Chapter 16

PEGASUS: Deriving Argument Structures after Syntax

Karen Jensen

Abstract

PEGASUS is the third component in the PLNLP analysis system, following syntax and reassignment. Its purpose is to produce a semantic representation, or logical form, for each input sentence or sentence fragment. To do this it computes: (a) the structure of arguments and adjuncts for each clause; (b) NP (pronoun)-anaphora; and (c) VP-anaphora (for elided VPs). While doing this, PEGASUS must maintain broad coverage (that is, accept and analyze unrestricted input text). More commonly in NLP systems, the computation of such meaning structures is considered impossible unless a particular domain is specified. This chapter explains these steps and then compares the PLNLP approach with other current approaches to defining predicate-argument structures.
16.1 Introduction

PEGASUS, the third component in the PLNLP analysis system after syntax and reassignment, provides a definitive move from syntax to semantics, where "semantics" is understood to involve, minimally, the definition of case frames or thematic roles (i.e., predicate-argument relations). The most obvious display of this is in its input and output representations. Input is shown as a parse tree; output, as a labeled, directed graph. (The underlying information is in the form of attribute-value record structures throughout.) A tree is primarily a syntactic representation, where linear ordering and categorial dominance are significant. A graph is a semantic representation; linear ordering is no longer relevant because whatever information it provided has been assigned to arc labels or features in the graph. This output can also be called a logical form.

In order to derive the logical form, PEGASUS must correctly make many challenging argument assignments, such as long-distance dependencies (e.g., assigning the right object for "ate" in "What did Mary say that John ate?"); functional control (e.g., assigning the proper subjects and objects to infinitives); the active/passive relationship (making sure that active and passive variants have the same underlying arguments); and so forth. The program must also identify meaningful relationships between head words of phrases and their modifiers or adjuncts. In addition, NP-anaphora (including pronoun and definite noun phrase referents) and VP-anaphora (assigning the correct arguments and adjuncts within elided VPs) must be completed; and the entire input string must be properly quantified. Currently PEGASUS does not handle definite NP reference or quantification, but does handle the other phenomena mentioned here.

16.2 Arguments and adjuncts

Consider the sentence, "After dinner, Mary gave a cake to John." Figure 1 shows the syntactic (tree) representation for that sentence after it has been processed by the first two analysis components, and figure 2 shows the semantic graph produced by PEGASUS for the same sentence.

A graph is produced by displaying only those attributes and values that are defined to be semantic. However, the underlying record structure contains all attributes resulting from the parse. In this fashion, all levels and types of information, from morphological to syntactic to semantic and beyond, are constantly available. This principle of accountability holds throughout the PLNLP system.