First of all: I don't think this task is an easy one. All my colleagues I've talked too are sad that they've chosen a "wrong word". So I did same!

In my case, the third meaning of the word was somehow a generalization over the second meaning. I assume that if I would treat them the same, I will receive much better results.

However, without such manipulations I was still able to implement a word count (which in fact had no impact over my train dataset, because it contained mostly short sentences with unique words). IDF helped me to increase quality by 2%.

It's strange that every sentence which was related to the machinery oil were treated as third sense when it fairly should be the second one. However, once again - those senses are very similar and if it would be fixed, overall quality would improve.