A comparison of addressee detection methods for multiparty conversations

Rieks op den Akker
Human Media Interaction
University of Twente
Enschede, the Netherlands
infrieks@cs.utwente.nl

David Traum
Institute for Creative Technologies
University of Southern California
Marina Del Rey, CA 90292 USA
traum@ict.usc.edu

Abstract

Several algorithms have recently been proposed for recognizing addressees in a group conversational setting. These algorithms can rely on a variety of factors including previous conversational roles, gaze, and type of dialogue act. Both statistical supervised machine learning algorithms as well as rule based methods have been developed. In this paper, we compare several algorithms developed for several different genres of multiparty dialogue, and propose a new synthesis algorithm that matches the performance of machine learning algorithms while maintaining the transparency of semantically meaningful rule-based algorithms.

1 Introduction

Detecting who is being addressed, i.e. who the speaker is talking to, is non-trivial in multi-party conversations. How speakers make clear who they address depends on the conversational situation, knowledge about other participants, inter-personal relations, and the available communication channels.

In this paper we present rule based methods for automatic addressee classification in four-participant face-to-face meetings. A rule based method is more transparent than the statistical classifiers. It synthesizes empirical findings of addressing behavior in face-to-face conversations. We have analysed addressing behavior in small design group meetings, and we have evaluated our methods using the multi-layered multi-modal annotated AMI meeting corpus (Carletta, 2007). The same multi-modal corpus has been used for developing statistical addressee classifiers using (Dynamic) Bayesian Networks (Jovanovic, 2007). The (Dynamic) Bayesian Network classifiers have performances ranging from 68-77%, depending on the types of features used and whether it is a static network, using Gold Standard (i.e. the manual annotated) values for addressees of previous acts, or dynamic, using own predicted values for addressees of previous acts in the dialogue. Our best performing rule-based method has an accuracy of 65%, which is 11% over the baseline (always predict that the group is addressed).

Performance measures don't tell much about the confidence we can have in the outcome in particular cases. A reliability analysis of the manually annotated data that is used for training and testing the machine classifier can reveal in what cases the outcomes are less reliable. In specific situations, such as when the speaker uses *"you"*, or when the speaker performs an initiating act, supported by visual attention directed to the addressed partner, the method outperforms the statistical methods. Our method uses speaker's gaze behavior (focus of attention), dialogue history, usage of address terms as well as information about the type of dialogue act performed by the speaker to predict who is being addressed.

2 How do speakers address others?

Addressing occurs in a variety of flavors, more or less explicitly, verbally or non-verbally. Thus, sometimes deciding whether or not the speaker addresses some individual partner in particular is far from a trivial exercise. Within a single turn, speakers can perform different dialogue acts (i.e. they can express different intentions), and these dialogue acts can be addressed to different participants. In small group discussions, like those in the AMI meetings with 4 participants, most contributions are addressed to the whole group. But sometimes speakers direct themselves to one listener in particular. Some important motivations for individual addressing are that the group mem-

bers bring in different expert knowledge and that they have different tasks in the design process. If someone says to a previous speaker *"can you clarify what you just said about ..."* it is clearly addressed to that previous speaker. This doesn't rule out that a non-addressed participant takes the next turn. But generally this will not happen in an unmarked way.

The basis of our concept of addressing originates from Goffman (Goffman, 1981). The addressee is the participant *"oriented to by the speaker in a manner to suggest that his words are particularly for them, and that some answer is therefore anticipated from them, more so than from the other ratified participants"*. Thus, according to Goffman, the addressee is the listener the speaker has selected because he expects a response from that listener. The addressee coincides with the one the speaker has selected to take the next turn. But addressing an individual does not always imply turn-giving, such as can be seen in (1), a fragment of Alice's speech, in a conversation between Alice, Ben and Clara.

(1) Yes, but, as Clara already said earlier
gaze: < Ben >

correct me if I'm wrong,
gaze: < Clara >

the price of working out *your* proposal is too high for us, so ...
gaze: < Ben >

In (1), the main dialogue act performed by Alice is addressed to Ben. Although Alice's contribution is to the whole group, it is meant especially as a reaction to the preceding proposal made by Ben, and she directs herself to Ben more than to the others. That is why we say that in this case *the dialogue act is addressed to* Ben. Note that *"your"* refers to Ben as well, and also Alice's gaze is directed at Ben. Alice is especially interested to see how Ben picks up and validates the concern that she expresses. The dialogue act expressed by the embedded phrase is addressed to Clara. Although, Alice explicitly invites Clara to correct her, which is indicated by the gaze shift during this clause, after mentioning her name, she doesn't yield the turn, but continues speaking.[1]

[1]The rules for dialogue act segmentation used in the AMI corpus do not cover dialogue act units embedded in other units, as is the case in this made up example.

Speakers use different procedures to make clear who they address. The selection of this procedure depends on (a) what the speaker believes of the attentiveness of the listener(s) to his talk, and (b) the speaker's expectation about the effect his speech has on the listener that he intends to address. For example if A just was just asked a question by B then A will assume that B is attending his answer. In a face-to-face meeting A will usually monitor how B takes up his answer and will now and then gaze at B as his visual focus of attention is not required for competing foci of interest. Lerner distinguished *explicit* addressing and *tacit* addressing. To characterize the latter he writes: *"When the requirements for responding to a sequence-initiating action limit eligible responders to a single participant, then that participant has been tacitly selected as next speaker. Tacit addressing is dependent on the situation and content."* (Lerner, 2003).

An example from our corpus is when a presenter says *"Next slide please"* during his presentation, a request that is clearly addressed to the one who operates the laptop. Tacit addressing is most difficult for a machine, since it requires to keep track of the parallel activities that participants are engaged in.

Explicit addressing is performed by the use of vocatives (*"John, what do you think?"*) or, when the addressee's attention need not be called, by a *deictic personal pronoun*: *"What do you think?"*. There is one form of address that always has the property of indicating addressing, but that does not itself uniquely specify *who* is being addressed: the *recipient reference term* "you" (Lerner, 2003). *The use of "you" as a form of person reference separates the action of "addressing a recipient" from the designation of just who is being addressed. In interactional terms, then, "you" might be termed a recipient indicator, but not a recipient designator. As such, it might be thought of as an incomplete form of address* (Lerner, 2003). Gaze or pointing gestures should complete this form of addressing. These analytical findings motivated the selection of rules for addressee detection.

3 Automatic Addressee Recognition

The starting point of our design of a rule based algorithm for addressee prediction was Traum's algorithm as presented in (Traum, 2004), shown in (2). This algorithm was meant to be used by virtual agents participating in a multi-party, multi-

conversation environment (Traum and Rickel, 2002; Rickel et al., 2002), in which conversations could be fluid in terms of starting and stopping point and the participants that are included. The algorithm only uses information from the previous and the current utterance; thus no information about uptake of the act performed by the current speaker. The method doesn't use speaker gaze. In initial versions of the virtual world, the agents did not have access to human gaze. Even when gaze is available, it is non-trivial to use it for addressee-prediction, because there are many other gaze targets in this dynamic world other than the addressee, including monitoring for expected events in the world and objects of discussion (Kim et al., 2005; Lee et al., 2007).

(2) 1 If utterance specifies a specific addressee (e.g. a vocative or utterance of just a name when not expecting a short answer or clarification of type person) then Addressee = *specified addressee*.

2 else if speaker of current utterance is the same as the speaker of the immediately previous utterance then *Addressee = previous addressee*

3 else if previous speaker is different from current speaker then *Addressee = previous speaker*

4 else if unique other conversational participant (i.e. a 2-party conversation) then *Addressee = that other participant*

5 else *Addressee = unknown*

Traum's algorithm had good performance in the Mission Rehearsal Exercise domain. (Traum et al., 2004) reports F-scores of from 65% to 100% in actual dialogues, using noisy speech recognition and NLU as input). In this paper we will examine to what degree this algorithm generalizes to a different sort of multi-party corpus, and what can be done to improve it.

4 The AMI meeting corpus

The manually annotated conversations that we analysed are from the AMI meeting corpus; (Carletta, 2007). There are 14 four-participant face-to-face meetings, where participants are mostly sitting at a rectangular table. Twelve of the 14 meetings were recorded in one meeting room, the other two in two other rooms.

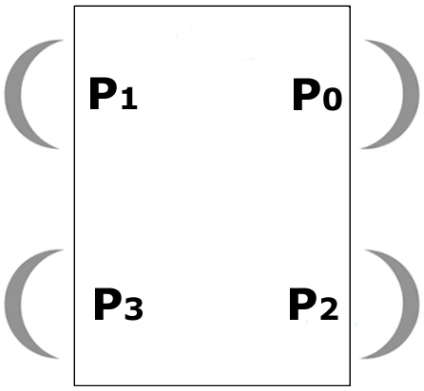

Figure 1: Fixed seating positions around a square table.

The 14 meetings were annotated with dialogue acts, addressee information as well as focus of attention of participants (FOA). Utterances are segmented in consecutive DA-segments. The segments are assigned a type. Dialogue acts types are: *Inform*, *Elicit-inform*, *Suggest*, *Offer*, *Elicit-offer-or-suggestion*, *Assess*, *Elicit-assessment*, *Comment-about-understanding*, *Elicit-comment-about-understanding*, *Be-positive*, and *Be-negative*. Other labels for DA-segments are *Backchannel*, *Stall*, and *Fragment*. The *Other* label was used when an utterance could not be labeled by one of the list of dialogue labels.

For important contentful dialogue acts (i.e. excluding Stall, Fragment and Backchannel[2] acts) the annotators have indicated whether the DA was addressed to the whole group (*G-addressed*), or to some individual (*I-addressed*), in which case they indicated who was being addressed. Annotators could also label the addressee as Unknown, but because there was very little reliability in this category, we combined it with the G-addressed category.

I-addressed acts are marked in terms of table position of the person being addressed: P0, P1, P2 or P3. Figure 1 shows the layout of the fixed seating positions in the meeting rooms.

Focus of attention (FOA) can be on one of these participants or at the white board, at the table, or at no specific target. Words and dialogue acts were time aligned, so that it can be computed what the

[2]Backchannel acts were assumed to be addressed to the "previous speaker" and were therefore not annotated in the AMI corpus.

focus of attention is of each of the participants during a specific time frame. Note that neither the addressee annotation, nor the FOA annotation allowes a multiple target label. This could be a possible cause of confusion between annotators in case a sub-group is addressed by the speaker. However, subgroup addressing hardly occurs in the data.

4.1 Reliability of the AMI annotations.

Since we based our models on the analysis of a human annotated corpus, and since we also tested them on manual annotated data, the question arises how much human annotators agree on the addressee labeling. How does the accuracy depend on the annotator? Are there specific situations in which results are more reliable than in others? (Jovanovic, 2007) (Chapter 3.4) contains a detailed examination of the inter-annotator agreement of the codings of the AMI corpus. We present some highlights here.

We compare three annotations of one and the same meeting in our corpus. Most confusions in the addressing labeling are between *I-addressed* and *G-addressed*, If annotators agree that the DA is I-addressed then they agree on the individual as well. We found that for both dialogue acts and addressee identification, reliability is higher for some decisions than others. Table 1 shows Krippendorff's alpha values (Krippendorff, 2004) for inter-annotator agreement for each pair of annotators. The statistics are computed on the subsets of pairwise agreed DA-segments: cases in which the annotators did not agree on the segmentation are left out of this analysis.[3]

Table 1 shows that annotators consistently agree more on the addressing of elicit acts (3rd column) than on DAs in general (2nd column). For the subset of elicit acts, when annotators agree that an elicit is *I-addressed* (which happens in 50-80% of the agreed elicit acts, depending on the annotators), than they agree on the individual that is addressed, without exception. Addressing is a complex phenomenon and we believe that the mediocre agreement between addressee annotations is due to this complexity. In particular, we observed that some annotators prefer to see a response act as I-addressed at the speaker of the initiating act, where for others the content is more decisive (does, for example, the question address an issue that is relevant for the whole group or does it only concern the speaker and his addressee?)

[3] A better analysis of addressing (dis)agreements might be based on speaker turns or sequences of dialogue acts, because (a) many segmentation disagreements do not affect addressing, and (b) the distribution of DA types over the set of agreed segments is different from the distribution of DA types over the whole corpus (agreed segments are shorter in the mean)

pair	adr	adr-eli	da	da-eli
a-b	0.56(412)	0.67(31)	0.62(756)	0.69
a-c	0.45(344)	0.58(32)	0.58(735)	0.64
b-c	0.46(430)	0.62(53)	0.55(795)	0.80

Table 1: Alpha values (and numbers of agreed DA segments) for the three pairs of annotators; for addressing, addressing of elicit acts only, dialog acts (all 15 DA classes), and elicit vs non-elicit acts (5th column).

As expected (because speakers FOA is an important indicator for addressing) annotators agree more on the addressee in situations with a clear speaker gaze at one person. We refer to (Reidsma et al., 2008) for more details.

Annotators agreed rather well in telling elicit acts from other types of dialogue acts, as is shown in Table 1, 5th column. This DA type information is thus quite reliable.

Focus of attention annotation was done with high agreement, so we can take gaze target information as reliable information, with a timing precision of about 0.5 sec. (See (Jovanovic, 2007) for a detailed reliability analysis of the FoA annotation.)

5 Dialog structure

Gupta et al. present experiments into the resolution of *"you"* in multi-party dialog, and they used the same part of the scenario based AMI meetings as we did. They distinguish between generic and referential uses of "you"; and, the referential uses, they try to classify automatically by identifying the referred-to addressee(s): either one of the participants, or the group. All results are achieved without the use of visual information. (Gupta et al., 2007). Gupta et. al. expected that *the structure of the dialog gives the most indicative cues to addressee: forward-looking dialog acts are likely to influence the addressee to speak next, while backward-looking acts might address a recent speaker*. In a similar way Galley et al. (Galley et al., 2004) also used the dialog structure present in adjacency pairs as indicative for ad-

dressees: the speaker of the a-part would likely be the addressee of the b-part and the addressee of the a-part would likely be the speaker of the b-part (dyadic pattern $ABBA$). In the one dimensional DA schema that we used on the AMI corpus there is no clear distinction between Backward Looking (BL) and Forward Looking (FL) "types" of dialogue acts. However, we may consider the *elicit types* as FL types of DAs. Typical BL DA types are *Comment about Understanding* and to a lesser extend *Assessments*. The other DA types can be assigned to BL as well as to FL utterances, but if an *Inform* act follows an *Elicit-Inform*, the last one more likely has a BL function. The AMI corpus is also annotated with dialog relation pairs, much like the classical adjacency pairs: they are typed relations (the type carries polarity information: is the response of the speaker positive or negative, or partial negative/positive the target act, or does the speaker express uncertainty), and related DAs need not be adjacent (i.e. there can be other DAs in between). In the AMI corpus the speaker addressee pattern $ABBA$ fits 60% of all adjacency pairs, which makes them a good feature for addressee prediction. We will however not use this adjacency pair information because this information is as hard to obtain automatically as addressee information.

The total number of DAs in our corpus is 9987, of which 6590 are contentful DAs (i.e. excluding *Stall*, *Fragment*, and *Backchannel*, which did not get an addressee label assigned). Of these, 2743 are addressed to some individual (*I-addressed*); the others are addressed to the Group (*G-addressed*).

In 1739 (i.e. 63%) cases of the 2743 *I-addressed* dialog acts, the addressed person is the next speaker (the current speaker might also perform additional dialogue acts before the next speaker's speech).

Forward looking DAs that are I-addressed are more selective for next speaker than I-addressed DAs in general. There are 652 elicit acts in our corpus. Of these, 387 are $I - addressed$. In 302 cases (78%) the addressee is the next speaker. This is indeed substantially more than the mean (63%) over all DA types.

Speaker's gaze is an important indication for whom they address their DA. (see (Kendon, 1967), (Kalma, 1992), (Vertegaal and Ding, 2002)). In our corpus, speakers gaze three times more at their addressee than at other listeners.

6 Algorithms for Addressee Identification in the AMI corpus

In this section, we compare several different algorithms for recognizing the addressee in the AMI corpus.

6.1 Jovanovich's DBN

In (Jovanovic, 2007), Dynamic Bayesian Networks, (D)BNs, were used to classify the addressee based on a number of features, including context (preceding addressee and dialogue acts, related dialogue acts), utterance features (personal pronouns, possessives, indefinite pronouns and proper names), gaze features, and the types of meeting actions, as well as topic and role information. The best performance for all features yielded roughly 77% accuracy on the AMI corpus. The best performing BNs uses "Gold Standard" values of addressees of previous and related DAs. The DBNs uses own predicted addressee values for these features. For comparison purposes, we recoded this approach using the Weka toolkit's implementation of BayesNets, using the same features as our other algorithms had available: no adjacency pair information, no topic role and role information. The BNs achieved accuracies of 62% and 67%.

6.2 Traum's algorithm

We re-implemented Traum's algorithm shown above in (2). While Traum's algorithm had good performance in the Mission Rehearsal Exercise domain it has very bad performance in the AMI domain, as shown in the next section. Why is this? Interaction styles are different across the two domains. Patterns of speaker turns are different and that is caused by the different scenarios. In the meeting scenario, there is a much more static environment, so gaze is a better predictor, which was not used in Traum's algorithm. More importantly, Traum's algorithm does not adequately account for speech addressed to a group rather than (primarily) to a single participant, while this formed the majority of the AMI data. Traum's algorithm indicates group, only if a group addressing term (e.g. "you all") is used , or the group is the previous addressee, or if the addressee is unknown. There are also more frequent uses of address terms in the Mission Rehearsal context than in the AMI

meetings.

6.3 GazeAddress

The method *gazeAddress* predicts the addressee of a DA using only information about speaker's cumulative focus of attention over the time period of the utterance of the speech act. It predicts the addressee as follows. If there is an individual B such that the speaker A gazes for more than 80% of the duration of his dialogue act in the direction of B, it is assumed that the dialogue act performed by A is I-addressed to B. Otherwise, the speaker is assumed to address the group (G). To obtain the best threshold value, we ran several tests with different values for the threshold and computed recall and precision for the Group class as well as for the individual class values. Going up from 50% to 80%, the precision and recall of the single addressee and group addressee identification slowly improves. After that the precision of the single addressee does not improve nor decline much. But the recall and precision of the group identification gets a lot worse. We used 80% as threshold value in subsequent experiments.

6.4 The Addressee Prediction Algorithm

Our Addressee Prediction Algorithm (APA) that returns the addressee of the current dialogue act (DA) runs as follows. It returns "G" when it predicts that DA is *G-addressed*. If it predicts that the DA is *I-addressed* it returns the table position of the individual participant.

```
(1) (address term used)
if (containsAddressTerm(DA)){
        return referredPerson;}

(2) (same speaker turn)
if (daSpeaker=prevDASpeaker) {
    if (gazeAddress=previousADR ){
        return previousADR;
    } else{
        return "G";}}

(3) (other speaker)
    if (daSpeaker=previousADR)
        return prevDASpeaker;
    if (gazeAddress!=null && you)
        return foa;
    if(gazeAddress=prevDASpeaker){
        return prevDASpeaker;}}
```

In (1) it is tested whether the speaker uses an address term (name or role name of a participant). If so, the referred person is returned as the addressee. Clause (2) fires when the current DA is by the same speaker as the previous one. If the *gazeAddress* method would return for an individual (the value of foa) and this is the same one as the person addressed in the previous act then this one is returned. Clause (3) fires when a speaker change occurred. If the previous speaker addressed the current speaker, then the previous speaker is the returned addressee. If not, when the DA contains "you" and the *gazeAddress* method returns some individual then this one is returned. If *gazeAddress* decided for an individual and this equals the previous speaker then this one is returned. Otherwise, the group is addressed. We experimented with some variations of this method. A slight improvement was obtained when we have a special treatment for forward looking DA types. Analyses of the corpus reveals that elicit acts are more frequently used as forward looking acts. In that case, the decision is based on *gazeAddress* not taking into account the previous speaker.

7 Results

Table 2 shows the performance of four methods from the previous section in terms of Recall, Precision, and F-score for group, participant P0 (the most challenging of the participants), and overall accuracy (i.e. percentage correct).

	Group			P_0			
Method	R	P	F	R	P	F	Acc
Traum's	12	92	22	70	31	44	36
BayesNet	65	73	69	62	45	52	62
GazeAdr	66	65	65	36	43	40	57
APA	89	65	75	26	62	36	65

Table 2: Performance table of the four methods for addressee prediction. N=6590 (DAs). Baseline (always Group) is 54%.

We can see from table 2 that APA has the highest overall accuracy for recognizing the addressees of each dialogue act in sequence. However it is the lowest of the four in recognizing P0. In table 3 we look at the importance of recognizing the previous addressee correctly, by supplying the Gold Standard value for this feature rather than the value calculated by the respective algorithms. Traum's algorithm shows the biggest improvement in this case, while APA improves the least.

Table 4 gives an overview of the performances of the two new methods - *gazeAddress* and APA - on various subclasses of the data set. ALL is the set of all contentful dialogue acts; ELI is the set of elicit acts; YOU is the set of acts containing

	Group			P_0			
Method	R	P	F	R	P	F	Acc
Traum	47	88	61	67	42	52	56
BayesNet	66	85	75	73	50	60	67
APA	86	68	76	34	61	44	67

Table 3: Performance table when using *Gold Standard* values for previous addressees of the three methods making use of previous addressee information.

	D A - S E T S			
	ALL	ELI	YOU	ELI-Y
N	6590	652	1061	166
Gaze	57	62	62	68
APA	65	62	68	69

Table 4: Accuracy values of methods on various sets of dialogue acts

"you"; ELI-YOU is the subset of eliciting acts that contains *"you"*.

We see that for the subsets of dialogue acts that contains *"you"* as well as for the mostly forward looking elicit acts APA performs better than the mean performance of APA over all DAs, and even better than the Dynamic Bayesian Networks. The average accuracy of APA for DAs with *"you"* over all the meetings is 68%.

The results vary over the set of meetings and a factor that causes this is the percentage of G-addressed DAs in the meeting. In general, the performance raises with the percentage of G-addressed DAs.

How does the performance depend on the annotators? For the one meeting IS1003d that was annotated by all three annotators involved, the accuracies of method APA were 61, 75 and 60. For the method *gazeAddress* they were 58, 66, 57, respectively. Also here the data annotated by the annotator who had a preference for the G-label over one of the individual labels has a higher accuracy.

7.1 Further research

A more detailed analyses of the results of method *gazeAddress* reveals that the recall and precision values depend on the position of the speaker as well as on the *relative position* of the person gazed at most by the speaker. In future work, we will examine both the role of the meeting participant and the physical locations in terms of their effect on performance and possibly augmentations to the algorithms. Using the same part of the AMI corpus, (Frampton et al., 2009) classify referential uses of "you" in terms of relative position of addressees from the view point of the speaker. They achieve good results in finding the I-addressee of those speech acts that contain such a referential use of "you". Note that our method does not identify if an occurrence of "you" is referential, so it is hard to compare the results.

8 Conclusion

We have seen that a rule based method can predict addressing with an accuracy that is comparable with that of the purely statistical methods using dynamic Bayesian networks. It is hard to obtain a high precision and recall for individual addressing. Although slight improvements can be expected if we take into account the relative positions of speakers and addressees when using gaze direction of speakers as indicator for who is being addressed, substantial improvements will likely be only possible when the system has more knowledge about what is going on in the meeting.

Knott and Vlugter implemented in their multi-agent language learning system a rule-based method for addressee detection which is similar to the one of Traum, see (Knott and Vlugter, 2008). In their system, agents make frequent use of address terms, and they do sub-group addressing, unlike the agents in the face-to-face meetings. Sub-group addressing remains a challenging issue for multi-agent dialogue systems.

Comparative analysis of various human annotations of the same data is very informative for clarifying such abstract and complex notions as addressing is. Such an analysis is important to improve our understanding of the phenomena and to sharpen the conceptual definitions that we use. Results inferred from statistics and patterns in relations between annotated data should take the difficulties that annotators have in applying the general notions in concrete new situations into account.

Acknowledgments

The authors are gratefull to the anonymous reviewers for their comments and suggestions. Traum's effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements

and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred. The work by R. op den Akker is supported by the EU IST Programme Project FP6-0033812 (AMIDA). This paper only reflects the authors views and funding agencies are not liable for any use that may be made of the information contained herein.

References

Jean C. Carletta. 2007. Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. *Language Resources and Evaluation*, 41(2):181–190, May.

Matthew Frampton, Raquel Fernández, Patrick Ehlen, Mario Christoudias, Trevor Darrell, and Stanley Peters. 2009. Who is "you"? combining linguistic and gaze features to resolve second-person references in dialogue. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 273–281, Athens, Greece, March. Association for Computational Linguistics.

Michel Galley, Kathleen McKeown, Julia Hirschberg, and Elizabeth Shriberg. 2004. Identifying agreement and disagreement in conversational speech: Use of Bayesian networks to model pragmatic dependencies. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, Barcelona, Spain.

Erving Goffman. 1981. Footing. In *Forms of Talk*, pages 124–159. Philadelphia: University of Pennsylvania Press.

Surabhi Gupta, John Niekrasz, Matthew Purver, and Daniel Jurafsky. 2007. Resolving "you" in multiparty dialog. In *Proceedings of 8th SigDial Workshop*, pages 227–230.

N. Jovanovic. 2007. *To whom it may concern. Addressee identification in face-to-face meetings*. Ph.D. thesis, University of Twente, Enschede, The Netherlands, March.

A. Kalma. 1992. Gazing in triads: A powerful signal in floor apportionment. *British Journal of Social Psychology*, 31(1):21–39.

A. Kendon. 1967. Some functions of gaze direction in social interaction. *Acta Psychologica*, 26:22–63.

Youngjun Kim, Randall W. Hill, and David R. Traum. 2005. Controlling the focus of perceptual attention in embodied conversational agents. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1097–1098, New York, NY, USA. ACM.

Alistair Knott and Peter Vlugter. 2008. Multi-agent human-machine dialogue: issues in dialogue management and referring expression semantics. *Artif. Intell.*, 172(2-3):69–102.

K. Krippendorff. 2004. *Content analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage, 2nd edition.

Jina Lee, Stacy Marsella, David R. Traum, Jonathan Gratch, and Brent Lance. 2007. The rickel gaze model: A window on the mind of a virtual human. In Catherine Pelachaud, Jean-Claude Martin, Elisabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé, editors, *IVA*, volume 4722 of *Lecture Notes in Computer Science*, pages 296–303. Springer.

Gene H. Lerner. 2003. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society*, 32:177–201.

D. Reidsma, D. K. J. Heylen, and H. J. A. op den Akker. 2008. On the contextual analysis of agreement scores. In J-C. Martin, P. Paggio, M. Kipp, and D. K. J. Heylen, editors, *Proceedings of the LREC Workshop on Multimodal Corpora, Marrakech, Morrocco*, pages 52–55, Paris, France, May. ELRA.

J. Rickel, S. Marsella, J. Gratch, R. Hill, D. Traum, and W. Swartout. 2002. Towards a new generation of virtual humans for interactive experiences. *Intelligent Systems*, 17:32–36.

David R. Traum and Jeff Rickel. 2002. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first International Joint conference on Autonomous Agents and Multiagent systems*, pages 766–773.

David R. Traum, Susan Robinson, and Jens Stephan. 2004. Evaluation of multi-party virtual reality dialogue interaction. In *Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1699–1702.

D. Traum. 2004. Issues in multi-party dialogues. In F. Dignum, editor, *Advances in Agent Communication*, pages 201–211. Springer-Verlag.

Roel Vertegaal and Yaping Ding. 2002. Explaining effects of eye gaze on mediated group conversations:: amount or synchronization? In *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 41–48, New York, NY, USA. ACM.