

CSC401
Natural Language Computing

Michael Boyadjian

January 14, 2021

Contents

1	Introduction to Natural Language Computing	3
1.1	Categories of Linguistic Knowledge	3
1.2	NLP as Artificial Intelligence	3
1.3	Overview of NLP	3
2	Corpora and Smoothing	5
3	Features and Classification	5
4	Neural Language Models	5
5	Hidden Markov Models	5
6	Statistical Machine Translation	5
7	Speech	5
8	Automatic Speech Recognition	5
9	Synthesis	5
10	Entropy Decisions	5
11	Intelligent Dialogue Agents	5

1 Introduction to Natural Language Computing

Natural language computing is getting computers to understand everything we say and write. We are interested in learning the statistics of language. Computers give insight into how humans process language, or generate language themselves.

1.1 Categories of Linguistic Knowledge

- **Phonology:** The study of patterns of speech sounds
- **Morphology:** How words can be changed by inflection or derivation
- **Syntax:** The ordering and structure between words and phrases (i.e. *grammar*)
- **Semantics:** The study of how meaning is created by words and phrases
- **Pragmatics:** The study of meaning in contexts

1.2 NLP as Artificial Intelligence

Natural language processing involves resolving ambiguity at all levels. In the early days knowledge was explicitly encoded in artificial symbolic systems by experts. Now, algorithms learn using probabilities (or pseudo-probabilities) to distinguish subtly different competing hypotheses.

- Is **Google** a noun or a verb?
- Examples where $Google \in Nouns$ (“*Google makes Android*”), does not mean that *Google* is never a verb (“*Go Google yourself*”)

$$P(Google \in Nouns) > P(Google \in Verbs) > 0$$

1.3 Overview of NLP

Is natural language processing (the discipline) hard?

- **Yes**, because natural language
 - is highly ambiguous at all levels
 - is complex and subtle
 - is fuzzy and probabilistic
 - involves real-world reasoning
- **No**, because computer science

- gives us many powerful statistical techniques,
- allows us to break the challenges down into more manageable features.

- 2 Corpora and Smoothing
- 3 Features and Classification
- 4 Neural Language Models
- 5 Hidden Markov Models
- 6 Statistical Machine Translation
- 7 Speech
- 8 Automatic Speech Recognition
- 9 Synthesis
- 10 Entropy Decisions
- 11 Intelligent Dialogue Agents