



Nothing beyond text

Description

In explaining the political philosopher Jean-Jacques Rousseau's (1712-1778) autobiographical book *Confessions*, the continental philosopher [Jacques Derrida](#) observed that there is nothing to support Rousseau's claims about his relationships with his family members, other than what the text discloses. Derrida generalises this observation to the controversial proposition that any attempts to ground spoken or written assertions in some authoritative reality beyond what people say or write inevitably take us on a journey through yet more texts, *ad infinitum*. Though an apparent affront to conventional empiricism, Derrida's acclamation of the power of texts finds surprising support from techniques of automated natural language processing (NLP) systems, the foundation of much current text-based AI.

NLP corpora

NLP systems are typically *trained* on very large corpora of texts, such as a subset of the resources of the Internet, including social media posts, blogs, web pages, articles, books, manuals and official government documents. These sources also include texts that are in the business of words, such as lexicons, dictionaries, and language guides. Whether small or massive, a corpus provides the basis for NLP training.

NLP systems process words in relationship to one another. Most NLP models assume no structuring, ontology, etymological derivation or semantic network of correct relationships and usages or the consequent actions of words other than what words and phrases reveal in their contexts of other words. NLP really brings to light the power of words, independent of their effects outside the world of language. Consistency between statements counts for something, but correspondence to an everyday lifeworld, material evidence, truth and falsehood seems to have little currency in the NLP world of words other than what word sequences themselves reveal: as Derrida said, *il n'y a pas de hors-texte*, there is no outside text, or there is nothing outside of the text [158].

What's in a word

In this wordy NLP universe, word characteristics include **frequency** (how often words appear in the corpus), the **proximity** of words to other words in sentences, paragraphs and whole documents (local **context**), the apparent binding of certain words to each other (roof tile, window frame, garden hose, great grandmother) (**clustering**), and word **order** in phrases, clauses and sentences. Such numerical textual parameters seem sufficient to implement passable automated natural language translation, chatbot conversations, question and answer sessions, creative AI writing and other [intimations](#) that meaningful communication is taking place between human and machine. How that happens is the subject of my next post. In the mean time, see post: [The imitation game](#).

References

- Coyne, Richard. *Derrida for Architects*. Abingdon: Routledge, 2011.
- Derrida, Jacques. *Of Grammatology*. Trans. Gayatri Chakravorty Spivak. Baltimore, Maryland: Johns Hopkins University Press, 1976.

Note

- The featured image is generated by MidJourney prompted by "There is nothing beyond the text."



Category

1. Artificial Intelligence

Tags

1. Derrida

2. natural language processing

Date Created

February 11, 2023

Author

rcoyne99

default watermark