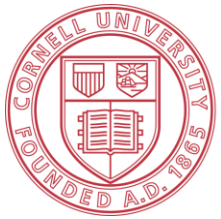
Domain-Independent Abstract Generation for Focused Meeting Summarization

Lu Wang and Claire Cardie
Department of Computer Science
Cornell University



Cornell University

Introduction

- Meetings are a common way to collaborate, share information and exchange opinions.

skype™

GoToMeeting®

Introduction

- Automatic meeting summarization can facilitate:
  - Quick access to the essential output
  - Digest large amount of information easily
  - Project management

Focused Meeting Summarization

- Generating summaries of the important outputs of a meeting

- Decisions
- Action Items
- Problems
- Progress

Focused Meeting Summarization

Dialogue Acts:

C: Looking at what we've got, we we want an LCD display with a spinning wheel.

B: You have to have some push-buttons, don't you?

C: Just spinning and not scrolling , I would say .

B: I think the spinning wheel is definitely very now.

A: but since LCDs seems to be uh a definite yes,

C: We're having push-buttons on the outside

C: and then on the inside an LCD with spinning wheel,

Decision Summary:

The remote will have push buttons outside, and an LCD and spinning wheel inside.

Related Work

- Existing meeting summarization systems are mostly extractive
  - Maximum entropy, conditional random fields (CRFs), and support vector machines (SVMs) are investigated for extracting utterances to make up a generic meeting summary (Buist et al., 2004; Galley, 2006; Xie et al., 2008).
  - Supervised methods are utilized to identify key phrases for inclusion in the decision summary (Fernandez et al., 2008; Bui et al., 2009)
- The need for abstractive summarization
  - Murray et al. (2010) show that users demonstrate a strong preference for abstractive summaries

Goal

Input

Dialogue Acts:

C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
C: We're having push-buttons on the outside
C: and then on the inside an LCD with spinning wheel,

Output

Final Summary:

The remote will have push buttons outside, and an LCD and spinning wheel inside.

Contribution

- We propose a fully automatic domain-independent abstract generation framework for focused meeting summarization.
- We rely on task-specific templates to guide abstract generation and present a novel template extraction algorithm based on Multiple Sequence Alignment.
- We use an Overgenerate-and-Rank strategy for abstract surface realization.

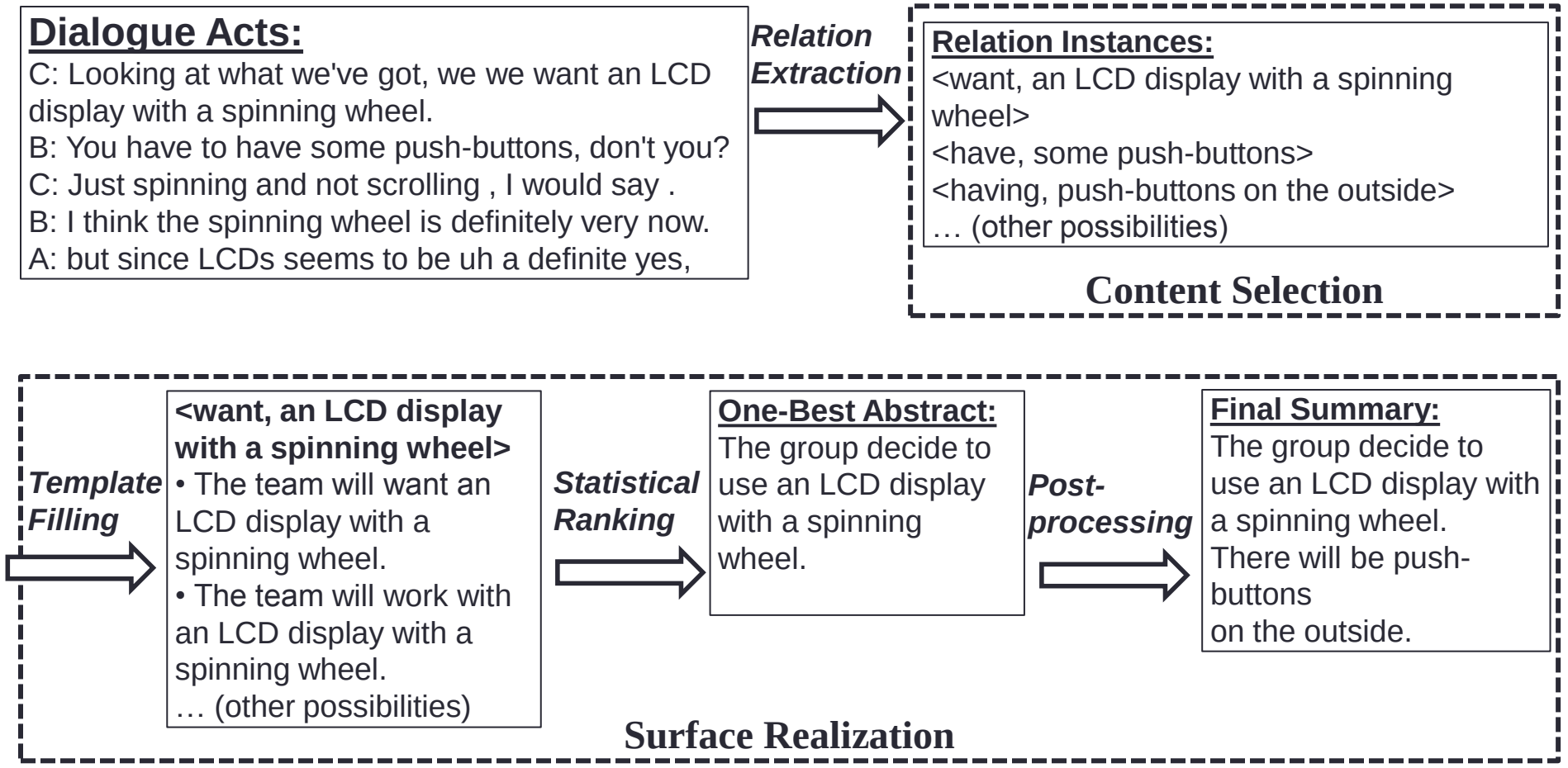
Related Work

- Beyond extractive meeting summarization
  - Murray et al. (2010a) present an abstraction system which first maps utterances to the summary types (e.g. decision), and then selects the ones covering most entities.
  - Sentence compression is studied to drop redundant words in Liu and Liu (2009).

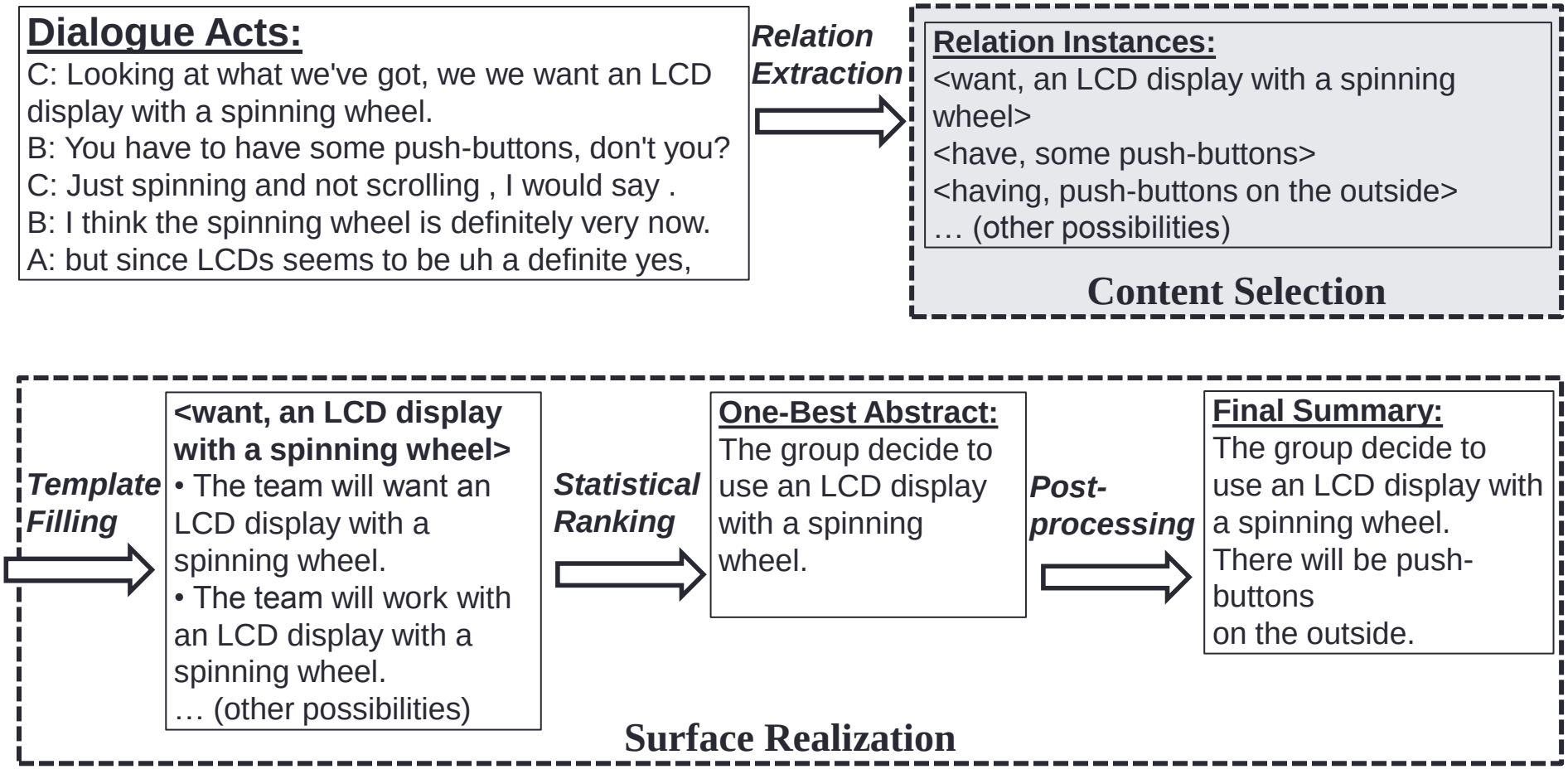
Related Work

- Concept-to-text generation
  - The generation process is usually decomposed into content selection (or text planning) and surface realization (Angeli et al., 2010; Konstas and Lapata, 2012).
  - Angeli et al. (2010) learn from structured database records and parallel textual descriptions, and then generate texts based on a series of decisions made to select the records, fields, and proper templates for rendering.

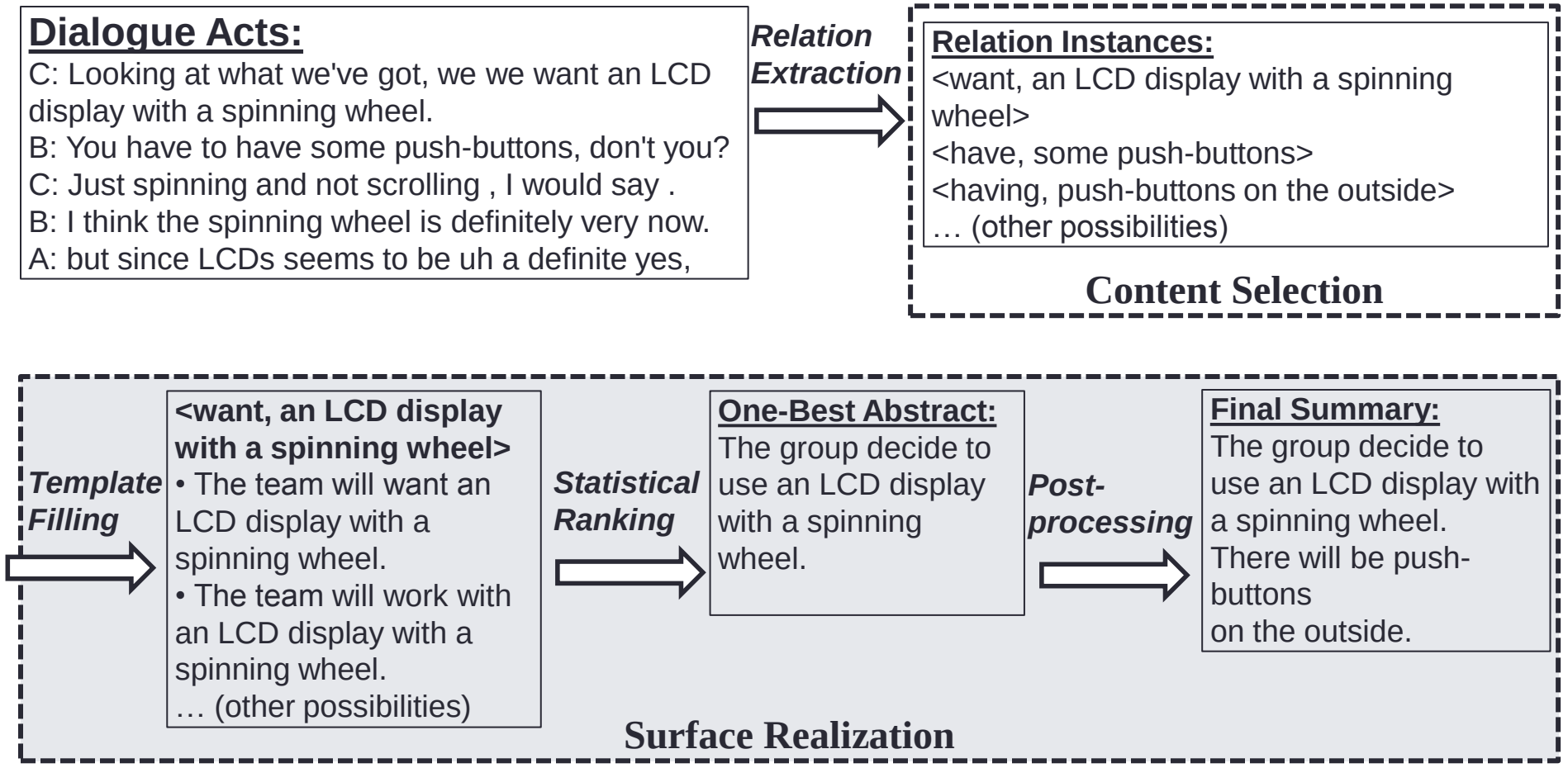
Framework


Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Relation Extraction
Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
… (other possibilities)
Content Selection
Template Filling
<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
… (other possibilities)
Statistical Ranking
One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.
Post-processing
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Surface Realization

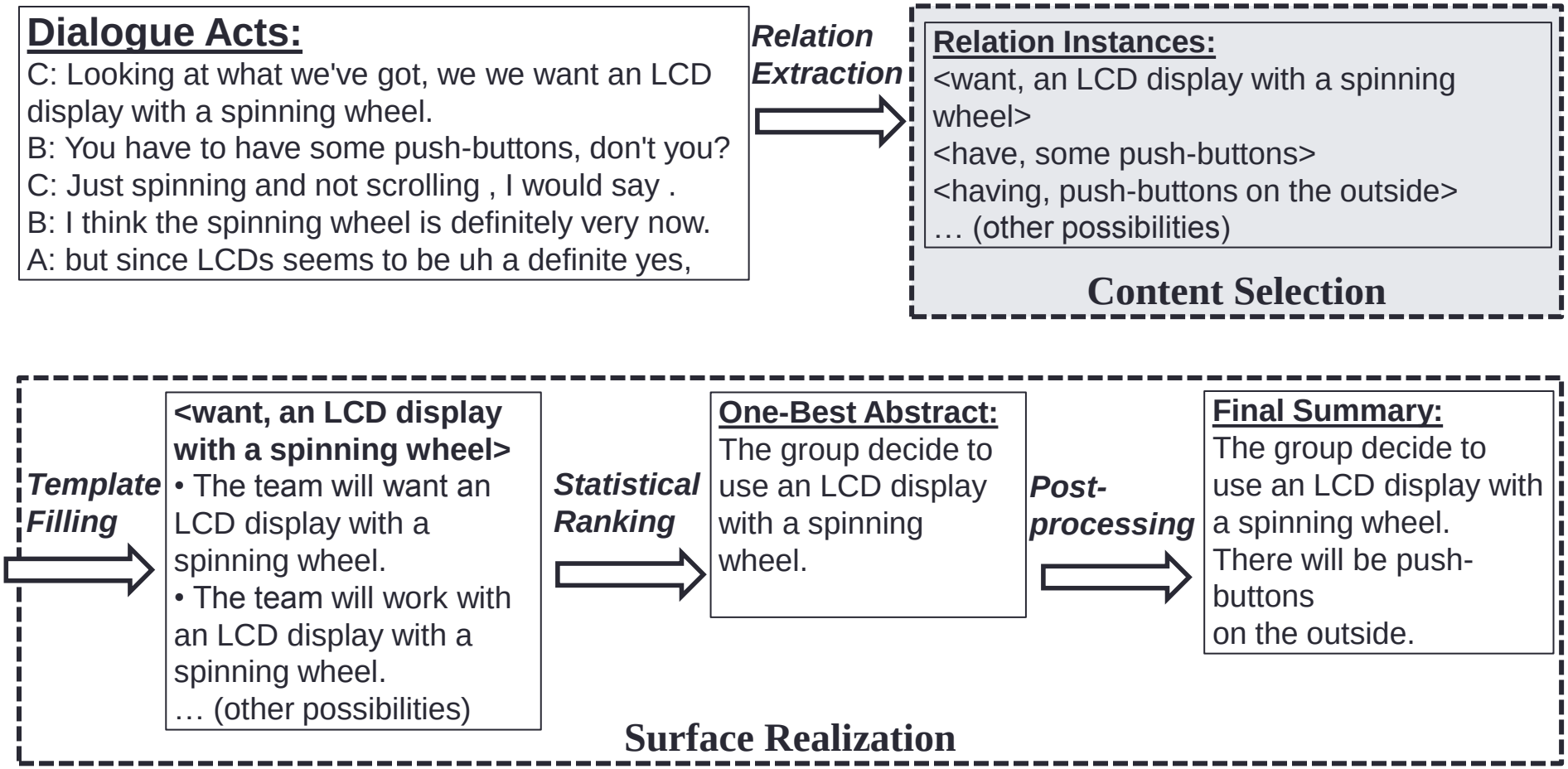
Framework


Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Relation Extraction
Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
… (other possibilities)
Content Selection
Template Filling
<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
… (other possibilities)
Statistical Ranking
One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.
Post-processing
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Surface Realization

Framework


Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Relation Extraction
Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
… (other possibilities)
Content Selection
Template Filling
<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
… (other possibilities)
Statistical Ranking
One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.
Post-processing
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Surface Realization

Framework


Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Relation Extraction
Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
… (other possibilities)
Content Selection
Template Filling
<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
… (other possibilities)
Statistical Ranking
One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.
Post-processing
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Surface Realization

Content Selection

Dialogue Acts:

C: Looking at what we've got, we we ***want*** **[an LCD display with a spinning wheel]**.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
C: We're having ***push-buttons*** **[on the outside]**
C: and then on the inside an LCD with spinning wheel,

Relation Instance:

<*indicator*, [argument]> pairs, where the ***indicator*** evokes a relation of interest and the **[argument]** is the target phrase containing the object.

Content Selection

Dialogue Acts:

C: Looking at what we've got, we we ***want*** **[an LCD display with a spinning wheel]**.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
C: We're having ***push-buttons*** **[on the outside]**
C: and then on the inside an LCD with spinning wheel,

Relation Instance:

<*indicator*, [argument]> pairs, where the ***indicator*** evokes a relation of interest and the **[argument]** is the target phrase containing the object.

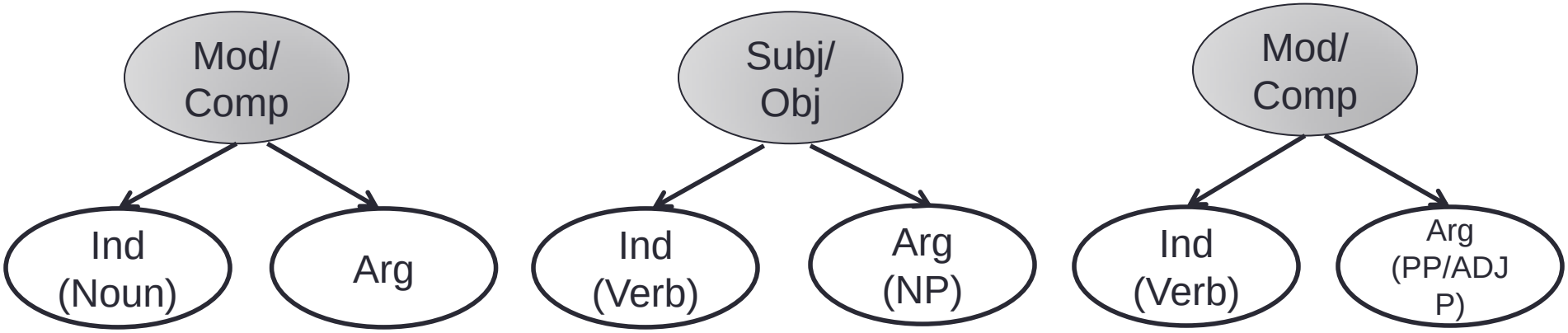
Sample Summary:

The group decided to use **an LCD display with a spinning wheel**. There will be **push-buttons on the outside**.

Content Selection via Relation Extraction

- Relation Instance Candidates:
  - **Indicator**: noun or verb
  - **Argument**: noun phrase (NP), prepositional phrase (PP) or adjectival phrase (ADJP).

- Contraints:



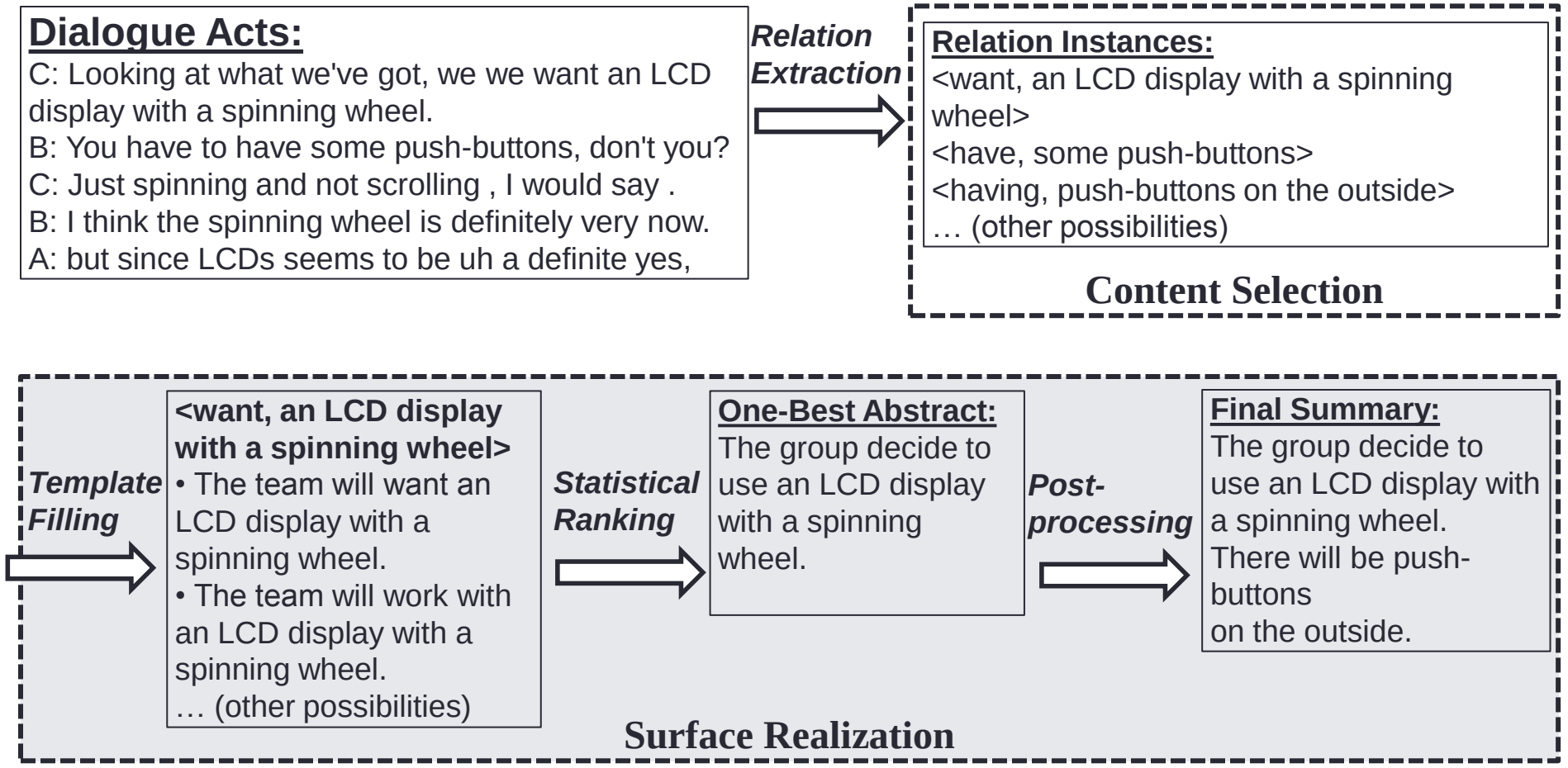
Content Selection via Relation Extraction

- Binary classifier based on Support Vector Machines
  - Summary-worthy vs. not summary-worthy

- Sample Features:

Basic Features	Discourse Features
TF/IDF/TFIDF	Main speaker or not?
Constituent tag	Shared by adjacency pair?
Dependency relation	Target DA is a positive response?

Framework


Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Relation Extraction
Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
… (other possibilities)
Content Selection
Template Filling
<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
… (other possibilities)
Statistical Ranking
One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.
Post-processing
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Surface Realization

Extracting Summary Templates

- Multiple Sequence Alignment (MSA)
  - MSA is commonly used in bioinformatics to identify equivalent fragments of DNAs (Durbin et al., 1998) and has also been employed for learning paraphrases (Barzilay and Lee, 2003)

Multiple Sequence Alignment (MSA)

S1 = T A T A G C T
S2 = T T A G C T
S3 = T A T C G C
S4 = A C A C C T

S1’	=	T	A	T	A	G	C	T
S2’	=	T	-	T	A	G	C	T
S3’	=	T	-	-	C	G	C	-
S4’	=	-	A	C	A	C	C	T

Multiple Sequence Alignment (MSA)

S1 = T A T A G C T
S2 = T T A G C T
S3 = T A T C G C
S4 = A C A C C T

Scorer(x1, x2)= { 1, if x1==x2
0, if x1=="-" or x2== "-"
-1, if x1!=x2

S1'	=	T	A	T	A	G	C	T
S2'	=	T	-	T	A	G	C	T
S3'	=	T	-	-	C	G	C	-
S4'	=	-	A	C	A	C	C	T

Extracting Summary Templates

- We first cluster the abstracts in the training data according to their lexical and structural similarity.
  - We first replace all appearances of dates, numbers, and proper names with generic semantic labels.
  - We further replace sequences that appear in both the abstract and supporting dialogue acts by a label indicating its phrase type.
  - Following Barzilay and Lee (2003), we approach the sentence clustering task by hierarchical clustering with a similarity metric based on word n-gram overlap.

Extracting Summary Templates

1) The group were not sure whether to include a recharger for the remote .
2) The group were not sure whether to use plastic and rubber or titanium for the case .
3) The group were not sure whether the remote control should include functions for controlling video .
4) They were not sure how much a recharger would cost to make .

1) The group were not sure whether to [include]$_{VP}$ [a recharger for the remote]$_{NP}$.
2) The group were not sure whether to use [plastic and rubber or titanium for the case]$_{NP}$.
3) The group were not sure whether [the remote control]$_{NP}$ should include [functions for controlling video]$_{NP}$.
4) They were not sure how much [a recharger]$_{NP}$ would cost to make .

1) The group were not sure whether to VP NP .
2) The group were not sure whether to use NP .
3) The group were not sure whether NP should include NP .
4) They were not sure how much NP would cost to make .

Extracting Summary Templates

1) The group were not sure whether to include a recharger for the remote .
2) The group were not sure whether to use plastic and rubber or titanium for the case .
3) The group were not sure whether the remote control should include functions for controlling video .
4) They were not sure how much a recharger would cost to make .

1) The group were not sure whether to [include]$_{VP}$ [a recharger for the remote]$_{NP}$.
2) The group were not sure whether to use [plastic and rubber or titanium for the case]$_{NP}$.
3) The group were not sure whether [the remote control]$_{NP}$ should include [functions for controlling video]$_{NP}$.
4) They were not sure how much [a recharger]$_{NP}$ would cost to make .

1) The group were not sure whether to VP NP .
2) The group were not sure whether to use NP .
3) The group were not sure whether NP should include NP .
4) They were not sure how much NP would cost to make .

Extracting Summary Templates

1) The group were not sure whether to include a recharger for the remote .
2) The group were not sure whether to use plastic and rubber or titanium for the case .
3) The group were not sure whether the remote control should include functions for controlling video .
4) They were not sure how much a recharger would cost to make .

1) The group were not sure whether to [include]$_{VP}$ [a recharger for the remote]$_{NP}$.
2) The group were not sure whether to use [plastic and rubber or titanium for the case]$_{NP}$.
3) The group were not sure whether [the remote control]$_{NP}$ should include [functions for controlling video]$_{NP}$.
4) They were not sure how much [a recharger]$_{NP}$ would cost to make .

1) The group were not sure whether to VP NP .
2) The group were not sure whether to use NP .
3) The group were not sure whether NP should include NP .
4) They were not sure how much NP would cost to make .

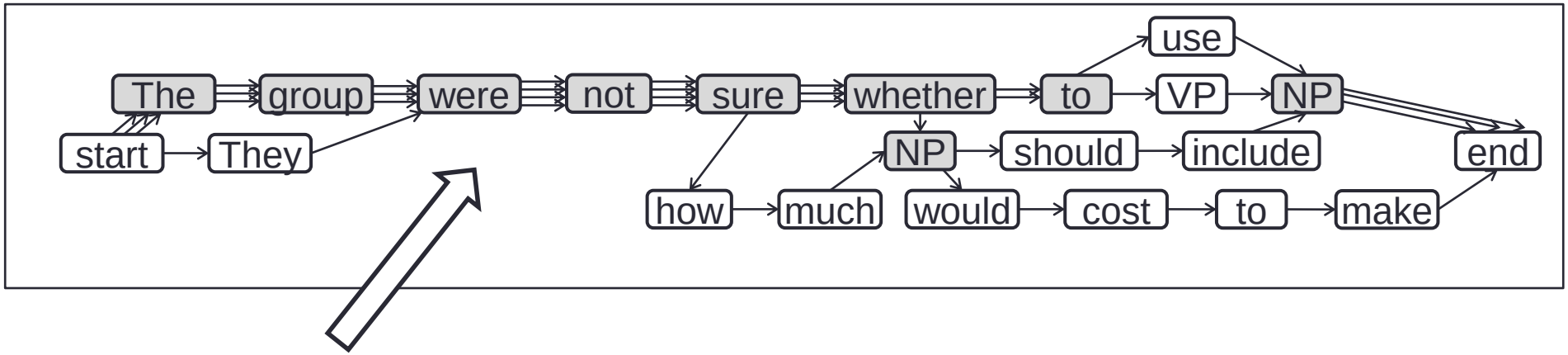
Extracting Summary Templates

- Multiple Sequence Alignment
  - Computing an optimal MSA is NP-complete (Wang and Jiang, 1994).
  - We implement an approximate algorithm (Needleman and Wunsch, 1970) that iteratively aligns two sequences each time and treats the resulting alignment as a new sequence

Extracting Summary Templates

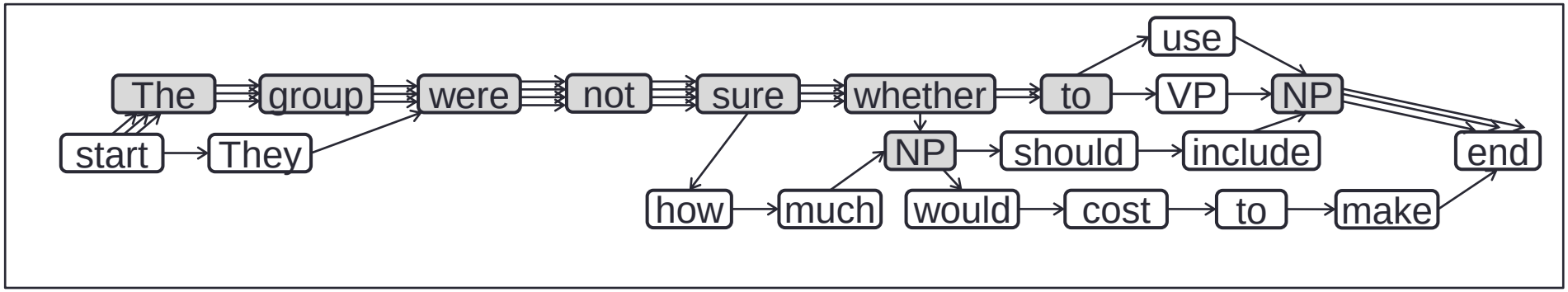
1) The group were not sure whether to VP NP .
2) The group were not sure whether to use NP .
3) The group were not sure whether NP should include NP .
4) They were not sure how much NP would cost to make .

MSA



The backbone nodes shared by at least 50% sentences are shaded.

Extracting Summary Templates



Template Induction

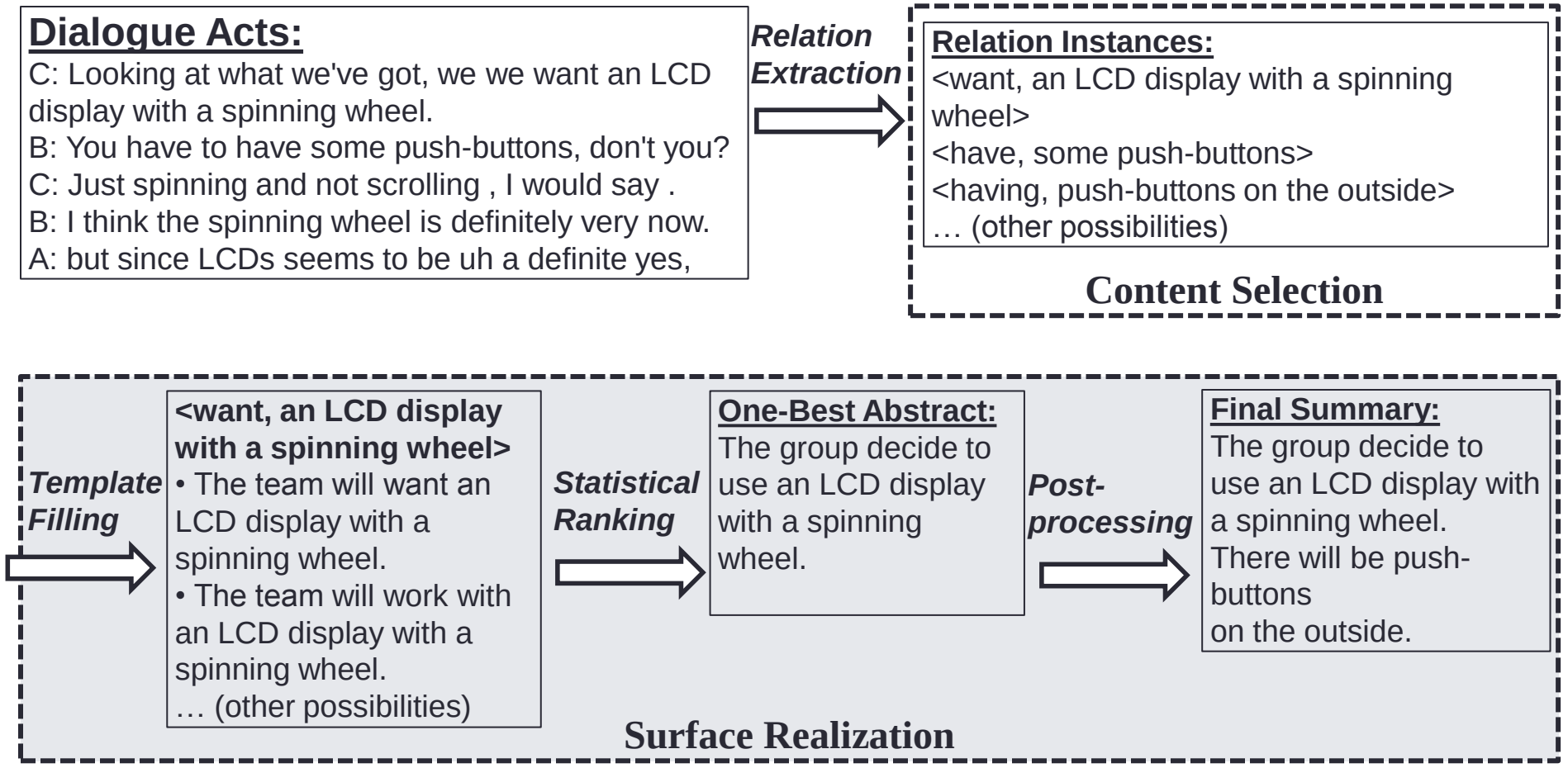
Template Examples:

T1: *The group were not sure whether to* VP NP .

T2: *The group were not sure whether NP* VP VP *NP* .

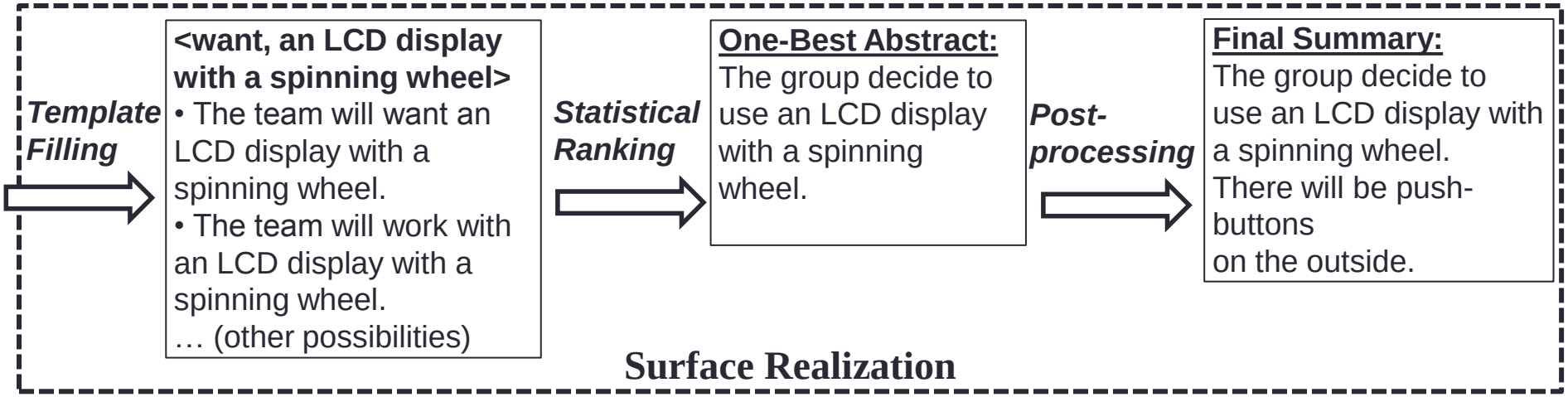
T3: NP *were not sure* WHADJP WHADJP NP VP VP VP VP VP .

Framework


Dialogue Acts:
C: Looking at what we've got, we we want an LCD display with a spinning wheel.
B: You have to have some push-buttons, don't you?
C: Just spinning and not scrolling , I would say .
B: I think the spinning wheel is definitely very now.
A: but since LCDs seems to be uh a definite yes,
Relation Extraction
Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
… (other possibilities)
Content Selection
Template Filling
<want, an LCD display with a spinning wheel>
• The team will want an LCD display with a spinning wheel.
• The team will work with an LCD display with a spinning wheel.
… (other possibilities)
Statistical Ranking
One-Best Abstract:
The group decide to use an LCD display with a spinning wheel.
Post-processing
Final Summary:
The group decide to use an LCD display with a spinning wheel. There will be push-buttons on the outside.
Surface Realization

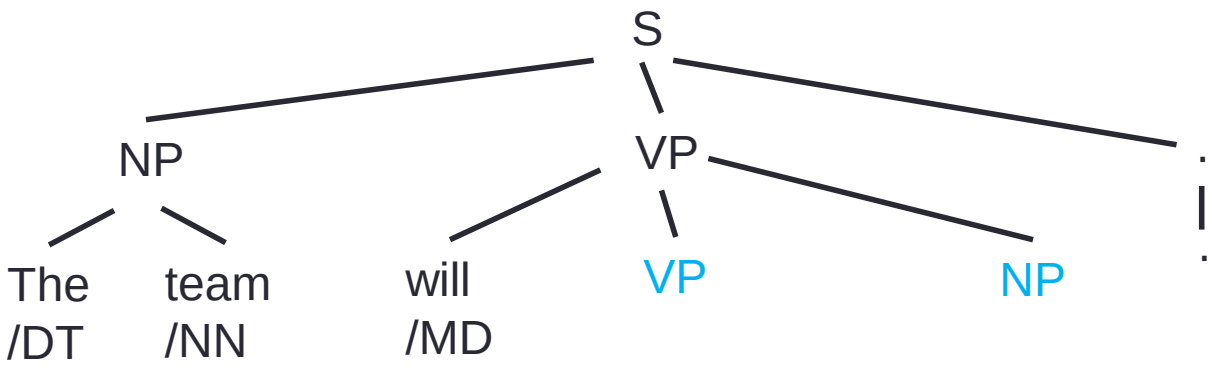
Surface Realization

- An Overgenerate-and-Rank Approach
  - Overgenerate-and-Rank strategy has been used for sentence planning (Walker et al., 2001) and questions generation (Heilman and Smith, 2010).



Template Filling

- The templates are represented by their parse tree.



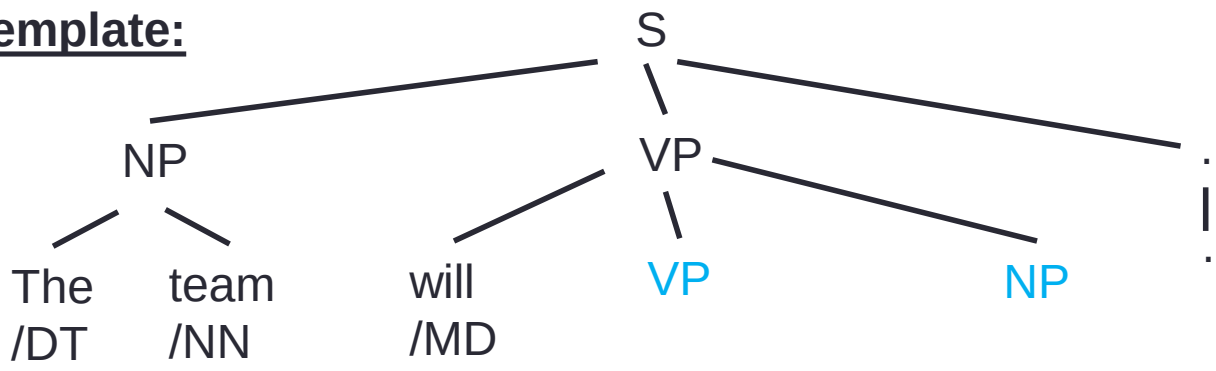
- We then fill the slots with relation instances.

Template Filling

- Constituent-level transformation of the relation instances

Template Filling

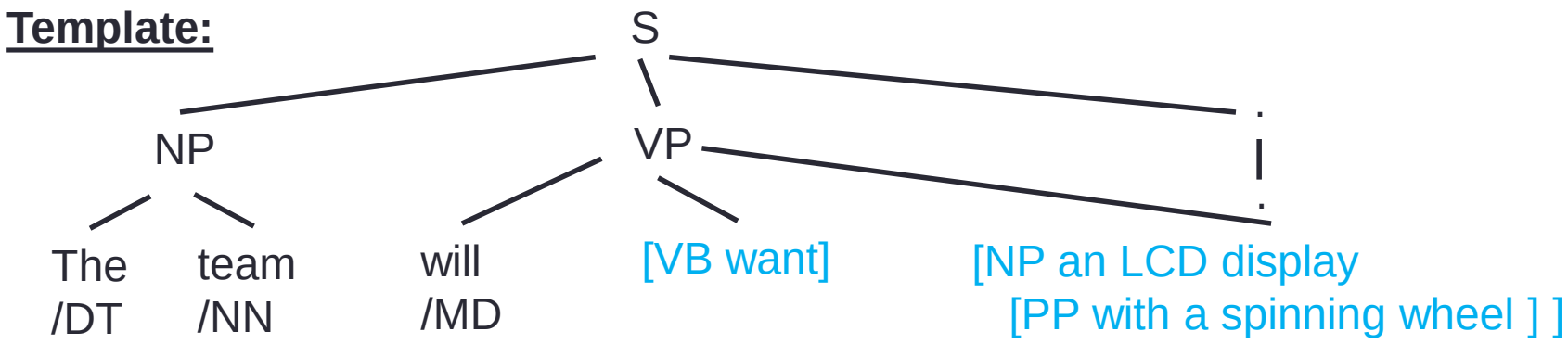
Template:



Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel]]

Full-Constituent Mapping

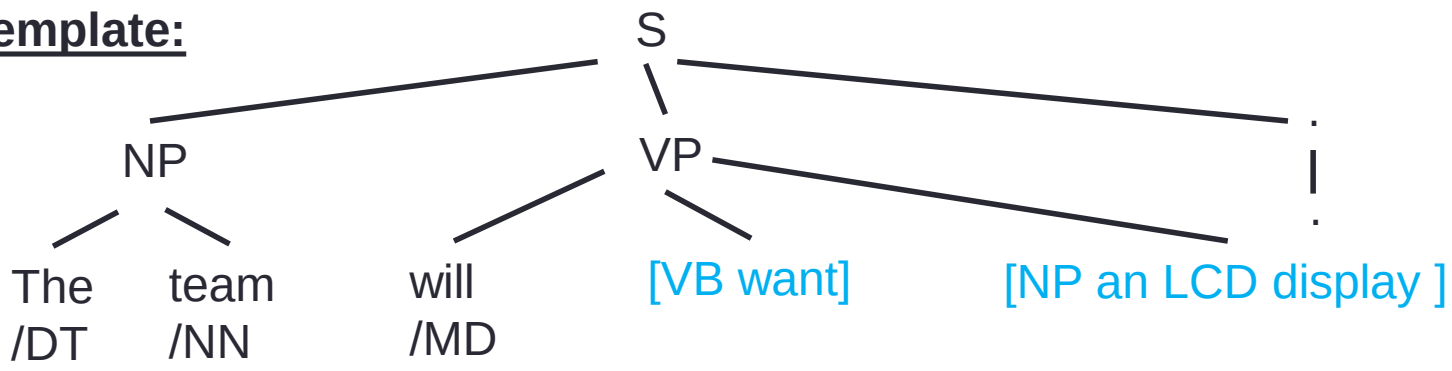


Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel]]

Sub-Constituent Mapping

Template:

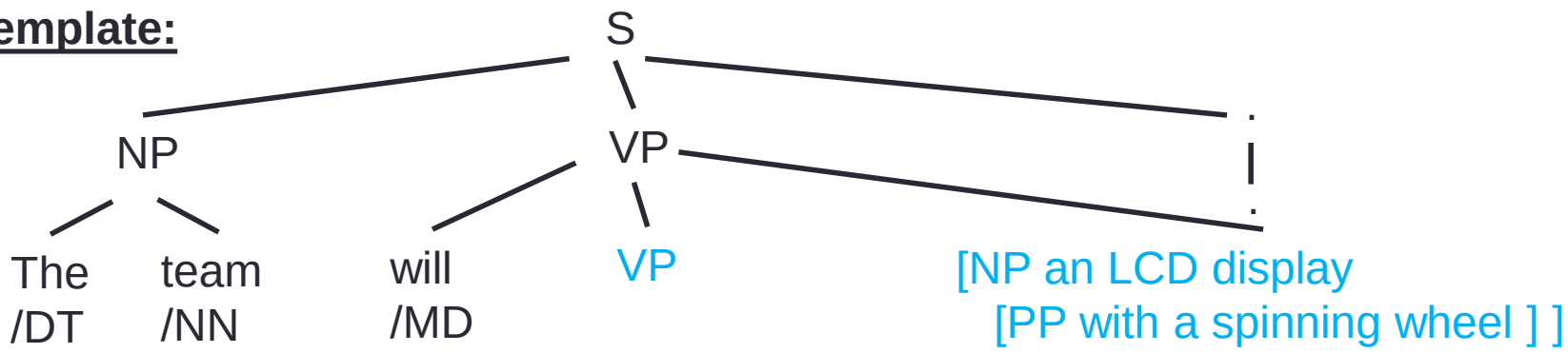


Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel]]

Removal

Template:



Relation instance:

Indicator: [VB want]
Argument: [NP an LCD display [PP with a spinning wheel]]

Template Filling

Templates:
T1: *The team will* VP NP .
T2: *The team will work with* NP .
…

Relation Instances:
<want, an LCD display with a spinning wheel>
<have, some push-buttons>
<having, push-buttons on the outside>
…

<want, an LCD display with a spinning wheel>
• *The team will* want an LCD display with a spinning wheel.
• *The team will work with* an LCD display with a spinning wheel.
… (other possibilities)

Statistical Ranking

- Support Vector Regression (SVR) (Smola and Scholkopf, 2004) is utilized to rank the generated abstracts.

- Sample Features

Basic Features	**Structure Features**
Number of words	Constituent tag for ind/arg
Number of new nouns	Dependency relation for ind/arg
Ind/Arg has content words?	The relation matches the template?
Realization Features	**Language Model Features**
Abstract has verb?	log p_{LM}(abstract)
Abstract starts with verb?	Bigram probability of ind/arg
Abstract has adjacent verb/adj?	Trigram probability of ind/arg

Redundancy Handling

- Post-selection aims to maximize the information coverage and minimize the redundancy of the summary.

- We use a greedy algorithm (Lin and Bilmes, 2010) to select a subset of the generated abstracts.

Experimental Setup

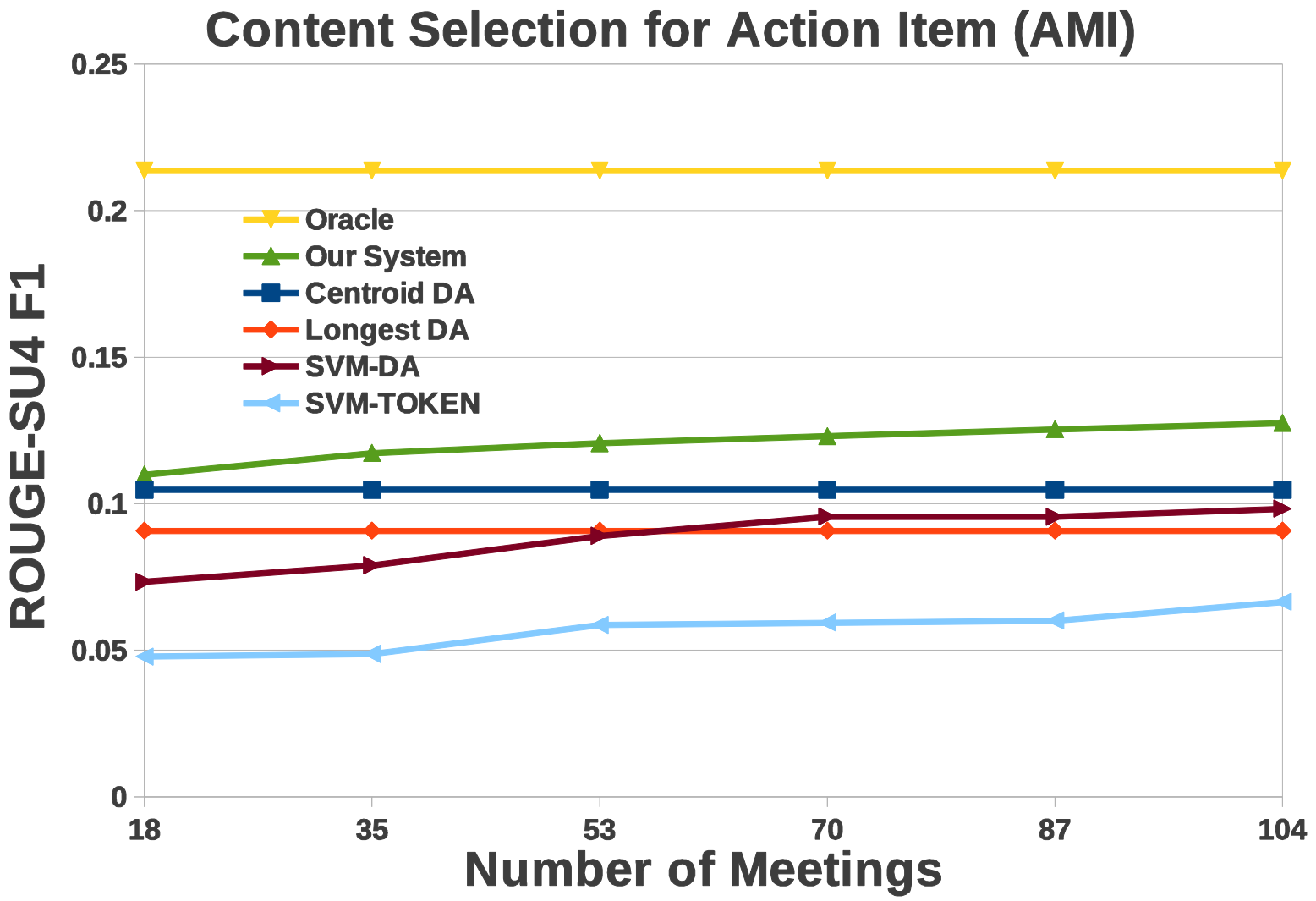
- Two disparate corpora
  - The AMI meeting corpus (McCowan et al., 2005)
    - 139 scenario-driven meetings
    - Contains abstracts for ***decisions, action items, and problems***
  - The ICSI meeting corpus (Janin et al., 2003)
    - 75 naturally occurring meetings
    - Contains abstracts for ***decisions, progress, and problems***

- System Input
  - True/System clusterings of summary related DAs

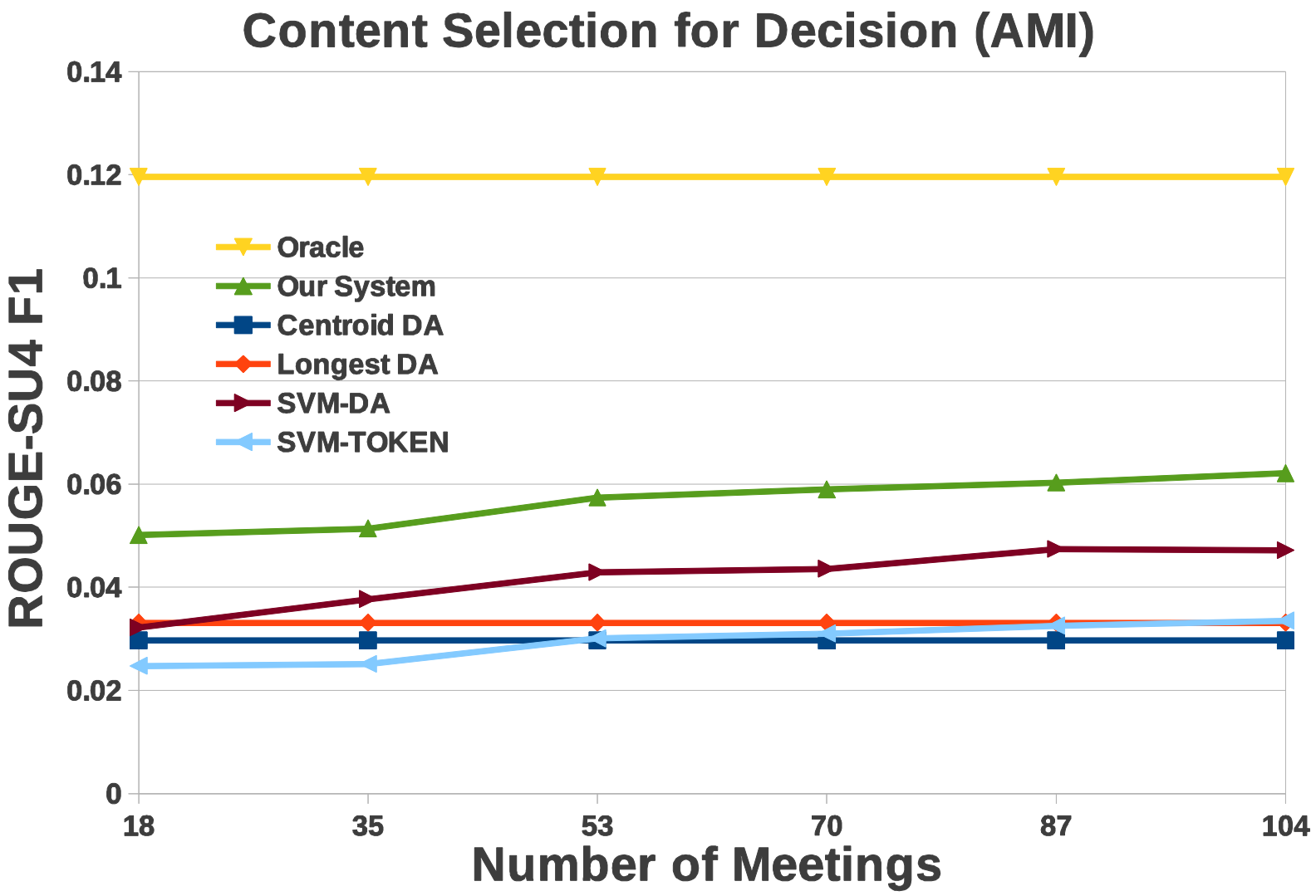
Content Selection Results

- Evaluation metric
  - ROUGE (Lin and Hovy, 2003) -- computes the ngram overlapping between the system summaries and the reference summaries.
  - It has been widely used for text and speech summarization (Dang, 2005; Xie et al., 2008).
- Comparison:
  - Baselines: Longest DA, Centroid DA
  - Supervised Learning: Utterance-level and token-level extractive summarization with SVMs (Xie et al., 2008; Sandu et al., 2010; Fernandez et al.,2008)

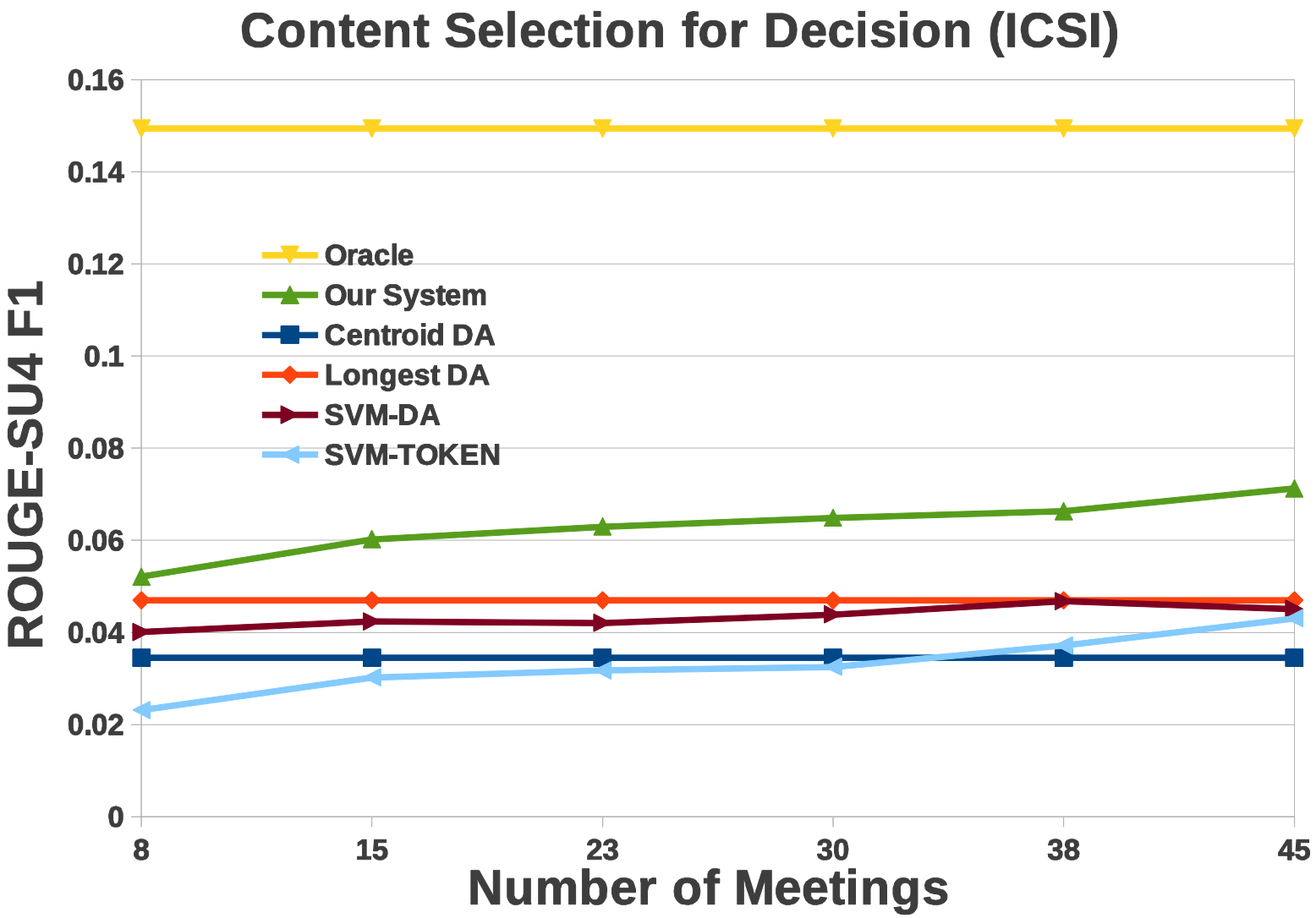
Content Selection Results



Content Selection Results



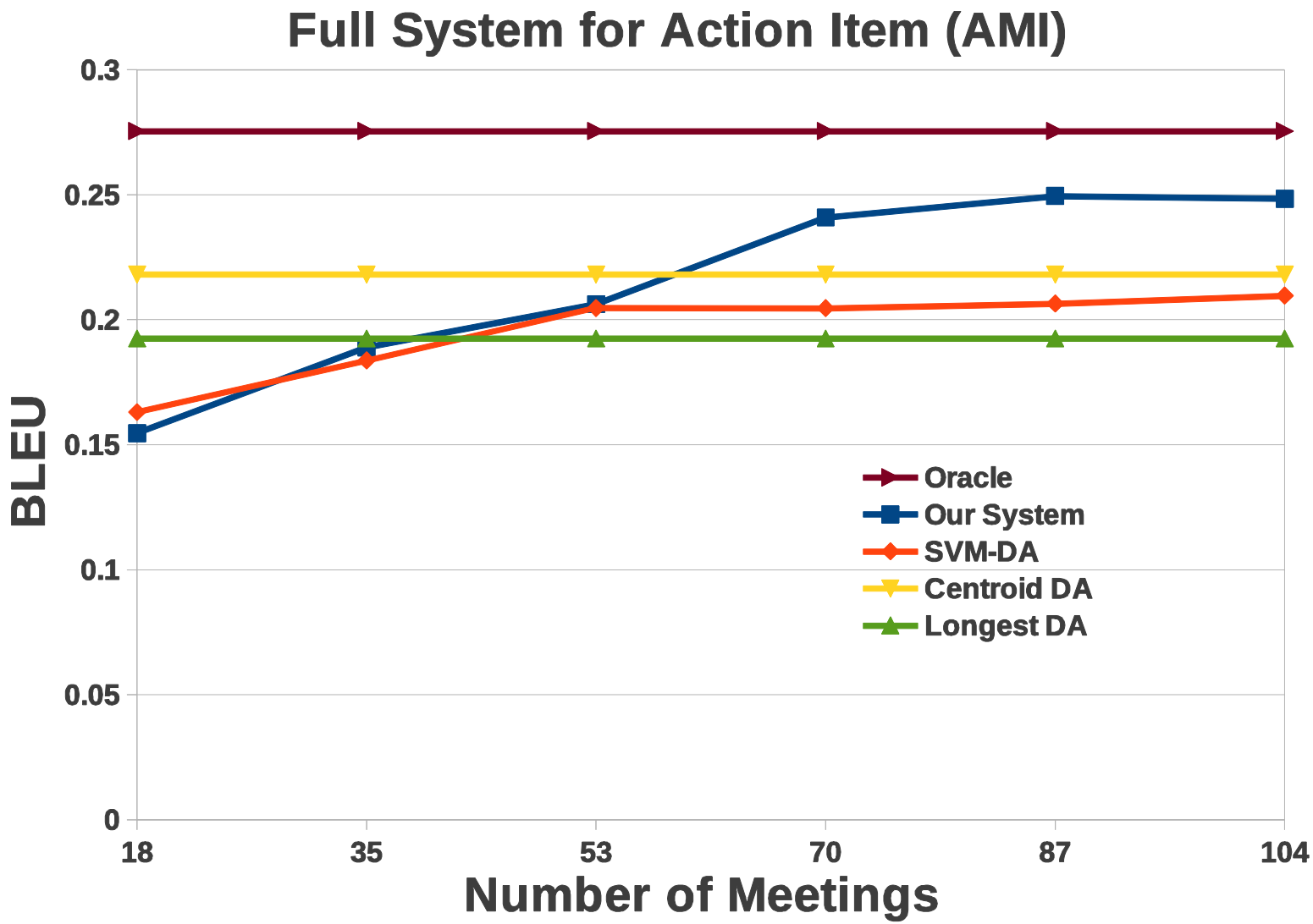
Content Selection Results



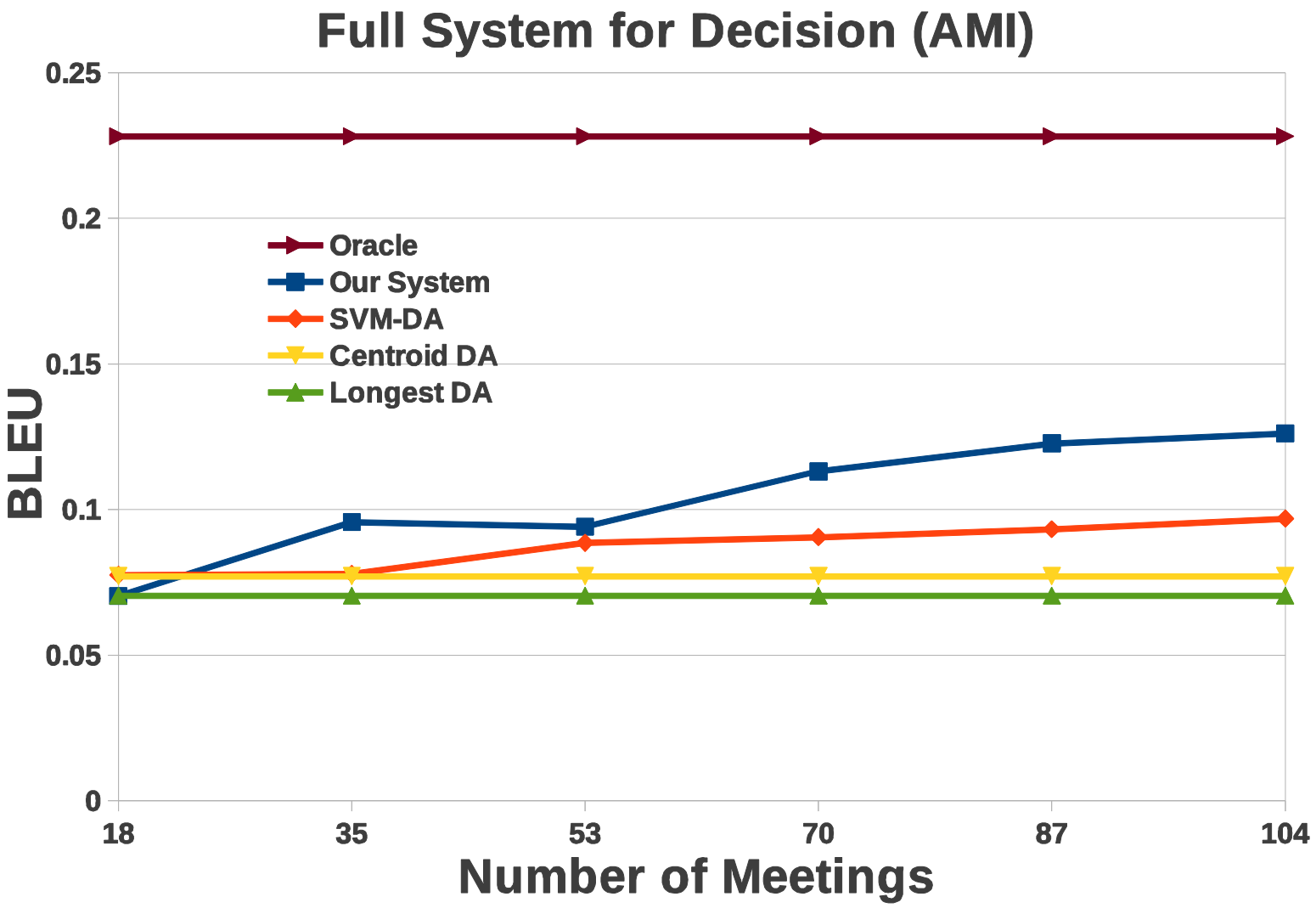
Abstract Generation Results

- Evaluation metrics
  - BLEU (Papineni et al., 2002) – computes the precision of unigrams and bigrams with a brevity penalty
    - has been used to evaluate a variety of language generation systems (Angeli et al., 2010; Konstas and Lapata, 2012).
  - Human evaluation
    - Fluency -- is the text grammatical?
    - Semantic correctness -- does the summary convey the gist of the DAs in the cluster?
    - Overall quality in content, conciseness and grammaticality.

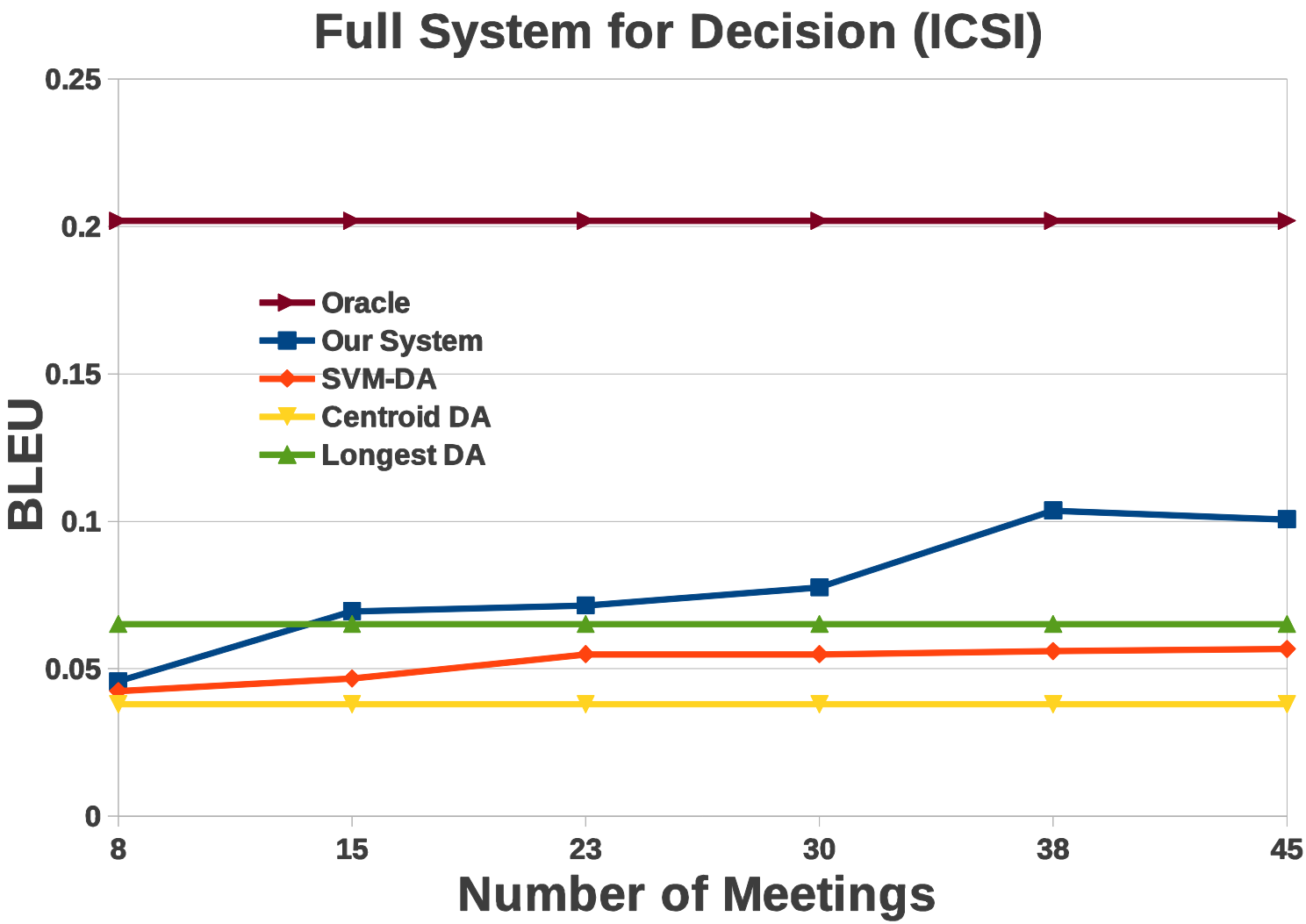
Abstract Generation Results



Abstract Generation Results



Abstract Generation Results



System	AMI Decision		ICSI Decision		AMI Problem		ICSI Problem	
	R-SU4	BLEU	R-SU4	BLEU	R-SU4	BLEU	R-SU4	BLEU
SVM-DA (In-Domain)	4.7	9.7	4.5	5.7	2.4	5.0	3.4	3.4
Our system (In-Domain)	**6.2**	**11.6**	**7.1**	**10.0**	**4.8**	**7.2**	**5.9**	**6.0**
Our system (Out-of-Domain)	**6.1**	**10.3**	**6.4**	7.8	**4.7**	6.2	**5.5**	5.3
Oracle	12.0	22.8	14.9	20.2	11.3	18.9	12.6	13.0

System	AMI Decision		ICSI Decision		AMI Problem		ICSI Problem	
	R-SU4	BLEU	R-SU4	BLEU	R-SU4	BLEU	R-SU4	BLEU
SVM-DA (In-Domain)	4.7	9.7	4.5	5.7	2.4	5.0	3.4	3.4
Our system (In-Domain)	**6.2**	**11.6**	**7.1**	**10.0**	**4.8**	**7.2**	**5.9**	**6.0**
Our system (Out-of-Domain)	**6.1**	**10.3**	**6.4**	7.8	**4.7**	6.2	**5.5**	5.3
Oracle	12.0	22.8	14.9	20.2	11.3	18.9	12.6	13.0

Human Evaluation

System	Fluency		Semantic Correctness		Avg. Length
	Mean	S.D.	Mean	S.D.	
Our system (In-Domain)	**3.67**	0.85	3.27	1.03	**23.65**
Our system (Out-of-Domain)	3.58	0.90	3.25	1.16	**24.17**
SVM-DA (In-Domain)	3.36	0.84	3.44	1.26	38.83

Human Evaluation

System	Fluency		Semantic Correctness		Avg. Length
	Mean	S.D.	Mean	S.D.	
Our system (In-Domain)	**3.67**	0.85	3.27	1.03	**23.65**
Our system (Out-of-Domain)	3.58	0.90	3.25	1.16	**24.17**
SVM-DA (In-Domain)	3.36	0.84	3.44	1.26	38.83

Judges selected our system as the best system in 62.3% scenarios (In-Domain: 35.6%, Out-of-Domain: 26.7%)

Sample Summary

Dialogue Acts:

C: Looking at what we've got, we we want an LCD display with a spinning wheel.

B: You have to have some push-buttons, don't you?

C: Just spinning and not scrolling , I would say .

B: I think the spinning wheel is definitely very now.

A: but since LCDs seems to be uh a definite yes,

C: We're having push-buttons on the outside

C: and then on the inside an LCD with spinning wheel,

Decision Summaries:

Human: The remote will have push buttons outside, and an LCD and spinning wheel inside.

Our System (In-Domain): The group decided to use an LCD display with a spinning wheel. There will be push-buttons on the outside.

Our System (Out-of-Domain): LCD display is going to be with a spinning wheel. It is necessary having push-buttons on the outside.

Conclusion

- We presented a domain-independent abstract generation framework for focused meeting summarization.
- Experimental results on two disparate meeting corpora show that our system can uniformly outperform the state-of-the-art supervised extraction-based systems in both automatic and manual evaluation.
- Our system also exhibits an ability to train on out-of-domain data to generate abstracts for a new target domain.

Thank you!