

# Multimodal annotation scheme for online written conversation analysis

November 21, 2014

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background and Related Work</b>	<b>1</b>
2.1	Background . . . . .	1
2.2	Speech act classification applied to forum posts . . . . .	2
2.2.1	Methodology . . . . .	2
2.2.2	Taxonomy . . . . .	2
2.2.3	Annotator agreement . . . . .	4
2.2.4	Classification . . . . .	5
2.2.5	Evaluation . . . . .	6
2.3	Sentence classification in message board posts . . . . .	6
2.3.1	Goals . . . . .	6
2.3.2	Message board specificities . . . . .	7
2.3.3	Classification . . . . .	7
2.3.4	Evaluation . . . . .	8
2.4	Segmentation of email message text . . . . .	9
2.4.1	Goals . . . . .	9
2.4.2	Email zones . . . . .	9
2.4.3	Classification . . . . .	9
2.4.4	Evaluation . . . . .	10
<b>3</b>	<b>Proposal</b>	<b>10</b>
3.1	Classes . . . . .	10
3.1.1	Boilerplate . . . . .	11
3.1.2	Quote . . . . .	12
3.1.3	Expository . . . . .	12

3.1.4	Speech . . . . .	14
3.1.5	Structure . . . . .	15

# 1 Introduction

Our goal is to develop an annotation scheme and protocol for the classification of email sentences or groups of sentences, which will have to be generalized to other forms of online written communication, in terms of speech acts. A further objective is to use such annotated data to produce a sample segmentation of our email corpus so that we can test the validity of the hypotheses made in [Hernandez and Salim, 2014].

## 2 Background and Related Work

In this section we introduce the concept of dialog act (DA) and provide details concerning a few works related to the field of speech act recognition and email segmentation.

### 2.1 Background

Speech act theory [Austin, 1975] attempts to describe utterances in terms of communicative function (e.g. question, answer, thanks...). Indeed, utterances are not limited to their semantic content, they also have a communicative function : a goal, and an effect. For Austin, they can be analyzed at three levels: locutory (the linguistic characteristics of the utterance), illocutory (the intention of the speaker) and perlocutory (the real-world effects of the utterance). When we refer to speech acts<sup>1</sup>, we are interested in the illocutory level of utterance analysis.

Thus, in most works, it is in terms of speech acts that interactions between participants of a conversation are modeled. Austin considers utterances as actions performed by a speaker ; this is based on the idea that every enunciation is the realization of a social act. Verbs that specify these actions are called *performative verbs*, such as when someone says "I grant you the title of captain". But speech acts are not only constituted by these kinds of verbs. [Searle, 1976] offers five classes of dialog acts : assertives (assertion...), directives (order, request, advice, etc.), commissives (promise, invitation, etc.), expressives (congratulations, thanks, etc.) and declarations (war declaration, nomination, baptism, etc.).

There are a number of different speech act taxonomies [Traum, 2000]. Table 1 details two founding taxonomies: Austin's and Searle's. Table 2

---

<sup>1</sup>Also known as dialog acts

Act	Description or examples	Reference
Verdictives	to condemn, to decree...	[Austin, 1975]
Exercitives	to command, to order, to forgive...	
Commissives	to promise, to guarantee, to bet, to swear...	
Behabitives	to apologize, to thank, to criticize...	
Expositives	to assert, to deny, to postulate...	
Assertives	asserts a state of fact	[Searle, 1976]
Directives	attempt to make an interlocutor do something	
Commissives	commitment on the part of the speaker	
Expressives	expression of a psychological state	
Declarations	statement with a direct impact	

Table 1: Foundational taxonomies for speech acts categorization

presents a few recent taxonomies used in the context of online conversation analysis.

## 2.2 Speech act classification applied to forum posts

Here we explain the work of [Kim et al., 2010].

### 2.2.1 Methodology

The authors annotate both threads and posts in a corpus built from the CNET forums<sup>2</sup>. Two annotators use a dedicated tool<sup>3</sup> for the annotation task. Before starting to annotate the overall data set, a pilot annotation was performed to ensure that both annotators had the same understanding of the taxonomy.

### 2.2.2 Taxonomy

First, they classify threads into two groups of classes: the basic group, which is further subdivided into *Solution Type* classes and *Problem Source* classes, as well as a miscellaneous group. The different classes are detailed in table 3. *Solution Type* and *Problem Source* classes can be used either separately or together, therefore there is an additional combined class set made of each combination of a *Solution Type* class and a *Problem Source* class (e.g.

<sup>2</sup><http://forums.cnet.com>

<sup>3</sup><http://mandrake.csse.unimelb.edu.au/rbp/cgi-bin/su/Classify.py>

Act	Corpus or kind of corpus	Reference
Disclosure Edification Advisement Confirmation Question Acknowledgment Interpretation Reflection	multi-domain	[Lampert et al., 2006]
Direct request Question-request Open question First person commitment First person expression of feeling First person other Other statements	corporate mail	[De Felice et al., 2013]
Accept response Acknowledge and appreciate Action motivator Polite mechanism Rhetorical question Open-ended question Or/or-clause question Wh-question Yes-no question Reject response Statement Uncertain response	BC3	[Ulrich et al., 2008]

Table 2: Examples of speech act taxonomies specific to online conversation analysis

Class	Group
Support	Solution Type
Documentation	Solution Type
Install	Solution Type
Search	Solution Type
Hardware	Problem Source
Operating System	Problem Source
Software	Problem Source
Media	Problem Source
Network	Problem Source
Programming	Problem Source
Spam	Miscellaneous
Other	Miscellaneous

Table 3: Thread classes in [Kim et al., 2010]

*Search-OS* or *Install-Network*). The combined class set being the superset of basic class sets, it was used to do the annotation.

Then, they annotate each individual post with one or several post classes. These classes are categorized into two groups, the answer group and the question group, plus three single classes. The different classes used are detailed in table 4.

### 2.2.3 Annotator agreement

The authors use Cohen’s Kappa to compute annotator agreement:

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)}$$

Where  $P(a)$  refers to the relative observed agreement between two annotators, and  $P(e)$  is the hypothetical probability of chance agreement between two annotators.

However, as they write: ” *The standard Cohen’s Kappa cannot be used in multi-class annotation tasks. By the time the annotation process started, no standard methodology for calculating Cohen’s Kappa for multi-class tasks was found. Therefore, an extended method for calculating  $\kappa$  value was proposed to address this problem.*”

They propose an improved Cohen’s Kappa for a multi-class situation, but the methodology is not applicable to our case because it takes into

Class	Group
Answer-Answer	Answer
Answer-Add	Answer
Answer-Correction	Answer
Answer-Confirmation	Answer
Answer-Objection	Answer
Question-Question	Question
Question-Add	Question
Question-Correction	Question
Question-Confirmation	Question
Resolution	-
Reproduction	-
Other	-

Table 4: Post classes in [Kim et al., 2010]

account the link information for each class (a concept we do not consider at all).

#### 2.2.4 Classification

The authors use four different models for thread classification: 1 Nearest Neighbor, Naïve Bayes and Support Vector Machine. A majority class model (ZeroR) and a random model (RAND) are used as baselines. Features are weighed using four different schemes: Information Gain (IG), Term Frequency-Inverse Document Frequency (TF-IDF), raw count and simple binary. For post classification, the authors use a conventional Maximum Entropy learner<sup>4</sup> as well as two structural learners : SVM-HMMs and CRFs (using *CRF++*<sup>5</sup>)

Features for threads are built by using the initial post or concatenating all posts in the thread to form the "text" of the thread. They then construct different bag-of-words feature sets by tuning different parameters such as input text, pre-processing, tokenization method, and  $n$ -gram length. For posts, features used include lexical, structural, contextual, and semantic features.

<sup>4</sup><http://maxent.sourceforge.net/>

<sup>5</sup><http://crfpp.sourceforge.net/>

Class	Definition
Expository Sentence	Presents or explain information to the reader
Commissive	Contains a stated commitment from the author
Directive	Contains an expectation that readers will do something as a response
Expressive	Contains a statement of the author’s psychological state
Representative	Commits the author to the truth of a certain proposition

Table 5: Taxonomy used in [Qadir and Riloff, 2011]

### 2.2.5 Evaluation

Results for both thread classification and post classification are evaluated using micro-average precision ( $P_\mu$ ), recall ( $R_\mu$ ), F-score ( $F_\mu$ ) and macro-averaged precision ( $P_M$ ), recall ( $R_M$ ) and F1-score ( $F_M$ ) on a stratified<sup>6</sup> 10-fold cross-validation.

## 2.3 Sentence classification in message board posts

Here we explain the work of [Qadir and Riloff, 2011].

### 2.3.1 Goals

The authors work with message board posts. They consider that not all sentences carry a speech act. Their primary goal is to distinguish between those that do and those that don’t (which they call *expository sentences*). Their second goal is to classify speech act sentences into four types: *commissives*, *directives*, *expressives* and *representatives*. These four classes are derived from the work of [Searle, 1976]. Searle’s fifth type of speech acts, *declarations*, was omitted because the authors found almost no examples of it in their corpus. This taxonomy is further explained in Table 5.

Their ultimate objective is to develop a system capable of classifying sentences in a topic-independent manner (the data they worked with was extracted from forums dealing with veterinary medicine).

<sup>6</sup>The authors make sure that “if a given post is contained in the test data for a given iteration, all other posts in that same thread are also in the test data (or more pertinently, not in the training data)”



### 2.3.2 Message board specificities

According to the authors, speech acts occur differently in message board posts than in spoken dialog.

They found that while the most common everyday occurrences of *commissives* are promises and threats, it is not so in forum data. Most occurrences of this type of speech act correspond to confirmations that the poster will perform some action in the future. The authors also consider that declarations from the poster that they will not do something are sentences carrying such a speech act.

They found that *directive* speech acts are common in message board posts, in particular under the form of a question or a request for assistance or advice. They attempt to weed out rhetorical questions from sentences classified as *directives*.

The authors found that typical examples of *expressive* speech acts occur when a poster thanks, welcomes or apologizes to a reader. They found this class to be very common in message board posts.

*Representative* speech acts were uncommon in the corpus they used. They chose to ignore most sentences stated as fact even though they might correspond to Searle's definition of a representative speech act and label them as *expository sentences* instead. They considered that sentences belonged in that class only when a doctor explicitly presented an opinion or hypothesis as their own (such as sentences starting with "I suspect that...").

### 2.3.3 Classification

The authors chose to approach the problem as a supervised classification task. Instead of a single classifier, they chose to train four different binary models, one for each class, and run them against the data. Therefore a single sentence can contain several speech acts. A preliminary classifier was trained just to filter out sentences containing no speech act (*expository sentences*).

They use a number of different feature sets for the classification task. First, they consider lexical and syntactic features. Such features include: unigrams, personal pronouns, tense, modals, plan phrases, punctuation, sentence position, number of verbs, and others. Second, they consider speech act word clues. In [Searle, 1976], the author listed a number of words that he considered to be indicative of speech acts. Here, the authors chose to discard a few that they considered too general and added new ones, such as a list of speech act verbs originally published by [Wierzbicka, 1987]. Their system recognizes all derivations of these word clues.

Class	Proportion
None	71.42%
Directive	15.90%
Expressive	9.92%
Representative	2.91%
Commissive	2.61%

Table 6: Repartition of speech acts in [Qadir and Riloff, 2011] test data

The classifiers used were Support Vector Machines (SVMs).

### 2.3.4 Evaluation

The data set used was constituted of 15,383 message board posts from the Veterinary Information Network (VIN)<sup>7</sup>. These posts covered three topics: cardiology, endocrinology, and feline internal medicine. The preprocessing included number tokenization, the removing of html tags and "basic cleaning" (???). Experiments were performed on a random sample of 150 threads from the collection. These threads contained 1,956 sentences (13.04 sentences per post).

The gold standard was built from the manual annotation of message posts by two human annotators provided with detailed annotation guidelines. A sample of 50 posts was independently annotated by them in order to calculate the annotator agreement. However, the authors ran into the same problem [Kim et al., 2010] did, that is the fact that the standard Cohen’s Kappa is not applicable to multi-class annotation tasks. Since they found only a small number of sentences classified as containing several speech acts, these were discarded and the kappa was calculated for the rest. The result was a score of .95, an extremely high agreement. However the "no speech act" class was by far the most predominant (70%), so the kappa score does not necessarily mean that annotators mostly agreed on speech act classification decisions. The authors therefore computed the kappa for each speech act class and found an average score of 87.25%, which is still very high. Table 6 shows the repartition of speech acts in their dataset.

The first classifier, the one filtering out sentences containing no speech act, recognized 83% of the speech act sentences with 86% precision, and 95% of the expository sentences with 93% precision. As for the speech

<sup>7</sup>[www.vin.com](http://www.vin.com)

act categorization task, the authors performed many experiments with a number of different feature combinations, for each speech act category. The authors found that many syntactic and lexical features had little impact on the classification results. Detailed results will not be presented here, but overall results were good for the identification of directive and expressive speech act sentences, less so for representative and commissive speech acts, which proved more difficult to detect.

## 2.4 Segmentation of email message text

Here we explain the work of [Lampert et al., 2009].

### 2.4.1 Goals

The authors attempt to segment emails into several prototypical areas, called *email zones*. They call their system "Zebra".

### 2.4.2 Email zones

There are three superclasses containing nine different classes of email zones.

*Sender zones* contain text written by the current sender. This superclass contains the following subclasses: *author* (new content by the sender), *greeting* (polite greetings at the beginning of a message), and *signoff* (closing words of a message).

*Quoted conversation zones* include content quoted from other messages. The subclasses for this superclass are: *reply* (content quoted from a previous message in the thread) and *forward* (content from an email outside the thread that has been forwarded by the current sender).

*Boilerplate zones* contain reusable message content, typically automatically appended to a message. Subclasses include: *signature* (contact information automatically appended at the end of a message), *advertising* (advertising material, typically found at the end of the message), *disclaimer* (legal disclaimers and privacy statements), and *attachment* (automated text indicating attached documents).

### 2.4.3 Classification

The authors consider that each line of text in the body of a message belongs to one of these *email zones*. The authors tried two approaches: in the first one, a classifier segments a message into zone fragments and then attempts

to classify the resulting fragments. In the second approach, the classifier simply works on a line-by-line basis.

In the first approach, the fragment-based one, the message is segmented based on detected *zone boundaries*. These boundaries are identified thanks to what the authors call "buffer lines", i.e. blank lines or lines containing only whitespace or punctuation characters.

The classifier is based on SVMs (Support Vector Machines) and uses graphic, orthographic and lexical features to categorize lines and text fragments.

#### 2.4.4 Evaluation

The training data for their classifier consists of almost 400 messages from the Enron email corpus [Klimt and Yang, 2004]. To build the gold standard, the authors chose to use only a single annotator since they believe the task to be relatively uncontroversial. Each line - blank lines excepted - was marked by the annotator as belonging to one of the nine zones. They used the resulting 7,922 annotated lines (out of 11,881) as training data for their classifier.

The results are calculated by 10-fold cross-validation.

The zone boundary detection system (used for the fragment based approach) reaches 90% accuracy.

Concerning zone classification, interestingly, the line-based approach outperformed the fragment-based one by a small margin, probably due to the 10% of inaccuracies in boundary detection. Zebra obtains a 91.53% accuracy for the classification of lines into the three superclasses, and 87.01% in all nine classes.

## 3 Proposal

Here we describe our findings and proposal.

### 3.1 Classes

Like [Qadir and Riloff, 2011], we consider the following two top-level classes of text units: expository text (**expository**) and speech act carrying text (**speech act**). The former present factual information while the latter contain conversational units bearing communicative acts between the writer and its readers. We add a third superclass: structural information (**structure**), which will represent all the information - redacted or automatic - included to facilitate the understanding of the message or the conversation. These

three classes correspond to a message’s body, its important content. We sort other content - such as quoted text, automatic signatures, privacy statements etc. - in two other classes: quoted text (**quote**) and boilerplate content (**boilerplate**). These two categories’ content’s definitions are based on the work of [Lampert et al., 2009].

The **speech act** superclass contains three main subclasses: **commissives**, **directives** and **expressives**. They are derived from the five classes defined in [Searle, 1976] (the fourth class, "representatives", was removed because making the distinction between sentences expressing the belief or a judgment of truth and expository sentences was not very useful to us, and the fifth class, "declarations", was omitted because we assume it will almost never appear in our corpus: [Qadir and Riloff, 2011] found almost no example of it in their datasets).

All these classes themselves contain a multitude of labels, all of which will be presented details in the following subsections of this document. We postulate that a sentence can carry several labels.

The following labels have been empirically extracted from our corpus. This was done by two researchers through the analysis of individual messages, on a sentence by sentence basis.

### 3.1.1 Boilerplate

Boilerplate text is non-content text found around the message’s body, typically reused without modification across multiple messages.

- Signature: the message author’s identity
  - Automatic signature: signature automatically appended to the message (e.g. *"Brian Lunergan Nepean, Ontario Canada"*)
  - Manual: signature written by the author for this specific message (e.g. *"Sean"*)
- Advertisement: advertising material automatically appended to the message (e.g. *"Novo Yahoo!"*)
- Disclaimer: legal disclaimers and privacy statements automatically appended to the message
- Contact: the author’s contact information (e.g. *"mns:renato4010591@hotmail.com"*)

### 3.1.2 Quote

Quoted text include both content quoted in reply to previous messages in the same conversation thread and forwarded content from other conversations.

- Repled text: content quoted from a previous message in the thread
- Forwarded text: content from a message outside the current conversation that has been forwarded by the current message (email only)
- Quote: quote from an outside source, such as a monologue document, a book, or a retranscription of famous spoken words

### 3.1.3 Expository

- Subjective report: account of something the author has observed, heard, done, or investigated
  - Action report: reporting of actions taken (e.g. *"I did try apt-cache search geotiff but it didn't work"*, *"I upgraded my system to 10.04 via clean install"*)
    - \* Solution research report: specific type of action report where the author took action to find a solution by himself (e.g. *"I checked the forums already"*)
  - Result report: reporting of the results of actions taken (e.g. *"I did try apt-cache search geotiff but it didn't work"*, *"No such luck"*)
    - \* Solution research result report: specific type of result report pertaining to the search for a solution (e.g. *"I couldn't find anything on the forums"*)
  - Observation: a remark, statement, or comment based on something one noticed ("It doesn't happen every time I move the cursor, it is completely random")
- Objective report: exact reporting of something without author interpretation
  - Computer text: copy-and-paste of code, log, or commands (e.g. *"usermod - G admin 2ndroot"*)
  - Quote: exact copy of a portion of text with an indication that one is not the original author

- Statement: definite and clear expression of the nature or truth of something
  - Problem statement: description of the issues that need to be addressed in order to consider the problem at hand resolved (e.g. *"The problem is (If its not clear from the picture), that the black area should be brown, or transparent or whatever as the background, but you probably guessed it."*)
  - Goal statement: statement that describes future state of affairs and provides general direction, purpose or intent of what needs to be accomplished (e.g. *"I would like to be able to send the whole package via email to acquaintances in ISO form and let them make their own DVDs"*)
- Description: objective account of a system, person or event
  - Event description: description of an event
  - System description: retranscription of a system's settings or specifications (e.g. *"My ubuntu 1 bandwidth settings are set to -1", "I am using ubuntu 12.04"*)
  - Author profile: description of the author's identity, habits, experiences, preferences or skills (e.g. *"Coming from a Winows world, I've probably been spoiled but..."*)
- Assimilation: expression of a belief in the similarity or non-similarity of two things
  - Problem assimilation: assimilation of a given problem to another (e.g. *"I had a similar problem"*)
- Assertion: a confident and forceful statement of fact (e.g. *"DD will work over the network", "Even OS X has nothing close to it, unfortunately"*)
- Guess: estimate or supposition without sufficient information to be sure of being correct (e.g. *" "*).
- Correction: rectification of an alleged error or inaccuracy

### 3.1.4 Speech

- Commissives: sentences containing a stated commitment from the author
  - Acknowledgment-Commitment: a commitment to do something in reaction to a previous message in the thread (e.g. *"Ok thanks I'll try that as soon as I get home"*, *"Hopefully it's explained in detail in there, I will search..."*)
  - Channel change: announcement that the author is going to switch or fork the discussion to a different communication channel (e.g. *"I'll file a wishlist bug for this"*)
- Directives: sentences containing an expectation that readers will do something as a response
  - Question: a sentence worded or expressed so as to explicitly elicit information (e.g. *"Should I just change them by hand?"*, *"Is there anyway to recover these short of recreating everything from scratch and restoring from backup?"*)
  - Clarification request: request for more specific information or confirmation that the author has correctly understood an utterance (e.g. *"do you mean the starterbar from gdesklets?"*, *"Which exact packages are you getting an error from?"*)
  - Request for assistance: call for help (e.g. *"please help!"*, *"can anyone help me???"*, *"i could use some help"*)
  - Suggestion: an idea submitted for consideration
  - Instruction: information telling how something should be done (e.g. *"Open a terminal"*)
- Expressives: sentences containing a statement of the author's psychological state
  - Greeting: polite words of salutation
    - \* General greeting: greeting directed to no one in particular (e.g. *"Hi all"*, *"Hello all"*)
    - \* Targeted greeting: greeting directed at a specific participant or group of participants (e.g. *"Hey Steve"*)
  - Signoff: the conclusion of a message (e.g. *"see you"*, *"-best -greg"*)



- Thanks: an expression of gratitude
  - \* Thanks in advance: an expression of anticipated gratitude (e.g. *"Thanks for any help,"*)
  - \* Thanks in reaction: an expression of gratitude for some specific assistance (e.g. *"Thanks I'll try that!"*)
- Opinion: expression of a subjective judgment on a thing or situation (e.g. *"what a nightmare"*, *"this is bullshit"*, *"this OS sucks"*)
  - \* Feedback: specific type of constructive opinion destined to be used as a basis for the improvement of a product or service
- False question: question whose function is not to elicit an answer
  - \* Rhetorical question: question asked in order to make a point (e.g. *"Are you seriously implying that he should reinstall Ubuntu because of a cosmetic issue?"*, *"Wow, CRTs to store data?"*)
  - \* Personal interrogation: question asked in order to express a sentiment (e.g. *"I wonder how it works?"*, *"Come on, how hard can this be?"*)
- Expression of solution confidence: expression of the author's level of confidence in a proposed solution (e.g. *"I don't know much about it, I just discovered it on gnomefiles.org"*, *"However, it may do what you need"*)
- Expression of solution preference: expression of the author's preferences (to varying degree) towards a specific type of solution

### 3.1.5 Structure

Structural informations are contained in sentences whose purpose is to explain what a message or sentence is and what is its role in the conversation.

- Metadiscourse: sentences that comments on surrounding text
  - Utterance introduction: sentences that introduces an upcoming sentence
    - \* Question introduction: sentence that introduces a question (e.g. *"the question f this mail is..."*)
    - \* Problem introduction: sentence that introduces a problem description (e.g. *"here's my problem"*)

- \* Instructions introduction: sentence that introduces a series of instructions (e.g. *"do this:", "Then release them and press the F2 key."*)
  - \* Quote introduction: sentence that introduces a quote - may be automatic (e.g. *"On 20/02/07, Larry wrote:"*)
  - Utterance qualifier: sentences that qualify a previous sentence (e.g. *"just a thought"*)
- Structure marks: non-sentences bearing structural information
  - Non-verbal separator: non-linguistic character sequence separating two parts of a message (e.g. *"–"*)
  - Heading: title announcing a new section of the message (e.g. *\*\*\*\*\*DISCLAIMER\*\*\*\*\**)
- Reference: sentences that mention or allude to something else
  - post reference: the author references to another post in the thread (e.g. *"Earlier Dago said that..."*)
  - item reference: the author references to a previously introduced item (e.g. *"what does that command do exactly?"*)
  - channel reference: the author references to another communication channel (e.g. *"Check the forums, we just discussed this subject and procedures for both dd and rsync"*)
- Conversational information: information on the conversation's flow
  - Answer acknowledgment: the author confirms the good reception and understanding of an answer (e.g. *"OK."*)
    - \* Answer confirmation: the sentence confirms that a proposed solution worked (e.g. *"Perfect!!"*)
    - \* Answer rejection: the sentence rejects an answer because the solution is inapplicable or didn't work (e.g. *"Thanks. But the CD still doesn't work"*)
  - Thread closure: the sentence announces that the conversation is over or that the problem was resolved or cannot be resolved
- Edition artifact: formal element showing the message has been modified, by the author or a moderator for example (e.g. *"[snip]"*)
- Hyperlink: hypertext link pointing toward another document (e.g. *"https://lists.ubuntu.com/mailman/listinfo/ubuntu-users"*)

## References

- [Austin, 1975] Austin, J. L. (1975). *How to do things with words*, volume 1955. Oxford university press.
- [De Felice et al., 2013] De Felice, R., Darby, J., Fisher, A., and Peplow, D. (2013). A classification scheme for annotating speech acts in a business email corpus. *ICAME Journal*, 37:71–105.
- [Hernandez and Salim, 2014] Hernandez, N. and Salim, S. (2014). Exploiting the human computational effort dedicated to message reply formatting for training discursive email segmenters. *LAW VIII*, page 110.
- [Kim et al., 2010] Kim, S. N., Wang, L., and Baldwin, T. (2010). Tagging and linking web forum posts. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL ’10, pages 192–202, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [Klimt and Yang, 2004] Klimt, B. and Yang, Y. (2004). The enron corpus: A new dataset for email classification research. In Boulicaut, J.-F., Esposito, F., Giannotti, F., and Pedreschi, D., editors, *ECML*, volume 3201 of *Lecture Notes in Computer Science*, pages 217–226. Springer.
- [Lampert et al., 2006] Lampert, A., Dale, R., and Paris, C. (2006). Classifying speech acts using verbal response modes.
- [Lampert et al., 2009] Lampert, A., Dale, R., and Paris, C. (2009). Segmenting email message text into zones. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, pages 919–928. Association for Computational Linguistics.
- [Qadir and Riloff, 2011] Qadir, A. and Riloff, E. (2011). Classifying sentences as speech acts in message board posts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 748–758. Association for Computational Linguistics.
- [Searle, 1976] Searle, J. R. (1976). *A taxonomy of illocutionary acts*. Linguistic Agency University of Trier.
- [Traum, 2000] Traum, D. R. (2000). 20 questions on dialogue act taxonomies. *Journal of semantics*, 17(1):7–30.

- [Ulrich et al., 2008] Ulrich, J., Murray, G., and Carenini, G. (2008). A publicly available annotated corpus for supervised email summarization. In *AAAI08 EMAIL Workshop*, Chicago, USA. AAAI.
- [Wierzbicka, 1987] Wierzbicka, A. (1987). *English speech act verbs: A semantic dictionary*. Academic Press Sydney.