# T1: 100 Things You Always Wanted to Know about Linguistics But Were Afraid to Ask*

## *... for fear of being told 1000 more

Emily M. Bender

## ABSTRACT

Many NLP tasks have at their core a subtask of extracting the dependencies---who did what to whom---from natural language sentences. This task can be understood as the inverse of the problem solved in different ways by diverse human languages, namely, how to indicate the relationship between different parts of a sentence. Understanding how languages solve the problem can be extremely useful in both feature design and error analysis in the application of machine learning to NLP. Likewise, understanding cross-linguistic variation can be important for the design of MT systems and other multilingual applications. The purpose of this tutorial is to present in a succinct and accessible fashion information about the structure of human languages that can be useful in creating more linguistically sophisticated, more language independent, and thus more successful NLP systems.

While many kinds of linguistic structure can be relevant to different NLP tasks, the focus of this tutorial will be on morphosyntax. The tutorial will take an explicitly typological perspective as an understanding of cross-linguistic variation can facilitate the design of more portable (language-independent) NLP systems. In order to help participants retain the information better, the tutorial will be structured interactively. I will ask participants for examples of tasks and data sets they work with, and then as a group we will brainstorm ways in which each of the linguistic properties discussed can related to feature design and/or error analysis for those tasks.

## OUTLINE

1. **Introduction:**

   Overview of tutorial, elicitation of examples of tasks and data sets

2. **Morphology:**

   Morphophonology and morphosyntax, the range of morphological processes found in the world's languages, what can be expressed through morphology

3. **Basic Syntax:**

   Part of speech and grammatical functions in cross-linguistic perspective, syntactic phenomena which can obscure the relationship between syntactic and semantic roles

4. **Syntactic Complications:**

   Phenomena beyond simple clauses, including long-distance dependencies, clausal modification, semantically empty function words, argument drop

5. **Resources:**

   Where to go to find out more, what to expect from morphological analyzers, dependency parsers and precision grammars

# BIO

**Emily M. Bender**
Department of Linguistics
University of Washington
Email: ebender@uw.edu

Emily M. Bender is an Associate Professor in the Department of Linguistics and Adjunct Associate Professor in the Department of Computer Science & Engineering at the University of Washington. Her primary research interests lie in multilingual grammar engineering and the incorporation of linguistic knowledge, especially from linguistic typology, in NLP. She is the PI of the Grammar Matrix project, which is developed in the context of the DELPH-IN Consortium (Deep Linguistic Processing with HPSG Initiative). More generally, she is interested in the intersection of linguistics and computational linguistics, from both directions: bringing computational methodologies to linguistic science and linguistic science to natural language processing.

Her PhD (in Linguistics) is from Stanford University. She has authored or co-authored papers in Linguistic Issues in Language Technology, the Journal of Research on Language and Computation, English Language and Linguistics, the Encyclopedia of Language and Linguistics, and the proceedings of ACL, COLING, IJCNLP and associated workshops.