SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

50.007 Machine Learning, Fall 2015
Design Project

Annotation Guideline

16 Oct 2015

This is the annotation guideline for annotating the noun phrases present in an SMS corpus.

1 Noun Phrase Definition

Simply said, a **noun phrase** is a sequence of words referring to a single object or entity. For the purpose of this annotation task, we need to annotate **the outermost noun phrases** found in the SMS corpus.

Being the *outermost noun phrase* means it is not part of any larger noun phrases. For example, in the text `I went to the Bank of China`, the noun phrases are `I` and `Bank of China`, while in the text `It is some scavenger hunt thingy my fren thought up`, the noun phrases are `It` and the whole `some scavenger hunt thingy my fren thought up` since the phrase `my fren thought up` is a dependent clause, explaining `some scavenger hunt thingy`.

Note that we do not tag `China` in the first example and `my fren` in the second example as a noun phrases, since they are part of a larger noun phrase.

1.1 Identifying Noun Phrases

A good way to determine whether a phrase is a noun phrase is to do replacement test. In this test, try to replace the phrase with a suitable pronoun (e.g., it, they). If the replaced phrases fit in the sentence, then they are noun phrases.

For example:

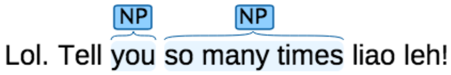
1. a. This sentence contains two noun phrases.
   b. **It** contains **them**.

2. a. The subject noun phrase that is present in this sentence is complex.
   b. **It** is complex.

2 Examples

2.1 Consecutive Noun Phrases

Some noun phrases can be immediately next to another noun phrase. If those noun phrases are not part of a single larger noun phrase, then tag those noun phrases as separate noun phrases. For example:

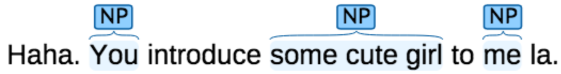


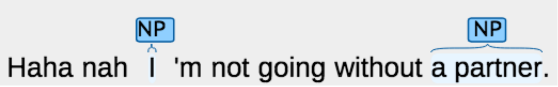
Lol. Tell you so many times liao leh!

2.2 Noun Modifiers

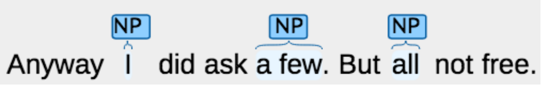
A noun phrase often includes the modifiers that comes before the head noun. This may include determiners (a, an the), adjectives (big, lovable, user-friendly, following), possessives (my, his, her, John's), and quantifiers (some of, all, every, no, none of). Those modifiers are included in the noun phrase.

So the annotations should be done as follows:


Haha. You introduce some cute girl to me la.

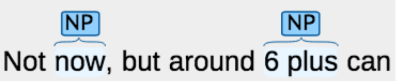

Haha nah I 'm not going without a partner.

And some quantifiers may appear by themselves. These, too, are noun phrases, for example the ”all” in the following example:


Anyway I did ask a few. But all not free.

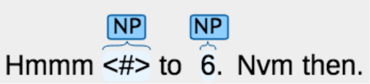
2.3 Numbers

Numbers are usually noun phrases, even when they represent time, phone number, or date. These are considered noun phrases.


Not now, but around 6 plus can

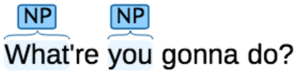
In our SMS dataset, some longer numbers are converted into the token <#>, these are still considered noun phrases.

For example:

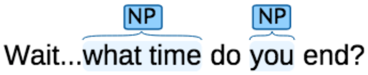

Hmmm <#> to 6. Nvm then.

2.4 Question Words

Some question words might appear as noun phrases, such as ”what” in the following example:


What're you gonna do?

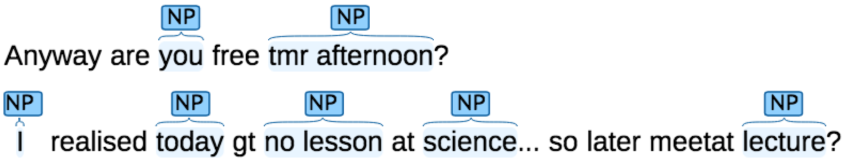
Be careful to include the whole noun phrase, such as in ”what time”, or ”which book”:


Wait...what time do you end?

Note that ”where” and ”when” are generally not considered noun phrases, as they replace prepositional phrases, such as ”at home” or ”on 9 December”. The words ”home” and ”9 December” themselves are noun phrases, but not ”at home” and ”on 9 December”.

2.5 Location and Time

Continuing the previous section, note that locations and time references are generally noun phrases, like the following examples:



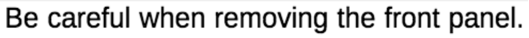
Anyway are you free tmr afternoon?

I realised today gt no lesson at science... so later meetat lecture?

2.6 Verbal Noun Phrase

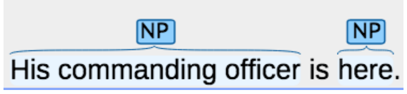
There are some verb phrases that serve as noun phrases. For the purpose of this task, verb phrases are **not** considered as noun phrases, *unless* the verb role in the phrase is adjectival.

So the ”removing the front panel” in the following example is not a noun phrase:



Be careful when removing the front panel.

But the ”commanding” in the following example is part of a noun phrase:



His commanding officer is here.

2.7 General

In general, those phrases starting with determiners are very likely to be noun phrases, so look out for determiners.

When in doubt, apply the replacement test mentioned in section 1.1 above.