Morphosyntactic Measures

Morphophonological Analysis

- **Utterance length**
- ***Phonemes-to-word ratio:*** e.g., Do speakers prefer long or short words?
- ***Content words***: e.g., Nouns, verbs, adjectives, and adverbs.
- ***Function Words***: e.g., Conjunctions, e.g., and, or, and, but Prepositions, e.g., de, in, pre and of, Determines, the and a/an, Pronouns such as he/she/it.
- ***Part of Speech Ratio***: Content to function word ratio nouns to vowels, etc.
- ***Morphological and semantic information about the gender (e.g., male, female, neuter), person (first, second, third), number (e.g., singular, plural), and time (e.g., present, past).***

Syntactic Analysis

- ***probability estimates of syntactic constituents*** (e.g., noun phrases and verb phrases),
- ***syntactic complexity*** (e.g., dependency depth),
- ***syntactic roles***
- ***the ratio of coordinated, subordinated, and reduced sentences,***
- ***the number of active and passive sentences,***
- ***counts of dependencies*** (e.g., average dependencies per sentence)

(Fraser et al., 2014, Themistocleous, Ficek, et al., 2021, Themistocleous, Webster, et al., 2020)

Semantic Measures

Semantic Analysis

- ***Name Entity Recognition (NER)*** is a process of information extraction that can be used to determine how semantic relationships are presented linguistically. For example, Napoleon [Person] was the king of France [Place].

- ***Word Embeddings***

target	response	semantic_score	target	response	semantic_score	target
cat	dog	0.88	book	dog	0.18	computer
cat	tiger	0.79	book	tiger	0.02	computer
cat	child	0.48	book	child	0.21	computer
cat	lion	0.72	book	lion	0.12	computer
cat	airplane	0.33	book	airplane	-0.03	computer
cat	computer	0.28	book	computer	0.25	computer
cat	house	0.33	book	house	0.12	computer
cat	animal	0.63	book	animal	0.10	computer
cat	tree	0.56	book	tree	0.12	computer
cat	lamp	0.35	book	lamp	0.06	computer
cat	chair	0.13	book	chair	0.09	computer
cat	table	0.28	book	table	0.19	computer
cat	fish	0.59	book	fish	-0.01	computer
cat	octapus		book	octapus		computer
cat	mop	0.19	book	mop	-0.10	computer
cat	ball	0.45	book	ball	0.06	computer
cat	book	0.24	book	book	1.00	computer
cat	cat	1.00	book	ipad	0.06	computer

Discourse Analysis

- ***Idea density:*** the number of ideas expressed in the number of words or sentences (Danner et al., 2001), and measures of rhetorical structure (e.g., Elaboration, Attribution, and Joint) (Abdalla et al., 2018). Topics in discourse
- ***Lexical richness***
  - *type-token ratio (TTR),*
  - *Herdan's C (Herdan, 1955),*
  - *Maas's TTR, Mean segmental TTR,*
  - *Moving-Average Type–Token Ratio (MATTR) (Covington & McFall, 2010), word variation index,*
  - *counts on function words, hapax legomena (i.e., words that appear once in the corpus), and*
- ***n-grams:*** which are sequences of n (2, 3, or 4, etc.) words that occur in a text and can be employed to identify the speaker characteristics in a text.
- ***Sentiment analysis:*** lexical and semantic analysis quantify subjective information from texts to analyze the emotional tone and can provide insights into the attitudes, such as the speaker's stance, and positive or negative emotions associated with pathology.