

# The MIT Press

## Journals

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


Home | Computational Linguistics | List Article navigation  
of Issues | Volume 35 , No. 1 |  
Unsupervised Type and Token  
Identification of Idiomatic Expressions



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

# Unsupervised Type and Token Identification of Idiomatic Expressions

Afsaneh Fazly, Paul Cook  
and Suzanne Stevenson

Posted Online March 12, 2009

<https://doi.org/10.1162/coli.08-010-R1-07-048>

© 2009 Association for Computational Linguistics

Computational Linguistics  
Volume 35 | Issue 1 | March 2009  
p.61-103

 **Download Options** >

## Abstract Authors

Idiomatic expressions are plentiful in everyday language, yet they remain mysterious, as it is not clear exactly how people learn and understand them. They are of special interest to linguists, psycholinguists, and lexicographers, mainly because of their syntactic and semantic idiosyncrasies as well as their unclear lexical status. Despite a great deal of research on the properties of idioms in the linguistics literature,

## Author Reprints

## Reader Resources

Rights and Permissions  
Most Read  
Most Cited

More About Computational Linguistics ▼

Metrics ▼



22 Total

citations

10 Recent

citations

6.1 Field Citation

Ratio

n/a Relative

Citation Ratio

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


there is not much agreement on which properties are characteristic of these expressions. Because of their peculiarities, idiomatic expressions have mostly been overlooked by researchers in computational linguistics. In this article, we look into the usefulness of some of the identified linguistic properties of idioms for their automatic recognition. Specifically, we develop statistical measures that each model a specific property of idiomatic expressions by looking at their actual usage patterns in text. We use these statistical measures in a type-based classification task where we automatically separate idiomatic expressions (expressions with a possible idiomatic interpretation) from similar-on-the-surface literal phrases (for which no idiomatic interpretation is possible). In addition, we use some of the measures in a token identification task where we distinguish idiomatic and literal usages of potentially idiomatic expressions in context.


## 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.)

### Most Cited

[See More](#)

globe. All articles are published under a [CC BY-NC-ND 4.0 license](#). For more information on allowed uses, please view the [CC license](#).

[Support OA at MITP](#)



[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. )

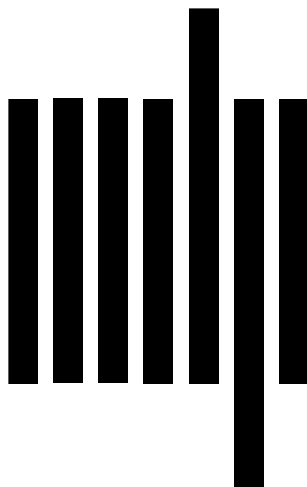
### Download > Options

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](#)

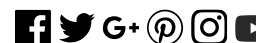
[US](#)

[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

U.S. Patent and Trademark Office

