

# The MIT Press

## Journals

[Sign In / Register](#)
[Books](#)
[Journals](#)
[Digital](#)
[Resources](#)
[About](#)
[Contact](#)


Home | Computational Linguistics | List Article navigation  
of Issues | Volume 33 , No. 4 |  
Unsupervised Acquisition of Predominant  
Word Senses



# Unsupervised Acquisition of Predominant Word Senses

[Diana McCarthy](#), [Rob Koeling](#), [Julie Weeds](#) and  
[John Carroll](#)

Posted Online December 03, 2007

<https://doi.org/10.1162/coli.2007.33.4.553>

© 2007 Massachusetts Institute of Technology

Computational Linguistics  
Volume 33 | Issue 4 | December 2007  
p.553-590

 **Download Options** >

## Abstract Authors

There has been a great deal of recent research into word sense disambiguation, particularly since the inception of the Senseval evaluation exercises. Because a word often has more than one meaning, resolving word sense ambiguity could benefit applications that need some level of semantic interpretation of language input. A major problem is that the accuracy of word sense disambiguation systems is strongly

Quarterly (March,  
June, September,  
December)

160pp. per issue

6 3/4 x 10

Founded: 1974

2018 Impact

Factor: 1.319

2018 Google

Scholar h5-index:  
32

ISSN: 0891-2017

E-ISSN: 1530-9312

## Journal

## Resources

[Editorial Info](#)

[Abstracting and  
Indexing](#)

[Release Schedule](#)

[Advertising Info](#)

## Author Resources

Submission Guidelines  
Publication Agreement  
Author Reprints

## Reader Resources

Rights and Permissions  
Most Read  
Most Cited

More About Computational Linguistics 

Metrics 

Open Access 



Computational Linguistics is Open Access. All content is freely available in electronic format (Full text HTML, PDF, and PDF Plus) to readers across the globe. All articles are published under a [CC BY-NC-ND 4.0 license](#).

dependent on the quantity of manually sense-tagged data available, and even the best systems, when tagging every word token in a document, perform little better than a simple heuristic that guesses the first, or predominant, sense of a word in all contexts. The success of this heuristic is due to the skewed nature of word sense distributions. Data for the heuristic can come from either dictionaries or a sample of sense-tagged data. However, there is a limited supply of the latter, and the sense distributions and predominant sense of a word can depend on the domain or source of a document. (The first sense of “star” for example would be different in the popular press and scientific journals). In this article, we expand on a previously proposed method for determining the predominant sense of a word automatically from raw text. We look at a number of different data sources and parameterizations of the method, using evaluation results and error analyses to identify where the method performs well and also where it does not. In particular, we find that the method does not work as well for verbs and adverbs as nouns and adjectives, but produces more accurate predominant sense information than the widely used SemCor corpus for nouns with low coverage in that corpus. We further show that the method is able to adapt successfully to domains when using domain specific corpora as input and where the input can either be hand-labeled for domain or automatically classified.

## Forthcoming

### Most Read

[See More](#)

 **Lexicon-Based Methods for Sentiment Analysis** (14057 times)  
Maite Taboada et al.  
Computational Linguistics  
Volume: 37, Issue: 2, pp. 267-307

 **Computational Linguistics and Deep Learning** (10535 times)  
Christopher D. Manning  
Computational Linguistics  
Volume: 41, Issue: 4, pp. 701-707

 **Near-Synonymy and Lexical Choice** (3670 times)  
Philip Edmonds et al.  
Computational Linguistics  
Volume: 28, Issue: 2, pp. 105-144

(Note that the Most Read numbers are based on the number of full text downloads over the last 12 months.)

For more information on allowed uses, please view the CC license.

[Support OA at MITP](#)

### Most Cited

[See More](#)

 **Lexicon-Based Methods for Sentiment Analysis** (436 times)  
Maite Taboada et al.  
Computational Linguistics  
Volume: 37, Issue: 2, pp. 267-307

 **A Systematic Comparison of Various Statistical Alignment Models** (174 times)  
Franz Josef Och et al.  
Computational Linguistics  
Volume: 29, Issue: 1, pp. 19-51

 **Opinion Word Expansion and Target Extraction through Double Propagation** (147 times)  
Guang Qiu et al.  
Computational Linguistics  
Volume: 37, Issue: 1, pp. 9-27

(Note that the Most Cited numbers are based on Crossref's [Cited-by service](#) and reflect citation information for the past 24 months. )

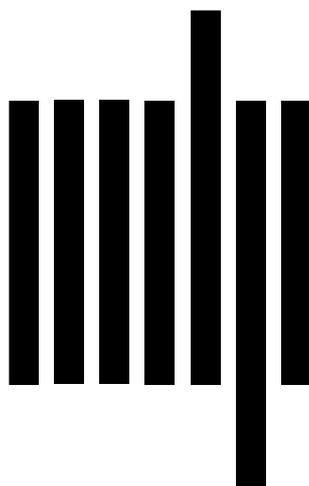
### Download Options >

Favorite  Sign up for Alerts 

Download Citation  RSS TOC 

RSS Citation  Submit your article

[Support OA at MITP](#) 



[Journals](#)

[Terms & Conditions](#)

[Privacy Statement](#)

[Contact Us](#)

[Books](#)

[US](#)

[UK](#)

[Connect](#)

One Rogers Street  
Cambridge MA  
02142-1209

Suite 2, 1 Duchess  
Street London,  
W1W 6AN, UK



© 2018 The MIT Press  
Technology Partner:  
[Atypon Systems, Inc.](#)  
[CrossRef Member](#)

[COUNTER Member](#)  
The MIT Press colophon is registered in the

