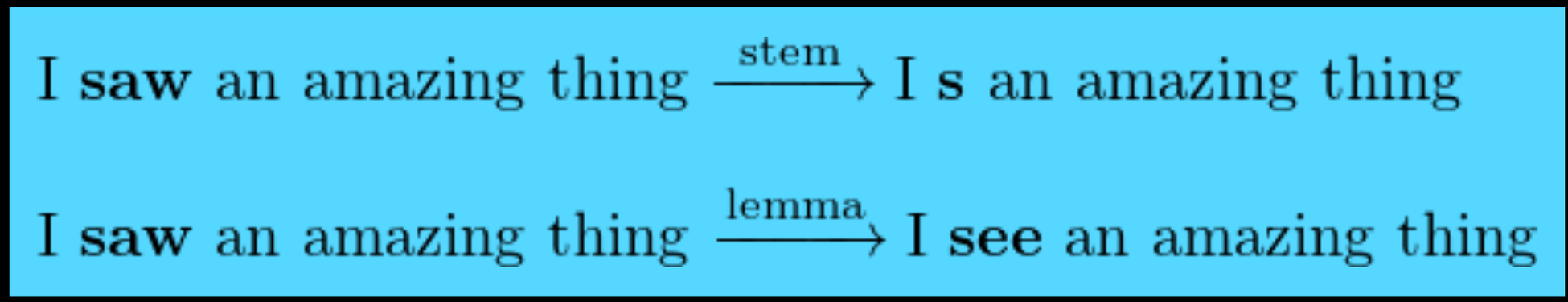
- The main challenge of NLP is **ambiguity**.
- This is largely due to **polysemy** and **homonymy**: Words that may seem unique can have different meanings depending on the context in which they are being evaluated.
  - Polysemy: “face” can mean the front of your head or confronting something.
  - Homonymy: “lie” can mean an untrue statement or a resting position.
- We can have words with different meanings in the same sentence **depending** on the way we interpret these words.
- This happens because of the difference between **signifier** (the way we represent the information, word) and **signified** (the meaning of that information, concept).



- We need to **normalize** text in order to analyze it.

- This means we want all the words to be lowercased or to convert plural terms into singular ones. We often do this with two methods:
  - **Stemming** chops off the ends of words in the hope of achieving this.
  - **Lemmatization** instead returns the base or dictionary form of a word, which is known as the lemma. It's the more robust approach than stemming.

- This means we want all the words to be lowercased or to convert plural terms into singular ones. We often do this with two methods:
  - **Stemming** chops off the ends of words in the hope of achieving this.
  - **Lemmatization** instead returns the base or dictionary form of a word, which is known as the lemma. It's the more robust approach than stemming.