe fondée sur une étude synthétique des spécificités de la langue et des entités nommées en amazighe. L'article met l'accent sur les choix méthodologiques à résoudre les ambiguïtés dues à la langue, en exploitant les technologies existantes pour d'autres langues.

## 1 Introduction

Depuis les débuts du TALN, la compréhension de texte fait l'objet d'un suivi particulier de plusieurs recherches. C'est en faveur du développement rapide de la tâche d'extraction d'information que la tâche de REN s'est manifestée. Elle consiste à rechercher les expressions référentielles (Ehrmann (2008)), qui recouvrent classiquement les noms désignant des personnes, des lieux, des organisations, des expressions temporelles et celles numériques, mais peuvent aussi se rapporter à des notions plus techniques comme les maladies. Dès la campagne MUC-6, la tâche de REN s'est ainsi polarisée sur trois types d'entités (Grishman et Sundheim (1996)), à savoir : ENAMEX (personnes, organisations et lieux), TIMEX (expressions temporelles), NUMEX (expressions numériques). Cette première définition a été étendue dans la campagne CoNLL (Tjong Kim Sang et De Meulder, 2003) où 4 classes ont été normalisées : personnes, organisations, lieux, Divers. Dans les campagnes d'évaluation ESTER2 (Galliano et al., 2009) 8 catégories ont été normalisées à savoir personnes, fonctions, organisations, lieux, productions humaines, dates, montants et événements.