Annotation and Analysis of Emotionally Relevant Behavior in the ISL Meeting Corpus

Kornel Laskowski and Susanne Burger

interACT

LREC 2006

At a Glance…

1. (1 slide of) Motivation
2. (3 slides about the) Data & Schema
3. (2 slides of) Annotation & Agreement Results
4. (4 slides of) Analysis

Motivation

1. Looking for ***naturally occurring*** emotion in multiparty conversation

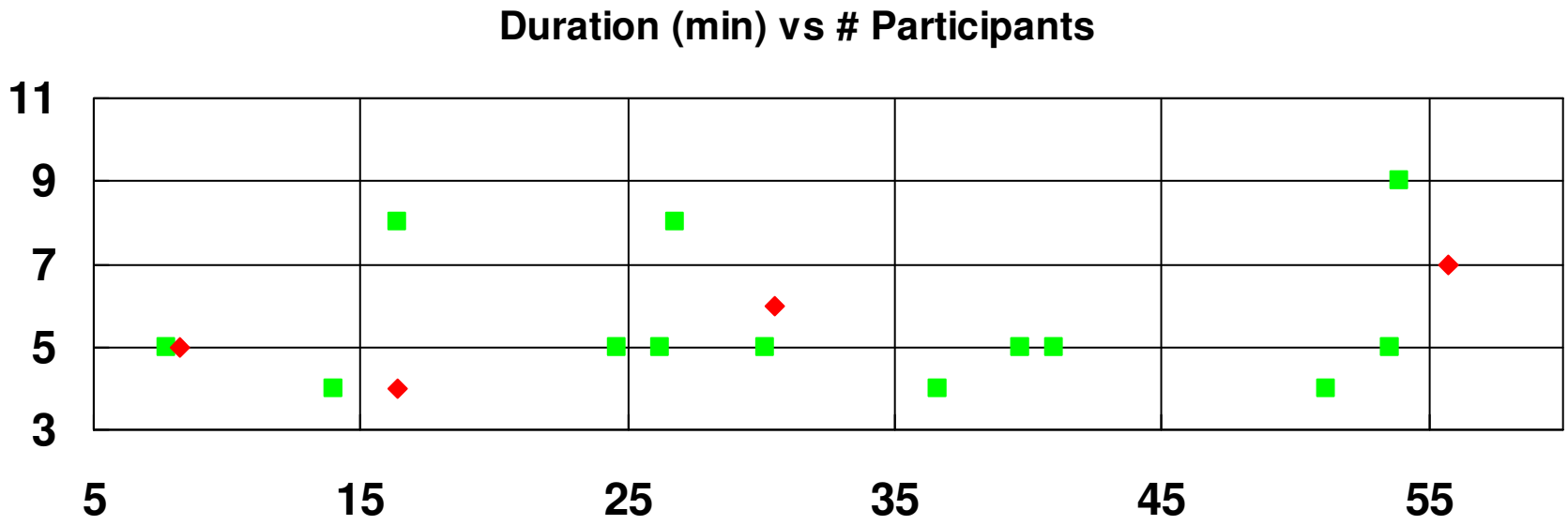
2. Pragmatic, computationally tractable scheme – theoretical validity of secondary importance
   - at the unit of the speaker contribution (utterance/turn/etc)

3. Application prototype: browsing meeting records by emotional terms
   - "Naïve" labelers: want anyone ***to be able to use the system***, not just experts

Data: ISL Meeting Corpus (Volume 1)

- 18 meetings
- 9hrs 38min of multichannel audio
- ~ 5.1 average # of participants
- 31 unique participants

- Contains both:
  - Natural, work-related
  - Induced, game-playing or discussion



Annotation Schema Development

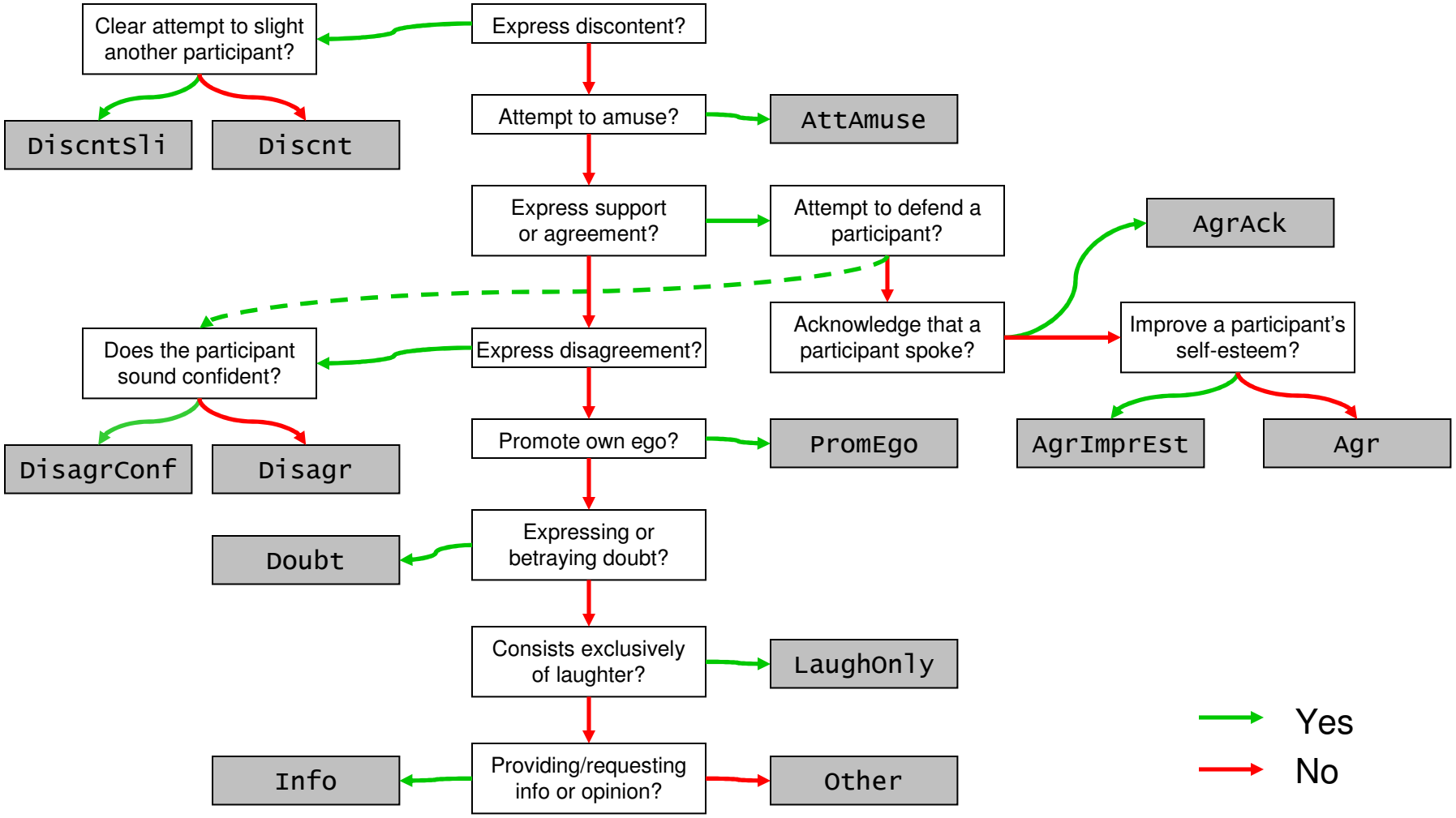
- 3 years of iterative development and annotation
- Pursued within the European projects PF-STAR and CHIL

- **Stage 1**
  - Explored assignment of “free”, open-set labels, by 3 annotators
  - Found that observers tend to use descriptors for ***what participants are doing*** (eg. complaining), rather than *how participants are feeling*

- **Stage 2**
  - Manually clustered the hundreds of labels thus obtained
  - Essentially a dialogue act annotation scheme, whose focus is ***the exchange of socio-emotional capital***, rather than of information (info-request, info-reply, etc)

- **Stage 3**
  - Placed the 13 labels in a decision tree for behavior annotation
  - A separate three-way discrete annotation of emotional valence

Behavior Annotation Scheme



Annotation of Emotionally Relevant Behavior

- 13221 speaker contributions, 3 annotators
  - 59.5%: all 3 annotators agree (unanimity)
  - 35.4%: 2 annotators agree (majority)
  - 6.1%: no agreement

- Pair-wise interlabeler kappa: $0.56 \leq \kappa \leq 0.59$

- Of those speaker contributions with a majority label:
  - 8% are `LaughOnly`
  - 85% are one of `Info`, `AgrAck` and `Agr`
  - 7% are all remaining behaviors
- All behaviors except `Discntsli` receive a vote at least 1% of the time

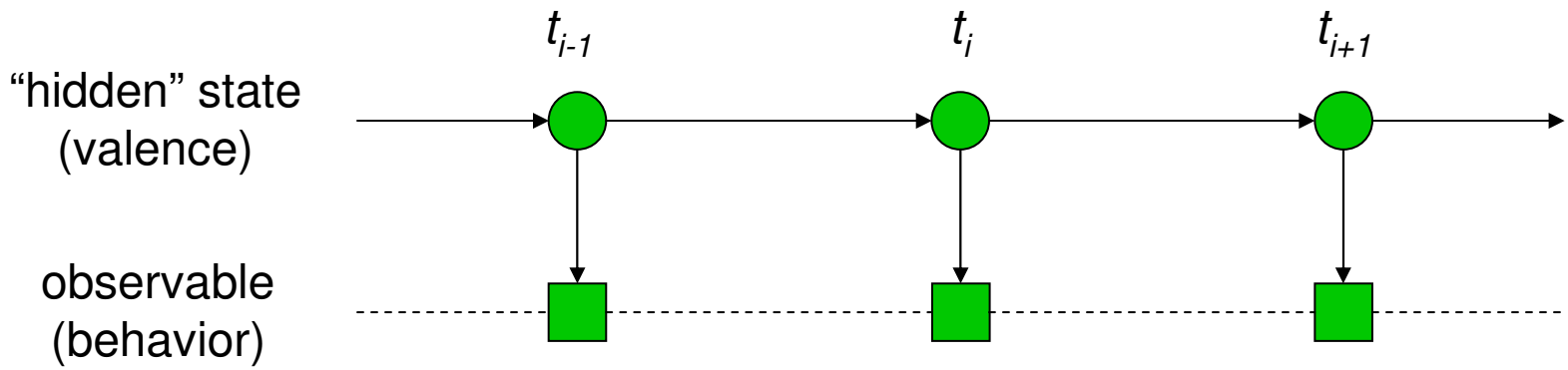
Annotation of Emotional Valence

Majority vs Minority Valence Labels

	Negative	Neutral	Positive	*Majority Votes*
Negative	**22**	**85**	**10**	*117*
Neutral	**354**	**9361**	**1142**	*10751*
Positive	**49**	**1887**	**235**	*2155*
Minority Votes	*403*	*1972*	*1152*	

- Of 13221 speaker contributions,
  - 99.4% exhibit a majority
  - ~ 81% are of `Neutral` valence
  - ~ 16% are of `Positive` valence
- Highest interlabeler kappa: 0.67
- We had a goat labeler, whose kappa values with the other two labelers were ~0.15
  - cf. paper for analysis with respect to a larger set of labelers

Intra-Speaker State-to-Action Association



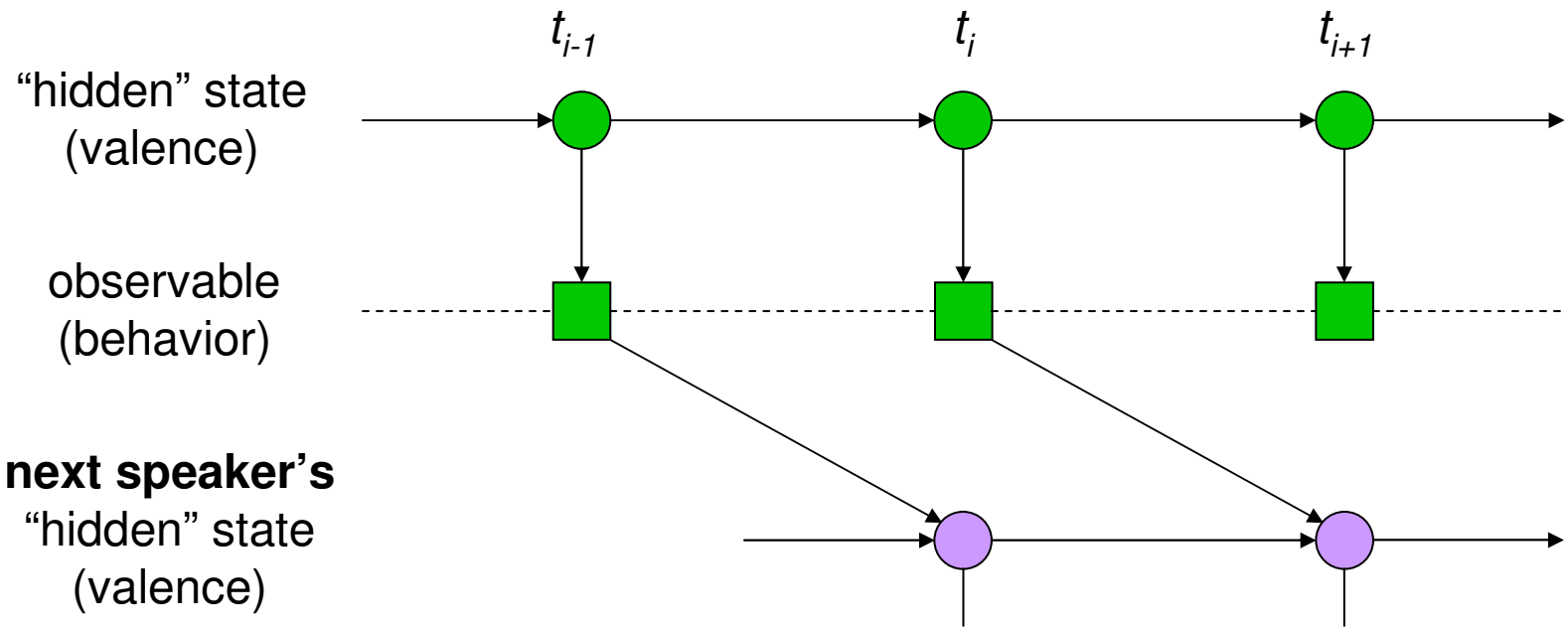
- We consider the evolution of participant valence over time
- Speaker contributions are observable actions, from which the (hidden) valence state must be **inferred**
- This corresponds to the task given to labelers: to describe the behavior embodied in speakers' dialogue contributions, and to infer their emotional valence from this behavior and its causal context

Intra-Speaker State-to-Action Association

	Negative	Neutral	Positive
DiscntSli	8	13	25
Discnt	37	132	69
DisagrConf	20	231	45
Disagr	11	258	39
Doubt	10	524	42
Other	5	218	21
Info	81	5452	897
AgrAck	10	1455	91
Agr	11	1398	218
PromEgo	5	138	24
AgrImprEst	3	174	64
AttAmuse	6	66	360
LaughOnly	6	56	998

- χ^2 test (13221 spkr contributions)
  H_0: no association between ***a given speaker's*** valence and behavior
- Found statistically significant association:
  - below chance, $p < 0.001$
  - below chance, $p < 0.01$
  - above chance, $p < 0.01$
  - above chance, $p < 0.001$
- Valence labels by labeler 2, behavior labels by labeler 3; results for other pairings similar

Inter-Speaker Action-to-State Association



- We additionally consider how one speaker's behavior affects the next speaker's state

Inter-Speaker Action-to-State Association

	Negative	Neutral	Positive
DiscntSli	3	28	22
Discnt	9	165	79
DisagrConf	14	275	71
Disagr	5	291	45
Doubt	6	261	48
Other	5	68	30
Info	107	6001	1319
AgrAck	6	471	87
Agr	19	761	167
PromEgo	6	120	26
AgrImprEst	3	135	64
AttAmuse	3	229	416
LaughOnly	4	200	288

- χ^2 test (11857 spkr contributions)

  H_0: no association between ***one speaker's*** behavior and the ***next speaker's*** valence

- Found statistically significant association:
  - below chance, $p < 0.001$
  - below chance, $p < 0.01$
  - above chance, $p < 0.01$
  - above chance, $p < 0.001$

- Valence labels by labeler 2, behavior labels by labeler 3; results for other pairings similar

Intra-Speaker vs Inter-Speaker

	Negative	Neutral	Positive
DiscntSli	8	13	25
Discnt	37	132	69
DisagrConf	20	231	45
Disagr	11	258	39
Doubt	10	524	42
Other	5	218	21
Info	81	5452	897
AgrAck	10	1455	91
Agr	11	1398	218
PromEgo	5	138	24
AgrImprEst	3	174	64
AttAmuse	6	66	360
LaughOnly	6	56	998

	Negative	Neutral	Positive
DiscntSli	3	28	22
Discnt	9	165	79
DisagrConf	14	275	71
Disagr	5	291	45
Doubt	6	261	48
Other	5	68	30
Info	107	6001	1319
AgrAck	6	471	87
Agr	19	761	167
PromEgo	6	120	26
AgrImprEst	3	135	64
AttAmuse	3	229	416
LaughOnly	4	200	288

Conclusions

- We've proposed a set of mutually exclusive, emotionally relevant behaviors:
  - Categories were obtained by analyzing "free", open-set annotations: data- rather than theory- driven
  - Nesting the categories in an annotation decision tree improved agreement
  - Annotator agreement is on par with similar tasks on similar data reported elsewhere

- In the ISL Meeting Corpus,
  - 15% of speaker contributions embody behaviors which are deemed emotionally relevant; half of these consist primarily of laughter
  - In 16% of speaker contributions, an annotator majority infers `Positive` valence in the speakers
  - Most behaviors show statistically significant association with valence
  - For a minority of behaviors, that association is also strong