

The MIT Press

Journals

Sign In / Register

Books

Journals

Digital

Resources

About

Contact



Home | Computational Linguistics | List Article navigation  
of Issues | Volume 28 , No. 2 | Empirical  
Methods for Exploiting Parallel Texts I. Dan  
Melamed (New York University)  
Cambridge, MA: MIT Press, 2001, xi+195  
pp; hardbound, ISBN 0-262-13380-6,  
\$32.95, £22.95



# Empirical Methods for Exploiting Parallel Texts I. Dan Melamed (New York University) Cambridge, MA: MIT Press, 2001, xi+195 pp; hardbound, ISBN 0-262-13380-6, \$32.95, £22.95

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

Ted Pedersen

**Journal  
Resources**

Editorial Info

Abstracting and  
Indexing

Posted Online March 30, 2006

<https://doi.org/10.1162/coli.2002.28.2.235>

© 2002 Association for Computational Linguistics

Computational Linguistics  
Volume 28 | Issue 2 | June 2002  
p.235-237

Release Schedule  
Advertising Info


 **Download Options** >

## 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  
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

## First Page Authors

### Empirical Methods for Exploiting Parallel Texts

**I. Dan Melamed**  
(New York University)

Cambridge, MA: MIT Press, 2001,  
xi+195 pp; hardbound, ISBN  
0-262-13380-6, \$32.95, £22.95

*Reviewed by*  
*Ted Pedersen*  
*University of Minnesota, Duluth*

Parallel translations of written texts have long been useful tools for the study of language and have begun to serve as an intriguing source of data for a variety of approaches to natural language processing. A source text and its translation are viewed as a coarse map between the two languages, and an intelligent and clever computer program may wish to refine that mapping so that individual sentences, phrases, and words are translations of one another.

Humans are very adept at finding such relations in parallel texts, even when one or both of the languages is unfamiliar, as can be seen in the following convincing exercise by Knight (1997). Although there was considerable interest in automatically identifying sentences in parallel text that are translations (e.g., Brown, Lai, and Mercer 1991, Gale and Church 1993), a variety of problems has emerged since that time.

*Empirical Methods for Exploiting Parallel Texts* is a revision of author's 1998 Ph.D. dissertation (University of Pennsylvania) and succeeds in addressing a range of problems inherent in parallel text. It presents a variety of interesting translation equivalents and demonstrates that once these are identified, they can be used to align text segments, detect omissions in translations, identify idiomatic compounds, and discriminate among word senses.

The book is arranged in three parts, the guiding organizational principle being the distinction between tokens and types. (A token represents an instance of a linguistic entity in a text, whereas a type consists of every occurrence of a particular entity.) The author casts his own work in terms of pattern recognition and machine learning, acquiring information about specific tokens and statistical learning on generalized models of word types on the basis of token data.

Part I consists of three chapters that focus on tokens and types. Chapter 2 presents an algorithm that finds a token-by-token mapping between two texts in which each token in the source text is aligned with its translation in the target text. This algorithm is presented in terms of pattern recognition concepts: generation, filtering, and search.




under a [CC BY-NC-ND 4.0 license](#). For more information on allowed uses, please view the [CC license](#). [Support OA at MITP](#)

The signals generated are candidate token alignments, and the nates identified in the parallel text and optionally from a user-su of word translations. High-frequency words create noise that is fi a localized search procedure that allows the algorithm to consid piece by piece as it performs token alignment. The next two chap that once such a mapping is available, it can be used to find seg parallel text and to detect portions of text that have been erroneo translation.

Forthcoming

Most Read




See More

 <b>Lexicon-Based Methods for Sentiment Analysis</b> (14087 times) Maite Taboada et al. Computational Linguistics Volume: 37, Issue: 2, pp. 267-307	 <b>Computational Linguistics and Deep Learning</b> (10542 times) Christopher D. Manning Computational Linguistics Volume: 41, Issue: 4, pp. 701-707	 <b>Near-Synonymy and Lexical Choice</b> (3675 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.)



Most Cited

See More


 <b>Lexicon-Based Methods for Sentiment Analysis</b> (436 times) Maite Taboada et al. Computational Linguistics Volume: 37, Issue: 2, pp. 267-307	 <b>A Systematic Comparison of Various Statistical Alignment Models</b> (174 times) Franz Josef Och et al. Computational Linguistics Volume: 29, Issue: 1, pp. 19-51	 <b>Opinion Word Expansion and Target Extraction through Double Propagation</b> (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. )

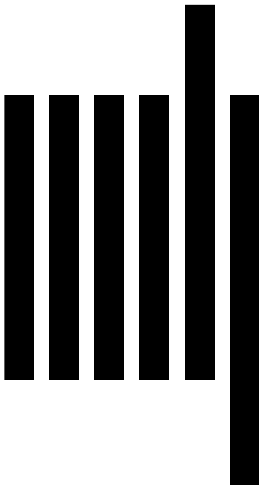
 **Download** >  
**Options**  
Sign up for

Favorite  Alerts 

Download  
Citation  RSS TOC 

RSS Citation  Submit your  
article

Support OA  
at MITP 



Journals

Books

US

One Rogers Street  
Cambridge MA  
02142-1209

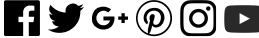
Terms &  
Conditions

UK

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

Privacy  
Statement

Connect



Contact  
Us

© 2018 The MIT  
Press  
Technology Partner:  
[Atypon Systems,  
Inc.](#)  
[CrossRef Member](#)  
[COUNTER Member](#)  
The MIT Press  
colophon is  
registered in the  
U.S. Patent and  
Trademark Office.  
[Site Help](#)