

DATA 340 4

Natural Language Processing

James M. Tucker, PhD
jmtucker02@wm.edu
(757) 525 1572

Lecture 02: Linguistic structure and properties of language

- Summary of problem set 0/questions
- Comments about course documentation - resources
- Linguistic terms and required background for NLP
- Jupyter Notebooks - Virtual Environments
- Some important NLP libraries to know about

Lecture 02: Linguistic structure and properties of language

- Summary of problem set 0/questions
- Comments about course documentation - resources
- Linguistic terms and required background for NLP
- Jupyter Notebooks - Virtual Environments
- Some important NLP libraries to know about

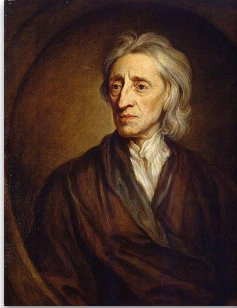
Lecture 02: Linguistic structure and properties of language

- Summary of problem set 0/questions
- Comments about course documentation - resources
- Linguistic terms and required background for NLP
- Jupyter Notebooks - Virtual Environments
- Some important NLP libraries to know about

Lecture 02: Linguistic structure and properties of language

- Summary of problem set 0/questions
- Comments about course documentation - resources
- Linguistic terms and required background for NLP
- Jupyter Notebooks - Virtual Environments
- Some important NLP libraries to know about

Linguistic turn of the 20th century



John Locke



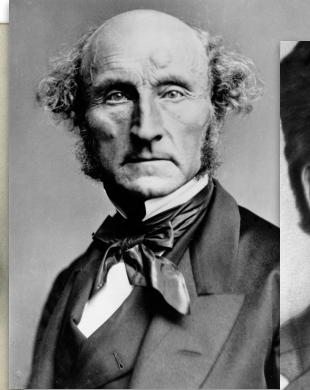
Francis Bacon



Thomas
Hobbes



Gottfried
Leibniz



John Stuart
Mill

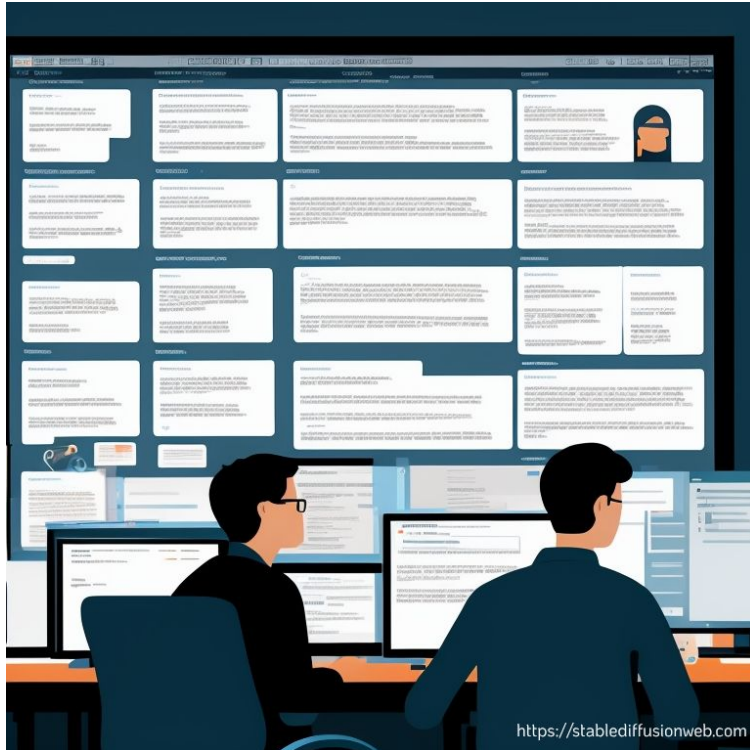


Ludwig
Wittgenstein



Ferdinand de
Saussure

Linguistic turn of the 21st century



Psycholinguistics

Neurolinguistics

Biolinguistics

Sociolinguistics

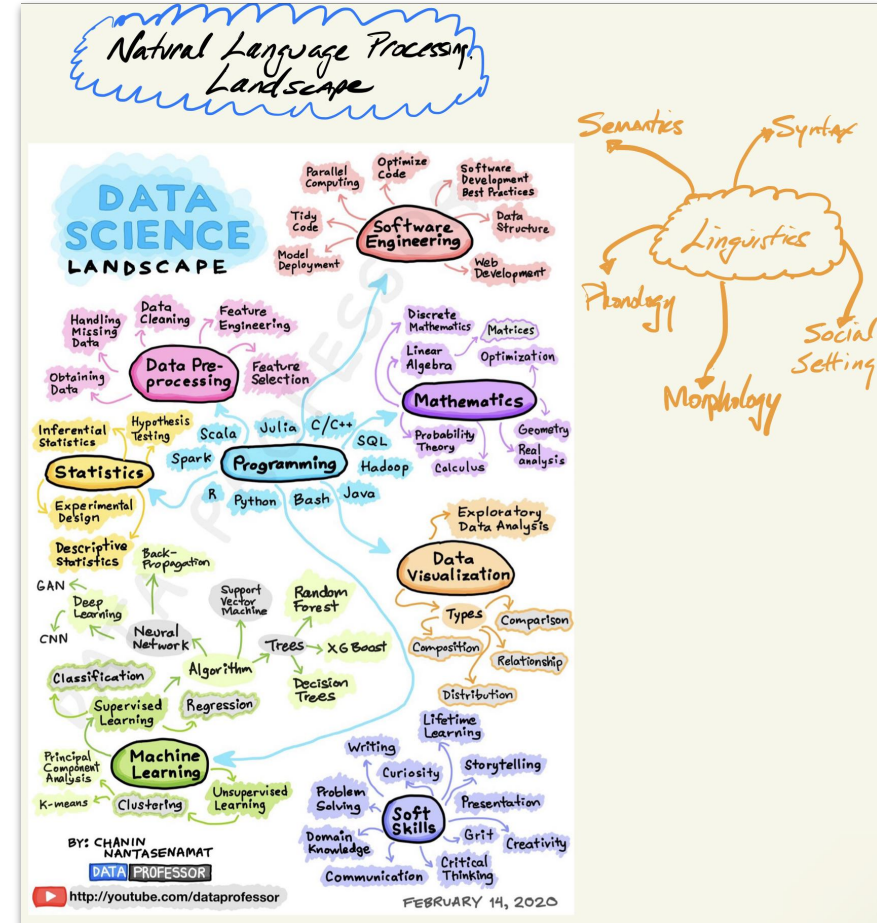
Computational Linguistics

Historical Linguistics

Semiotics

Linguistic essentials

- Phonology
- Morphology
- Syntax
- Semantics
- Social setting - “grounding”

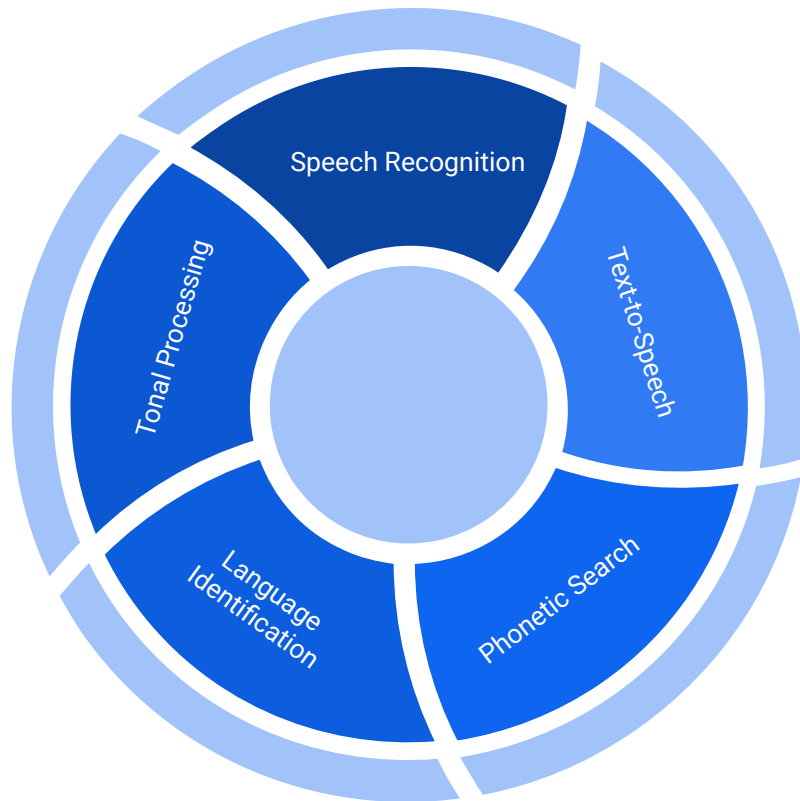




Linguistic essentials: Phonology

- Phonology is a subfield of linguistics that studies the systematic organization of sounds in languages. It focuses on the way sounds function within a particular language or across languages to encode meaning. Unlike phonetics, which is concerned with the physical production and perception of sounds, phonology deals with the abstract, cognitive aspects of sound.
- **Phonemes:** These are the smallest units of sound that can differentiate meaning in a language.
- **Allophones:** These are the different realizations of a phoneme. For example, the "p" sound in "pat" is slightly different from the "p" sound in "spot," but both are allophones of the /p/ phoneme in English.
- **Syllables:** These are units of sound that typically consist of a vowel at their center, with optional initial and final consonants.
- **Prosody:** This refers to the rhythm, stress, and intonation of speech.
- **Phonotactics:** This is the study of the rules governing the possible phoneme combinations in a language.
- **Minimal Pairs:** These are pairs of words that differ in only one phonological element and have distinct meanings. They are used to demonstrate that two sounds contrast in a particular language. For example, "bat" and "pat" are a minimal pair in English.

NLP Use Cases: Phonology

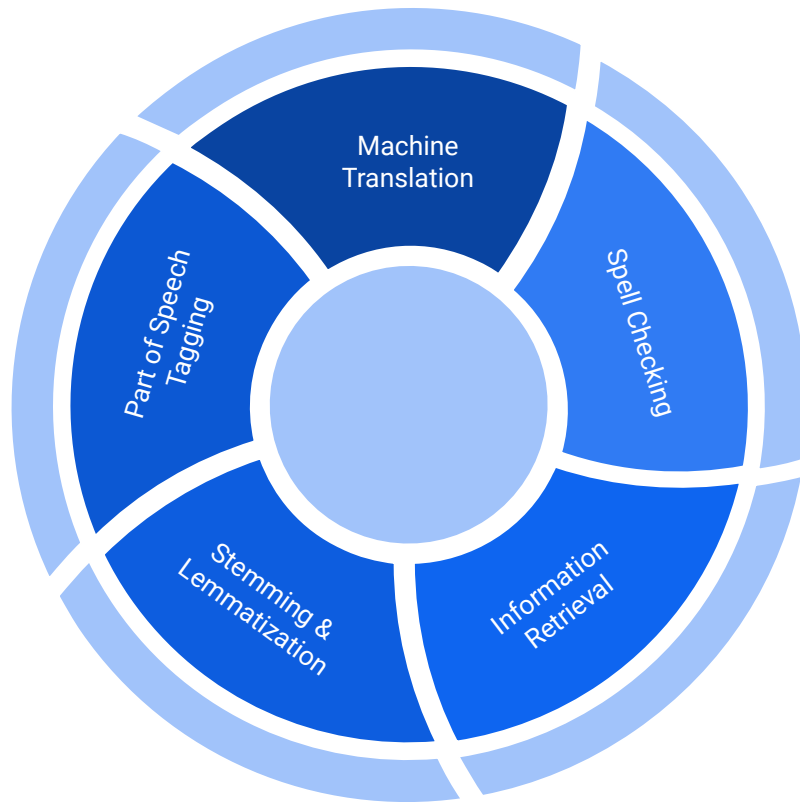




Linguistic essentials: Morphology

- Morphology is the study of the structure and formation of words in a language. It delves into how words are constructed from smaller units, known as morphemes, which are the smallest grammatical units that carry meaning. Morphology bridges the gap between phonology, which deals with sounds, and syntax, which concerns sentence structure.
- **Morphemes:** These are the fundamental units of meaning in a language.
- **Roots, Affixes, Infixes:** A root is the primary lexical unit of a word, which carries the main meaning. Affixes and Infixes are morphemes added to a root to modify its meaning or function.
- **Inflectional vs. Derivational Morphology:** Inflectional morphemes modify a word to fit its role in a sentence without changing its lexical category or core meaning (e.g., "run" -> "running"). Derivational morphemes, on the other hand, often change the core meaning or the lexical category of a word (e.g., "happy" -> "unhappy" or "beauty" -> "beautiful").
- **Compounding:** This is the process of combining two or more words to form a new word, such as "toothbrush" or "mailbox."
- **Morphological Typology:** Languages can be classified based on how they use morphemes.
- **Morphological Analysis:** This involves breaking down a word into its constituent morphemes and understanding their individual and combined meanings.

NLP Use Cases: Morphology

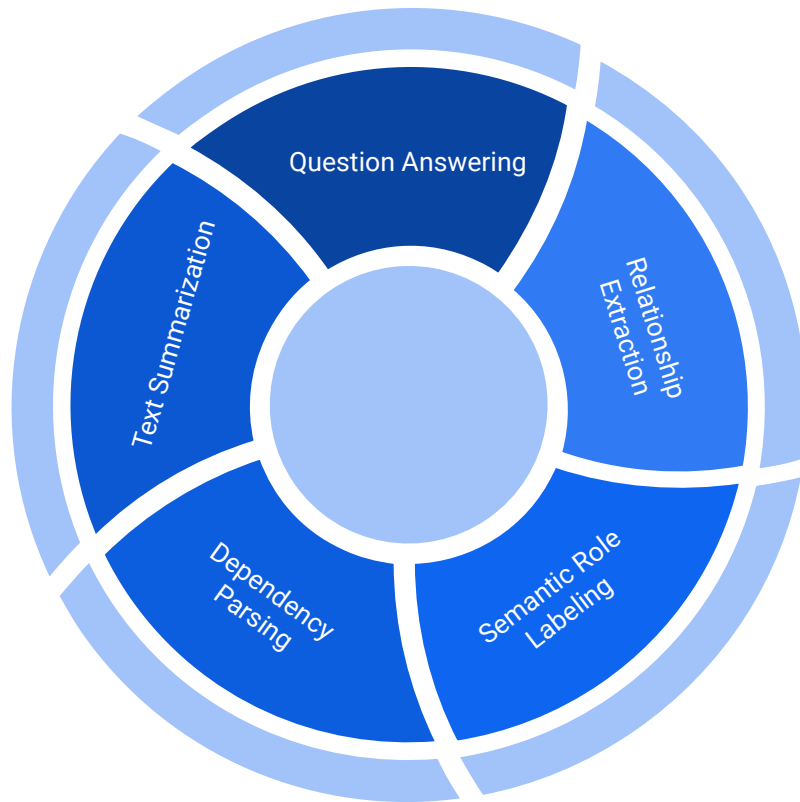




Linguistic essentials: Syntax

- Syntax is a central subfield of linguistics that studies the structure of sentences and the rules governing the arrangement of words within them. It delves into how words and phrases combine to form coherent sentences in a language. Syntax seeks to uncover the underlying principles that dictate the order and combination of words, ensuring meaningful and grammatically correct communication.
- **Constituency:** Sentences can be broken down into constituents, or groups of words that function as a single unit.
- **Phrase Structure:** This refers to the hierarchical organization of constituents in a sentence. noun phrases (NP), verb phrases (VP), and other phrase types.
- **Grammar:** Set of rules that governs the structure of sentences in a particular language.
- **Transformational Rules:** Proposed in transformational-generative grammar, these rules account for the various ways a single underlying structure can surface as different sentence types, such as passive or interrogative sentences.
- **Dependency Relations:** Sentence structure is represented as a network of relations between words, indicating which words depend on others for their function.
- **Case and Role:** In many languages, the grammatical role of a noun (e.g., subject, object) is indicated by its case. Understanding case systems is crucial for parsing sentences in languages like Latin or German.

NLP Use Cases: Syntax





Linguistic essentials: Semantics

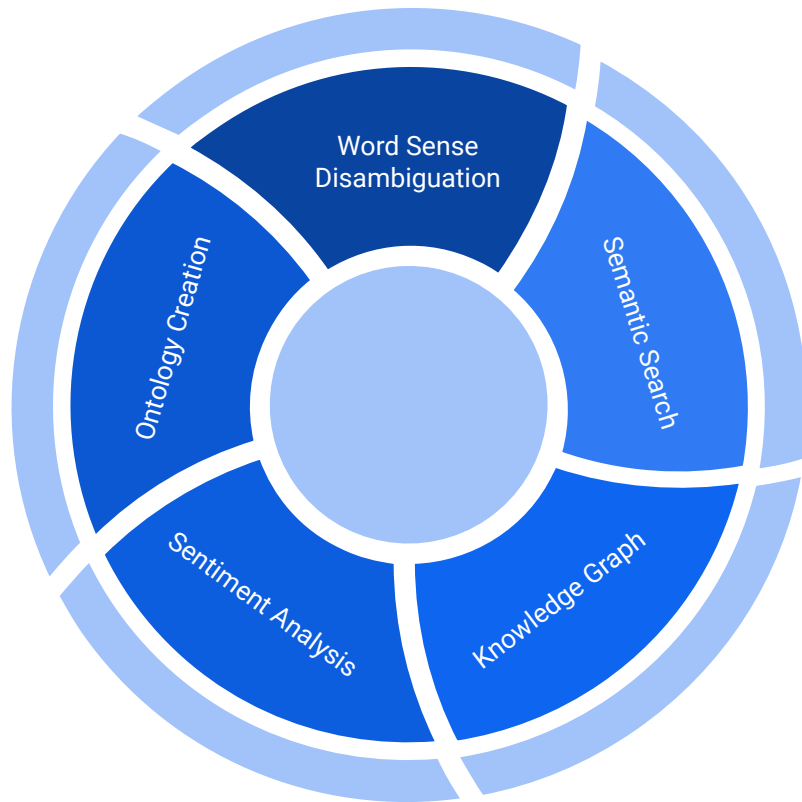
- Semantics is the branch of linguistics that studies the meaning of words, phrases, sentences, and texts. While syntax focuses on the structural arrangement of words and phrases, semantics delves into their inherent meaning and the relationships between them. It seeks to understand how language conveys meaning and how listeners or readers interpret that meaning.
- **Lexical Semantics:** This subfield focuses on the meaning of individual words and the relationships between them. synonymy (words with similar meanings, e.g., "big" and "large"), antonymy (words with opposite meanings, e.g., "hot" and "cold"), and hyponymy (hierarchical relationships, e.g., "sparrow" is a hyponym of "bird").
- **Compositional Semantics:** This area studies how individual word meanings combine to form the meanings of larger syntactic units, such as phrases or sentences.
- **Truth Conditions:** Conditions under which a sentence or statement can be considered true or false in a given context.
- **Semantic Roles:** These are the roles that entities play in relation to the main verb of a sentence, such as agent (the doer of an action), patient (the entity undergoing an action), or instrument (the means by which an action is performed).



Linguistic essentials: Semantics

- Semantics is the branch of linguistics that studies the meaning of words, phrases, sentences, and texts. While syntax focuses on the structural arrangement of words and phrases, semantics delves into their inherent meaning and the relationships between them. It seeks to understand how language conveys meaning and how listeners or readers interpret that meaning.
- **Pragmatics:** Pragmatics studies how context influences the interpretation of meaning. It considers factors like speaker intent, cultural norms, and situational context.
- **Ambiguity:** This occurs when a word, phrase, or sentence has multiple possible meanings. Semantics helps in disambiguating such instances based on context and structure.
- **Reference and Denotation:** Semantics examines how words refer to specific entities in the world. For instance, the word "dog" denotes a particular type of animal.

NLP Use Cases: Semantics





Linguistic essentials: Social Setting / Grounding

- “Grounding is the grammaticized means of relating the thing profiled by a nominal, or the process profiled by a finite clause, to the ground (the speech event and its participants). As narrowly defined, grounding elements make very fundamental specifications of a basically epistemic nature.” Langacker, [Cognitive Grammar](#).

FrameNet Index of Lexical Units

This page is an index to alphabetical lists of the names of the lexical units (LUs).

Each LU name is followed by the part of speech, the name of the relevant frame, and its status. If a HTML files for the lexical entry and the annotated sentences. Lexical units on which work has not and their Syntactic Realizations; and Valence Patterns.

<#> | [A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#) | [All](#) |

Search: testify

- testify.v ([Evidence](#)) **FN1_Sent** [Lexical Entry](#) [Annotation](#)

[FrameNet](#)

Valence Patterns:

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns		
1 TOTAL	Proposition	Proposition	Support
(1)	NP Ext	PP[to] Dep	INI --
19 TOTAL	Proposition	Support	
(1)	AVP Dep	NP Ext	
(10)	PP[to] Dep	NP Ext	
(1)	PP[to] Dep	Sfin Ext	
(2)	Sfin Dep	NP Ext	



Semantic systems



Claude Shannon

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design." *A Mathematical Theory of Communication*, 1.

Lecture 02: Linguistic structure and properties of language

- Summary of problem set 0/questions
- Comments about course documentation - resources
- Linguistic terms and required background for NLP
- Jupyter Notebooks - Virtual Environments
- Some important NLP libraries to know about