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

Kornel Laskowski[*] **and Susanne Burger**[†]

[*]interACT
University of Karlsruhe, Karlsruhe, Germany
kornel@ira.uka.de
[†]interACT
Carnegie Mellon University, Pittsburgh PA, USA
sburger@cs.cmu.edu

Abstract

We present an annotation scheme for emotionally relevant behavior at the speaker contribution level in multiparty conversation. The scheme was applied to a large, publicly available meeting corpus by three annotators, and subsequently labeled with emotional valence. We report inter-labeler agreement statistics for the two schemes, and explore the correlation between speaker valence and behavior, as well as that between speaker valence and the previous speaker's behavior. Our analyses show that the co-occurrence of certain behaviors and valence classes significantly deviates from what is to be expected by chance; in isolated cases, behaviors are predictive of valence.

1. Introduction

Multiparty conversation, and meetings in particular, are currently one of the most intensely studied types of speech corpora. They offer an opportunity for studying spontaneous speech and language in one of their most natural and unconstrained contexts, and a broad paradigm for application development. As base technologies such as speech recognition improve in this domain, researchers are turning to higher-level analysis of group interactions, including applications in summarization and understanding (Renals & Ellis, 2003). Such analysis is likely to become important in the synthesis of appropriate speech-related behaviors in machine agents, expected to assist and/or participate in human-human interaction.

Emotion represents one such higher-level phenomenon. Frequently studied in a synthetic context, and less so in realistic domains (Batliner et al., 2000) such as call centers (Steininger et al., 2002), tutoring dialogues (Litman & Forbes-Riley, 2004a) and entertainment robots (Batliner et al., 2004), emotional vocal behavior has received little attention in computational work with multiparty conversation corpora. In large part this is due to the near-absence in social settings of overt displays of the so-called "canonical" emotions (happiness, sadness, etc.), giving some credence to claims that emotion is simply not present. Even when natural speech corpora do contain expressions of emotions, they tend to exhibit much lower agreement among observers (Russell et al., 2003) than is reported for more objective categorizations of vocal activity. Additionally, it has been posited that such expressions are both directed at specific receivers (Russell et al., 2003) and that they are highly context-sensitive (Cauldwell, 2000). In spite of these and other difficulties, many researchers continue to feel that an emotional subtext plays an important role in understanding verbal human-human interactions, in particular as motivations for actions. There is therefore a clear need for tools to characterize the emotional aspects of multiparty communication.

In this paper we present the assessment of such a characterization tool, namely an annotation scheme for emotional behavior in meetings at the speaker contribution level. We also report on the annotation of meetings with emotional valence, and on the correlation between the two schemes. Previous work on the annotation of socio-emotional phenomena in meetings has given them a minor role in large, relatively complex dialogue act annotation schemes (Shriberg et al., 2004), explored multispeaker activation at the talk-spurt level in the form of hot spots (Wrede & Shriberg, 2003a), and considered participant interaction in meeting acts spanning multiple speaker contributions (Bates et al., 2005).

2. Data

The data used in this research is the entire ISL Meeting Corpus (Burger et al., 2002), as available publicly from the LDC. The corpus consists of 18 meetings, with an average duration of 34 minutes and with 5 participants on average. The conversations are natural, spanning a spectrum from work-related meetings that would have been held anyways, to recordings of ISL lab members in social settings such as game playing. The data were recorded as part of a previous project and not originally intended for research on emotion. They are accompanied by manual segmentation at the speaker contribution level and orthographic transcriptions.

3. Annotation Scheme

The set of emotionally relevant behavior labels used in this work is the outcome of a previous manual clustering exercise, following open label set annotation of a small number of ISL meetings by three naive labelers. Surprisingly, we found that when allowed to annotate in this way, labelers tend to describe how people are *behaving* rather than how people are *feeling* (Laskowski & Burger, 2005). The applied annotation scheme is shown in Figure 1; placing the mutually exclusive labels at the leaves of a decision tree, and using cryptic one-letter names rather than a flat ontology with real words as class names, appears to improve

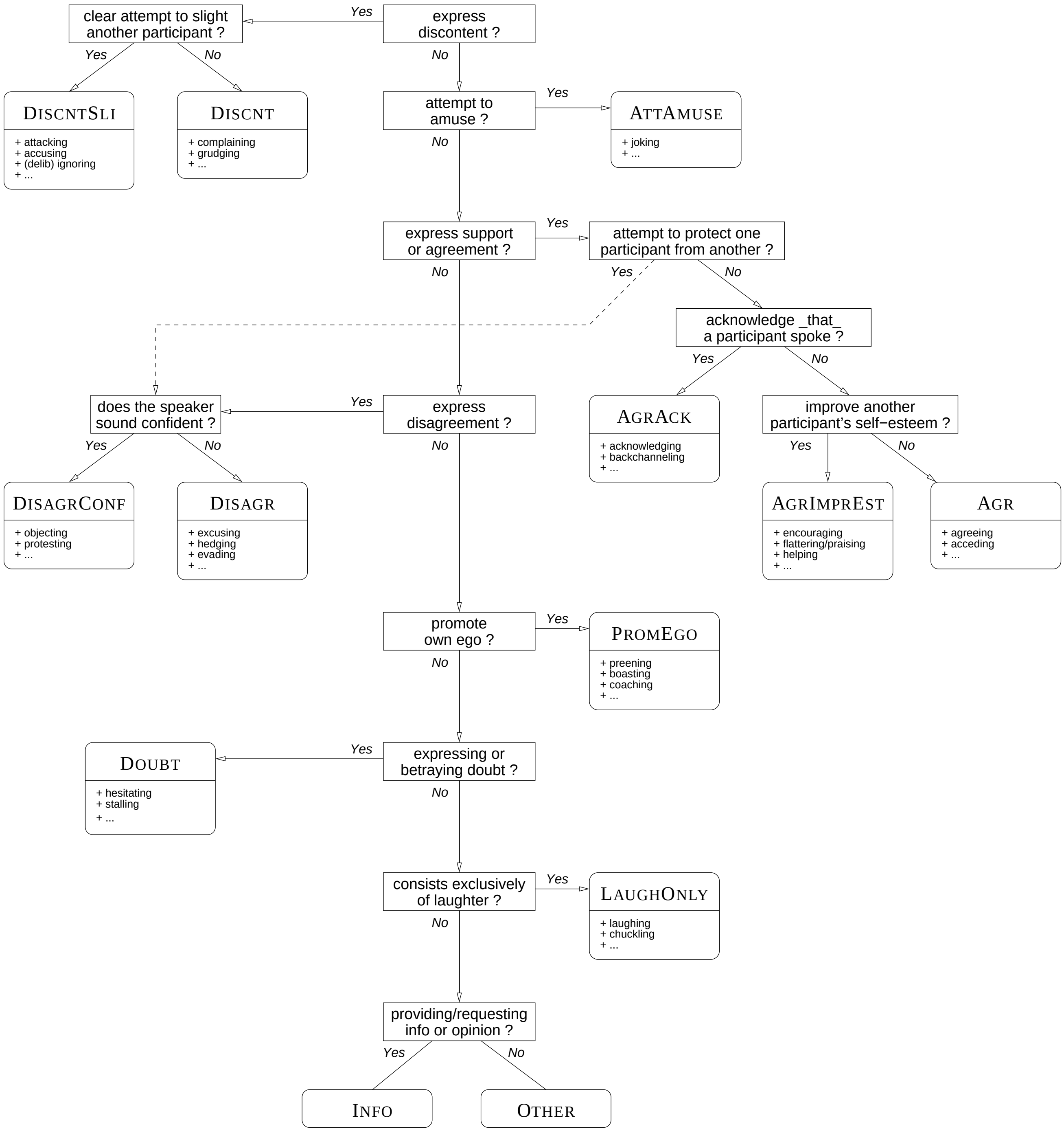

Figure 1: *Decision tree for annotation of emotionally relevant behavior. The version given to labelers used random letters at the leaves, without examples.*

inter-rater agreement. We hypothesize that this is because annotators are forced to traverse the tree, answering questions at each node, rather than rely on their interpretation of lexical class names. Each class is thereby defined by the specific sequence of yes/no answers given.

The structure of the tree is a deliberate attempt to focus on emotionally interesting phenomena. For example, the first question, *Express discontent?*, is meant to identify emotionally important behaviors, which are rare and which might otherwise be lost if the labeler was allowed to answer other questions first (ie. *Providing/requesting info or opinion?*). Certain categories exist to eliminate certain behaviors prior to further questioning. In general, behaviors which we felt to be positive interaction behaviors are on the right side of the tree, those which we felt to be negative are on the left (Lazarus, 1991). The single dashed line from the right side to the left expresses our impression that when protecting one participant from another, a third party objects to the first participant's behavior in the same way they might object to an opinion.

For emotional valence, we chose a three-way distinction between Positive, Negative, and Neutral, as frequently

	DiscntSli	Discnt	DisagrConf	Disagr	Doubt	Other	Info	AgrAck	Agr	PromEgo	AgrImprEst	AttAmuse	LaughOnly	Majority Votes per behavior
DiscntSli	**8**	5	1	1	0	0	8	0	0	0	1	1	1	26
Discnt	1	**3**	0	1	1	5	12	7	0	0	0	1	0	31
DisagrConf	3	2	**49**	26	4	1	106	2	4	0	0	0	0	197
Disagr	0	1	12	**9**	6	2	42	3	3	0	0	0	0	78
Doubt	0	0	1	3	**34**	29	35	17	6	0	0	0	0	125
Other	0	12	0	2	26	**155**	35	13	1	1	5	0	6	256
Info	21	139	195	192	203	191	**4973**	229	477	121	103	276	4	7124
AgrAck	0	13	4	4	120	53	161	**901**	428	0	13	1	11	1709
Agr	0	1	3	6	9	4	337	498	**687**	2	13	4	2	1566
PromEgo	0	0	0	0	0	0	3	0	0	**2**	0	0	0	5
AgrImprEst	0	0	0	0	0	0	14	7	5	0	**6**	0	0	32
AttAmuse	3	1	1	1	0	1	161	1	1	1	0	**41**	0	212
LaughOnly	0	0	1	0	1	22	4	23	1	0	0	1	**1004**	1057
Minority Votes per behavior	28	174	218	236	370	308	918	800	926	125	135	284	24	

Table 1: *Annotator majority (rows) vs annotator minority (columns) voting, emotionally relevant behavior, for the 93.9% of speaker contributions for which an annotator majority does exist. The rightmost column represents the number of instances of a majority, per behavior category; the number of instances of a strict minority, per behavior category, is shown in the bottom row. Numbers along the diagonal represent unanimity among the three annotators.*

proposed elsewhere, ie. (Litman & Forbes-Riley, 2004a). We chose not to annotate emotional activation, studied in the context of meetings in (Wrede & Shriberg, 2003a), as there was not as much intra-speaker variability in our data relative to the seemingly larger differences between baselines for different speakers. It has been reported (Russell et al., 2003) that for naturally occurring speech, listeners find it more easy to distinguish between activation levels than they do between valence levels, which we focus on in this work.

4. Annotation Process

The task of annotating the entire ISL Meeting Corpus, in terms of both emotionally relevant behavior and valence as described, was given to three labelers. The labelers did not know the participants in the corpus meetings (except one participant). While experienced with orthographic transcription of speech, they were not previously exposed to the annotation of emotion. They were given only the decision tree with no other guidelines, and mechanical instructions for recording their labels. Emotionally relevant behavior was annotated first, with labelers having access to both orthographic transcription and the audio; emotional valence was annotated second, with labelers having access to orthographic transcription, audio, and their own behavior labels.

5. Interlabeler Agreement

5.1. Emotionally Relevant Behavior

Of the 13221 speaker contributions in the corpus, 803 (6.1%) exhibit no majority (each of three labelers assigns a different category). Of the remaining 12418 speaker contributions for which a majority does exist, 7872 (63.4%, or 59.5% of the total) exhibit complete agreement among all three labelers. The distribution of the assignments of the minority labeler vs the assignments of the majority labelers is shown in Table 1. For lack of space, we do not show absolute agreement matrices for individual labeler pairs.

As can be seen, the majority of speaker contributions are devoid of behaviors usually associated with emotion, consisting primarily of Info, AgrAck and Agr. We note however that annotators rigorously agree on the exclusive presence of laughter: speaker contributions receiving the LaughOnly label by at least two annotators make up 8% of the corpus. Other behaviors which are interesting from an emotion point of view are more rare, together accounting for just over 7% of all speaker contributions, given majority label voting and excluding laughter. However, all categories with the exception of DiscntSli receive a vote at least 1% of the time. In general, such minority labels are often concurrent with the majority voting for one of Info, AgrAck and Agr.

In Table 2 we show absolute agreement, chance agreement (assuming labeler independence), and chance-corrected agreement in the form of the kappa statistic, for each labeler pair. The κ values for our untrained labelers lie in a tight range of 0.56 to 0.59, which we consider acceptable (Cohen, 1960), (Carletta, 1996). Agreement is notably worse than that reported for dialog act/structure coding schemes involving practiced labelers, ie. $0.75 \leq \kappa \leq 0.86$ for 4 classes in (Carletta et al., 1997) and $0.75 \leq \kappa \leq 0.82$ for 6 classes in (Shriberg et al., 2004), but it is on par with more subjective distinctions in meetings such as agreement/disagreement in talk-spurts, where $\kappa = 0.63$ for 4 classes (Galley et al., 2004).

Labelers	1&2	1&3	2&3
Absolute agreement	0.72	0.71	0.70
Chance agreement	0.34	0.29	0.32
κ coefficient	0.58	0.59	0.56

Table 2: *Interlabeler agreement on the entire ISL Meeting Corpus (13221 speaker contributions), emotionally relevant behavior.*

5.2. Emotional Valence

Of the 13221 speaker contributions, only 76 (0.58%) exhibit no majority. Of the remaining 13145 speaker contributions for which a majority does exist, 9526 exhibit unanimity. The distribution of the assignments of the minority labeler vs the assignments of the majority labelers is given in Table 3. As is shown, neutral valence accounts for 81% of speaker contributions, with an annotator majority agreeing that the proportion of negative speaker contributions is less than 1%. However, over 16% of speaker contributions receive a positive valence label from an annotator majority, which is more than was expected.

	NEGATIVE	NEUTRAL	POSITIVE	Majority Votes per valence
NEGATIVE	**22**	85	10	117
NEUTRAL	354	**9361**	1142	10751
POSITIVE	49	1887	**235**	2155
Minority Votes per valence	403	1972	1152	

Table 3: *Annotator majority (rows) vs annotator minority (columns) voting, emotional valence, for the 99.4% of speaker contributions for which an annotator majority does exist. Conventions as in Table 1.*

In Table 4 we show absolute agreement, chance agreement (assuming labeler independence), and chance-corrected agreement kappa for each labeler pair. Agreement between labelers 2 and 3 is similar to that reported elsewhere, ie. utterance-level hot spots in meetings ($0.35 \leq \kappa \leq 0.79$, 4 classes) in (Wrede & Shriberg, 2003a), as well as for valence in other domains, including tutoring dialogues ($0.40 \leq \kappa \leq 0.68$, 3 classes) in (Litman & Forbes-Riley, 2004a) and automated travel planning systems (0.47, 5 classes) in (Ang et al., 2002). However, agreement for pairs involving labeler 1 is close to chance. In spite of using naive labelers, we expected significantly better than chance agreement on what passes for non-neutral valence. In the remainder of this section, we assess this labeler's behavior in the context of more labelers which had labeled a pilot subset (Laskowski & Burger, 2005) of the ISL Meeting Corpus.

Labeler	1&2	1&3	2&3
Absolute agreement	0.77	0.79	0.89
Chance agreement	0.73	0.76	0.65
κ coefficient	0.15	0.14	0.67

Table 4: *Inter-labeler agreement on the entire ISL Meeting Corpus (13221 speaker contributions), emotional valence.*

Our earlier study pilot corpus consisted of 5 meetings, and was annotated for emotional valence in the same manner by three annotators, referred to here as A, B, and C (labeler C in that work was the same as labeler 2 in the current study). This set of labels, for 2558 speaker contributions, allows for a more general account of what is typical for this task. We show pair-wise inter-labeler kappas for all 6 label tracks in Table 5. As can be seen, the average kappa for all labeler pairs involving labeler 1 is different from all other averages, suggesting that the behavior of labeler 1 is in fact atypical.

labeler	A	B	C	1	2	3
A		0.48	0.68	0.10	0.66	0.59
B	0.48		0.49	0.08	0.48	0.45
C	0.68	0.49		0.11	0.73	0.64
1	0.10	0.08	0.11		0.11	0.11
2	0.66	0.48	0.73	0.11		0.64
3	0.59	0.45	0.64	0.11	0.64	
mean	0.50	0.40	0.53	0.10	0.52	0.49
excl L1	0.60	0.46	0.61	—	0.60	0.58

Table 5: *Pairwise inter-labeler agreement kappas for 5 meetings in the ISL Meeting Corpus (2558 speaker contributions), emotional valence. The last two rows represent average kappas, both including and excluding labeler L1.*

Using the same pilot subcorpus of 5 meetings, we also performed a signal detection theory analysis (Green & Swets, 1966). We excluded labeler C and considered each labeler's sensitivity to the NEUTRAL vs NONNEUTRAL distinction in the data, comparing their assignment per speaker contribution to the majority vote. The results suggest that labeler 1 was far more conservative in pronouncing a speaker contribution as NONNEUTRAL than the other two labelers. When averaged over all 5 meetings, the sensitivity criterion for this labeler was 2.19, relative to 0.56 and 0.43 for labelers 2 and 3, respectively; a similar trend appears when comparing each labeler's sensitivity to the sensitivity mean for all five labelers. Additionally, the numbers for individual meetings suggest that labeler sensitivity to speaker valence varies across meeting genres; validation of this finding awaits future work.

6. Intra-Speaker State-to-Action Association

In order to validate the emotional relevance of the proposed behavior classes, we explored the association between behavior label assignments and the assignments of emotional valence. We show a sample crosstabulation analysis, between behavior as assigned by annotator 3 and valence as assigned by annotator 2, in Table 6. In addition to the absolute counts, we report the significance of deviation from the null hypothesis of no association, at both the $p < 0.01$ and $p < 0.001$ levels.

As the table shows, the informational behaviors which comprise the majority in this corpus have an association with Neutral valence which is significantly higher than that expected by chance, and their association with Positive valence is significantly lower. As anticipated, AttAmuse and LaughOnly exhibit the reverse trend. It is interesting to note that while there is 50% more of DiscntSli, Discnt, DisagrConf and Disagr when the four are taken together than there is of AttAmuse, only the latter is perceived by observers to be co-occurring with non-Neutral valence in a large majority of cases. Co-occurrence with Negative valence of the four behaviors expressing discontent or disagreement is significantly above chance, but all four co-occur with Neutral valence more than they do with Negative valence. This suggests that meeting participants may be suppressing their Negative valence more effectively than their Positive valence, or alternately that the vocal expression of Negative valence is more recipient-specific and not perceptible to outside observers (labelers).

Finally, we note that DiscntSli and Discnt exhibit significant above chance association with Positive in addition to that with Negative valence. We attribute this to "teasing" behaviors, in which participants display discontent towards each other mixed with, or covered by, humor, or in which they enjoy complaining. Crosstabulation analyses involving different pairings of labelers reveal a similar pattern (except those involving valence from labeler 1, whose valence assignments we disregard for reasons mentioned earlier).

	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

Table 6: *Co-occurrence of emotional valence as assigned by labeler 2 with the same speaker's emotionally relevant behavior as assigned by labeler 3 for the entire ISL Meeting Corpus (13221 speaker contributions), absolute counts. + and − represent rejection of the null hypothesis of no association, based on a* χ^2 *test. ++ and + identify counts which are significantly above that expected by chance; −− and − identify counts significantly below chance. Significance is at the* $p < 0.001$ *level for ++/−−, and at the* $p < 0.01$ *level for +/−.*

7. Inter-Speaker Action-to-State Association

In the previous section, we assessed the degree to which, from an external observer's point of view, speaker emotional valence correlates with the *same* speaker's concurrent behavior. In this section, we attempt to assess the degree to which a speaker's emotional valence correlates with his/her interlocutor's previous behavior. To do so, we needed to identify pragmatic adjacency (Levinson, 1983) for each speaker contribution; the ISL Meeting Corpus is not annotated with adjacency pairs. However, using a large subset of the ICSI Meeting Corpus (Janin et al., 2003) for which adjacency pair annotation does exist, (Galley et al., 2004) reported that selecting the most recent speaker yields a correct antecedent identification accuracy of 79.8%. Using this simple algorithm with minor extensions to resolve overlapping speaker contributions, we show the corresponding crosstabulation analysis in Table 7. Speaker contributions which had been split prior to annotation, and for which we had no segmentation, are excluded, resulting in a total of 11857 speaker contributions with identified antecedents. Note that it is possible for a given speaker contribution to be the antecedent of zero, one or more other speakers' contributions.

	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

Table 7: *Adjacency of emotional valence as assigned by labeler 2 with the antecedent speaker's emotionally relevant behavior as assigned by labeler 3 for the ISL Meeting Corpus (11857 speaker contributions), absolute counts. + and − represent rejection of the null hypothesis of no association; notation as in Table 6.*

Table 7 shows a pattern similar to that of Table 6, in that the antecedent speaker's Info, AgrAck and Agr show significant above chance co-occurrence with the current speaker's Neutral valence and below chance co-occurrence with the current speaker's Positive valence. Similarly, AttAmuse and LaughOnly exhibit the opposite association. In contrast to Table 6, this crosstabulation analysis reveals that the association between Negative valence and the antecedent speaker's DiscntSli or Discnt is not significantly different from chance. This suggests that complaining or criticizing behaviors, which are rare to begin with, may not lead to negative valence in other meeting participants. However, they appear to have the same significantly above chance association with their hearers' POSITIVE valence as with their speaker's. Finally, we note that AGRIMPREST appears to be effective. As in the previous section, patterns for crosstabulation analyses with different labeler pairs show similar results.

8. Conclusion

We have presented an annotation scheme for emotionally relevant behavior at the level of speaker contributions, in which the classes were originally constructed by manually clustering the open set labels produced by naive labelers. Inter-rater agreement for this scheme, presented in the form of a decision tree, is on par with similar work. In spite of an atypical labeler in the annotation of emotional valence, 99.4% of speaker contributions in our corpus exhibited at least a 2:1 label majority. In both schemes, which are complimentary, 20% of speaker contributions in the ISL Meeting Corpus are perceived by a labeler majority as not emotionally neutral.

When comparing assignments by different annotators, correlation of the two schemes shows that while there are more expressions of discontent or disagreement than attempts to amuse, the former are not predictive of perceived negative valence, whereas the latter are predictive of perceived positive valence in the speakers. Furthermore, certain speaker behaviors show significant above chance correlation with specific valence categories in subsequent speakers. Isolated laughter and attempts to amuse appear to be predictive of positive valence in both cases.

9. Acknowledgments

We would like to thank our labelers Matthew Bell, Julie Whitehead, Timothy Notari, Maria Kernecker, Joseph P. Friday, and Ryan Schultz.

This research was funded in part by European projects PF-STAR `http://pfstar.itc.it` (IST 2001 37599) and CHIL `http://chil.server.de` (IP 506909).

10. References

J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke. 2002. Prosody-based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog. In *Proceedings ICSLP*, Denver CO, USA.

R. Bates, P. Menning, E. Willingham, and C. Kuyper. 2005. Meeting Acts: A Labelling System for Group Interaction in Meetings. In *Proceedings Eurospeech*, Lisbon, Portugal.

A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. 2000. Desperately Seeking Emotions or: Actors, Wizards, and Human Beings. In *Proceedings ISCA Workshop on Speech and Emotion*, Belfast, N Ireland.

A. Batliner, C. Hacker, S. Steidl, E. Nöth, S. D'Arcy, M. Russell, and M. Wong. 2004. You Stupid Tin Box — Children Interacting with the AIBO Robot: A Cross-Linguistic Emotional Speech Corpus. In *Proceedings LREC*, Lisbon, Portugal.

S. Burger, V. MacLaren, and H. Yu. 2002. The ISL Meeting Corpus: The Impact of Meeting Type on Speech Style. In *Proceedings ICSLP*, Denver CO, USA.

J. Carletta. 1996. Assessing Agreement on Classification Tasks: the Kappa Statistic. *Computational Linguistics* 22(2):249–254.

J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson. 1997. The Reliability of a Dialogue Act Coding Scheme. *Computational Linguistics* 23(1):13–31.

R. Cauldwell. 2000. Where Did the Anger Go? The Role of Context in Interpreting Emotion in Speech. In *Proceedings ISCA Workshop on Speech and Emotion*, Belfast, N Ireland.

J. Cohen. 1960. A Coefficient of Agreement for Nominal Scales. In *Educational and Psychological Measurements*, 20:37–46.

M. Galley, K. McKeown, J. Hirschberg, and E. Shriberg. 2004. Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies. In *Proceedings 42nd Meetings of the ACL*, Barcelona, Spain.

D. M. Green, and J. A. Swets. 1966. *Signal Detection Theory and Psychophysics*. John Wiley and Sons, Inc.

A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbard, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters. 2003 The ICSI Meeting Corpus. In *Proceedings IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Hong Kong, China.

K. Laskowski and S. Burger. 2005. Annotation Scheme for Emotionally Relevant Behavior in Multiparty Conversation. Presented at *MLMI 2005*, Edinburgh, Scotland.

R. Lazarus. 1991. *Emotion and Adaptation*. Oxford University Press.

S. Levinson. 1983. *Pragmatics*. Cambridge University Press.

D. Litman, and K. Forbes-Riley. 2004. Annotating Student Emotional States in Spoken Tutoring Dialogues. In *Proceedings 5th SIGdial Workshop on Discourse and Dialogue*, Boston MA, USA.

D. Litman, and K. Forbes-Riley. 2004. Predicting Student Emotions in Computer-Human Tutoring Dialogues. In *Proceedings 42nd Meetings of the ACL*, Barcelona, Spain.

S. Renals, and D. Ellis. 2003. Audio Information Access from Meeting Rooms. In *Proceedings Intl. Conf. Acoustics, Speech and Signal Processing*, Hong Kong, China.

J. Russell, J.-A. Bachorowski, and J.-M. Fernandez-Dols. 2003. Facial and Vocal Expressions of Emotion. In *Annual Review of Psychology*, 54:329-349.

E. Shriberg, R. Dhillon, J. Bhagat, J. Ang, and H. Carvey. 2004. The ICSI Meeting Recorder Dialog Act (MRDA) Corpus. In *Proceedings 5th SIGdial Workshop on Discourse and Dialogue*, Boston MA, USA.

S. Steininger, F. Schiel, O. Dioubina, and S. Rabold. 2002. Development of User-State Conventions for the Multimodal Corpus in SmartKom. In *Proceedings Workshop on Multimodal Resources and Multimodal Systems Evaluation*, Las Palmas, Gran Canaria.

B. Wrede, and E. Shriberg. 2003. Spotting "Hot Spots" in Meetings: Human Judgements and Prosodic Cues. In *Proceedings Eurospeech*, Geneva, Switzerland.