

## Introduction

Ibn Mada was the first grammarian ever to use the term *dependency* in the grammatical sense that we use it today. He was born in 1119 in Cordoba, studied in Sevilla and Ceuta, and died 1195 in Sevilla. He is known for his only book, *Radd: the refutation of grammarians*, in which he tackles subjects that still seem modern today: He criticizes the use of ellipsis, i.e. underlying invisible forms, in the analyses of grammatical phenomena, for example when discussing whether the unmarked nominative form consists of a zero marker. He further speaks out against semantic interpretation and justification of grammatical rules, and thus in favor of an independence of syntax, semantics, and cognitive interpretation: “*And why is the agent in the nominative?*” *The correct answer is [...] : “This is how the Arabs speak”*.

And he used the term *تعلق* *Ta'alluq* which translates as *being suspended to, dependent upon, connected with; attachment; love of the world; dependence ; connection; relation; relationship ; concern, reference, regard ; consideration, reflection ; commerce ; means of support, employment, office ; property, possession ; a manor ; a small division of a district*, when referring to the relation between verbs and their direct and indirect *dependents*. He prefers this term to *عمل* ‘*amal* ('operation', 'government'), the commonly used term at his time for relations between *governing* and *dependent* words, because, following Ibn Mada, the *head* word does not *operate* on its dependents, but he only sees a *relation*, a *dependency*. He goes as far as calling it *heretic* to use *amal* because words cannot act on other words and *cause* an inflection. Since this was merely a change in vocabulary, the use of *dependency* did not catch on until the 20<sup>th</sup> century. The importance of a dependency type analysis for the description of language, however, was well-established in the Arabic grammatical tradition before Ibn Mada and can even partially be traced back to Panini.

So why, you might ask, do we need a conference on Dependency Linguistics in Barcelona, when grammarians have done dependency linguistics in Spain for 1000 years?

The generative grammatical tradition that, in its origins, solely attempts to construct a system that distinguishes grammatical from ungrammatical sentences, left linguistics in a state where the grammatical analysis, phrase structure, was difficult to connect to deeper (semantic, conceptual) structures. The result was a complete separation between, on one side, Natural Language Processing that needed deeper analyses, for translation, classification, generation etc. and, on the other side, generative linguistics that built complex structures with the declared goal to model Language as a whole, where the structures got more and more complicated the further the described language is from English. In the second half of the 20<sup>th</sup> century, only a few linguists, often referring themselves to Lucien Tesnière, continued to describe language in terms of dependency, mainly because they were working on free word order languages, where the use of phrase structure is more clearly maladaptive.

Since the 1990s, NLP is turning towards dependency analysis, and in the past five years dependency has become quasi-hegemonic: The very large majority of parsers presented in recent NLP conferences are explicitly dependency-based. It seems, however, that the connection between computational linguists and dependency linguists remains sporadic: What happens commonly is that someone transfers an existing tree bank into a dependency format that fits his or her needs, and other researchers attempt to reproduce this annotation, with statistical or rule-based grammars. Not that the situation was any better when parsers still automatized phrase structure construction and linguistics discussed *move alpha*. Yet, we believe that the situation is different today and dependency linguists and computational linguists have a lot to share:

We know that statistical parsers give better results if we have a linguistically coherent corpus analysis. We need to know what the differences are between surface and deep dependency. How to define dependency? What are the units that appear in dependency analysis? What set of labels (particularly syntactic functions) do we use? Do we agree on the same syntactic representations? Or simply, what are the others doing? What kind of analysis works for which application? How to link dependency to structures to the lexicon and to semantics?



Not all questions will find a direct answer, but we believe that Dependency Linguistics 2011 provides a forum allowing for an interchange between researchers on the theoretical and the applicative sides of current linguistics, on various languages, and on various underlying models.

The conference is organized in thematic sessions:

We will start with the *theoretical foundations* of dependency: What types of dependency exist, how to define dependency, how to handle coordination and discontinuities, how to relate dependency and morphology, as well as, more specifically, how to handle clitics, and finally how to translate Tesnièrean notions into a grammar formalism.

In the *semantics* session, we learn about the relations of dependency structures to Frame Semantics, about the semantic analysis of person names and about semantic structures on learner corpora.

A big part of the work presented at this conference concerns *treebanks*: A syntactic annotation scheme for Spanish, error analysis for Finnish, a multi-layer corpus annotated in terms of the Generative Lexicon, the analysis of coordination on a learner corpus of English, the detection of code switching in an English-German by means of the dependency distance, user-centered syntactic annotation for Finnish, and the extraction of valency patterns from a Chinese treebank.

*Linguistic issues* include the relationship of grammar and lexicon, the definition of unmarked word order in Czech, the Prodrop problem in Arabic, the interrogative clitic of Turkish, wh-copying in German, free word order in Japanese noun phrases, and parallels between syntax and discourse.

The session on *formal topics* presents the prosody syntax interface in an analysis of Hebrew, statistical language generation, and categorical dependency grammars, as well as tools for their development.

Last but not least, *dependency parsing* will be presented under its various aspects: A comparison of graph-based and transition-based parsers, incremental parsing, improving dependency label accuracy, a comparison of rule-based and data-driven parsers, and a rule-based dependency parser for Russian.

Overall, these proceedings include 33 articles from 16 countries: Canada, China, Czech Republic, Denmark, Egypt, Finland, France, Germany, Great Britain, Ireland, Israel, Italy, Japan, Russia, Spain, and the United States.

We would like to thank Igor Mel'čuk and Joakim Nivre for having accepted our invitation to give talks on the two fundamental sides of our conference: dependency analysis and dependency parsing. We would also like to thank the program committee for their participation and their astonishingly thorough reviews that have certainly contributed to the quality of the papers presented here:

Margarita Alonso Ramos	University of La Coruña
Lorraine Baqué	Autonomous University of Barcelona
David Beck	University of Alberta, Edmonton
Xavier Blanco	Autonomous University of Barcelona
Bernd Bohnet	Stuttgart University
Igor Boguslavsky	Polytechnical University of Madrid
Marie Candito	University Paris 7
Éric de la Clergerie	University Paris 7
Michael Collins	Columbia University, New York
Benoit Crabbé	University Paris 7
Denys Duchier	University of Orléans
Jason Eisner	Johns Hopkins University, Baltimore
Dina El Kassas	Miniya University
Gülşen Cebiroğlu Eryiğit	Istanbul Technical University
Charles J. Fillmore	University of California, Berkeley
Koldo Gojenola	University of the Basque Country, Bilbao
Jan Hajič	Charles University in Prague



Hans-Jürgen Heringer	University of Augsburg
Richard Hudson	University College London
Leonid Iomdin	Russian Academy of Sciences, Moscow
Lidija Iordanskaja	University of Montreal
Aravind Joshi	University of Pennsylvania, Philadelphia
Sylvain Kahane	University Paris Ouest
Marco Kuhlmann	Uppsala University
François Lareau	Macquarie University, Sydney
Alessandro Lenci	University of Pisa
Leonardo Lesmo	University of Turin
Haitao Liu	Zhejiang University, Hangzhou
Henning Lobin	University of Gießen
Chris Manning	Stanford University
Igor Mel'čuk	University of Montreal
Wolfgang Menzel	University of Hamburg
Kemal Oflazer	Carnegie Mellon University, Qatar
Ryan McDonald	Google Research, New York
Piet Mertens	University of Leuven
Jasmina Milićević	Dalhousie University, Halifax
Dipti Misra Sharma	IIT, Hyderabad
Henrik Høeg Muller	Copenhagen Business School
Jee-Sun Nam	Hankuk University of Foreign Studies, Seoul
Alexis Nasr	University of Marseille
Joakim Nivre	Uppsala University
Gertjan van Noord	University of Groningen
Martha Palmer	University of Colorado, Boulder
Jarmila Panevova	Charles University in Prague
Alain Polguère	Nancy University
Prokopis Prokopidis	ILSP, Athens
Owen Rambow	Columbia University, New York
Ines Rehbein	Saarland University, Saarbrücken
Petr Sgall	Charles University in Prague
Davy Temperley	University of Rochester
Robert Van Valin	Heinrich Heine University, Düsseldorf

Many thanks also to Joana Clotet and Bea Abad for taking care of practically all matters related to the local organization of the conference, to Simon Mille for assisting them and to all the other members of the local organization team: Stefan Bott, Alicia Burga, Gerard Casamayor, Gaby Ferraro, Estela Mosquera, Luz Rello, Orsi Vincze, and Alexandra Vorobyova. Financial support for Depling was provided by the Natural Language Processing research group TALN of the Pompeu Fabra University (UPF), the Department of Communication and Information Technologies, UPF, and the Department of French and Romance Philology at the Autonomous University of Barcelona.

Kim Gerdes, Eva Hajičová, and Leo Wanner  
September 2011

