

# **Supertagging: Using Complex Lexical Descriptions in Natural Language Processing**

**Edited by Srinivas Bangalore and Aravind K. Joshi**

**A Bradford Book  
The MIT Press  
Cambridge, Massachusetts  
London, England**

© 2010 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

For information about special quantity discounts, please email  
special\_sales@mitpress.mit.edu

This book was set in LaTeX by the authors.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Supertagging : using complex lexical descriptions in natural language processing  
/ edited by Srinivas Bangalore and Aravind K. Joshi.

p. cm. — (A Bradford Book)

Includes bibliographical references and index.

ISBN 978-0-262-01387-1 (hardcover : alk. paper) 1. Natural language processing  
(Computer science) 2. Computational linguistics—Statistical methods. 3.  
Semantics. 4. Lexicology. I.

Bangalore, Srinivas, 1969– II. Joshi, Aravind K. (Aravind Krishna), 1929–

QA76.9.N38S86 2010

006.3'5—dc22

2009028194

10 9 8 7 6 5 4 3 2 1

---

## Preface

The conventional (mathematical) wisdom in specifying a grammar formalism is to start with basic primitive structures as simple as possible and then introduce various operations for constructing more complex structures. This has the consequence that information (e.g., syntactic and semantic) about a lexical item (word) is distributed over more than one primitive structure. Therefore, the information associated with a lexical item is not captured locally, that is within the domain of a primitive structure. An alternate approach is to start with complex (more complicated) primitives for a word, which capture directly some crucial linguistic properties of the word and then introduce some general operations for composing these complex structures (primitive or derived). All the pieces of information associated with a lexical item have to be represented in the local domains of the primitive structures. In this approach, non-local dependencies are pushed to become local that is, these dependencies start out in the basic primitive structure. We refer to the primitive structure that is associated with a word as a *supertag* and characterize this approach as *complicate locally, simplify globally* (CLSG).

During this past decade, the CLSG approach has been explored in different grammar formalisms and the consequences of such a localization for computational, linguistic and psycholinguistic models of natural language processing has been extensively studied. This book is a collection of some of the research that investigates this theme. We expect that this book will be of special interest to computational linguists and researchers in speech and language processing for its perspective on the representation and its implications on computation of linguistic structure. For the machine learning community interested in language applications, we expect this book to provide an opportunity for exploring novel machine learning techniques that can exploit the richer feature space provided by supertag representations. The close coupling of lexical and syntactic information afforded by the supertag representation and its impact on language processing would be of interest to researchers of psycholinguistics interested in human sentence processing.

The book is broadly organized into five parts. The first part highlights issues concerning the creation and organization of supertags in Lexicalized Tree-adjoining Grammar (LTAG) framework. Research concerning the models for

supertag disambiguation and their relation to parsing in the context of LTAG supertags are included in the second part of the book. The third part presents different instantiations of the notion of supertags in a range of grammar formalisms. Research work on some of the linguistic and psycholinguistic issues related to supertags constitute the fourth part. And finally, some of the speech and language applications that exploit supertags are highlighted in the last part of the book.

We are most grateful to the members of the XTAG group at University of Pennsylvania who helped develop the idea of supertags in LTAG formalism and for the effort in building a wide-coverage grammar of English that has served as the basis for the initial experiments. We thank all the chapter authors who kindly consented to contribute their research to the theme of this book. We would like to thank AT&T Labs-Research for continued support in this research agenda and during the preparation of the book.