Can Requests-for-Action and Commitments-to-Act be Reliably Identified in Email Messages?

Andrew Lampert †‡
†CSIRO ICT Centre
North Ryde NSW Australia
Andrew.Lampert@csiro.au

Cécile Paris
CSIRO ICT Centre
North Ryde NSW Australia
Cecile.Paris@csiro.au

Robert Dale
‡Centre for Language Technology
Macquarie University
NSW Australia
rdale@ics.mq.edu.au

Abstract *This paper reports on the results of an exploratory annotation task where three coders classified the presence and strength of Requests-for-Action (requests) and Commitments-to-Act (promises) in workplace email messages. The purpose of our annotation task was to explore levels of human agreement to establish whether this is a repeatable task that lends itself to automation. The results from our annotation task suggest that there is relatively high agreement about which sentences embody Requests-for-Action ($\kappa = 0.78$), but poorer agreement about Commitments-to-Act ($\kappa = 0.54$). Analysis of cases of coder disagreement highlighted several areas of systematic disagreement which we believe can be addressed through refining our annotation guidelines. Given this scope for improving agreement, we believe the results presented here are encouraging for our intention to perform larger-scale annotation work leading to automation of the detection and classification of Requests-for-Action and Commitments-to-Act in email communication.*

Keywords Email, document workflow, document management, Speech Acts, task management

1 Introduction

It is well documented that users routinely use email for managing requests and commitments in the workplace (e.g., [15, 7]). With the volume of email ever increasing, previous studies have highlighted that when facing the need to manage multiple ongoing tasks through email, users regularly feel overloaded [21, 3].

As a consequence of feelings of overload, many users struggle to give appropriate attention to tasks hidden in email that require action. The problem is severe enough that some people have even advocated giving up on managing the mountain of pending email tasks altogether by declaring ”Email Bankruptcy”[1] and wiping the slate clean. Of course, this serves only to restart the process, and all too soon important requests and promises are likely to be unintentionally overlooked or ignored in the noise of an overloaded inbox.

Our research aims to help email users by providing them with tools that can automatically detect Requests-for-Action and Commitments-to-Act within their incoming and outgoing email messages. As we explain in Section 4.1 and 4.2, we define a Request-for-Action to be a sentence that creates some form of obligation for an email recipient to do something (a request), and a Commitment-to-Act as some form of commitment from the email sender to perform some action (a promise). An example Request-for-Action from our data is *Please call when you have a chance.*; an example Commitment-to-Act is *I'll keep you posted on any changes.*

Requests-for-Action and Commitments-to-Act in email are particularly interesting because they occur in the context of textual conversations. Because they form part of a conversation, email messages differ from most other forms of written documents that have been more widely studied in the IR and NLP communities. One such difference is that email messages involve interaction between two or more participants, and as a result, the structure of email messages and conversations includes patterns borrowed from verbal conversation. We believe this structure can be usefully exploited to improve the efficiency and effectiveness of managing tasks in email. An example is the role of adjacency pairs of Speech Acts across email messages, which we believe can be used to help associate requests or promises with relevant responses from later email messages within the same conversation.

Proceedings of the 10th Australasian Document Computing Symposium, Melbourne, Australia, December 10, 2007.

[1] See, for example, http://www.wired.com/culture/lifestyle/news/2004/06/63733

There are also a variety of ways in which information about Requests-for-Action and Commitments-to-Act can be used to augment and improve existing tools for managing email, some of which we have presented previously [12]. We hope to couple information about Requests-for-Action and Commitments-to-Act with notions of conversation structure to provide tools that allow users to be more efficient and effective in identifying, prioritising, acting on and monitoring actionable content in email messages.

In the work we present in this paper, we focus on an experiment involving human coders manually identifying Requests-for-Action and Commitments-to-Act in workplace email messages. To automatically identify these actions is a non-trivial language processing task, because of the variety of surface forms that people can use to phrase their requests and promises. We want to investigate how closely humans agree in interpreting requests and promises in email. The problem of ambiguous surface forms is closely related to challenges encountered when automatically classifying Speech Acts associated with utterances. Similarly, difficulties that arise in recognising implicit, indirect Speech Acts must also be resolved when attempting to identify Requests-for-Action or Commitments-to-Act that are expressed implicitly or whose interpretation depends on other aspects of the email message context such as relationships between the sender and recipient, or the previous conversation history.

Our paper is structured as follows. First, in Section 2, we describe related work that has looked at email from an action-oriented perspective, and explain how our work both builds on and differs from these activities. In Section 3, we describe how we have extracted and processed data from the Enron Email Corpus for use in our pilot annotation task that is described in Section 4. We then present the results of our annotation experiment in Section 5 and discuss the levels of human agreement in identifying Requests-for-Action and Commitments-to-Act in Section 6 before making some concluding remarks and indicating our directions for future work in Section 7.

2 Related Work

In considering email from a conversational point of view, our work differs from the approach of many existing search engines and email systems which routinely treat email messages as simple bags-of-words, ignoring any conversational structure. The idea of identifying and exploiting patterns of communicative acts in textual conversation is, however, not new. Two decades ago, Flores and Winograd [22] proposed that workplace workflow can be seen as a process of creating and maintaining networks of conversations in which requests and promises lead to successful completion of work. Our research is attempting to build on these influential ideas to provide intelligent, automated assistance to email users. The work presented in this paper represents some early exploratory steps towards this goal.

While our approach differs from most existing email software, we are certainly not alone in looking at email from an action-oriented point of view. In particular, there is a growing body of research that has taken ideas from Speech Act Theory [1, 19] and applied them to analysing and enhancing email communication. This existing work differs as to whether the Speech Acts (or more properly, Speech-Act-inspired units) should be annotated at the message level (as in [5, 14]) or at the utterance/sentence level (as in [6]). Our thesis is that a single email message may contain multiple Requests-for-Action and Commitments-to-Act on a range of tasks and topics, and thus our work focuses on sentence-level classification. It is possible that another classification unit is also appropriate, such as clause-level annotation, but for simplicity the annotation task we report on here is performed at the sentence level.

The most similar work to our own is that of Corston-Oliver et. al. [6], whose SmartMail system attempted to automatically extract and reformulate action items from email messages for the purpose of adding them to a user's to-do list. While their task is similar in nature, our work differs in a number of aspects. First, Corston-Oliver et. al. only attempt to identify tasks that "looked like an appropriate item to add to an on-going 'to do' list". In particular, they note that factual questions are not annotated as tasks because responding fulfills any associated obligation. In contrast, we annotate factual questions, and indeed any non-rhetorical questions, as Requests-for-Action, because some level of obligation to respond is imposed on the recipient, and thus the user may benefit from tools that can aid in identifying and tracking such requests.

A more significant distinction is that we are focused on identifying both Commitments-to-Act and Requests-for-Action, due to our broader ideas about tasks that a user may want to monitor. While Corston-Oliver et. al. do include a *promise* category in their annotation taxonomy, no mention is made of its use, and no definition is given for what they consider to be *promises*. Additionally, they restrict coders to applying a single tag to each sentence, meaning that a sentence cannot embody both a *task* and a *promise*. In the results of our annotation task, we see that this is a potentially artificial restriction.

Our motivation for considering Commitments-to-Act in addition to Requests-for-Action (which subsume the *tasks* that Corston-Oliver et. al. were identifying) is well explained by Murray's observations borne out of her ethnographic research into the use of electronic messaging at IBM in the early 1990s [16]. In that work, she notes that "They [managers] would like to be able to track outstanding promises they have made, promises made to them, requests they've made that have not been met and requests made of them that they have not fulfilled." More recent studies of email usage by Bellotti et. al. in their work on the TaskMaster system [3] also identified similar problems with "keeping track of lots of concurrent actions: One's own to-dos and to-dos one expects from others" using existing email clients. To allow users to keep track of actions in this way requires an ability to identify both Requests-for-Action (requests) and Commitments-to-Act (promises) in email messages.

Work by Kushmerick and Lau [11] is also similar to our work in terms of both their focus on ongoing activities that occur via email, and their aim to provide high-level overviews of ongoing activities to help users manage these tasks more effectively. The application of their ideas is, however, restricted to repeated, structured processes such as those that occur with e-commerce transactions, whereas our interest lies beyond these in the more informal exchange of Requests-for-Action and Commitments-to-Act that occurs in task-oriented workplace email conversations.

3 Preparation of Email for Annotation

To prepare for our pilot annotation task, we manually extracted 54 email messages containing 310 sentences, which we stored in MySQL database tables. Email messages were selected to represent a variety of syntactic styles of expressing possible Requests-for-Action and Commitments-to-Act. In particular, we attempted to select examples of sentences representing both explicit and implicit Requests-for-Action and Commitments-to-Act, and sentences with and without explicit task addressivity (i.e., sentences that do and do not address Requests-for-Action to a specific, named recipient).

3.1 Email Corpus

Our set of 54 messages for the pilot annotation task were selected from a random subset of the Enron Email corpus [9]. For this exploratory stage of annotation, selected messages were constrained to be of less than 12 sentences, to reduce the effort associated with coding longer messages.

We used the version of the Enron corpus released as a MySQL database dump[2] by Andrew Fiore and Jeff Heer at the University of California, Berkeley. This database version of the corpus has had a substantial amount of processing performed on the contents of the raw Enron corpus. This processing includes removing duplicate email messages and normalizing names of senders and recipients. The end result is a corpus of just over 250,000 email messages. Like all publicly released versions of the Enron Email corpus, no attachments are included with any of the email messages.

3.2 Processing Email Body Text

We pre-processed the text in the body of each email message to try to remove email signatures and all quoted or forwarded email content, leaving only text written by the author of the current email message. We will refer to such original content written by the current email sender as *author text*.

We used the Jangada software [4] to identify signature blocks and reply lines. Unfortunately, in using this software, we identified similar shortcomings to those identified recently by Estival et. al. [8] who used Jangada on their own email corpus. In particular, Jangada did not accurately identify forwarded or reply material in the email messages we used for our annotation task. We posit that at least one factor in the poor performance of Jangada is due to it being trained on data from Usenet newsgroups, which generally uses different syntactic markers for forwarded and quoted material. We believe this is a significant factor in the systematic errors that Jangada makes in failing to identify quoted reply and forwarded content represented in the style used by Microsoft Outlook. Unfortunately, Outlook is the most common email client used to compose messages in the Enron corpus.

Unlike Estival et. al., whose work we only became aware of after running our pilot annotation, we did not develop our own document parser for email messages. Given a lack of other available email processing tools, we used the Jangada software despite its shortcomings, and allowed coders to flag processing errors during the annotation process.

3.3 Sentence Splitting

Because our annotation is performed at the sentence level, once we had attempted to remove quoted reply content, forwarded content and email signatures, we segmented the body of each email message into sentences. For this purpose, we used the SentParBreak sentence and paragraph segmenter [18]. SentParBreak uses heuristic rules for identifying the boundaries of sentences and paragraphs. We have not yet attempted to refine these rules for email data, and instead used SentParBreak without modification. We applied SentParBreak to the bodies of all 250,000 email messages in our corpus and produced just over 3 million sentences of probable author text. Due to the Jangada processing errors, however, we know that some proportion of these sentences actually contain quoted reply or forwarded email content.

Although we haven't formally evaluated the performance of the SentParBreak as a sentence segmenter, we allowed coders to flag sentence segmentation errors during the annotation process, in the same way as we did for Jangada processing errors.

4 Annotation Task

The purpose of our pilot annotation task was to explore the levels of human agreement in identifying Requests-for-Action and Commitments-to-Act within email messages. We gave three coders (the authors of this paper) a set of annotation guidelines, described in Section 4,

[2] Available at: http://bailando.sims.berkeley.edu/enron/enron.sql.gz

and asked them to independently classify Requests-for-Action and Commitments-to-Act within a corpus of 310 sentences from 54 email messages.

The annotation guidelines were written by the first author with comments received from the other authors. The set of guidelines was finalised before any annotation took place.

Our pilot annotation task was performed using a custom web-based tool, developed using Ruby-on-Rails. This annotation tool displays email messages from the Enron Email database using a look and feel that approximates the way email messages are displayed in Microsoft Outlook. Each email message was presented with the usual preceding header fields and values: *From, Date, To, Cc, Bcc, and Subject*. Below the header information, in the email content pane, the author text of the email message was presented as a sequence of sentences, one per line. Paragraph breaks in the original email were represented by a single uncodable blank line in the annotation tool.

For each sentence, coders were required to select annotation values from a number of aligned drop-down menus. Using these menus, coders performed three actions for each sentence:

1. First coders indicated whether the sentence expressed a Request-for-Action for the Specified Recipient, and, if so, whether the Request-for-Action was weak, medium or strong. (See below for an explanation of Specified Recipient.)

2. Next, coders indicated whether the sentence expressed a Commitment-to-Act from the sender, and if so, whether the Commitment-to-Act was weak, medium or strong.

3. Finally, coders could optionally flag any processing errors with the sentence. Flaggable processing errors included sentence segmentation problems, and the inclusion of quoted or forwarded email material (non author text).

Coders were instructed to annotate each sentence in the context of the original email message, rather than in isolation. At the top of each message, one recipient to whom the email message was originally sent – the Specified Recipient – was noted, and coders were instructed to approach the annotation task from the point of view of that person. For the purposes of our pilot annotation task, the first non-sender recipient of the message was chosen as the Specified Recipient. The 'non-sender' constraint was introduced to counter cases where the first recipient was actually the sender (presumably copying their own email to themself for action or recall purposes). Where no explicit recipients were identifiable, as in the case of an email message whose recipients are all Bcc'd, coders were instructed to annotate the email message from the point of view of any recipient. In general, Requests-for-Action and Commitments-to-Act in such messages tend to be addressed to all recipients, meaning that annotating from a the point of view of a generic recipient is acceptable. Coders were instructed not to mark any Requests-for-Action directed explicitly to recipients other than the Specified Recipient as Requests-for-Action.

To explain or comment on any aspect of the annotation task coders were able to make notes using a comments field for each email message. This comments field was also used in combination with the *Other* category of processing errors to highlight processing or display problems other than segmentation and author text related issues. Coders were instructed to use the comments field to explain any annotation decisions which they felt were conditional or context-sensitive. For example, if a coder's decision depended on potentially ambiguous interpretation of the email message or its context, they were instructed to explain the basis for their annotation.

Finally, we also noted to coders that it was possible for a single sentence to contain both a Request-for-Action and a Commitment-to-Act: e.g., *Please send the document today, so I can get comments back to you by Monday*. The annotation tool made it possible for any sentence to be annotated as both a Request-for-Action and a Commitment-to-Act.

4.1 Requests-for-Action

A Request-for-Action places some form of obligation on the recipient to respond or act. An example Request-for-Action is: *Do you have an outage calendar for 2002?*.

A simple test for identifying a Request-for-Action is: Is this sentence asking me to do something? Examples of actions can include, but are not restricted to:

- answering a question, in email or otherwise;
- forwarding the message to a new recipient; or
- performing some action in the real world, such as preparing a document or gathering some data.

Instructions to coders were to annotate any sentence that carried an expectation that the Specified Recipient of the email message should respond or take some action as a Request-for-Action.

As we explained in Section 4, coders were also asked to indicate the strength of each Request-for-Action they identified. The different strengths were explained as follows:

- **Strong:** Action or response from the Specified Recipient is considered important and/or mandatory.

- **Medium:** The sender expects a response or action from the Specified Recipient.

- **Weak:** Action or response from the Specified Recipient is optional or conditional; the sender would find it reasonable if the Specified Recipient took no action.
- **None:** No Request-for-Action is expressed.

4.2 Commitments-to-Act

A Commitment-to-Act represents a promise from the author that action will be taken. An example Commitment-to-Act is: *Enron has committed to processing and paying any expenses incurred prior to transaction closing*.

A simple test for identifying a Commitment-to-Act is: Is this sentence promising to do something?

Commitments-to-Act occur both when an action is to be taken by the sender, or when the sender promises action on behalf of another person. The reason for including such indirect Commitments-to-Act is based on the intuition that such delegated promises might occur frequently because of the hierarchical nature of many workplaces, and are likely to be an important part of workplace conversations for action. An example of an indirect Commitment-to-Act is: *Peter will call to let you know the final arrangements*.

Coders were instructed to annotate any sentence that carried an expectation that the sender of the email message will take responsibility for some action being taken as a Commitment-to-Act.

As for Requests-for-Action, coders were also asked to indicate the strength of each Commitment-to-Act that they identified. The different strengths were explained as follows:

- **Strong:** Action from the Sender is considered important and/or mandatory.
- **Medium:** The Specified Recipient expects a response or action from the Sender.
- **Weak:** Action or response from the Sender is optional or conditional; the Specified Recipient would find it reasonable if no action was taken.
- **None:** No Commitment-to-Act is expressed.

4.3 Audience Types

For each email message in our pilot annotation corpus, we also manually classified the nature of the recipient audience. Each email was classified with one of the following message types:

- **Single Recipient:** Addressed to a single recipient. We also consider email messages that are addressed specifically to a single recipient but Cc'd or Bcc'd to another recipient to belong to this group.
- **Closed Group:** Addressed explicitly to a specified group of recipients. Each recipient must be identifiable from the email headers.
- **Broadcast:** Addressed to an unspecified group of recipients, such as a group alias or mailing list.

We used information about the audience type of email messages in analysing intercoder agreements for our annotation task, as we describe below in Section 5.

5 Results and Discussion

The results of our pilot annotation are shown in Table 1 and Table 2. Note that all κ values referred to in this section, apart from the pairwise κ values in Tables 1 and 2, refer to agreement between all three coders calculated using the generalization of Cohen's Kappa to more than two coders, as specified by Krippendorff [10].

The results in Table 1 show pairwise and three-way interannotator agreement between coders for our pilot annotation task. Separate *binary* and *strength* κ values are given for each measurement.

Binary agreement refers to interannotator agreement about which sentences contain a Request-for-Action, ignoring any indication of strength. To calculate these κ scores, we collapsed all three strengths of annotated Requests-for-Action into a single Request-for-Action class. Thus, disagreement about the strength of a Request-for-Action is ignored in binary κ scores.

Strength agreement, which is always lower than binary agreement, refers to interannotator agreement for the more fine-grained strength categories of Requests-for-Action. Thus, it represents agreement between coders over which sentences contain a Strong, Medium, Weak or No Request-for-Action. Disagreement about the strength of an identified Request-for-Action is considered a complete disagreement for the strength κ scores.

The results in Table 2 similarly show both pairwise and three-way interannotator agreement between coders for classifying Commitments-to-Act. The same separate measures of binary and strength agreement are used.

Finally, both Table 1 and Table 2 also show separate κ scores for subsets of the annotation corpus, grouped according to the email audience type (see Section 4.3).

From these results, we can draw several tentative conclusions about human agreement for identifying Requests-for-Action and Commitments-to-Act in email:

1. There is good agreement ($\kappa = 0.78$) about which sentences embody a Request-for-Action.
2. There is some tentative agreement ($\kappa = 0.60$) about the strength of Requests-for-Action.
3. There is poorer agreement about which sentences embody a Commitment-to-Act ($\kappa = 0.54$) and poor agreement about the strength of those commitments ($\kappa = 0.37$).

	Coder A & B		*Coder A & C*		*Coder B & C*		*3-Way* κ	
Message Types	Binary	Strength	Binary	Strength	Binary	Strength	Binary	Strength
All	0.83	0.57	0.79	0.66	0.74	0.56	0.78	0.60
Single Recipient	0.85	0.54	0.79	0.66	0.80	0.55	**0.81**	0.58
Closed Group	0.79	0.55	0.79	0.58	0.66	0.61	0.75	0.58
Broadcast	0.82	0.66	0.77	0.70	0.70	0.53	0.76	0.63

Table 1: Pairwise and 3-way Kappa agreements for classifying Requests-for-Action

	Coder A & B		*Coder A & C*		*Coder B & C*		*3-Way* κ	
Message Types	Binary	Strength	Binary	Strength	Binary	Strength	Binary	Strength
All	0.45	0.30	0.62	0.44	0.55	0.37	0.54	0.37
Single Recipient	0.51	0.28	0.76	0.54	0.52	0.41	**0.60**	0.41
Closed Group	0.14	0.06	0.51	0.32	0.41	0.19	0.35	0.19
Broadcast	0.52	0.41	0.52	0.40	0.66	0.41	0.57	0.41

Table 2: Pairwise and 3-way Kappa agreements for classifying Commitments-to-Act

4. The level of agreement appears to vary depending upon the audience type of the email message.

A particularly interesting aspect of our results is the variation in interannotator agreements across the different audience types that we identified in Section 4.3. As can be seen both in Tables 1 and 2, agreement between our three coders about the presence and strength of both Requests-for-Action and Commitments-to-Act is highest for Single Recipient email messages, and lowest for Closed Group email messages. In the case of Commitments-to-Act, this difference is particularly marked. Analysing the cases of disagreement has not yet revealed a recurring reason for these differences. Some coders did make observations that the strength of Requests-for-Action may be determined, to some extent, by the probability that it would apply to the Specified Recipient when an email message has multiple recipients. It's unclear how to objectively judge such a probability, but perhaps observations such as this shed some light on the reduced agreement for Closed Group email messages.

In addition to the annotations made regarding Requests-for-Action and Commitments-to-Act, we allowed users to flag segmentation errors, as noted in Section 3.3. Although we don't consider this method a rigorous way to evaluate the SentParBreak sentence splitter, it is interesting to note that the results from this error flagging suggest that segmentation error for our pilot annotation data is at least 10%. This is a much greater error rate than the 0.997352 precision and 0.995093 recall results that were apparently achieved using SentParBreak over the Genia corpus [17]. Such a difference in performance serves to highlight the need for standard NLP tools like sentence segmenters to be re-trained, re-tuned or otherwise tailored when working with email data, due to differences in the nature of textual content.

6 Error Analysis

In analysing cases of disagreement for classifying Requests-for-Action and Commitments-to-Act, it is clear that a significant proportion of problems can be solved by improving the annotation guidelines. Below we discuss several highly-represented classes of disagreement between coders.

6.1 Conditional Offers

One systematic source of disagreement between coders in identifying Commitments-to-Act was how to classify sentences that embody an *offer*. An example from one of our email messages is: *If you wish, I could provide you questions in advance to maximize your time*. Some coders classified sentences such as these as containing only a Request-for-Action for the recipient, while others judged that it contained both a Request-for-Action for the recipient and a Commitment-to-Act for the speaker, who implicitly promises to make good on their offer if the recipient accepts or acts on it.

This difference of opinion is actually not surprising when we look into the Speech Act literature. Offers were originally assigned to either be directive acts (which in our terminology are Requests-for-Action) or commissive acts (Commitments-to-Act) [19, 1]. Subsequent studies, however, have differed greatly over whether to categorise *offers* as commissive acts (e.g., [13]), directive/requestive acts (e.g. [20]), or some hybrid category in-between these (e.g. [2]).

Although the issue is still under discussion for our future annotation tasks, we are currently leaning towards specifying that *offers* should not be annotated as Commitments-to-Act, since no obligation to act is enacted until the recipient accepts the offer. At that point, we would classify the acceptance as a Request-for-Action from the original recipient. Doing so relies on the conversational nature of email communication and being able to identify a response to an offer, perhaps using adjacency pairing cues. Another possibility is to introduce the idea of a conditional Commitment-to-Act, which would qualify Commitments-to-Act in a manner orthogonal to the strength (i.e., we could have strong, medium or weak conditional Commitments-to-Act). As an aside, we could similarly have conditional Requests-for-Action. Regardless of how we decide to proceed, our annotation guidelines will be updated to provide specific instructions to coders specifying how *offers* should be annotated. This will improve our κ agreements considerably, since almost 40% of our disagreements over the presence of Commitments-to-Act are for sentences containing *offers*.

6.2 Implicit Requests

Another common source of disagreement occurs for implicit requests. In our pilot annotation corpus, these occur frequently as statements about meetings. As an example, one email from our corpus contains the following sentence: *Ken wants to have a meeting this afternoon in regard to California from a PR standpoint.* Our coders disagreed as to whether this statement represents a request to attend a meeting, and thus a Request-for-Action. In the original email this sentence is followed by an explicitly directive sentence: *Please let me know if you will be able to attend*. This second sentence is clearly a Request-for-Action, but it creates some confusion over the status of the first sentence. In the absence of the second directive sentence, the first sentence should, we would argue, be interpreted as a Request-for-Action, in essence acting as an indirect Speech Act. When followed by the second sentence, however, this interpretation as an indirect Speech Act is no longer necessary, as the Request-for-Action has moved to the explicit directive sentence. For our pilot annotation task, we directed coders to annotate each sentence in the context of the entire email message, which would therefore indicate that the first sentence should not be annotated as a Request-for-Action. Clearly, however, our annotation guidelines need to provide more specific guidance to coders for how to interpret such situations.

6.3 Relationship Context

In our pilot annotation task, we did not include any information that indicated the organisational hierarchy or relationships between senders and recipients. This lack of specified relationship context for each email message sometime lead to ambiguity in interpreting Requests-for-Action and Commitments-to-Act.

Some coders noted that differences in the relationship between the sender and the Specified Recipient (specifically, differences in the “importance” of the sender) would affect the strength they associated with Requests-for-Action or Commitments-to-Act.

One example sentence on which coders commented is: *You might like to organize the paper from a broad overview of the electrical market in the west (including basic descriptions, timelines, fundamentals etc.) down to a specific description of what you did at Enron*. Two coders commented that their interpretation of the strength of any Request-for-Action would be different depending on the organisational relationship between sender and receiver. If this represented, for example, a manager assigning a task to a direct report, then the sentence would be interpreted to have a different strength than if it were a Request-for-Action from a peer.

7 Conclusions and Future Work

We regard our current results as encouraging for automation of this classification task.

Based on analysis of the results from our pilot annotation task, we are working to refine our annotation guidelines and the annotation task in a number of ways. In addition to the issues discussed in Section 6, it is clear from our interannotator agreement results that we need to improve the way we define and distinguish strength for both Requests-for-Action and Commitments-to-Act. Our current plans in this regard are to revise our finer-grained classification of Requests-for-Action and Commitments-to-Act to capture conditionality and explicitness rather than require our coders to directly annotate the strength of requests and promises.

Once we have refined our annotation guidelines, we plan to run a larger annotation task using a wider pool of annotators. We hope to use our corpus of manually annotated data to bootstrap and refine automated classifiers and to iteratively use these classifiers to apply active learning techniques to select maximally informative email data for future stages of manual annotation.

We also plan to resolve some limitations of our existing email pre-processing steps. In particular, rather than removing non-author text such as quoted and forwarded material from email messages that we present to coders, we plan to include it as content that can be viewed but not annotated. This will be necessary as we expand our work to begin looking at aspects of conversational structure, so that coders can correctly identify features such as adjacency pairs that require knowledge of the email conversation history.

With reference to the interesting variations in interannotator agreement across different Audience Types, it is difficult to say anything conclusive, given the relatively small amount of data annotated in our pilot study. This is something that we plan to explore further in future annotation work.

Overall, given the scope for refining our annotation guidelines together with the current levels of human agreement, we believe that humans can reliably identify and classify Requests-for-Action and Commitments-to-Act in email messages, and thus that our task is repeatable.

Acknowledgements We thank Jeroen Geertzen for the use of his multi-annotator Kappa calculator[3] in calculating three-way interannotator agreement values for our annotation data.

References

[1] John L Austin. *How to do things with words*. Harvard University Press., 1962.

[2] Kent Bach and Robert M Harnish. *Linguistic Communication and Speech Acts*. The MIT Press, 1979.

[3] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard and Ian Smith. Taking email to task: The design and evaluation of a task management centred email tool. In *Computer Human Interaction Conference*, CHI, Ft Lauderdale, Florida, USA, April 5-10 2003.

[4] Vitor R Carvalho and William W Cohen. Learning to extract signature reply lines from email. In *Proceedings of First Conference on Email and Anti-Spam (CEAS)*, Mountain View, CA, July 30-31 2004.

[5] William W Cohen, Vitor R Carvalho and Tom M Mitchell. Learning to classify email into "speech acts". In Dekang Lin and Dekai Wu (editors), *Conference on Empirical Methods in Natural Language Processing*, pages 309–316, Barcelona, Spain, 2004. Association for Computational Linguistics.

[6] Simon H Corston-Oliver, Eric Ringger, Michael Gamon and Richard Campbell. Task-focused summarization of email. In *ACL-04 Workshop: Text Summarization Branches Out*, July 2004.

[7] Nicolas Ducheneaut and Victoria Bellotti. E-mail as habitat: an exploration of embedded personal information management. *Interactions*, Volume 8, Number 5, pages 30–38, September/October 2001.

[8] Dominique Estival, Tanja Gaustad, Son Bao Pham, Will Radford and Ben Hutchinson. Author profiling for english emails. In *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, pages 263–272, Melbourne, Australia, July 2007.

[9] Bryan Klimt and Yiming Yang. Introducing the enron corpus. In *First Conference on Email and Anti-Spam (CEAS)*, 2004.

[10] Klaus Krippendorff. *Content Analysis: An Introduction to its Methodology*. Sage Publications, Beverley Hills, CA, USA, 1980.

[11] Nicholas Kushmerick and Tessa Lau. Automated email activity management: An unsupervised learning approach. In *Intelligent User Interfaces Conference (IUI)*, San Diego, USA, 2005.

[12] Andrew Lampert. Managing obligations and commitments in email. Presentation at NGS07 - the Second HCSNet Workshop of the Next Generation Search Priority Area, Sydney, Australia, 19-20 July 2007.

[13] Geoffrey N Leech. *Principles of Pragmatics*. Longman Publishing Group, 1983.

[14] Anton Leuski. Context features in email archives. In *Proceedings of IRiX workshop at SIGIR'05*, 2005.

[3] http://www.cosmion.net/jeroen/software/kappa/

[15] Wendy E Mackay. More than just a communication system: Diversity in the use of electronic mail. In *ACM conference on Computer-supported cooperative work*, pages 344–353, Portland, Oregon, USA, 1988. MIT, ACM Press.

[16] Denise E Murray. *Conversation for Action: The Computer Terminal As Medium of Communication*, Volume 196 pages. John Benjamins Publishing Co, 1991.

[17] Scott Piao. Sentparbreaker web page, 2007. `http://text0.mib.man.ac.uk:8080/scottpiao/sent_detector`, Accessed: 5/10/2007.

[18] Scott S L Piao, Andrew Wilson and Tony McEnery. A multilingual corpus toolkit. In *Proceedings of AAACL*, 2002.

[19] John R Searle. *Speech Acts : An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

[20] Amy B M Tsui. *English Conversation*. Oxford University Press, 1994.

[21] Steve Whittaker and Candace Sidner. Email overload: exploring personal information management of email. In *ACM Computer Human Interaction conference*, pages 276–283. ACM Press, 1996.

[22] Terry Winograd and Fernando Flores. *Understanding Computers and Cognition*. Ablex Publishing Corporation, Norwood, New Jersey, USA, 1st edition, 1986. ISBN: 0-89391-050-3.