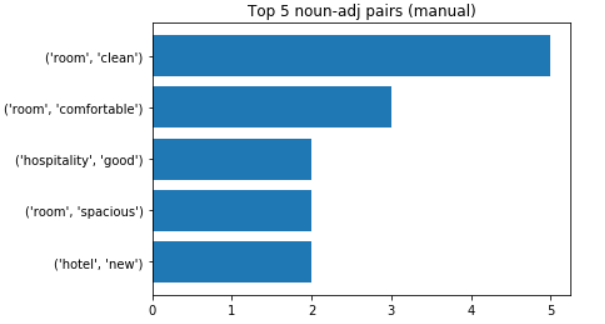
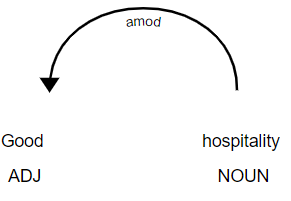
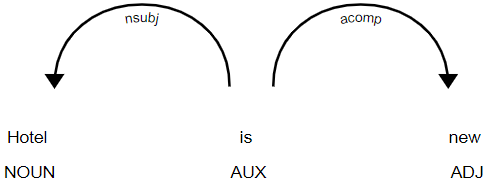
# Manual Identification

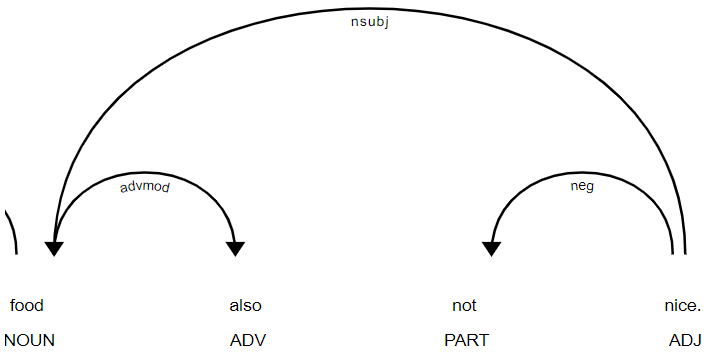
For part 3.2 of our assignment, we got 30 reviews of a hotel on Booking.com. We then manually identified all the noun-adjective pairs from the reviews. From there, we found the top 5 noun – adjective pairs, as shown below.

# Automatic Identification

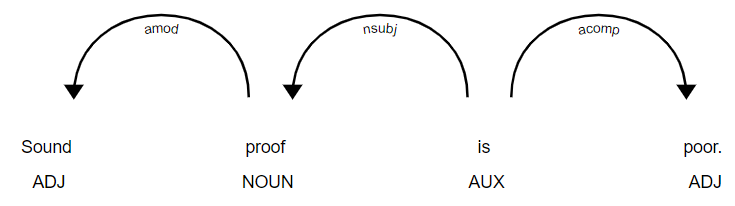
We then developed a program to find these noun – adjective pairs using spaCy, an open-source Natural Language Processing library in Python. We did so by finding the words that matched any one of three rules, which allowed us to find most of the noun – adjective pairs. For each rule, we searched for the children of the noun with a *compound* dependency and added the words to the noun. This allowed us to find the whole noun phrase (e.g. parking slots) instead of just one part of the phrase (e.g. slots). We also searched for children of the adjective with a *conj* dependency, which allowed us to extract the pairs <room – clean> and <room – comfortable> from the sentence “The room was clean and comfortable”.

The first rule looks for words with the *amod* dependency. In this case, our program will find the adjective “good”, and take the head of the token as the noun, giving the noun – adjective pair <hospitality – good>.

The second rule looks for words having children with both *nsubj* and *acomp* dependencies. In this case, our program will identify “is”, which has “hotel” with a *nsubj* dependency and “new” with an *acomp* dependency. This will give us the noun – adjective pair <hotel – new>.

The third rule is similar to the first rule, covering the edge cases where the noun has a *nsubj* dependency, and the head of the noun is an adjective. In this case, the program finds the adjective “nice”, and searches its children for a word with a *nsubj* dependency. In this case, an additional step taken as there is a *neg* dependency. The final pair returned will be <food – not nice>.

# Challenges Faced

One of the challenges faced was that the reviews were written informally, with poor spelling, and sentence structures. This could lead to incorrect POS tagging being applied to the sentence. In this case, the review had the sentence “Sound proof is poor”. “Soundproof” is supposed to be a noun, but as the review wrote it as two words, spaCy’s dependency tagger interpreted “Sound” as an *amod* of “proof”, leading to the incorrect pairs of <proof – sound> and <proof – poor>.

This problem was worsened by the fact that the size of the dataset was too small. Many noun – adjective pairs that we found only appeared once throughout all the reviews. This results in a relatively larger proportion of misspellings, which would affect the results greatly.