International Journal of E-Services and Mobile Applications
Volume 13 • Issue 4 • October-December 2021

Application of Morphosyntactic Cues in Detection of GOAL Semantic Role

Nives Mikelic Preradovic, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Tomislava Lauc, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia
Danijela Unic, VERN' University of Applied Sciences, Croatia

ABSTRACT

This paper analyzes the semantics of verbs with the prefix "do-" and explains the adlativity feature based on the morpho-syntactically annotated corpus hrWaC and handcrafted verb valency frames. The work aims to automatically add all types of adlativity to Croatian verb valency lexicon. As a result, it was revealed that if a language resource encodes "do-" as the adlative prefix in Croatian as a source language, then the adlative meaning in the target language can be assumed as well. Using the valency frame transition rules for language pairs, it is possible to design matching verb valency frames in other languages and consequently describe each verb and its translation by semantic roles (agent, patient, direction-to, and goal) and by selectional restrictions.

KEYWORDS

Adlativity, Croatian Verb Valency Lexicon CroVallex, Cross-Linguistic Rules, Frame Transition Rules, Machine Translation Post-Editing, Motion Verbs, Semantic Role Labeling, Verb Valency

INTRODUCTION

Numerous research projects in linguistics, psychology and computer science deal with how languages manage to encode motion, i.e. the way languages encode changing the position of an entity in relation to another entity (Vulchanova & van der Zee, 2013, p. 2).

In this paper, we deal with the semantics of motion verbs with the prefix "do-" and the adlativity feature of the motion prefixed verbs in Croatian, while exploring the possibility for a cross-linguistic approach to semantic role labelling (more precisely the labelling of DIR3 (Direction-To) and Goal semantic roles in different languages), as well as solving semantic ambiguities in machine translation post-editing or post-processing.

The paper aims to present how semantic role labeling of adlativity (expressing motion and direction) could help in computational linguistics for further text processing and in practical implementation, as in machine translation where morphological and syntactic errors influence the perceived quality of machine-translated text (Seljan et al., 2015).

Semantic role labeling is a natural processing task that aims to reveal who did what to whom, when and why.

DOI: 10.4018/IJESMA.2021100103

From the practical viewpoint, semantic role labelling is valuable for a whole range of natural processing tasks, such as question answering (Yih et al., 2016; Fitzgerald, He, & Zettlemoyer, 2018), event detection (Mikelic Preradovic et al, 2013), machine reading comprehension, textual entailment and context-oriented dialogue (Zhang et al., 2018), since deep learning models do not really understand the natural language texts (Mudrakarta et al., 2018). Also, semantic role labelling proved to be useful for machine translation (Shi et al., 2016) and discourse relation sense classification (Mihaylov & Frank, 2016).

The main contribution of this paper is to present the framework of semantic role labelling for adlative meaning in 7 languages, as presented in Table 2.

Motion encoding in the spatial language (the part of language describing perceived space) refers to how languages encode path of motion (i.e. direction) which has a beginning point (source of motion), a middle point (trajectory) and an endpoint or goal (Luraghi, Nikitina & Zanchi, 2017, p. 2). Directional meaning can be categorized as follows: (1) the adlative that expresses direction "moving closer to"; (2) the ablative that expresses direction "moving away from" and (3) the perlative that expresses motion which is directionally unspecified and represents only the path of the moving object (Mikelić Preradović, Boras & Lauc, 2013). Talmy (2000) differentiates satellite-framed languages (S-languages) where the path is lexicalized as a "satellite" to the verb (e.g. English, Slavic and Germanic languages) and verb-framed languages (V-languages) where the path is lexicalized in the root of the motion verb, while the meaning of a verb is provided by its prefix (e.g. Romance languages). The satellite may be either a verb prefix (like Russian or Croatian inseparable verb prefixes or German separable and inseparable verb prefixes) or a free word (like English verb particles). Regarding differences in verb prefixation of the motion verb, Dickey (2010) distinguishes the North Slavic allative prefixation (e.g. Russian and Czech) and South Slavic allative prefixation (e.g. Croatian and Serbian), where the first group of languages generalizes "pri-" ("to") as the prefix of the motion verbs that signal the crossing of the boundary of the goal, while the latter generalizes "do-" ("to") as the prefix of the motion verbs that do not assert the crossing of the boundary of the goal, but only the traversal of a trajectory up to the goal. In other words, apart from the difference between "pri-" and "do-", goal prepositional phrases in the North Slavic languages specify the crossing of the boundary, while the goal prepositional phrases in the South Slavic languages can specify the boundary crossing or just reaching the boundary.

Considering verb prefixation, it is possible to differentiate prefixes in regard to their productivity in language. Thus, "do-" proved to be a very productive verbal prefix in Croatian and other South Slavic languages and there is a vast literature discussing it (Pranjković, 2009; Sarić, 2008; Brala, 2012). Besides, previous research on verb prefixation for Croatian (Mikelić Preradović, Boras & Lauc, 2013) showed that most of the verbs sharing the same meaning of the specific prefix, belong to the same semantic class sharing the same valency frame. Also, according to recent studies (Cheikhrouhou, 2014; Ghorbel, 2017) formal grammars, created for verbs of communication that take into consideration the difference in syntactic constructions and the types of complements, solve semantic ambiguities which represent the main source of errors in automatic translation of verbs.

In general, errors in automatic translation are language-dependent and influenced by the type of text as well as the engine itself, and usually follow a pattern. There are three main machine translation (MT) approaches: rule-based machine translation (RMBT), statistical machine translation (SMT) and neural machine translation (NMT). The last approach has been developed more recently, while rule-based and statistical approaches have a long history. Statistical machine translation systems use probabilities between segments in a source and target language document to propose candidates for translation (Dunder, 2015). These systems include a statistical model based on the corpus analysis, while a rule-based approach uses language rules to transfer the meaning to the target language (Koglin & Cunha, 2019).

Considering the repetitive nature of errors made by rule-based machine translation systems, Simard, Ueffing, Isabelle and Kuhn (2007) have introduced a statistical phrase-based machine