

## Chapter 8.15

# Dependency Parsing: Recent Advances

**Ruket Çakıcı**  
*University of Edinburgh, UK*

### INTRODUCTION

Annotated data have recently become more important, and thus more abundant, in computational linguistics. They are used as training material for machine learning systems for a wide variety of applications from Parsing to Machine Translation (Quirk et al., 2005). Dependency representation is preferred for many languages because linguistic and semantic information is easier to retrieve from the more direct dependency representation. Dependencies are relations that are defined on words or smaller units where the sentences are divided into its elements called heads and their arguments, e.g. verbs and objects. Dependency parsing aims to predict these dependency relations between lexical units to retrieve information,

mostly in the form of semantic interpretation or syntactic structure.

Parsing is usually considered as the first step of Natural Language Processing (NLP). To train statistical parsers, a sample of data annotated with necessary information is required. There are different views on how informative or functional representation of natural language sentences should be. There are different constraints on the design process such as: 1) how intuitive (natural) it is, 2) how easy to extract information from it is, and 3) how appropriately and unambiguously it represents the phenomena that occur in natural languages.

In this article, a review of statistical dependency parsing for different languages will be made and current challenges of designing dependency tree-banks and dependency parsing will be discussed.

DOI: 10.4018/978-1-60960-818-7.ch8.15

## DEPENDENCY GRAMMAR

The concept of dependency grammar is usually attributed to Tesnière (1959) and Hays (1964). The dependency theory has since developed, especially with the works of Gross (1964), Gaiffman (1965), Robinson (1970), Mel'čuk (1988), Starosta (1988), Hudson (1984, 1990), Sgall et al. (1986), Barbero et al. (1998), Duchier (2001), Menzel and Schröder (1998), Kruijff (2001).

Dependencies are defined as links between lexical entities (words or morphemes) that connect heads and their dependants. Dependencies may have labels, such as *subject*, *object*, and *determiner* or they can be unlabelled. A dependency tree is often defined as a directed, acyclic graph of links that are defined between words in a sentence. Dependencies are usually represented as trees where

the root of the tree is a distinct node. Sometimes dependency links cross. Dependency graphs of this type are non-projective. Projectivity means that in surface structure a head and its dependants can only be separated by other dependants of the same head (and dependants of these dependants). Non-projective dependency trees cannot be translated to phrase structure trees unless treated specially. We can see in Table 1 that the notion of non-projectivity is very common across languages although distribution of it is usually rare in any given language. The fact that it is rare does not make it less important because it is this kind of phenomena that makes natural languages more interesting and that makes all the difference in the generative capacity of a grammar that is suggested to explain natural languages.

Table 1. Treebank information; #T = number of tokens \* 1000, #S = number of sentences \* 1000, #T/#S = tokens per sentence, %NST = % of non-scoring tokens (only in CoNLL-X), %NPR = % of non-projective relations, %NPS = % of non-projective sentences, IR = has informative root labels

Language	#T		#S		#T/#S		%NST		%NPR		%NPS		IR	
Arabic	54	112	1.5	2.9	37.2	38.3	8.8	-	0.4	0.4	11.2	10.1	Y	-
Basque	-	51	-	3.2	-	38.3	-	-	-	2.9	-	26.2	-	-
Bulgarian	190		12.8	-	14.8	-	14.4	-	0.4	-	5.4	-	N	-
Catalan	-	431	-	15	-	28.8	-	-	-	0.1	-	2.9	-	-
Chinese	337	337	57	57	5.9	5.9	0.8	-	0.0	0.0	0.0	0.0	N	-
Czech	1249	432	72.7	25.4	17.2	17.0	14.9	-	1.9	1.9	23.2	23.2	Y	-
Danish	94	-	5.2	-	18.2	-	13.9	-	1.0	-	15.6	-	N	-
Dutch	195	-	13.3	-	14.6	-	11.3	-	5.4	-	36.4	-	N	-
English	-	447	-	18.6	-	24.0	-	-	-	0.3	-	6.7	-	-
German	700	-	39.2	-	17.8	-	11.5	-	2.3	-	27.8	-	N	-
Greek	-	65	-	2.7	-	24.2	-	-	-	1.1	-	20.3	-	-
Hungarian	-	132	-	6.0	-	21.8	-	-	-	2.9	-	26.4	-	-
Italian	-	71	-	3.1	-	22.9	-	-	-	0.5	-	7.4	-	-
Japanese	151	-	17	-	8.9	-	11.6	-	1.1	-	5.3	-	N	-
Portuguese	207	-	9.1	-	22.8	-	14.2	-	1.3	-	18.9	-	Y	-
Slovene	29	-	1.5	-	18.7	-	17.3	-	1.9	-	22.2	-	Y	-
Spanish	89	-	3.3	-	27	-	12.6	-	0.1	-	1.7	-	N	-
Swedish	91	-	11	-	17.3	-	11.0	-	1.0	-	9.8	-	N	-
Turkish	58	65	5	5.6	11.5	11.6	33.1	-	1.5	5.5	11.6	33.3	N	-

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/dependency-parsing-recent-advances/56245](http://www.igi-global.com/chapter/dependency-parsing-recent-advances/56245)

## Related Content

---

### Hate and Aggression Detection in Social Media Over Hindi English Language

Kapil Pareek, Arjun Choudhary, Ashish Tripathi, K. K. Mishra and Namita Mittal (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-20).

[www.irma-international.org/article/hate-and-aggression-detection-in-social-media-over-hindi-english-language/300357](http://www.irma-international.org/article/hate-and-aggression-detection-in-social-media-over-hindi-english-language/300357)

### Design and Implementation of a Cognitive User-Support System for Skin Progress Analysis Using a Smart Phone

Hiroyuki Nishiyama and Fumio Mizoguchi (2013). *International Journal of Software Science and Computational Intelligence* (pp. 51-62).

[www.irma-international.org/article/design-and-implementation-of-a-cognitive-user-support-system-for-skin-progress-analysis-using-a-smart-phone/88991](http://www.irma-international.org/article/design-and-implementation-of-a-cognitive-user-support-system-for-skin-progress-analysis-using-a-smart-phone/88991)

### On the Cognitive Complexity of Software and its Quantification and Formal Measurement

Yingxu Wang (2012). *Software and Intelligent Sciences: New Transdisciplinary Findings* (pp. 264-286).

[www.irma-international.org/chapter/cognitive-complexity-software-its-quantification/65134](http://www.irma-international.org/chapter/cognitive-complexity-software-its-quantification/65134)

### Modeling the Ecological Footprint of Nations via Evolutionary Computation and Machine Learning Models

Mohamed M. Mostafa (2012). *Machine Learning: Concepts, Methodologies, Tools and Applications* (pp. 2081-2100).

[www.irma-international.org/chapter/modeling-ecological-footprint-nations-via/56243](http://www.irma-international.org/chapter/modeling-ecological-footprint-nations-via/56243)

### Petri Nets and Discrete Events Systems

Juan L. G. Guirao and Fernando L. Pelayo (2011). *International Journal of Software Science and Computational Intelligence* (pp. 13-22).

[www.irma-international.org/article/petri-nets-discrete-events-systems/60746](http://www.irma-international.org/article/petri-nets-discrete-events-systems/60746)