# Impact of noun-prep-noun collocations on parsing results

Eric Wehrli

LATL-CUI

University of Geneva

Eric.Wehrli@unige.ch

August 25, 2014

## Abstract

In a previous workshop, I argued that collocational knowledge has a positive impact on parsing results, a claim supported by evaluation data based on verb-object collocations in English. The present paper is also concerned with the impact of collocations on parsing, but this time the evaluation is extended to French data and takes into account a much larger set of collocations, including the noun-preposition-noun class, a particularly common type of collocations in Romance languages.

Collocation knowledge can help the parser in two ways. The improvement comes either from a better choice of lexical element (in case of ambiguous words), or from a more felicitous phrase attachment. Both cases are illustrated below. Collocations of the noun-noun type containing syntactically ambiguous words (in the sense that they can be assigned more than one lexical category) as in (1):

(1)  austerity measures
     opinion polls
     tax cuts
     protest marches

The noun-noun collocations in (1) all have a noun head which is ambiguous between a nominal and a verbal reading. Such examples, which are

taken from a newspaper corpus, are likely to be wrongly interpreted (ie. with the verbal reading) when parsed with the identification procedure turned off.

The second way in which collocational knowledge can help the parser has to do with structural ambiguities. This concerns particularly collocations which include a prepositional phrase, such as the noun-preposition-noun collocations, as in (2):

(2)  bone of contention
     state of emergency
     struggle for life
     flag of convenience

The attachment of prepositional phrases is known to be a very difficult task for parsers. So, knowing that a particular prepositional phrase is part of a collocation (and giving priority to analyses containing collocations over alternative analyses) is an effective way to solve a large number of PP attachment ambiguities.

Although quite common in English, noun-preposition-noun collocations are much more frequent in Romance languages. This is due to a large extent to the fact that Romance languages lack the noun-noun construction of Germanic languages and use instead the noun-preposition-noun construction. Thus, many of the noun-noun collocations of English have a noun-preposition-noun equivalent in French, as illustrated in (3).

(3)a.  ligne de partage des eaux – watershed

   b.  force de maintien de la paix – peacekeeping force

   c.  organisation de protection de l'environnement – environment protection agency

We will present an evaluation comparing parse results for French documents obtained by our Fips parser with and without the collocation detection component.

To simplify the (manual) evaluation, the output of the parser will be displayed in POS-tag mode, that is one word per line, along with dependency information specifying the attachment site of a prepositional complement.