

Unsupervised Learning of Verb Argument Structures

Thiago Alexandre Salgueiro Pardo¹, Daniel Marcu², Maria das Graças Volpe Nunes¹

¹ Núcleo Interinstitucional de Linguística Computacional (NILC)
CP 668 – ICMC-USP, 13.560-970 São Carlos, SP, Brasil
<http://www.nilc.icmc.usp.br>

² Information Sciences Institute (ISI)
4676 Admiralty Way, Suite 1001, Marina del Rey, CA 90292
<http://www.isi.edu>

tasparado@gmail.com, marcu@isi.edu, gracan@icmc.usp.br

Abstract. We present a statistical generative model for unsupervised learning of verb argument structures. The model was used to automatically induce the argument structures for the 1,500 most frequent verbs of English. In an evaluation carried out for a representative sample of verbs, more than 90% of the induced argument structures were judged correct by human subjects. The induced structures also overlap significantly with those in PropBank, exhibiting some correct patterns of usage that are not present in this manually developed semantic resource.

1 Introduction

Inspired by the impact that the availability of Penn Treebank (Marcus et al., 1993; Marcus, 1994) had on syntactic parsing, several efforts have recently focused on the creation of semantically annotated resources. The annotation of verb arguments, their roles, and preferential linguistic behaviors represents a significant fraction of these efforts. The annotations that we are focusing on here pertain to the argument structures of a verb. In particular, we look for the words/concepts that constitute the arguments required by the verbs when these are used in real sentences.

The determination of verb argument structures has been shown to be a hard task for several reasons. Little agreement exists with respect to (a) how many canonical usages a verb has, (b) which arguments are really required by a verb and (c) in what order they may be realized in sentences. For instance, examples (1)-(3) show some patterns of usage for the verb *bought*.

- (1) He had bought them gifts.
- (2) He bought it 40 years ago.
- (3) About 8 million home water heaters are bought each year.

Intuitively, one can induce from these examples that the object/thing that is bought (“gifts” in sentence (1), “it” in sentence (2), and “about 8 million home water heaters” in sentence (3)) is more likely to be a required argument for the verb than the time when the buying event occurred, since the thing bought is specified in all the cases