**The laboratory work 8 (Feature-vectors)**

Use the feature-vector approach for the disambiguation approach.

As an example of this type of feature-encoding, consider the situation where we need to disambiguate the lexeme *bass* in the following example. An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.

A feature-vector consisting of the two words to the right and left of the target

word, along with their respective parts-of-speech, would yield the following

vector.

[guitar, NN1, and, CJC, player, NN1, stand, VVB]

The second type of feature consists of co-occurrence data about neighboring words, ignoring their exact position. In this approach, the words themselves (or their roots) serve as features. The value of the feature is the number of times the word occurs in a region surrounding the target word.

This region is most often defined as a fixed size window with the target word at the center. To make this approach manageable, a small number of frequently used content words are selected for use as features. This kind of feature is effective at capturing the general topic of the discourse in which the target word has occurred. This, in turn, tends to identify senses of a word that are specific to certain domains.

For example, a co-occurrence vector consisting of the 12 most frequent content words from a collection of *bass* sentences drawn from the WSJ corpus would have the words as features: *fishing, big, sound, player, fly, rod,* *pound, double, runs, playing, guitar, band*. Using these words as features, the vector will be represented as the following vector.

[0,0,0,1,0,0,0,0,0,0,1,0]

1. I went out in the woods and found a ***bat***.
2. All servants **bow** to the king
3. I found a **can** of paint outside
4. Make a **right** turn at the light
5. He quickly **rose** from his seat