

# MARK MYSLÍN – DATA SCIENTIST

805.448.7690 · [mmyslín@ucsd.edu](mailto:mmyslín@ucsd.edu)

## WORK EXPERIENCE

---

### Senior data scientist

San Francisco, CA

ERNST & YOUNG, LLP

2014-

Design and build new technologies for linguistic and unstructured data for a wide range of clients and use cases: natural language processing, machine learning, speech recognition, search optimization, information retrieval and classification. Manage projects and teams spanning legal, financial, advisory, and process improvement sectors.

### Doctoral researcher

San Diego, CA

COMPUTATIONAL PSYCHOLINGUISTICS LABORATORY, UC SAN DIEGO

2010-

Build computational models of natural language production and comprehension. Analysis of large datasets using R and Python (NLP, machine learning, Bayesian statistics). Conduct behavioral research including experimental design in the lab and online (HTML, PHP, JavaScript, Amazon Mechanical Turk crowdsourcing platform, eyetracking).

## EDUCATION

---

### UC San Diego, Departments of Linguistics & Cognitive Science

San Diego, CA

Ph.D. in LINGUISTICS and COGNITIVE SCIENCE (Interdisciplinary Program), *expected* June 2016. Overall GPA: 4.0

**Thesis:** Causal inference in language comprehension

**External grants:** National Science Foundation Research Fellowship (\$120K), Javits Fellowship (\$150K)

**Summer school:** Linguistic Society of America Summer Institute, University of Colorado, Boulder (2011)

### UC Santa Barbara

Santa Barbara, CA

B.A. in LINGUISTICS and SPANISH, 2010. Overall GPA: 3.99

**Thesis:** A computational model of multilingual discourse (*Chancellor's Award for Best Undergraduate Thesis*)

**Relevant coursework:** Natural language processing, machine learning, corpus linguistics, sentiment analysis, computational models of cognition, Bayesian statistics, probability theory, experimental design, discourse analysis

## TEACHING

---

**Course instructor:** Language, Cognition & Computation (designed and taught, College of the Redwoods, Eureka, CA).

**TA:** LING 165 Computational Linguistics, LING 170 Psycholinguistics, LISP 1BX Analysis of Spanish (UC San Diego).

## INDUSTRY PUBLICATIONS

---

**Myslín, Mark**, et al. 2016. Audio analytics: New opportunities in litigation and regulation. **Ernst & Young, LLP**.

## SCIENTIFIC PUBLICATIONS

---

**Myslín, Mark**, & Roger Levy. 2016. Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. **Cognition**.

**Myslín, Mark**, & Roger Levy. 2015. Codeswitching and predictability of meaning in discourse. **Language**.

Seyfarth, Scott & **Mark Myslín**. 2014. Discriminative learning predicts human recognition of English blend sources. **Cognitive Science Proceedings**.

**Myslín, Mark**, et al. 2013. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. **Journal of Medical Internet Research**. *2014 Outstanding Research Article in Biosurveillance (Finalist)*

**Myslín, Mark**, & Stefan Gries. 2010. Corpus study of Spanish Internet orthography. **Literary & Linguistic Computing**.

## CONFERENCE PRESENTATIONS

---

Architectures & Mechanisms for Language Processing, CUNY Sentence Processing Conference, Cognitive Science, Corpus Linguistics, Gradient Symbolic Computation, Linguistic Society of America, Psychonomics, Stanford Psychology of Language.

## CODING EXPERTISE

---

R, ggplot2, lme4, BUGS/JAGS, Python, numpy, scipy, HTML, PHP, CSS, JavaScript, Amazon Mechanical Turk crowdsourcing platform, Weka, TGrep2/tregex, Praat, svn, slurm, LaTeX, SRILM.