

An Efficient Multi-Agent System Combining POS-Taggers for Arabic Texts

Chiraz Ben Othmane Zribi¹, Aroua Torjmen¹, Mohamed Ben Ahmed¹

¹ RIADI laboratory, National School of Computer Sciences, 2010,
University of La Manouba, Tunisia
{Chiraz.benothmane, Aroua.torjmen,
Mohamed.benahmed}@riadi.rnu.tn

Abstract. In this paper, we address the problem of Part-Of-Speech tagging of Arabic texts with vowel marks. After the description of the specificities of Arabic language and the induced difficulties on the task of POS-tagging, we propose an approach combining several methods. One of these methods, based on sentences patterns, is original and very attractive. We present, afterward, the multi-agent architecture that we adopted for the conception and the realization of our POS-tagging system. The multi-agent architecture is justified by the need for collaboration, parallelism and competition between the different agents. Finally, we expose the implementation and the evaluation of the system implemented.

1 Introduction

The process of Part-Of-Speech tagging was widely automated for English and French and for many others European languages giving a rate of accuracy ranging from 95 % to 98 %. We find on the Web, many tagged corpora as well as programs of POS-tagging for these languages. The methods used by these POS-taggers are various, namely stochastic approaches such as the Hidden Markov Model [1], the decision trees [2], the maximum entropy model [3], rules-based approaches inspired in their majority of the transformation rules-based POS-tagging [4], hybrid approaches [5] (statistics and rules-based), or combined ones [6] and [7].

Unfortunately, the situation is different for Arabic as there are neither POS-taggers nor tagged corpora available. Nevertheless, some Arabic POS-taggers [8], [9] and [10] started to appear with an accuracy going from 85% to 90% on average for texts with vowel marks and by about 65% for texts without vowel marks.

This gap noted for Arabic language is especially due to, its particular characteristics, which, involve firstly, a rate of grammatical ambiguity relatively more significant than for other languages, and secondly, make impossible the application of existing POS-taggers without any change. Thus, obtaining improving accuracy remains a challenge to reach for Arabic language.

Accordingly, we propose a POS-tagging system for Arabic texts. Due to the complexity of the problem, and in order to decrease grammatical ambiguity, we have restricted the scope of our investigation: we only treat texts with vowels marks.