Lexical Simplification

Matthew Shardlow

Initial lexical simplification system developed as a simplicity ranking task [1]

Portuguese language simplification advances initial work [2]

Word sense disambiguation investigated [3,4,5]

Present Day

1998

2007 - 2010

2010 - 2012

2012

Automated simplification of lexicons has been occurring since 1998.
Lexical simplification sits between text simplification and assistive technology.
Simple is a subjective concept and difficult to define.
This research will survey and improve current lexical simplification systems.
The pipeline below shows the current research challenges in a typical system.

The Lexical Simplification Pipeline

Input: The feline sat on the mat

Identification of Complex Words
The feline sat on the mat

Generation of Substitutions
Feline: Cat, Moggie, Kitten, Tomcat

Word Sense Disambiguation
Cat, Moggie, Kitten, Tomcat

Ranking Synonyms by Simplicity
#1 Cat #2 Moggie

Result: The cat sat on the mat

Difficult terms must be identified before simplification.
Techniques borrowed from linguistics exhibit poor performance.
Challenging as context affects complexity.

A set of synonyms for a difficult term is obtained.
Typically WordNet has been used.
Recent systems have learnt substitutions from corpora.

Replacements which make little sense are discarded.
Context must be taken into account.
Few techniques currently exist.

The easiest to understand substitution is discovered.
Affected by factors such as context and audience.
Most research into simplification has focussed on this task.

Who Could Benefit?

Second Language Learners
Easy to read texts in target language.

Lay Readers of Technical Documents
Difficult terms replaced with familiar.

Academics
Out-of-domain technical content simplified.

Low Literacy Readers
Automatic access to news/web content.

People suffering Cognitive Impairment
Rehabilitative access to information.

Future Work

Research will focus on optimising the above pipeline.
Novel evaluation methods will be developed.
When appropriate, new techniques will be introduced.
A proof of concept system will be developed.
Evaluation will involve system’s target user groups.

References: