

Thesis

Philip White

November 4, 2011

1 Grammatical Relations

The simplest classification approach considered the relative frequency of different grammatical relations. For this approach, the governor and the dependent of the dependencies were ignored, with only the relation itself being used.

Each data set instance contained 52 numerical attributes, one for each relation in the Stanford Dependency system. For each attribute A_r corresponding to the relation r , the corresponding value was n_r/n_t , where n_r and n_t were the number of occurrences of the relation r and the total number of relations in the text, respectively.

A C4.5 decision tree classifier trained on these instances produces the decision tree shown in Algorithm 1.1. The full names for the seven relations are shown in Table 1.1. The following subsections explore the linguistic reasons why these particular relations should be so useful in classifying the texts.

Algorithm 1.1 C4.5 decision tree classifier

```
if complm  $\leq$  0.011635 then
  if purpcl  $\leq$  0.000856 then
    if rcmod  $\leq$  0.012254 then
      en(34.0)
    else
      if prt  $\leq$  0.002113 then
        es(4.0/1.0)
      else
        en(6.0)
  else
    if purpcl  $\leq$  0.001191 then
      if advmod  $\leq$  0.045825 then
        es(6.0)
      else
        en(2.0)
    else
      en(7.0)
else
  if mark  $\leq$  0.00808 then
    en(6.0)
  else
    if aux  $\leq$  0.044037 then
      en(6.0/1.0)
    else
      en(60.0)
```

Table 1.1: Relation abbreviations

<i>advmod</i>	adverbial modifier
<i>aux</i>	auxiliary
<i>complm</i>	complementizer
<i>mark</i>	marker
<i>prt</i>	phrasal verb particle
<i>purpcl</i>	purpose clause modifier
<i>rcmod</i>	relative clause modifier

1.1 Adverbial Modifier

1.2 Auxiliary

1.3 Complementizer

1.4 Marker

1.5 Phrasal Verb Particle

The phrasal verb particle relation ties the head word of a phrasal verb to its particle. See Example 1.1.

prt
└──┘
(1.1) ...the reduction of superfluous proteins will free up resources ...

1.6 Purpose Clause Modifier

1.7 Relative Clause Modifier

1.1

Witten and Frank [2005]

References

Marie-Catherine de Marneffe and Christopher D. Manning. Stanford typed dependencies manual, September 2008. URL http://nlp.stanford.edu/software/dependencies_manual.pdf.

Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, second edition, 2005.