

Detecting Action Items in Multi-Party Meetings: Annotation and Initial Experiments

Matthew Purver, Patrick Ehlen, and John Niekrasz

Center for the Study of Language and Information
Stanford University
Stanford, CA 94305, USA
{mpurver,ehlen,niekrasz}@stanford.edu

Abstract. This paper presents the results of initial investigation and experiments into automatic *action item detection* from transcripts of multi-party human-human meetings. We start from the flat action item annotations of [1], and show that automatic classification performance is limited. We then describe a new hierarchical annotation schema based on the roles utterances play in the action item assignment process, and propose a corresponding approach to automatic detection that promises improved classification accuracy while also enabling the extraction of useful information for summarization and reporting.

1 Introduction

A great deal of everyday human-human interaction takes place in meetings, and their content can be important: information is exchanged, plans and decisions are made, new tasks assigned, deadlines changed and so on. There is consequently a great deal of interest in the automatic processing, understanding, summarization and reporting of meetings. While there may be many useful outputs that could be reported, including automatically produced transcripts or notes, user studies [2, 3] suggest that amongst the most important are records of the *decisions* made and the associated *action items* assigned. This paper concentrates on the automatic detection of *action items*, specific kinds of decisions common in meetings and characterized by the concrete assignment of tasks together with certain properties such as an associated timeframe and responsible party. Our aims are firstly to detect the regions of discourse which establish action items, so that their surface form can be used as the basis of a targeted report or summary; and secondly, to identify the important properties of the action items themselves (such as the associated tasks and deadlines), so that we can work towards more concise and informative semantically-based reporting (for example, adding task specifications to a user’s calendar or to-do list). Our claim in this paper is that both of these aims are facilitated by an approach which takes into account the roles which different utterances play in the decision-making process – in short, a shallow notion of discourse structure.

We first discuss an existing set of action item annotations [1] which take a simple approach, tagging the relevant utterances as action-item-related, but not

distinguishing them further. We show that while these annotations allow some useful automatic detection performance even with simple classification methods, the accuracy is limited. We then describe a richer, hierarchical approach to action item annotation, in which utterances are not only tagged as being related to a particular action item, but classified according to the role they play in the process of establishing and agreeing on that action item. Our newer approach assigns relevant utterances to one or more of a small set of decision-making acts. Initial experiments indicate that significant performance improvements can be gained by using this deeper information, detecting the individual acts independently, and then detecting action items via the presence of multiple act types. As well as improving performance, this method allows important information to be extracted to enable more detailed understanding and reporting of the detected action items.

2 Background

Action Items. In institutions where group projects and collaborative problem-solving are an important element of the institution's purpose, meetings are a common (perhaps too common [4]) occurrence where important steps are taken toward achieving both individual and institutional goals. The type of communication that takes place in these meetings can include briefings, brainstorming, problem-solving, and planning, just to name a few. For the great majority of these, the ultimate goals are to share information and make group decisions.

Action items are a common form of these group decisions that make an interesting subject for investigation for a number of reasons. First, they often embody the transfer of group responsibility to that of the individual, an extremely important component of the institutional problem-solving process. Also, a common (although by no means universal) practice in meetings is to reiterate or review them during a specific period, when individuals will summarize their commitments to others to make explicit the tasks which have been assigned to them. In addition, because they are group decisions which result in individual commitment, the committed owner has a great responsibility to the others to be sure that the commitment is properly represented, recorded, remembered, and fulfilled.

The person committing to be responsible for the action item need not be the person who actually performs the action (they might, say, delegate the task to a subordinate), but publicly commits to seeing that the action is carried out; we call this person the *owner* of the action item. Because this action is a social action that is coordinated by more than one person, its initiation is reinforced by *agreement* and uptake among the owner and other participants that the action should and will be done. And to distinguish this action from immediate actions that occur during the meeting and from more vague future actions that are still in the planning stage, an action item will be specified as expected to be carried out within a *timeframe* that begins at some point after the meeting and extends no further than the not-too-distant future. So an action item, as a type of social

action, often comprises four components: a *task description*, a *timeframe*, an *owner*, and a round of *agreement* among the owner and others.

Automatic Detection. While meeting minute-takers and managers will often summarize assigned action items, and a plethora of assistive technologies are available to facilitate that, we are unaware of any previous attempt to do this automatically for spoken discourse. There is precedent in text processing: [5] attempted to identify action items in e-mails, using classifiers trained on annotations of individual sentences within each e-mail. Sentences were annotated with one of a set of “dialogue” act classes; one class **Task** corresponded to any sentence containing items that seemed appropriate to add to an ongoing to-do list. They report good inter-annotator agreement over their general tagging exercise ($\kappa > 0.8$), although individual figures for the **Task** class are not given. They then concentrated on those sentences classed as **Task**, establishing a set of predictive features (in which word n-grams emerged as “highly predictive”) and achieved reasonable per-sentence classification performance (with f-scores around 0.6).

However, for multi-party spoken dialogue (as occurs in meetings), the closest related work is probably in the area of dialogue act detection, where the dialogue act taxonomy chosen often includes decision-related classes. Many such annotation schema are ultimately based on the DAMSL annotation scheme [6] which includes tags for utterances like **Action-Directive** (a sub-class of **Influencing-Addressee-Future-Action**) and **Commit** (a sub-class of **Committing-Speaker-Future-Action**). But reliability (in terms of inter-annotator agreement) on the latter category has been found to be low ([7] report $\kappa = 0.15$, partly due to the ambiguity in distinguishing **acknowledgements** from **acceptances** – e.g., does “okay” mean “I understand” or “I’ll do that”?). The ICSI MRDA annotation schema [8] incorporates a **commit** dialogue act type, but finds them in only 0.24% of utterances in meetings. And to date, most attempts to automatically tag MRDA-based dialog acts concentrate on five general high-level dialogue act classes [9, 10], rather than tagging at a level low enough to distinguish **commit** acts from other statements. More importantly for current purposes, though, these commitment acts do not in any case capture the distinction between action items and more general commitments (i.e. commitments to general courses of action or approaches, as well as to specific concrete tasks).

From these studies, it may be productive to surmise that finding action items in meetings involves identifying an interactive process that dialog acts by themselves do not capture. After all, a dialog act corresponds to the illocutionary force of one utterance made by one person. The process of establishing an action item in a meeting, however, is better represented as a type of group action – or *social action* [11] – that is often coordinated by multiple participants over multiple utterances using multiple sign-systems or modalities. That coordinated action entails a public commitment by a specific person or group to be responsible for a specific action to be carried out within a specific timeframe; and that commitment is made in the presence of, and is acknowledged by, others. As we discuss below, this means that detection is best carried out by trying to de-

tect such group actions – here, by looking for multiple complementary utterance types, but potentially also by including information from other modalities.

3 Baseline Experiments

In [1], an initial annotation of action item subdialogues (and topic segmentations) was performed on 65 meetings from the ICSI and ISL meeting corpora [12, 13]. In this exercise, action items were defined as tasks that would be entered on a to-do list, and identified simply as sets of utterances with a brief textual description. The two annotators identified a total of 921 and 1267 utterances respectively as belonging to action items, and inter-annotator agreement was rather low ($\kappa = .36$, where κ is the kappa statistic as formulated in [14]).¹ This approach is therefore roughly parallel to that of [5] to email classification, although note that the inter-annotator agreement seems much lower on our discourse data. We therefore performed a similar experiment to examine automatic classification performance. Like [5] we used support vector machines [15] via the classifier *SVMlight* [16]; their full set of features are not available to us as many are text- or email-specific, but we experimented with combinations of words and n-grams. Performance, however, was poor, with precision, recall and f-score all below 0.25 (perhaps unsurprisingly, given the low human inter-annotator agreement).

Partly to examine the effect on performance of using a smaller, more homogeneous dataset, and partly in order to compare with our later results (see below), we applied this simple flat annotation schema to a separate sequence of 5 short related meetings produced as part of the CALO project. These meetings were simulated according to a given general scenario, but were not scripted. In order to avoid entirely data- or scenario-specific results (and also to provide an acceptable amount of training data), we then added a random selection of 6 ICSI meetings and 1 ISL meeting from [1]’s annotations. We assessed classification performance via a 5-fold validation on each of the CALO meetings; in each case, we trained classifiers on the other 4 meetings in the CALO sequence, plus the fixed ICSI/ISL training selection. Performance is shown in Table 1; these figures were obtained using *SVMlight* with words (unigrams, after text normalization and stemming) as features – we also investigated other discriminative classifier methods, and the use of 2- and 3-grams as features, but no improvements were gained.

Overall f-score figures do improve, but are still poor; while recall now may be enough to provide useful feedback to a user (over 50% in most cases), precision is low (probably low enough to make this feedback confusing at best). We did obtain higher precision (in some cases over 50%) by using a simple n-gram-based classification method (we trained a trigram language model, and manually set a suitable likelihood threshold), but this was not consistent across all 5 meetings, and came at the cost of much lower recall (c.10%).

¹ Agreement here is calculated simply with regard to the binary classification of utterances as being action-item-related utterances or not; their classification as belonging to *the same* action item has not been tested.

Table 1. Baseline Classification Performance

Meeting	Number of AI Utterances	Precision	Recall	F-Score
1	22	0.31	0.50	0.38
2	27	0.36	0.33	0.35
3	18	0.28	0.55	0.37
4	15	0.20	0.60	0.30
5	9	0.19	0.67	0.30

4 Hierarchical Annotations

Two problems will be apparent: firstly, the accuracy is lower than desired; secondly, the identification of related utterances does not in itself go very far towards allowing us to identify the action items themselves, and to extract their associated properties (deadline, owner etc.). It became apparent during this phase that the utterances in question form a very heterogeneous class, including some distinct sub-classes which perform different discourse functions and have their own distinct features. Treating these distinct classes as one leads to a classifier which must be too general to give good precision, or too specific to give good recall. Treating them separately and combining the results (along the lines of [17]) might allow better performance. Our next step was therefore to produce an annotation schema which incorporates an explicit distinction between these distinct utterance sub-classes. As discussed above, we decided on the following criteria to determine if an exchange of dialogue specifies an action item:

1. The content of the exchange specifies a concrete future action discussed in the meeting that someone would write down on a to-do list.
2. There is an explicit person or persons who will carry out the action item, and agreement by that person(s) to do so.
3. There is a fairly specific timeframe for when that action is expected to be performed.

These criteria yielded four classes, as shown in Table 2. The first three correspond to the discussion and assignment of the individual properties of the action item: the associated **task description**, the **timeframe** for completion of that task, and the **owner** or party responsible for it. The final class is **agreement**, which covers utterances which explicitly show that the action item is accepted or agreed upon. The classes are shown in Table 2, and annotation examples in Figs. 1 and 2.

More specifically, annotation with the **task description** subclass includes any utterances that specify what action is to be done, including the utterances that provide required antecedents for anaphoric references: as “notion of the preliminary patent” does for the statement “we should apply for that right away” in Fig. 2. In short, it includes any utterances that contain the actual words that would be used to put together a short description of the task.

Annotation with the **owner** subclass includes any utterances that explicitly specify who is responsible for ensuring that the action is carried out, as with “you

Table 2. Utterance Sub-classes

Key	Class	Description
D	task description	Utterances containing an explicit description of the task to be carried out.
O	owner	Utterances containing an explicit reference to the responsible party.
T	timeframe	Utterances containing an explicit reference to the timeframe for completion.
A	agreement	Utterances explicitly signalling acceptance or agreement.

should ...start moving in that direction”, but not e.g. those whose function (rather than explicit surface form) might be taken to do so implicitly (such as agreements by the responsible party). The **timeframe** subclass includes any utterances that explicitly refer to when a task may start or when it is expected to be finished; this is often not specified with an exact date – as with, “by the end of next week,” or “before the trip to Aruba” – but the time that the action is expected to be performed should still be fairly clear.

Finally, the **agreement** subclass includes any utterances in which people agree that the action should and will be done. These are often acknowledgements by the owner to carry out the task, but can be utterances made when other people express their agreement that an action should be done or that a particular person should do it.

Note that a single utterance may be assigned to more than one of these classes: “**John, you** should do that **by next Monday**” might count as **owner** and **timeframe**. Likewise, there may be more than one utterance of each subclass for a single action item: John’s response “OK, **I’ll** do that” would also be classed as **owner** (as well as **agreement**).

While we are not currently requiring all of these subclasses to be present for a set of utterances to qualify as denoting an action item, we expect any action item to include most of them. Figure 3 shows an example of a less concrete task decision, *not* classified as an action item in this case. As we annotate more data, we hope to get a more concrete idea of the effect of the presence/absence of the individual classes, and hope to use that to gain more insight into the distinction between the specific action items we concentrate on here, and more general decisions and commitments.

To date, we have applied this annotation schema to 12 meetings, using the NOMOS annotation software [18]: the sequence of 5 related short CALO meetings, and a random selection of 6 ICSI meetings and 1 ISL meeting (as described in the baseline section above). An initial assessment of reliability between 2 annotators on the single ISL meeting (chosen as it presented a significantly more complex set of action items and annotations than the others in this current set) was encouraging, with reasonable figures for the kappa statistic for each of the distinct sub-classes. The best agreement was achieved on **timeframe** utterances ($\kappa = .86$), with **owner** utterances slightly less good (between $\kappa = .77$), and

Speaker	Utterance	D	O	T	A
CYA	yeah. also, 'cause you said you were gonna send me an email about how to set up our travel.	x	x		
HHI	yeah, I'm gonna send- yeah, I'll send you the email uhm uhm when I go back. send you the email. uhm and you're gonna have to contact him, and they have a travel agency.	x	x	x	x
CYA	okay.				x

Fig. 1. A nice neat example, where most of the work is done by one utterance. Here, the desired task description might be something like “send CYA an email about setting up travel”; the timeframe “when I go back”, and the owner is HHI.

Speaker	Utterance	D	O	T	A
SAQ	not really. the there was the uh notion of the preliminary patent , that uh	x			
FDH	yeah, it is a cheap patent.				
SAQ	yeah.				
CYA	okay.				
SAQ	which is				
FDH	so, it is only seventy five dollars.				
SAQ	and it is it is e an e				
CYA	hm, that is good.				
HHI	talk to				
SAQ	yeah and and it is really broad, you don't really have to define it as w as much as in in a you know, a uh				
FDH	yeah.				
HHI	I actually think we should apply for that right away.	x	x	x	
CYA	yeah, I think that is a good idea.				x
HHI	I think you should, I mean, like, this week , s start moving in that direction. just 'cause that is actually good to say, when you present your product to the it gives you some instant credibility.		x	x	
SAQ	[Noise]				
SAQ	mhm.				x
CYA	right.				x

Fig. 2. A messier example. Here the desired task description might be something like “apply for preliminary patent”, and so the **description** utterances must include SAQ’s original mention of “preliminary patent” as well as HHI’s proposal to “apply for” it. The timeframe would be “right away, this week”, and the owner seems to be “you” (whoever HHI is addressing).

agreement and **description** utterances worse but still acceptable ($\kappa = .73$). Further annotation and refinement of the schema is in progress.

Speaker	Utterance	D	O	T	A
CYA	so, both the charger and the interface need to be designed , 'cause you need to figure out how you're gonna attach the charger to the batteries.	x	x		
SAQ	yeah , [Smack] yeah . and then that that is where we'll address the issue of the parallel versus series of configuration of the batteries.		x		x
AOF	yeah, good call . [Noise]				x
CYA	right .				x
CYA	I think some of these things obviously we wanna get as much done before the meeting as possible, but some of them can uh will have to wait, like marketing will have to wait. the testing can wait. we don't really need to get into that yet. and the battery charger is something that				
FDH	mhm.				
HHI	we don't need to do that anyway.				
CYA	exactly. because we can just put fresh batteries in right now, if we need to.				

Fig. 3. An example of a less concrete task decision. Here, there is a decision, a joint commitment that there is a task which needs doing (“design the charger and interface”, perhaps), and it seems to get agreed - so a conscientious note-taker might add that to a to-do list. But there is no concrete assignment to a person, and no definite time frame (in fact, they seem to decide NOT to take action on this task immediately, but leave it till later).

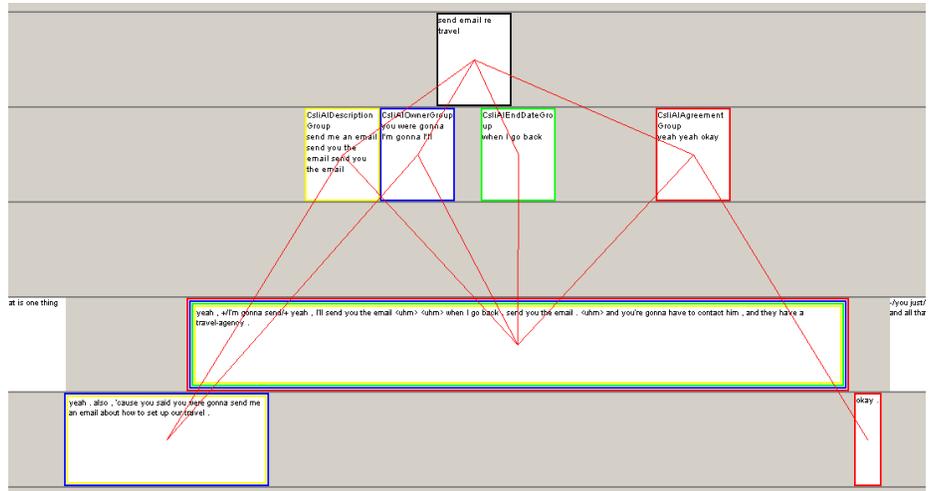


Fig. 4. An example of the hierarchical annotation scheme for a single action item.

5 Experiments

Given the small amount of data currently available, a full evaluation of the proposed classification and detection approach is not possible, but we can perhaps

get some indications. We first trained individual classifiers for each of the utterance sub-classes. For **agreement** utterances, we used a naive n-gram classifier very similar to that of [10] for dialogue act detection, scoring utterances via a set of most predictive n-grams of length 1–3 and making a classification decision by comparing the maximum score to a threshold (where the n-grams, their scores and the threshold are automatically extracted from the training data). For **owner**, **timeframe** and **task description** utterances, we used support vector machines as before, with word unigrams as the features (2- and 3-grams gave no improvement – we expect that this is due to the small amount of training data). Again, we cross-validated by testing on each of the 5 CALO meetings separately, with the training set in each case being the other 4 CALO meetings, plus the fixed ICSI/ISL set. Performance varied greatly by sub-class (see Table reftab:res1), with some (e.g. agreement) achieving higher accuracy than the baseline flat classifications, but others being worse. As there is now significantly less training data available to each sub-class than there was for all utterances grouped together in the baseline experiment, worse performance might indeed be expected; it is encouraging that some sub-classes do better. The worst performing class is **owner**; we suspect parse information may help here (see below).

Table 3. Sub-class Classification Performance

Class	Precision	Recall	F-Score
D	0.23	0.41	0.29
O	0.12	0.28	0.17
T	0.19	0.38	0.26
A	0.48	0.44	0.40

However, even with poor performance for some of the individual sub-classifiers, we should still be able to combine them to get a benefit as long as their true positives correlate better than their false positives (intuitively, if they make mistakes in different places). So far we have only conducted an initial naive experiment, in which we combine the individual classifier decisions in a weighted sum over a window (currently set to 5 utterances). If the sum over the window reaches a given threshold, we hypothesize an action item, and take the highest-confidence utterance given by each sub-classifier in that window to provide the corresponding property. As shown in Table 4, this gives reasonable performance on most meetings, although it does badly on meeting 5 (apparently because no explicit agreement takes place, while our manual weights emphasised agreement).² Most encouragingly, the correct examples provide some useful “best” sub-class utterances, from which the relevant properties could be extracted – see Fig. 5.

We are confident that these results can be improved significantly: rather than sum over the binary classification outputs of each classifier, we can use their

² Accuracy here is currently assessed only over correct detection of an action item in a window, not correct assignment of all sub-classes.

Table 4. A first experiment at combined classification

Meeting	Number of AIs	Correct	False Pos	False Neg	F-Score
1	3	2	1	1	0.67
2	4	1	0	3	0.40
3	5	2	1	3	0.50
4	4	4	0	0	1.00
5	3	0	1	3	0.00

T the start of week three just to
O reconfirm everything and at that
time jack i'd like you to come back
to me with the
D the details on the printer and server
A okay
O so jack uh for i'd like you to
D have one more meeting on um uh
T in in a couple days about uh
A okay

Fig. 5. Examples from meeting 4, with “best” sub-class utterances in dialogue order

confidence scores or posterior probabilities, and learn the combination weights to give a more robust approach. There is still a long way to go to evaluate this approach over more data, and to evaluate the accuracy and utility of the resulting sub-class utterance hypotheses.

6 Discussion and Future Work

We have shown that taking a notion of structure into account seems advantageous when detecting action items in spoken discourse. Without one, classification accuracy is limited; with one, we believe that accuracy can be improved, and the detected utterances can be used to provide the properties of the action item itself. An interesting question is how and whether the notion of structure we use here relates to notions of discourse structure in more general use. If a relation exists, this would help shed light on the decision-making process we are attempting to (begin to) model; and might provide us with a way of using other more plentiful annotated data.

The main priority for our current and future efforts is the annotation of more meetings in order to obtain sufficiently large training and test sets. This effort will concentrate on those meetings from the ICSI, ISL, and CALO corpora which contain decision-making dialogues (in some types of meeting, action items are very sparse). Once more annotated data is available, we will also be able to examine the CALO and ICSI corpora for correlations with existing annotations for other related phenomena, such as meeting acts [19] and dialog acts [8], which may add useful information for features not currently being modelled.

Another priority is to examine the effect on performance when working from speech recognition hypotheses (as opposed to the human transcripts used in this paper), and the best way to incorporate multiple hypotheses (either as n-best

lists or word confusion networks). This will allow us to incorporate action item detection into a working system, e.g. the CALO assistant.

We are also actively investigating alternative approaches to sub-classifier combination: the method used so far is rather ad-hoc and manually defined by trial and error, and better performance (and a more robust and trainable overall system) might be obtained by using a Bayesian network, or a maximum entropy classifier as used by [17].

Another avenue of research we will be pursuing in collaboration with vision and speech researchers on the CALO project will be to integrate multimodal and paralinguistic information as model features. In particular, we expect gaze, head pose, and prosody to help in distinguishing action item agreement and assignment utterances from less relevant classes (e.g. backchannels); and we are examining the incorporation of written and drawn information (in particular, milestones drawn on project sketches) to improve deadline extraction.

We are also developing an interface to a new large-vocabulary version of the Gemini parser [20], allowing us to use semantic parse information, firstly as features in the individual sub-class classifiers, and secondly to extract entity and event representations from the classified timeframe, owner and task description utterances – eventually working towards a full semantic representation for the action item [21]. This can then be supplied to other agents within the CALO system to provide useful functionality for a user, such as the automatic addition of entries to calendars and to-do lists.

Acknowledgements

This work was supported by the CALO project, DARPA grant NBCH-D-03-0010. The content of the information in this publication does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

1. Gruenstein, A., Niekrasz, J., Purver, M.: Meeting structure annotation: Data and tools. In: Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue. (2005) 117–127
2. Lisowska, A., Popescu-Belis, A., Armstrong, S.: User query analysis for the specification and evaluation of a dialogue processing and retrieval system. In: Proceedings of the 4th International Conference on Language Resources and Evaluation. (2004)
3. Banerjee, S., Rosé, C., Rudnicky, A.: The necessity of a meeting recording and playback system, and the benefit of topic-level annotations to meeting browsing. In: Proceedings of the 10th International Conference on Human-Computer Interaction. (2005) 643–656
4. Nicholas C. Romano, J., Jay F. Nunamaker, J.: Meeting analysis: Findings from research and practice. In: Proceedings of the 34th Hawaii International Conference on System Sciences. (2001)

5. Corston-Oliver, S., Ringger, E., Gamon, M., Campbell, R.: Task-focused summarization of email. In: Proceedings of the Text Summarization Branches Out ACL Workshop. (2004)
6. Allen, J., Core, M.: Draft of DAMSL: Dialog act markup in several layers (1997) <http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/>.
7. Core, M., Allen, J.: Coding dialogues with the DAMSL annotation scheme. In Traum, D., ed.: AAAI Fall Symposium on Communicative Action in Humans and Machines. (1997) 28–35
8. Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., Carvey, H.: The ICSI Meeting Recorder Dialog Act Corpus. In: Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue. (2004) 97–100
9. Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Ess-Dykema, C.V., Martin, R., Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics* **26**(3) (2000) 339–373
10. Webb, N., Hepple, M., Wilks, Y.: Dialogue act classification using intra-utterance features. In: Proceedings of the AAAI Workshop on Spoken Language Understanding. (2005)
11. Goodwin, C.: Action and embodiment within situated human interaction. *Journal of Pragmatics* **32** (2000) 1489–1522
12. Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., Wooters, C.: The ICSI Meeting Corpus. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). (2003) 364–367
13. Burger, S., MacLaren, V., Yu, H.: The ISL Meeting Corpus: The impact of meeting type on speech style. In: Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP 2002). (2002)
14. Siegel, S., N. J. Castellan, J.: *Nonparametric Statistics for the Behavioral Sciences*. 2nd edn. McGraw-Hill (1988)
15. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer (1995)
16. Joachims, T.: Making large-scale SVM learning practical. In Schölkopf, B., Burges, C., Smola, A., eds.: *Advances in Kernel Methods – Support Vector Learning*. MIT Press (1999)
17. Klein, D., Toutanova, K., Ilhan, H.T., Kamvar, S.D., Manning, C.D.: Combining heterogeneous classifiers for word-sense disambiguation. In: Proceedings of the ACL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions. (2002) 74–80
18. Niekrasz, J., Gruenstein, A.: A software framework for semantic annotation of multimodal discourse. In: The 5th Conference on Language Resources and Evaluation (LREC '06). (2006)
19. Bates, R., Menning, P., Willingham, E., Kuyper, C.: Meeting acts: A labeling system for group interaction in meetings. In: The 9th European Conference on Speech Communication and Technology (Interspeech '05). (2005) 1589–1592
20. Dowding, J., Gawron, J.M., Appelt, D., Bear, J., Cherny, L., Moore, R., Moran, D.: Gemini: A natural language system for spoken language understanding. In: Proc. 31st Annual Meeting of the Association for Computational Linguistics. (1993)
21. Niekrasz, J., Purver, M., Dowding, J., Peters, S.: Ontology-based discourse understanding for a persistent meeting assistant. In: *Persistent Assistants: Living and Working with AI: Papers from the 2005 AAAI Spring Symposium*. (2005)