

The background is a low-angle, teal-tinted photograph of a city street. Tall buildings line both sides of the street, creating a sense of depth. A street lamp is visible in the center foreground. Large, white, stylized brackets are positioned on either side of the title text.

Introduction to NLP



Goals

1. To appreciate the conceptual foundations of NLP
2. To appreciate the technical foundations of NLP



Conceptual Foundations

Where does NLP come from?

Linguistics

- Linguistics offers the notions and theories that guide us through natural language analysis.
- Without linguistic notions and theories, we would get lost in the data 'easily.'

Computational linguistics

- Computational linguistics uses math to mobilize the notions and theories linguistics brings about
- In other words, computational linguistics bridges the disconnect between natural language (i.e., words) and formal language (i.e., math)

Linguistics

Foci

- Pragmatics
 - The study of how 'context' contributes to words and sentences' meaning
- Phonetics
 - The study of the physical properties of speeches
- Semantics
 - The study of meanings
- Syntax
 - The study of grammatical relations and constituency

WORD

JOURNAL OF THE INTERNATIONAL LINGUISTIC ASSOCIATION

[Home](#) [About WORD](#) [Committees](#) [Submissions](#) [News & Events](#) [Past Issues](#) [Subscription & Permission](#) [Contact us](#)

Volumes 61 -

Volumes 1 - 60

Volumes 61 -

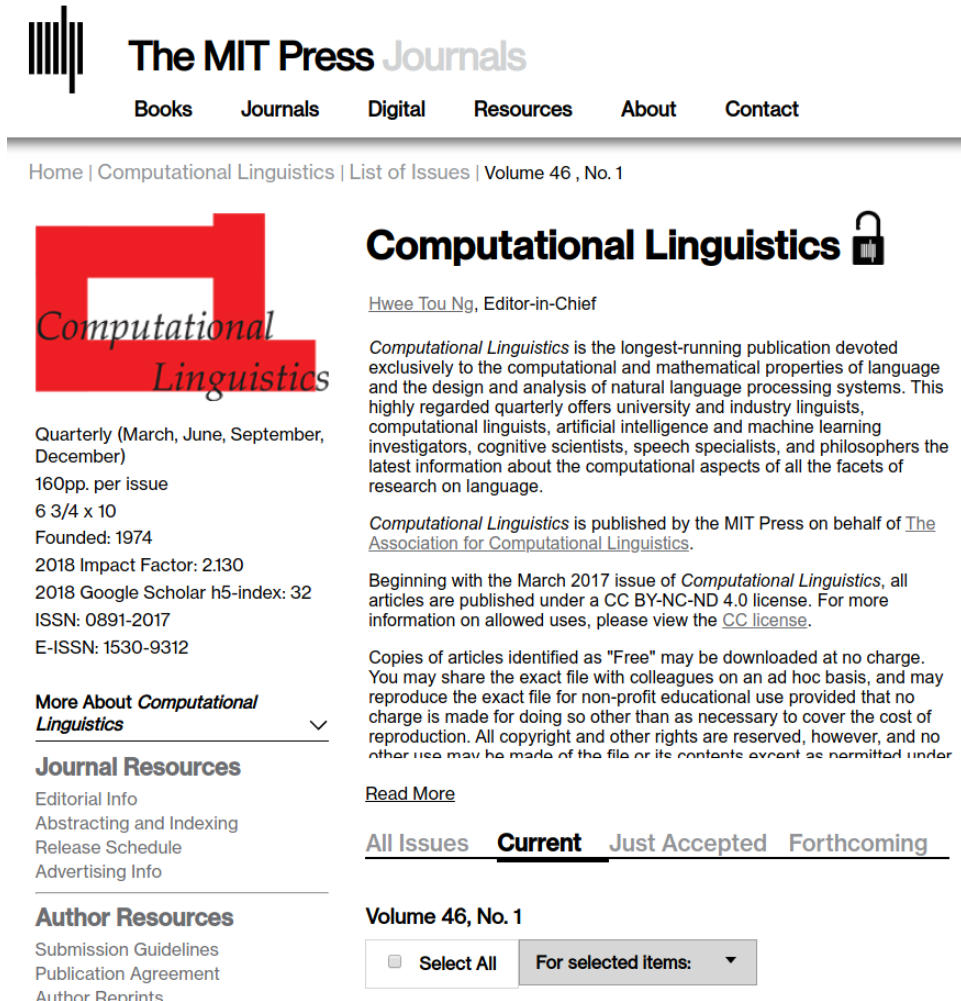
Volume 61 (2015)

| Issue | Type | Item | Author(s) / Reviewer(s) | Page No. |
|--------|---------|---|------------------------------|----------|
| 61.1 | Article | The effect of Italo-Romance contact on the Greek cluster vy in Corsica and the implications for sound change in Italiot Greek in Southern Italy | Nick Nicholas and John Hajek | 2 |
| | | Aspects of cultural intelligence in idiomatic Asian cultural scripts | Jyh Wee Sew | 12 |
| | | PRO theory | Ala Al-Kajela | 25 |
| | | Participle clauses between adverbial and complement | Hendrik De Smet | 39 |
| | | A framework for studying languages in contact: a prolegomenon to a theory | Eric S. Wheeler | 75 |
| Review | | Eugen Helimsky, Ulrike Kahrs and Monika Schötschel, eds: Mari und Mordwinen im heutigen Rußland: Sprache, Kultur, Identität | Alan R. Libert | 87 |
| | | Alexandra Y. Aikhenvald: Language contact in Amazonia | Edward J. Vajda | 89 |

Computational linguistics

Focus

- Designing algorithms for the analysis of pragmatics, phonetics, semantics, syntax



The screenshot displays the website for 'The MIT Press Journals', specifically the page for 'Computational Linguistics'. The header includes the MIT Press logo and navigation links for Books, Journals, Digital, Resources, About, and Contact. Below the header, a breadcrumb trail reads 'Home | Computational Linguistics | List of Issues | Volume 46, No. 1'. The main content area is divided into two columns. The left column features a large red graphic with the text 'Computational Linguistics' and a list of publication details: Quarterly (March, June, September, December), 160pp. per issue, 6 3/4 x 10, Founded: 1974, 2018 Impact Factor: 2.130, 2018 Google Scholar h5-index: 32, ISSN: 0891-2017, and E-ISSN: 1530-9312. Below this is a section titled 'More About Computational Linguistics' with a dropdown arrow, followed by 'Journal Resources' (Editorial Info, Abstracting and Indexing, Release Schedule, Advertising Info) and 'Author Resources' (Submission Guidelines, Publication Agreement, Author Reprints). The right column features the journal title 'Computational Linguistics' with a lock icon, the Editor-in-Chief 'Hwee Tou Ng', and a description of the journal's focus on computational and mathematical properties of language. It also mentions the journal is published by the MIT Press on behalf of the Association for Computational Linguistics. A section titled 'Beginning with the March 2017 issue of Computational Linguistics, all articles are published under a CC BY-NC-ND 4.0 license. For more information on allowed uses, please view the CC license.' is followed by a statement about the availability of 'Free' articles for download and sharing. At the bottom, there is a 'Read More' link, a navigation bar with 'All Issues', 'Current' (highlighted), 'Just Accepted', and 'Forthcoming', and a section for 'Volume 46, No. 1' with a 'Select All' button and a 'For selected items:' dropdown menu.

The MIT Press Journals

Books Journals Digital Resources About Contact

Home | Computational Linguistics | List of Issues | Volume 46, No. 1

Computational Linguistics

Hwee Tou Ng, Editor-in-Chief

Computational Linguistics is the longest-running publication devoted exclusively to the computational and mathematical properties of language and the design and analysis of natural language processing systems. This highly regarded quarterly offers university and industry linguists, computational linguists, artificial intelligence and machine learning investigators, cognitive scientists, speech specialists, and philosophers the latest information about the computational aspects of all the facets of research on language.

Computational Linguistics is published by the MIT Press on behalf of [The Association for Computational Linguistics](#).

Beginning with the March 2017 issue of *Computational Linguistics*, all articles are published under a CC BY-NC-ND 4.0 license. For more information on allowed uses, please view the [CC license](#).

Copies of articles identified as "Free" may be downloaded at no charge. You may share the exact file with colleagues on an ad hoc basis, and may reproduce the exact file for non-profit educational use provided that no charge is made for doing so other than as necessary to cover the cost of reproduction. All copyright and other rights are reserved, however, and no other use may be made of the file or its contents except as permitted under

[Read More](#)

[All Issues](#) **[Current](#)** [Just Accepted](#) [Forthcoming](#)

Volume 46, No. 1

For selected items:



Technical Foundations

Where does NLP come from?

Statistical analysis of natural language

- The use of statistics has greatly facilitated the application of computational linguistics algorithms to large text corpora

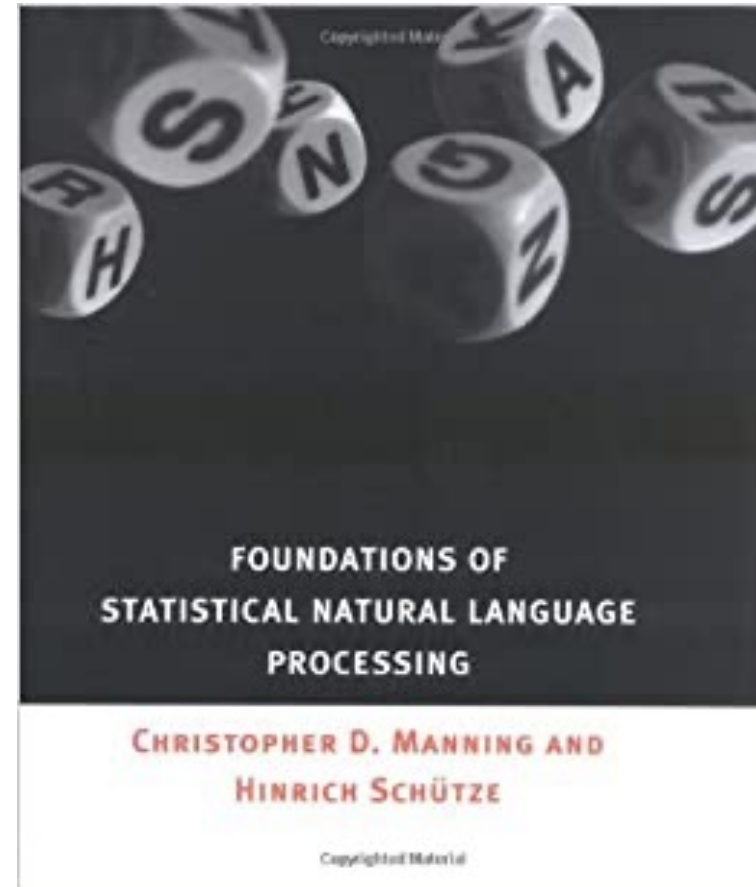
ML and DL with natural language

- ML facilitates NLP tasks such as text classification
- ML can also be used to post-process the outcome of NLP pipelines
- Thanks to DL, now we have granular representation of a word meanings based on humungous text corpora

Statistical analysis of natural language

Examples of foci

- Identification of n-grams that is, the combination of words occurring next to each other and taking specific meanings (e.g., decision-making)
- Application of Markov Chains to Part-of-Speech tagging



ML and DL with natural language

ML

- ML plays a central role in NLP
- For example, SVM is a popular tool in text classification
- Manifold learning techniques are useful to explore and visualize the semantic similarity between text corpora

DL

- DL is the engine that powers embedding models
- Without DL, we would not be able to capture and to represent the semantic attributes of words, sentences, and broader text corpora

Efficient Estimation of Word Representations in Vector Space

Tomas Mikolov

Google Inc., Mountain View, CA
tmikolov@google.com

Greg Corrado

Google Inc., Mountain View, CA
gcorrado@google.com

Kai Chen

Google Inc., Mountain View, CA
kaichen@google.com

Jeffrey Dean

Google Inc., Mountain View, CA
jeff@google.com

Abstract

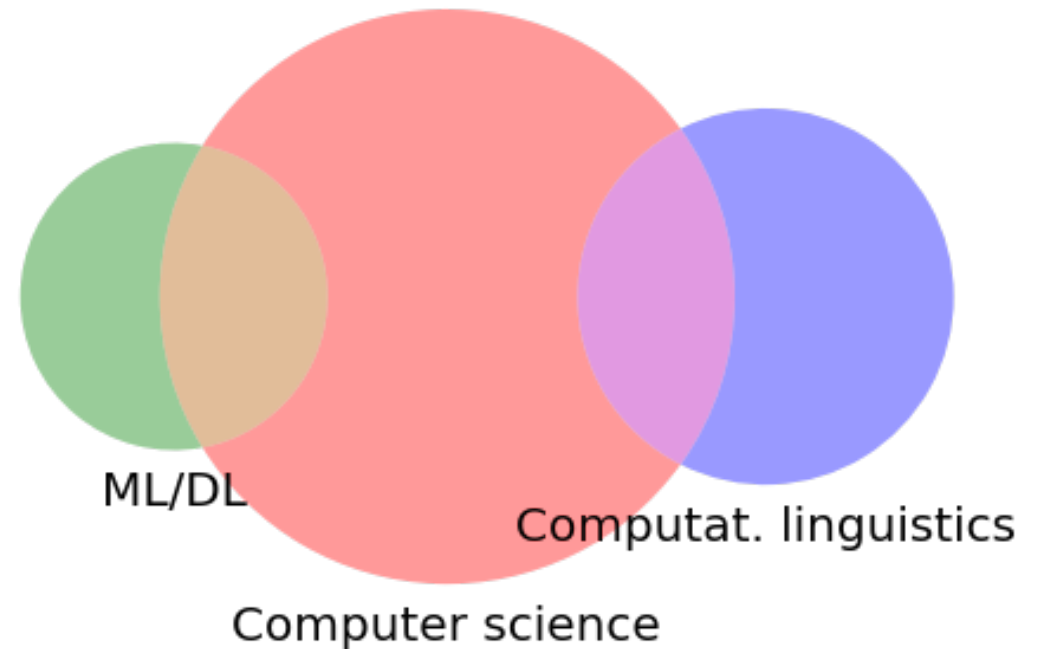
We propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, i.e. it takes less than a day to learn high quality word vectors from a 1.6 billion words data set. Furthermore, we show that these vectors provide state-of-the-art performance on our test set for measuring syntactic and semantic word similarities.

**Mikolov et al's
paper draws a line
between the past
and the present of
NLP**

The past of NLP (pre-2013)

Before Mikolov et al's paper

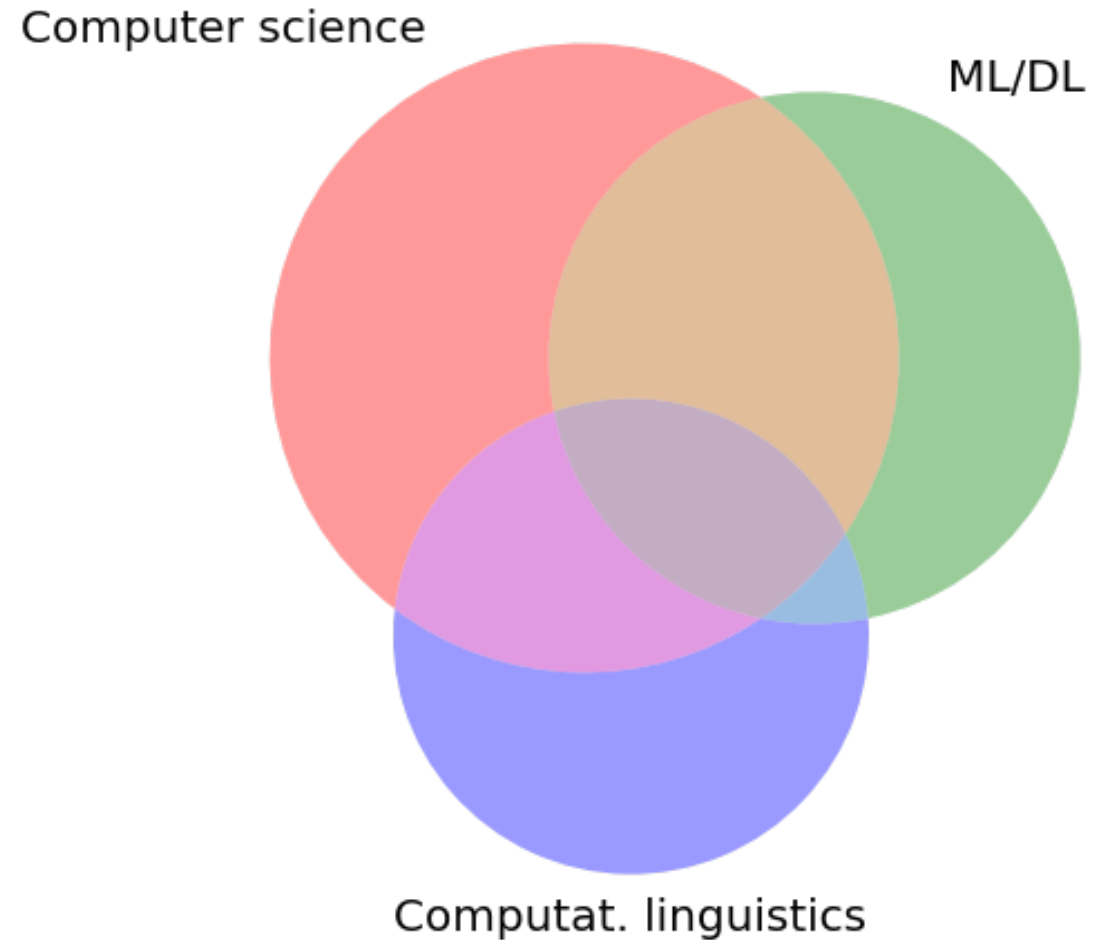
- Some computational linguists rely on computer science tools to manipulate large amounts of textual data
- The fields of ML/DL and computational linguistics are disconnected



The present of NLP (after 2013)

After Mikolov et al's paper

- NLP emerges at the intersection of computational linguistics, computer science and ML/DL
- Researching and using NLP requires a genuine cross-disciplinary approach



Wrap-up

Main points

NLP draws on multiple bodies of knowledge and disciplines.

Traditionally, linguistics has provided NLP with key conceptual foundations to NLP, while statistics and computer science have facilitated the application of linguistics' notions and frameworks.

Now, NLP emerges at the intersection of linguistics, computer science, and ML/DL.

Given the cross-disciplinary nature of NLP, developers and users may want to approach the subject with a curious and open-minded approach.