Constructing the CODA Corpus: a Parallel Corpus of Monologues and Expository Dialogues

Svetlana Stoyanchev and Paul Piwek

The Open University

Motivation

- CODA project (2009-2011): COherent Dialogue Automatically generated from text

Why dialogue?

- Education: observing dialogue motivates students to ask deeper questions (Craig et al. 2000)
- Accessible and entertaining information presentation for children, museum visitors, patients, etc.

Why generate dialogue automatically?

- Most information currently is available in text (books, articles, leaflets, description of art at museums)
- Most people are skilled monologue writers
- Fewer have experience writing dialogues

Uses of Generated Dialogue

- Multimodal presentation of automatically generated dialogues in Second Life

Monologue-to-Dialogue Conversion Approaches

- Manual: construct transformation rules (Piwek et al., 2008) based on Rhetorical relations
  - ATTRIBUTION ("John said", "that he likes apples") →
  - A: What did John say?
  - B: that he likes apples
- Empirical: extract transformation rules from a parallel monologue/dialogue corpus
  - Analyse how rhetorical relations are realised in a dialogue
  - Learn dialogue generation rules from parallel dialogue-monologue

Parallel Corpus of Dialogues and Monologues

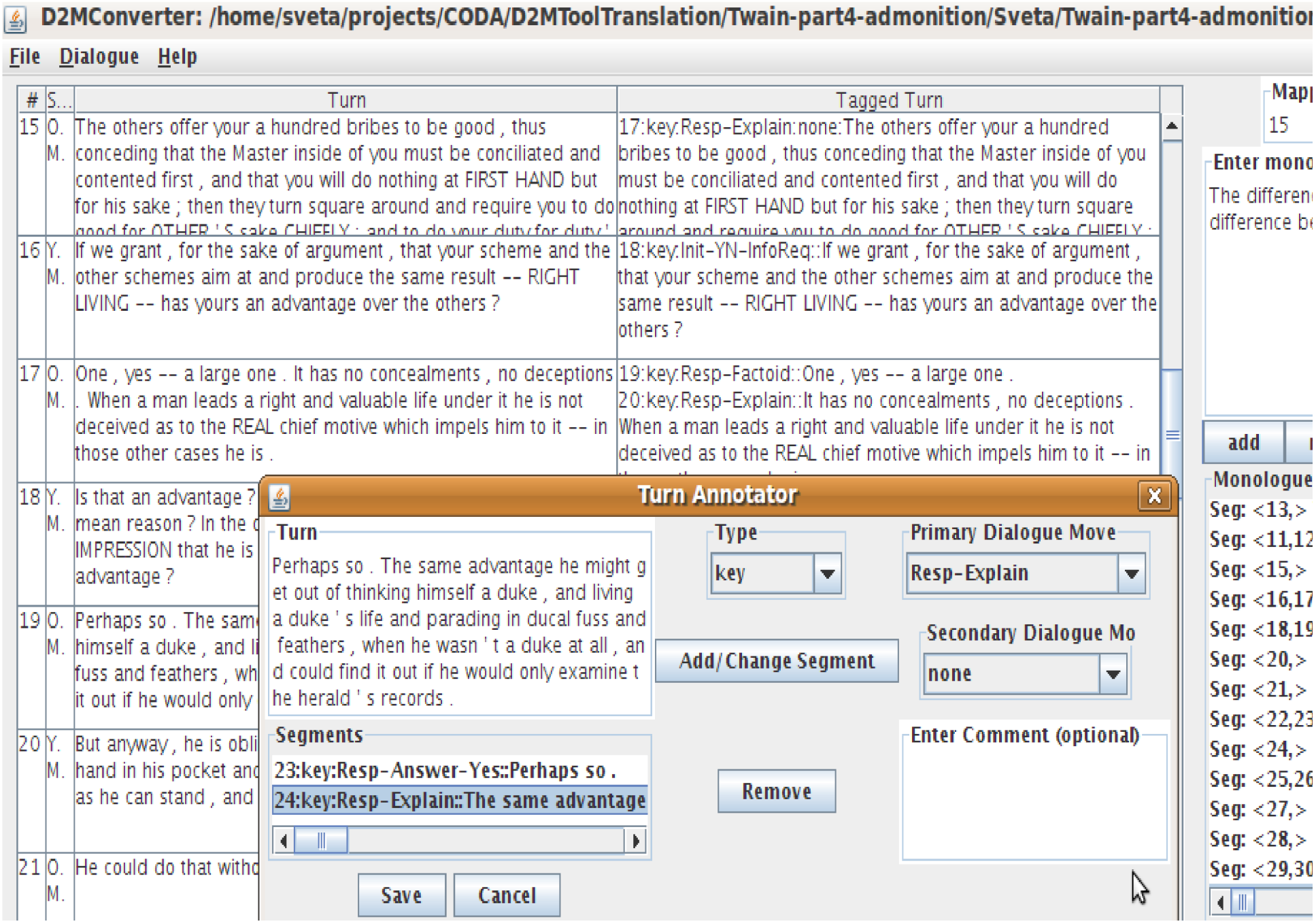
- Take existing dialogues and convert them to monologue
  - Use the dialogues written by the professional authors
  - Use corpora which can be released to research community
  - Choose dialogues that are easy to paraphrase as monologues, these are expository and non-dramatic dialogues:
    - Mark Twain ("Who is Man" 19th century literature)
    - Berkeley (18th century philosophical literature)
    - Lewis (20th century, published in a philosophical journal)
    - Gurevich(20th century, published in a computer science conference proceedings)

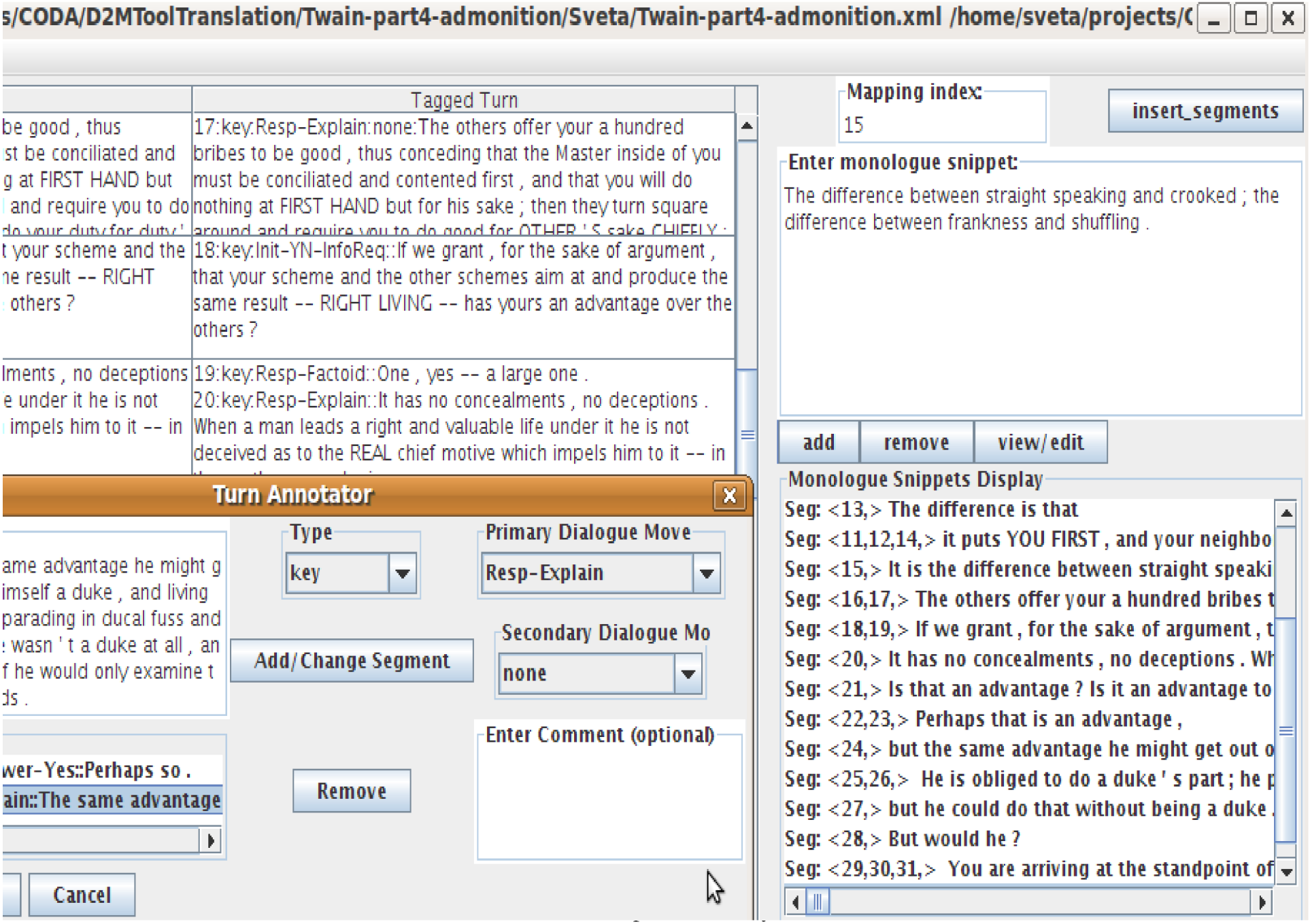
Alternatively, we could take monologues and convert them to dialogue, but then we would miss the *professional author* requirement

Corpus Construction

1. Dialogue annotation and translation
   - Segment dialogue
   - Annotate dialogue acts
   - Translate dialogue into monologue
2. Monologue annotation
   - Annotate discourse relations in monologue

Created a custom-designed annotation tool


D2MConverter: /home/sveta/projects/CODA/D2MToolTranslation/Twain-part4-admonition/Sveta/Twain-part4-admonitio
File
Dialogue
Help
S... Turn Tagged Turn
15 O. M. The others offer your a hundred bribes to be good , thus conceding that the Master inside of you must be conciliated and contented first , and that you will do nothing at FIRST HAND but for his sake ; then they turn square around and require you to do
17:key:Resp-Explain:none:The others offer your a hundred bribes to be good , thus conceding that the Master inside of you must be conciliated and contented first , and that you will do nothing at FIRST HAND but for his sake ; then they turn square
16 Y. M. If we grant , for the sake of argument , that your scheme and the other schemes aim at and produce the same result -- RIGHT LIVING -- has yours an advantage over the others ?
18:key:Init-YN-InfoReq::If we grant , for the sake of argument , that your scheme and the other schemes aim at and produce the same result -- RIGHT LIVING -- has yours an advantage over the others ?
17 O. M. One , yes -- a large one . It has no concealments , no deceptions . When a man leads a right and valuable life under it he is not deceived as to the REAL chief motive which impels him to it -- in those other cases he is .
19:key:Resp-Factoid::One , yes -- a large one .
20:key:Resp-Explain::It has no concealments , no deceptions . When a man leads a right and valuable life under it he is not deceived as to the REAL chief motive which impels him to it -- in
18 Y. M. Is that an advantage ?
mean reason ? In the
IMPRESSION that he is
advantage ?
19 O. M. Perhaps so . The sam
himself a duke , and li
fuss and feathers , wh
it out if he would only
20 Y. M. But anyway , he is obli
hand in his pocket and
as he can stand , and
21 O. M. He could do that witho
Turn Annotator
Turn
Perhaps so . The same advantage he might g et out of thinking himself a duke , and living a duke ' s life and parading in ducal fuss and feathers , when he wasn ' t a duke at all , an d could find it out if he would only examine t he herald ' s records .
Type
key
Primary Dialogue Move
Resp-Explain
Secondary Dialogue Mo
none
Add/Change Segment
Segments
23:key:Resp-Answer-Yes::Perhaps so .
24:key:Resp-Explain::The same advantage
Remove
Enter Comment (optional)
Save
Cancel
Mapp
15
Enter mono
The differen
difference be
add
Monologue
Seg: <13,>
Seg: <11,12
Seg: <15,>
Seg: <16,17
Seg: <18,19
Seg: <20,>
Seg: <21,>
Seg: <22,23
Seg: <24,>
Seg: <25,26
Seg: <27,>
Seg: <28,>
Seg: <29,30


s/CODA/D2MToolTranslation/Twain-part4-admonition/Sveta/Twain-part4-admonition.xml /home/sveta/projects/C
Tagged Turn
be good , thus
st be conciliated and
g at FIRST HAND but
and require you to do
17:key:Resp-Explain:none:The others offer your a hundred bribes to be good , thus conceding that the Master inside of you must be conciliated and contented first , and that you will do nothing at FIRST HAND but for his sake ; then they turn square
t your scheme and the
ne result -- RIGHT
others ?
18:key:Init-YN-InfoReq::If we grant , for the sake of argument , that your scheme and the other schemes aim at and produce the same result -- RIGHT LIVING -- has yours an advantage over the others ?
lments , no deceptions
e under it he is not
impels him to it -- in
19:key:Resp-Factoid::One , yes -- a large one .
20:key:Resp-Explain::It has no concealments , no deceptions . When a man leads a right and valuable life under it he is not deceived as to the REAL chief motive which impels him to it -- in
Mapping index:
15
insert_segments
Enter monologue snippet:
The difference between straight speaking and crooked ; the difference between frankness and shuffling .
add
remove
view/edit
Monologue Snippets Display
Seg: <13,> The difference is that
Seg: <11,12,14,> it puts YOU FIRST , and your neighbo
Seg: <15,> It is the difference between straight speaki
Seg: <16,17,> The others offer your a hundred bribes t
Seg: <18,19,> If we grant , for the sake of argument , t
Seg: <20,> It has no concealments , no deceptions . Wh
Seg: <21,> Is that an advantage ? Is it an advantage to
Seg: <22,23,> Perhaps that is an advantage ,
Seg: <24,> but the same advantage he might get out o
Seg: <25,26,> He is obliged to do a duke ' s part ; he p
Seg: <27,> but he could do that without being a duke
Seg: <28,> But would he ?
Seg: <29,30,31,> You are arriving at the standpoint of
Turn Annotator
ame advantage he might g
imself a duke , and living
parading in ducal fuss and
wasn ' t a duke at all , an
f he would only examine t
ds .
wer-Yes::Perhaps so .
ain::The same advantage
Type
key
Primary Dialogue Move
Resp-Explain
Secondary Dialogue Mo
none
Add/Change Segment
Remove
Enter Comment (optional)
Cancel

Dialogue Annotations

1. Identify *Decorative* and *key* segments
   - *Decorative* information is included by the dialogue author to liven up the dialogue and create natural transitions in dialogue (e.g. “Wait!”, “Well ”, “You think so?”)
     - Decorative segments are not translated into dialogue
     - Correspond to dialogue management acts (Bunt's DIT theory)
   - *Key* information contributes to the main idea that the author is trying to get across. It is directly relevant to the main purpose of the dialogue
     - Key segments are translated into dialogue
     - Most segments are *key*
2. Assign dialogue act to *key* segments

Dialogue Act Annotations

Dialogue Act annotations are based on the scheme originally defined by Carletta et al. (1997)

1. Initiation Dialogue acts:
   - Yes/no info request
   - Factoid info request (who, when, where question)
   - Complex info request (why, how, what question)
   - Clarification request
2. Responding dialogue acts:
   - Agree
   - Contradict
   - Answer-factoid
   - Answer-yes/no
3. Explanation

Compute inter-annotator agreement:

- Dialogue act segmentation agreement: 91 %
- Overall dialogue act tagging agreement kappa=.82

Translation Guidelines

- Annotators (authors) write a close paraphrase of the dialogue
  - No new information is introduced in the monologue
  - No information that is in dialogue is omitted in the monologue

Original Dialogue	Annotation	
Utterance	Dialogue Act	Monologue
OM: He felt well? YM: One can not doubt it.	YN-Info-Request Explain	One can not doubt that he felt well.
OM: Are the metals suddenly deposited in the ores ? YM: No -- YM: it is the patient work of countless ages.	Init-YN-InfoReq Resp-Answer-No Explain	The metals are not suddenly deposited in the ores. It is the patient work of countless ages.

Monologue Annotation

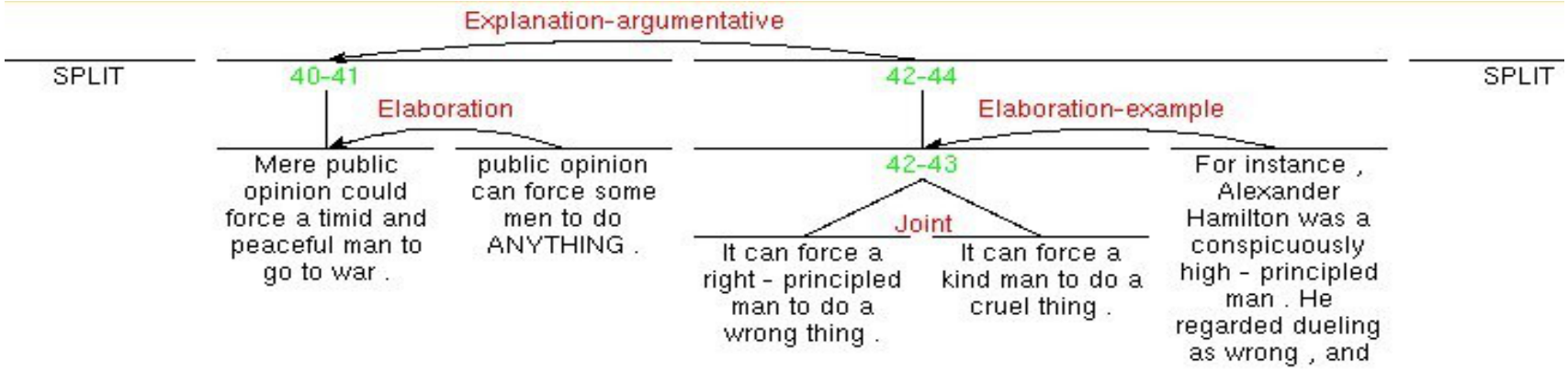
Based on Carlson et al. (2001) RST annotation scheme which defines

- Text segmentation into elementary discourse units
- Discourse (rhetorical) relation assignment between spans of text

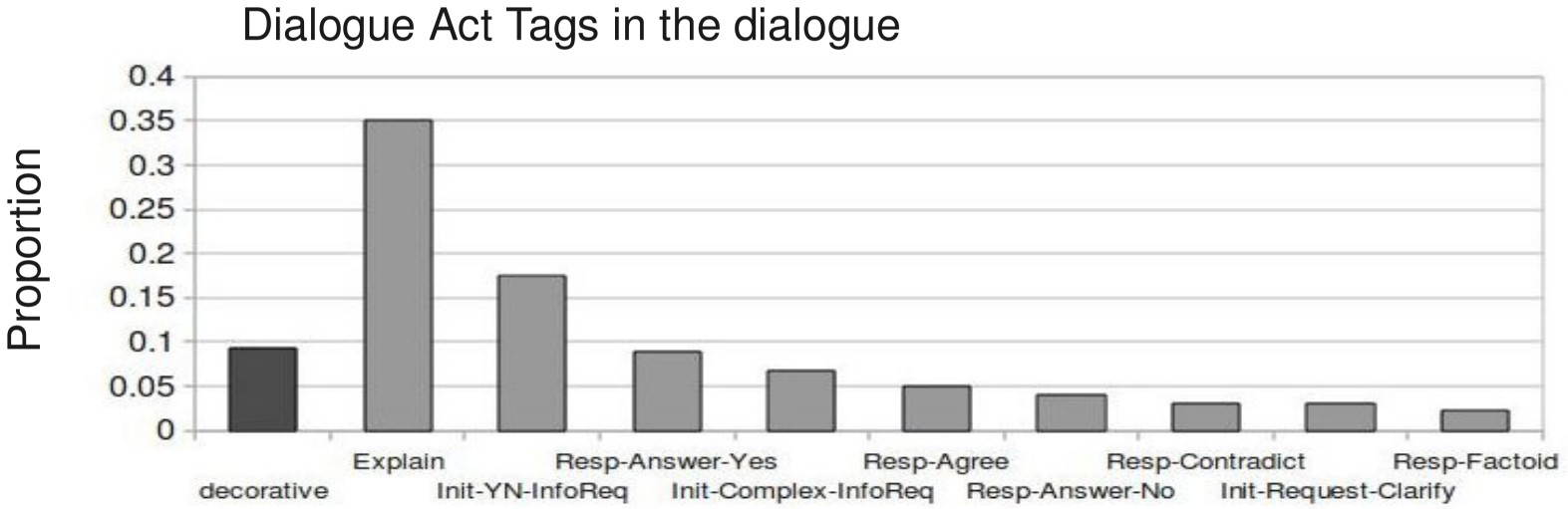
Modifications to the scheme:

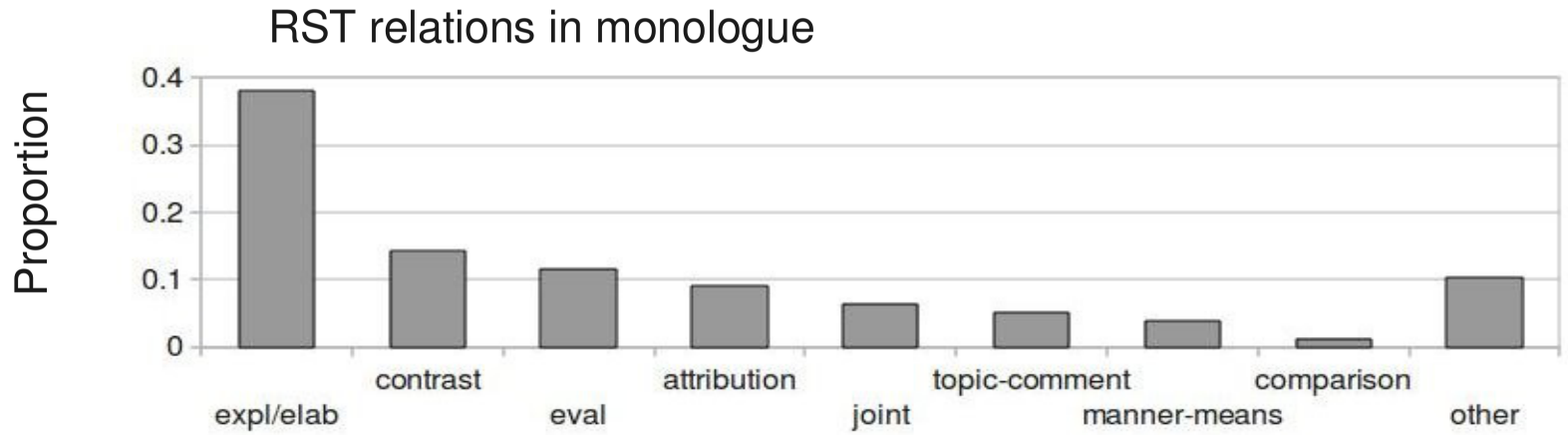
- Adapted the set of tags: 2-level annotation (coarse-grain 17 tags, fine-grain 26 tags)
- Segment monologue into paragraphs, annotate relations within a paragraph.
- Inter-annotator agreement kappa=.68 (with the pair of most-confusable tags merged)

Use O'Donnel (2000) annotation tool:



Corpus Statistics





Examples of Structural M2D Rules

	RST relation in monologue	*→ Dialogue act sequence*
Rule	**Attribution**	YN-Info-Request Explain
Example (Twain)	[*One can not doubt*] [that he felt well]	He felt well? One can not doubt it.

Rule	**Contrast**	Init-YN-InfoReq Resp-Answer-No Explain
Example (Twain)	[The metals are not suddenly deposited in the ores.] [It is the patient work of countless ages.]	Are the metals suddenly deposited in the ores ? No -- it is the patient work of countless ages.

Applying M2D Mapping Rules

- Examples of Attribution relations extracted from the corpus, applied to

Attribution(*"John said", "that he likes apples"*)

- **RULE1: Attribution → A: Init-YN-InfoReq; B: Resp-Yes Explain**
  - A: Does John like apples?
  - B: Yes, he said so.

- **RULE2: Attribution → A: Init-YN-InfoReq; B: Resp-Yes**
  - A: Did John say that he like apples?
  - B: Yes

- **RULE3: Attribution → A: Explain; B: Resp-Agree**
  - A: John said that he likes apples.
  - B: Yes, he did.

Conclusions

- Described and motivated construction of a parallel monologue-dialogue corpus
- Adopted, applied, and evaluated annotation schemes
  - dialogue annotation scheme to authored dialogues
  - RST annotations to monologue
- Final corpus will contain 1000 dialogue turns and will be available in June.
- http://computing.open.ac.uk/coda/

Thank you

Svetlana Stoyanchev and Paul Piwek

http://computing.open.ac.uk/coda/

Extra slides

Rule	Attribution	Explain Resp-Agree
Example (Twain)	[*A minute ago I said*] [that Hamilton fought that duel to get PUBLIC approval .]	A minute ago you said Hamilton fought that duel to get PUBLIC approval . I did