

# The Subjectivity of Lexical Cohesion in Text

Jane Morris and Graeme Hirst, *University of Toronto*

## 1 Research questions

**General:** What degree of subjectivity exists in text understanding?

**Specific:** Degree of subjectivity in readers' perceptions of lexical cohesion.

**Future:** How does this subjectivity reflect the reader's attitude?

## 2 Lexical cohesion

Intersentence groups of related words indicate the structure of topical units:  
*call, notify, send, demand* ⇒ communication.

*James Bond, Hollywood, filmmakers, actors* ⇒ movies.

## 3 Lexical semantic relations

**"Classical":** hyponymy, synonymy, antonymy, meronymy, ...

**"Non-classical":** All the rest: Non-hierarchical relations with no standard classification. Heavily used by readers but under-represented in lexical resources.

*James Bond / Hollywood:* Related because James Bond films are made in Hollywood.

## 4 Our approach

Investigate readers' perceptions of lexical cohesion.

**Basic methodology:** Measure subjectivity as individual differences.

## 5 Experiment

**Pilot:** 5 readers, 1.5 pages of *Reader's Digest* article. Subjects marked the word groups and the related word pairs they perceived, and then described how they thought the word pairs were related.

Call the police! Notify the censors! James Bond was caught *smoking a cigar!* Who do we get in touch with? Should we send a letter to the Department of Role Model Development in the Hollywood Ministry of Socialist Realism? How dare filmmakers show actors *smoking a cigar.* We demand to know how this happened.

Word Numbers: 4, 2 Related Word Pair: *James Bond, Hollywood*  
What is the relation between the words? *James Bond movies are made in Hollywood.*

Subjects used colored pencils to mark words that they perceived to be related. Then they explained the relationship that they saw.

## 6 Results

- (1) Pairwise agreement on word membership in each group.
- (2) Word pairs agreed on by  $\geq 50\%$  of the subjects for "core" words.
- (3) Agreement on relation types for pairs used by  $\geq 2$  readers.

Gloss of word group	Pairwise agreement (%)	Word pairs agreed on (%)	Relationship agreement (%)
Movies	71	10	75
Communications <sup>a</sup>	69	12	20
Smoking	73	13	85
Groups and causes	63	18	69
Bad behaviors	41	12	100

<sup>a</sup>Only 3 subjects used this group.

## 7 Subjectivity and attitude

Subjects largely agree on the core words of "bad behaviors" group:

Core words (chosen by $\geq 3/5$ subjects)
5 shooting
4 sex
4 drinking
4 dangerous
3 drag racing
3 irresponsible [behaviors]

One subject adds idiosyncratic attitude-bearing choices:

"Law/order/authority" outliers (all chosen by one single subject)

- 1 police
- 1 caught
- 1 British Intelligence Service
- 1 gun control lobby
- 1 Department of Role Model Development
- 1 MADD
- 1 spies

## 8 Discussion

There is a "common core" of words within word groups, but some subjectivity is observed.

The word pair agreement is low, but the relation type agreement is high.

The subjects reported that marking word groups and explaining relations in context is easy, but they had difficulty choosing word pairs from the word groups.

## 9 Future work

Complete the study with 3 texts, 26 subjects. (Present results are for pilot only.)

Focus on non-classical relations: Is there a "smallish" set of common non-classical relations that readers use?

Use different types of readers and texts.

Investigate the potential of Hasan's "cohesive harmony".

Hasan, Ruqaiya. (1984). Coherence and cohesive harmony. In J. Flood (ed.), *Understanding reading comprehension*. International Reading Association.

Further explore how attitude is reflected in subjectivity.

This work was supported by the Natural Sciences and Engineering Research Council of Canada. We are grateful to Clare Beghtol for helpful discussions.