Anaphora resolution and coreference resolution: still a hard nut to crack? How far has it gone, what is its impact on NLP and what are the ways forward?

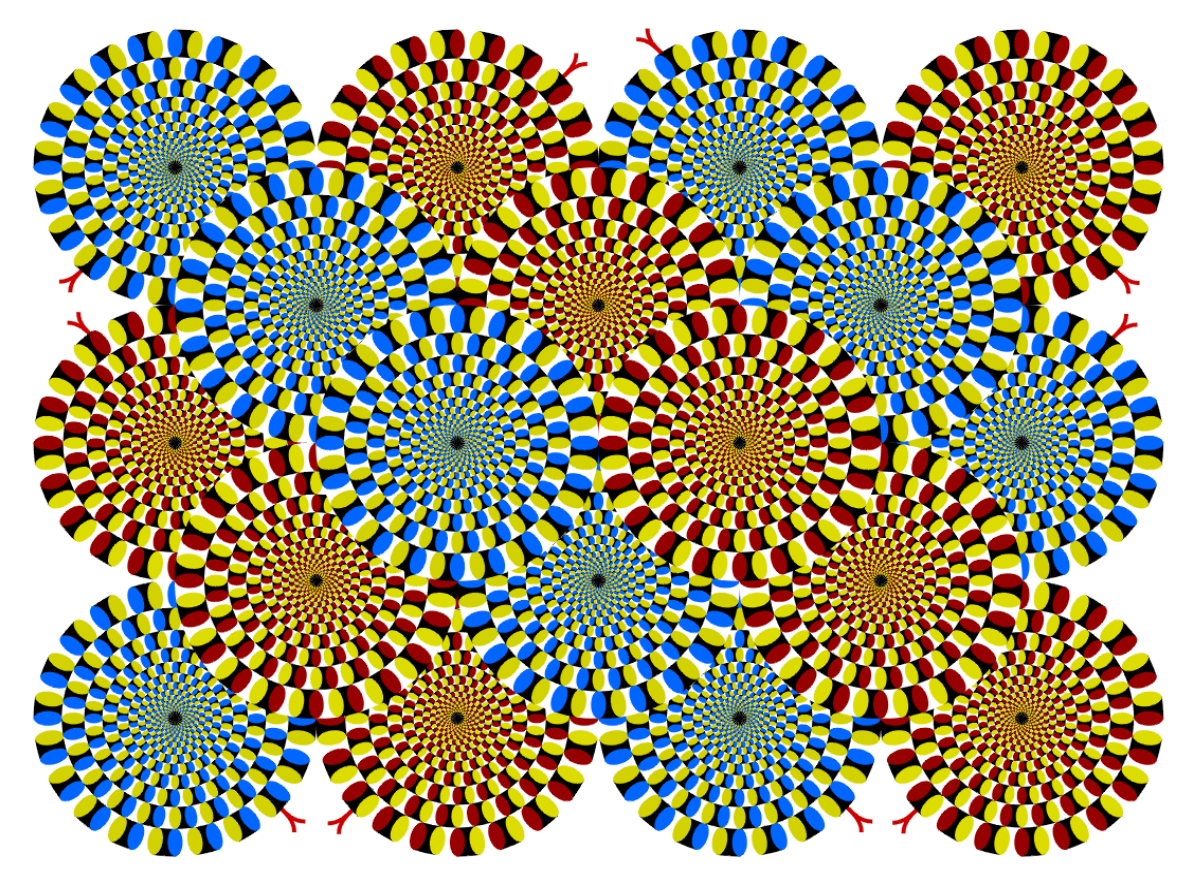
0011 0010 1010 1101 0001 0100 1011

Ruslan Mitkov

Research Group in Computational Linguistics

University of Wolverhampton

Outline of the presentation



- Terminology
- How far has anaphora resolution gone?
- The impact on NLP applications
- Ways forward and my latest research

5 minutes

Please excuse the speed of this presentation as we have to go through around 100

This presentation is partially based on….

- My invited talk at the EACL'2017 Workshop on Coreference Resolution beyond OntoNotes in Valencia

Anaphora and coreference resolution: three perennial questions

1. Are (automatic) anaphora resolution and coreference resolution beneficial to NLP applications?
2. Do we know how to evaluate anaphora resolution algorithms?
3. Which are the coreferential links most difficult to resolve?

Anaphora: basic notions and terminology

- **Cohesion**

- Sipping the last of the bitter cordial, *the businessman* was presented with the cheque. A look of incredulity crept over *his* face.
- Sipping the last of the bitter cordial, the businessman was presented with the cheque. A look of incredulity crept over her face.
- Sipping the last of the bitter cordial, the businessman was presented with the cheque. This lecture is about anaphora.

Basic notions and terminology (2)

- **Anaphora** (Haliday & Hasan 76): cohesion which points back to some previous item
- Pointing back, NOT *referring*!
- **Anaphor:** the "pointing back" word
- **Antecedent: t**he entity to which it refers or for which it stands
- **Anaphora resolution:** the process of determining the antecedent of an anaphor

Basic notions and terminology (3)

- **Coreference**: the act of picking out the same referent in the real world.
- Anaphors and antecedents are said to be **coreferential** if they have the same referent in the real world

Example

- *Sophia Loren* says *she* will always be grateful to Bono. *The actress* revealed that the U2 singer helped *her* calm down when *she* became scared by a thunderstorm while travelling on a plane.
- Coreferential chains: {Sophia Loren, she, the actress, her, she}, {Bono, the U2 singer}, {a thunderstorm}, {a plane}

Anaphora vs. coreference

- Anaphora and coreference are not identical phenomena
- Anaphora which is not coreference: identity of sense anaphora
- The man who gave his paycheck to his wife was wiser than the man who gave it to his mistress
- Coreference which is not anaphora:
- Cross-document coreference

Anaphora (and coreference) resolution

- Anaphora resolution: tracking down the antecedent of an anaphor
- Coreference resolution: identification of all coreference classes (chains).

Anaphora (and coreference) resolution (2)

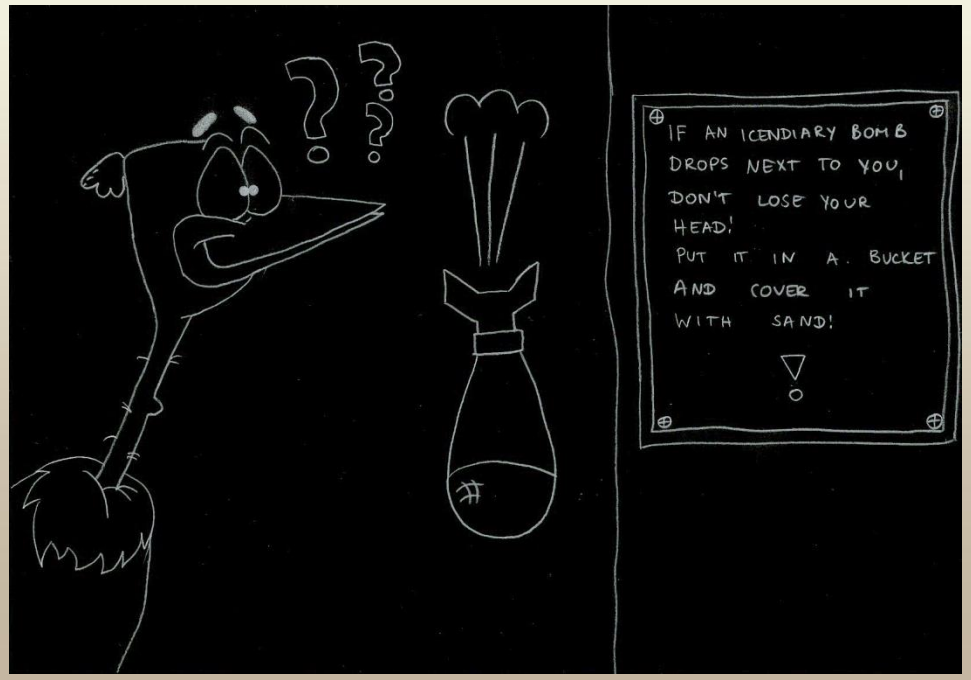
- Crucial for virtually every NLP application: Machine Translation, Summarisation, Information Extraction, Question Answering, Textual Entailment, Term Extraction…
- Incorrect identification of anaphoric (and coreferential) relations could be costly….

Anaphora and Machine Translation

- The monkey ate the banana because it was hungry.
- The monkey ate the banana because it was ripe.
- The monkey ate the banana because it was tea time.
- Der Affe hat die Banane gefressen, weil er hungrig war.
- Der Affe hat die Banane gefressen, weil sie reif war.
- Der Affe hat die Banane gefressen, weil es Zeit zum Abendessen war.

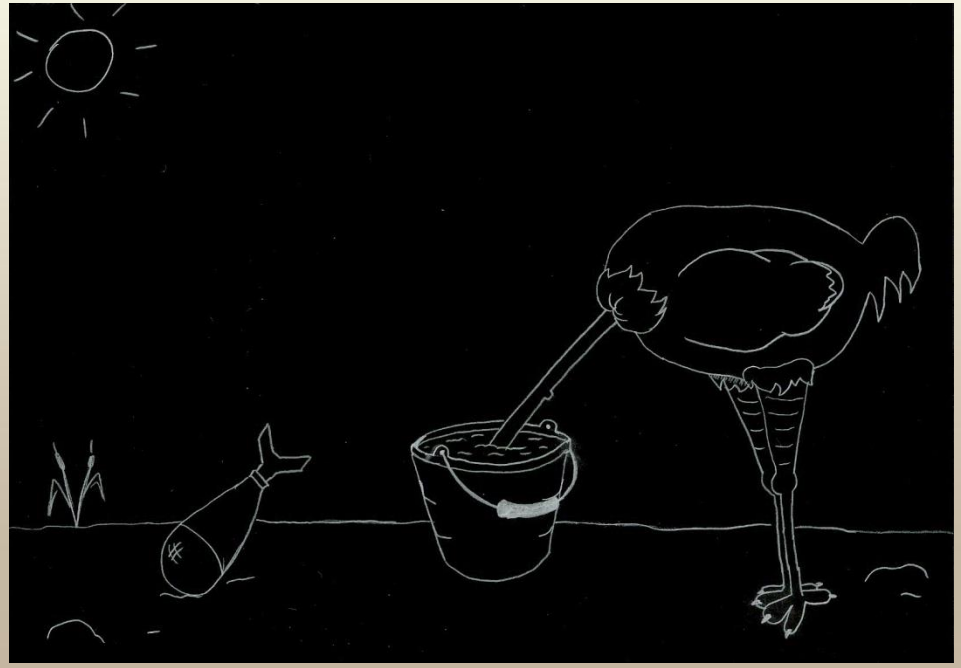
Anaphora and anaphora resolution (2)

- If an incendiary bomb drops next to you, don’t lose your head. Put it in a bucket and cover it with sand.

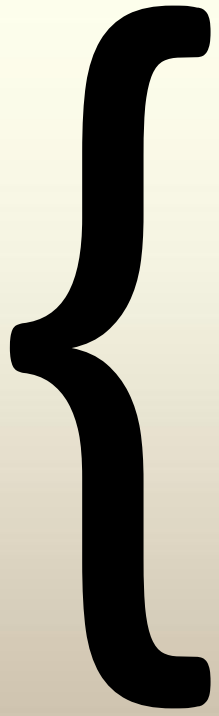


Anaphora and anaphora resolution (2)

- If an incendiary bomb drops next to you, don't lose your head. Put it in a bucket and cover it with sand.



How far has anaphora resolution gone (and the mystery of the evaluation results)



Intrinsic evaluation results

- MARS: success rate 45-65%
- Over this data: 46.63% (MARS'02), 49.47% (MARS'06)
- Our study of knowledge-poor approaches and full-parser approaches on 2,597 anaphors and 3 genres (Mitkov and Hallett 2007):
  - MARS: 57.03%
  - Kennedy and Boguraev: 52.08%
  - Baldwin's CogNIAC: 37.66%
  - Hobbs' naïve algorithm: 60.07%
  - Lappin and Leass RAP: 60.65%
  - Baselines: 30.07%-14.56%

The mystery of the original results

The mystery of the original results

- Differences between results presented in the original papers and the results obtained in our study
- Hobbs (1976): 31.63%
- Lappin and Leass (1998): 25.35%
- Boguraev and Kennedy (1996): 22.92%
- Mitkov (1996, 1998): 31.97%
- Baldwin (1997): 54.34%

Why are results so different?

- Different genres (computer science manuals: ill-structured)
- Procedure fully automatic
- Lack of domain-specific NER

The issue of complexity of evaluation data

- Some evaluation data may contain anaphors which are more difficult to resolve such as
  - anaphors that are ambiguous and require real-world knowledge
  - anaphors that have a high number of competing candidates
  - anaphors that have their antecedents far away
- Other data may have most of their anaphors with single candidates for antecedent ⇒
- Resolution complexity has to be quantified for every evaluation data

Quantifying the complexity via the evaluation workbench

- Average referential distance in NPs between the anaphor and its antecedent (for each sample or all anaphors)
- Average referential distance in sentences between the anaphor and its antecedent (for each sample or all anaphors).

Mysteries in evaluation

No sufficient evaluation details

Not clear what is the degree of automation of the system

Transparency, honesty?

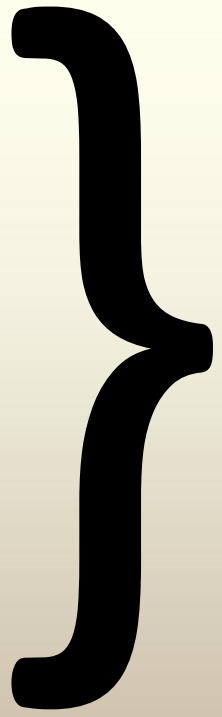
Objectivity?

- How objective is evaluation?
- How objective are (annotated) corpora?
- How objective/reliable is human judgement?
- Interannotator agreement can be as low as 60% (Mitkov et al. 2000)

Reluctance...

- ... to publish modest or negative results
- Publishing negative results is also worthwhile!

The mystery of the evaluation results



Evaluation in anaphora resolution: status quo

- Intrinsic evaluation: accounts for the performance of an algorithm or system
- Extrinsic evaluation: accounts for the impact of an algorithm or a system within a wider architecture/application
- Anaphora resolution research has focused almost exclusively on intrinsic evaluation

Objectives of Study 1

- To integrate a pronoun resolution system (MARS) within 3 NLP applications (text summarisation, term extraction, text categorisation)
- To evaluate these applications with and without a pronoun resolution module
- To establish of impact of pronoun resolution on these NLP applications

Objectives of Study 2

- To integrate a coreference pronoun resolution system (BART) within 3 NLP applications (text summarisation, text categorisation, recognising textual entailment)
- To evaluate these applications with and without the coreference resolution module
- To establish of impact of coreference resolution on these NLP applications

Study 1

- Mitkov's knowledge-poor pronoun resolution algorithm
- MARS'02 and MARS'06
- Evaluation data
- Extrinsic evaluation results
- Discussion
- Recent related research

Study 1

- Mitkov's knowledge-poor pronoun resolution algorithm
- MARS'02 and MARS'06
- Evaluation data
- Extrinsic evaluation results
- Discussion
- Recent related research

MARS – fully automatic pronoun resolution

Based on Mitkov's (1996; 1998) knowledge-poor approach:

- Text processed by a part-of-speech tagger and an NP extractor
- Locates noun phrases which precede the anaphor within a distance of 2 sentences
- Checks gender and number agreement
- Applies antecedent indicators to remaining
- Noun phrase with highest composite score proposed as antecedent

Antecedent indicators

- Boosting indicators apply a positive score to an NP, reflecting a positive likelihood that it is the antecedent of the current pronoun.
- Impeding ones apply a negative score to an NP, reflecting a lack of confidence that it is the antecedent of the current pronoun.

Example boosting indicator

- First Noun Phrases: A score of +1 is assigned to the first NP in a sentence.
- Collocation Match: A score of +2 is assigned to those NPs that have an identical collocation pattern to the pronoun.

Press the key down and turn the volume up... Press it again.

Example impeding indicator

- Prepositional Noun Phrases: NPs appearing in prepositional phrases are assigned a score of -1.

  Insert the cassette into the VCR making sure it is suitable for the length of recording.

Study 1

- Mitkov's knowledge-poor pronoun resolution algorithm
- MARS'02 and MARS'06
- Evaluation data
- Extrinsic evaluation results
- Discussion
- Recent related research

Improvements in MARS'2002

Three new indicators:

- Boost pronoun
- Syntactic parallelism (FDG parser)
- Frequent candidates
- Other indicators implemented differently (term preference, relative distance)

Improvements in MARS'2002 (2)

- Incorporation of a program for identifying non-nominal anaphora
- Incorporation of a program for animacy identification
- Implementation of intra-sentential syntax constraints (Lappin and Leass 1994; Kennedy and Boguraev 1996)

Improvements in MARS (2006)

- MARS'2006 caters for number and gender disagreement
  i. Collective nouns
  ii. NPs whose gender is underspecified
  iii. Quantified nouns/indefinite pronouns
  iv. Organisation names
- Example
  – If there is *a doctor* on board, could they please make *themselves* known to the crew
- Selectional restriction preference

Study 1

- Mitkov's knowledge-poor pronoun resolution algorithm
- MARS'02 and MARS'06
- Evaluation data
- Extrinsic evaluation results
- Discussion
- Recent related research

Evaluation data

- Newspaper articles published in *New Scientist* (55 texts from BNC)
- Short enough to be manually annotated
- Suitable for all extrinsic evaluation tasks performed
- Articles manually categorised into six classes – “Being Human”, “Earth”, “Fundamentals”, “Health”, “Living World”, and “Opinion”
- Caution: MARS was not specially tuned to these genres!

Evaluation data (2)

- 1,200 3rd person pronouns; over 48,000 words
- Very short and very long texts filtered out
- Annotation: PALinkA (Orasan, 2003)
- Several layers of annotations:
  - Coreference
  - Important sentences
  - Terms
  - Topics

Study 1

- Mitkov's knowledge-poor pronoun resolution algorithm
- MARS'02 and MARS'06
- Evaluation data
- Extrinsic evaluation results
- Discussion
- Recent related research

Study 1

- Mitkov's knowledge-poor pronoun resolution algorithm
- MARS'02 and MARS'06
- Evaluation data
- Extrinsic evaluation results
- Discussion
- Recent related research

Extrinsic evaluation

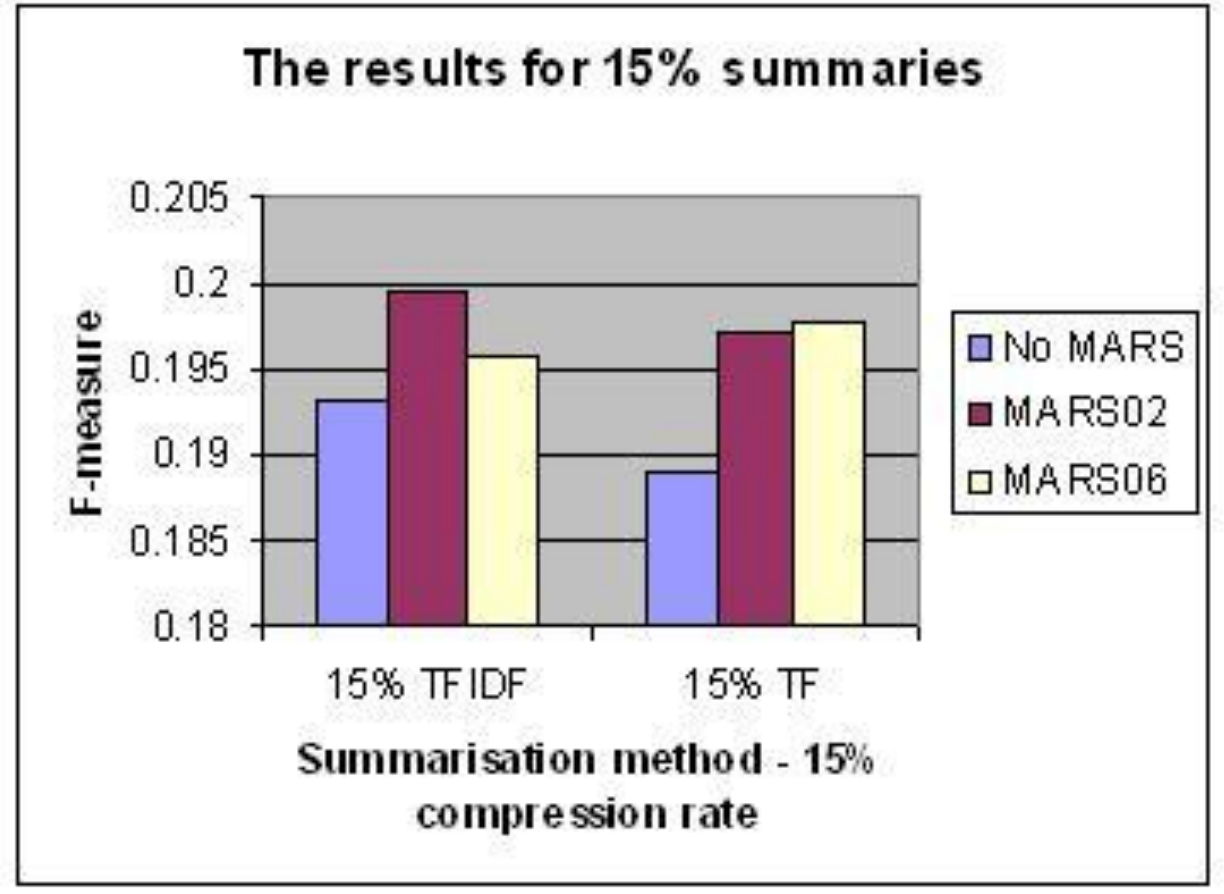
- Text summarisation
- Term extraction
- Text categorisation

Text summarisation

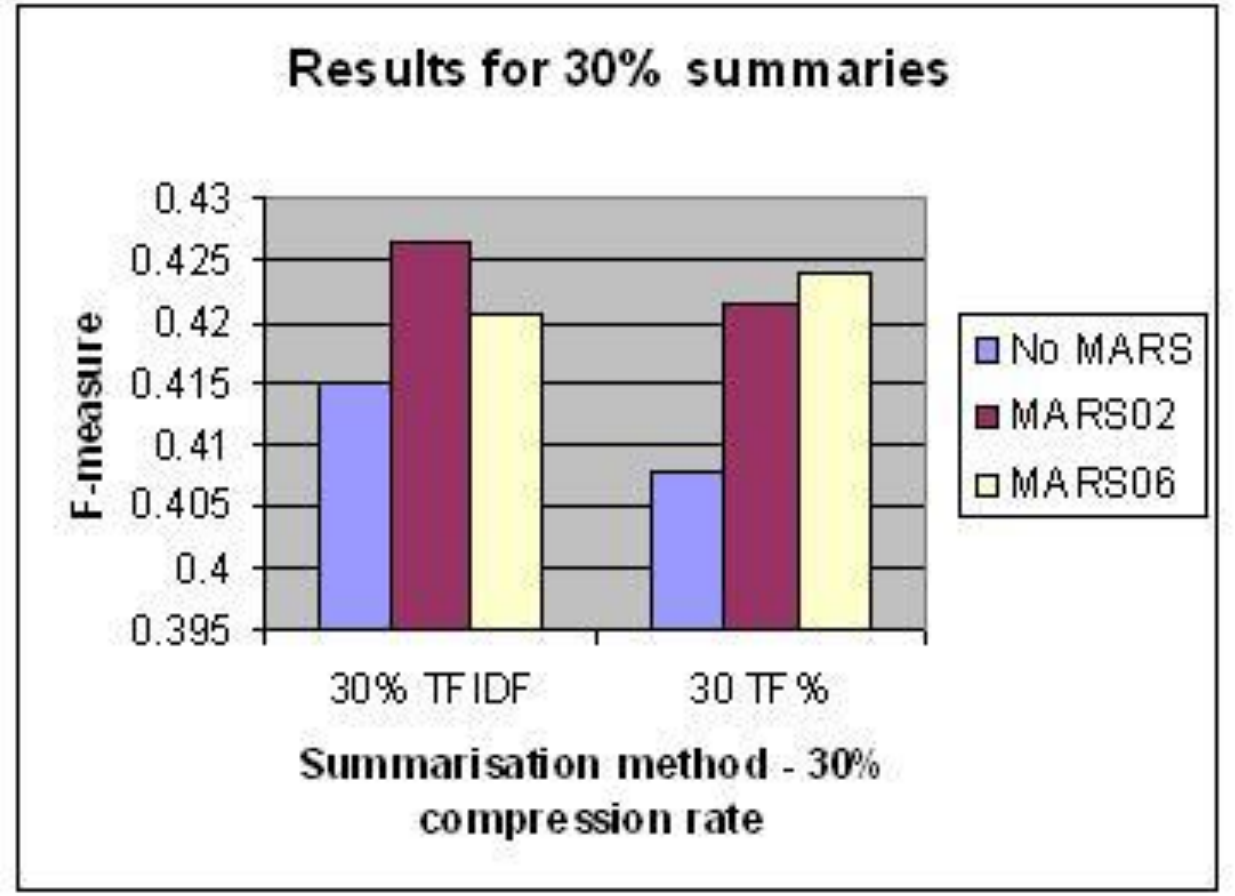
Summarisation

- Two term weighting methods investigated: term frequency and TF*IDF
- Evaluation measures: precision, recall and F-measure
- Evaluation performed for two (15% and 30%) compression rates

Summarisation (2)



Summarisation (3)

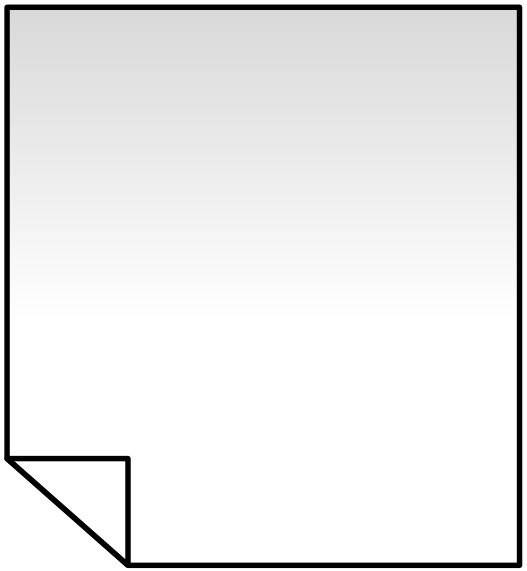


Summarisation (4)

- F-measure increases when anaphora resolution method employed
- Increase not statistically significant (T-test)
- Term frequency: results better for MARS'06
- TF.IDF: results better for MARS'02

Term extraction

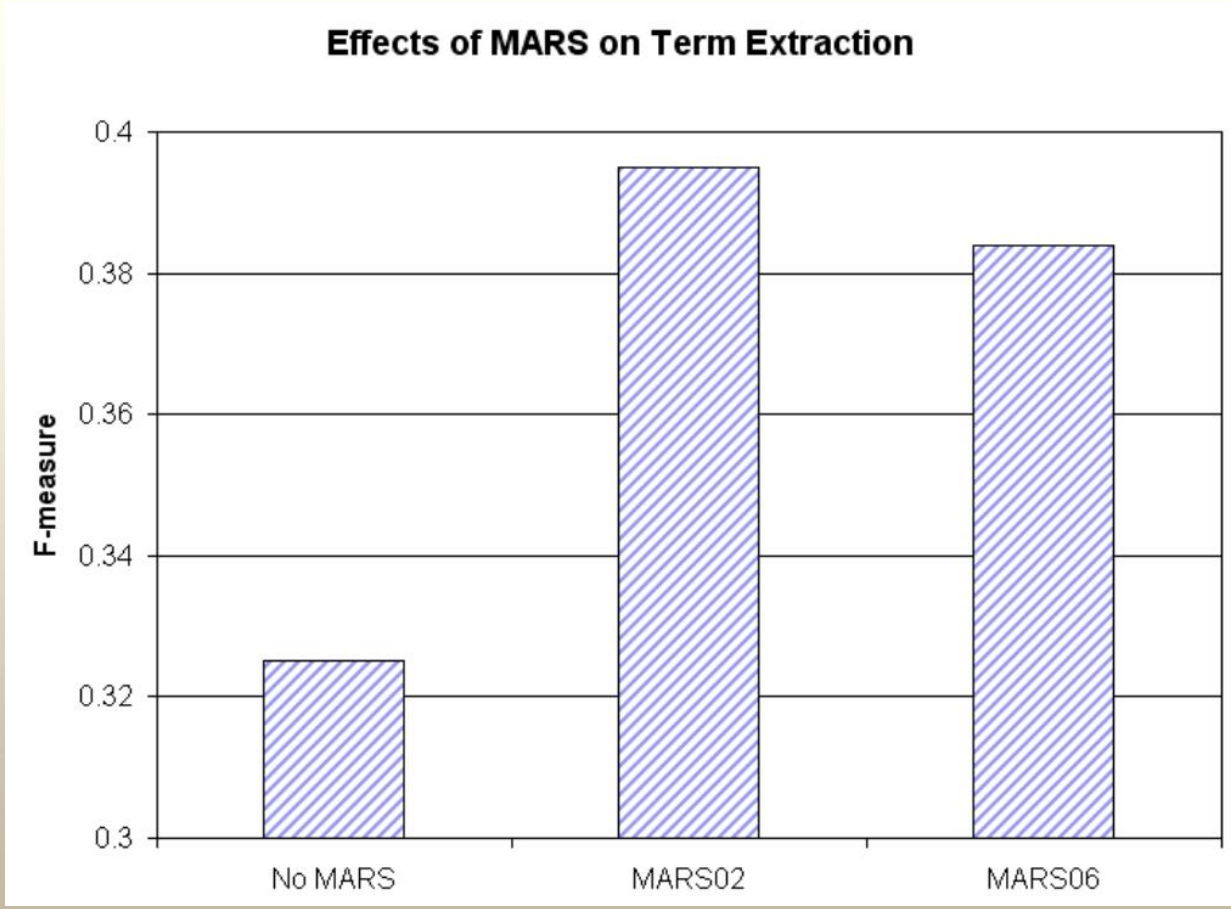
Natural language processing (NLP) is a field of computer science, artificial intelligence and linguistics concerned with the interactions between computers and human (natural) languages.



Term extraction

- Hybrid approach which combines statistical and lexical-syntactic filters in line with (Justeson and Katz 1986) and (Hulth 2003).
- Evaluation measures: precision, recall and F-measure.

Term extraction (2)



Term extraction (3)

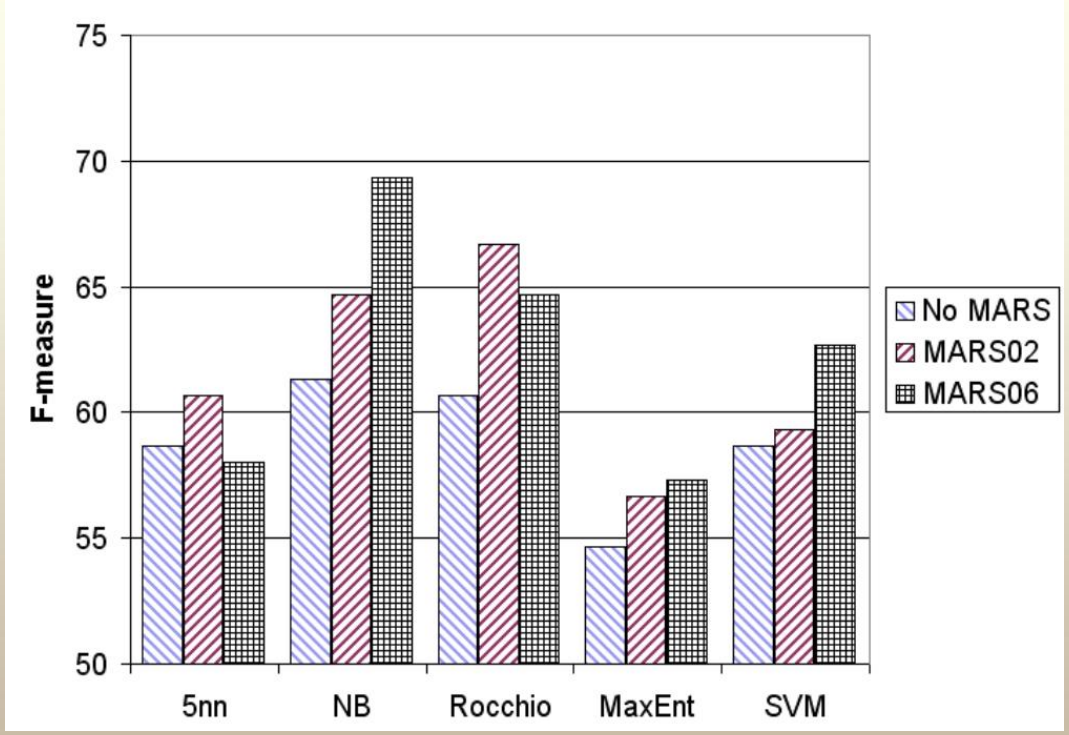
- F-measure increases when anaphora resolution method employed
- Increase not statistically significant (T-test)
- MARS'02 fares better in general
- MARS'02 improves both precision and recall
- MARS'06 improves mostly recall

Text categorisation

Text categorisation

- 5 different text classification methods: k nearest neighbours, Naïve Bayes, Rocchio, Maximum Entropy, and Support Vector Machines.
- Evaluation measures: precision, recall and F-measure

Text categorisation (2)



Text categorisation (3)

- F-measure increases in *most cases* when anaphora resolution method employed
- Increase not statistically significant for any of the methods

Structure of the presentation

- Mitkov's knowledge-poor pronoun resolution algorithm
- MARS'02 and MARS'06
- Evaluation data
- Extrinsic evaluation results
- Discussion
- Recent related research

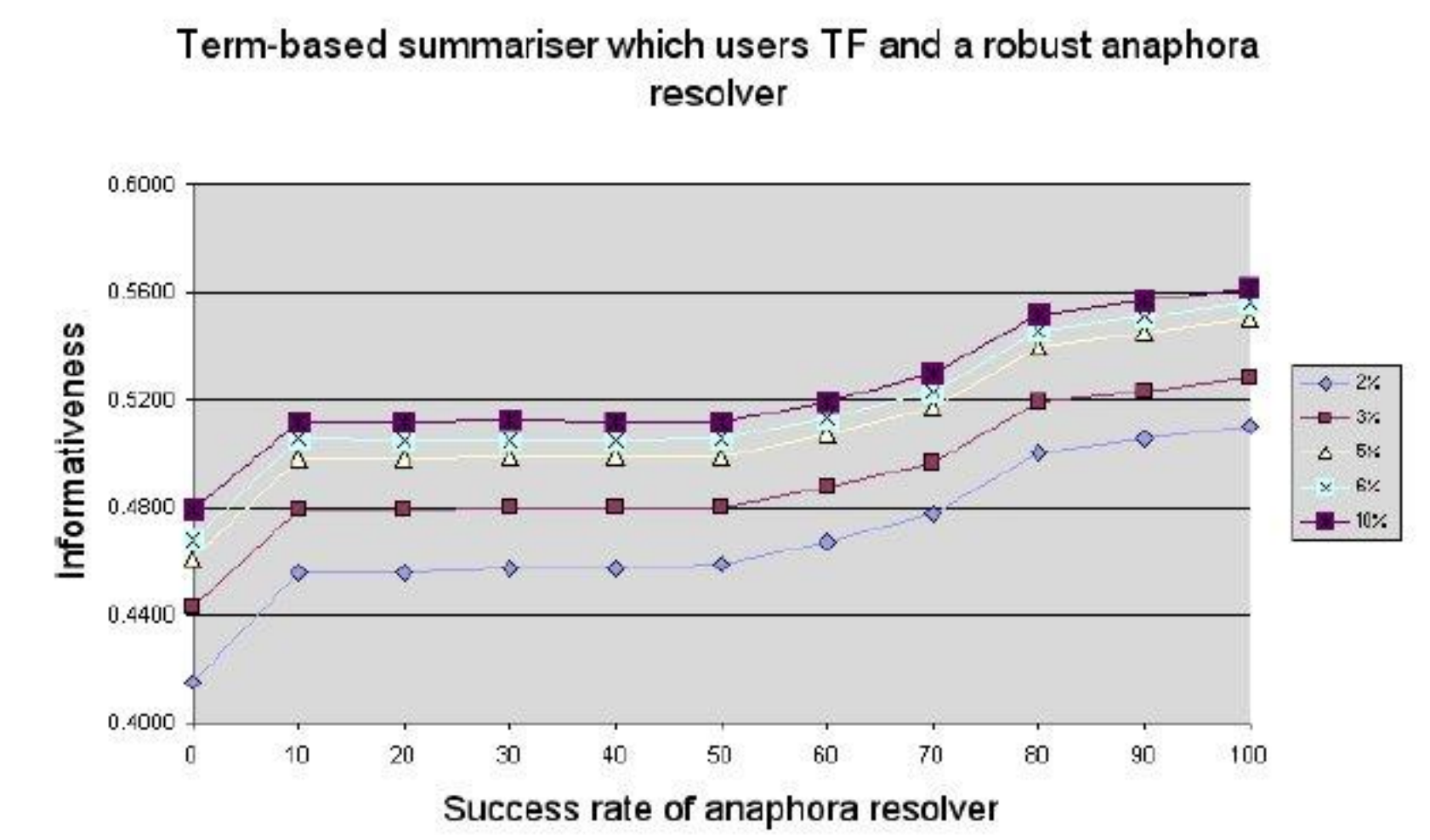
Discussion

- By and large deployment of MARS has positive but limited impact
- Slight improvement in performance of MARS'06 over MARS'02
- Observation in Mitkov's DAARC'98 paper confirmed: unpredictability of anaphora resolution with regard to difference of data/files
- Would dramatic improvement in anaphora resolution lead to a marked improvement of NLP applications?

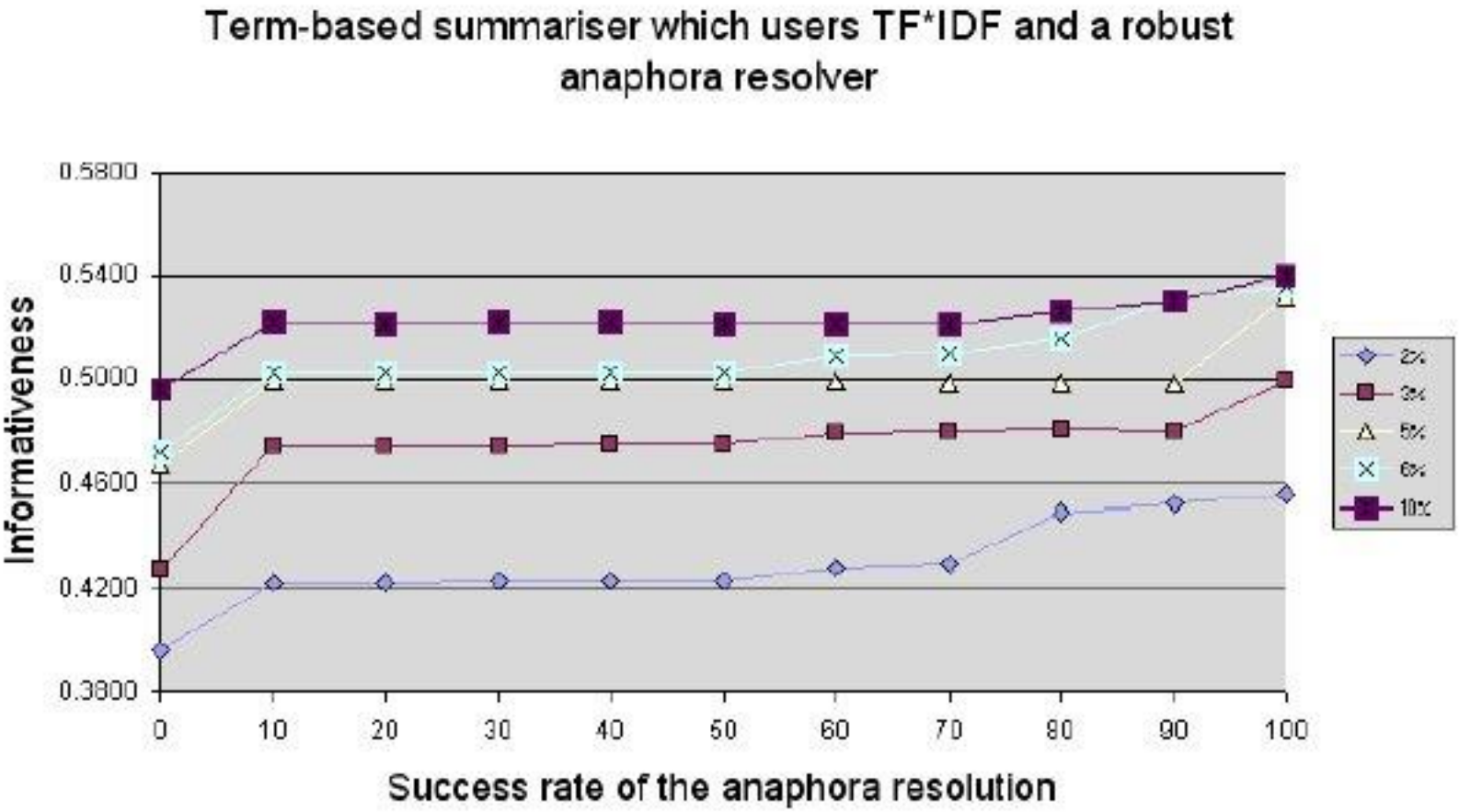
Would dramatic improvement in anaphora resolution lead to a marked improvement of NLP applications?

- Experiments on text summarisation (Orasan 2006)
- On a corpus of scientific articles anaphora resolution helps ….
  - TF summarisation if performance over 60-70%
  - TF.IDF summarisation if performance above 80%

Would dramatic improvement in anaphora resolution lead to a marked improvement of NLP applications? (2)



Would dramatic improvement in anaphora resolution lead to a marked improvement of NLP applications? (3)



Study 2

- BART coreference resolution system
- Text summarisation experiments
- Text categorisation experiments
- Recognising textual entailment experiments

The impact of coreference resolution on NLP applications

- Investigating the impact on:
  - Text summarisation
  - Text classification
  - Textual entailment

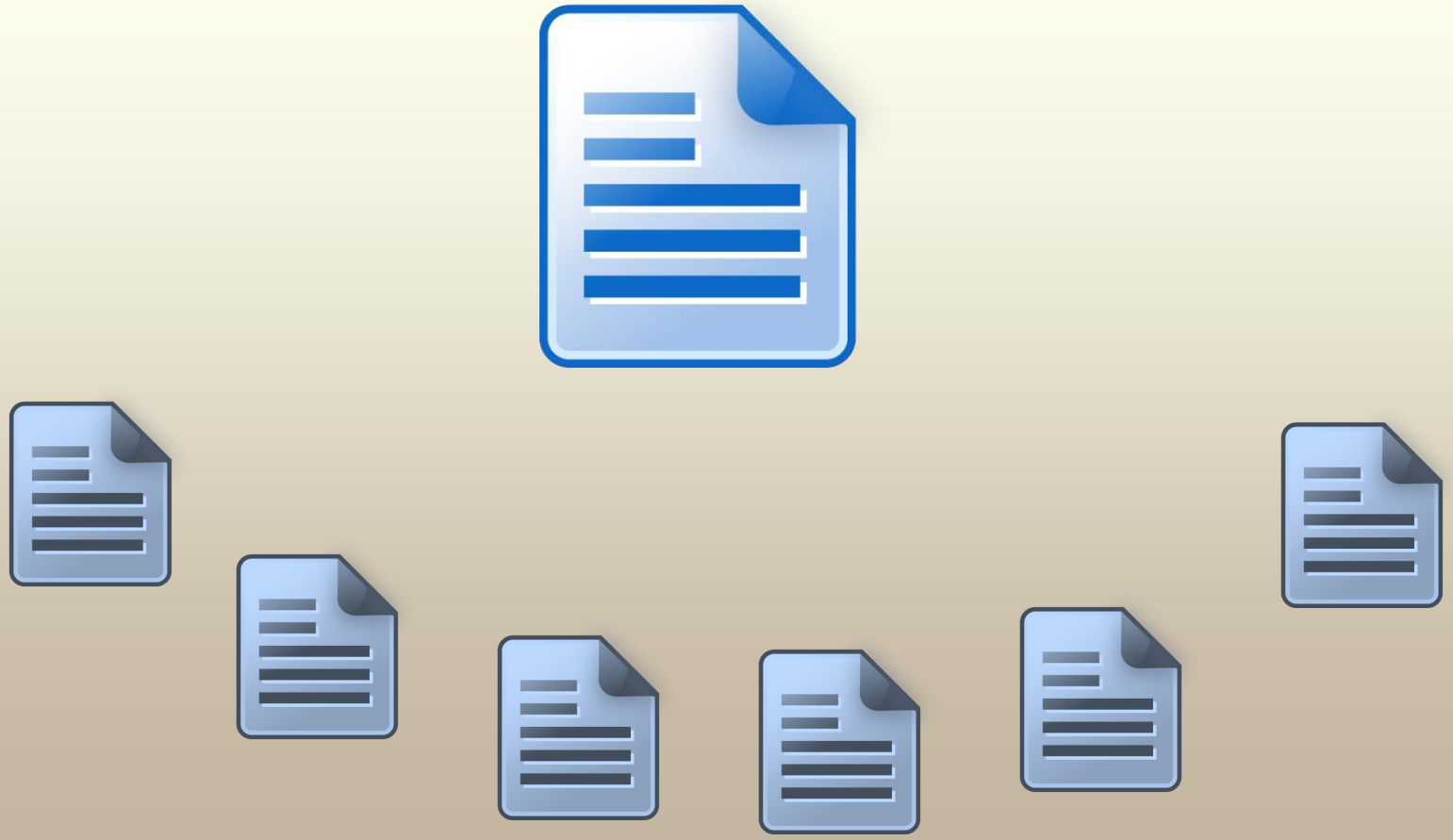
BART Toolkit

- The BART toolkit was used in the current work
  - Offers state of the art performance in coreference resolution, particularly in the resolution of pronouns
    - Reports a recall of 73.4% in pronoun resolution

BART Toolkit

- BART's algorithm:
  - First detects references:
    - Pronouns, noun chunks, base NPs, named entities
  - For each anaphor:
    - Extracts pairs consisting of the anaphor and each potential antecedent
      - Pairs are represented using a feature set that includes:
        - features proposed by Soon et al. (2001)
        - features encoding the syntactic relation between the anaphor and the potential antecedent
        - Features based on knowledge extracted from Wikipedia
  - Machine learning is then used to classify the pairs as coreferential or otherwise.

Text summarisation



The summarisation experiment

- Information from coreference resolver is used to increase score of each sentence by
  - Setting 1: score of longest mention in chain
  - Setting 2: highest score of mention in chain

  for each coreferential chain traversing the sentence

- Chains with one element (singletons) discarded

- Score of words calculated using their frequency in document without any morphological processing and with the stopwords filtered

The summarisation experiment (II)

- Corpus:
  - 89 randomly selected texts from the CAST corpus (http://clg.wlv.ac.uk/projects/CAST/corpus/)
  - Each text annotated with information about the importance of each sentence:
    - 15% marked as ESSENTIAL
    - a further 15% marked as IMPORTANT
- Evaluation:
  - Precision, recall, f-measure
  - Produced summaries of 15% and 30% compression rate

Results and discussion
summarisation experiment

Compression rate	15%	30%
Without BART	32.88%	46.34%
With BART – setting 1	28.62%	45.88%
With BART – setting 2	27.14%	45.19%

- Performance of summarisation decreases when coreference information is added
- Drop is less for 30% summaries
- Decrease in performance can be explained by the errors introduced by the coreference resolver

Text categorisation

Text classification experiments

- Documents represented as weighted feature vectors (bag of words representation, Sebastiani, 2002)
  - stop words removed, Porter's stemming
  - both unigrams and bigrams used
  - *tfidf* weights
  - feature selection χ^2 and minimum occurrences; only 10% of features used
- Coreference information:
  - Terms t_k which occur in a coreference chain c are given more weight:

$$\mathrm{tf}^{coref}(t_k, d_j) = \mathrm{tf}(t_k, d_j) + \sum_{c \in C_k} \mathrm{len}(c)$$

- Binary SVM models (one-vs-all) (Joachims, 1998)

Text classification experiments (2)

- Evaluation corpus: a subset of Reuters-21578
  - ModApte contains 10,788 documents
  - Only use the 10 most frequent categories (R10)
- Runs:
  - run-bow: standard *tfidf* weights
  - run-bart: boost weights of terms in coref. chains
- Difference is not statistically significant (McNemar's test)

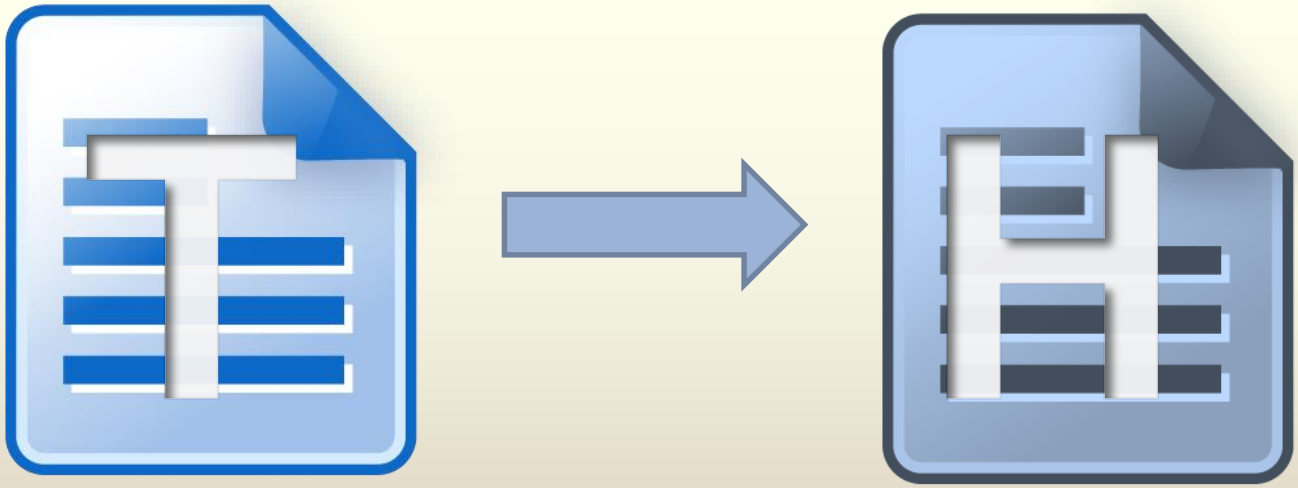
Results and Discussion

Text classification experiments

	P	R	F1
run-bow	95.59%	60.89%	74.39%
run-bart	95.70%	61.05%	74.54%

- Boosting *tfidf* weights of terms occurring in coreference chains **does not** significantly improve text classification performance
- Approach limitations:
  - Limited BART performance -> coreference information is noisy
  - BART biased towards named entities -> coreference chains are incomplete; common nouns could be more important
  - Feature selection -> could discard boosted terms
  - Results are quite high (95% macro averaged precision); perhaps a more challenging classification task would benefit more from coreference information

Textual entailment



Textual entailment experiments

- Classifier is trained on **similarity metrics**
  - Lexical similarity metrics (e.g. Precision, Recall)
  - BLEU (Papineni et al., 2002)
  - METEOR (Denkowski and Lavie, 2011)
  - TINE (Rios et al., 2011)
- Coreference chains processed: each mention in a chain is substituted by the longest (**most informative**) mention (Castillo 2010)
- Train/Test RTE two-way benchmark datasets

Results

Textual entailment experiments

- Accuracy with 10-fold-cross validation
- Comparison: model with coreference information and model without coreference information

Dataset	Model coref	Model no-coref
RTE-1	54.14	56.61
RTE-2	58.50	60
RTE-3	60.25	67.25

Results
Textual entailment experiments (2)

- Accuracy with test datasets
- Comparison: model with coreference information and model without coreference information

Dataset	Model coref	Model no-coref
RTE-1	56.87	56.87
RTE-2	57.12	59.12
RTE-3	60.25	61.75

Final word

- First study of extrinsic evaluation in the context of anaphora resolution and coreference resolution for more than one NLP application
- Impact of anaphora resolution (MARS) on three NLP applications (text summarisation, term extraction and text categorisation) explored
- Deployment of anaphora resolution has positive albeit limited impact
- Alternative models as to how benefit from the anaphora resolution information, appear to be promising
- Higher performance, domain-tuned anaphora resolution should be considered

Final word (2)

- For coreference resolution, impact of BART investigated
- BART has no positive impact
- Alternative models for coreference resolution should be considered as well
- Not-so-high performing anaphora or coreference resolution is not an encouraging option

Anaphora and coreference are really a hard nut to crack…

If Peter Mandelson had been in Tony Blair's shoes **he** would have demanded **his** resignation the day the Prime Minister forced **him** to leave the Cabinet.

Ways forward

- Exploiting semantic knowledge
- Exploiting statistical as well as latest DL and word representation methods
- Exploiting other knowledge sources (gaze data)

My latest research

- Focuses improving anaphora resolution performance through:
- Employment of Linear regression/ Word Representation/Deep Learning / techniques
- Use of gaze data
- Joint research with V Yaneva, R Evans and L A Ha

My latest research:
seeking to improve non-anaphoric recognition

- Annotate instances of anaphoric and pleonastic *it*
- Derive information from gaze data

My latest research: seeking to improve MARS

- Annotate antecedents
- Annotate antecedent indicators
- Optimise the values of antecedent indicators using the annotation and linear regression/word2vec techniques
- Further optimise the values using gaze data

Anaphora and Eye Tracking

- Study 1: Do readers process pleonastic and anaphoric cases of *"it"* in the same way?

- Study 2: Can gaze data be used to enhance anaphora resolution systems? (optimising MARS antecedent indicators)

Eye-tracking corpus

- The GECO eye-tracking corpus (Cop *et al*., 2017)
- Agatha Christie's *"The Mysterious Affair at Styles"*: 54,364 tokens and 5,012 types (English version)
- Read by 14 English monolingual students
- Contains 48 early and late eye-tracking features

Study 1: Pleonastic vs Anaphoric *It*

- All 1,070 cases of the pronoun *it* in GECO were annotated, marking them as either *pleonastic* (340) or *anaphoric* (712).

- We average the values of the 46 gaze features across participants

- The two classes were compared for all gaze features using Wilcoxon Signed Rank test.

Study 1: Preliminary Results

- 25 gaze features resulted in significant differences between the pronoun classes.
- This indicates that pleonastic and anaphoric cases of *it* are processed differently.
- Future work includes training ML models to distinguish between the classes

Feature	P value
WORD FIXATION COUNT	0.000
WORD FIXATION %	0.000
WORD FIRST RUN START TIME	0.001
WORD FIRST RUN END TIME	0.001
WORD FIRST RUN FIXATION %	0.000
WORD GAZE DURATION	0.030
WORD SECOND RUN FIX. COUNT	0.007
WORD SECOND RUN FIX. Perc	0.001
WORD FIRST FIXATION INDEX	0.001
WORD FIRST FIX. RUN INDEX	0.001
WORD FIRST FIXATION TIME	0.001
WORD FIRST FIX. VIS. W. COUNT	0.000
WORD FIRST FIXATION Y	0.005
WORD SECOND FIX. DURATION	0.000
WORD SECOND FIXATION RUN	0.024
WORD SECOND FIXATION X	0.011
WORD THIRD FIX. DURATION	0.033
WORD THIRD FIXATION RUN	0.030
WORD THIRD FIXATION TIME	0.022
WORD LAST FIXATION RUN	0.005
WORD LAST FIXATION TIME	0.002
WORD SELECTIVE GO PAST TIME	0.008
WORD TOTAL READING TIME	0.002
WORD TOTAL READING TIME %	0.000
WORD SKIP	0.014

Study 2: Antecedent Indicators

All NPs are annotated manually. The data is divided in 3 parts and annotated by 3 annotators.

First Noun Phrases (FNP)	1 if the NP is the one closest to the pronoun, 0 otherwise.
Indicating Verbs (IV)	1 if "*analyse, assess, check, consider, cover, define, describe, develop, discuss, examine, explore, highlight, identify, illustrate, investigate, outline, present, report, review, show, study, summarise, survey, synthesise"* are present.
Lexical Reiteration 1 (LR1)	1 if the NP is repeated once, 0 otherwise
Lexical Reiteration 2 (LR2)	1 if the NP is repeated more than once, 0 otherwise
Collocation Match (CM)	1 if the NP precedes or follows the same word that the pronoun precedes or follows. 0 otherwise.
Immediate Reference (IR)	1 if constructions of the form '...(You) V1 NP ... con (you) V2 it (con (you) V3 it)', where con ϵ {and/or/before/after/until/so that...}. 0 otherwise.
Term Preference (TP)	1 if the NP is an important concept/term in the text, 0 otherwise.
Indefiniteness (IND)	1 if the NP is indefinite, 0 otherwise.
Prepositional NPs (PNP)	1 if the NP is prepositional, 0 otherwise.
Referential Distance 0 (RD0)	1 if the NP occurs in the same sentence to the pronoun. 0 otherwise.
Referential Distance 1 (RD1)	1 if the NP occurs in the previous sentence to the pronoun. 0 otherwise.
Referential Distance 2 (RD2)	1 if the NP occurs two sentences prior to the pronoun. 0 otherwise
Syntactic Parallelism (SP)	1 if the NP has the same syntactic role as the pronoun. 0 otherwise.

Study 2: Preliminary results

Weight optimisation for each indicator using linear regression

Inidcators	Original weight	Weight Optimized on Annotator 1 data portion	Weight Optimized on Annotator 2 data portion	Weight Optimized on Annotator 3 data portion	Weight Optimized using all data
FNP	1	0.14	0.15	0.14	0.12
IV	1	0.5	0.69	0.5	0.25
L1	1	0.2	0.12	0.2	0.17
L2	1	0	0.2	0	0
CM	2	0.3	0.56	0.3	0.18
IR	2	0.58	0.53	0.58	0.6
TP	1	0.23	0.46	0.23	0.36
IND	-1	0	0	0	0
PNP	-1	-0.2	-0.13	-0.2	-0.16
RD0	2	0	0	0	0
RD1	1	0	-0.13	0	0
RD2	-1	-0.1	-0.2	-0.1	-0.1

Study 2: Preliminary results

System accuracy for the annotated data using optimised weights

		Original weight	Weight Optimised on Annotator 1 data	Weight Optimised on Annotator 2 data	Weight Optimised on Annotator 3 data	Weight Optimised using all data
Annot. 1	Correct	124	139	136	130	135
	Accuracy	0.61	0.69	0.67	0.64	0.67
Annot. 2	Correct	84	92	92	90	90
	Accuracy	0.64	0.70	0.70	0.68	0.68
Annot. 3	Correct	62	70	73	87	80
	Accuracy	0.496	0.56	0.584	0.696	0.64
All data	Correct	270	301	301	307	305
	Accuracy	0.59	0.66	0.66	0.67	0.66

Conclusion

- The preliminary results are very encouraging

Contact details

- My email: R.Mitkov@wlv.ac.uk
- My webpage: www.wlv.ac.uk/~le1825
- My research group web page: rgcl.wlv.ac.uk

The mission of the Research Group (RGCL) and the Research Institute (RIILP) is to:

- produce world-leading research
- offer first-class research supervision and postgraduate teaching in the interdisciplinary areas of information and language processing
- to deliver cutting-edge practical applications with far-reaching societal impact

RGCL
Research Group in Computational Linguistics

Anaphora resolution and coreference resolution: still a hard nut to crack? How far has it gone, what is its impact on NLP and what are the ways forward?

0011 0010 1010 1101 0001 0100 1011

Ruslan Mitkov

Research Group in Computational Linguistics

University of Wolverhampton