

# Conversational text composition through commonsense detection

Angel Rendon<sup>[0000–0003–3900–9582]</sup>

Universidad Nacional de Colombia  
amrendonsa@unal.edu.co  
<http://unal.edu.co/>

**Abstract.** Natural Language Processing techniques allows us to process text in wide range ways, making possible to extract key information out of texts and even proposing machine translators systems. One of those possibilities is tied to having well trained systems to have smart enough conversations with humans. This work aims to analyze the state-of-the-art techniques and implement them in the construction of a system that using different methods, make it possible to sustain a basic conversation on general topics.

**Keywords:** commonsense knowledge · natural language processing · machine learning · semantic association

## 1 Introduction

One of the artificial intelligence keystones would be definitively be having fully conversational systems to interact with people for several applications ranging from recommender systems, expert systems, to specialized chatbots and assistants<sup>1</sup>.

Several techniques based on Machine Learning (e.g. Bayesian models, SVM, supervised and unsupervised learning methods), and statistical model methods (e.g. word frequency, text rank, and inverse document frequency), have been used for a long time, with promising results.

However, systems based on these techniques rely on well formed corpora. As an example, WordNet [2] has the following synset for *cat*:

S: (n) computerized tomography, computed tomography, CT, computerized axial tomography, computed axial tomography, CAT (a method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis)

This simple example could lead to think that it would be possible to miss that specific synset when talking about *computerized tomography* when using

---

<sup>1</sup> <https://www.youtube.com/watch?v=d40jgFZ5hXk>

*cat* in a medical text, showing instead the most probable definition *feline*. That scenario is plausible since the knowledge is strongly dependent on the quality of either unstructured texts or its scale and domain-specific knowledge [1].

Lexical semantic understanding, sustained by commonsense knowledge, enriches the meaning of words and sentences

## References

1. Ghazvininejad, M., Brockett, C., Chang, M.W., Dolan, B., Gao, J., Yih, W.t., Galley, M.: A Knowledge-Grounded Neural Conversation Model pp. 5110–5117 (2017), <http://arxiv.org/abs/1702.01932>
2. Miller, G.A., Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: WordNet: An on-line lexical database. INTERNATIONAL JOURNAL OF LEXICOGRAPHY (1990)