A Three-stage Disfluency Classifier for Multi Party Dialogues

Margot Mieskes[1] and Michael Strube[2]

[1]http://www.eml-d.de/english/homes/mieskes

[2]http://www.eml-research.de/∼strube

[1]European Media Laboratory GmbH, Heidelberg, Germany

[2]EML Research gGmbH, Heidelberg, Germany

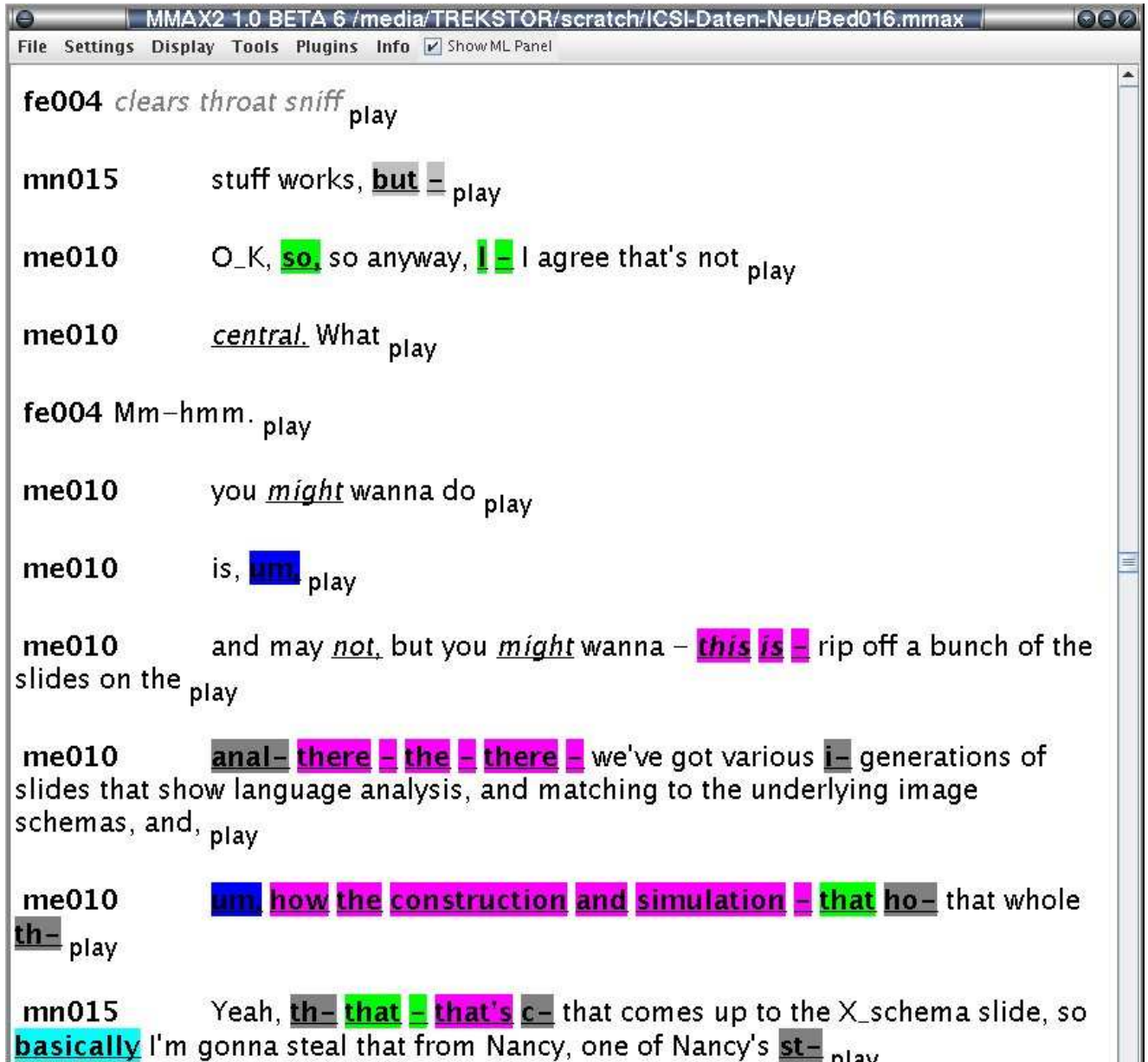
Outline

- Data
- Manual Annotation
- Interannotator Agreement κ and κ_j
- Experiments on automatic detection and classification
- Conclusion & Outlook

Disfluency Classes

- Non-lexicalized Filled Pauses (NLFP): *um, uh, ah*
- Lexicalized Filled Pauses (LFP): *like, well*
- repairs (repai): Well *they – they have s- they have* the close talking microphones for each of us
- verbatim repetitions (repet): I know *you were – you were* doing that
- abandoned words (abw): *w-, h-, shou-*
- abandoned utterances (abutt): the newest version after you comments, *and –*

Disfluency Classes



Manual Annotation Evaluation

type	relative frequency
NLFP	23.6
LFP	23.4
repet	14.5
repai	17.9
abw	7.0
abutt	13.5
κ	0.952

Manual Annotation Evaluation

Token(s)		abutt	abw	nlfp	lfp	repet	repai	none
like					3			
I'm		2					1	
Eh-				3				
tried to -		2					1	
and that would		2					1	
um-			1	2				
So w-		1					1	1
Well -		3						
somebody'll		3						
that's uh		1					1	1
and that would		1				1		1
and then		3						

Manual Annotation Evaluation

Token(s)		abutt	abw	nlfp	lfp	repet	repai	none
like					3			
I'm		2					1	
Eh-				3				
tried to -		2					1	
and that would		2					1	
um-			1	2				
So w-		1					1	1
Well -		3						
somebody'll		3						
that's uh		1					1	1
and that would		1				1		1
and then		3						
κ	0.322							

Manual Annotation Evaluation

Token(s)		abutt	abw	nlfp	lfp	repet	repai	none
like					3			
I'm		2					1	
Eh-				3				
tried to -		2					1	
and that would		2					1	
um-			1	2				
So w-		1					1	1
Well -		3						
somebody'll		3						
that's uh		1					1	1
and that would		1				1		1
and then		3						
κ/κ_j	0.322	0.33	-0.02	0.76	1.0	-0.02	0.16	0.09

Manual Annotation Evaluation

Token(s)		abutt	abw	nlfp	lfp	repet	repai	none
like					3			
I'm		2					1	
Eh-				3				
tried to -		2					1	
and that would		2					1	
um-			1	2				
So w-		1					1	1
Well -		3						
somebody'll		3						
that's uh		1					1	1
and that would		1				1		1
and then		3						
κ/κ_j Example	0.322	0.33	-0.02	0.76	1.0	-0.02	0.16	0.09
κ/κ_j Dataset	0.952	0.85	0.96	0.99	0.98	0.98	0.78	

Automatic Classification – Script Based

- Detects nlfp based on lexicon and POS tags
- Detects abw based on transcription with “-”
- Detects repet based on a script
  - not limited in length – potentially 0.5*length of utterance long
  - iterative process: one-item repet, two-item repet, ...
- Upon detection and classification disfluency is removed for further analysis

Automatic Classification – Script Based

- Detects nlfp based on lexicon and POS tags
- Detects abw based on transcription with “-”
- Detects repet based on a script
  - not limited in length – potentially 0.5*length of utterance long
  - iterative process: one-item repet, two-item repet, ...
- Upon detection and classification disfluency is removed for further analysis

DisflType	prec	rec	f
nlfp	89.56	98.66	93.89
repet	74.64	93.36	82.95
abw	89.99	99.19	94.37

Machine Learning Based

- part-of-speech tag
- length of the utterance considered
- gender of the speaker
- native or non-native speaker
- position of the current utterance in the meeting
- talkativity features like average length of segments, number of segments uttered etc.

Decision Tree based learner/classifier

Binary Classification

type	accuracy	prec	rec	f
non oversampled				
disfluent	88.5	75.3	55.8	64.1
non-disfluent		90.6	95.9	93.1
oversampled				
disfluent	84.3	61.9	70.2	65.8
non-disfluent		91.5	88.1	89.8

Binary Classification

type	accuracy	prec	rec	f
non oversampled				
disfluent	89.7	80.7	58.4	67.7
non-disfluent		91.1	96.8	93.9
oversampled				
disfluent	80.5	54.3	60.8	57.4
non-disfluent		88.9	86.0	87.4

Full Classification

disfl class	accuracy	prec	rec	f
NLFP	86.4	55.5	45.5	50.0
LFP		64.3	51.4	57.1
abutt		29.8	4.5	7.8
abw		67.3	79.6	72.9
repai		45.2	12.6	19.7
repet		64.7	50.0	56.4
none		89.8	97.3	93.2

Full Classification

Classification using previous knowledge

disfl class	prec	rec	f
NLFP	89.56	98.66	93.89
REPET	74.64	93.36	82.95
ABW	89.99	99.19	94.37

Full Classification

Classification using previous knowledge

disfl class	prec	rec	f
NLFP	89.56	98.66	93.89
REPET	74.64	93.36	82.95
ABW	89.99	99.19	94.37
LFP	83.4	91.1	87.1
abutt	76.2	73.0	74.6
repai	84.3	77.0	80.5

Feature Ranks

- POS tags
  - current
  - preceding
  - following
- length of the current utterance
- distance to previous disfluency
- average length of utterances by the current speaker
- . . .
- distance to previous
  - NLFP
  - REPET
  - ABW
- . . .
- gender

Example Rule 1

if

segmentLength <= 11 **&** tag = UH **&** 1prevTag = CC **&** previousDisfl = yes

THEN ABUTT

Example Rule 1

if

segmentLength <= 11 **&** tag = UH **&** 1prevTag = CC **&** previousDisfl = no

THEN LFP

Example Rule 2

if

segmentLength <= 11 **&** tag = INP **&** 1prevTag = IN **&** 2nextTag = INP **&** 1nextTag = IN **&**

distanceToDisflStart <= 1

THEN ABUTT

Example Rule 2

if

segmentLength <= 11 **&** tag = INP **&** 1prevTag = IN **&** 2nextTag = INP **&** 1nextTag = IN **&**
distanceToDisflStart > 1 **&**
distanceToDisflStart <= 3 **&**
segmentsSF <= 48

THEN ABUTT

Example Rule 2

if

segmentLength <= 11 **&** tag = INP **&** 1prevTag = IN **&**
2nextTag = INP **&** 1nextTag = IN **&**
distanceToDisflStart > 1 **&**
distanceToDisflStart <= 3 **&**
segmentsSF > 48 **&**
gender = f
THEN LFP

Example Rule 2

if

segmentLength <= 11 **&** tag = INP **&** 1prevTag = IN **&**
2nextTag = INP **&** 1nextTag = IN **&**
distanceToDisflStart > 1 **&**
distanceToDisflStart <= 3 **&**
segmentsSF > 48 **&**
gender = m **&**
averageSegment <= 7

THEN LFP

Example Rule 2

if

segmentLength <= 11 **&** tag = INP **&** 1prevTag = IN **&**
2nextTag = INP **&** 1nextTag = IN **&**
distanceToDisflStart > 1 **&**
distanceToDisflStart <= 3 **&**
segmentsSF > 48 **&**
gender = m **&**
averageSegment > 7

THEN ABUTT

Conclusion & Outlook

- more detailed analysis of the manual annotation procedure
- three stage procedure for detection and classification of disfluencies
- more fine-grained distinction than in previous work
- better performance than comparison work
- comparison to descriptive work on the phenomenon of disfluencies
- features inspired by descriptive work were not relevant for the detection (e.g. gender)
- might be due to two party vs. multi party dialogues

Acknowledgments

Thanks to

- Deutsche Forschungsgemeinschaft
- Klaus Tschira Stiftung
- Our annotators

Software and Data

Annotation Tool MMAX2: `http://mmax2.sourceforge.net/`

Octave/Matlab Script for κ_j calculation:

`http://projects.villa-bosch.de/nlpsoft/`

Disfluency Annotation:

`http://www.eml-r.org/english/research/nlp/download/index.php`