Requirement Extraction from Transport Policy Documents

Oliver Mason, Aleksandar Trklja, Dominik Vajn
The University of Birmingham
Project Aims

- reduce the amount of text to read
- work towards improving requirement documents
Problems

- no clear definition of ‘requirement’
- ‘shall’ statements?!
- subjectivity
- different levels: policy, strategy, requirement
Approach

- linguistic analysis of a set of policy docs
- evaluated on an unseen paper annotated by a systems engineer
- manual
Approach

- lexical triggers (modal verbs, requires, requirement)
- structures (it is X that/to-inf)
- no machine learning involved
Issues

- small amounts of data
- rule-based vs statistical
- no clear definition
Results

• Local Grammar matches 189 sentences (25% of the text)
• Human - Requirements identified: 90
• System - Correct matches: 72
• Precision: 0.38
• Recall: 0.8
Train operators AGENT must improve ACTION [their efficiency ] TARGET in ways that are visible on the bottom line.

For our railway to be more affordable, it AGENT must also be efficient KEY, ...
Analysis

[We ] AGENT must also [help ] ACTION [to ] WITH_CLAUSE [reduce ] INF the demand on taxpayer subsidy...
Analysis

... [we ] AGENT need [to ]
WITH_CLAUSE [look ] INF seriously
at the possibility of rewarding
passengers who do not travel on the
most crowded trains ...
Analysis

... , which exemplifies the increasing recognition that innovation needs to be [managed] ACTION and encouraged;
Analysis

The rail industry and the unions now need to consider how working practices could become more efficient.
Consideration also needs to be given to whether more streamlined training programmes can be developed which can deliver the equivalent level of skill over a shorter time period.
Analysis

X will play an important role in Y
X is an essential part of our strategy
X could be a good way to Y
our ambition is to Y
X must : <bullet list>
Conclusions

- using local grammar is a feasible approach
- development of the grammar is iterative
- scope of patterns needs to be broadened
- alternative ways of expressing reqs
Conclusions

• many patterns/paths are still too general
• semantic classes needed to filter false hits
• lexical items need to replace POS classes
  • eg *it is ADJ that/to-inf*: ADJ is too general