Chapter 9

Translational Mismatches Involving Clitics (Illustrated from Serbian ~ Catalan Language Pair)

Jasmina Milićević
Dalhousie University, Canada

Àngels Catena
Universitat Autònoma de Barcelona, Spain

ABSTRACT

Translation of sentences featuring clitics often poses a problem to machine translation systems. In this chapter, we illustrate, on the material from a Serbian ~ Catalan parallel corpus, a rule-based approach to solving translational structural mismatches between linguistic representations that underlie source- and target language sentences containing clitics. Unlike most studies in this field, which make use of phrase structure formalisms, ours has been conducted within the dependency framework of the Meaning-Text linguistic theory. We start by providing a brief description of Catalan and Serbian clitic systems, then introduce the basics of our framework to finally illustrate Serbian ~ Catalan translational mismatches involving the operations of clitic doubling, clitic climbing, and clitic possessor raising.

INTRODUCTION

This chapter focuses on translational mismatches between Serbian and Catalan sentences featuring clitics. A clitic is a deficient wordform in that it lacks inherent stress and has to lean prosodically on a stressed wordform (or phrase) in the clause, called the host of this clitic. Prototypical examples of clitics are clitic pronouns, but there are also clitic auxiliaries, conjunctions, particles, etc. In addition to having outstanding phonological behavior, clitics can be set apart form “normal” wordforms by their morphology (special inflection and/or external sandhis) and syntax (clustering, rigid linear placement even in so called free word-order languages). The literature on clitics is huge; for a general and typological

DOI: 10.4018/978-1-4666-8690-8.ch009
Translational Mismatches Involving Clitics

perspective on clitics, see in particular Zwicky (1977), Klavans (1995), Halpern (1995) and Spencer and Louis (2012); for work on Serbian and Catalan clitics, as well as for illustrative examples, see below.

By translational mismatches, or divergences, we mean non-trivial correspondences between linguistic representations that underlie a source language sentence and its equivalent in the target language. More specifically, we are interested in structural mismatches, which represent cases of violation of isomorphism between linguistic representations (Mel’čuk 2006, pp. 105-106). Two linguistic representations, say two syntactic dependency trees, are isomorphic if 1) all their nodes are in one-to-one correspondence based on the identity or semantic equivalence of their lexical labels, and 2) the dependency inside any two corresponding pairs of nodes has the same direction (i.e., there is no inversion of subordination, or head-switching). An influential typology of structural mismatches was proposed in Dorr (1993, 1994) and further developed in Mel’čuk and Wanner (2006); more on this will be said in due course.

For example, the (surface) syntactic structures of the Serbian sentence (1a) and its Catalan equivalent (1b) feature a structural mismatch involving clitic pronouns (boldfaced); for the structures themselves, see Figure 1.1

(1) a. Ponudiše ga votkom
   they.treated he-ACC.SG vodka-INSTR.SG
   ‘They treated him with vodka’.

   b. Li van oferir vodka
   he-DAT.SG they.go offer vodka
   ‘They offered him vodka’.

While the Serbian verb takes the Recipient (= the person getting the offer) as the DirO, for its Catalan equivalent the same complement is an IndirO, which results in different case assignment to the clitics (accusative vs. dative). Moreover, in the Catalan sentence, the IndirO clitic is raised to the past auxiliary, van, and linearly positioned with respect to it. Roughly speaking, raising is a syntactic operation whereby a syntactic dependent changes its governor for one that is higher up in the syntactic hierarchy (in our case, the dative clitic li, a dependent of the verb oferir, becomes a dependent of the governor of oferir).2

Translation of clitic sequences represents a difficult NLP task, as can be concluded from many publications dealing with text generation or machine translation in which the problem is mentioned, among others, Russo (2010), Larasati (2012) and Din (2013). A major clitic related problem encountered when

---

1. See Figure 1 for the syntactic structures of sentences (1).
2. For more details on raising and head-switching, see Mel’čuk (2006, pp. 66-67).