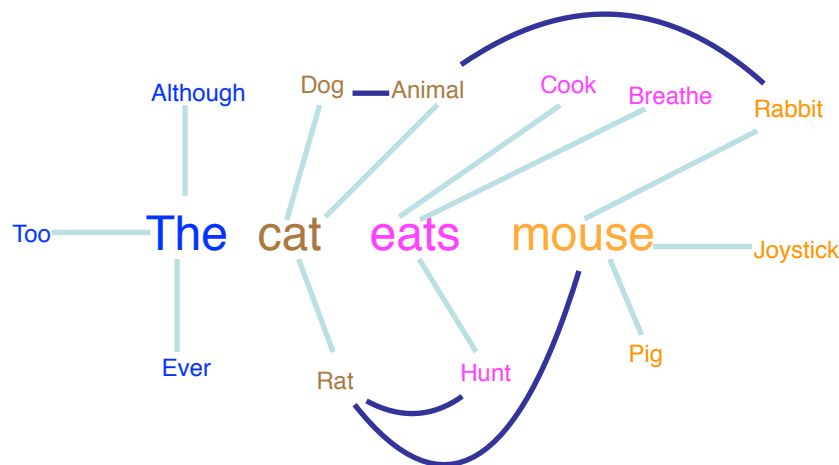
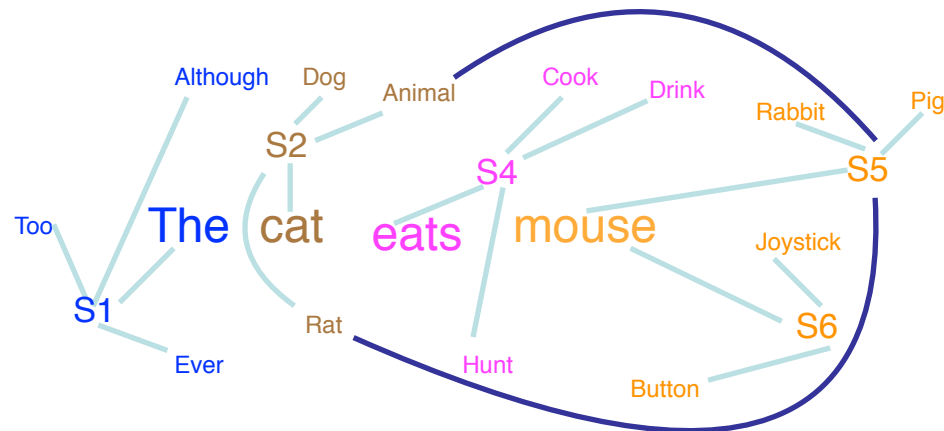


# Preparing Contextualized Embedding Using Distributional Thesaurus

- **Idea:** Given a sentence, prepare a local graph using Distributional Thesaurus (DT), Apply Network Embedding (NE) methods on that graph to get the embeddings of each word, as well as the full sentence (using a sentence marker [SM]).
- Two probable approaches to prepare the local graph
  - **Approach 1:** Take the n hop neighbour (from DT) of each word present in a sentence and prepare a local network where each node will be word only.
    - For example, in the following sentence, we add 'Rabbit, Pig, Joystick' for 'mouse' from DT network. After adding all the neighbours of each words in the sentence we prepare the following induced local graph by adding the extra edges (in 'deep blue') borrowed from DT network..



- **Approach 2:** Take the sense clusters (from sense clustered DT) of each word in a sentence, add the words in the sense cluster as well in the local network. Here the nodes would be of two types - words and sense clusters.
- For example, in the following sentence, we add 'S6 (may represent [Computer hardware](#) sense)' and 'S5 ([animal](#) sense)' for 'mouse' and the words in the clusters from DT network. After adding all such senses of each words in the sentence we prepare the following induced local graph by adding the extra edges (in 'deep blue') borrowed from DT network.



- We can experiment with different network embeddings like node2vec, deepwalk etc.
- After obtaining the contextualized word embeddings, we need to evaluate them using different NLP tasks like - paraphrase detection, question-answering, natural language inference/entailment detection, measuring semantic textual similarity, etc.
- Can be formulated as a Bachelor or Master thesis
- Contact Person - [Özge Sevgili Ergüven](#), [Steffen Remus](#), [Abhik Jana](#)