LSP 2007, Hamburg
Monday 27 August 2007

The Languages of Emotion and Financial News

Ann Devitt
Khurshid Ahmad

Specialised Language in Global Communication

- Language of international financial transactions
- across linguistic and cultural barriers
- uses specialist vocabulary

Specialised Language of Financial News

Reuters Group fell after the news and information company confirmed it was in discussions about a possible combination with rival Thomson Corp. There has been concern about the uncertainty about regulatory approval, and execution problems if the combination is approved.

Sentiment and the Markets

Ann Devitt
Trinity College Dublin

Specialised Language in Global Communication

"I remember '29 very well," Steinbeck writes (2002: 17), "We had it made…I remember the drugged and happy faces of people who built paper fortunes in stocks they couldn't possibly have paid for […]." Then, however, "came panic, and panic changed to dull shock…People remembered their little bank balances, the only certainties in a treacherous world. They rushed to draw the money out. There were fights and riots and lines of policemen. Some banks failed; rumors began to fly"

Proponents of market economics insist that financial transactions are conducted between rational agents.

Research into the causes of the failures of financial institutions, regular boom-&-bust cycles, suggests that 'sentiment' plays a key role here.

Specialised Language of Financial News

Reuters Group **fell** after the news and information company confirmed it was in discussions about a possible combination with **rival** Thomson Corp. There has been **concern** about the **uncertainty** about regulatory approval, and execution **problems** if the combination is approved.

Specialised Language in Global Communication

Sentiment analysis is now an important sub-discipline of financial studies.

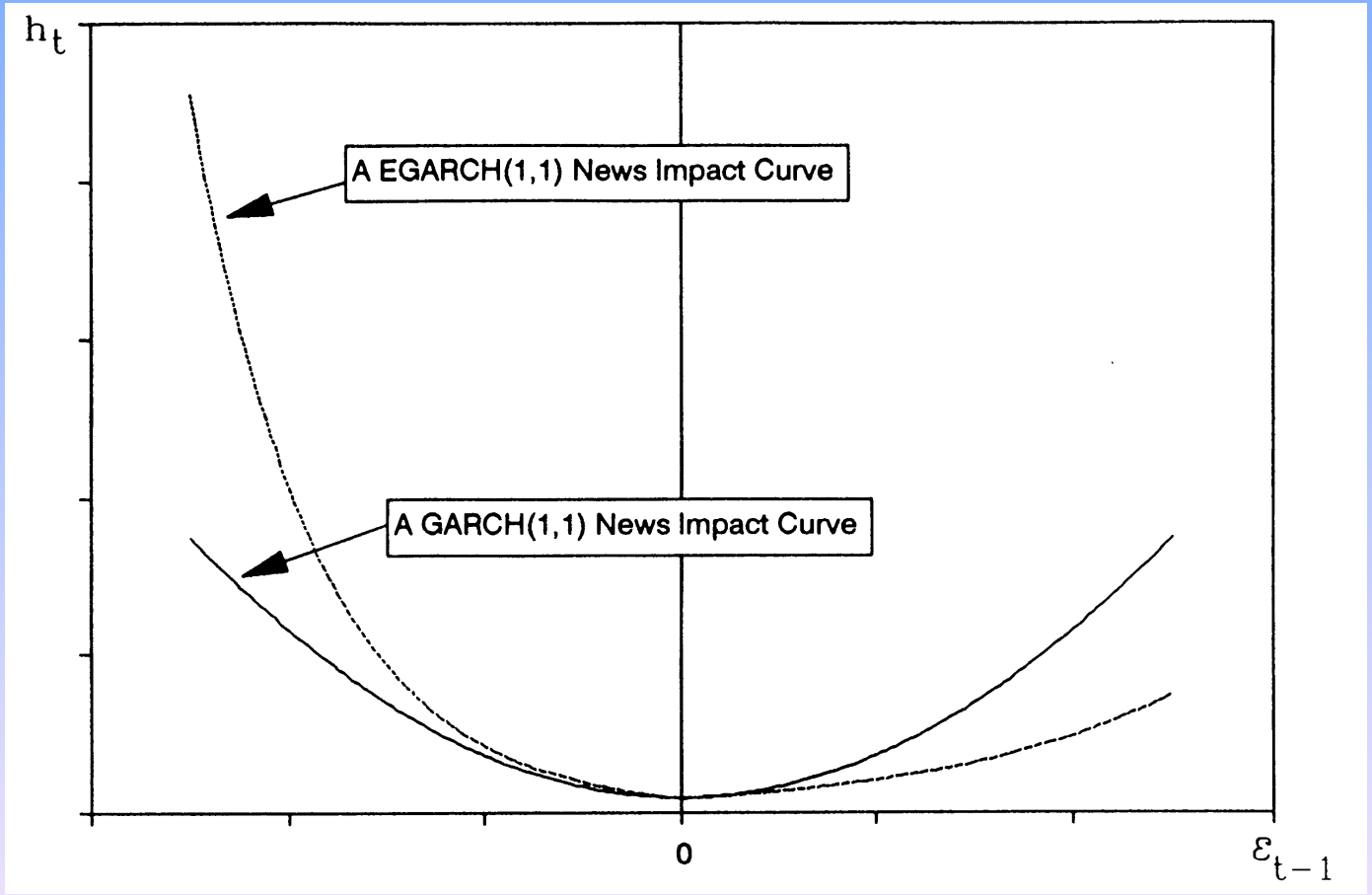
The words *sentiment* and *emotion* are inextricably linked.

We are focusing on the role of (specialised) language in financial transactions and, in particular, how the transactions are influenced by and influence what is written about the markets.

Sentiment and the Markets

The induction of an emotional public response in order to generate feedback effects constitutes a priority of today's mass media, also of the business media. There is thus absolutely no doubt that the news flow of the mass media does not follow a random pattern. A homogenization of the market response is thereby certainly not prevented.

Engle Ng (1993) Asymmetry Curve

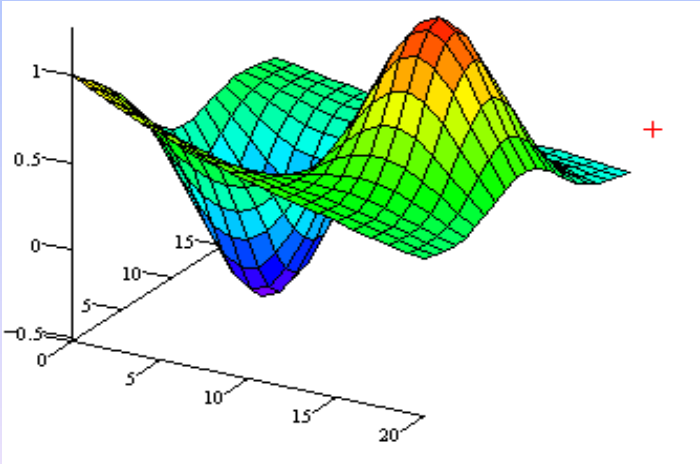


Outline

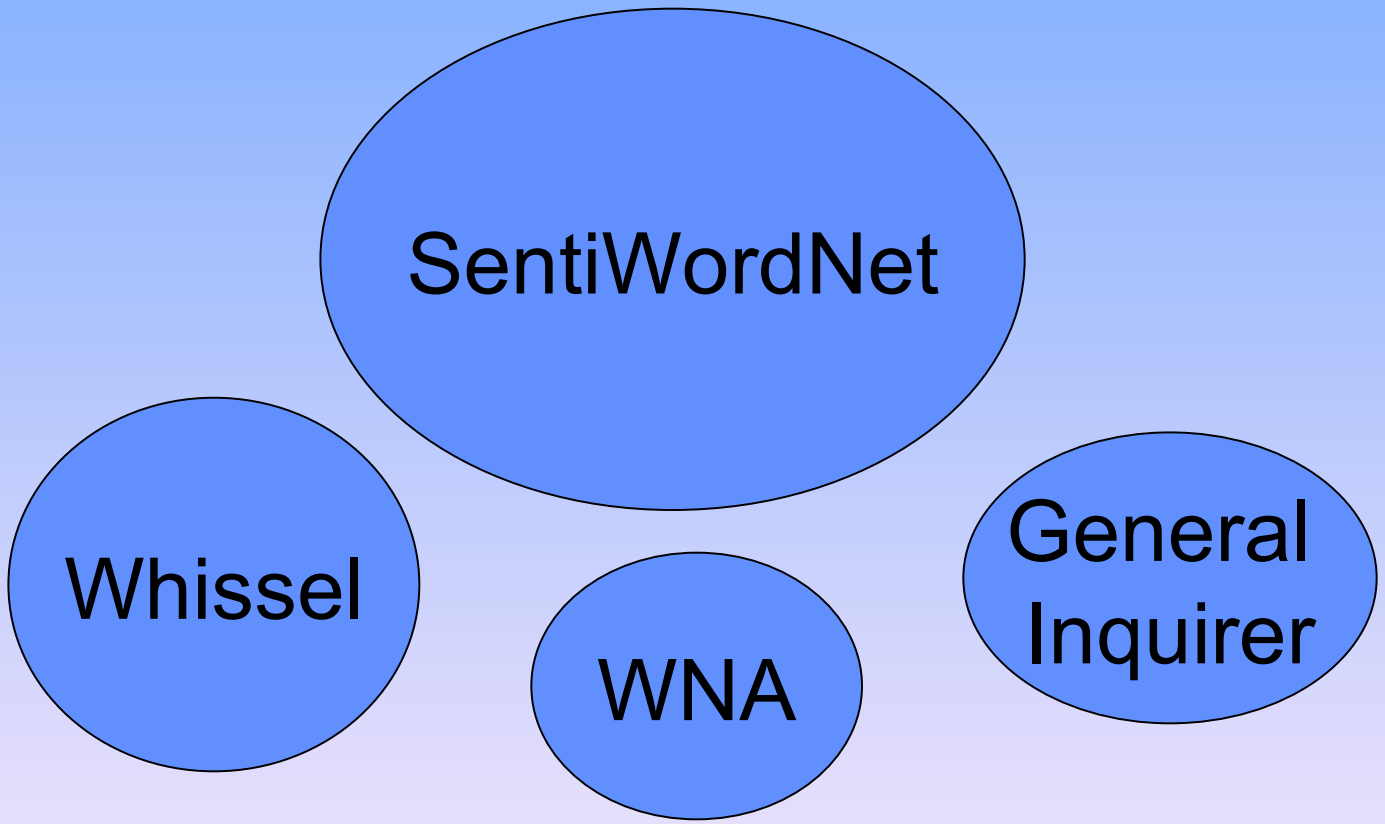
- Current psychological theory of emotion
- Evaluation of lexical "emotion" resources
- Comparative corpus analysis of "emotion"

Cognitive Theory of Emotion: Dimensions

- Osgood / Russell
  - Evaluation
  - Activity
  - Potency
- Mehabrian PAD
  - Pleasure
  - Activation
  - Dominance

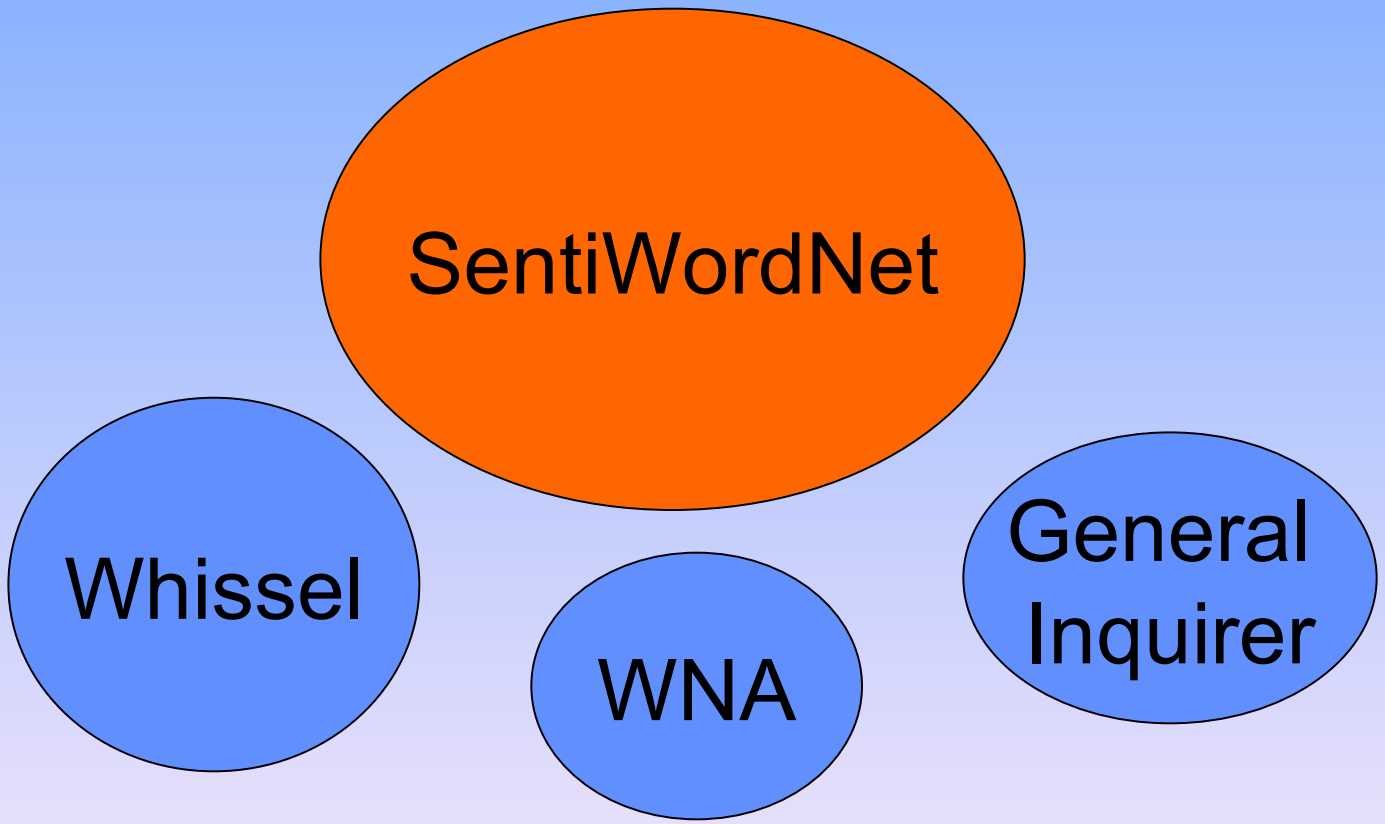


Lexical Resource Evaluation



Lexical Resource Evaluation

Senti WordNet



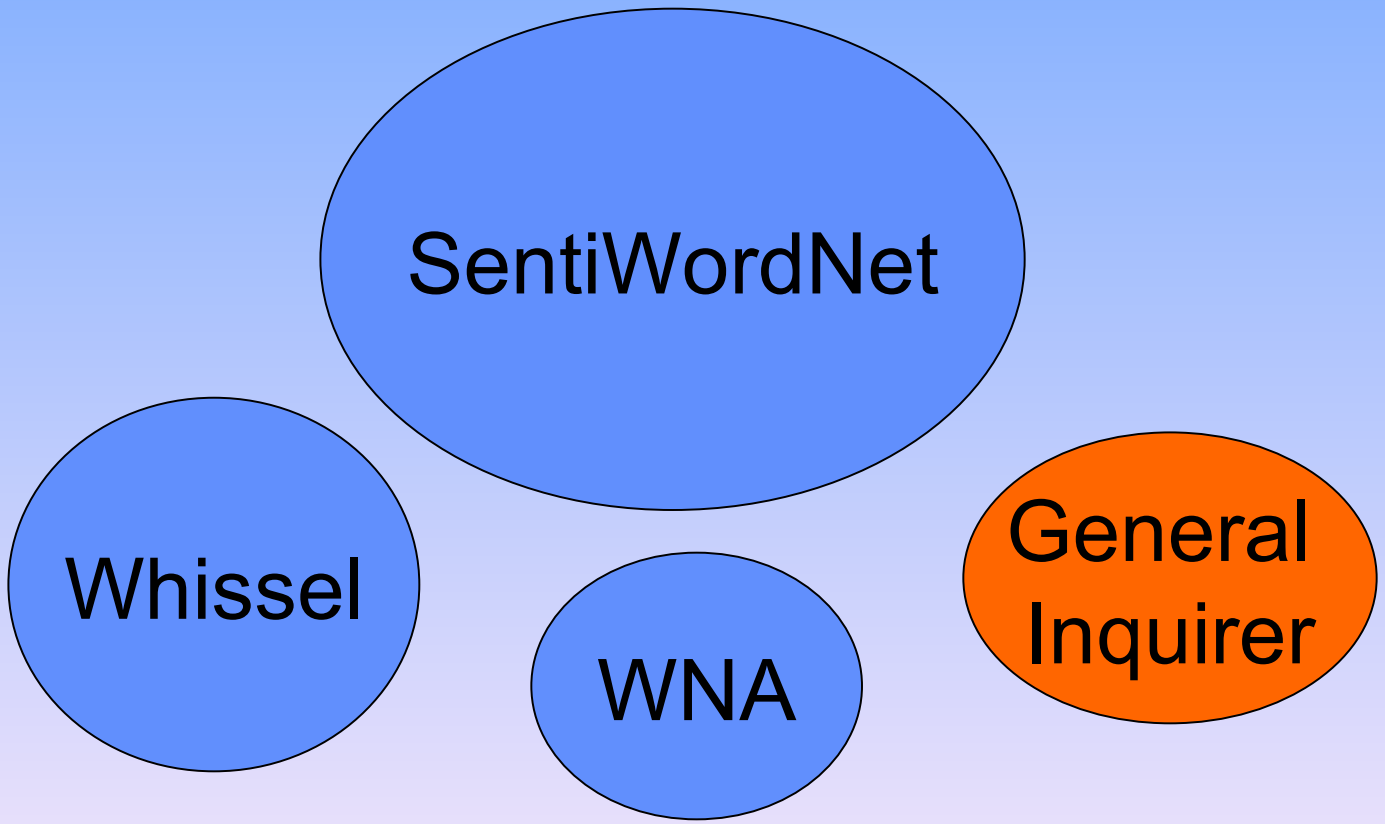
Lexical Resource Evaluation

Senti WordNet

Word	PositiveVal	NegativeVal
Happy	0.9	0.0
Sad	0.0	0.9

- 39066 terms
- Evaluation dimension scale: 0 - 1
- Low average: Pos=0.18, Neg=0.23
- More extreme Neg values
- Error-prone: rude (pos 0.875), gladsome (neg 0.875)

Lexical Resource Evaluation
General Inquirer



Lexical Resource Evaluation

General Inquirer

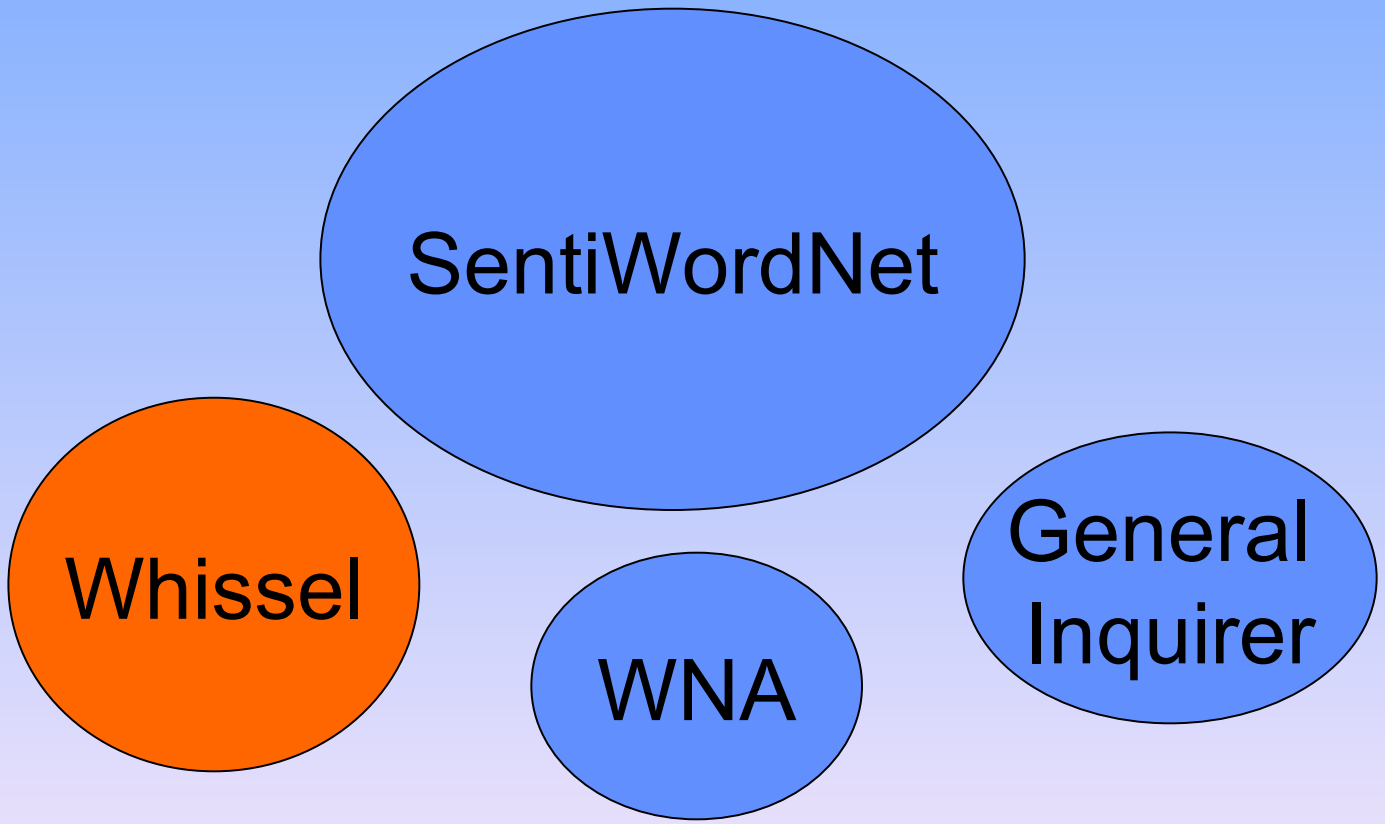
ECSTATIC Pos Pleasure

SORROWFUL Neg Pain

- 8641 terms
- 184 binary categories (including MAB dimensions)
- Negative > Positive
- Active > Passive
- Strong > Weak

Lexical Resource Evaluation

Whissel Dictionary of Affect



Lexical Resource Evaluation

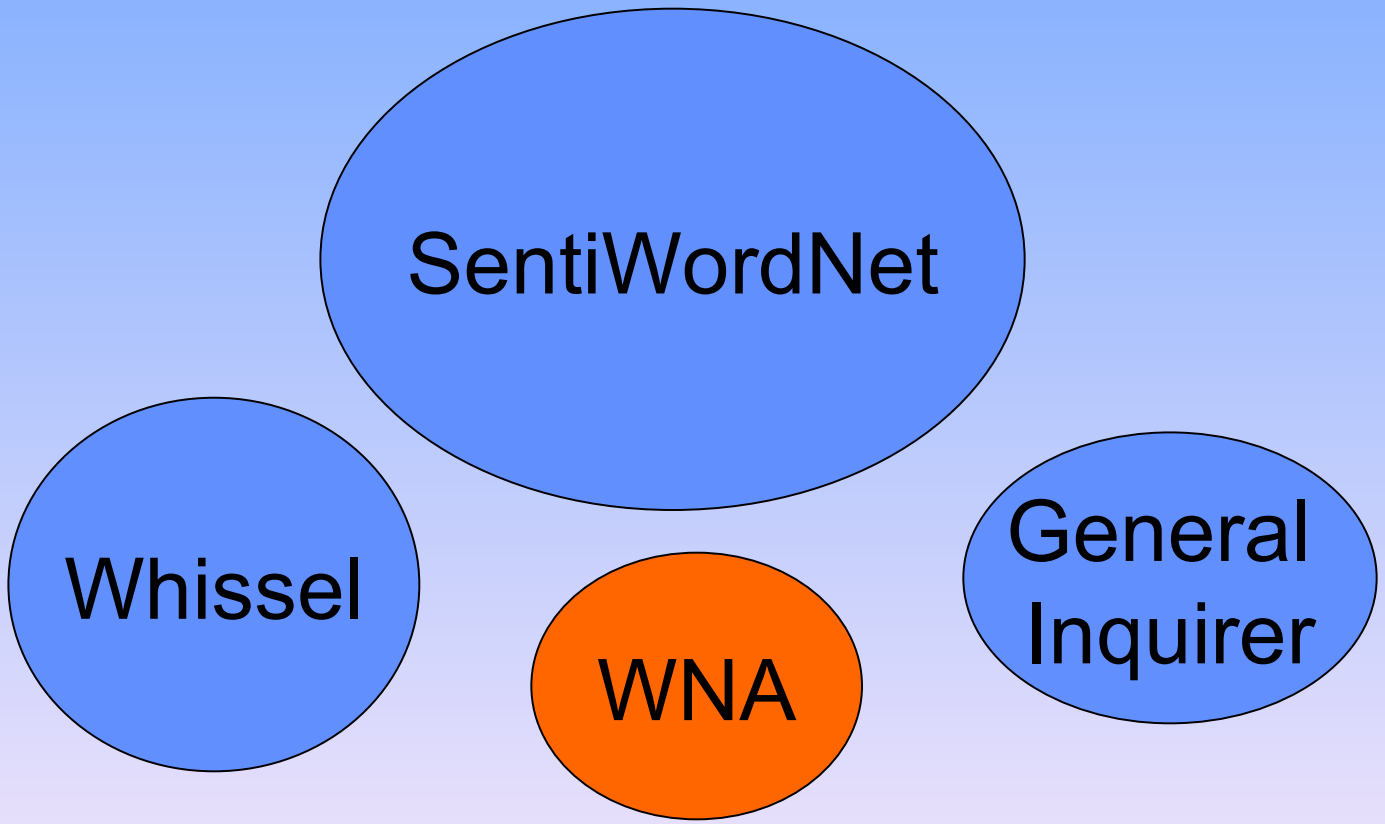
Whissel Dictionary of Affect

Word	Eval	Activ	Imag
great	2.6250	2.1250	1.0
disastrous	1.4444	2.4000	2.0

- 8742 terms
- Dimensional representation: 1-3 scale
- Evaluation, Activation, Imagery

Lexical Resource Evaluation

WordNet Affect



Lexical Resource Evaluation

WordNet Affect

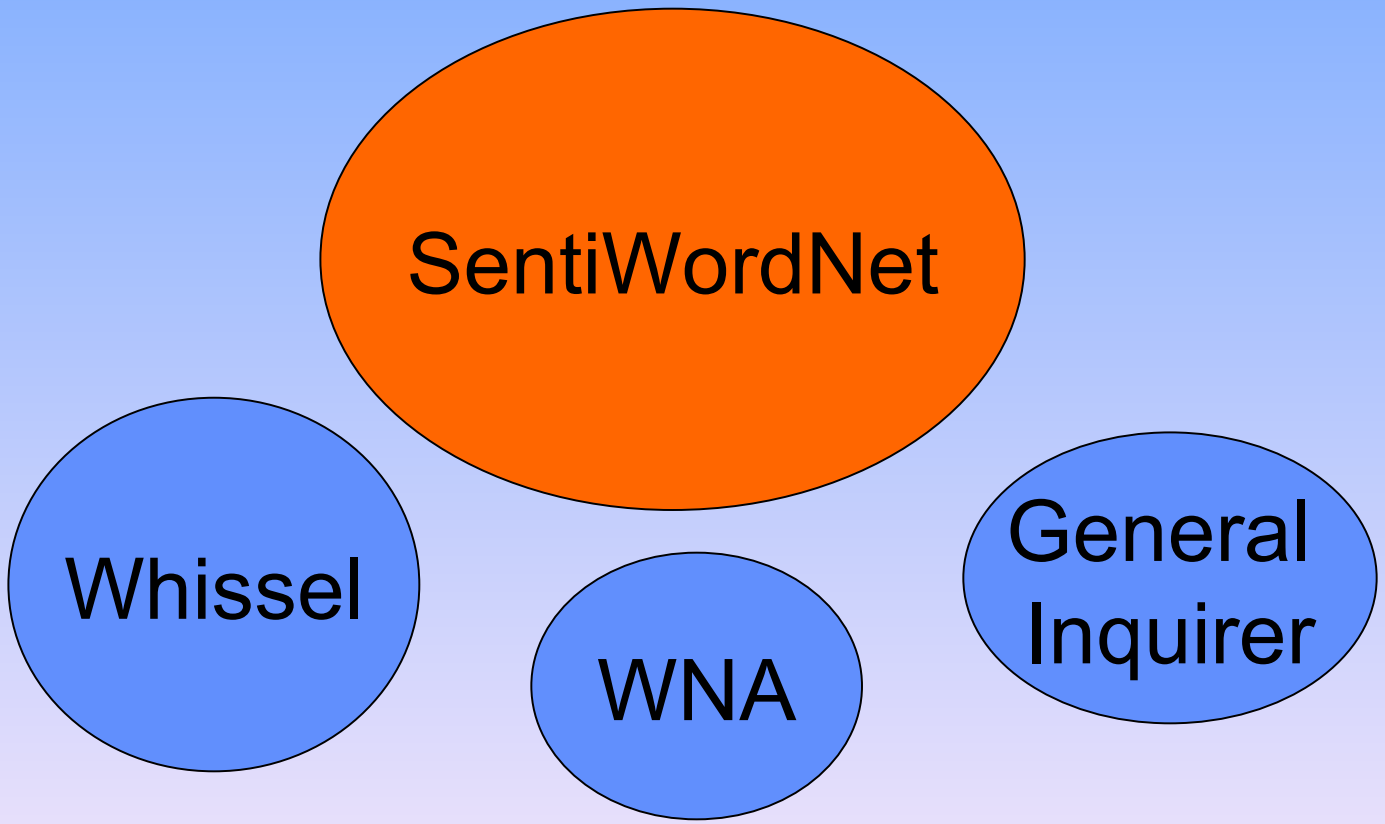
Word	BinaryFeatures
Loneliness	cognitive state, emotion
Happiness	cognitive state, emotion

- 5432 terms
- Domains of emotional experience
- No Polarity
- Short-term: Mood, Manner
- Long-term: Attribute, Trait

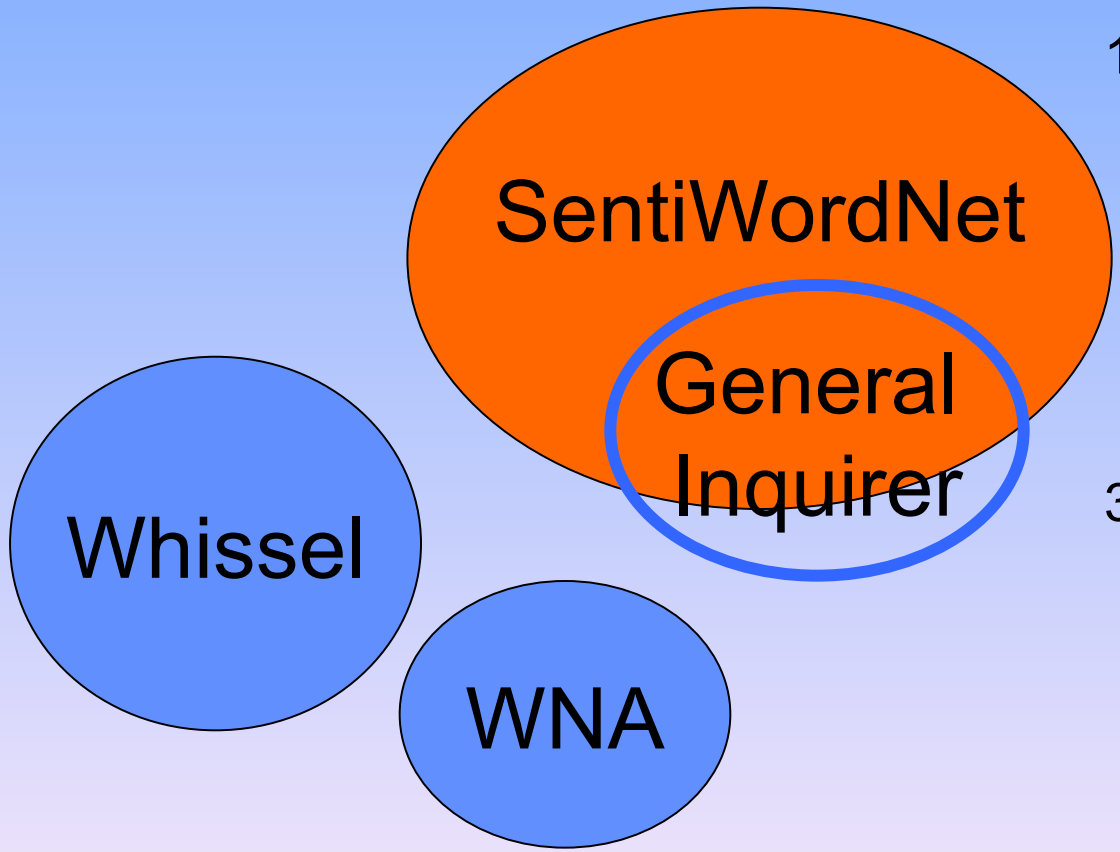
Lexical Resource Evaluation
Lexical Overlap

- Are the lexica consistent?
- Are they mutually exclusive?
- Dice, Jaccard and Asymmetric coefficients

Lexical Resource Evaluation
Lexical Overlap



Lexical Resource Evaluation
Lexical Overlap



1. Statistically significant agreement for Polarity Assignment (Chi square test)

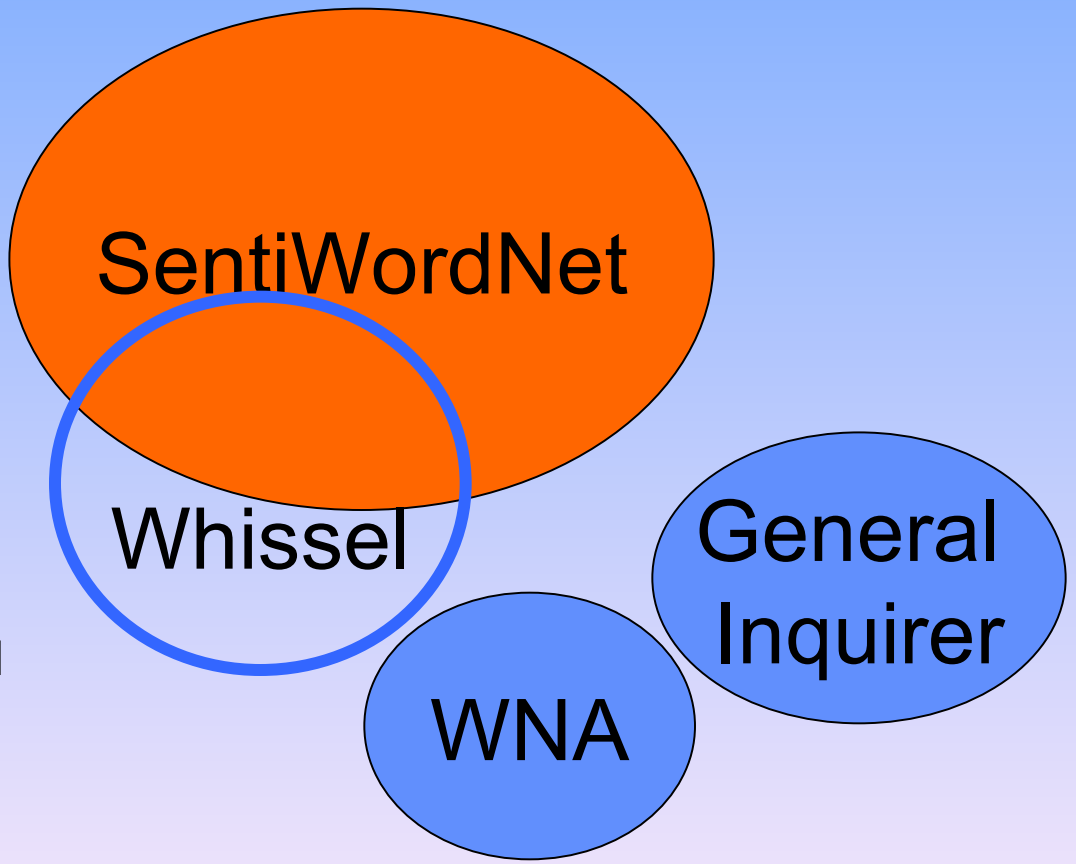
3. Very weak correlation for activation features.

Lexical Resource Evaluation
Lexical Overlap

1. **Weak** correlation of SWN with Whissel evaluation

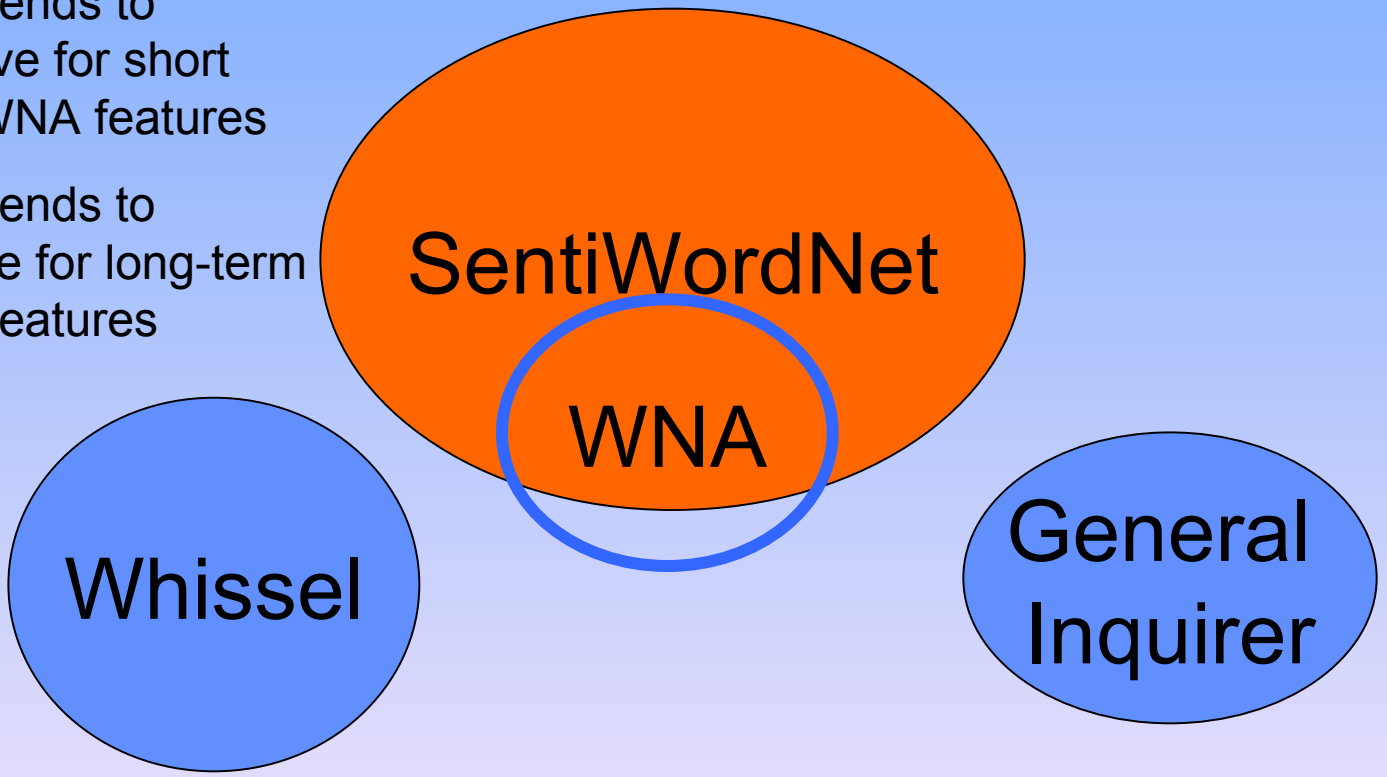
2. **No** correlation with Whissel activation dimension

3. SWN positive **negatively** correlated with imageability

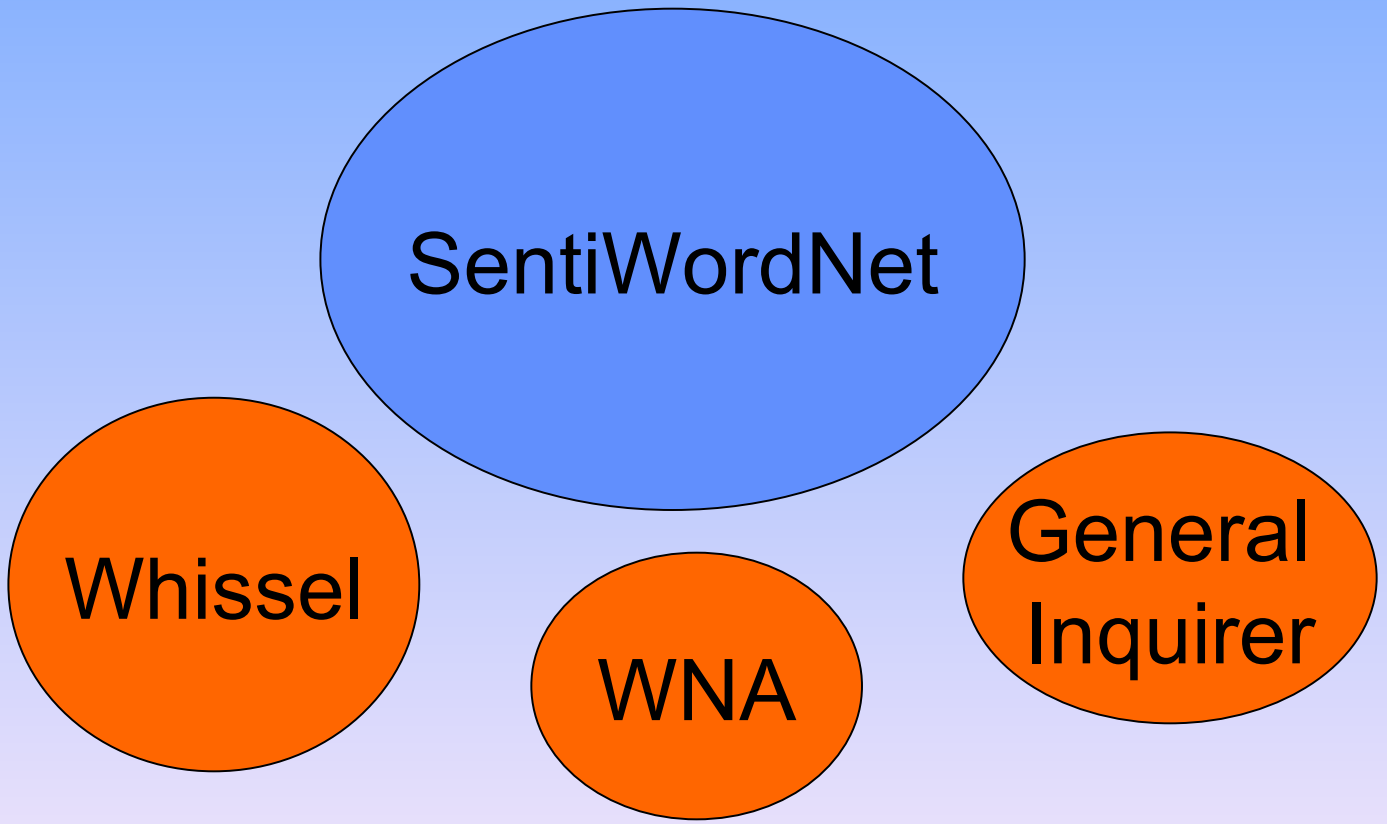


Lexical Resource Evaluation
Lexical Overlap

1. SWN tends to negative for short term WNA features
2. SWN tends to positive for long-term WNA features



Lexical Resource Evaluation
Lexical Overlap



Lexical Resource Evaluation
Lexical Overlap

- WNA feature division:

Short-term	**Long-term**
Negative	Positive
Physical	**Cognitive**
More active	Less active
Internal	**External**
Less abstract	More concrete

Lexical Resource Evaluation
Some conclusions

- The lexica:
  - Are quite consistent
  - Can be used in combination
- SentiWN: Largely unexplored territory

Comparative Corpus Analysis Aims

- Examine affective term use
- Identify statistically different distributions
- Is there a dominant feature/polarity?

Comparative Corpus Analysis
The Data

- Financial Language
  - 2 million words
  - On-line financial news:
    - Reuters, CNN, Bloomberg
    - Newspapers
- General Language

  BNC
  - 100 million words
  - Balanced, broad corpus

Comparative Corpus Analysis
The Data

- BNC sub-corpora

- Imaginative written English
  - 16 million words

- Informative written English
  - 70 million words

Comparative Corpus Analysis Methodology

- Compare proportions of Sentiment Features
- χ^2 Test of Independence
- H_0: $\pi_{\text{FinCorpus}} = \pi_{\text{BNC}}$

Comparative Corpus Analysis Methodology

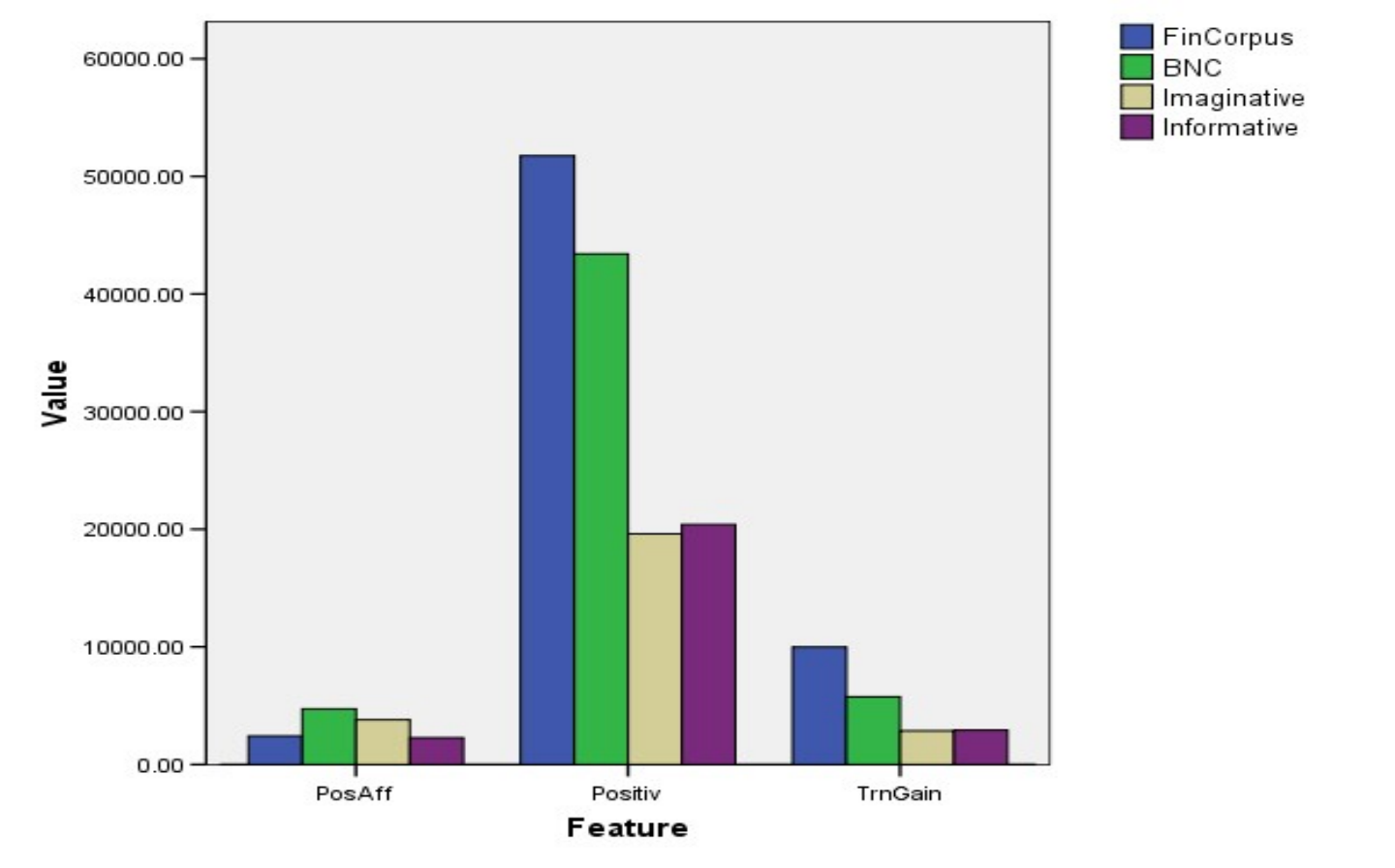
- Statistical significance of different proportion
  - $\chi^2 > 7.8794$
  - $p >= 0.005$

- Features:
  - 41 Lexicon Sentiment Features from 4 lexica
  - Frequency per million words

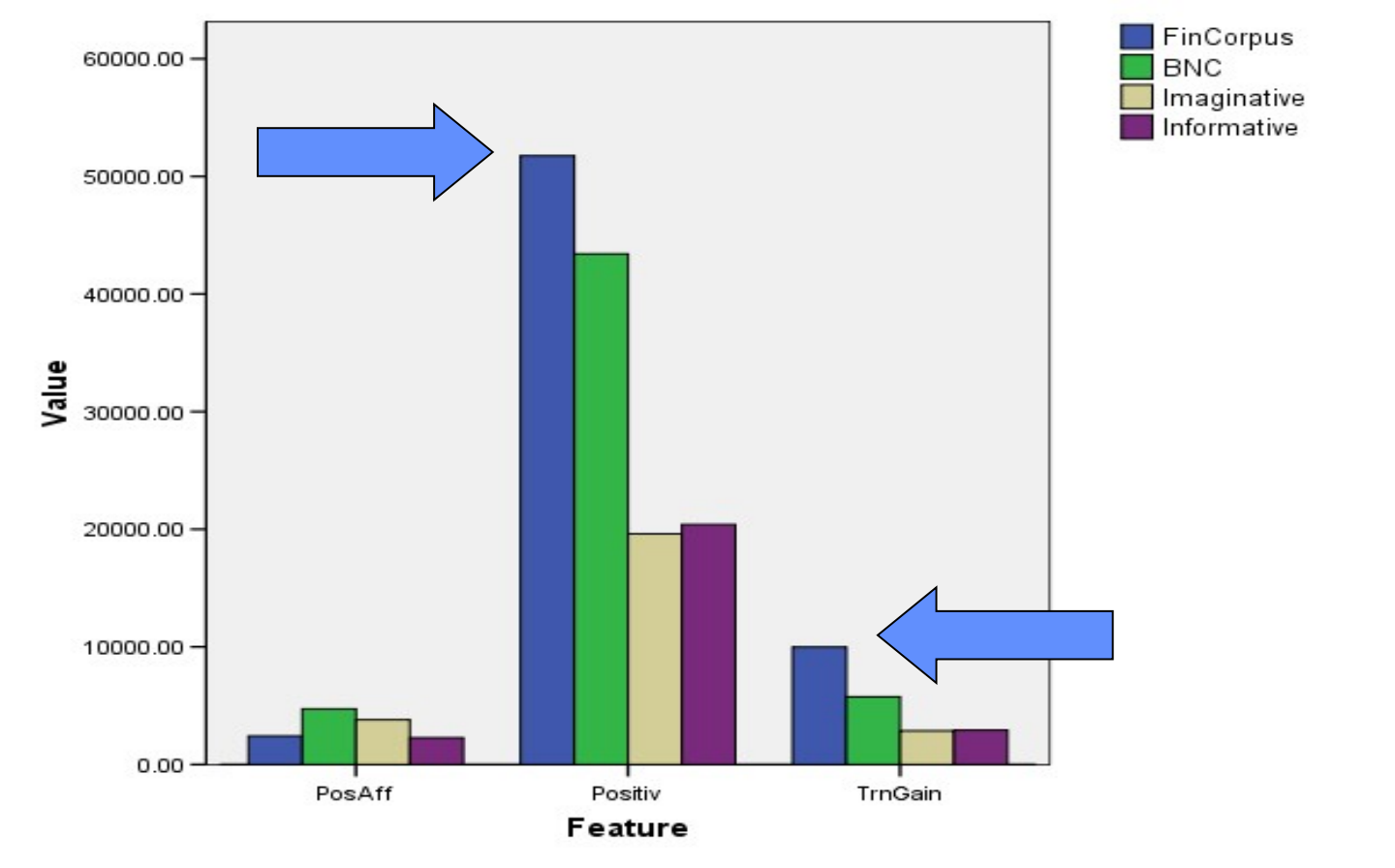
Comparative Corpus Analysis
Financial Corpus

- WRT Imaginative: More affective terms
- WRT Informative: Many more affective terms
- WRT BNC:
  - Dependent on feature type
  - Distributions are statistically distinct

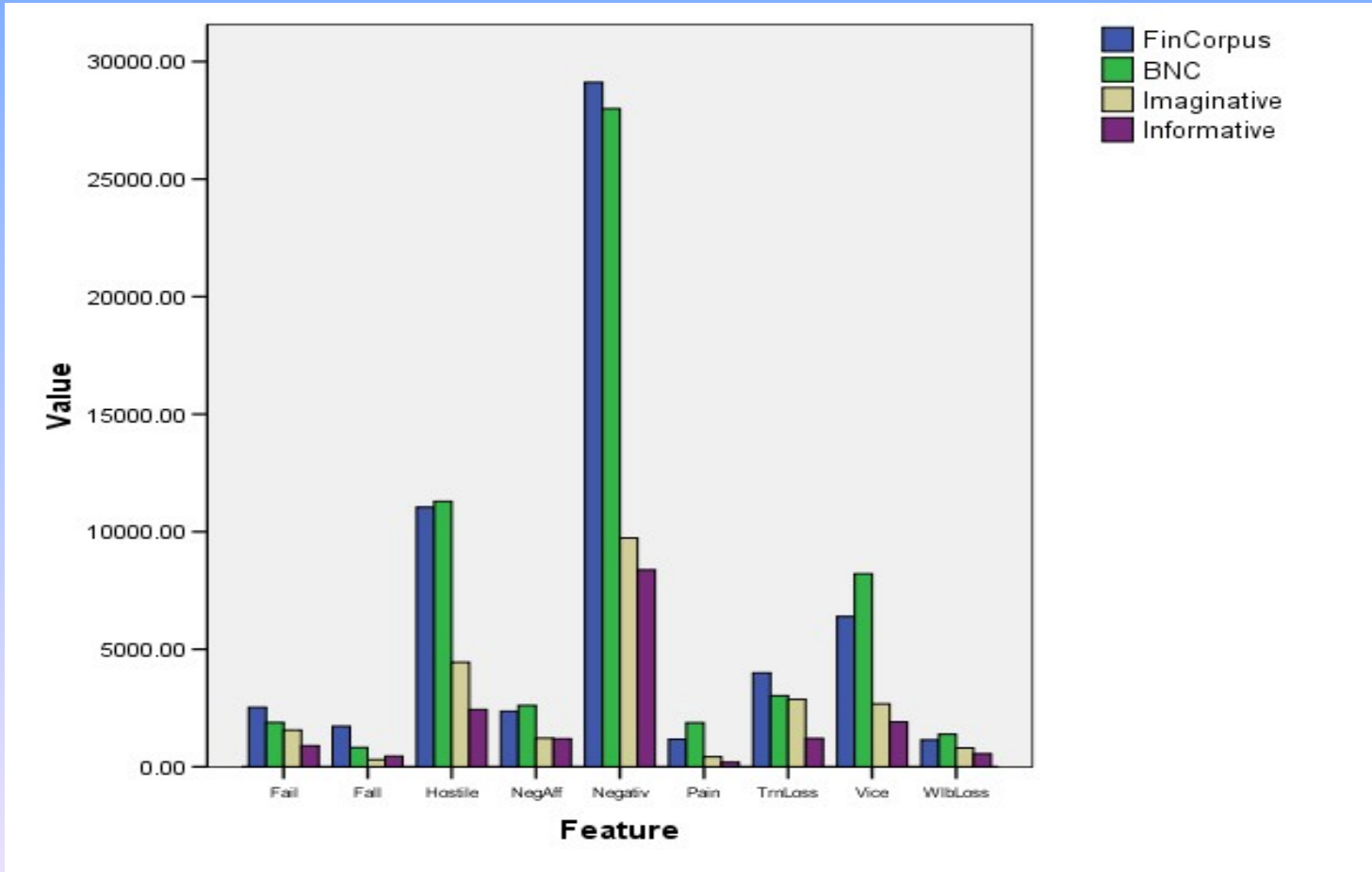
Comparative Corpus Analysis
Positive GI Features



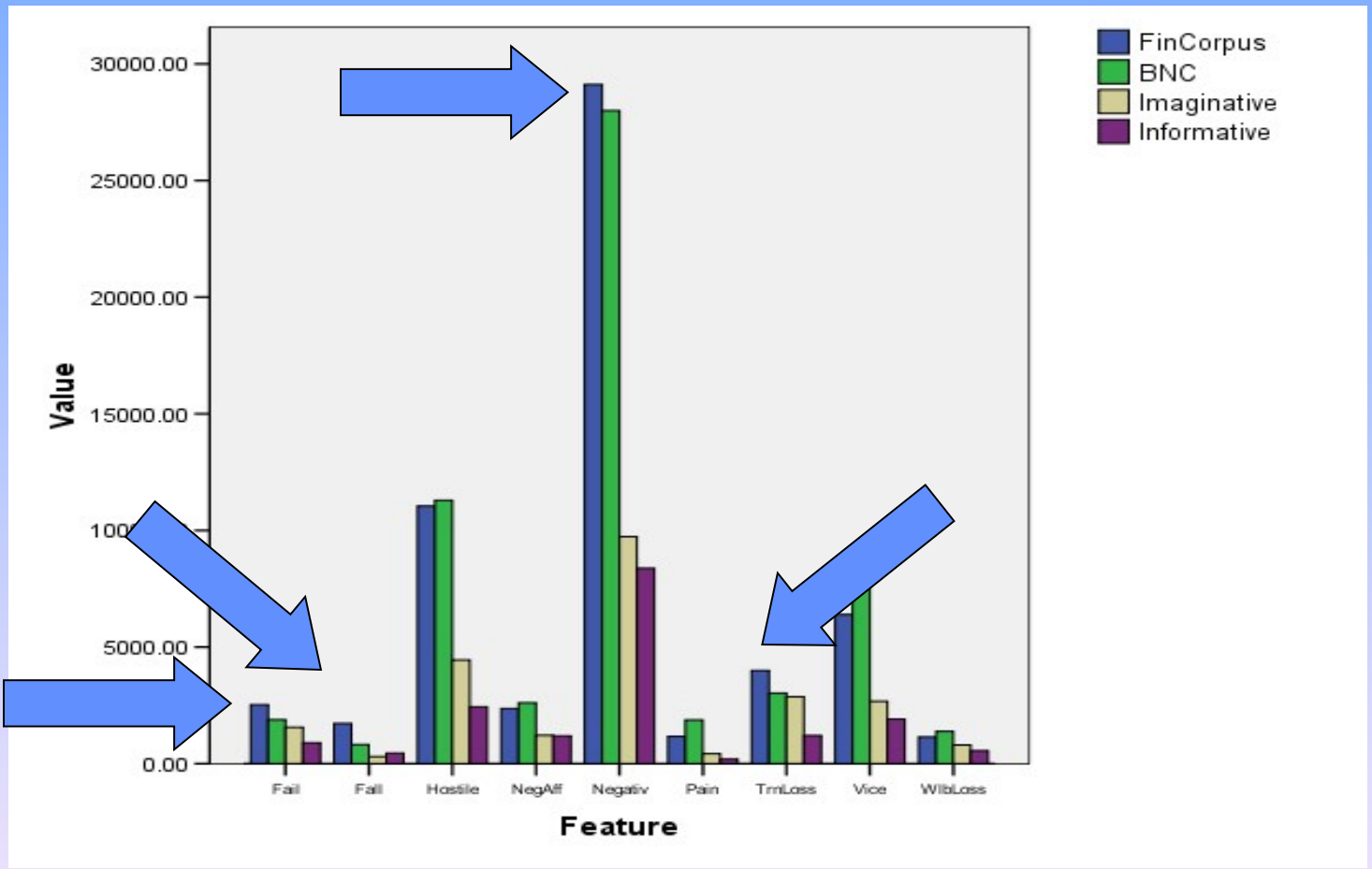
Comparative Corpus Analysis
Positive GI Features



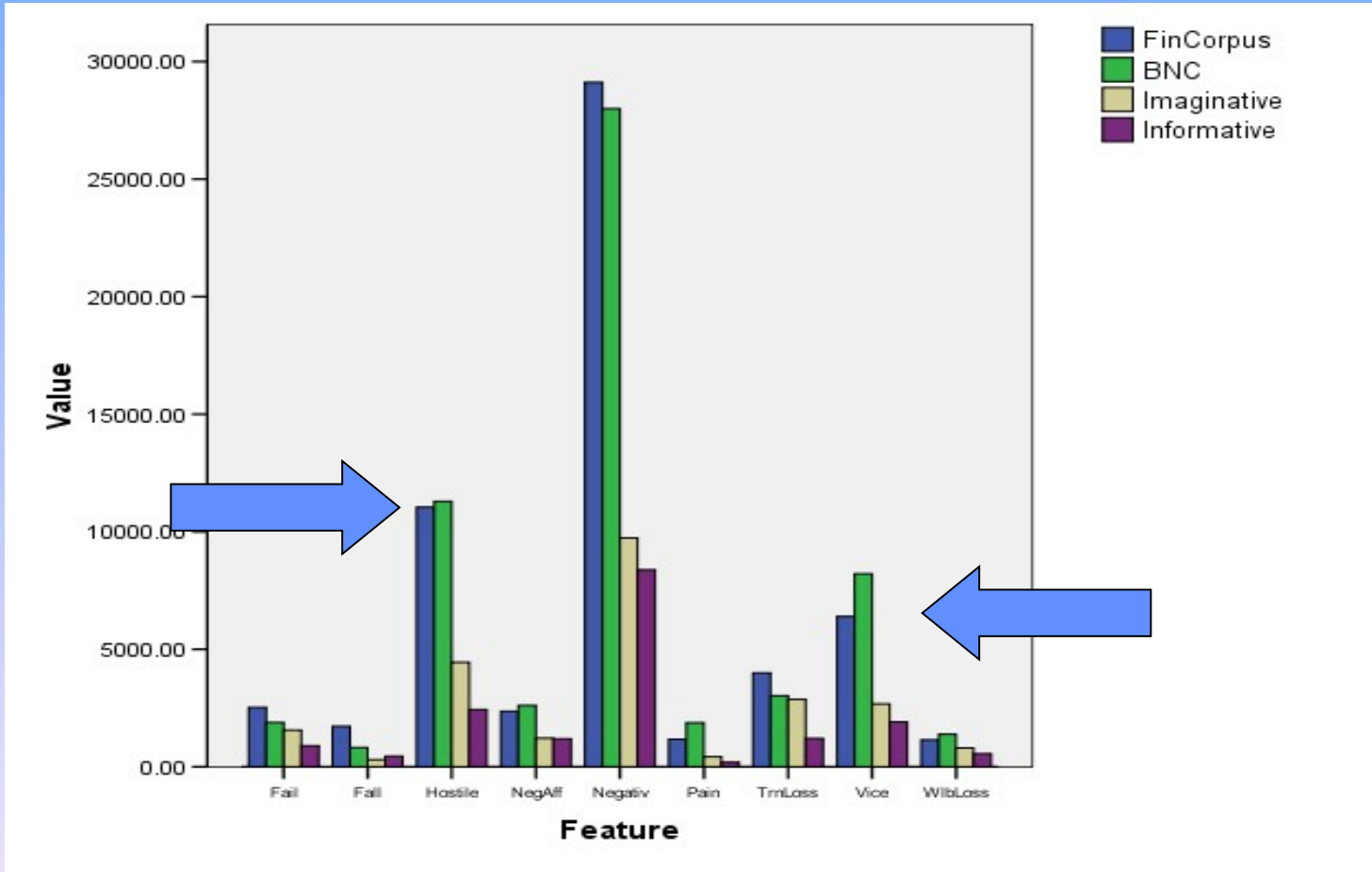
Comparative Corpus Analysis
Negative GI Features



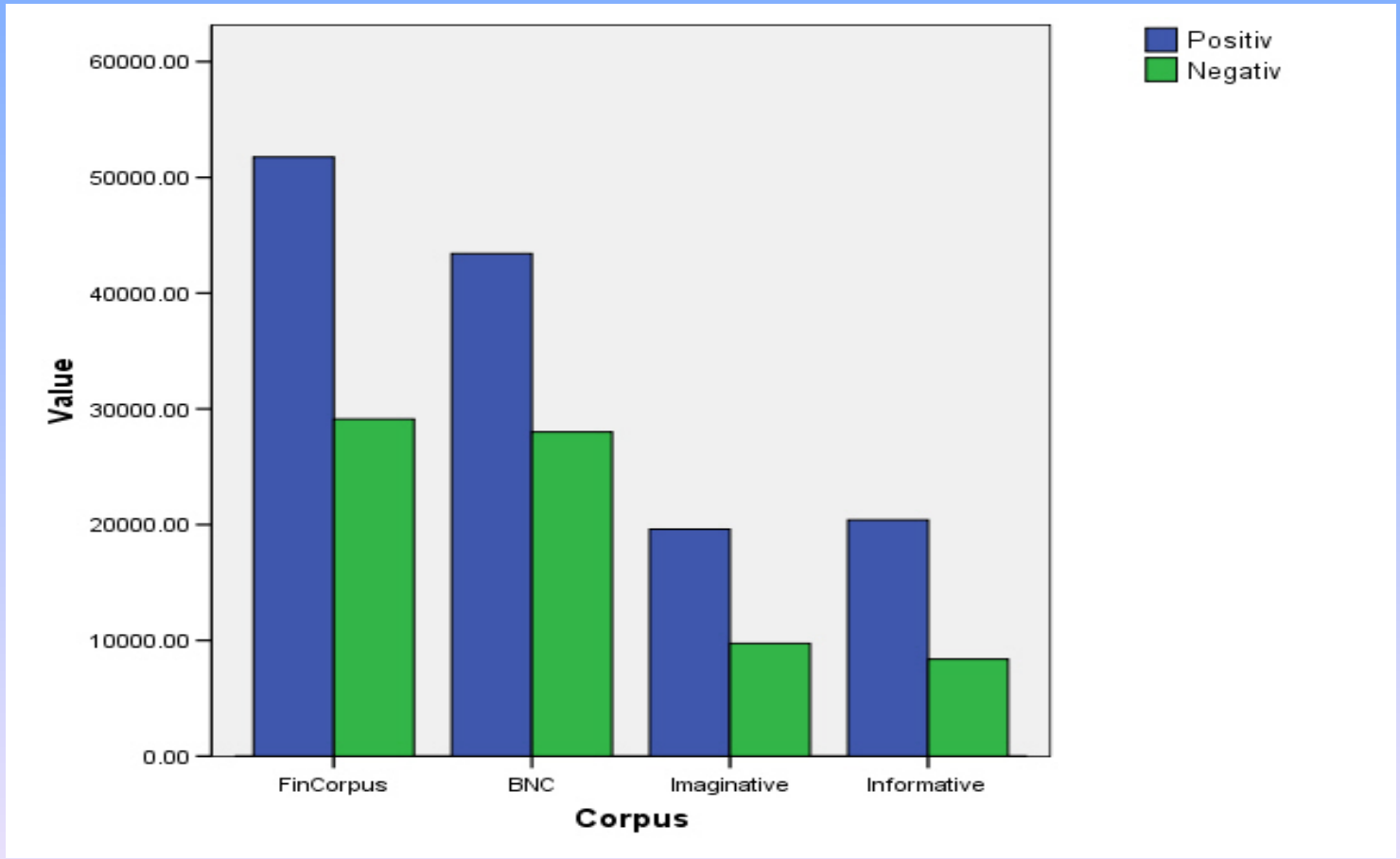
Comparative Corpus Analysis
Negative GI Features



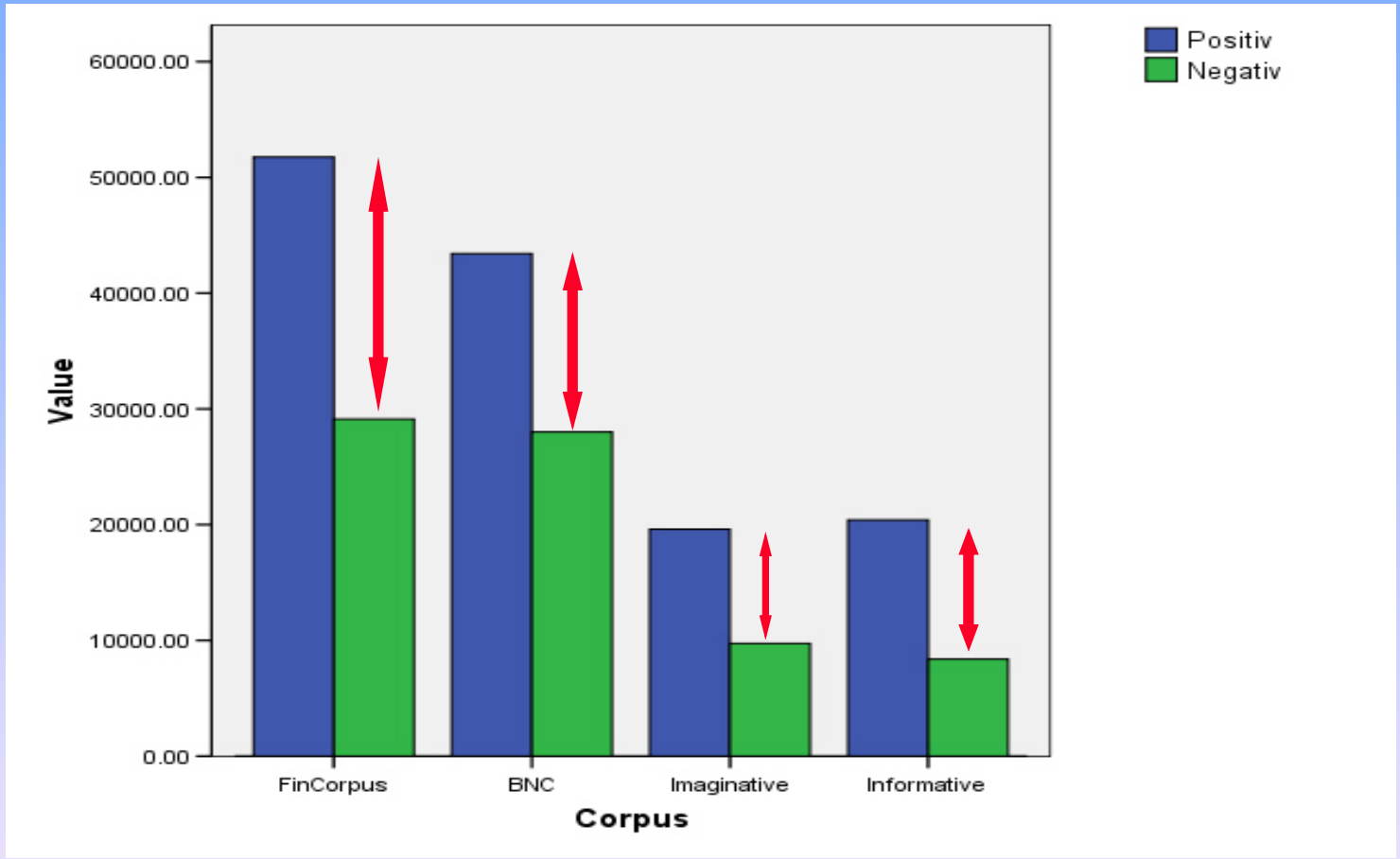
Comparative Corpus Analysis
Negative GI Features



Comparative Corpus Analysis: Sentiment Polarity Asymmetry



Comparative Corpus Analysis: Sentiment Polarity Asymmetry

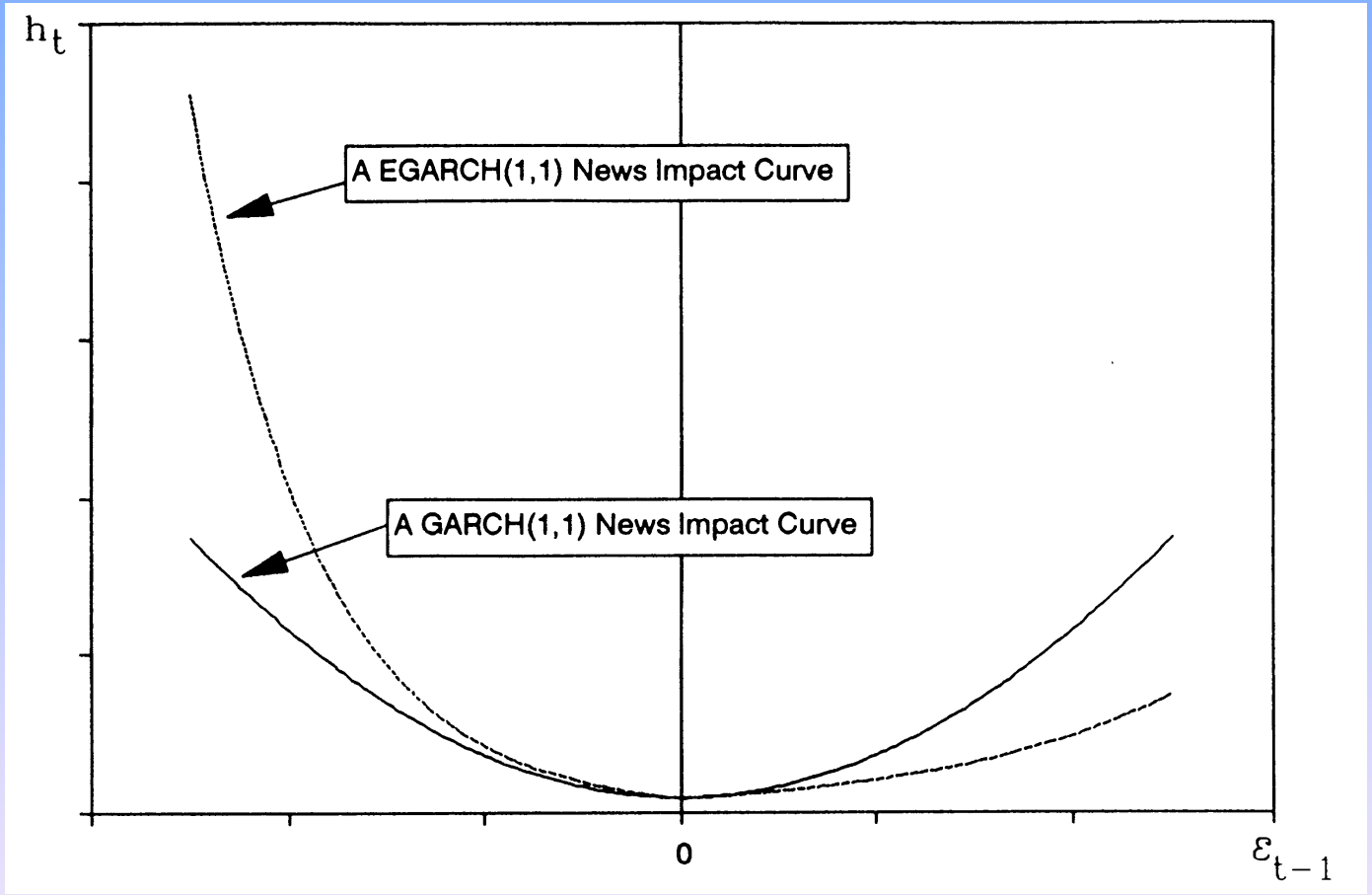


Ann Devitt
Trinity College Dublin

Afterword

- Lexical resources for sentiment are consistent
- Financial news is a sub-language:
  - Affective content is statistically distinct relative to general language
- Text polarity is asymmetric, positive skew

Engle Ng (1993) Asymmetry Curve



Afterword

- **Econometric analysis has suggested that the effect of news on financial transactions is *asymmetric* → negative news has longer lasting effect than the positive news.**
- Typically in econometrics, 'news' is interpreted in fairly narrow terms – restricted to the time of the announcement (scheduled/unscheduled) or the frequency of terms in a hand-crafted lexicon of sentiment words.
- **We have some confirmation of the asymmetry – both in general language and in financial news.**
- Our corpus-based work, and work based on a generic lexicon of emotion indicates perhaps that we might use such a lexicon rather than hand-crafting one.

Thank You!

Ann.Devitt@cs.tcd.ie