NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

IMPLEMENTATION AND PERFORMANCE EXPLORATION OF A CROSS-GENRE PART OF SPEECH TAGGING METHODOLOGY TO DETERMINE DIALOG ACT TAGS IN THE CHAT DOMAIN

by

J.R. Hitt

September 2010

Thesis Advisor:	Craig H. Martell
Second Reader:	Joel D. Young

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
22–9–2010	Master's Thesis	2008-09-01—2010-09-30

4. TITLE AND SUBTITLE

Implementation and Performance Exploration of a Cross-Genre Part of Speech Tagging Methodology to Determine Dialog Act Tags in the Chat Domain

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

J.R. Hitt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Department of the Navy

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. **IRB Protocol Number: n/a**

14. ABSTRACT

Internet Relay Chat is a popular means of communication. Because chat data does not follow established grammatical rules, traditional machine learning algorithms perform poorly in tasks such as part-of-speech and dialog-act tagging, and yet the volume of data created makes human analysis impractical. We present a cross-genre part-of-speech tagging methodology and analyze its effectiveness in determining the dialog-act classes of chat posts. Previous methods for determining part-of-speech tags focused on accuracy, were computationally expensive and required human verification. We show that our cross-genre maximum likelihood estimation part-of-speech tagging performs virtually identically to hand-tagged parts-of-speech and that accurate part-of-speech tags are not required for acceptable automatic dialog-act determination. Furthermore, we show that a simple naïve Bayes classifier achieves the same performance in a fraction of the time as a carefully trained neural network.

15. SUBJECT TERMS

Part of Speech Tagging, Chat Dialog Act Tagging, NPS Chat Corpus, Naïve Bayes Classifier, Internet Relay Chat, Tactical Military Chat, Cross-Genre POS Tagging, Cheap POS, Maximum Likelihood Estimation POS Tagging

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	UU	129	

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

IMPLEMENTATION AND PERFORMANCE EXPLORATION OF A CROSS-GENRE PART OF SPEECH TAGGING METHODOLOGY TO DETERMINE DIALOG ACT TAGS IN THE CHAT DOMAIN

J.R. Hitt
Captain, United States Navy
B.S., Santa Clara University, 1987

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
September 2010

Author: J.R. Hitt

Approved by: Craig H. Martell
Thesis Advisor

Joel D. Young
Second Reader

Peter J. Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Internet Relay Chat is a popular means of communication. Because chat data does not follow established grammatical rules, traditional machine learning algorithms perform poorly in tasks such as part-of-speech and dialog-act tagging, and yet the volume of data created makes human analysis impractical. We present a cross-genre part-of-speech tagging methodology and analyze its effectiveness in determining the dialog-act classes of chat posts. Previous methods for determining part-of-speech tags focused on accuracy, were computationally expensive and required human verification. We show that our cross-genre maximum likelihood estimation part-of-speech tagging performs virtually identically to hand-tagged parts-of-speech and that accurate part-of-speech tags are not required for acceptable automatic dialog-act determination. Furthermore, we show that a simple naïve Bayes classifier achieves the same performance in a fraction of the time as a carefully trained neural network.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgements

I dedicate this work to the memory of my father, Chuck Hitt, whose example I will forever strive to emulate. He devoted his life to serving others. Though he always remained humble about it, at the age of 17, he volunteered for the U.S. Navy during World War II. Dad then positively impacted the lives of innumerable young students as a teacher and principal while raising four children to become citizens of this great nation. I love you and miss you, Dad.

My mother, Claire, who has always been there for her children. Whether tending to our scrapes and bruises or cheering us on during innumerable concerts or sporting events, you were always there with loving words of encouragement. I love you, Mom, for all you have done to inspire me.

To Stephanie, the love of my life. I couldn't ask for a better teammate during this career. You have sacrificed much at every duty station, including NPS. Thank you for always putting a smile on...I could not have done this without you. I love you always!

To my daughter, Taylor: you are truly special, exeptionally intelligent and beautiful. I enjoyed the times discussing homework. I am very proud of the woman you have become. I love you, Princess!

Professor Martell, I will always be grateful for the patience and guidance you have provided over the last two years. I was initially impressed with your teaching style and benefitted tremendously from your persistent encouragement. Congratulations on tenure and thank you for letting me run with this idea.

To Lieutenant Colonel Young: You are the consummate instructor and mentor. The professionalism you display in and out of the classroom is, simply put, outstanding. Your meticulous attention to detail is a trait I wish I possessed. Thank you for all of your efforts to keep me on track through this challenge!

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1: INTRODUCTION

1.1 The Chat Domain

Since its introduction in the late 1980s, Internet Relay Chat (IRC) has become popular world-wide as a means of real-time communications. With hundreds of thousands of users each day, the volume of data created is overwhelming for complete human analysis. Natural Language Processing (NLP) techniques can be applied for applications such as social networking analysis, data-mining and detection of illicit uses.

1.1.1 General Chat Characteristics

IRC chat "rooms" are hosted on servers around the world. Some of these rooms are devoted to specific topics, while others are simply gathering places for social interaction. Users log in to rooms of their choosing and select an alias for self-identification. They are then free to type their inputs, which are then broadcasted to all participants in the room. There are also functions available that allow users to hold "private" conversations with other selected users. In these private rooms, only those users invited to participate see others' posts.

Due to IRC's synchronous nature, users may provide their inputs at any time. There is no requirement for turn-taking as commonly found in spoken dialog. Hence, IRC data streams frequently consist of multiple, interleaved conversations further complicating analysis. For example, when a user presents a question it is available to all users present in the respective chat room. The next item appearing in the stream may not be a response to this question and may, in fact, be another, unrelated question by a different participant. Correlating questions and subsequent answers becomes a difficult task, particularly in active chat rooms with many participants. The problem of identifying who said what to whom is called *conversational thread extraction*. A good source for understanding this problem and other chat specific issues is Adams [1].

Because chat users are not generally constrained by strict language semantics or structure, the task of identifying questions amongst other types of posts is also difficult. While traditional written language contains punctuation (question marks) that identify illocutionary (or dialog) acts as questions, these clues are frequently missing in chat messages. Identifying questions, as

opposed to other dialog acts, is therefore a difficult task. The ad-hoc nature of chat usage also results in unique features, including symbols intended to convey emotions ("emoticons") and intentionally misspelled words, not typically found in traditional language usage. As a result, parsing algorithms that are trained on structured language examples perform poorly in the chat domain.

Previous work in the chat domain has focused on part of speech and dialog act tagging as a foundation for higher-level analysis. These tasks include conversational thread extraction, data-mining and social-networking analysis. Due to the aforementioned structural differences between chat data and oral or written data, these tasks are not easily automated and human interaction is frequently required. The volume of data created by heavily populated chat servers, however, makes such human involvement infeasible. Development of NLP techniques to assist in these tasks, in this particular domain, is therefore desirable.

While this thesis is focused on IRC data, the techniques apply to any chat system such as Yahoo, AOL Instant Messenger or even military applications such as tactical chat.

Chat in the Military Domain

Just as chat is a popular form of communication for the general public, tactical military chat has become an important command and control tool for forces operating around the world [2]. The topics discussed in these chat sessions are more focused toward tactical situations and are structured with user names derived from assigned user duties. This additional structure may provide information useful for higher level analysis such as post-event reconstruction. The information derived from this data may then be used to document lessons learned for follow-on tactical performance improvements.

Eovito provided functional requirements for tactical military chat. Eovito's work included items that we believe would benefit from inclusion of dialog act information such as "Thread Population/Repopulation," "Suppress System Event Messages," and "User Access to Chat Logs."[2] Consider the possibility of a chat participant being able to determine who has asked what questions and what answers were provided without interrupting other users. These functions may serve to filter undesired noise from the conversation thereby increasing the rate of acquiring situational awareness. We believe that such an enhanced filter may benefit from automatically produced dialog act information.

1.2 Purpose of this Thesis

This thesis provides an improved method for dialog act tagging chat posts. We show that the use of maximum likelihood estimation part of speech tags nearly equal the performance of computationally expensive, human verified parts of speech in determining dialog act tags in the chat domain. More importantly, our methodology demonstrates that maximum likelihood estimation part of speech tags from a fundamentally different, labeled domain work very well in the chat domain. This is very important, not just for analysis of chat, as it bodes well for new domains of Internet communications as they are invented, deployed and developed.

In fact, this thesis represents new work in the important field of cross-genre machine learning. We show that previous, human-involved investments in another genre can be effectively applied to produce acceptable results in the chat domain. Our work should serve as a foundation for other research in the rapidly expanding field of computer communications.

1.3 Organization of Thesis

This thesis is organized as follows:

- Chapter 1 discusses computer-mediated communications and motivation for this thesis. We include a brief overview of the chat domain and specific challenges to analysis of the data found there.

- Chapter 2 contains background information on Internet Relay Chat, previous research into chat analysis and the machine learning techniques used in this work.

- Chapter 3 includes our experimental approach to dialog act tagging chat posts. This chapter includes discussions on the sources of data for our part of speech tagger, training and test data. We describe a cross-genre methodology (one that uses data derived from a different domain) that effectively determines dialog act tags in the chat domain. Also included are specific details about feature selection and our experimental approach.

- Chapter 4 provides the results of our work in dialog act tagging chat posts. We also provide statistical significance test results for our data. Additionally, we include results of experiments to that our results are not skewed by individual author contributions to the chat data.

- In Chapter 5, we summarize our results and provide recommendations for future work in improving machine learning approaches to determine dialog act tags in chat posts.

CHAPTER 2:
BACKGROUND

2.1 Online Chat

The proliferation of computers and increased Internet availability have produced new means for connecting socially and professionally. Some of these new forms of information exchange, in which users pass typed messages to one or more other users, are referred to as computer-mediated communications [3]. Internet Relay Chat (IRC) is one popular form of computer-mediated communication.

Chat "rooms" provide a stage upon which users can express thoughts via typewritten messages called "chat posts" or, simply, "posts." These posts are broadcast to all subscribers logged into the respective chat room. Posts may be composed at any time and are broadcast in the order they are received, interlacing conversations between distinct users and general announcements meant for all participants.

Previous works by Herring and Kucukyilmaz noted that the structure of chat posts differs from that of written text and also from that of spoken language [3, 4]. Examples of specific differences include the use of emoticons (see Appendix A), flexible grammatical rules including punctuation and spelling, and the intentional use of misspelled words to convey emotion or emphasis. These differences present unique challenges when analyzing higher-order characteristics of chat posts such as classification of dialog act and semantic meaning.

2.2 Prior and Related Work

In 2006, Lin collected and preserved over 477,000 chat posts from an Internet chat site. The source material was saved from chat rooms that were organized by user age groups, and this organization was maintained. These chat rooms were not limited to particular topics [5]. The goal of Lin's work was an attempt to identify any sexual predators actively participating in these chat rooms.

Forsyth followed Lin with a primary goal of using machine learning algorithms to apply part-of-speech tags to chat posts, and secondary goal of exploring potential techniques for automatic dialog act tagging of chat posts. In the course of his work, Forsyth removed all personally iden-

tifiable information from 10,567 chat posts sampled from different chat rooms. This privatized subset of Lin's work has become known as the Naval Postgraduate School (NPS) chat corpus. Forsyth tagged the NPS chat corpus with parts-of-speech and dialog act tags using a bootstrapping method followed by verification by humans [6]. For part-of-speech tagging, he used the

CC	Coordinating conjunction	PRP$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential there	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	to
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
POS	Possessive ending	WP$	Possessive wh-pronoun
PRP	Personal pronoun	WRB	Wh-adverb

Table 2.1: Penn Treebank Tagset. From [7].

Penn Treebank POS tagset (see Table 2.1), and dialog act tagged the NPS chat corpus using Wu et al.'s 15 post act categories (see Table 2.2). Forsyth compared the performance of taggers based on n-grams, hidden Markov models (both discussed in the next section) and Brill taggers [6]. Using his implementation of a Brill tagger trained on the NPS Chat, Wall Street Journal, and Switchboard corpora, Forsyth achieved a 90.8% POS tagging accuracy. In his dialog act tagging effort, Forsyth developed 27 features including lexical and temporal characteristics of chat posts and the number of chat users participating in the chat room of interest (see Table 2.3). He compared the performance of naïve Bayes (discussed in the next section) and back-propagation neural networks in dialog act tagging accuracy. Forsyth recorded an 83.2% dialog act tagging accuracy using a back-propagation neural network with 23 of these features [6].

Tag	Example	Count	Percent
Statement	I'll check after class	3185	30.14%
System	Tom[JADV 11.22.33.44] has left#sacbal	2632	24.91%
Greet	Hi, Tom	1363	12.90%
Emotion	lol	1106	10.47%
Yes-No-Question	Are you still there?	550	5.20%
Wh-Question	Where are you?	533	5.04%
Accept	I agree	233	2.20%
Bye	See you later	195	1.85%
Emphasis	I do believe he is right.	190	1.80%
Continuer	And	168	1.59%
Reject	I don't think so.	159	1.50%
Yes-Answer	Yes, I am.	108	1.02%
No-Answer	No, I'm not.	72	0.68%
Clarify	Wrong spelling	38	0.36%
Other	********	35	0.33%

Table 2.2: 15 Post Act Classification for Chat. After [8]. - Statistics from NPS Chat Corpus

2.3 Machine Learning Techniques

When we use computers to analyze data derived from experience and use this information to predict (in our case, to classify) new data, we are performing machine learning [9]. The volume of Internet traffic, specifically IRC data, necessitates use of computers for any meaningful attempt at analysis of the information being transmitted. Because IRC data is a form of written communication using human language, the analysis of chat data generalizes to a form of natural language processing (NLP). One general goal of NLP is that of classification, where we attempt to determine some higher-level grouping of data. Examples of this effort include dialog act tagging, authorship detection and topic detection.

In the use of computers to process this type of information, we identify features (e.g., words, parts-of-speech, semantic or syntactic structure) from which to draw and test hypotheses.

2.3.1 Features

The basis for classification of text data must be some set of features whose analysis sufficiently identifies a particular example's class as opposed to non-classes. One common approach to feature selection in natural language processing is to use the lexical items, sentences, phrases or words, in documents of interest. The basis for probabilistic methods used in NLP involves

Feature	Definition	Rationale
f0	Number of posts ago the poster last posted	Indicator for a Continuer act
f1	Number of posts ago the poster made a spelling error	Indicator for a Clarify act
f2	Number of posts ago that a post contained a '?' but no WRB or WP POS tag	Indicator for a Yes / No Answer act
f3	Number of posts in the future that contained a Yes of No word	Indicator for a Yes / No Question act
f4	Number of posts ago that contained a Greet word	Indicator for a Greet act
f5	Number of posts in the future that contained a Greet word	Indicator for a Greet act
f6	Number of posts ago that contained a Bye word	Indicator for a Bye act
f7	Number of posts in the future that contained a Bye word	Indicator for a Bye act
f8	Number of posts ago that a post was a JOIN	Indicator for a Greet act
f9	Number of posts in the future that a post is PART	Indicator for a Bye act
f10	Total number of words in post	Longer posts may be Statements and Questions, shorter posts may be Emotions and Greets/Byes, etc.
f11	First word is a conjunction, preposition, or ellipses (POS tag of 'CC,' 'IN,' or ':')	Indicator for a Continuer act
f12	A word contains emotion variants such as lol, ;-), etc.	Indicator for an Emotion act
f13	A word contains hello or variants	Indicator for a Greet act
f14	A word contains goodbye or variants	Indicator for a Bye act
f15	A word contains yes or variants	Indicator for Yes or Accept acts
f16	A word contains no or variants	Indicator for No or Reject acts
f17	A word POS tag is WRB or WP	Indicator for a Wh-Question act
f18	A word contains one or more '?'	Indicator for Wh- or Yes/No Question acts
f19	A word contains one or more '!' (but not a '?')	Indicator for an Emphasis act
f20	A word POS tag is 'X'	Indicator for an Other act
f21	A word is a system command (. or ! With SYM POS tag)	Indicator for a System act
f22	A word is a system word, e.g. JOIN, MODE, ACTION, etc.	Indicator for a System act
f23	A word is an 'any' variant, e.g. 'anyone,' 'n e,' etc.	Indicator for a Yes/No Question act
f24	A word is in all caps, but not a system word like JOIN	Indicator for an Emphasis act
f25	A word is an 'even' or 'mean' variant	Indicator for a Clarify act
f26	Total number of users currently in the chat room	More users may stretch out distances between adjacency pairs

Table 2.3: Initial Post Feature Set (27 Features). From [6].

counting the number of occurrences of selected features in their respective classes.

For example, given a chat post $D =$ "he bought the purple dog," we could compute the probability of D as one item:

$$P(D) = \frac{\text{number of occurrences of } D}{\text{total number of posts in corpus}}$$

or, if we consider each word as a random variable, we could simplify this task by estimating the

probability of D using the chain rule for joint probability:

$$\begin{aligned} &P(\text{he bought the purple dog}) = \\ &\quad P(\text{dog}|\text{he bought the purple}) \times P(\text{purple}|\text{he bought the}) \\ &\quad \times P(\text{the}|\text{he bought}) \times P(\text{bought}|\text{he}) \times P(\text{he}|\text{start}) \times P(\text{start}) \end{aligned}$$

But this becomes cumbersome in that it would require us to maintain all the probabilities of all words given all observed previous words. We can simplify this further by making the assumption that the probability of each word is dependent only on a limited number of previous words. This is known as the Markov assumption and it is used frequently in NLP [10]. For example, if we estimate the probability of D based on only using one previous feature (or word in this example):

$$\begin{aligned} &P(\text{he bought the purple dog}) \approx \\ &\quad P(\text{dog}|\text{purple}) \times P(\text{purple}|\text{the}) \times P(\text{the}|\text{bought}) \\ &\quad \times P(\text{bought}|\text{he}) \times P(\text{he}|\text{start}) \times P(\text{start}) \end{aligned}$$

or, more generally:

$$P(f_1 f_2 ... f_n) \approx P(f_1) \prod_{k=1}^{n} P(f_k|f_{k-1}) \tag{2.1}$$

If we choose to estimate the probability of sentences based on zero previous words, we simply maintain the probability of each individual word and multiply using the chain rule. In this case, our features are called a "bag of words" since the order is not important.

We primarily use n-grams where $n \in \{1, 2, 3\}$ and indicates the number of individual data elements included in each feature. Throughout this document, 1-grams (or items in the aforementioned "bag of words") are referred to as unigrams, 2-grams as bigrams (these correspond to equation 2.1) and 3-grams as trigrams [10]. In addition to using n-grams made up of individual words, we examine the potential of classifying posts by dialog act using part-of-speech n-grams. For our experiments, we examined the use of 1,2 and 3-grams consisting of parts-of-speech tags and 1 and 2-grams of lexical items (the words themselves.)

Parts of Speech (POS)

Because of the aforementioned relaxation of spelling, grammar and punctuation rules in chat, automatic POS tagging in the chat domain has been the focus of other efforts [6, 11]. Traditional

POS taggers apply a variety of approaches to identify each word's part-of-speech as determined by the nature of the word and the context in which it is used. Many words in the English language are appropriately tagged with different parts of speech depending on how they are used. For example, "flies" may either be tagged as a plural noun if it is used to refer to common insects ("He swatted the flies.") or a present tense verb when describing what an airplane does ("An airplane flies.") This disambiguation is, in general, computationally expensive.

Forsyth part-of-speech tagged the anonymized portion of the NPS chat corpus using the Penn Treebank system of tags (see Table 2.1.) For his work, Forsyth used 27 features to compare performance of naïve Bayes, hidden Markov model, and Brill taggers in the determination of chat parts-of-speech classification.

Dialog Acts

Stolcke suggested that a useful, first level of detail in the analysis of discourse structure is dialog act identification [12]. For example, because of chat's aforementioned broadcast structure and interlaced conversations, dialog acts have been shown to provide some assistance in conversational thread extraction [11] or determination of conversation meaning [13].

2.3.2 Naïve Bayes Classifiers

Naïve Bayes classifiers are a form of supervised learning. This type of algorithm requires labeled data for training. Across all of the labeled classes, we can determine the probability of each dialog act by counting the example posts of each category and dividing by the total number of posts used for training:

$$P(C_j) = \frac{\text{number of training set examples of } C_j}{\text{total number of posts in training set}}$$

This value is referred to as the "prior" probability of the class C_j in the training set and it is an important part of our classifier as seen below.

We label the count of feature f_i as $count(f_i)$. Then the probability of f_i occurring in dialog act class C_j of words is $P(f_i) = \frac{count(f_i)}{\text{total number of words in } C_j}$. We use these counts in the form of a feature vector $\vec{F} = \{P(f_1), P(f_2), ..., P(f_n)\}$. Because we have computed these feature counts from each dialog act class, we condition the counts on the feature giving us $P(\vec{F}|C)$. However, our classification task requires us to compute $P(C|\vec{F})$. To do this we apply Bayes Rule, which

states:

$$P(C|\vec{F}) = \frac{P(C)P(\vec{F}|C)}{P(\vec{F})}$$

The task for our Naïve Bayes classifier is then to find the class C_j that maximizes $P(C|\vec{F})$. We call the class so identified by our classifier $\hat{C}$ where:

$$\hat{C} = \underset{C \in Classes}{\operatorname{argmax}} \frac{P(C)P(\vec{F}|C)}{P(\vec{F})}$$

Note that we compare the probability of a feature vector through all the classes. Thus the denominator, our feature vector, does not change between classes. Because the denominator behaves as a constant, and division by a constant does not change the relative results across classes, we can simplify the equation for our Naïve Bayes classifier as:

$$\hat{C} = \underset{C \in Classes}{\operatorname{argmax}} P(C)P(\vec{F}|C)$$

A critical assumption made in the use of Naïve Bayes classifiers is that each feature in the feature vector is independent of every other feature. This assumption means that:

$$P(\vec{F}|C) = \prod_{f_k \in \vec{F}} P(f_k|C)$$

and our final equation for the Naïve Bayes classifier becomes:

$$\hat{C} = \underset{C \in Classes}{\operatorname{argmax}} P(C) \prod_{k=1}^{n} P(f_k|C)$$

Note that the first term in the equation is our class prior as discussed above.
One limitation of digital computers arises here. Note that because we are potentially multiplying many probabilities, and probabilities are less than or equal to one, we may rapidly generate a number that is too small for a computer to represent. Hence it is common to map these probabilities via logarithms and, exploiting the properties of logarithms, we add these log-probabilities instead of multiplying the actual probabilities. Our equation becomes:

$$\hat{C} = \underset{C \in Classes}{\operatorname{argmax}} \log P(C) + \sum_{k=1}^{n} \log P(f_k|C)$$

2.3.3 Smoothing

One issue with applying naïve Bayes as above is that we must address is the probability of encountering features not seen during training (or "unseen" events). If we simply try to assign these new features no value (or 0), our product rule would produce zero for an entire case when encountering an unseen event. Similarly, since the logarithm of a zero value is undefined, our summation including the $\log$ of zero is undefined.

In order to account for the possibility that we will encounter events unseen in training, we implement techniques that assign some minute probability to these features. This process is called "smoothing." Because we are dealing with probabilities and they must sum to 1, the idea in smoothing techniques is to take a small amount of probability mass from the features we have seen and give it to the features we have not seen [14].

Add-One Smoothing (Also Known as "Laplace Smoothing")

This method introduces some variability in the science of our data. Add-One smoothing, as the name implies, adds one to every count. The features we have seen are treated as if they have been seen one additional time and unseen features (that had zero counts in training) are given a value as if we had seen each one time. Typically, we define Add-One Smoothing in terms of:

T : the number of unique types we have observed
N : the total number of tokens we have observed
V : the size of the vocabulary
Z : the number of types we have not seen ($Z = V - T$)

Because we added one to every feature, our total count must now be $N + V$ to make room in the total probability for all our features. We denote the smoothed probability as P^*. Our smooth probability of feature f_i now becomes $P_i^* = \frac{count(f_i)}{N+V}$ and we assign all unseen features the probability $\frac{1}{N+V}$ [14].

Witten-Bell Smoothing

This type of smoothing uses a frequentist approach in an attempt to capture an estimate of the probability of seeing a feature for the first time. Using the same notation as above, the sum of the probabilities of seeing features for the first time is assigned as $\frac{T}{N+T}$. As above, Z is all the vocabulary words we have not seen (and thus have no probability data for.) Then each

unseen word will be assigned $(\frac{1}{Z})^{th}$ the total value or $\frac{T}{Z(N+T)}$. Using Witten-Bell smoothing, for features we have seen we use $\frac{count(f_i)}{N+T}$ for the probability of feature f_i [15]. Succinctly:

$$P^*(f_i) = \begin{cases} \frac{T}{Z(N+T)} & \text{if } count(f_i) = 0 \\ \frac{count(f_i)}{N+T} & \text{if } count(f_i) > 0 \end{cases}$$

2.3.4 Hidden Markov Models

Hidden Markov models (HMMs) are frequently used in part-of-speech tagging with the hidden states emitting the respective POS. We are not interested in POS tagging. Instead we implemented HMMs in order to determine if they could provide useful information in dialog act classification. HMMs are discussed in [9, 10, 14, 16].

Baum-Welch Algorithm (input training sequence O, output HMM $H = (\pi, A, B, N, M)$)
Goal: Iteratively estimate model parameters A, B, π
Define: $p_t(i,j), 1 \leq t \leq T, 1 \leq i,j \leq N$ where T is length of O, N is number of hidden states
Step 1: Let initial model be $\mu_0 = (A_0, B_0, \pi_0)$
Step 2:

$$p_t(i,j) = \frac{P(O_t = i, O_{t+1} = j, O|\mu)}{P(O|\mu)} = \frac{\alpha_i(t) a_{ij} b_{ijo_t} \beta_j(t+1)}{\sum_{m=1}^{N} \alpha_m(t)\beta_m(t)}$$

Define:

$$\gamma_t(i) = \sum_{j=1}^{N} P_t(i,j) \text{ the probability of being in state } i \text{ at time } t \text{ given } O.$$

$$\sum_{t=1}^{T} \gamma_i(t) = \text{expected number of transitions from state } i \text{ in } O$$

$$\sum_{t=1}^{T} p_t(i,j) = \text{expected number of transitions from state } i \text{ to } j \text{ in } O$$

$$\hat{\pi}_i = \gamma_i(1) = \text{expected frequency in state } i \text{ at time } t = 1$$

$$\hat{a}_{ij} = \frac{\text{expected number of transitions from state } i \text{ to state } j}{\text{expected number of transitions from } i}$$

$$\hat{b}_{ijk} = \frac{\text{expected number of transitions from } i \text{ to } j \text{ with observed token } k}{\text{expected number of transitions from } i \text{ to } j}$$

If $\log P(O|\hat{\mu}) - \log P(O|\mu_0) < \epsilon$ return $\hat{\mu}$
else $\mu_0 = \hat{\mu}$ and goto Step 2.

Figure 2.1: Baum-Welch Algorithm. After [16].

An HMM H consists of a five-tuple $H = \{\Pi, A, B, N, M\}$ where Π represents the probabilities for each initial state, A is the set of state transition probabilities, B is the set of emission probabilities, N is a set of hidden states, and M is the symbol alphabet.
H is trained by use of a sequence of tokens derived from all tokens observed during training.

The parameters of the language model $\mu = \{\Pi, A, B\}$ are learned through a form of expectation maximization. In this methodology, we begin by estimating the parameters (the expectation step) and then use the maximization step to determine the likelihood of the training sequence given the estimated parameters. We then determine relative importance of the proposed model's transition and emission probabilities and use this information to produce new parameters for the model. By iteratively improving μ's parameters, we improve the overall performance of the HMM until the magnitude of the changes falls below a defined threshold. For HMMs, this iterative algorithm is called the Baum-Welch or Forward-Backward algorithm (see Figure 2.1) [16].

Though subjected to settling at local maxima, the expectation maximization approach has proven effective for use in the training of Hidden Markov Models. When the language model has been determined through training, the HMM uses the calculated μ and processes observation sequences (O where $O = (o_1, o_2, ..., o_n)$ and $o_k \in M$) derived from test cases (for this work these consist of individual chat posts.) The Viterbi algorithm, a form of dynamic programming, is then used to determine the probability of observing a respective test case given H.

2.3.5 Support Vector Machines

Support Vector Machines (SVMs) are discussed in [9, 10, 14]. In general, SVMs produce a discriminant classifier that attempts to find boundaries that separate two distinct classes of data. Because we may have an infinite number of boundaries that satisfy this requirement (see Figure 2.3), SVM further refines the solution to the boundary that maximizes the distance between the data points closest to the proposed boundary (see Figure 2.4). Hence, SVM is also referred to as a maximum margin classifier. Note that the data points closest to the boundary, those whose margin we are maximizing, are called the support vectors. The boundaries produced by SVM classifiers are of dimension $n - 1$ where n corresponds to the dimensions of the data points themselves. Therefore, for two dimensional space, SVM attempts to find a line separating the class examples from the non-class examples, and for three dimensional data, the algorithm attempts to find a boundary in the form of a plane. Above three dimensions, SVM boundaries are called hyperplanes.

Viterbi Algorithm (O, HMM H = (Π, A, B, N, M))
Goal: Find the most probably state sequence $\hat{X} = \text{argmax}_X\, P(X|O, \mu)$
it is sufficient to maximize $\text{argmax}_X\, P(X, O|\mu)$
define:

$$\delta_j(t) = \max_{X1...X_{t-1}} P(X_1, ..., X_{t-1}, o_1, ..., o_{t-1}, X_t = j|\mu)$$

Step 1: Initialization

$$\delta_j(1) = \pi_j, 1 \leq j \leq N$$

Step 2: Induction

$$\delta_j(s+1) = \max_{1 \leq i \leq N} \delta_i(t) a_{ij} b_{ijo_t}, 1 \leq j \leq N$$

Store backtrace

$$\psi_j(t+1) = \operatorname*{argmax}_{1 \leq i \leq N} \delta_i(t) a_{ij} b_{ijo_t}, 1 \leq j \leq N$$

Step 3: Termination and path readout (by backtracking). The most likely state sequence is worked out from the right backwards:

$$\hat{X}_{T+1} = \operatorname*{argmax}_{1 \leq i \leq N} \delta_i(T+1)$$
$$\hat{X}_t = \psi_{\hat{X}_{t+1}}(t+1)$$
$$P(\hat{X}) = \max_{1 \leq i \leq N} \delta_i(T+1)$$

Figure 2.2: Viterbi Algorithm. After [16].

If the data is not linearly separable, support vector machines may apply a kernel function to the data points. This results in added dimensionality of the resulting data and may provide linearly separable points in the new feature space [18].

2.3.6 Decision Trees

Classification using decision trees is discussed in [9, 14, 19, 20, 21]. This classification method uses successive questions about dataset attributes to reduce the possible selections for our classifier until a determination is achieved. At the root of the tree, all classes are considered possible and a question is asked regarding the data features. For a binary decision tree, this is a yes or no question whose answer leads to another node with a subsequent question. Ideally, when a node's question determines the class of a test case, the answer leads to a leaf node that returns

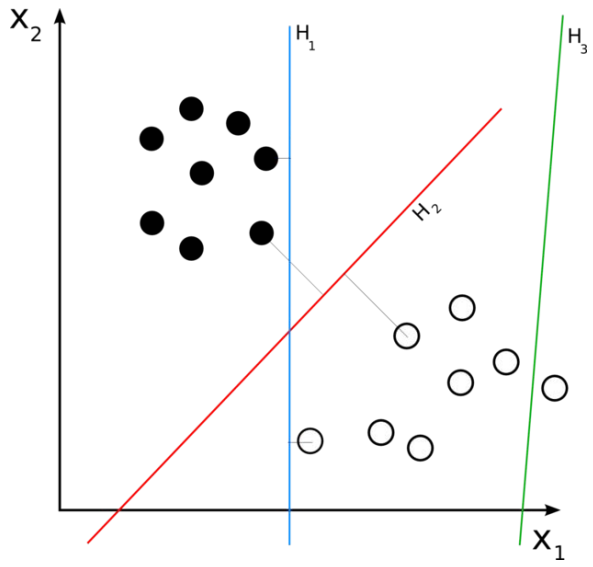


Figure 2.3: Example Separators. From [17].

the classifier's result.

Rather than constructing a decision tree by randomly selecting questions, Quinlan advocates a method of inductively creating decision trees based on using a measure of maximum information gain [22]. Called the ID3 algorithm, it starts at the root of the tree and develops from the top down recursively. If, at a node, the data belongs to only one subset, the tree classifies test data leading to this node as belonging to that subset. If questions are available to divide the subset further, the question providing the highest information gain is selected and the new subsets become nodes on the next lower level. If there are no questions that further segregate the data, the node becomes a leaf and classifies and examples that lead to this leaf as belonging to the most likely class included in the remaining subsets.

Decision trees are hampered by several issues including overfitting and "...handling continuous attributes, choosing an appropriate attribute selection measure, handling training data with missing attribute values, handling attributes with differing costs, and improving computational efficiency" [20]. To address some these issues, Quinlan modified ID3 by using reduced-error pruning. This method considers each node of the tree and if removal of the node does not reduce the performance of the tree when validation data is tested, it is removed and a leaf node that returns the most likely of the classes remaining is installed. Note that this requires the training data to be divided into a training set and a validation set which is not desirable for small training sets.

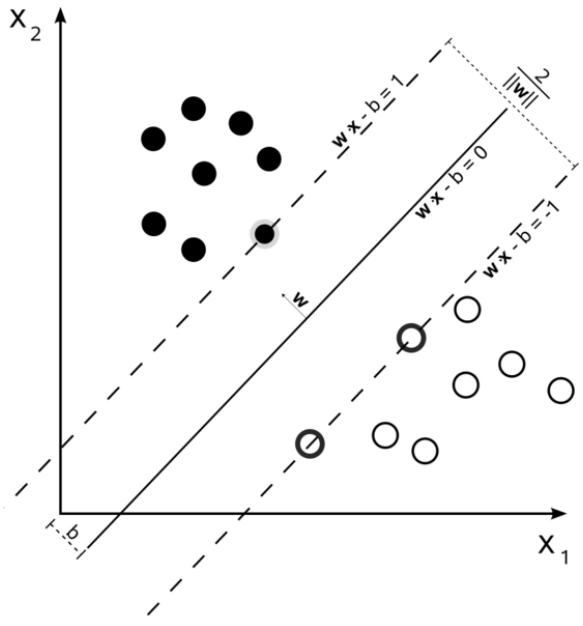


Figure 2.4: Maximum Margin Hyperplane. From [17].

In 1993, Quinlan introduced the C4.5 algorithm as an extension of ID3. This change functions as in ID3 but refines the resulting tree by creating a rule for each path in the tree, generalizing each rule if possible then sorting the rules by comparing their estimated accuracy. In his estimate of rule accuracy in C4.5, Quinlan uses the training set to determine each rule's accuracy and applies a penalty to better estimate test performance [20]. The test data is then classified using these rules.

2.3.7 Maximum Entropy

Application of Maximum Entropy techniques in NLP are discussed in [23, 24]. These techniques are based on making no arbitrary assumptions about the data to be classified. Given no information about a data set with N classes, in order to avoid making undue assumptions, we would require that the probability of an element x belonging to class c_j is uniformly distributed across all classes, thus $p(c_j|x) = \frac{1}{n}$ where $1 \leq j \leq N$. If we discover some piece of evidence during training that would indicate that x is more likely to belong to a subset of one classes, then the probabilities of the classes belonging to this subset are promoted [23]. The classes not in this subset are subsequently reduced in order to maintain total probability equal to one. These models continue to be updated throughout training.

To develop a model, these techniques are used to develop "features" which consist of binary

functions based on observations made during training. These functions, with respect to the class distributions discovered during training, are then used in classification. These statistics, when determined important to the classification task, are then used as constraints to which prospective models must adhere. Those models that violate a constraint are discarded from consideration [23].

Consider the NPS chap corpus domain where we have 15 distinct dialog act classes. With no other information, by the principle of maximum entropy, we would assume a uniform distribution of assign the probability of a particular post belonging to our categories as $p(c) = \frac{1}{15} = 0.067$. If, during training, we discover that half the time we observe the word "how" in a chat post the post belongs to the whQuestion or ynQuestion classes, then we would update our model. Because we have no other information between the two Question classes, we evenly distribute the update across them giving

$$p(\text{whQuestion}|\text{“how”}) = p(\text{ynQuestion}|\text{“how”}) = 0.25$$

and

$$p(\text{all other classes}|\text{“how”}) = 0.0385$$

By repeatedly comparing a test case's data with multiple constraints, the classifier predicts to which class the test case belongs.

2.3.8 Evaluation Criteria

Accuracy

Accuracy is a frequently used metric for comparing the performance of classifiers. Accuracy reports the percentage of items classified correctly. The formula for accuracy is:

$$Accuracy = \frac{TruePositives + TrueNegatives}{TruePositives + FalsePositives + TrueNegatives + FalseNegatives} \tag{2.2}$$

Where:

True Positives: the number of posts in the class of interest that were correctly classified
False Positives: the number of posts incorrectly called members of the class of interest
True Negatives: the number of posts correctly classified as non-class

False Negatives: the number of posts that were members of the class of interest but that were incorrectly classified as non-class

For our work, *Accuracy* is the number of chat posts our classifier correctly labeled divided by the total number of chat posts in the test set

Precision, Recall and F-score

Precision is the proportion of the items a classifier labeled as class c_i correctly versus the total number of it classified as c_i. In essence, precision is a measure of how reliable the output of a classification scheme is. The precision formula is:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

Consider, however, that if our classifier selects one correct example out of many ($TruePositives = 1$), but selects no others ($FalsePositives = 0$), we would achieve a precision of 1.00. Clearly, precision alone is an insufficient measure of performance. *Recall* is the proportion of items a classifier labeled as class c_i versus the total number of examples of c_i in the testing set. The formula for recall is:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

Similar to precision, *Recall* has a shortcoming in that if we select everything, we can achieve a recall of 1.00 because we have classified no false negatives. Because algorithmic approaches may be biased in favor of either precision or recall, and these biases frequently sacrifice one for the other, we provide an F-score for our results [16]. The F-score is a harmonic mean and is given by the formula:

$$F\text{-}score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}.$$

Confusion Matrices

While *Accuracy*, *Precision*, *Recall* and *F-score* provide a high-level indication of a classifiers performance, they provide no utility in determining where the classifier erred. A confusion matrix can be useful in error analysis by displaying truth information in columns and classifier results in rows. Cell (x,y) then represents the number of items in class y that our classifier labeled as x [10]. A confusion matrix is then:

		TRUTH	
		+	-
LABELS	+	True Positives	False Positives
	-	False Negatives	True Negatives

Table 2.4: Example Confusion Matrix

Note that the cell entries in Table 2.4 directly correspond to the terms used in *Accuracy*, *Precision* and *Recall* above.

Consider an example binary classification task performed on a set consisting of 100 test cases with 10 belonging to class c_1 and 90 belonging to class c_2. If our classifier correctly labels 5 cases that belong to c_1 and mislabeled no cases belonging to c_2, our confusion matrix would be:

		TRUTH	
		c_1	c_2
LABELS	c_1	5	0
	c_2	5	90

Table 2.5: Confusion Matrix with Sample Data

We have 5 correctly labeled examples as shown in cell (c_1, c_1). In other terms, we have *True Positives* = 5. Cell (c_1, c_2) shows that we did not mislabel any examples of c_2 as belonging to class c_1 (*False Positives* = 0) and cell (c_2, c_1) indicates that we have mislabeled 5 c_1 cases as not belonging to c_1, or *False Negatives* = 5. Finally, cell (c_2, c_2) shows that we correctly identified all c_2 cases (*True Negatives* = 90).

From our formulas above:

$$\begin{aligned} \textit{Precision} &= \frac{5}{5} = 1.00 \\ \textit{Recall} &= \frac{5}{10} = 0.50 \\ \textit{F-score} &= \frac{2}{\frac{1}{1.0} + \frac{1}{0.5}} = 0.667 \end{aligned}$$

Additionally, we can see a shortcoming of using *Accuracy* as a measure of performance when there are many non-examples in a test set. In this case, *Accuracy* = $\frac{5+90}{100} = 0.95$. While a measure of 95% seems satisfactory, it obfuscates the fact that our classifier missed half of the example cases we may have been interested in.

One's choice in evaluation criteria is clearly important in determining the true performance of any classifier.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:
TECHNICAL APPROACH

3.1 Introduction

Part of speech tagging is useful in dialog act tagging as shown in Forsyth [6] and Wu et al. [8]. Unfortunately, at the current state-of-the-art, accurate grammatical tagging requires hand-annotation in the chat domain. We hypothesize that by using an MLE part of speech tags, similar dialog act tagging performance is achievable with significantly less effort vis-a-vis hand POS tagging.

In this chapter, we describe the data sources and experimental design.

3.2 Sources of Data

We elected to generate our MLE part of speech tags from a domain outside of chat in order to test the viability of our cross-genre approach.

3.2.1 *Wall Street Journal* and Brown Corpora

In order to produce a cross-genre, maximum likelihood estimation (MLE) part of speech tagger we counted the number of words and their corresponding parts of speech in the *Wall Street Journal* and Brown corpora. For the MLE tags we applied, for each word in the CPOS dictionary, the part of speech that had the highest count in the combined corpora. We refer to these tags as "cheap" part-of-speech (or "CPO") tags. Because the tag set used in the Brown corpora was larger, we mapped some of the Brown tags to their Wall Street Journal equivalents. In addition, all words in the CPOS dictionary were converted to lower case.

To reduce the size of the CPOS dictionary, tokens that consisted of cardinal numbers (POS tagged as "CD") were removed and later recognized by regular expressions. Our methodology resulted in a dictionary with 74,034 entries. Note that we did not use any chat corpus data in creating this dictionary.

3.2.2 NPS Chat Corpus

The chat data originally collected by Lin in 2006 is described in Lin [5]. She collected over 477,000 individual posts by 3,290 unique authors. A portion of this corpus was anonymized by

Forsyth who masked personally identifiable information such as names and ages. Users' chat aliases were replaced with templates assigned based on chat room (including age group), date and order that each user joined the respective chat room.

Forsyth part-of-speech and dialog act tagged the anonymized portion of the Lin corpus consisting of 10,567 chat posts [6]. This subset is known as the NPS chat corpus. We considered his tags, both POS and dialog act, as "ground truth" and compared the performance of our dialog act classifier based on his parts of speech and our cheap parts of speech. Table 3.1 shows the

	Post Count	**Percent of Total**
Statement	3185	30.14%
System	2632	24.91%
Greet	1363	12.90%
Emotion	1106	10.47%
ynQuestion	550	5.20%
whQuestion	533	5.04%
Accept	233	2.20%
Bye	195	1.85%
Emphasis	190	1.80%
Continuer	168	1.59%
Reject	159	1.50%
yAnswer	108	1.02%
nAnswer	72	0.68%
Clarify	38	0.36%
Other	35	0.33%

Table 3.1: Number of Posts in NPS Chat Corpus by Dialog Act

breakdown of posts by dialog act class in the entire NPS chat corpus. Note the disparities in the sizes of the different dialog act classes as shown in column two. Naïve Bayes classifiers use class priors ($P(C)$). These are displayed in column three. The large differences in class priors will significantly skew our classifier results toward the Statement and System dialog act classes.

3.2.3 Division of Data

In order to directly compare our classifier results with Forsyth's, we considered each chat post independently and held-out ten percent of the posts for testing. This test set was not used in training. Actual dialog act tags were maintained in the test set data in order to determine classifier performance.

We tested over 50 such divisions. This resulted in an average of 9,513.34 posts (90.02% of total) for training and 1,053.66 (9.98%) posts for testing.

3.3 Classification Tasks

Our task was to determine the effectiveness of cheap parts of speech in determining dialog act class by use of a naïve Bayes classifier. We performed a multi-class classification task over the 15 dialog act classes. Our results contain a comparison of performance between computationally expensive techniques with human verification to determine accurate POS tags versus "cheap" POS tags.

3.4 Feature Selection

Rather than repeating Forsyth's approach of using temporal and specific lexical features of the data (see Table 2.3), we elected to use a more traditional, token-based approach for our naïve Bayes classifier. We used unigrams, bigrams and trigrams from POS tags only as well as bigrams made up of pairs of word/POS pairs.

3.4.1 Features

Naïve Bayes Classifier Features:

1. Actual Part of Speech unigrams, bigrams, trigrams (for comparison)
2. Cheap Part of Speech unigrams, bigrams, trigrams
3. Word, Actual POS pair bigrams (for comparison)
4. Word, Cheap POS pair bigrams
5. Word Bigrams

Figures 3.1, 3.2 and 3.3 show the total counts of features in all 10,567 posts in the NPS chat corpus. We observe that the number of training features for each dialog act class is skewed toward Statement and System classes. Though there are more posts tagged as Emotion than either of the Question classes, the count of features in the Emotion dialog act class is lower. We can infer that posts in the Emotion class are generally shorter than Question posts.

3.5 Experiment Setup

For our experiments, we read in all 10,567 posts in the NPS chat corpus. Training posts were segregated into two data sets, one of which retained actual POS tags and one that replaced these

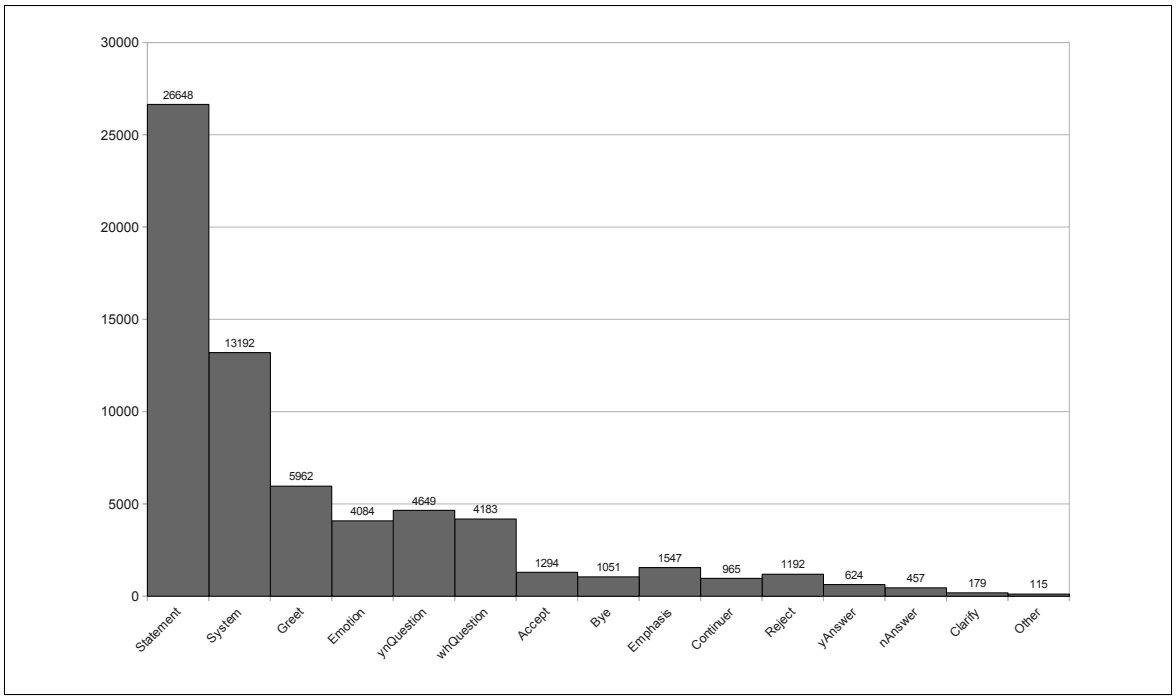


Figure 3.1: Number of Unigram Features by Dialog Act

with CPOS tags. These two structures produced the feature vectors used for testing. Test posts were similarly separated into two data structures one retaining the actual POS tags, the other utilizing CPOS tags. Feature vectors were calculated for each individual post in the test data structures.

We chose to use naïve Bayes classifiers with our different features due to their speed.

Each POS tag was associated with an integer that functioned as an index into arrays that maintained the feature counts.

For each test post, we computed:

$$\hat{C} = \underset{C_i \in Classes}{\operatorname{argmax}} = \log P(C_i) \sum_j \log P^*(f_j|C_i)$$

Noting the disparity in between the class populations, we expected that the class prior probabilities would affect the performance of a naïve Bayes classifier. In order to help overcome this

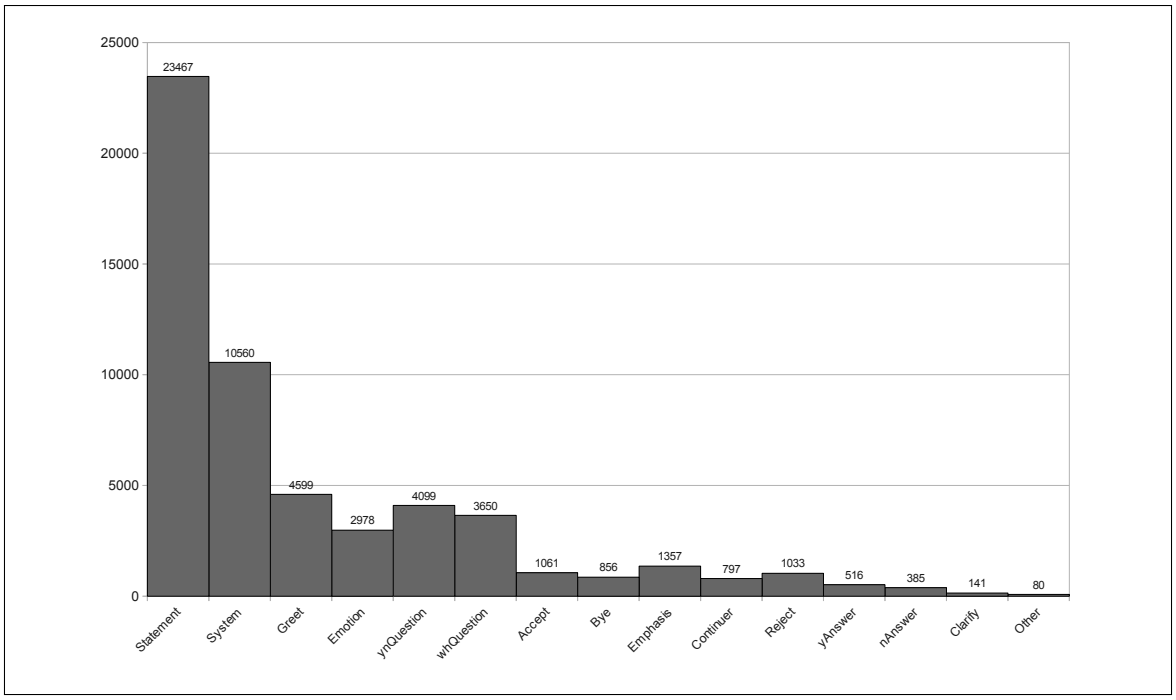


Figure 3.2: Number of Bigram Features by Dialog Act

disparity, we also computed:

$$\hat{C} = \underset{C_i \in Classes}{\operatorname{argmax}} \log P(C_i) \sum_j \log P^*(wpp_j|C_i) + \sum_k \log P^*(pb_k|C_i) \tag{3.1}$$

where wpp_j is the word/POS pair bigram j and pb is the POS bigram k.

Overall Accuracy was computed as $\frac{\textit{True Positives}}{\text{Number of Test Posts}}$ for comparison with Forsyth and because of the large number of *True Negatives* skews the *accuracy* (as shown in equation 2.2) calculations toward 1.00 so as to make them useless.

We noted that Witten-Bell smoothing performed better than LaPlace for our experiments. We provide results for Witten-Bell smoothed unigrams, bigrams and trigrams and LaPlace smoothing of bigrams for comparison.

3.5.1 Data Preprocessing

In processing both the actual and cheap data structures, we converted all word tokens to lower case to match our CPOS dictionary. The parts of speech applied by Forsyth were not changed in the actual data structure. In the cheap data structure, we replaced the actual parts of speech with the cheap parts of speech found in the CPOS dictionary. For each post, start-of-post and

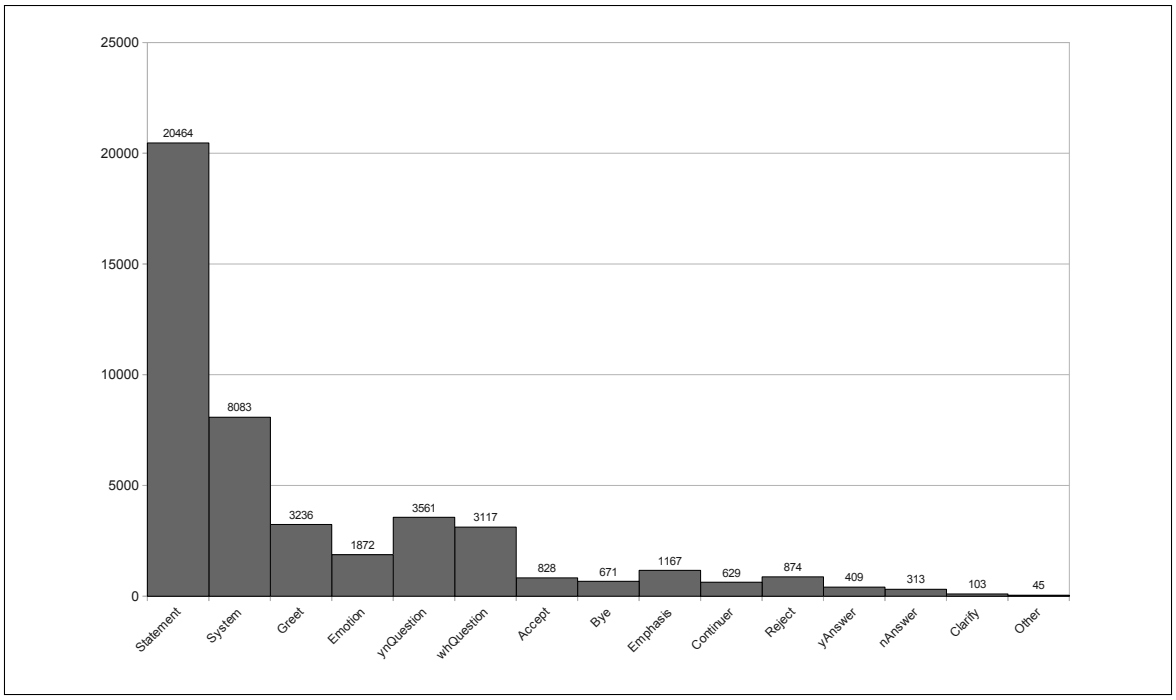


Figure 3.3: Number of Trigram Features by Dialog Act

end-of-post markers were added to preserve context in bigram and trigram classification tasks.

Because none of the emoticons were contained in either the WSJ or Brown corpora, these were initially assigned the CPOS tag of "UNK" or unknown. In addition to providing results with no effort to recognize emoticons, we augmented the CPOS dictionary to recognize emoticons in order to compare performance with the added context provided by these chat features.

Emoticons were assigned the interdiction ("UH") POS by Forsyth, we compared performance of our classifier with "UH" and other POS tags. In addition to marking emoticons with "UNK" (not found in the CPOS dictionary) and "UH," we followed Forsyth's recommendation and tested our classifier marking these features with the unique POS tag "EMO" [6]. We further divided the emoticons into two categories, those found in Appendix A and those composed of phrase abbreviations such as "lol." We provide results of our experiments using all emoticon tagging schemes in Chapter 4.

For our experiments, because we were not interested in identifying individuals, we further masked all user names in training and test posts with a unique word. Because this word was not found in the CPOS dictionary, we automatically assigned the POS tag "NNP" for accurate performance comparison.

Though our effort was not focused on POS tagging accuracy, we noted that our CPOS tagging methodology produced an accuracy ranging from 68.16% to 71.36% depending on our selection of emoticon POS marks. Figure 3.4 provides an example of the difference introduced by the

Post with Actual POS tags: with/IN an/DT answer/NN like/IN that/DT .../: nope/UH ..../: lol/UH
Post with Cheap POS tags: with/IN an/DT answer/NN like/IN that/IN .../: nope/UH ..../UNK lol/EMO

Figure 3.4: Example Post Displaying Differences in POS Markings

CPOS methodology. Start- and end-of-post markings have been removed for clarity. Note that the actual POS tagged post includes the POS tags as applied by Forsyth. The same post, with CPOS tags applied, shows that "with," "an," "answer," and "like" are most often used in the Wall Street Journal and Brown corpora with the same tags as Forsyth applied. "That," however, is most frequently tagged as "IN" (Preposition/subordinating conjunction) in the WSJ and Brown corpora and is marked as such by our CPOS dictionary. In fact, "that" is POS tagged as "IN" 6,682 times and as "DT" 4,373 times in WSJ and Brown. The string "..." is recognized by the CPOS dictionary, however when it includes extra characters, it is not and is given the "UNK" tag as can be seen above. Note also that the popular emoticon "lol" (laugh out loud) is marked with our "EMO" tag as specified in the settings used in this particular experiment.

For illustration, actual POS bigrams for this sample post would produce:
(IN,DT), (DT,NN), (NN,IN), (IN,DT), (DT,:), (:,UH), (UH,:), (:,UH).
Using cheap POS with no emoticon recognition would result in:
(IN,DT), (DT,NN), (NN,IN), (IN,IN), (IN,:), (:,UH), (UH,UNK), (UNK,UNK).
Augmenting our CPOS dictionary to tag emoticons with our "EMO" tag gives:
(IN,DT), (DT,NN), (NN,IN), (IN,IN), (IN,:), (:,UH), (UH,UNK), (UNK,EMO).

3.5.2 Random Trials

We conducted 50 random trials in which 10% of the chat posts were held-out for testing. Confusion matrices for selected experiment runs are included in Appendix C.

Having completed the discussion of our technical approach, we present our results in the next chapter.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:
RESULTS AND ANALYSIS

4.1 Introduction

In this chapter, we present the results of our experiments. Comparison between the performance of the naïve Bayes classifier with various settings and feature selections are provided. For additional comparison, consider that Forsyth achieved a top dialog act tagging accuracy of 83.2% using a time-consuming process that included 300 iterations by a neural network incorporating 24 features. These results were achieved after human verified part of speech tags were applied. Due to limited time available for this work, we could not recreate Forsyth's experiments over our training/testing splits. We noted that each of our experiment runs completed in an average of 27.5 seconds on a desktop machine equipped with an Intel Core i7 and 8 gigabytes of ram. Note that this includes loading all dictionary and chat data, training and testing on both actual POS tagged posts and cheap POS tagged posts.

4.2 Results

For all experiments, we considered the human-verified dialog act tags applied by Forsyth to be ground truth. The results provided in this chapter refer to the performance of the classifier using these tags as "actual" results. These are provided for comparison with the four emoticon tagging schemes below. Note that the actual POS results do not change between experiment sets. In all confusion matrices and summaries, the results derived when using actual POS are provided with the results of cheap POS application for easy reference.

Our results include performance metrics from naïve Bayes classifiers using part of speech unigrams, bigrams, trigrams, word bigrams, and word/POS pair bigrams, all using Witten-Bell smoothing. We also provide LaPlace smoothed results for POS bigrams for comparison to Witten-Bell for these experiments.

We considered our results separately according to the tagging scheme applied to emoticons. Appendix B provides some insight into how our tags grouped features differently. Essentially, we are binning words by their maximum likelihood estimation parts of speech.

No other changes were made to the algorithm between these sets of results. We initially made

no effort to recognize emoticon features noting that none appeared in the cheap POS dictionary. In our first set of 50 experiments, these were automatically assigned the "UNK" part of speech tag.

4.2.1 Emoticons Not Recognized

Making no effort to recognize emoticons results in our cheap POS tagging achieving an accuracy of 68.16%. Essentially, these features are counted with all other unrecognized words, a set that includes misspelled words, unusual use of punctuation (e.g. "...."), etc.

Run Number:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Training Posts	9581	9521	9526	9537	9477	9516	9468	9493	9517	9511	9525	9558	9501	9477	9562	9519	9512	9473
Test Posts	986	1046	1041	1030	1090	1051	1099	1074	1050	1056	1042	1009	1066	1090	1005	1048	1055	1094
MLE performance	0.307	0.285	0.296	0.312	0.312	0.310	0.306	0.304	0.307	0.309	0.316	0.295	0.298	0.298	0.299	0.293	0.282	0.283
Actual POS Unigrams	0.662	0.672	0.663	0.678	0.672	0.684	0.693	0.694	0.696	0.670	0.677	0.681	0.672	0.694	0.689	0.677	0.681	0.683
Cheap POS Unigrams	0.657	0.657	0.628	0.652	0.661	0.656	0.669	0.673	0.652	0.667	0.649	0.654	0.646	0.641	0.662	0.656	0.651	0.654
LaPlace Actual POS 2-grams	0.717	0.721	0.720	0.729	0.720	0.736	0.746	0.742	0.750	0.730	0.727	0.736	0.733	0.741	0.732	0.736	0.735	0.740
LaPlace Cheap POS 2-grams	0.723	0.725	0.709	0.714	0.725	0.731	0.733	0.734	0.724	0.727	0.718	0.714	0.705	0.720	0.721	0.720	0.722	0.723
Actual POS Bigrams	0.729	0.723	0.729	0.742	0.726	0.733	0.744	0.754	0.747	0.741	0.731	0.745	0.727	0.734	0.738	0.736	0.741	0.744
Cheap POS Bigrams	0.729	0.734	0.719	0.724	0.734	0.740	0.748	0.746	0.740	0.743	0.725	0.719	0.720	0.729	0.731	0.730	0.727	0.723
Word/Actual-POS pair 2-grams + POS 2-grams	**0.829**	**0.820**	**0.822**	**0.826**	**0.854**	**0.839**	**0.854**	**0.846**	**0.840**	**0.838**	**0.850**	**0.832**	**0.841**	**0.850**	**0.836**	**0.824**	**0.829**	**0.836**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.834**	**0.822**	**0.828**	**0.820**	**0.845**	**0.842**	**0.854**	**0.834**	**0.835**	**0.833**	**0.839**	**0.832**	**0.833**	**0.842**	**0.825**	**0.819**	**0.827**	**0.824**
Actual POS Trigrams	0.809	0.811	0.810	0.817	0.828	0.816	0.826	0.820	0.823	0.823	0.837	0.813	0.821	0.836	0.823	0.811	0.819	0.820
Cheap POS Trigrams	0.815	0.805	0.803	0.807	0.828	0.826	0.832	0.807	0.811	0.819	0.829	0.811	0.818	0.823	0.807	0.806	0.811	0.814
word 2-grams	0.822	0.823	0.810	0.822	0.849	0.821	0.850	0.823	0.835	0.823	0.840	0.826	0.832	0.837	0.821	0.819	0.824	0.821

Run Number:	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Training Posts	9554	9527	9542	9518	9492	9547	9523	9534	9491	9546	9553	9497	9496	9506	9480	9563	9498	9458
Test Posts	1013	1040	1025	1049	1075	1020	1044	1033	1076	1021	1014	1070	1071	1061	1087	1004	1069	1109
MLE performance	0.316	0.303	0.286	0.299	0.311	0.298	0.291	0.300	0.299	0.299	0.296	0.297	0.288	0.293	0.315	0.282	0.275	0.298
Actual POS Unigrams	0.677	0.680	0.685	0.684	0.687	0.690	0.671	0.684	0.676	0.679	0.686	0.676	0.673	0.647	0.695	0.671	0.675	0.682
Cheap POS Unigrams	0.640	0.651	0.654	0.663	0.649	0.679	0.664	0.661	0.648	0.657	0.665	0.664	0.653	0.627	0.669	0.643	0.659	0.656
LaPlace Actual POS 2-grams	0.735	0.721	0.738	0.735	0.730	0.745	0.725	0.733	0.724	0.728	0.732	0.727	0.725	0.715	0.739	0.735	0.728	0.738
LaPlace Cheap POS 2-grams	0.710	0.713	0.722	0.720	0.716	0.742	0.711	0.732	0.714	0.718	0.730	0.727	0.713	0.704	0.733	0.713	0.724	0.710
Actual POS Bigrams	0.736	0.729	0.734	0.741	0.725	0.750	0.731	0.724	0.741	0.738	0.735	0.739	0.727	0.730	0.741	0.734	0.728	0.739
Cheap POS Bigrams	0.728	0.725	0.726	0.741	0.719	0.749	0.720	0.735	0.723	0.728	0.732	0.745	0.721	0.716	0.753	0.725	0.737	0.717
Word/Actual-POS pair 2-grams + POS 2-grams	**0.831**	**0.847**	**0.837**	**0.845**	**0.832**	**0.851**	**0.827**	**0.832**	**0.840**	**0.836**	**0.845**	**0.836**	**0.826**	**0.833**	**0.847**	**0.839**	**0.837**	**0.844**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.819**	**0.829**	**0.837**	**0.830**	**0.832**	**0.851**	**0.832**	**0.834**	**0.837**	**0.831**	**0.852**	**0.823**	**0.813**	**0.830**	**0.835**	**0.841**	**0.845**	**0.834**
Actual POS Trigrams	0.819	0.825	0.825	0.824	0.815	0.845	0.812	0.817	0.819	0.820	0.831	0.826	0.796	0.807	0.834	0.824	0.816	0.834
Cheap POS Trigrams	0.821	0.811	0.819	0.817	0.812	0.841	0.817	0.820	0.825	0.817	0.830	0.817	0.798	0.816	0.833	0.810	0.819	0.823
word 2-grams	0.810	0.838	0.828	0.831	0.816	0.835	0.823	0.826	0.828	0.829	0.835	0.816	0.810	0.825	0.833	0.829	0.833	0.841

Run Number:	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Training Posts	9513	9560	9528	9484	9548	9477	9436	9471	9495	9577	9511	9445	9485	9538
Test Posts	1054	1007	1039	1083	1019	1090	1131	1096	1072	990	1056	1122	1082	1029
MLE performance	0.309	0.327	0.278	0.302	0.295	0.303	0.304	0.302	0.311	0.287	0.307	0.295	0.291	0.296
Actual POS Unigrams	0.680	0.684	0.679	0.682	0.654	0.701	0.691	0.675	0.683	0.683	0.676	0.688	0.658	0.684
Cheap POS Unigrams	0.646	0.652	0.639	0.646	0.629	0.671	0.655	0.637	0.660	0.651	0.644	0.651	0.638	0.669
LaPlace Actual POS 2-grams	0.733	0.737	0.726	0.733	0.714	0.760	0.730	0.727	0.737	0.713	0.737	0.744	0.704	0.733
LaPlace Cheap POS 2-grams	0.701	0.726	0.713	0.704	0.709	0.734	0.714	0.702	0.726	0.697	0.716	0.711	0.697	0.726
Actual POS Bigrams	0.741	0.749	0.723	0.741	0.736	0.761	0.744	0.719	0.735	0.728	0.750	0.740	0.712	0.733
Cheap POS Bigrams	0.715	0.741	0.719	0.712	0.722	0.749	0.717	0.706	0.737	0.697	0.734	0.719	0.707	0.716
Word/Actual-POS pair 2-grams + POS 2-grams	**0.832**	**0.831**	**0.838**	**0.837**	**0.799**	**0.839**	**0.848**	**0.829**	**0.838**	**0.829**	**0.820**	**0.834**	**0.821**	**0.831**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.832**	**0.833**	**0.842**	**0.830**	**0.809**	**0.834**	**0.828**	**0.829**	**0.830**	**0.823**	**0.816**	**0.831**	**0.816**	**0.824**
Actual POS Trigrams	0.807	0.812	0.823	0.813	0.795	0.826	0.831	0.813	0.828	0.814	0.808	0.826	0.806	0.826
Cheap POS Trigrams	0.804	0.815	0.826	0.806	0.786	0.818	0.821	0.800	0.811	0.798	0.795	0.812	0.810	0.809
word 2-grams	0.823	0.831	0.833	0.834	0.805	0.825	0.843	0.813	0.838	0.818	0.809	0.822	0.821	0.819

Mean	Max	Min
9513.3	9581	9436
1053.7	1131	986
0.2993	0.327	0.275
0.6795	0.701	0.647
0.6534	0.679	0.627
0.7315	0.760	0.704
0.7183	0.742	0.697
0.7359	0.761	0.712
0.7279	0.753	0.697
0.8356	**0.854**	**0.799**
0.8315	**0.854**	**0.809**
0.8196	0.845	0.795
0.8146	0.841	0.786
0.8263	0.850	0.805

Figure 4.1: Summary of Results with Emoticons Unrecognized

The emphasized row in figure 4.1 shows that our best results were achieved using equation 3.1. These rows represent using the sum of feature probabilities of word/POS pair bigrams and of POS bigrams in our naïve Bayes classifier. We can see that using actual POS tags we were able to provide better overall accuracy than was achieved by Forsyth. In fact, using cheap POS, which require no preprocessing time or effort, nearly equaled the previous work.

In order to determine if we would achieve better dialog act classification accuracy with different emoticon tags, we attempted three new tagging schemes.

4.2.2 Emoticons Labeled as Interjections

One of the decisions made by Forsyth in developing the NPS chat corpus was that emoticons should be labeled as interjections ("UH"). We used regular expressions to identify both types of emoticons and augmented our cheap POS dictionary to also label them as interjections.

Using this scheme, our MLE part of speech tagger achieved its highest level of accuracy matching the truth POS tags only 71.36% of the time.

Run Number:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Training Posts	9581	9521	9526	9537	9477	9516	9468	9493	9517	9511	9525	9558	9501	9477	9562	9519	9512	9473
Test Posts	986	1046	1041	1030	1090	1051	1099	1074	1050	1056	1042	1009	1066	1090	1005	1048	1055	1094
MLE performance	0.307	0.285	0.296	0.312	0.312	0.310	0.306	0.304	0.307	0.309	0.316	0.295	0.298	0.298	0.299	0.293	0.282	0.283
Actual POS Unigrams	0.662	0.672	0.663	0.678	0.672	0.684	0.693	0.694	0.696	0.670	0.677	0.681	0.672	0.694	0.689	0.677	0.681	0.683
Cheap POS Unigrams	0.631	0.638	0.605	0.648	0.650	0.634	0.660	0.661	0.645	0.643	0.628	0.644	0.630	0.649	0.646	0.642	0.636	0.654
LaPlace Actual POS 2-grams	0.717	0.721	0.720	0.729	0.720	0.736	0.746	0.742	0.750	0.730	0.727	0.736	0.733	0.741	0.732	0.736	0.735	0.740
LaPlace Cheap POS 2-grams	0.686	0.696	0.692	0.706	0.715	0.707	0.716	0.723	0.713	0.705	0.699	0.699	0.687	0.713	0.707	0.714	0.708	0.716
Actual POS Bigrams	0.729	0.723	0.729	0.742	0.726	0.733	0.744	0.754	0.747	0.741	0.731	0.745	0.727	0.734	0.738	0.736	0.741	0.744
Cheap POS Bigrams	0.704	0.705	0.701	0.710	0.717	0.716	0.723	0.723	0.728	0.717	0.696	0.704	0.700	0.720	0.720	0.720	0.713	0.718
Word/Actual-POS pair 2-grams + POS 2-grams	**0.829**	**0.820**	**0.822**	**0.826**	**0.854**	**0.839**	**0.854**	**0.846**	**0.840**	**0.838**	**0.850**	**0.832**	**0.841**	**0.850**	**0.836**	**0.824**	**0.829**	**0.836**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.836**	**0.824**	**0.827**	**0.820**	**0.842**	**0.842**	**0.851**	**0.832**	**0.838**	**0.831**	**0.841**	**0.833**	**0.835**	**0.841**	**0.825**	**0.822**	**0.823**	**0.822**
Actual POS Trigrams	0.809	0.811	0.810	0.817	0.828	0.816	0.826	0.820	0.823	0.823	0.837	0.813	0.821	0.836	0.823	0.811	0.819	0.820
Cheap POS Trigrams	0.813	0.804	0.804	0.808	0.828	0.820	0.830	0.806	0.809	0.815	0.834	0.809	0.819	0.824	0.811	0.806	0.811	0.812
word 2-grams	0.822	0.823	0.810	0.822	0.849	0.821	0.850	0.823	0.835	0.823	0.840	0.826	0.832	0.837	0.821	0.819	0.824	0.821

Run Number:	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Training Posts	9554	9527	9542	9518	9492	9547	9523	9534	9491	9546	9553	9497	9496	9506	9480	9563	9498	9458
Test Posts	1013	1040	1025	1049	1075	1020	1044	1033	1076	1021	1014	1070	1071	1061	1087	1004	1069	1109
MLE performance	0.316	0.303	0.286	0.299	0.311	0.298	0.291	0.300	0.299	0.299	0.296	0.297	0.288	0.293	0.315	0.282	0.275	0.298
Actual POS Unigrams	0.677	0.680	0.685	0.684	0.687	0.690	0.671	0.684	0.676	0.679	0.686	0.676	0.673	0.647	0.695	0.671	0.675	0.682
Cheap POS Unigrams	0.635	0.644	0.642	0.651	0.647	0.666	0.650	0.644	0.630	0.631	0.652	0.644	0.643	0.618	0.655	0.633	0.646	0.650
LaPlace Actual POS 2-grams	0.735	0.721	0.738	0.735	0.730	0.745	0.725	0.733	0.724	0.728	0.732	0.727	0.725	0.715	0.739	0.735	0.728	0.738
LaPlace Cheap POS 2-grams	0.711	0.704	0.709	0.701	0.709	0.723	0.691	0.710	0.697	0.708	0.715	0.706	0.703	0.694	0.710	0.700	0.703	0.706
Actual POS Bigrams	0.736	0.729	0.734	0.741	0.725	0.750	0.731	0.724	0.741	0.738	0.735	0.739	0.727	0.730	0.741	0.734	0.728	0.739
Cheap POS Bigrams	0.723	0.715	0.703	0.715	0.709	0.730	0.700	0.715	0.710	0.716	0.715	0.723	0.710	0.700	0.729	0.706	0.717	0.707
Word/Actual-POS pair 2-grams + POS 2-grams	**0.831**	**0.847**	**0.837**	**0.845**	**0.832**	**0.851**	**0.827**	**0.832**	**0.840**	**0.836**	**0.845**	**0.836**	**0.826**	**0.833**	**0.847**	**0.839**	**0.837**	**0.844**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.821**	**0.833**	**0.838**	**0.831**	**0.828**	**0.847**	**0.831**	**0.834**	**0.834**	**0.832**	**0.848**	**0.826**	**0.817**	**0.834**	**0.836**	**0.843**	**0.843**	**0.838**
Actual POS Trigrams	0.819	0.825	0.825	0.824	0.815	0.845	0.812	0.817	0.819	0.820	0.831	0.826	0.796	0.807	0.834	0.824	0.816	0.834
Cheap POS Trigrams	0.816	0.809	0.822	0.820	0.814	0.842	0.815	0.821	0.823	0.820	0.832	0.821	0.797	0.812	0.834	0.818	0.819	0.819
word 2-grams	0.810	0.838	0.828	0.831	0.816	0.835	0.823	0.826	0.828	0.829	0.835	0.816	0.810	0.825	0.833	0.829	0.833	0.841

Run Number:	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Training Posts	9513	9560	9528	9484	9548	9477	9436	9471	9495	9577	9511	9445	9485	9538
Test Posts	1054	1007	1039	1083	1019	1090	1131	1096	1072	990	1056	1122	1082	1029
MLE performance	0.309	0.327	0.278	0.302	0.295	0.303	0.304	0.302	0.311	0.287	0.307	0.295	0.291	0.296
Actual POS Unigrams	0.680	0.684	0.679	0.682	0.654	0.701	0.691	0.675	0.683	0.683	0.676	0.688	0.658	0.684
Cheap POS Unigrams	0.641	0.646	0.625	0.644	0.621	0.659	0.637	0.625	0.652	0.630	0.653	0.634	0.619	0.654
LaPlace Actual POS 2-grams	0.733	0.737	0.726	0.733	0.714	0.760	0.730	0.727	0.737	0.713	0.737	0.744	0.704	0.733
LaPlace Cheap POS 2-grams	0.693	0.713	0.700	0.705	0.702	0.733	0.714	0.686	0.711	0.688	0.723	0.705	0.678	0.715
Actual POS Bigrams	0.741	0.749	0.723	0.741	0.736	0.761	0.744	0.719	0.735	0.728	0.750	0.740	0.712	0.733
Cheap POS Bigrams	0.700	0.726	0.706	0.707	0.711	0.740	0.717	0.694	0.719	0.693	0.736	0.716	0.692	0.708
Word/Actual-POS pair 2-grams + POS 2-grams	**0.832**	**0.831**	**0.838**	**0.837**	**0.799**	**0.839**	**0.848**	**0.829**	**0.838**	**0.829**	**0.820**	**0.834**	**0.821**	**0.831**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.828**	**0.829**	**0.843**	**0.829**	**0.806**	**0.833**	**0.833**	**0.828**	**0.834**	**0.824**	**0.815**	**0.832**	**0.820**	**0.827**
Actual POS Trigrams	0.807	0.812	0.823	0.813	0.795	0.826	0.831	0.813	0.828	0.814	0.808	0.826	0.806	0.826
Cheap POS Trigrams	0.805	0.815	0.827	0.808	0.785	0.814	0.824	0.805	0.814	0.802	0.795	0.810	0.813	0.808
word 2-grams	0.823	0.831	0.833	0.834	0.805	0.825	0.843	0.813	0.838	0.818	0.809	0.822	0.821	0.819

Mean	Max	Min
9513.3	9581	9436
1053.7	1131	986
0.2993	0.327	0.275
0.6795	0.701	0.647
0.6413	0.666	0.605
0.7315	0.760	0.704
0.7053	0.733	0.678
0.7359	0.761	0.712
0.7129	0.740	0.692
0.8356	**0.854**	**0.799**
0.8316	**0.851**	**0.806**
0.8196	0.845	0.795
0.8149	0.842	0.785
0.8263	0.850	0.805

Figure 4.2: Summary of Results with Emoticons Tagged as Interjection

Table 4.2 shows that, again, using a combination of feature vectors described in equation 3.1 provided the highest average $accuracy$. Forsyth's decision to tag emoticons as interjections performs better than grouping them into the cheap "UNK" category.

We then explored the use of a unique tag for emoticons. We used regular expressions to identify common emoticons and augmented our dictionary to tag recognized emoticons with "EMO."

4.2.3 Two Types of Emoticons as One Part of Speech

We hypothesized that emoticons may deserve their own part of speech tag and, if so, that our dialog act classification accuracy may improve with this added information. To this point, we have seen that putting all unrecognized words into one category provides less accuracy then identifying emoticons as interjections. We decided to give emoticons a unique cheap POS tag and elected to tag them with "EMO."

Run Number:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Training Posts	9581	9521	9526	9537	9477	9516	9468	9493	9517	9511	9525	9558	9501	9477	9562	9519	9512	9473
Test Posts	986	1046	1041	1030	1090	1051	1099	1074	1050	1056	1042	1009	1066	1090	1005	1048	1055	1094
MLE performance	0.307	0.285	0.296	0.312	0.312	0.310	0.306	0.304	0.307	0.309	0.316	0.295	0.298	0.298	0.299	0.293	0.282	0.283
Actual POS Unigrams	0.662	0.672	0.663	0.678	0.672	0.684	0.693	0.694	0.696	0.670	0.677	0.681	0.672	0.694	0.689	0.677	0.681	0.683
Cheap POS Unigrams	0.656	0.665	0.630	0.661	0.664	0.659	0.673	0.679	0.669	0.675	0.655	0.663	0.659	0.661	0.669	0.662	0.658	0.664
LaPlace Actual POS 2-grams	0.717	0.721	0.720	0.729	0.720	0.736	0.746	0.742	0.750	0.730	0.727	0.736	0.733	0.741	0.732	0.736	0.735	0.740
LaPlace Cheap POS 2-grams	0.720	0.728	0.718	0.721	0.730	0.735	0.736	0.740	0.733	0.738	0.721	0.717	0.712	0.728	0.727	0.729	0.728	0.731
Actual POS Bigrams	0.729	0.723	0.729	0.742	0.726	0.733	0.744	0.754	0.747	0.741	0.731	0.745	0.727	0.734	0.738	0.736	0.741	0.744
Cheap POS Bigrams	0.732	0.736	0.726	0.729	0.731	0.742	0.747	0.747	0.749	0.751	0.726	0.721	0.727	0.735	0.735	0.740	0.734	0.735
Word/Actual-POS pair 2-grams + POS 2-grams	**0.829**	**0.820**	**0.822**	**0.826**	**0.854**	**0.839**	**0.854**	**0.846**	**0.840**	**0.838**	**0.850**	**0.832**	**0.841**	**0.850**	**0.836**	**0.824**	**0.829**	**0.836**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.831**	**0.825**	**0.831**	**0.819**	**0.846**	**0.846**	**0.853**	**0.834**	**0.839**	**0.832**	**0.841**	**0.834**	**0.834**	**0.839**	**0.822**	**0.823**	**0.826**	**0.824**
Actual POS Trigrams	0.809	0.811	0.810	0.817	0.828	0.816	0.826	0.820	0.823	0.823	0.837	0.813	0.821	0.836	0.823	0.811	0.819	0.820
Cheap POS Trigrams	0.815	0.802	0.802	0.811	0.831	0.825	0.835	0.809	0.809	0.816	0.827	0.814	0.823	0.823	0.809	0.807	0.816	0.814
word 2-grams	0.822	0.823	0.810	0.822	0.849	0.821	0.850	0.823	0.835	0.823	0.840	0.826	0.832	0.837	0.821	0.819	0.824	0.821

Run Number:	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Training Posts	9554	9527	9542	9518	9492	9547	9523	9534	9491	9546	9553	9497	9496	9506	9480	9563	9498	9458
Test Posts	1013	1040	1025	1049	1075	1020	1044	1033	1076	1021	1014	1070	1071	1061	1087	1004	1069	1109
MLE performance	0.316	0.303	0.286	0.299	0.311	0.298	0.291	0.300	0.299	0.299	0.296	0.297	0.288	0.293	0.315	0.282	0.275	0.298
Actual POS Unigrams	0.677	0.680	0.685	0.684	0.687	0.690	0.671	0.684	0.676	0.679	0.686	0.676	0.673	0.647	0.695	0.671	0.675	0.682
Cheap POS Unigrams	0.652	0.664	0.667	0.669	0.661	0.690	0.674	0.666	0.656	0.656	0.671	0.674	0.663	0.641	0.680	0.650	0.665	0.666
LaPlace Actual POS 2-grams	0.735	0.721	0.738	0.735	0.730	0.745	0.725	0.733	0.724	0.728	0.732	0.727	0.725	0.715	0.739	0.735	0.728	0.738
LaPlace Cheap POS 2-grams	0.724	0.724	0.737	0.727	0.727	0.751	0.717	0.737	0.723	0.735	0.732	0.734	0.721	0.716	0.741	0.723	0.731	0.722
Actual POS Bigrams	0.736	0.729	0.734	0.741	0.725	0.750	0.731	0.724	0.741	0.738	0.735	0.739	0.727	0.730	0.741	0.734	0.728	0.739
Cheap POS Bigrams	0.735	0.738	0.735	0.745	0.731	0.755	0.730	0.738	0.730	0.742	0.736	0.750	0.727	0.725	0.756	0.729	0.746	0.727
Word/Actual-POS pair 2-grams + POS 2-grams	**0.831**	**0.847**	**0.837**	**0.845**	**0.832**	**0.851**	**0.827**	**0.832**	**0.840**	**0.836**	**0.845**	**0.836**	**0.826**	**0.833**	**0.847**	**0.839**	**0.837**	**0.844**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.822**	**0.835**	**0.839**	**0.830**	**0.833**	**0.847**	**0.832**	**0.836**	**0.836**	**0.832**	**0.848**	**0.822**	**0.816**	**0.835**	**0.838**	**0.843**	**0.841**	**0.839**
Actual POS Trigrams	0.819	0.825	0.825	0.824	0.815	0.845	0.812	0.817	0.819	0.820	0.831	0.826	0.796	0.807	0.834	0.824	0.816	0.834
Cheap POS Trigrams	0.813	0.809	0.821	0.822	0.817	0.844	0.816	0.823	0.824	0.819	0.832	0.820	0.796	0.814	0.835	0.815	0.821	0.820
word 2-grams	0.810	0.838	0.828	0.831	0.816	0.835	0.823	0.826	0.828	0.829	0.835	0.816	0.810	0.825	0.833	0.829	0.833	0.841

Run Number:	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Training Posts	9513	9560	9528	9484	9548	9477	9436	9471	9495	9577	9511	9445	9485	9538
Test Posts	1054	1007	1039	1083	1019	1090	1131	1096	1072	990	1056	1122	1082	1029
MLE performance	0.309	0.327	0.278	0.302	0.295	0.303	0.304	0.302	0.311	0.287	0.307	0.295	0.291	0.296
Actual POS Unigrams	0.680	0.684	0.679	0.682	0.654	0.701	0.691	0.675	0.683	0.683	0.676	0.688	0.658	0.684
Cheap POS Unigrams	0.660	0.655	0.647	0.654	0.632	0.676	0.659	0.650	0.662	0.647	0.653	0.652	0.649	0.680
LaPlace Actual POS 2-grams	0.733	0.737	0.726	0.733	0.714	0.760	0.730	0.727	0.737	0.713	0.737	0.744	0.704	0.733
LaPlace Cheap POS 2-grams	0.713	0.725	0.721	0.719	0.711	0.750	0.734	0.713	0.729	0.709	0.728	0.725	0.707	0.738
Actual POS Bigrams	0.741	0.749	0.723	0.741	0.736	0.761	0.744	0.719	0.735	0.728	0.750	0.740	0.712	0.733
Cheap POS Bigrams	0.723	0.742	0.729	0.723	0.723	0.758	0.737	0.720	0.735	0.707	0.742	0.730	0.714	0.732
Word/Actual-POS pair 2-grams + POS 2-grams	**0.832**	**0.831**	**0.838**	**0.837**	**0.799**	**0.839**	**0.848**	**0.829**	**0.838**	**0.829**	**0.820**	**0.834**	**0.821**	**0.831**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.831**	**0.830**	**0.842**	**0.830**	**0.807**	**0.835**	**0.831**	**0.828**	**0.833**	**0.823**	**0.816**	**0.831**	**0.823**	**0.828**
Actual POS Trigrams	0.807	0.812	0.823	0.813	0.795	0.826	0.831	0.813	0.828	0.814	0.808	0.826	0.806	0.826
Cheap POS Trigrams	0.809	0.815	0.825	0.805	0.786	0.813	0.826	0.803	0.814	0.801	0.798	0.812	0.814	0.811
word 2-grams	0.823	0.831	0.833	0.834	0.805	0.825	0.843	0.813	0.838	0.818	0.809	0.822	0.821	0.819

Mean	Max	Min
9513.3	9581	9436
1053.7	1131	986
0.2993	0.327	0.275
0.6795	0.701	0.647
0.6613	0.690	0.630
0.7315	0.760	0.704
0.7267	0.751	0.707
0.7359	0.761	0.712
0.7347	0.758	0.707
0.8356	**0.854**	**0.799**
0.8323	**0.853**	**0.807**
0.8196	0.845	0.795
0.8157	0.844	0.786
0.8263	0.850	0.805

Figure 4.3: Summary of Results with Emoticons Tagged as “EMO”

As can be seen in Figure 4.3, we improved the accuracy of our classifier slightly. This suggests that emoticons may serve better as this new part of speech rather than as interjections. Examining the emoticons in use today, there appear to be two distinct types, those that are made of combinations of punctuation (“smileys” such as “:)”) and those that are acronyms like “lol” for “laugh[ing] out loud.”

4.2.4 Two Types of Emoticon Tags

In order to determine if our dialog act classifier performance would improve if we recognized the different types of emoticons as two different parts of speech, we augmented the POS dictionary as such. Emoticons based on acronyms were assigned the part of speech tag of “EMO2.”

Run Number:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Training Posts	9581	9521	9526	9537	9477	9516	9468	9493	9517	9511	9525	9558	9501	9477	9562	9519	9512	9473
Test Posts	986	1046	1041	1030	1090	1051	1099	1074	1050	1056	1042	1009	1066	1090	1005	1048	1055	1094
MLE performance	0.307	0.285	0.296	0.312	0.312	0.310	0.306	0.304	0.307	0.309	0.316	0.295	0.298	0.298	0.299	0.293	0.282	0.283
Actual POS Unigrams	0.662	0.672	0.663	0.678	0.672	0.684	0.693	0.694	0.696	0.670	0.677	0.681	0.672	0.694	0.689	0.677	0.681	0.683
Cheap POS Unigrams	0.656	0.666	0.629	0.660	0.663	0.662	0.674	0.679	0.670	0.675	0.654	0.664	0.659	0.661	0.669	0.662	0.658	0.666
LaPlace Actual POS 2-grams	0.717	0.721	0.720	0.729	0.720	0.736	0.746	0.742	0.750	0.730	0.727	0.736	0.733	0.741	0.732	0.736	0.735	0.740
LaPlace Cheap POS 2-grams	0.720	0.728	0.719	0.722	0.730	0.735	0.735	0.740	0.731	0.736	0.721	0.720	0.711	0.728	0.726	0.729	0.729	0.730
Actual POS Bigrams	0.729	0.723	0.729	0.742	0.726	0.733	0.744	0.754	0.747	0.741	0.731	0.745	0.727	0.734	0.738	0.736	0.741	0.744
Cheap POS Bigrams	0.735	0.736	0.723	0.730	0.732	0.739	0.749	0.745	0.747	0.749	0.726	0.721	0.727	0.736	0.733	0.742	0.734	0.733
Word/Actual-POS pair 2-grams + POS 2-grams	**0.829**	**0.820**	**0.822**	**0.826**	**0.854**	**0.839**	**0.854**	**0.846**	**0.840**	**0.838**	**0.850**	**0.832**	**0.841**	**0.850**	**0.836**	**0.824**	**0.829**	**0.836**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.833**	**0.827**	**0.828**	**0.821**	**0.846**	**0.840**	**0.854**	**0.837**	**0.839**	**0.833**	**0.841**	**0.836**	**0.833**	**0.839**	**0.824**	**0.823**	**0.827**	**0.824**
Actual POS Trigrams	0.809	0.811	0.810	0.817	0.828	0.816	0.826	0.820	0.823	0.823	0.837	0.813	0.821	0.836	0.823	0.811	0.819	0.820
Cheap POS Trigrams	0.816	0.806	0.798	0.811	0.831	0.821	0.834	0.808	0.809	0.817	0.825	0.818	0.820	0.822	0.812	0.807	0.817	0.813
word 2-grams	0.822	0.823	0.810	0.822	0.849	0.821	0.850	0.823	0.835	0.823	0.840	0.826	0.832	0.837	0.821	0.819	0.824	0.821

Run Number:	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Training Posts	9554	9527	9542	9518	9492	9547	9523	9534	9491	9546	9553	9497	9496	9506	9480	9563	9498	9458
Test Posts	1013	1040	1025	1049	1075	1020	1044	1033	1076	1021	1014	1070	1071	1061	1087	1004	1069	1109
MLE performance	0.316	0.303	0.286	0.299	0.311	0.298	0.291	0.300	0.299	0.299	0.296	0.297	0.288	0.293	0.315	0.282	0.275	0.298
Actual POS Unigrams	0.677	0.680	0.685	0.684	0.687	0.690	0.671	0.684	0.676	0.679	0.686	0.676	0.673	0.647	0.695	0.671	0.675	0.682
Cheap POS Unigrams	0.654	0.665	0.666	0.670	0.662	0.691	0.673	0.666	0.656	0.657	0.670	0.672	0.663	0.641	0.681	0.650	0.666	0.665
LaPlace Actual POS 2-grams	0.735	0.721	0.738	0.735	0.730	0.745	0.725	0.733	0.724	0.728	0.732	0.727	0.725	0.715	0.739	0.735	0.728	0.738
LaPlace Cheap POS 2-grams	0.724	0.723	0.737	0.725	0.728	0.750	0.717	0.736	0.722	0.734	0.731	0.735	0.721	0.716	0.741	0.724	0.733	0.720
Actual POS Bigrams	0.736	0.729	0.734	0.741	0.725	0.750	0.731	0.724	0.741	0.738	0.735	0.739	0.727	0.730	0.741	0.734	0.728	0.739
Cheap POS Bigrams	0.736	0.736	0.735	0.745	0.731	0.756	0.729	0.737	0.730	0.742	0.737	0.750	0.724	0.723	0.760	0.729	0.748	0.723
Word/Actual-POS pair 2-grams + POS 2-grams	**0.831**	**0.847**	**0.837**	**0.845**	**0.832**	**0.851**	**0.827**	**0.832**	**0.840**	**0.836**	**0.845**	**0.836**	**0.826**	**0.833**	**0.847**	**0.839**	**0.837**	**0.844**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.822**	**0.832**	**0.838**	**0.830**	**0.831**	**0.848**	**0.832**	**0.835**	**0.836**	**0.835**	**0.850**	**0.822**	**0.815**	**0.832**	**0.837**	**0.843**	**0.843**	**0.837**
Actual POS Trigrams	0.819	0.825	0.825	0.824	0.815	0.845	0.812	0.817	0.819	0.820	0.831	0.826	0.796	0.807	0.834	0.824	0.816	0.834
Cheap POS Trigrams	0.814	0.809	0.820	0.816	0.813	0.845	0.817	0.821	0.823	0.821	0.832	0.819	0.795	0.813	0.837	0.816	0.822	0.820
word 2-grams	0.810	0.838	0.828	0.831	0.816	0.835	0.823	0.826	0.828	0.829	0.835	0.816	0.810	0.825	0.833	0.829	0.833	0.841

Run Number:	37	38	39	40	41	42	43	44	45	46	47	48	49	50	Mean	Max	Min
Training Posts	9513	9560	9528	9484	9548	9477	9436	9471	9495	9577	9511	9445	9485	9538	9513.3	9577	9436
Test Posts	1054	1007	1039	1083	1019	1090	1131	1096	1072	990	1056	1122	1082	1029	1053.7	1131	990
MLE performance	0.309	0.327	0.278	0.302	0.295	0.303	0.304	0.302	0.311	0.287	0.307	0.295	0.291	0.296	0.2993	0.327	0.275
Actual POS Unigrams	0.680	0.684	0.679	0.682	0.654	0.701	0.691	0.675	0.683	0.683	0.676	0.688	0.658	0.684	0.6795	0.701	0.647
Cheap POS Unigrams	0.659	0.655	0.649	0.656	0.631	0.678	0.658	0.649	0.664	0.646	0.654	0.651	0.649	0.680	0.6616	0.691	0.629
LaPlace Actual POS 2-grams	0.733	0.737	0.726	0.733	0.714	0.760	0.730	0.727	0.737	0.713	0.737	0.744	0.704	0.733	0.7315	0.760	0.704
LaPlace Cheap POS 2-grams	0.713	0.725	0.720	0.718	0.710	0.750	0.734	0.712	0.729	0.709	0.728	0.725	0.706	0.738	0.7265	0.750	0.706
Actual POS Bigrams	0.741	0.749	0.723	0.741	0.736	0.761	0.744	0.719	0.735	0.728	0.750	0.740	0.712	0.733	0.7359	0.761	0.712
Cheap POS Bigrams	0.724	0.743	0.729	0.720	0.724	0.757	0.736	0.719	0.735	0.710	0.743	0.730	0.717	0.730	0.7345	0.760	0.710
Word/Actual-POS pair 2-grams + POS 2-grams	**0.832**	**0.831**	**0.838**	**0.837**	**0.799**	**0.839**	**0.848**	**0.829**	**0.838**	**0.829**	**0.820**	**0.834**	**0.821**	**0.831**	**0.8356**	**0.854**	**0.799**
Word/Cheap-POS pair 2-grams + POS 2-grams	**0.831**	**0.827**	**0.841**	**0.829**	**0.808**	**0.835**	**0.831**	**0.829**	**0.835**	**0.826**	**0.816**	**0.831**	**0.823**	**0.826**	**0.8323**	**0.854**	**0.808**
Actual POS Trigrams	0.807	0.812	0.823	0.813	0.795	0.826	0.831	0.813	0.828	0.814	0.808	0.826	0.806	0.826	0.8196	0.845	0.795
Cheap POS Trigrams	0.809	0.814	0.829	0.802	0.787	0.812	0.826	0.804	0.813	0.802	0.798	0.811	0.811	0.811	0.8154	0.845	0.787
word 2-grams	0.823	0.831	0.833	0.834	0.805	0.825	0.843	0.813	0.838	0.818	0.809	0.822	0.821	0.819	0.8263	0.850	0.805

Figure 4.4: Summary of Results with Emoticons Separated into Two Groups

Figure 4.4 shows a similar performance when we segregate the two emoticon types. Separating the emoticons into two groups based on type actually increased our classifiers performance by 0.003%. We suspect that this may indicate that the different types serve different syntactic purposes. Further analysis of this phenomenon was not completed due to time constraints.

4.3 Analysis

We have demonstrated that using equation 3.1 provided the best accuracy for our cheap POS method and that our method equals or improves accuracy depending on which tags are applied to emoticons as compared to Forsyth [6].

We note that classification based on word bigrams gives an overall accuracy of 82.63%, actual POS bigrams result in 73.59% and actual POS trigrams 81.96%. This suggests that sentence structure rather than content carries the dialog act signal. Cheap POS bigrams achieve an accuracy of 73.47% when all emoticons are given a common tag. Cheap POS trigrams with this tagging scheme result in an overall accuracy of 81.57%, only 0.39% less than actual POS trigrams. Our cheap POS trigrams carry the dialog act signal virtually as well as actual POS trigrams.

Appendix C contains tables showing the effects of our various POS tagging schemes. We provide the counts of each POS by dialog act type. Figure B.1 contains the counts of these tags as applied by Forsyth. Figures B.2, B.3, B.4 and B.5 show the cheap POS counts as applied by our methodology. We note the shifts in "UNK," "UH," "EMO" and "EMO2" counts according to tagging scheme as each figure's caption indicates. These experiments serve as a preliminary exploration of Harris' "Distributional Hypothesis" [25].

In order to demonstrate statistical significance in our experiments, we chose to compare the performance of word bigrams, cheap POS and actual POS for this task, we chose the Wilcoxon Signed-Rank Pair test and selected a confidence level of 99%.

4.3.1 Statistical Significance with Emoticons Unrecognized

Figure 4.5 displays the distributions of overall accuracies when emoticons are not recognized and are therefore tagged as "UNK." We can see overlap in the performance of our classifier using our selected feature sets.

We applied the Wilcoxon Signed-Rank Pair test between word bigrams and cheap POS results using equation 3.1 with a resulting p value of 0.0000181. There is only a remote possibility that

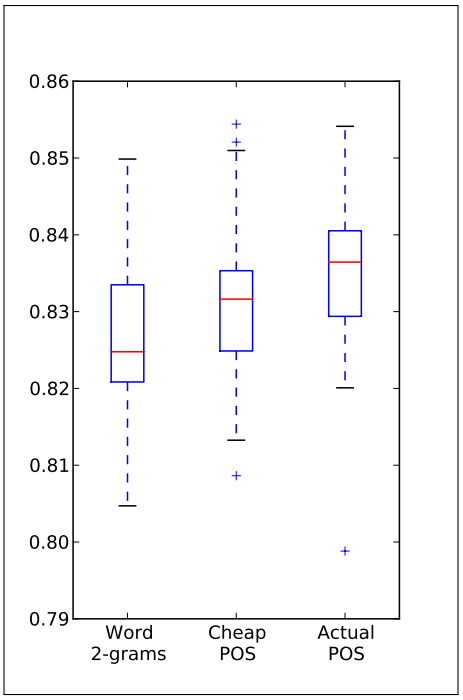


Figure 4.5: Bar Plot of Accuracies with Emoticons Unrecognized

this is a result of random chance.

Applying the same test between the cheap and actual POS results also exceeded our 99% confidence level with a p value of 0.00025. We conclude that we have strong statistical significance in our method's performance.

4.3.2 Statistical Significance with Emoticons Tagged as Interjections

Figure 4.6 shows the distributions of overall accuracies when we concur with Forsyth's decision to tag emoticons as interjections. The word bigrams and features using actual POS marks data show no change from the previous figure and are provided for easy reference. We see the general improvement in cheap POS feature performance as a slight upward trend.

We applied the Signed-Rank Pair test between word bigram and cheap POS performance and computed a p value of 0.0000051. We conclude statistical significance in our method.

We also applied the test between cheap POS and actual POS data with a resulting p value of 0.00017.

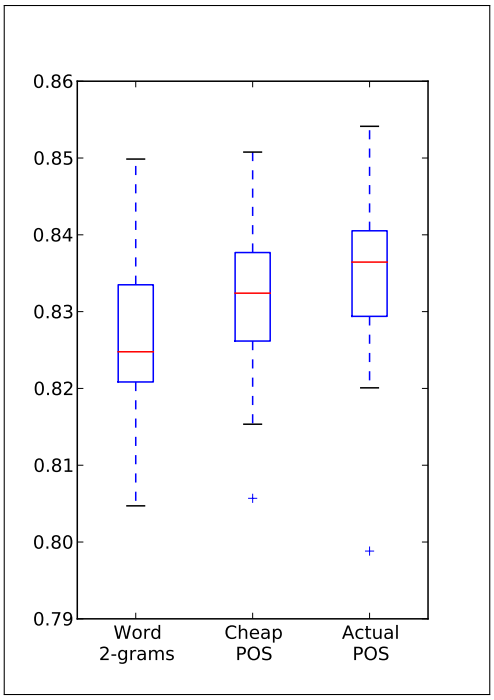


Figure 4.6: Bar Plot of Accuracies with Emoticons Tagged as Interjections

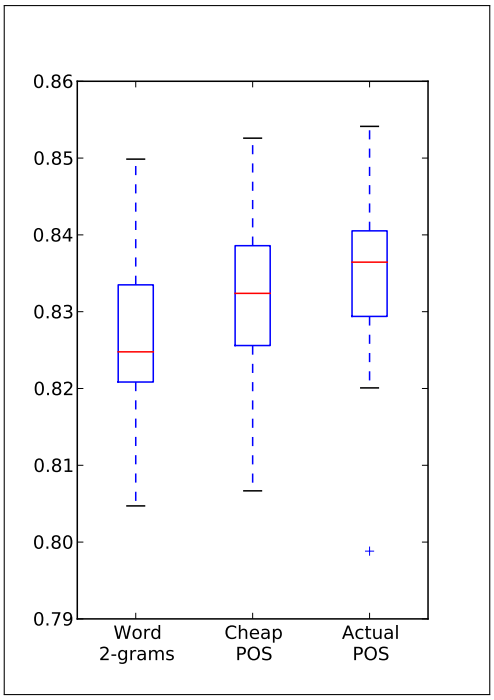


Figure 4.7: Bar Plot of Accuracies with All Emoticons Tagged as “EMO”

4.3.3 Statistical Significance with Emoticons Tagged with "EMO"

We find in Figure 4.7 that marking emoticons with a single, unique tag gives better results than using the interjection tag. In fact, this emoticons tagging scheme produces better average accuracy than the previous work.

We continued to use the Wilcoxon Signed-Rank Pair test with a confidence level of 99%. When we compared word bigrams with the performance of cheap POS, our p value was 0.0000013. The test resulted in a p value of 0.00191 when comparing cheap POS performance to actual POS performance.

We continue to demonstrate statistical significance.

4.3.4 Statistical Significance with Emoticons Tagged as Two Types

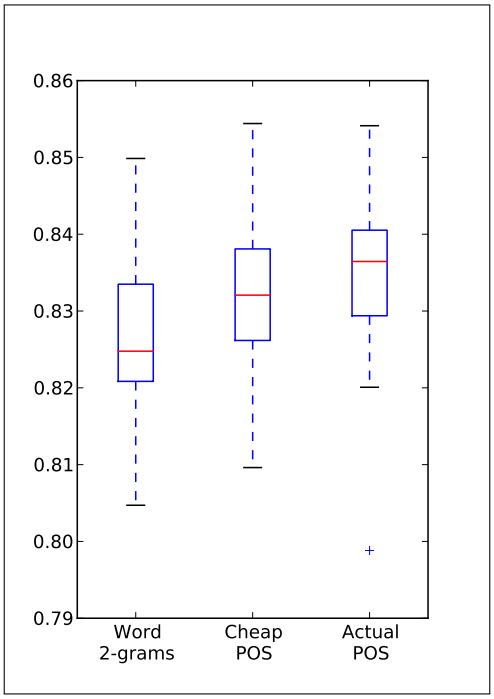


Figure 4.8: Bar Plot of Accuracies with Emoticons Tagged as "EMO" and "EMO2"

Figure 4.8 represents our final emoticon tagging scheme and shows a very slight increase in average accuracy over the previous scheme. This average also matches Forsyth's best. Significance testing, conducted as in previous experiments, gives a p value of 0.000019 in comparison of word bigrams and cheap POS. We achieved a p value of 0.00061 when testing cheap and actual POS performances.

4.3.5 Dialog Act or Authorship Identification?

In order to determine that our classifier results were not influenced by the characteristics of prolific chat participants, we used Forsyth's original data to map masked user names to their screen names. We were able to attribute 9,856 posts to 1,122 individuals. We split the correlated data using 90% of the identified authors for training and the other 10% for testing. No posts from the tested authors were included in the training set. We performed testing over 10 such splits with the overall accuracies provided in Table 4.1: Note that the number of posts used

Run Number:	Mean	Max	Min
Training Posts	8828.8	9277	8653
Test Posts	1027.2	1203	579
MLE performance	0.2988	0.3310	0.2418
Actual POS Unigrams	0.6920	0.7513	0.6540
Cheap POS Unigrams	0.6802	0.7427	0.6238
LaPlace Actual POS 2-grams	0.7406	0.7910	0.7037
LaPlace Cheap POS 2-grams	0.7357	0.7807	0.6852
Actual POS Bigrams	0.7395	0.7997	0.7027
Cheap POS Bigrams	0.7410	0.7841	0.6988
Actual word/POS 2-grams + POS 2-grams	**0.8350**	**0.8722**	**0.8051**
Cheap POS word/POS 2-grams + POS 2-grams	**0.8337**	**0.8756**	**0.7973**
Actual POS Trigrams	0.8160	0.8411	0.7836
Cheap POS Trigrams	0.8131	0.8549	0.7700
word 2-grams	0.8231	0.8549	0.7856

Table 4.1: Average Dialog Act Tagging Accuracies Leaving 10% of Authors Out

for testing varies significantly with a maximum of 1,203 and minimum of 579. This is due to the wide variation in individual user contributions. Figure 4.9 provides a histogram of the number of authors with post count bins on the x-axis. Note that 913 authors (81.4%) of the identifiable authors produced 10 or less posts while the most prolific author provided more than 130 posts. Splitting the data set by author and the disparity in levels of author participation are responsible for our test population variance. Table 4.1 shows that it is unlikely that our dialog act classification method is influenced by author characteristics.

We have demonstrated a technique that provides improved dialog act tagging accuracy in the chat domain. We have also shown statistical significance in our method's performance and that our results are not skewed by author characteristics. While prior work in this domain has relied on time consuming, human-verified part of speech tagging, our method demonstrates that this

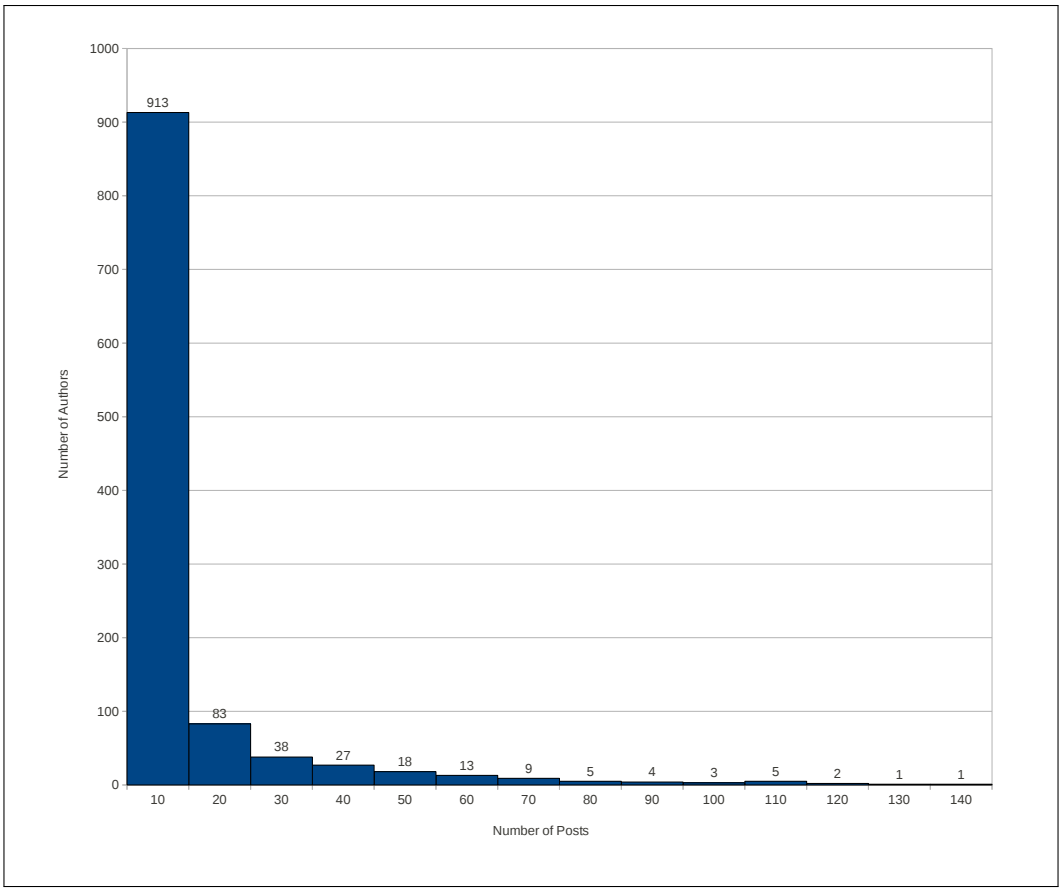


Figure 4.9: Histogram of Author Post Counts

investment is not required for effective dialog act tagging in the chat domain.

With our presentation of experiment results and analysis complete, we provide our conclusions and recommendations for future work in Chapter 5.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:
CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Part of speech tagging is useful in dialog act tagging as shown in Forsyth [6] and Wu et al. [8]. Unfortunately, at the current state-of-the-art, accurate grammatical tagging requires hand-annotation in the chat domain. We hypothesize that by using cross-domain MLE part of speech tags, similar dialog act tagging performance is achievable with significantly less effort vis-a-vis hand POS tagging.

The methodology presented in Chapter 3 performs virtually as well as using actual, hand-tagged part of speech tags without the preprocessing time and effort. Our experiments show that for the chat domain, accurate POS tags are not required to effectively determine chat post dialog act tags. Though our results show a minimal decrease in overall accuracy when compared to same experiments using actual parts of speech, we required no preprocessing nor hand-tagging of parts of speech. Further, cheap POS tagging is extremely fast. We also showed, through statistical significance testing that our method's performance, with high probability, is not the result of chance.

While using actual POS tags performed only 0.3% better than cheap POS tags, for accurate dialog act determination, we required only the processing time required to load our POS dictionary and apply these tags.

5.1.1 Uses for Dialog Acts

Stolcke suggests that consensus is building in the Natural Language Processing community that dialog act tags are useful for higher-order linguistic analysis [12]. Dialog act tags have been used in multi-party meeting summarization (Yang et al. [26]) and spoken dialog systems (Walker and Passionneau [27]). Spoken dialog systems can use dialog act tags to improve response accuracy.

5.1.2 Implications for Tactical Military Chat

Eovito's thesis provides a list of functional requirements for tactical military chat. We believe some of these requirements may benefit from automatically determined dialog act information. For example, Eovito' core requirements include Thread Population/Repopulation. This function

is designed to provide new or returning users a recapitulation of recent tactical chat events [2]. Rather than present these participants with a temporally indiscriminant list of messages, we believe that dialog acts could be used to filter the information provided to include dialog acts of interest. For example, the system could be configured to display recent questions and their corresponding answers. Direction from higher-authority in the form of statements could also be highlighted thus filtering unnecessary noise and providing improved situational awareness with less effort required by the user.

An additional requirement identified by Eovito is Chat Logging, or preserving chat data for historical record [2]. While this may simply involve saving files, we believe that post-processing tasks would benefit from our methodology. By automatically identifying dialog acts and using these new features, we could separate the inherently interleaved conversations thus automatically providing a summary of who said what to whom and when. We believe this information could then be used to generate lessons learned for individuals, units and operational planners.

While our method will require addition of further functionality to achieve these goals, we believe that we provide an enabling foundation for further development.

5.2 Contributions

Our experiments serve to expand the field and include:

- We developed a cross-genre POS tagging methodology. This pushes the field forward in that it was previously known that MLE *within genre* works well; our contribution shoes that MLE cross-genre is effective in the chat domain. We refer to this as "Cheap" POS (or CPOS) tagging. This opens the door for more research in domains where there is little labeled data.

- We further validated the benefits of CPOS tagging by comparing it against hand-tagged POS for dialog act prediction. Our research shows that the extra work required for hand labeling is unnecessary. Simply using pre-existing labeled data from other genres is as effective without the time and cost investment.

- We empirically verified Harris' "Distribution Hypothesis" as applied to emoticons. When we treat emoticons as distinct parts-of-speech, with their own n-gram distributions, our results are better.

- We accomplished significant feature engineering to discover effective combinations of features for dialog act tagging. Further research is needed, but we believe these features will be useful for down-stream analysis.

5.3 Future Work

While we have provided useful results, we recommend the following research with the goal of improving on this foundation:

- For most machine learning techniques, more training data is generally desired. We recommend continuing Forsyth's work in privacy masking and tagging more of the chat corpus. Results for this and other methods would benefit from expanded training data. Our work should prove useful in expanding the size of the NPS chat corpus after anonymizing a larger portion of the raw data collected by Lin.

- Traditional Naïve Bayes classifiers use the formula

$$\hat{C} = \underset{C \in Classes}{\operatorname{argmax}} \log P(C) + \sum_{i} \log P(f_i|C).$$

  In the course of our experimentation, we noted that our classifier determined the correct class within the top two results over 89% of the time using the formula

$$\hat{C} = \underset{C \in Classes}{\operatorname{argmax}} \log P(C) + \sum_{i} \log P(wpp_i|C) + \sum_{j} \log P(pb_j|C).$$

  where wpp_i is word/POS pair i and pb_j is POS bigram j. Our recommendation is an exploration of cascading naïve Bayes results to another classifier in order to improve dialog act tagging accuracy. As noted in Chapter 4, our results decreased slightly when we segregated emoticons that differed in form by tagging them with different parts of speech. Additional training data is needed to determine if this decrease in performance is due to these features similarity or if segregating them reduced our classifiers ability to overcome the widely disparate dialog act class prior probabilities.

- Per Forsyth's recommendations, we showed that emoticons may be better tagged with a POS tag different from "UH" [6]. An exploration of this phenomenon should include other potential tagging schemes for these features.

- The use of this method of dialog act determination should be explored in the tactical military chat domain. Additional effort should be directed to thread extraction in this critical command and control subsystem to provide the functionality proposed above.
- We also believe that our method of dialog act tagging chat posts would translate to similar results on Short Message Service (SMS) data. The popularity of this form of computer mediated communications continues to grow. A corpus of privatized text messages should be constructed for analysis.
- We initially hypothesized that cheap POS tags could be useful in authorship identification. While we performed no work to validate this theory, we believe that it should be explored.

5.4 Final Conclusion

We present a methodology that capitalizes on previous, human-involved POS tagging efforts to effectively determine dialog acts in the chat domain. We hypothesize that methods similar to ours are useful for analysis of emerging domains. This research is an initial foray into the cross-genre POS domain providing a foundation to improve methods in other areas of interest for natural language processing.

APPENDIX A: EMOTICON DICTIONARY

:-)	XD	v.v	:-b	:-/	O:-)	B-)
:)	=D	:-9	:b	:/	0:3	8)
:o)	=3	;-)	:-0	:	O:)	8-)
:]	<=3	;)	:0	=/	:'(	D:<
:3	<=8	*)	0_0	=	;*(	>:(
:c)	:-(	;]	o_o	:S	T_T	D-:<
:>	:(	;D	8O	:—	TT_TT	>:-(
=]	:c	:-P	OwO	d:-)	T.T	:-@
8)	:<	:P	O-O	qB-)	:-*	;(
=)	:[	XP	O-o	:)	:*	' _ '
:D	D:	:-p	O3O	:-)>...	ô)	D<
C:	D8	:p	o0o	:-X	>:)	<3
( )	D;	=p	;o_o;	:X	>;)	<333
:-D	D=	:-	o...o	:=#	>:-)	8D
B)	DX	:	0w0	:#		

Table A.1: Partial Emoticon Dictionary from Wikipedia

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B:
EFFECTS OF CHEAP POS METHOD

This appendix contains tables showing the redistribution of POS tags by our different emoticon tagging schemes.

Figure B.1 shows the distribution of POS tag counts across all dialog act classes as tagged by Forsyth [6] and serves as a baseline for comparison.

Actual POS counts:

	NN	NNP	NNPS	NNS	JJ	JJR	JJS	VB	VBD	VBG	VBN	VBP	VBZ	RB	RBR	RBS	RP	PDT	POS	PRP	PRP$	IN	TO	DT	UH	BES
Statement	2058	1473	12	579	1155	33	20	1230	585	388	151	983	450	1240	23	1	149	3	21	2469	311	1220	378	1214	1330	91
System	956	689	3	150	159	9	2	2233	43	77	49	58	377	133	2	1	37	0	38	143	77	315	91	244	95	13
Greet	110	1127	0	36	31	0	0	25	4	4	2	15	2	27	0	0	9	0	2	41	8	21	27	47	1437	1
Emotion	17	339	0	14	23	0	0	17	0	6	2	9	10	20	0	0	2	0	0	30	3	22	5	10	1195	1
ynQuestion	341	312	3	127	144	2	2	231	59	49	26	249	70	185	5	0	25	0	0	354	25	187	56	259	150	5
whQuestion	226	298	0	43	130	1	2	70	80	66	18	169	88	85	0	0	32	0	5	263	20	152	45	130	132	21
Accept	33	73	0	10	49	0	2	22	14	7	3	42	17	48	0	0	2	0	0	95	3	21	6	35	222	5
Bye	42	90	0	14	18	0	0	38	5	8	3	25	9	33	0	0	7	0	0	38	6	12	10	34	193	0
Emphasis	118	69	1	23	53	2	1	63	22	19	9	51	19	64	2	0	11	1	1	134	23	56	10	56	74	4
Continuer	61	23	1	17	25	2	5	40	13	14	1	34	10	29	1	1	3	0	0	72	6	45	15	45	31	2
Reject	64	56	0	28	27	0	0	81	14	20	2	54	9	87	1	1	12	0	0	93	14	36	10	60	86	2
yAnswer	21	44	0	4	12	0	0	7	12	3	2	18	9	12	1	0	2	0	0	50	5	10	2	20	108	0
nAnswer	20	25	0	3	10	0	0	14	6	3	2	14	3	32	0	0	1	0	0	37	0	13	5	23	66	0
Clarify	9	8	0	6	2	0	0	5	4	1	1	10	1	8	0	0	0	0	0	13	0	0	1	4	16	0
Other	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	1	14	0

	CC	CD	EX	FW	GW	HVS	MD	SYM	WDT	WP	WRB	X	UNK	:	,	.	EMO	EM2	)	(	LS	$	#	WP$	MD*
Statement	413	226	18	7	12	3	296	73	42	63	90	2	0	772	317	339	0	0	18	19	1	0	0	0	0
System	93	234	0	0	1	0	19	645	3	16	4	6	0	290	72	472	0	0	39	40	0	0	0	0	0
Greet	24	11	15	1	2	0	1	9	0	13	7	1	0	52	26	96	0	0	1	1	0	0	0	0	0
Emotion	1	4	0	0	0	0	5	1	0	1	1	1	0	39	10	84	0	0	0	0	0	0	0	0	0
ynQuestion	53	45	4	0	3	0	45	2	1	12	10	0	0	62	32	411	0	0	2	1	0	0	0	0	0
whQuestion	51	11	3	4	0	0	16	0	6	268	210	1	0	64	41	366	0	0	0	0	0	0	0	0	0
Accept	4	5	2	0	0	0	7	4	1	3	3	0	0	43	23	22	0	0	1	1	0	0	0	0	0
Bye	4	1	0	0	0	0	2	4	0	1	2	1	0	30	12	19	0	0	0	0	0	0	0	0	0
Emphasis	7	8	0	0	3	0	14	6	0	6	2	2	0	44	16	173	0	0	0	0	0	0	0	0	0
Continuer	56	13	0	2	0	0	9	1	2	2	4	0	0	23	5	16	0	0	0	0	0	0	0	0	0
Reject	17	2	2	1	2	0	8	0	1	2	0	0	0	27	20	35	0	0	0	0	0	0	0	0	0
yAnswer	4	0	0	0	0	0	5	1	0	0	4	1	0	27	11	13	0	0	0	0	0	0	0	0	0
nAnswer	4	0	0	0	0	0	3	0	0	1	0	0	0	14	10	4	0	0	0	0	0	0	0	0	0
Clarify	1	0	0	0	0	0	1	6	0	0	0	0	0	5	0	1	0	0	0	0	0	0	0	0	0
Other	0	5	0	0	0	0	0	0	0	0	0	14	0	3	2	1	0	0	0	0	0	0	0	0	0

Figure B.1: Actual POS Counts

Figure B.2 shows the distribution of POS tag counts when emoticons are unrecognized and are tagged as "UNK". We can observe the changes in POS tag counts resulting from our cheap POS methodology. Specifically, note the increased size of the UNK category. Shifts in the noun and verb categories are also evident as a result of our maximum likelihood estimation approach to POS tagging.

Cheap POS counts:

	NN	NNP	NNPS	NNS	JJ	JJR	JJS	VB	VBD	VBG	VBN	VBP	VBZ	RB	RBR	RBS	RP	PDT	POS	PRP	PRP$	IN	TO	DT	UH	BES
Statement	2172	1580	4	528	1113	52	50	1248	493	313	252	463	403	1186	8	0	192	0	116	2081	379	1409	384	976	211	0
System	1962	546	2	238	140	11	7	1191	39	64	67	33	219	129	2	0	44	0	52	135	85	314	91	233	12	0
Greet	128	1155	0	31	98	0	0	26	4	1	2	7	2	26	0	0	16	0	4	30	11	22	27	6	969	0
Emotion	37	343	0	18	34	0	2	19	0	6	4	4	6	20	0	0	3	0	2	20	10	22	5	10	41	0
ynQuestion	357	315	2	108	126	7	8	317	57	37	38	75	68	171	0	0	30	0	5	304	32	211	53	228	52	0
whQuestion	229	328	0	41	130	1	4	118	67	54	28	91	85	94	2	0	29	0	20	196	21	248	45	99	34	0
Accept	44	83	0	9	92	0	5	32	15	4	3	16	18	68	0	0	2	0	6	85	6	34	6	17	61	0
Bye	60	127	0	14	35	0	1	39	3	6	5	15	7	26	11	0	11	0	1	30	6	10	10	16	45	0
Emphasis	114	70	1	22	57	4	2	66	20	11	16	24	18	61	1	0	7	0	8	114	28	69	11	44	12	0
Continuer	59	34	2	14	26	3	6	49	10	12	2	11	10	29	0	0	3	0	4	59	7	49	15	36	1	0
Reject	72	77	0	20	29	1	1	69	9	15	12	21	10	89	0	0	19	0	2	78	12	41	10	70	16	0
yAnswer	25	40	0	5	18	1	2	12	9	3	4	11	9	58	0	0	0	0	0	42	11	12	2	14	20	0
nAnswer	20	27	0	4	8	0	0	15	6	2	4	8	2	26	0	0	1	0	0	28	1	16	5	61	10	0
Clarify	11	8	0	4	2	0	0	11	4	0	1	1	0	9	0	0	0	0	0	13	1	0	1	3	4	0
Other	1	4	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0

	CC	CD	EX	FW	GW	HVS	MD	SYM	WDT	WP	WRB	X	UNK	:	,	.	EMO	EMO2	)	(	LS	$	#	WP$	MD*
Statement	313	277	66	13	0	0	254	10	47	0	90	0	2647	256	316	337	0	0	18	19	0	1	0	0	1
System	61	207	3	1	0	0	18	1	12	0	4	0	674	176	72	1025	0	0	26	30	0	0	2	0	0
Greet	14	14	18	3	0	0	3	0	2	0	6	0	491	25	26	67	0	0	1	1	0	0	0	0	0
Emotion	1	4	2	0	0	0	4	0	1	0	1	0	1173	17	10	53	0	0	0	0	0	0	0	0	0
ynQuestion	36	58	15	3	0	0	42	0	5	0	10	0	362	30	32	352	0	0	2	1	0	0	0	0	0
whQuestion	46	15	7	2	0	0	20	0	134	0	197	0	364	18	41	308	0	0	0	0	0	0	0	1	0
Accept	4	6	4	0	0	0	6	0	1	0	3	0	132	17	23	24	0	0	1	1	0	0	0	0	0
Bye	2	2	0	1	0	0	3	0	0	0	2	0	134	11	12	16	0	0	0	0	0	0	0	0	0
Emphasis	6	8	2	0	0	0	12	1	4	0	3	0	205	20	16	109	0	0	0	0	0	0	0	0	1
Continuer	54	14	1	0	0	0	9	0	2	0	4	0	78	3	5	18	0	0	0	0	0	0	0	0	0
Reject	15	4	3	0	0	0	7	0	2	0	0	0	105	14	20	31	0	0	0	0	0	0	0	0	0
yAnswer	4	0	0	0	0	0	5	0	0	0	5	0	62	12	11	11	0	0	0	0	0	0	0	0	0
nAnswer	4	0	0	0	0	0	3	0	0	0	0	0	39	9	10	4	0	0	0	0	0	0	0	0	0
Clarify	1	0	0	0	0	0	1	1	0	0	0	0	25	1	0	1	0	0	0	0	0	0	0	0	0
Other	0	5	0	0	0	0	0	0	0	0	0	0	26	2	2	0	0	0	0	0	0	0	0	0	0

Figure B.2: POS Counts with Emoticons Unrecognized

Figure B.3 shows the changes in the “UH” category as emoticons were moved from the “UNK” POS counts.

Cheap POS counts:

	NN	NNP	NNPS	NNS	JJ	JJR	JJS	VB	VBD	VBG	VBN	VBP	VBZ	RB	RBR	RBS	RP	PDT	POS	PRP	PRP$	IN	TO	DT	UH	BES
Statement	2172	1580	4	528	1113	52	50	1248	493	313	252	463	403	1186	8	0	192	0	116	2081	379	1409	384	976	604	0
System	1962	546	2	238	140	11	7	1191	39	64	67	33	219	129	2	0	44	0	52	135	85	314	91	233	27	0
Greet	128	1155	0	31	98	0	0	26	4	1	2	7	2	26	0	0	16	0	4	30	11	22	27	6	1118	0
Emotion	37	343	0	18	34	0	2	19	0	6	4	4	6	20	0	0	3	0	2	20	10	22	5	10	757	0
ynQuestion	357	315	2	108	126	7	8	317	57	37	38	75	68	171	0	0	30	0	5	304	32	211	53	228	85	0
whQuestion	229	328	0	41	130	1	4	118	67	54	28	91	85	94	2	0	29	0	20	196	21	248	45	99	70	0
Accept	44	83	0	9	92	0	5	32	15	4	3	16	18	68	0	0	2	0	6	85	6	34	6	17	89	0
Bye	60	127	0	14	35	0	1	39	3	6	5	15	7	26	11	0	11	0	1	30	6	10	10	16	104	0
Emphasis	114	70	1	22	57	4	2	66	20	11	16	24	18	61	1	0	7	0	8	114	28	69	11	44	30	0
Continuer	59	34	2	14	26	3	6	49	10	12	2	11	10	29	0	0	3	0	4	59	7	49	15	36	11	0
Reject	72	77	0	20	29	1	1	69	9	15	12	21	10	89	0	0	19	0	2	78	12	41	10	70	25	0
yAnswer	25	40	0	5	18	1	2	12	9	3	4	11	9	58	0	0	0	0	0	42	11	12	2	14	30	0
nAnswer	20	27	0	4	8	0	0	15	6	2	4	8	2	26	0	0	1	0	0	28	1	16	5	61	19	0
Clarify	11	8	0	4	2	0	0	11	4	0	1	1	0	9	0	0	0	0	0	13	1	0	1	3	9	0
Other	1	4	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2	0

	CC	CD	EX	FW	GW	HVS	MD	SYM	WDT	WP	WRB	X	UNK	:	,	.	EMO	EMO2	)	(	LS	$	#	WP$	MD*
Statement	313	277	66	13	0	0	254	10	47	0	90	0	2254	256	316	337	0	0	18	19	0	1	0	0	1
System	61	207	3	1	0	0	18	1	12	0	4	0	659	176	72	1025	0	0	26	30	0	0	2	0	0
Greet	14	14	18	3	0	0	3	0	2	0	6	0	342	25	26	67	0	0	1	1	0	0	0	0	0
Emotion	1	4	2	0	0	0	4	0	1	0	1	0	457	17	10	53	0	0	0	0	0	0	0	0	0
ynQuestion	36	58	15	3	0	0	42	0	5	0	10	0	329	30	32	352	0	0	2	1	0	0	0	0	0
whQuestion	46	15	7	2	0	0	20	0	134	0	197	0	328	18	41	308	0	0	0	0	0	0	0	1	0
Accept	4	6	4	0	0	0	6	0	1	0	3	0	104	17	23	24	0	0	1	1	0	0	0	0	0
Bye	2	2	0	1	0	0	3	0	0	0	2	0	75	11	12	16	0	0	0	0	0	0	0	0	0
Emphasis	6	8	2	0	0	0	12	1	4	0	3	0	187	20	16	109	0	0	0	0	0	0	0	0	1
Continuer	54	14	1	0	0	0	9	0	2	0	4	0	68	3	5	18	0	0	0	0	0	0	0	0	0
Reject	15	4	3	0	0	0	7	0	2	0	0	0	96	14	20	31	0	0	0	0	0	0	0	0	0
yAnswer	4	0	0	0	0	0	5	0	0	0	5	0	52	12	11	11	0	0	0	0	0	0	0	0	0
nAnswer	4	0	0	0	0	0	3	0	0	0	0	0	30	9	10	4	0	0	0	0	0	0	0	0	0
Clarify	1	0	0	0	0	0	1	1	0	0	0	0	20	1	0	1	0	0	0	0	0	0	0	0	0
Other	0	5	0	0	0	0	0	0	0	0	0	0	24	2	2	0	0	0	0	0	0	0	0	0	0

Figure B.3: POS Counts with Emoticons Tagged as Interjections

In Figure B.4, we have tagged all emoticons with the unique “EMO” tag. Note the changes from the “UH” category to the “EMO” column.

Cheap POS counts:

	NN	NNP	NNPS	NNS	JJ	JJR	JJS	VB	VBD	VBG	VBN	VBP	VBZ	RB	RBR	RBS	RP	PDT	POS	PRP	PRP$	IN	TO	DT	UH	BES
Statement	2172	1580	4	528	1113	52	50	1248	493	313	252	463	403	1186	8	0	192	0	116	2081	379	1409	384	976	211	0
System	1962	546	2	238	140	11	7	1191	39	64	67	33	219	129	2	0	44	0	52	135	85	314	91	233	12	0
Greet	128	1155	0	31	98	0	0	26	4	1	2	7	2	26	0	0	16	0	4	30	11	22	27	6	969	0
Emotion	37	343	0	18	34	0	2	19	0	6	4	4	6	20	0	0	3	0	2	20	10	22	5	10	41	0
ynQuestion	357	315	2	108	126	7	8	317	57	37	38	75	68	171	0	0	30	0	5	304	32	211	53	228	52	0
whQuestion	229	328	0	41	130	1	4	118	67	54	28	91	85	94	2	0	29	0	20	196	21	248	45	99	34	0
Accept	44	83	0	9	92	0	5	32	15	4	3	16	18	68	0	0	2	0	6	85	6	34	6	17	61	0
Bye	60	127	0	14	35	0	1	39	3	6	5	15	7	26	11	0	11	0	1	30	6	10	10	16	45	0
Emphasis	114	70	1	22	57	4	2	66	20	11	16	24	18	61	1	0	7	0	8	114	28	69	11	44	12	0
Continuer	59	34	2	14	26	3	6	49	10	12	2	11	10	29	0	0	3	0	4	59	7	49	15	36	1	0
Reject	72	77	0	20	29	1	1	69	9	15	12	21	10	89	0	0	19	0	2	78	12	41	10	70	16	0
yAnswer	25	40	0	5	18	1	2	12	9	3	4	11	9	58	0	0	0	0	0	42	11	12	2	14	20	0
nAnswer	20	27	0	4	8	0	0	15	6	2	4	8	2	26	0	0	1	0	0	28	1	16	5	61	10	0
Clarify	11	8	0	4	2	0	0	11	4	0	1	1	0	9	0	0	0	0	0	13	1	0	1	3	4	0
Other	1	4	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0

	CC	CD	EX	FW	GW	HVS	MD	SYM	WDT	WP	WRB	X	UNK	:	,	.	EMO	EM2	)	(	LS	$	#	WP$	MD*
Statement	313	277	66	13	0	0	254	10	47	0	90	0	2254	256	316	337	393	0	18	19	0	1	0	0	1
System	61	207	3	1	0	0	18	1	12	0	4	0	659	176	72	1025	15	0	26	30	0	0	2	0	0
Greet	14	14	18	3	0	0	3	0	2	0	6	0	342	25	26	67	149	0	1	1	0	0	0	0	0
Emotion	1	4	2	0	0	0	4	0	1	0	1	0	457	17	10	53	716	0	0	0	0	0	0	0	0
ynQuestion	36	58	15	3	0	0	42	0	5	0	10	0	329	30	32	352	33	0	2	1	0	0	0	0	0
whQuestion	46	15	7	2	0	0	20	0	134	0	197	0	328	18	41	308	36	0	0	0	0	0	0	1	0
Accept	4	6	4	0	0	0	6	0	1	0	3	0	104	17	23	24	28	0	1	1	0	0	0	0	0
Bye	2	2	0	1	0	0	3	0	0	0	2	0	75	11	12	16	59	0	0	0	0	0	0	0	0
Emphasis	6	8	2	0	0	0	12	1	4	0	3	0	187	20	16	109	18	0	0	0	0	0	0	0	1
Continuer	54	14	1	0	0	0	9	0	2	0	4	0	68	3	5	18	10	0	0	0	0	0	0	0	0
Reject	15	4	3	0	0	0	7	0	2	0	0	0	96	14	20	31	9	0	0	0	0	0	0	0	0
yAnswer	4	0	0	0	0	0	5	0	0	0	5	0	52	12	11	11	10	0	0	0	0	0	0	0	0
nAnswer	4	0	0	0	0	0	3	0	0	0	0	0	30	9	10	4	9	0	0	0	0	0	0	0	0
Clarify	1	0	0	0	0	0	1	1	0	0	0	0	20	1	0	1	5	0	0	0	0	0	0	0	0
Other	0	5	0	0	0	0	0	0	0	0	0	0	24	2	2	0	2	0	0	0	0	0	0	0	0

Figure B.4: POS Counts with Emoticons Tagged with our EMO Tag

Finally, Figure B.5 displays the changes in POS tag counts when emoticons are separated into two groups.

Cheap POS counts:

	NN	NNP	NNPS	NNS	JJ	JJR	JJS	VB	VBD	VBG	VBN	VBP	VBZ	RB	RBR	RBS	RP	PDT	POS	PRP	PRP$	IN	TO	DT	UH	BES
Statement	2172	1580	4	528	1113	52	50	1248	493	313	252	463	403	1186	8	0	192	0	116	2081	379	1409	384	976	211	0
System	1962	546	2	238	140	11	7	1191	39	64	67	33	219	129	2	0	44	0	52	135	85	314	91	233	12	0
Greet	128	1155	0	31	98	0	0	26	4	1	2	7	2	26	0	0	16	0	4	30	11	22	27	6	969	0
Emotion	37	343	0	18	34	0	2	19	0	6	4	4	6	20	0	0	3	0	2	20	10	22	5	10	41	0
ynQuestion	357	315	2	108	126	7	8	317	57	37	38	75	68	171	0	0	30	0	5	304	32	211	53	228	52	0
whQuestion	229	328	0	41	130	1	4	118	67	54	28	91	85	94	2	0	29	0	20	196	21	248	45	99	34	0
Accept	44	83	0	9	92	0	5	32	15	4	3	16	18	68	0	0	2	0	6	85	6	34	6	17	61	0
Bye	60	127	0	14	35	0	1	39	3	6	5	15	7	26	11	0	11	0	1	30	6	10	10	16	45	0
Emphasis	114	70	1	22	57	4	2	66	20	11	16	24	18	61	1	0	7	0	8	114	28	69	11	44	12	0
Continuer	59	34	2	14	26	3	6	49	10	12	2	11	10	29	0	0	3	0	4	59	7	49	15	36	1	0
Reject	72	77	0	20	29	1	1	69	9	15	12	21	10	89	0	0	19	0	2	78	12	41	10	70	16	0
yAnswer	25	40	0	5	18	1	2	12	9	3	4	11	9	58	0	0	0	0	0	42	11	12	2	14	20	0
nAnswer	20	27	0	4	8	0	0	15	6	2	4	8	2	26	0	0	1	0	0	28	1	16	5	61	10	0
Clarify	11	8	0	4	2	0	0	11	4	0	1	1	0	9	0	0	0	0	0	13	1	0	1	3	4	0
Other	1	4	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0

	CC	CD	EX	FW	GW	HVS	MD	SYM	WDT	WP	WRB	X	UNK	:	,	.	EMO	EM2	)	(	LS	$	#	WP$	MD*
Statement	313	277	66	13	0	0	254	10	47	0	90	0	2254	256	316	337	94	299	18	19	0	1	0	0	1
System	61	207	3	1	0	0	18	1	12	0	4	0	659	176	72	1025	2	13	26	30	0	0	2	0	0
Greet	14	14	18	3	0	0	3	0	2	0	6	0	342	25	26	67	41	108	1	1	0	0	0	0	0
Emotion	1	4	2	0	0	0	4	0	1	0	1	0	457	17	10	53	102	614	0	0	0	0	0	0	0
ynQuestion	36	58	15	3	0	0	42	0	5	0	10	0	329	30	32	352	6	27	2	1	0	0	0	0	0
whQuestion	46	15	7	2	0	0	20	0	134	0	197	0	328	18	41	308	4	32	0	0	0	0	0	1	0
Accept	4	6	4	0	0	0	6	0	1	0	3	0	104	17	23	24	7	21	1	1	0	0	0	0	0
Bye	2	2	0	1	0	0	3	0	0	0	2	0	75	11	12	16	2	57	0	0	0	0	0	0	0
Emphasis	6	8	2	0	0	0	12	1	4	0	3	0	187	20	16	109	3	15	0	0	0	0	0	0	1
Continuer	54	14	1	0	0	0	9	0	2	0	4	0	68	3	5	18	2	8	0	0	0	0	0	0	0
Reject	15	4	3	0	0	0	7	0	2	0	0	0	96	14	20	31	2	7	0	0	0	0	0	0	0
yAnswer	4	0	0	0	0	0	5	0	0	0	5	0	52	12	11	11	0	10	0	0	0	0	0	0	0
nAnswer	4	0	0	0	0	0	3	0	0	0	0	0	30	9	10	4	1	8	0	0	0	0	0	0	0
Clarify	1	0	0	0	0	0	1	1	0	0	0	0	20	1	0	1	1	4	0	0	0	0	0	0	0
Other	0	5	0	0	0	0	0	0	0	0	0	0	24	2	2	0	0	2	0	0	0	0	0	0	0

Figure B.5: POS Counts with Emoticons Separated into Two Groups

APPENDIX C:
CONFUSION MATRICES

This appendix contains confusion matrices for selected experiments. These are separated by specific emoticon tagging schemes and experiment numbers as found in the caption of each table.

Figures C.1 through C.10 show the results of corresponding experiment runs with emoticons unrecognized (tagged as “UNK”).

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**301**	4	4	11	10	4	6	6	13	11	6	3	2	2	1	0.7839	0.8853	0.8315	85.41%
System	1	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9821	0.9892	
Greet	14	0	**118**	2	0	0	0	1	0	0	0	1	0	0	0	0.8676	0.9593	0.9112	
Emotion	2	0	0	**97**	0	0	0	0	1	1	0	0	0	0	0	0.9604	0.8584	0.9065	
ynQuestion	4	1	1	0	**43**	6	0	0	1	0	1	0	0	0	0	0.7544	0.7288	0.7414	
whQuestion	1	0	0	0	6	**47**	0	0	1	0	0	0	0	0	0	0.8545	0.8246	0.8393	
Accept	6	0	0	1	0	0	**12**	0	0	0	0	2	0	0	0	0.5714	0.5714	0.5714	
Bye	1	0	0	1	0	0	0	**18**	1	0	0	0	0	0	0	0.8571	0.7200	0.7826	
Emphasis	2	0	0	1	0	0	1	0	**5**	0	0	0	0	0	0	0.5556	0.2174	0.3125	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	1	0	0.5714	0.2353	0.3333	
Reject	6	0	0	0	0	0	1	0	1	1	**3**	0	1	0	0	0.2308	0.3000	0.2609	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**5**	0	0	0	0.8333	0.4167	0.5556	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.4000	0.5714	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**297**	1	13	1	11	0	5	1	3	2	5	0	0	0	1	0.8735	0.7655	0.8159	84.59%
System	3	**277**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9893	0.9893	0.9893	
Greet	5	0	**117**	0	1	0	0	0	0	0	0	0	0	0	0	0.9512	0.8797	0.9141	
Emotion	13	2	1	**96**	0	0	0	0	1	0	0	0	0	0	0	0.8496	0.9697	0.9057	
ynQuestion	14	0	0	0	**38**	6	0	0	0	1	0	0	0	0	0	0.6441	0.6909	0.6667	
whQuestion	4	0	1	0	4	**48**	0	0	0	0	0	0	0	0	0	0.8421	0.8889	0.8649	
Accept	7	0	0	0	0	0	**11**	0	1	0	0	1	1	0	0	0.5238	0.5500	0.5366	
Bye	5	0	1	0	0	0	0	**19**	0	0	0	0	0	0	0	0.7600	0.9048	0.8261	
Emphasis	13	0	0	1	1	0	0	1	**5**	1	1	0	0	0	0	0.2174	0.5000	0.3030	
Continuer	11	0	0	1	0	0	0	0	0	**4**	1	0	0	0	0	0.2353	0.4000	0.2963	
Reject	8	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0	0.2000	0.2222	0.2105	
yAnswer	3	0	0	0	0	0	4	0	0	1	0	**4**	0	0	0	0.3333	0.8000	0.4706	
nAnswer	2	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0.6000	0.7500	0.6667	
Clarify	2	0	0	0	0	0	0	0	0	1	0	0	0	**0**	0	0.0000	undef	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Figure C.1: Experiment Run 5: Emoticons Unrecognized

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	3	5	15	11	6	9	4	10	7	14	8	3	6	0	0.7363	0.8650	0.7955	83.81%
System	0	**249**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9960	0.9881	0.9920	
Greet	12	0	**136**	2	0	1	0	0	0	1	0	0	0	0	0	0.8947	0.9444	0.9189	
Emotion	3	0	1	**93**	0	0	0	0	0	0	0	0	0	0	0	0.9588	0.8378	0.8942	
ynQuestion	6	0	1	0	**36**	4	0	0	0	0	0	1	0	0	0	0.7500	0.7059	0.7273	
whQuestion	3	0	0	1	1	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.7755	0.8261	
Accept	9	0	0	0	0	0	**17**	0	0	0	0	3	0	0	0	0.5862	0.6296	0.6071	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.7500	0.8571	
Emphasis	4	0	0	0	2	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.3125	0.3571	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	1.0000	0.2353	0.3810	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.2500	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**278**	4	4	16	11	3	8	4	13	7	12	8	2	6	1	0.7374	0.8528	0.7909	83.33%
System	2	**248**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9880	0.9841	0.9861	
Greet	9	0	**136**	0	0	1	0	0	0	1	0	0	0	0	0	0.9252	0.9444	0.9347	
Emotion	4	0	2	**94**	0	0	1	0	0	0	0	1	0	0	1	0.9126	0.8468	0.8785	
ynQuestion	6	0	1	0	**35**	3	0	0	0	0	0	1	0	0	0	0.7609	0.6863	0.7216	
whQuestion	4	0	0	1	3	**42**	0	0	0	0	0	0	0	0	0	0.8400	0.8571	0.8485	
Accept	9	0	0	0	0	0	**16**	0	0	0	0	2	0	0	0	0.5926	0.5926	0.5926	
Bye	2	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0.8571	0.7500	0.8000	
Emphasis	4	0	0	0	1	0	0	0	**2**	0	0	1	0	0	0	0.2500	0.1250	0.1667	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**5**	0	3	0	0	0.4545	0.2778	0.3448	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**4**	0	0	0	0.8000	0.2353	0.3636	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3750	0.4615	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	1.0000	0.5000	0.6667	

Figure C.2: Experiment Run 10: Emoticons Unrecognized

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**265**	6	8	21	11	5	5	4	8	9	11	3	4	0	0	0.7361	0.8833	0.8030	83.58%
System	1	**235**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9874	0.9711	0.9792	
Greet	8	0	**119**	1	1	0	0	2	0	0	0	0	0	0	0	0.9084	0.9084	0.9084	
Emotion	3	0	0	**90**	0	0	1	0	0	0	0	0	0	0	0	0.9574	0.7826	0.8612	
ynQuestion	9	1	1	0	**42**	3	0	0	0	0	0	0	0	0	0	0.7500	0.6885	0.7179	
whQuestion	2	0	0	1	5	**40**	0	0	1	0	0	0	0	0	0	0.8163	0.8333	0.8247	
Accept	6	0	0	1	0	0	**6**	0	0	1	1	0	0	0	0	0.4000	0.3750	0.3871	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.6667	0.8000	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4500	0.5455	
Continuer	2	0	0	0	2	0	0	0	1	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	1	0	0	0	0	0	1	0	1	0	**5**	0	1	0	0	0.5556	0.2778	0.3704	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**6**	0	0	0	0.6000	0.6667	0.6316	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**4**	0	0	0.6667	0.4444	0.5333	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**256**	4	5	24	16	6	4	4	11	9	9	4	3	0	0	0.7211	0.8533	0.7817	82.49%
System	2	**238**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9835	0.9835	0.9835	
Greet	6	0	**122**	0	1	0	0	1	0	0	0	0	0	0	0	0.9385	0.9313	0.9349	
Emotion	5	0	0	**89**	0	0	1	0	0	0	0	0	0	0	0	0.9368	0.7739	0.8476	
ynQuestion	11	0	1	0	**36**	1	0	0	0	0	0	0	0	0	0	0.7347	0.5902	0.6545	
whQuestion	4	0	0	1	7	**40**	0	0	1	0	0	1	0	0	0	0.7407	0.8333	0.7843	
Accept	7	0	0	0	0	0	**7**	0	0	1	1	2	0	0	0	0.3889	0.4375	0.4118	
Bye	0	0	0	0	0	0	0	**13**	0	0	0	0	0	0	0	1.0000	0.7222	0.8387	
Emphasis	2	0	1	1	0	0	0	0	**7**	0	0	0	0	0	0	0.6364	0.3500	0.4516	
Continuer	4	0	0	0	1	0	0	0	0	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	2	0	0	0	0	1	1	0	1	0	**7**	0	1	0	0	0.5385	0.3889	0.4516	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2222	0.2667	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**5**	0	0	0.7143	0.5556	0.6250	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	1.0000	1.0000	

Figure C.3: Experiment Run 15: Emoticons Unrecognized

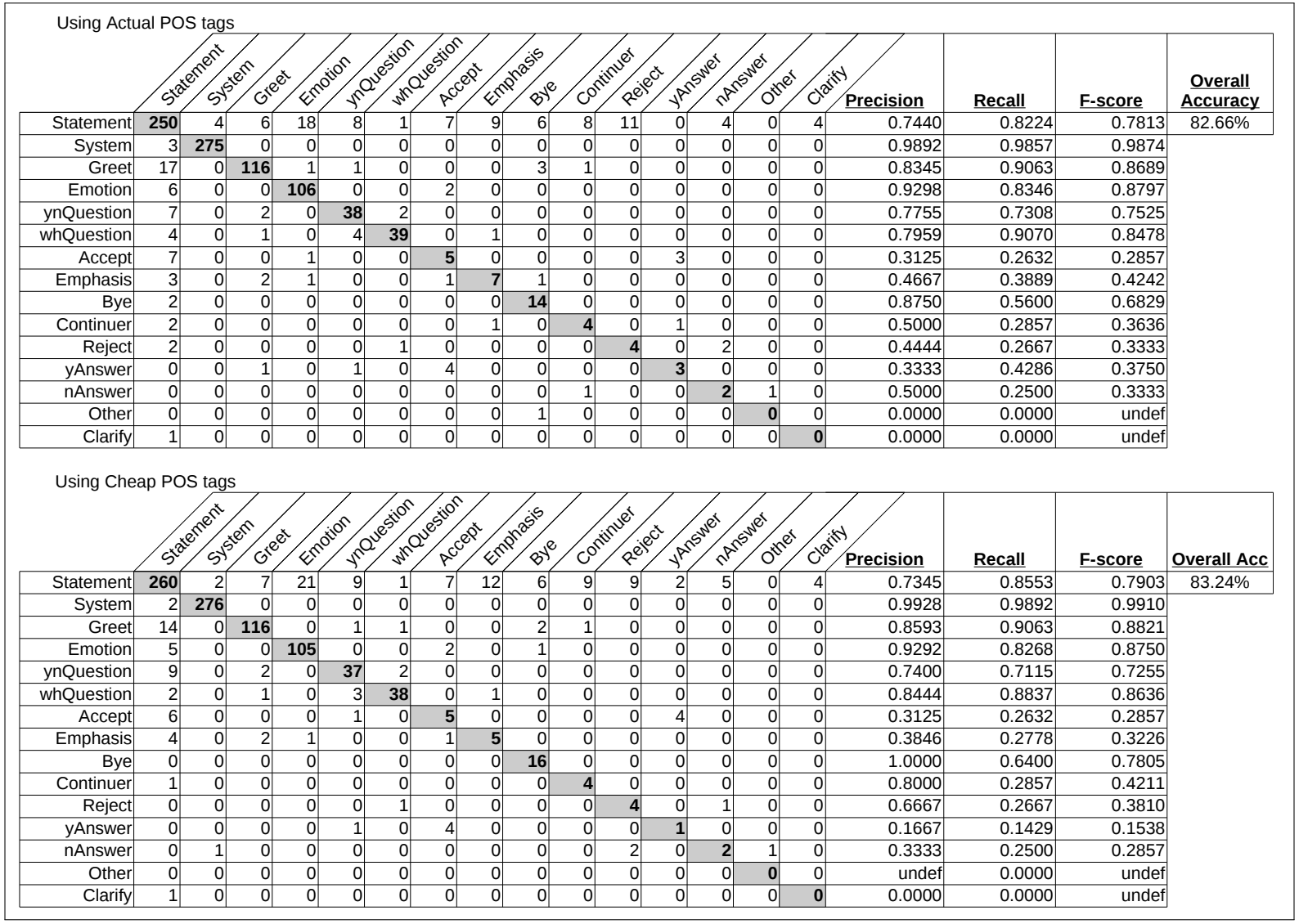
Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	4	8	17	10	4	8	5	13	9	6	2	1	0	1	0.7622	0.8952	0.8234	84.71%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	11	0	**131**	4	0	1	0	1	0	1	1	0	0	0	0	0.8733	0.8973	0.8851	
Emotion	1	0	1	**81**	0	0	2	0	0	0	0	0	0	0	0	0.9529	0.7714	0.8526	
ynQuestion	2	0	2	0	**37**	1	0	0	1	0	0	0	0	0	0	0.8605	0.7400	0.7957	
whQuestion	2	0	0	0	3	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.8636	0.8736	
Accept	5	0	0	1	0	0	**15**	0	1	0	3	0	0	0	0	0.6000	0.5556	0.5769	
Bye	1	0	1	0	0	0	1	**13**	0	0	0	0	0	0	0	0.8125	0.6842	0.7429	
Emphasis	2	0	2	1	0	0	0	0	**11**	0	1	0	0	0	0	0.6471	0.4074	0.5000	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	1	0	0	1	0	0	1	0	0	0	0	**6**	0	0	0	0.6667	0.6000	0.6316	
nAnswer	0	0	0	0	0	0	0	0	0	1	2	0	**1**	0	0	0.2500	0.5000	0.3333	
Clarify	1	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.8000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**276**	4	9	21	8	4	8	5	16	10	5	3	1	0	3	0.7399	0.8762	0.8023	82.88%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	6	0	**131**	2	0	1	0	0	0	1	0	0	0	0	0	0.9291	0.8973	0.9129	
Emotion	2	0	1	**78**	0	0	1	0	0	0	1	0	0	0	0	0.9398	0.7429	0.8298	
ynQuestion	6	0	1	0	**33**	1	0	0	1	0	0	0	0	0	1	0.7674	0.6600	0.7097	
whQuestion	3	0	0	1	9	**37**	0	0	0	0	0	0	0	0	0	0.7400	0.8409	0.7872	
Accept	7	0	0	0	0	0	**17**	0	1	0	3	1	0	0	0	0.5862	0.6296	0.6071	
Bye	0	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.9333	0.7368	0.8235	
Emphasis	3	0	2	2	0	0	0	0	**8**	0	0	0	0	0	0	0.5333	0.2963	0.3810	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	2	0	0	1	0	1	1	0	0	0	0	**4**	0	0	0	0.4444	0.4000	0.4211	
nAnswer	3	0	0	0	0	0	0	0	0	0	4	0	**1**	0	0	0.1250	0.5000	0.2000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.2000	0.3333	

Figure C.4: Experiment Run 20: Emoticons Unrecognized



Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**250**	4	6	18	8	1	7	9	6	8	11	0	4	0	4	0.7440	0.8224	0.7813	82.66%
System	3	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9892	0.9857	0.9874	
Greet	17	0	**116**	1	1	0	0	0	3	1	0	0	0	0	0	0.8345	0.9063	0.8689	
Emotion	6	0	0	**106**	0	0	2	0	0	0	0	0	0	0	0	0.9298	0.8346	0.8797	
ynQuestion	7	0	2	0	**38**	2	0	0	0	0	0	0	0	0	0	0.7755	0.7308	0.7525	
whQuestion	4	0	1	0	4	**39**	0	1	0	0	0	0	0	0	0	0.7959	0.9070	0.8478	
Accept	7	0	0	1	0	0	**5**	0	0	0	0	3	0	0	0	0.3125	0.2632	0.2857	
Emphasis	3	0	2	1	0	0	1	**7**	1	0	0	0	0	0	0	0.4667	0.3889	0.4242	
Bye	2	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0.8750	0.5600	0.6829	
Continuer	2	0	0	0	0	0	0	1	0	**4**	0	1	0	0	0	0.5000	0.2857	0.3636	
Reject	2	0	0	0	0	1	0	0	0	0	**4**	0	2	0	0	0.4444	0.2667	0.3333	
yAnswer	0	0	1	0	1	0	4	0	0	0	0	**3**	0	0	0	0.3333	0.4286	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	1	0	0	**2**	1	0	0.5000	0.2500	0.3333	
Other	0	0	0	0	0	0	0	0	1	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**260**	2	7	21	9	1	7	12	6	9	9	2	5	0	4	0.7345	0.8553	0.7903	83.24%
System	2	**276**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9928	0.9892	0.9910	
Greet	14	0	**116**	0	1	1	0	0	2	1	0	0	0	0	0	0.8593	0.9063	0.8821	
Emotion	5	0	0	**105**	0	0	2	0	1	0	0	0	0	0	0	0.9292	0.8268	0.8750	
ynQuestion	9	0	2	0	**37**	2	0	0	0	0	0	0	0	0	0	0.7400	0.7115	0.7255	
whQuestion	2	0	1	0	3	**38**	0	1	0	0	0	0	0	0	0	0.8444	0.8837	0.8636	
Accept	6	0	0	0	1	0	**5**	0	0	0	0	4	0	0	0	0.3125	0.2632	0.2857	
Emphasis	4	0	2	1	0	0	1	**5**	0	0	0	0	0	0	0	0.3846	0.2778	0.3226	
Bye	0	0	0	0	0	0	0	0	**16**	0	0	0	0	0	0	1.0000	0.6400	0.7805	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.8000	0.2857	0.4211	
Reject	0	0	0	0	0	1	0	0	0	0	**4**	0	1	0	0	0.6667	0.2667	0.3810	
yAnswer	0	0	0	0	1	0	4	0	0	0	0	**1**	0	0	0	0.1667	0.1429	0.1538	
nAnswer	0	1	0	0	0	0	0	0	0	0	2	0	**2**	1	0	0.3333	0.2500	0.2857	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.5: Experiment Run 25: Emoticons Unrecognized

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**288**	3	7	20	9	6	16	7	7	12	13	1	1	0	1	0.7366	0.9057	0.8124	83.64%
System	1	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9891	0.9927	
Greet	12	0	**141**	1	0	2	1	0	1	1	0	0	0	0	0	0.8868	0.9400	0.9126	
Emotion	3	0	0	**72**	0	0	0	0	0	0	1	0	0	0	0	0.9474	0.7423	0.8324	
ynQuestion	4	0	1	0	**34**	6	0	0	0	1	0	0	0	0	0	0.7391	0.6667	0.7010	
whQuestion	0	0	0	0	7	**45**	0	0	0	0	0	1	0	0	0	0.8491	0.7627	0.8036	
Accept	3	0	0	2	1	0	**7**	0	0	0	2	1	0	0	0	0.4375	0.2593	0.3256	
Emphasis	4	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.3000	0.3000	0.3000	
Bye	0	0	1	0	0	0	1	0	**14**	0	0	0	0	0	0	0.8750	0.6087	0.7179	
Continuer	1	0	0	0	0	0	0	0	0	**5**	0	1	0	0	0	0.7143	0.2500	0.3704	
Reject	2	0	0	1	0	0	0	0	0	0	**3**	1	2	0	0	0.3333	0.1579	0.2143	
yAnswer	0	0	0	0	0	0	2	0	0	0	0	**3**	0	0	0	0.6000	0.3750	0.4615	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	1	0.6667	0.4000	0.5000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	1.0000	1.0000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.6667	0.8000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**284**	4	10	20	11	8	17	7	6	9	11	4	1	1	4	0.7154	0.8931	0.7944	82.34%
System	1	**272**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9963	0.9855	0.9909	
Greet	8	0	**136**	0	0	1	1	0	0	1	0	0	0	0	0	0.9252	0.9067	0.9158	
Emotion	4	0	2	**74**	0	0	0	0	1	0	0	0	0	0	0	0.9136	0.7629	0.8315	
ynQuestion	7	0	1	0	**30**	4	0	0	0	1	1	0	0	0	0	0.6818	0.5882	0.6316	
whQuestion	0	0	0	0	9	**46**	0	0	0	1	0	1	1	0	0	0.7931	0.7797	0.7863	
Accept	5	0	0	1	1	0	**7**	0	0	0	2	1	0	0	0	0.4118	0.2593	0.3182	
Emphasis	5	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.2727	0.3000	0.2857	
Bye	1	0	1	0	0	0	0	0	**15**	0	0	0	0	0	0	0.8824	0.6522	0.7500	
Continuer	2	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0.7500	0.3000	0.4286	
Reject	1	0	0	1	0	0	0	0	0	0	**3**	0	1	0	0	0.5000	0.1579	0.2400	
yAnswer	0	0	0	0	0	0	2	0	0	1	0	**2**	0	0	0	0.4000	0.2500	0.3077	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**2**	0	1	0.4000	0.4000	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.1667	0.2857	

Figure C.6: Experiment Run 30: Emoticons Unrecognized

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**263**	4	0	20	8	7	17	5	9	9	12	2	4	4	0	0.7225	0.8946	0.7994	83.72%
System	2	**298**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9933	0.9868	0.9900	
Greet	8	0	**114**	0	0	0	0	0	2	2	1	0	0	0	1	0.8906	0.9421	0.9157	
Emotion	4	0	4	**84**	0	0	0	0	0	0	0	0	0	0	0	0.9130	0.7778	0.8400	
ynQuestion	6	0	0	0	**39**	4	0	0	0	1	0	1	0	0	0	0.7647	0.7647	0.7647	
whQuestion	1	0	0	0	1	**52**	0	0	0	0	0	0	0	0	0	0.9630	0.8254	0.8889	
Accept	7	0	0	0	1	0	**8**	0	1	0	2	3	0	0	0	0.3636	0.2963	0.3265	
Bye	0	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.9333	0.7368	0.8235	
Emphasis	0	0	2	4	1	0	0	0	**6**	0	0	1	0	0	0	0.4286	0.3000	0.3529	
Continuer	0	0	0	0	0	0	1	0	1	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**3**	0	3	0	0	0.3333	0.1667	0.2222	
yAnswer	1	0	0	0	0	0	1	0	0	0	0	**3**	0	0	0	0.6000	0.2727	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	1.0000	0.3000	0.4615	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**269**	3	0	18	7	9	15	4	11	8	10	2	4	4	1	0.7370	0.9150	0.8164	84.47%
System	0	**299**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9967	0.9901	0.9934	
Greet	2	0	**115**	2	0	1	0	0	0	2	0	0	0	0	0	0.9426	0.9504	0.9465	
Emotion	5	0	3	**83**	0	0	0	0	2	0	1	0	0	0	0	0.8830	0.7685	0.8218	
ynQuestion	6	0	0	0	**38**	3	0	0	0	2	0	1	0	0	0	0.7600	0.7451	0.7525	
whQuestion	1	0	0	0	4	**50**	0	0	0	0	0	0	0	0	0	0.9091	0.7937	0.8475	
Accept	5	0	0	0	1	0	**11**	0	1	0	2	3	0	0	0	0.4783	0.4074	0.4400	
Bye	1	0	1	0	0	0	0	**15**	0	0	0	0	0	0	0	0.8824	0.7895	0.8333	
Emphasis	0	0	2	4	0	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.2500	0.3125	
Continuer	1	0	0	0	0	0	1	0	0	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	1	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0.7500	0.2727	0.4000	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3000	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Figure C.7: Experiment Run 35: Emoticons Unrecognized

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**279**	3	7	10	15	3	10	4	8	7	16	4	1	0	6	0.7480	0.8532	0.7971	83.66%
System	3	**264**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9814	0.9814	0.9814	
Greet	16	0	**131**	1	0	0	0	0	0	0	1	0	0	0	0	0.8792	0.9034	0.8912	
Emotion	6	0	1	**99**	0	0	0	0	1	0	0	0	0	1	1	0.9083	0.8839	0.8959	
ynQuestion	4	2	1	0	**44**	1	0	0	1	0	0	0	0	0	0	0.8302	0.7097	0.7652	
whQuestion	1	0	1	0	2	**44**	0	0	1	0	0	1	0	0	0	0.8800	0.8980	0.8889	
Accept	6	0	0	0	1	0	**13**	0	1	1	1	2	0	0	0	0.5200	0.5200	0.5200	
Bye	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0	1.0000	0.7778	0.8750	
Emphasis	3	0	2	2	0	0	0	0	**6**	0	0	0	0	0	0	0.4615	0.3158	0.3750	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	3	0	0	0	0	0	0	0	1	0	**2**	0	1	0	0	0.2857	0.0952	0.1429	
yAnswer	1	0	0	0	0	1	2	0	0	0	0	**5**	0	0	0	0.5556	0.4167	0.4762	
nAnswer	0	0	0	0	0	0	0	0	0	1	1	0	**0**	0	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	3	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**276**	2	5	10	16	4	10	3	11	8	15	5	0	0	7	0.7419	0.8440	0.7897	83.01%
System	2	**265**	3	1	0	0	0	0	0	0	0	0	0	0	0	0.9779	0.9851	0.9815	
Greet	11	0	**130**	0	0	0	0	0	0	0	1	0	0	0	0	0.9155	0.8966	0.9059	
Emotion	8	0	2	**100**	0	0	1	0	1	0	0	0	0	1	0	0.8850	0.8929	0.8889	
ynQuestion	10	1	3	0	**41**	2	0	0	1	0	0	0	0	0	0	0.7069	0.6613	0.6833	
whQuestion	0	1	0	0	4	**42**	0	0	0	0	0	1	0	0	0	0.8750	0.8571	0.8660	
Accept	11	0	0	0	1	0	**12**	0	1	1	1	2	0	0	0	0.4138	0.4800	0.4444	
Bye	1	0	0	0	0	0	1	**15**	0	0	0	0	0	0	0	0.8824	0.8333	0.8571	
Emphasis	2	0	2	1	0	0	0	0	**5**	0	0	0	0	0	0	0.5000	0.2632	0.3448	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	1	0	0	0	0	0	0	0	0	0	**3**	0	1	0	0	0.6000	0.1429	0.2308	
yAnswer	0	0	0	0	0	1	1	0	0	0	0	**4**	0	0	0	0.6667	0.3333	0.4444	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**1**	0	0	0.3333	0.5000	0.4000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	2	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.8: Experiment Run 40: Emoticons Unrecognized

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**285**	4	7	19	8	6	5	4	7	11	11	3	2	1	0	0.7641	0.8559	0.8074	83.77%
System	4	**270**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9818	0.9854	0.9836	
Greet	11	0	**109**	2	0	1	0	0	0	1	0	0	0	0	0	0.8790	0.8790	0.8790	
Emotion	4	0	2	**94**	0	0	0	0	1	0	0	0	0	0	0	0.9307	0.7899	0.8545	
ynQuestion	10	0	1	0	**35**	2	0	0	0	0	0	0	0	1	0	0.7143	0.7609	0.7368	
whQuestion	0	0	3	0	2	**48**	0	0	0	1	0	0	0	0	0	0.8889	0.8421	0.8649	
Accept	8	0	0	0	0	0	**14**	0	0	0	0	2	1	0	0	0.5600	0.6087	0.5833	
Bye	1	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	0.9412	0.7273	0.8205	
Emphasis	2	0	2	2	0	0	0	0	**8**	1	1	0	0	0	0	0.5000	0.4706	0.4848	
Continuer	4	0	0	0	1	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	1	0	0	0	0	0	0	0	1	0	**1**	0	0	0	0	0.3333	0.0714	0.1176	
yAnswer	0	0	0	1	0	0	3	0	0	1	0	**3**	0	0	0	0.3750	0.3750	0.3750	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	2	0	0	0	0	0	0	1	0	0	0	0	0	**1**	0	0.2500	0.3333	0.2857	
Other	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**4**	0.8000	1.0000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**289**	4	9	22	9	3	5	5	8	10	12	4	2	2	2	0.7487	0.8679	0.8039	83.02%
System	2	**270**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9926	0.9854	0.9890	
Greet	8	0	**110**	1	0	1	0	0	0	1	0	0	0	0	0	0.9091	0.8871	0.8980	
Emotion	5	0	1	**90**	0	0	0	0	1	0	0	0	0	0	0	0.9278	0.7563	0.8333	
ynQuestion	11	0	0	0	**29**	1	0	0	0	1	0	0	0	1	0	0.6744	0.6304	0.6517	
whQuestion	1	0	2	0	8	**52**	0	0	0	1	0	0	0	0	0	0.8125	0.9123	0.8595	
Accept	8	0	0	0	0	0	**14**	0	1	0	0	2	0	0	0	0.5600	0.6087	0.5833	
Bye	0	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	1.0000	0.7273	0.8421	
Emphasis	2	0	2	3	0	0	0	0	**6**	1	1	0	0	0	0	0.4000	0.3529	0.3750	
Continuer	4	0	0	1	0	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	2	0	0	0	0	0	0	0	1	0	**0**	0	1	0	0	0.0000	0.0000	undef	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2500	0.2857	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	0	0	0	0	0	0	0	1	0	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Other	0	0	0	2	0	0	0	0	0	0	0	0	0	0	**2**	0.5000	0.5000	0.5000	

Figure C.9: Experiment Run 45: Emoticons Unrecognized

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**266**	4	7	13	11	3	6	7	9	11	7	4	4	1	2	0.7493	0.8721	0.8061	83.09%
System	0	**238**	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000	0.9835	0.9917	
Greet	10	0	**131**	0	0	0	0	0	0	1	0	1	0	0	1	0.9097	0.9161	0.9129	
Emotion	3	0	1	**92**	0	0	1	0	1	0	0	0	1	0	0	0.9293	0.8288	0.8762	
ynQuestion	2	0	1	0	**29**	6	0	0	1	0	0	0	1	1	0	0.7073	0.6744	0.6905	
whQuestion	1	0	2	0	3	**57**	0	0	0	0	0	0	0	0	0	0.9048	0.8382	0.8702	
Accept	8	0	0	1	0	0	**5**	0	0	0	2	4	0	0	0	0.2500	0.3333	0.2857	
Bye	0	0	0	1	0	0	0	**10**	0	0	0	0	0	0	0	0.9091	0.5556	0.6897	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4286	0.5294	
Continuer	4	0	0	0	0	0	1	0	0	**5**	0	0	0	0	0	0.5000	0.2941	0.3704	
Reject	6	0	0	0	0	1	0	1	0	0	**3**	0	2	0	0	0.2308	0.2143	0.2222	
yAnswer	0	0	0	0	0	1	2	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**3**	0	0	0.6000	0.2727	0.3750	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.3333	0.4000	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**266**	5	7	16	12	3	5	7	11	9	8	4	4	1	3	0.7368	0.8721	0.7988	82.41%
System	0	**237**	0	0	0	0	0	0	0	1	0	0	0	0	0	0.9958	0.9793	0.9875	
Greet	7	0	**131**	0	0	1	0	0	0	1	0	0	0	0	0	0.9357	0.9161	0.9258	
Emotion	5	0	1	**90**	0	0	1	0	1	0	0	0	1	0	0	0.9091	0.8108	0.8571	
ynQuestion	4	0	2	0	**25**	4	0	0	1	0	0	0	0	1	0	0.6757	0.5814	0.6250	
whQuestion	1	0	1	0	6	**57**	0	1	0	0	0	0	0	0	0	0.8636	0.8382	0.8507	
Accept	9	0	0	1	0	0	**7**	0	0	0	2	5	0	0	0	0.2917	0.4667	0.3590	
Bye	0	0	0	0	0	0	0	**10**	0	0	0	0	0	0	0	1.0000	0.5556	0.7143	
Emphasis	2	0	1	1	0	0	0	0	**7**	0	1	0	0	0	0	0.5833	0.3333	0.4242	
Continuer	5	0	0	0	0	0	1	0	0	**6**	0	0	0	0	0	0.5000	0.3529	0.4138	
Reject	3	0	0	0	0	1	0	0	0	0	**2**	0	3	0	0	0.2222	0.1429	0.1739	
yAnswer	0	0	0	0	0	2	1	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.7500	0.2727	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.3333	0.4000	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

Figure C.10: Experiment Run 50: Emoticons Unrecognized

Figures C.11 through C.20 show the results of corresponding experiment runs with emoticons tagged as “UH” per Forsyth’s methodology.

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**301**	4	4	11	10	4	6	6	13	11	6	3	2	2	1	0.7839	0.8853	0.8315	85.41%
System	1	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9821	0.9892	
Greet	14	0	**118**	2	0	0	0	1	0	0	0	1	0	0	0	0.8676	0.9593	0.9112	
Emotion	2	0	0	**97**	0	0	0	0	1	1	0	0	0	0	0	0.9604	0.8584	0.9065	
ynQuestion	4	1	1	0	**43**	6	0	0	1	0	1	0	0	0	0	0.7544	0.7288	0.7414	
whQuestion	1	0	0	0	6	**47**	0	0	1	0	0	0	0	0	0	0.8545	0.8246	0.8393	
Accept	6	0	0	1	0	0	**12**	0	0	0	0	2	0	0	0	0.5714	0.5714	0.5714	
Bye	1	0	0	1	0	0	0	**18**	1	0	0	0	0	0	0	0.8571	0.7200	0.7826	
Emphasis	2	0	0	1	0	0	1	0	**5**	0	0	0	0	0	0	0.5556	0.2174	0.3125	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	1	0	0.5714	0.2353	0.3333	
Reject	6	0	0	0	0	0	1	0	1	1	**3**	0	1	0	0	0.2308	0.3000	0.2609	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**5**	0	0	0	0.8333	0.4167	0.5556	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.4000	0.5714	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**297**	4	5	12	14	4	7	4	14	11	8	3	2	2	1	0.7655	0.8735	0.8159	84.22%
System	1	**276**	0	2	0	0	0	0	0	0	0	0	0	0	0	0.9892	0.9857	0.9875	
Greet	13	0	**117**	3	0	1	0	1	0	0	0	0	0	0	0	0.8667	0.9512	0.9070	
Emotion	2	0	0	**94**	0	0	0	0	1	1	0	0	0	0	0	0.9592	0.8319	0.8910	
ynQuestion	9	0	1	0	**38**	4	0	0	1	0	0	0	0	0	0	0.7170	0.6441	0.6786	
whQuestion	0	0	0	0	6	**48**	0	0	0	0	0	0	0	0	0	0.8889	0.8421	0.8649	
Accept	6	0	0	0	0	0	**11**	0	0	0	0	4	0	0	0	0.5238	0.5238	0.5238	
Bye	1	0	0	0	0	0	0	**19**	1	0	0	0	0	0	0	0.9048	0.7600	0.8261	
Emphasis	3	0	0	2	0	0	1	0	**4**	0	0	0	0	0	0	0.4000	0.1739	0.2424	
Continuer	2	0	0	0	1	0	0	0	1	**4**	0	1	0	1	0	0.4000	0.2353	0.2963	
Reject	5	0	0	0	0	0	0	0	1	1	**2**	0	0	0	0	0.2222	0.2000	0.2105	
yAnswer	0	0	0	0	0	0	1	1	0	0	0	**4**	0	0	0	0.6667	0.3333	0.4444	
nAnswer	0	0	0	0	0	0	1	0	0	0	0	0	**3**	0	0	0.7500	0.6000	0.6667	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Figure C.11: Experiment Run 5: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	3	5	15	11	6	9	4	10	7	14	8	3	6	0	0.7363	0.8650	0.7955	83.81%
System	0	**249**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9960	0.9881	0.9920	
Greet	12	0	**136**	2	0	1	0	0	0	1	0	0	0	0	0	0.8947	0.9444	0.9189	
Emotion	3	0	1	**93**	0	0	0	0	0	0	0	0	0	0	0	0.9588	0.8378	0.8942	
ynQuestion	6	0	1	0	**36**	4	0	0	0	0	0	1	0	0	0	0.7500	0.7059	0.7273	
whQuestion	3	0	0	1	1	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.7755	0.8261	
Accept	9	0	0	0	0	0	**17**	0	0	0	0	3	0	0	0	0.5862	0.6296	0.6071	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.7500	0.8571	
Emphasis	4	0	0	0	2	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.3125	0.3571	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	1.0000	0.2353	0.3810	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.2500	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**279**	4	5	16	12	4	8	4	13	7	12	9	2	6	1	0.7304	0.8558	0.7881	83.14%
System	2	**248**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9880	0.9841	0.9861	
Greet	10	0	**136**	2	0	0	1	0	0	1	0	0	0	0	0	0.9067	0.9444	0.9252	
Emotion	3	0	1	**91**	0	0	0	0	0	0	0	0	0	0	1	0.9479	0.8273	0.8835	
ynQuestion	6	0	1	0	**34**	2	0	0	0	0	0	1	0	0	0	0.7727	0.6667	0.7158	
whQuestion	4	0	0	1	3	**43**	0	0	0	0	0	0	0	0	0	0.8431	0.8776	0.8600	
Accept	9	0	0	0	0	0	**16**	0	0	0	0	2	0	0	0	0.5926	0.5926	0.5926	
Bye	1	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0.9231	0.7500	0.8276	
Emphasis	4	0	0	0	1	0	0	0	**2**	0	0	1	0	0	0	0.2500	0.1250	0.1667	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**5**	0	3	0	0	0.4545	0.2778	0.3448	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**4**	0	0	0	0.8000	0.2353	0.3636	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3750	0.4615	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**2**	0.6667	0.5000	0.5714	

Figure C.12: Experiment Run 10: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**265**	6	8	21	11	5	5	4	8	9	11	3	4	0	0	0.7361	0.8833	0.8030	83.58%
System	1	**235**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9874	0.9711	0.9792	
Greet	8	0	**119**	1	1	0	0	2	0	0	0	0	0	0	0	0.9084	0.9084	0.9084	
Emotion	3	0	0	**90**	0	0	1	0	0	0	0	0	0	0	0	0.9574	0.7826	0.8612	
ynQuestion	9	1	1	0	**42**	3	0	0	0	0	0	0	0	0	0	0.7500	0.6885	0.7179	
whQuestion	2	0	0	1	5	**40**	0	0	1	0	0	0	0	0	0	0.8163	0.8333	0.8247	
Accept	6	0	0	1	0	0	**6**	0	0	1	1	0	0	0	0	0.4000	0.3750	0.3871	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.6667	0.8000	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4500	0.5455	
Continuer	2	0	0	0	2	0	0	0	1	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	1	0	0	0	0	0	1	0	1	0	**5**	0	1	0	0	0.5556	0.2778	0.3704	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**6**	0	0	0	0.6000	0.6667	0.6316	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**4**	0	0	0.6667	0.4444	0.5333	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**258**	4	5	24	16	6	4	5	11	9	10	4	4	0	0	0.7167	0.8600	0.7818	82.49%
System	2	**238**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9917	0.9835	0.9876	
Greet	9	0	**124**	0	1	0	0	1	0	0	0	0	0	0	0	0.9185	0.9466	0.9323	
Emotion	2	0	0	**88**	0	0	1	0	0	0	0	0	0	0	0	0.9670	0.7652	0.8544	
ynQuestion	11	0	1	0	**36**	1	0	0	0	0	0	0	0	0	0	0.7347	0.5902	0.6545	
whQuestion	3	0	0	1	7	**40**	0	0	1	0	0	1	0	0	0	0.7547	0.8333	0.7921	
Accept	6	0	0	0	0	0	**7**	0	0	1	1	2	0	0	0	0.4118	0.4375	0.4242	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.6667	0.8000	
Emphasis	2	0	1	1	0	0	0	0	**7**	0	0	0	0	0	0	0.6364	0.3500	0.4516	
Continuer	4	0	0	0	1	0	0	0	0	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	2	0	0	0	0	1	1	0	1	0	**6**	0	1	0	0	0.5000	0.3333	0.4000	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2222	0.2667	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**4**	0	0	0.6667	0.4444	0.5333	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	1.0000	0.6667	

Figure C.13: Experiment Run 15: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	4	8	17	10	4	8	5	13	9	6	2	1	0	1	0.7622	0.8952	0.8234	84.71%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	11	0	**131**	4	0	1	0	1	0	1	1	0	0	0	0	0.8733	0.8973	0.8851	
Emotion	1	0	1	**81**	0	0	2	0	0	0	0	0	0	0	0	0.9529	0.7714	0.8526	
ynQuestion	2	0	2	0	**37**	1	0	0	1	0	0	0	0	0	0	0.8605	0.7400	0.7957	
whQuestion	2	0	0	0	3	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.8636	0.8736	
Accept	5	0	0	1	0	0	**15**	0	1	0	3	0	0	0	0	0.6000	0.5556	0.5769	
Bye	1	0	1	0	0	0	1	**13**	0	0	0	0	0	0	0	0.8125	0.6842	0.7429	
Emphasis	2	0	2	1	0	0	0	0	**11**	0	1	0	0	0	0	0.6471	0.4074	0.5000	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	1	0	0	1	0	0	1	0	0	0	0	**6**	0	0	0	0.6667	0.6000	0.6316	
nAnswer	0	0	0	0	0	0	0	0	0	1	2	0	**1**	0	0	0.2500	0.5000	0.3333	
Clarify	1	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.8000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**276**	4	8	21	8	3	8	4	15	10	5	3	1	0	3	0.7480	0.8762	0.8070	83.27%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	7	0	**132**	2	0	1	0	0	0	1	1	0	0	0	0	0.9167	0.9041	0.9103	
Emotion	1	0	1	**78**	0	0	1	0	0	0	0	0	0	0	0	0.9630	0.7429	0.8387	
ynQuestion	6	0	1	0	**33**	1	0	0	1	0	0	0	0	0	1	0.7674	0.6600	0.7097	
whQuestion	3	0	0	1	9	**38**	0	0	0	0	0	0	0	0	0	0.7451	0.8636	0.8000	
Accept	7	0	0	1	0	0	**17**	0	1	0	3	1	0	0	0	0.5667	0.6296	0.5965	
Bye	0	0	1	0	0	0	0	**15**	0	0	0	0	0	0	0	0.9375	0.7895	0.8571	
Emphasis	3	0	2	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6000	0.3333	0.4286	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	2	0	0	1	0	1	1	0	0	0	0	**4**	0	0	0	0.4444	0.4000	0.4211	
nAnswer	3	0	0	0	0	0	0	0	0	0	4	0	**1**	0	0	0.1250	0.5000	0.2000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.2000	0.3333	

Figure C.14: Experiment Run 20: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**250**	4	6	18	8	1	7	9	6	8	11	0	4	0	4	0.7440	0.8224	0.7813	82.66%
System	3	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9892	0.9857	0.9874	
Greet	17	0	**116**	1	1	0	0	0	3	1	0	0	0	0	0	0.8345	0.9063	0.8689	
Emotion	6	0	0	**106**	0	0	2	0	0	0	0	0	0	0	0	0.9298	0.8346	0.8797	
ynQuestion	7	0	2	0	**38**	2	0	0	0	0	0	0	0	0	0	0.7755	0.7308	0.7525	
whQuestion	4	0	1	0	4	**39**	0	1	0	0	0	0	0	0	0	0.7959	0.9070	0.8478	
Accept	7	0	0	1	0	0	**5**	0	0	0	0	3	0	0	0	0.3125	0.2632	0.2857	
Emphasis	3	0	2	1	0	0	1	**7**	1	0	0	0	0	0	0	0.4667	0.3889	0.4242	
Bye	2	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0.8750	0.5600	0.6829	
Continuer	2	0	0	0	0	0	0	1	0	**4**	0	1	0	0	0	0.5000	0.2857	0.3636	
Reject	2	0	0	0	0	1	0	0	0	0	**4**	0	2	0	0	0.4444	0.2667	0.3333	
yAnswer	0	0	1	0	1	0	4	0	0	0	0	**3**	0	0	0	0.3333	0.4286	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	1	0	0	**2**	1	0	0.5000	0.2500	0.3333	
Other	0	0	0	0	0	0	0	0	1	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**260**	3	7	20	8	1	7	11	8	9	9	2	5	0	4	0.7345	0.8553	0.7903	83.14%
System	2	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9928	0.9857	0.9892	
Greet	16	0	**116**	0	1	1	0	0	3	1	0	0	0	0	0	0.8406	0.9063	0.8722	
Emotion	4	0	0	**106**	0	0	2	0	0	0	0	0	0	0	0	0.9464	0.8346	0.8870	
ynQuestion	8	0	2	0	**37**	2	0	0	0	0	0	0	0	0	0	0.7551	0.7115	0.7327	
whQuestion	2	0	1	0	4	**38**	0	1	0	0	0	0	0	0	0	0.8261	0.8837	0.8539	
Accept	6	0	0	0	1	0	**5**	0	0	0	0	4	0	0	0	0.3125	0.2632	0.2857	
Emphasis	4	0	2	1	0	0	1	**6**	0	0	0	0	0	0	0	0.4286	0.3333	0.3750	
Bye	0	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	1.0000	0.5600	0.7179	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.8000	0.2857	0.4211	
Reject	0	0	0	0	0	1	0	0	0	0	**4**	0	1	0	0	0.6667	0.2667	0.3810	
yAnswer	0	0	0	0	1	0	4	0	0	0	0	**1**	0	0	0	0.1667	0.1429	0.1538	
nAnswer	0	1	0	0	0	0	0	0	0	0	2	0	**2**	1	0	0.3333	0.2500	0.2857	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.15: Experiment Run 25: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**288**	3	7	20	9	6	16	7	7	12	13	1	1	0	1	0.7366	0.9057	0.8124	83.64%
System	1	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9891	0.9927	
Greet	12	0	**141**	1	0	2	1	0	1	1	0	0	0	0	0	0.8868	0.9400	0.9126	
Emotion	3	0	0	**72**	0	0	0	0	0	0	1	0	0	0	0	0.9474	0.7423	0.8324	
ynQuestion	4	0	1	0	**34**	6	0	0	0	1	0	0	0	0	0	0.7391	0.6667	0.7010	
whQuestion	0	0	0	0	7	**45**	0	0	0	0	0	1	0	0	0	0.8491	0.7627	0.8036	
Accept	3	0	0	2	1	0	**7**	0	0	0	2	1	0	0	0	0.4375	0.2593	0.3256	
Emphasis	4	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.3000	0.3000	0.3000	
Bye	0	0	1	0	0	0	1	0	**14**	0	0	0	0	0	0	0.8750	0.6087	0.7179	
Continuer	1	0	0	0	0	0	0	0	0	**5**	0	1	0	0	0	0.7143	0.2500	0.3704	
Reject	2	0	0	1	0	0	0	0	0	0	**3**	1	2	0	0	0.3333	0.1579	0.2143	
yAnswer	0	0	0	0	0	0	2	0	0	0	0	**3**	0	0	0	0.6000	0.3750	0.4615	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	1	0.6667	0.4000	0.5000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	1.0000	1.0000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.6667	0.8000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**285**	3	9	19	11	8	17	7	6	9	11	4	1	1	4	0.7215	0.8962	0.7994	82.62%
System	1	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9891	0.9927	
Greet	8	0	**138**	1	0	1	1	0	1	1	0	0	0	0	0	0.9139	0.9200	0.9169	
Emotion	4	0	0	**73**	0	0	0	0	0	0	0	0	0	0	0	0.9481	0.7526	0.8391	
ynQuestion	5	0	1	0	**30**	4	0	0	0	1	1	0	0	0	0	0.7143	0.5882	0.6452	
whQuestion	0	0	0	0	9	**46**	0	0	0	1	0	1	1	0	0	0.7931	0.7797	0.7863	
Accept	5	0	0	2	1	0	**7**	0	0	0	2	1	0	0	0	0.3889	0.2593	0.3111	
Emphasis	5	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.2727	0.3000	0.2857	
Bye	1	0	2	0	0	0	0	0	**15**	0	0	0	0	0	0	0.8333	0.6522	0.7317	
Continuer	2	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0.7500	0.3000	0.4286	
Reject	1	0	0	1	0	0	0	0	0	0	**3**	0	1	0	0	0.5000	0.1579	0.2400	
yAnswer	1	0	0	0	0	0	2	0	0	1	0	**2**	0	0	0	0.3333	0.2500	0.2857	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**2**	0	1	0.4000	0.4000	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.1667	0.2857	

Figure C.16: Experiment Run 30: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**263**	4	0	20	8	7	17	5	9	9	12	2	4	4	0	0.7225	0.8946	0.7994	83.72%
System	2	**298**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9933	0.9868	0.9900	
Greet	8	0	**114**	0	0	0	0	0	2	2	1	0	0	0	1	0.8906	0.9421	0.9157	
Emotion	4	0	4	**84**	0	0	0	0	0	0	0	0	0	0	0	0.9130	0.7778	0.8400	
ynQuestion	6	0	0	0	**39**	4	0	0	0	1	0	1	0	0	0	0.7647	0.7647	0.7647	
whQuestion	1	0	0	0	1	**52**	0	0	0	0	0	0	0	0	0	0.9630	0.8254	0.8889	
Accept	7	0	0	0	1	0	**8**	0	1	0	2	3	0	0	0	0.3636	0.2963	0.3265	
Bye	0	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.9333	0.7368	0.8235	
Emphasis	0	0	2	4	1	0	0	0	**6**	0	0	1	0	0	0	0.4286	0.3000	0.3529	
Continuer	0	0	0	0	0	0	1	0	1	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**3**	0	3	0	0	0.3333	0.1667	0.2222	
yAnswer	1	0	0	0	0	0	1	0	0	0	0	**3**	0	0	0	0.6000	0.2727	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	1.0000	0.3000	0.4615	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**269**	3	0	21	7	9	15	5	11	8	10	2	4	3	1	0.7310	0.9150	0.8127	84.28%
System	0	**299**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9967	0.9901	0.9934	
Greet	3	0	**114**	0	0	1	0	0	0	2	1	0	0	0	0	0.9421	0.9421	0.9421	
Emotion	3	0	4	**82**	0	0	0	0	2	0	0	0	0	0	0	0.9011	0.7593	0.8241	
ynQuestion	6	0	0	0	**38**	3	0	0	0	2	0	1	0	0	0	0.7600	0.7451	0.7525	
whQuestion	1	0	0	0	4	**50**	0	0	0	0	0	0	0	0	0	0.9091	0.7937	0.8475	
Accept	5	0	0	0	1	0	**11**	0	1	0	2	3	0	0	0	0.4783	0.4074	0.4400	
Bye	1	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.8750	0.7368	0.8000	
Emphasis	0	0	2	4	0	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.2500	0.3125	
Continuer	1	0	0	0	0	0	1	0	0	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	2	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0.6000	0.2727	0.3750	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3000	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.2500	0.3333	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Figure C.17: Experiment Run 35: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**279**	3	7	10	15	3	10	4	8	7	16	4	1	0	6	0.7480	0.8532	0.7971	83.66%
System	3	**264**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9814	0.9814	0.9814	
Greet	16	0	**131**	1	0	0	0	0	0	0	1	0	0	0	0	0.8792	0.9034	0.8912	
Emotion	6	0	1	**99**	0	0	0	0	1	0	0	0	0	1	1	0.9083	0.8839	0.8959	
ynQuestion	4	2	1	0	**44**	1	0	0	1	0	0	0	0	0	0	0.8302	0.7097	0.7652	
whQuestion	1	0	1	0	2	**44**	0	0	1	0	0	1	0	0	0	0.8800	0.8980	0.8889	
Accept	6	0	0	0	1	0	**13**	0	1	1	1	2	0	0	0	0.5200	0.5200	0.5200	
Bye	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0	1.0000	0.7778	0.8750	
Emphasis	3	0	2	2	0	0	0	0	**6**	0	0	0	0	0	0	0.4615	0.3158	0.3750	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	3	0	0	0	0	0	0	0	1	0	**2**	0	1	0	0	0.2857	0.0952	0.1429	
yAnswer	1	0	0	0	0	1	2	0	0	0	0	**5**	0	0	0	0.5556	0.4167	0.4762	
nAnswer	0	0	0	0	0	0	0	0	0	1	1	0	**0**	0	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	3	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**276**	2	6	11	16	4	10	4	11	8	15	5	0	0	7	0.7360	0.8440	0.7863	82.92%
System	2	**265**	3	1	0	0	0	0	0	0	0	0	0	0	0	0.9779	0.9851	0.9815	
Greet	17	0	**131**	0	0	0	1	0	0	0	1	0	0	0	0	0.8733	0.9034	0.8881	
Emotion	3	0	1	**99**	0	0	0	0	1	0	0	0	0	1	0	0.9429	0.8839	0.9124	
ynQuestion	9	1	2	0	**41**	2	0	0	1	0	0	0	0	0	0	0.7321	0.6613	0.6949	
whQuestion	0	1	0	0	4	**42**	0	0	0	0	0	1	0	0	0	0.8750	0.8571	0.8660	
Accept	10	0	0	0	1	0	**12**	0	1	1	1	2	0	0	0	0.4286	0.4800	0.4528	
Bye	1	0	0	0	0	0	1	**14**	0	0	0	0	0	0	0	0.8750	0.7778	0.8235	
Emphasis	2	0	2	1	0	0	0	0	**5**	0	0	0	0	0	0	0.5000	0.2632	0.3448	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	1	0	0	0	0	0	0	0	0	0	**3**	0	1	0	0	0.6000	0.1429	0.2308	
yAnswer	1	0	0	0	0	1	1	0	0	0	0	**4**	0	0	0	0.5714	0.3333	0.4211	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**1**	0	0	0.3333	0.5000	0.4000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	2	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.18: Experiment Run 40: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	**Precision**	**Recall**	**F-score**	**Overall Accuracy**
Statement	**285**	4	7	19	8	6	5	4	7	11	11	3	2	1	0	0.7641	0.8559	0.8074	83.77%
System	4	**270**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9818	0.9854	0.9836	
Greet	11	0	**109**	2	0	1	0	0	0	1	0	0	0	0	0	0.8790	0.8790	0.8790	
Emotion	4	0	2	**94**	0	0	0	0	1	0	0	0	0	0	0	0.9307	0.7899	0.8545	
ynQuestion	10	0	1	0	**35**	2	0	0	0	0	0	0	0	1	0	0.7143	0.7609	0.7368	
whQuestion	0	0	3	0	2	**48**	0	0	0	1	0	0	0	0	0	0.8889	0.8421	0.8649	
Accept	8	0	0	0	0	0	**14**	0	0	0	0	2	1	0	0	0.5600	0.6087	0.5833	
Bye	1	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	0.9412	0.7273	0.8205	
Emphasis	2	0	2	2	0	0	0	0	**8**	1	1	0	0	0	0	0.5000	0.4706	0.4848	
Continuer	4	0	0	0	1	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	1	0	0	0	0	0	0	0	1	0	**1**	0	0	0	0	0.3333	0.0714	0.1176	
yAnswer	0	0	0	1	0	0	3	0	0	1	0	**3**	0	0	0	0.3750	0.3750	0.3750	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	2	0	0	0	0	0	0	1	0	0	0	0	0	**1**	0	0.2500	0.3333	0.2857	
Other	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**4**	0.8000	1.0000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	**Precision**	**Recall**	**F-score**	**Overall Acc**
Statement	**292**	4	9	20	9	3	5	6	8	10	12	4	2	2	2	0.7526	0.8769	0.8100	83.40%
System	2	**270**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9926	0.9854	0.9890	
Greet	8	0	**109**	2	0	1	0	0	0	1	0	0	0	0	0	0.9008	0.8790	0.8898	
Emotion	3	0	1	**91**	0	0	0	0	0	0	0	0	0	0	0	0.9579	0.7647	0.8505	
ynQuestion	11	0	1	0	**29**	1	0	0	0	1	0	0	0	1	0	0.6591	0.6304	0.6444	
whQuestion	1	0	2	0	8	**52**	0	0	0	1	0	0	0	0	0	0.8125	0.9123	0.8595	
Accept	8	0	0	0	0	0	**14**	0	1	0	0	2	0	0	0	0.5600	0.6087	0.5833	
Bye	0	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	1.0000	0.7273	0.8421	
Emphasis	2	0	2	3	0	0	0	0	**7**	1	1	0	0	0	0	0.4375	0.4118	0.4242	
Continuer	4	0	0	1	0	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	1	0	0	0	0	0	0	0	1	0	**0**	0	1	0	0	0.0000	0.0000	undef	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2500	0.2857	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	2	0	0	0	0	0	0	0	0	0	0	**2**	0.5000	0.5000	0.5000	

Figure C.19: Experiment Run 45: Emoticons Assigned “UH” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**266**	4	7	13	11	3	6	7	9	11	7	4	4	1	2	0.7493	0.8721	0.8061	83.09%
System	0	**238**	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000	0.9835	0.9917	
Greet	10	0	**131**	0	0	0	0	0	0	1	0	1	0	0	1	0.9097	0.9161	0.9129	
Emotion	3	0	1	**92**	0	0	1	0	1	0	0	0	1	0	0	0.9293	0.8288	0.8762	
ynQuestion	2	0	1	0	**29**	6	0	0	1	0	0	0	1	1	0	0.7073	0.6744	0.6905	
whQuestion	1	0	2	0	3	**57**	0	0	0	0	0	0	0	0	0	0.9048	0.8382	0.8702	
Accept	8	0	0	1	0	0	**5**	0	0	0	2	4	0	0	0	0.2500	0.3333	0.2857	
Bye	0	0	0	1	0	0	0	**10**	0	0	0	0	0	0	0	0.9091	0.5556	0.6897	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4286	0.5294	
Continuer	4	0	0	0	0	0	1	0	0	**5**	0	0	0	0	0	0.5000	0.2941	0.3704	
Reject	6	0	0	0	0	1	0	1	0	0	**3**	0	2	0	0	0.2308	0.2143	0.2222	
yAnswer	0	0	0	0	0	1	2	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**3**	0	0	0.6000	0.2727	0.3750	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.3333	0.4000	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**269**	5	7	16	12	2	5	9	11	8	8	4	4	1	3	0.7390	0.8820	0.8042	82.70%
System	0	**237**	0	0	0	0	0	0	0	1	0	0	0	0	0	0.9958	0.9793	0.9875	
Greet	7	0	**131**	0	0	1	0	0	0	1	0	0	0	0	0	0.9357	0.9161	0.9258	
Emotion	2	0	1	**90**	0	0	1	0	1	0	0	0	1	0	0	0.9375	0.8108	0.8696	
ynQuestion	4	0	2	0	**25**	4	0	0	1	1	0	0	0	1	0	0.6579	0.5814	0.6173	
whQuestion	1	0	1	0	6	**58**	0	0	0	0	0	0	0	0	0	0.8788	0.8529	0.8657	
Accept	9	0	0	1	0	0	**7**	0	0	0	2	5	0	0	0	0.2917	0.4667	0.3590	
Bye	0	0	0	0	0	0	0	**9**	0	0	0	0	0	0	0	1.0000	0.5000	0.6667	
Emphasis	2	0	1	1	0	0	0	0	**7**	0	1	0	0	0	0	0.5833	0.3333	0.4242	
Continuer	5	0	0	0	0	0	1	0	0	**6**	0	0	0	0	0	0.5000	0.3529	0.4138	
Reject	3	0	0	0	0	1	0	0	0	0	**2**	0	3	0	0	0.2222	0.1429	0.1739	
yAnswer	0	0	0	0	0	2	1	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.7500	0.2727	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.3333	0.4000	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

Figure C.20: Experiment Run 50: Emoticons Assigned “UH” Tag

Figures C.21 through C.30 show the results of corresponding experiment runs with emoticons tagged as “EMO”.

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**301**	4	4	11	10	4	6	6	13	11	6	3	2	2	1	0.7839	0.8853	0.8315	85.41%
System	1	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9821	0.9892	
Greet	14	0	**118**	2	0	0	0	1	0	0	0	1	0	0	0	0.8676	0.9593	0.9112	
Emotion	2	0	0	**97**	0	0	0	0	1	1	0	0	0	0	0	0.9604	0.8584	0.9065	
ynQuestion	4	1	1	0	**43**	6	0	0	1	0	1	0	0	0	0	0.7544	0.7288	0.7414	
whQuestion	1	0	0	0	6	**47**	0	0	1	0	0	0	0	0	0	0.8545	0.8246	0.8393	
Accept	6	0	0	1	0	0	**12**	0	0	0	0	2	0	0	0	0.5714	0.5714	0.5714	
Bye	1	0	0	1	0	0	0	**18**	1	0	0	0	0	0	0	0.8571	0.7200	0.7826	
Emphasis	2	0	0	1	0	0	1	0	**5**	0	0	0	0	0	0	0.5556	0.2174	0.3125	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	1	0	0.5714	0.2353	0.3333	
Reject	6	0	0	0	0	0	1	0	1	1	**3**	0	1	0	0	0.2308	0.3000	0.2609	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**5**	0	0	0	0.8333	0.4167	0.5556	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.4000	0.5714	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**297**	3	5	13	14	4	7	5	13	11	8	3	2	2	1	0.7655	0.8735	0.8159	84.59%
System	1	**277**	0	2	0	0	0	0	0	0	0	0	0	0	0	0.9893	0.9893	0.9893	
Greet	13	0	**117**	1	0	1	0	1	0	0	0	0	0	0	0	0.8797	0.9512	0.9141	
Emotion	1	0	0	**96**	0	0	0	0	1	1	0	0	0	0	0	0.9697	0.8496	0.9057	
ynQuestion	11	0	1	0	**38**	4	0	0	1	0	0	0	0	0	0	0.6909	0.6441	0.6667	
whQuestion	0	0	0	0	6	**48**	0	0	0	0	0	0	0	0	0	0.8889	0.8421	0.8649	
Accept	5	0	0	0	0	0	**11**	0	0	0	0	4	0	0	0	0.5500	0.5238	0.5366	
Bye	1	0	0	0	0	0	0	**19**	1	0	0	0	0	0	0	0.9048	0.7600	0.8261	
Emphasis	3	0	0	1	0	0	1	0	**5**	0	0	0	0	0	0	0.5000	0.2174	0.3030	
Continuer	2	0	0	0	1	0	0	0	1	**4**	0	1	0	1	0	0.4000	0.2353	0.2963	
Reject	5	0	0	0	0	0	0	0	1	1	**2**	0	0	0	0	0.2222	0.2000	0.2105	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**4**	0	0	0	0.8000	0.3333	0.4706	
nAnswer	0	0	0	0	0	0	1	0	0	0	0	0	**3**	0	0	0.7500	0.6000	0.6667	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Figure C.21: Experiment Run 5: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	3	5	15	11	6	9	4	10	7	14	8	3	6	0	0.7363	0.8650	0.7955	83.81%
System	0	**249**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9960	0.9881	0.9920	
Greet	12	0	**136**	2	0	1	0	0	0	1	0	0	0	0	0	0.8947	0.9444	0.9189	
Emotion	3	0	1	**93**	0	0	0	0	0	0	0	0	0	0	0	0.9588	0.8378	0.8942	
ynQuestion	6	0	1	0	**36**	4	0	0	0	0	0	1	0	0	0	0.7500	0.7059	0.7273	
whQuestion	3	0	0	1	1	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.7755	0.8261	
Accept	9	0	0	0	0	0	**17**	0	0	0	0	3	0	0	0	0.5862	0.6296	0.6071	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.7500	0.8571	
Emphasis	4	0	0	0	2	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.3125	0.3571	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	1.0000	0.2353	0.3810	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.2500	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**279**	4	5	16	11	4	8	4	13	7	12	9	2	6	1	0.7323	0.8558	0.7893	83.24%
System	2	**248**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9880	0.9841	0.9861	
Greet	10	0	**136**	2	0	0	1	0	0	1	0	0	0	0	0	0.9067	0.9444	0.9252	
Emotion	3	0	1	**91**	0	0	0	0	0	0	0	0	0	0	1	0.9479	0.8273	0.8835	
ynQuestion	6	0	1	0	**35**	2	0	0	0	0	0	1	0	0	0	0.7778	0.6863	0.7292	
whQuestion	4	0	0	1	3	**43**	0	0	0	0	0	0	0	0	0	0.8431	0.8776	0.8600	
Accept	9	0	0	0	0	0	**16**	0	0	0	0	2	0	0	0	0.5926	0.5926	0.5926	
Bye	1	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0.9231	0.7500	0.8276	
Emphasis	4	0	0	0	1	0	0	0	**2**	0	0	1	0	0	0	0.2500	0.1250	0.1667	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**5**	0	3	0	0	0.4545	0.2778	0.3448	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**4**	0	0	0	0.8000	0.2353	0.3636	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3750	0.4615	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**2**	0.6667	0.5000	0.5714	

Figure C.22: Experiment Run 10: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**265**	6	8	21	11	5	5	4	8	9	11	3	4	0	0	0.7361	0.8833	0.8030	83.58%
System	1	**235**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9874	0.9711	0.9792	
Greet	8	0	**119**	1	1	0	0	2	0	0	0	0	0	0	0	0.9084	0.9084	0.9084	
Emotion	3	0	0	**90**	0	0	1	0	0	0	0	0	0	0	0	0.9574	0.7826	0.8612	
ynQuestion	9	1	1	0	**42**	3	0	0	0	0	0	0	0	0	0	0.7500	0.6885	0.7179	
whQuestion	2	0	0	1	5	**40**	0	0	1	0	0	0	0	0	0	0.8163	0.8333	0.8247	
Accept	6	0	0	1	0	0	**6**	0	0	1	1	0	0	0	0	0.4000	0.3750	0.3871	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.6667	0.8000	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4500	0.5455	
Continuer	2	0	0	0	2	0	0	0	1	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	1	0	0	0	0	0	1	0	1	0	**5**	0	1	0	0	0.5556	0.2778	0.3704	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**6**	0	0	0	0.6000	0.6667	0.6316	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**4**	0	0	0.6667	0.4444	0.5333	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**256**	4	5	24	16	6	4	5	11	9	9	4	4	0	0	0.7171	0.8533	0.7793	82.19%
System	2	**238**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9835	0.9835	0.9835	
Greet	9	0	**122**	0	1	0	0	1	0	0	0	0	0	0	0	0.9173	0.9313	0.9242	
Emotion	2	0	0	**88**	0	0	1	0	0	0	0	0	0	0	0	0.9670	0.7652	0.8544	
ynQuestion	12	0	1	0	**36**	1	0	0	0	0	0	0	0	0	0	0.7200	0.5902	0.6486	
whQuestion	3	0	0	1	7	**40**	0	0	1	0	0	1	0	0	0	0.7547	0.8333	0.7921	
Accept	7	0	0	0	0	0	**7**	0	0	1	1	2	0	0	0	0.3889	0.4375	0.4118	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.6667	0.8000	
Emphasis	2	0	1	1	0	0	0	0	**7**	0	0	0	0	0	0	0.6364	0.3500	0.4516	
Continuer	4	0	0	0	1	0	0	0	0	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	2	0	0	0	0	1	1	0	1	0	**7**	0	1	0	0	0.5385	0.3889	0.4516	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2222	0.2667	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**4**	0	0	0.6667	0.4444	0.5333	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	1.0000	0.6667	

Figure C.23: Experiment Run 15: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	4	8	17	10	4	8	5	13	9	6	2	1	0	1	0.7622	0.8952	0.8234	84.71%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	11	0	**131**	4	0	1	0	1	0	1	1	0	0	0	0	0.8733	0.8973	0.8851	
Emotion	1	0	1	**81**	0	0	2	0	0	0	0	0	0	0	0	0.9529	0.7714	0.8526	
ynQuestion	2	0	2	0	**37**	1	0	0	1	0	0	0	0	0	0	0.8605	0.7400	0.7957	
whQuestion	2	0	0	0	3	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.8636	0.8736	
Accept	5	0	0	1	0	0	**15**	0	1	0	3	0	0	0	0	0.6000	0.5556	0.5769	
Bye	1	0	1	0	0	0	1	**13**	0	0	0	0	0	0	0	0.8125	0.6842	0.7429	
Emphasis	2	0	2	1	0	0	0	0	**11**	0	1	0	0	0	0	0.6471	0.4074	0.5000	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	1	0	0	1	0	0	1	0	0	0	0	**6**	0	0	0	0.6667	0.6000	0.6316	
nAnswer	0	0	0	0	0	0	0	0	0	1	2	0	**1**	0	0	0.2500	0.5000	0.3333	
Clarify	1	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.8000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**276**	4	8	21	7	3	8	4	15	10	5	3	1	0	3	0.7500	0.8762	0.8082	83.46%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	7	0	**132**	2	0	1	0	0	0	1	1	0	0	0	0	0.9167	0.9041	0.9103	
Emotion	1	0	1	**79**	0	0	1	0	0	0	0	0	0	0	0	0.9634	0.7524	0.8449	
ynQuestion	6	0	1	0	**34**	1	0	0	1	0	0	0	0	0	1	0.7727	0.6800	0.7234	
whQuestion	3	0	0	1	9	**38**	0	0	0	0	0	0	0	0	0	0.7451	0.8636	0.8000	
Accept	7	0	0	0	0	0	**17**	0	1	0	3	1	0	0	0	0.5862	0.6296	0.6071	
Bye	0	0	1	0	0	0	0	**15**	0	0	0	0	0	0	0	0.9375	0.7895	0.8571	
Emphasis	3	0	2	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6000	0.3333	0.4286	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	2	0	0	1	0	1	1	0	0	0	0	**4**	0	0	0	0.4444	0.4000	0.4211	
nAnswer	3	0	0	0	0	0	0	0	0	0	4	0	**1**	0	0	0.1250	0.5000	0.2000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.2000	0.3333	

Figure C.24: Experiment Run 20: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**250**	4	6	18	8	1	7	9	6	8	11	0	4	0	4	0.7440	0.8224	0.7813	82.66%
System	3	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9892	0.9857	0.9874	
Greet	17	0	**116**	1	1	0	0	0	3	1	0	0	0	0	0	0.8345	0.9063	0.8689	
Emotion	6	0	0	**106**	0	0	2	0	0	0	0	0	0	0	0	0.9298	0.8346	0.8797	
ynQuestion	7	0	2	0	**38**	2	0	0	0	0	0	0	0	0	0	0.7755	0.7308	0.7525	
whQuestion	4	0	1	0	4	**39**	0	1	0	0	0	0	0	0	0	0.7959	0.9070	0.8478	
Accept	7	0	0	1	0	0	**5**	0	0	0	0	3	0	0	0	0.3125	0.2632	0.2857	
Emphasis	3	0	2	1	0	0	1	**7**	1	0	0	0	0	0	0	0.4667	0.3889	0.4242	
Bye	2	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0.8750	0.5600	0.6829	
Continuer	2	0	0	0	0	0	0	1	0	**4**	0	1	0	0	0	0.5000	0.2857	0.3636	
Reject	2	0	0	0	0	1	0	0	0	0	**4**	0	2	0	0	0.4444	0.2667	0.3333	
yAnswer	0	0	1	0	1	0	4	0	0	0	0	**3**	0	0	0	0.3333	0.4286	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	1	0	0	**2**	1	0	0.5000	0.2500	0.3333	
Other	0	0	0	0	0	0	0	0	1	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**260**	2	7	21	9	1	7	11	7	9	9	2	5	0	4	0.7345	0.8553	0.7903	83.24%
System	2	**276**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9928	0.9892	0.9910	
Greet	16	0	**116**	0	1	1	0	0	2	1	0	0	0	0	0	0.8467	0.9063	0.8755	
Emotion	3	0	0	**105**	0	0	2	0	1	0	0	0	0	0	0	0.9459	0.8268	0.8824	
ynQuestion	9	0	2	0	**37**	2	0	0	0	0	0	0	0	0	0	0.7400	0.7115	0.7255	
whQuestion	2	0	1	0	3	**38**	0	1	0	0	0	0	0	0	0	0.8444	0.8837	0.8636	
Accept	6	0	0	0	1	0	**5**	0	0	0	0	4	0	0	0	0.3125	0.2632	0.2857	
Emphasis	4	0	2	1	0	0	1	**6**	0	0	0	0	0	0	0	0.4286	0.3333	0.3750	
Bye	0	0	0	0	0	0	0	0	**15**	0	0	0	0	0	0	1.0000	0.6000	0.7500	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.8000	0.2857	0.4211	
Reject	0	0	0	0	0	1	0	0	0	0	**4**	0	1	0	0	0.6667	0.2667	0.3810	
yAnswer	0	0	0	0	1	0	4	0	0	0	0	**1**	0	0	0	0.1667	0.1429	0.1538	
nAnswer	0	1	0	0	0	0	0	0	0	0	2	0	**2**	1	0	0.3333	0.2500	0.2857	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.25: Experiment Run 25: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**288**	3	7	20	9	6	16	7	7	12	13	1	1	0	1	0.7366	0.9057	0.8124	83.64%
System	1	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9891	0.9927	
Greet	12	0	**141**	1	0	2	1	0	1	1	0	0	0	0	0	0.8868	0.9400	0.9126	
Emotion	3	0	0	**72**	0	0	0	0	0	0	1	0	0	0	0	0.9474	0.7423	0.8324	
ynQuestion	4	0	1	0	**34**	6	0	0	0	1	0	0	0	0	0	0.7391	0.6667	0.7010	
whQuestion	0	0	0	0	7	**45**	0	0	0	0	0	1	0	0	0	0.8491	0.7627	0.8036	
Accept	3	0	0	2	1	0	**7**	0	0	0	2	1	0	0	0	0.4375	0.2593	0.3256	
Emphasis	4	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.3000	0.3000	0.3000	
Bye	0	0	1	0	0	0	1	0	**14**	0	0	0	0	0	0	0.8750	0.6087	0.7179	
Continuer	1	0	0	0	0	0	0	0	0	**5**	0	1	0	0	0	0.7143	0.2500	0.3704	
Reject	2	0	0	1	0	0	0	0	0	0	**3**	1	2	0	0	0.3333	0.1579	0.2143	
yAnswer	0	0	0	0	0	0	2	0	0	0	0	**3**	0	0	0	0.6000	0.3750	0.4615	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	1	0.6667	0.4000	0.5000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	1.0000	1.0000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.6667	0.8000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**281**	3	10	19	11	8	17	7	6	9	11	4	1	1	4	0.7168	0.8836	0.7915	82.24%
System	1	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9891	0.9927	
Greet	8	0	**137**	1	0	1	1	0	1	1	0	0	0	0	0	0.9133	0.9133	0.9133	
Emotion	8	0	1	**74**	0	0	0	0	0	0	0	0	0	0	0	0.8916	0.7629	0.8222	
ynQuestion	6	0	1	0	**30**	4	0	0	0	1	1	0	0	0	0	0.6977	0.5882	0.6383	
whQuestion	1	0	0	0	9	**46**	0	0	0	1	0	1	1	0	0	0.7797	0.7797	0.7797	
Accept	5	0	0	1	1	0	**7**	0	0	0	2	1	0	0	0	0.4118	0.2593	0.3182	
Emphasis	4	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.3000	0.3000	0.3000	
Bye	1	0	1	0	0	0	0	0	**15**	0	0	0	0	0	0	0.8824	0.6522	0.7500	
Continuer	2	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0.7500	0.3000	0.4286	
Reject	1	0	0	1	0	0	0	0	0	0	**3**	0	1	0	0	0.5000	0.1579	0.2400	
yAnswer	0	0	0	0	0	0	2	0	0	1	0	**2**	0	0	0	0.4000	0.2500	0.3077	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**2**	0	1	0.4000	0.4000	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.1667	0.2857	

Figure C.26: Experiment Run 30: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**263**	4	0	20	8	7	17	5	9	9	12	2	4	4	0	0.7225	0.8946	0.7994	83.72%
System	2	**298**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9933	0.9868	0.9900	
Greet	8	0	**114**	0	0	0	0	0	2	2	1	0	0	0	1	0.8906	0.9421	0.9157	
Emotion	4	0	4	**84**	0	0	0	0	0	0	0	0	0	0	0	0.9130	0.7778	0.8400	
ynQuestion	6	0	0	0	**39**	4	0	0	0	1	0	1	0	0	0	0.7647	0.7647	0.7647	
whQuestion	1	0	0	0	1	**52**	0	0	0	0	0	0	0	0	0	0.9630	0.8254	0.8889	
Accept	7	0	0	0	1	0	**8**	0	1	0	2	3	0	0	0	0.3636	0.2963	0.3265	
Bye	0	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.9333	0.7368	0.8235	
Emphasis	0	0	2	4	1	0	0	0	**6**	0	0	1	0	0	0	0.4286	0.3000	0.3529	
Continuer	0	0	0	0	0	0	1	0	1	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**3**	0	3	0	0	0.3333	0.1667	0.2222	
yAnswer	1	0	0	0	0	0	1	0	0	0	0	**3**	0	0	0	0.6000	0.2727	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	1.0000	0.3000	0.4615	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**269**	3	1	21	7	9	15	5	11	8	10	2	4	3	1	0.7290	0.9150	0.8115	84.10%
System	0	**299**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9967	0.9901	0.9934	
Greet	3	0	**114**	2	0	1	0	0	0	2	1	0	0	0	0	0.9268	0.9421	0.9344	
Emotion	3	0	3	**80**	0	0	0	0	2	0	0	0	0	0	0	0.9091	0.7407	0.8163	
ynQuestion	6	0	0	0	**38**	3	0	0	0	2	0	1	0	0	0	0.7600	0.7451	0.7525	
whQuestion	1	0	0	0	4	**50**	0	0	0	0	0	0	0	0	0	0.9091	0.7937	0.8475	
Accept	5	0	0	0	1	0	**11**	0	1	0	2	3	0	0	0	0.4783	0.4074	0.4400	
Bye	1	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.8750	0.7368	0.8000	
Emphasis	0	0	2	4	0	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.2500	0.3125	
Continuer	1	0	0	0	0	0	1	0	0	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	2	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0.6000	0.2727	0.3750	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3000	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.2500	0.3333	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Figure C.27: Experiment Run 35: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**279**	3	7	10	15	3	10	4	8	7	16	4	1	0	6	0.7480	0.8532	0.7971	83.66%
System	3	**264**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9814	0.9814	0.9814	
Greet	16	0	**131**	1	0	0	0	0	0	0	1	0	0	0	0	0.8792	0.9034	0.8912	
Emotion	6	0	1	**99**	0	0	0	0	1	0	0	0	0	1	1	0.9083	0.8839	0.8959	
ynQuestion	4	2	1	0	**44**	1	0	0	1	0	0	0	0	0	0	0.8302	0.7097	0.7652	
whQuestion	1	0	1	0	2	**44**	0	0	1	0	0	1	0	0	0	0.8800	0.8980	0.8889	
Accept	6	0	0	0	1	0	**13**	0	1	1	1	2	0	0	0	0.5200	0.5200	0.5200	
Bye	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0	1.0000	0.7778	0.8750	
Emphasis	3	0	2	2	0	0	0	0	**6**	0	0	0	0	0	0	0.4615	0.3158	0.3750	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	3	0	0	0	0	0	0	0	1	0	**2**	0	1	0	0	0.2857	0.0952	0.1429	
yAnswer	1	0	0	0	0	1	2	0	0	0	0	**5**	0	0	0	0.5556	0.4167	0.4762	
nAnswer	0	0	0	0	0	0	0	0	0	1	1	0	**0**	0	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	3	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**276**	2	5	11	14	4	10	4	11	8	15	5	0	0	7	0.7419	0.8440	0.7897	83.01%
System	2	**265**	3	1	0	0	0	0	0	0	0	0	0	0	0	0.9779	0.9851	0.9815	
Greet	16	0	**131**	0	0	0	1	0	0	0	1	0	0	0	0	0.8792	0.9034	0.8912	
Emotion	3	0	1	**99**	1	0	0	0	1	0	0	0	0	1	0	0.9340	0.8839	0.9083	
ynQuestion	10	1	3	0	**42**	2	0	0	1	0	0	0	0	0	0	0.7119	0.6774	0.6942	
whQuestion	0	1	0	0	4	**42**	0	0	0	0	0	1	0	0	0	0.8750	0.8571	0.8660	
Accept	11	0	0	0	1	0	**12**	0	1	1	1	2	0	0	0	0.4138	0.4800	0.4444	
Bye	1	0	0	0	0	0	1	**14**	0	0	0	0	0	0	0	0.8750	0.7778	0.8235	
Emphasis	2	0	2	1	0	0	0	0	**5**	0	0	0	0	0	0	0.5000	0.2632	0.3448	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	1	0	0	0	0	0	0	0	0	0	**3**	0	1	0	0	0.6000	0.1429	0.2308	
yAnswer	0	0	0	0	0	1	1	0	0	0	0	**4**	0	0	0	0.6667	0.3333	0.4444	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**1**	0	0	0.3333	0.5000	0.4000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	2	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.28: Experiment Run 40: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**285**	4	7	19	8	6	5	4	7	11	11	3	2	1	0	0.7641	0.8559	0.8074	83.77%
System	4	**270**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9818	0.9854	0.9836	
Greet	11	0	**109**	2	0	1	0	0	0	1	0	0	0	0	0	0.8790	0.8790	0.8790	
Emotion	4	0	2	**94**	0	0	0	0	1	0	0	0	0	0	0	0.9307	0.7899	0.8545	
ynQuestion	10	0	1	0	**35**	2	0	0	0	0	0	0	0	1	0	0.7143	0.7609	0.7368	
whQuestion	0	0	3	0	2	**48**	0	0	0	1	0	0	0	0	0	0.8889	0.8421	0.8649	
Accept	8	0	0	0	0	0	**14**	0	0	0	0	2	1	0	0	0.5600	0.6087	0.5833	
Bye	1	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	0.9412	0.7273	0.8205	
Emphasis	2	0	2	2	0	0	0	0	**8**	1	1	0	0	0	0	0.5000	0.4706	0.4848	
Continuer	4	0	0	0	1	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	1	0	0	0	0	0	0	0	1	0	**1**	0	0	0	0	0.3333	0.0714	0.1176	
yAnswer	0	0	0	1	0	0	3	0	0	1	0	**3**	0	0	0	0.3750	0.3750	0.3750	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	2	0	0	0	0	0	0	1	0	0	0	0	0	**1**	0	0.2500	0.3333	0.2857	
Other	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**4**	0.8000	1.0000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**291**	4	8	21	9	3	5	6	9	10	12	4	2	2	2	0.7500	0.8739	0.8072	83.30%
System	2	**270**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9926	0.9854	0.9890	
Greet	8	0	**110**	2	0	1	0	0	0	1	0	0	0	0	0	0.9016	0.8871	0.8943	
Emotion	3	0	1	**90**	0	0	0	0	0	0	0	0	0	0	0	0.9574	0.7563	0.8451	
ynQuestion	11	0	1	0	**29**	1	0	0	0	1	0	0	0	1	0	0.6591	0.6304	0.6444	
whQuestion	1	0	2	0	8	**52**	0	0	0	1	0	0	0	0	0	0.8125	0.9123	0.8595	
Accept	8	0	0	0	0	0	**14**	0	0	0	0	2	0	0	0	0.5833	0.6087	0.5957	
Bye	0	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	1.0000	0.7273	0.8421	
Emphasis	2	0	2	3	0	0	0	0	**7**	1	1	0	0	0	0	0.4375	0.4118	0.4242	
Continuer	4	0	0	1	0	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	2	0	0	0	0	0	0	0	1	0	**0**	0	1	0	0	0.0000	0.0000	undef	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2500	0.2857	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	2	0	0	0	0	0	0	0	0	0	0	**2**	0.5000	0.5000	0.5000	

Figure C.29: Experiment Run 45: Emoticons Assigned “EMO” Tag

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**266**	4	7	13	11	3	6	7	9	11	7	4	4	1	2	0.7493	0.8721	0.8061	83.09%
System	0	**238**	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000	0.9835	0.9917	
Greet	10	0	**131**	0	0	0	0	0	0	1	0	1	0	0	1	0.9097	0.9161	0.9129	
Emotion	3	0	1	**92**	0	0	1	0	1	0	0	0	1	0	0	0.9293	0.8288	0.8762	
ynQuestion	2	0	1	0	**29**	6	0	0	1	0	0	0	1	1	0	0.7073	0.6744	0.6905	
whQuestion	1	0	2	0	3	**57**	0	0	0	0	0	0	0	0	0	0.9048	0.8382	0.8702	
Accept	8	0	0	1	0	0	**5**	0	0	0	2	4	0	0	0	0.2500	0.3333	0.2857	
Bye	0	0	0	1	0	0	0	**10**	0	0	0	0	0	0	0	0.9091	0.5556	0.6897	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4286	0.5294	
Continuer	4	0	0	0	0	0	1	0	0	**5**	0	0	0	0	0	0.5000	0.2941	0.3704	
Reject	6	0	0	0	0	1	0	1	0	0	**3**	0	2	0	0	0.2308	0.2143	0.2222	
yAnswer	0	0	0	0	0	1	2	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**3**	0	0	0.6000	0.2727	0.3750	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.3333	0.4000	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**269**	5	7	16	12	2	5	8	10	8	8	4	4	1	3	0.7431	0.8820	0.8066	82.80%
System	0	**237**	0	0	0	0	0	0	0	1	0	0	0	0	0	0.9958	0.9793	0.9875	
Greet	7	0	**132**	0	0	1	0	0	0	1	0	0	0	0	0	0.9362	0.9231	0.9296	
Emotion	2	0	0	**90**	0	0	1	0	1	0	0	0	1	0	0	0.9474	0.8108	0.8738	
ynQuestion	4	0	2	0	**24**	4	0	0	1	1	0	0	0	1	0	0.6486	0.5581	0.6000	
whQuestion	1	0	1	0	7	**58**	0	1	0	0	0	0	0	0	0	0.8529	0.8529	0.8529	
Accept	9	0	0	1	0	0	**7**	0	0	0	2	5	0	0	0	0.2917	0.4667	0.3590	
Bye	0	0	0	0	0	0	0	**9**	0	0	0	0	0	0	0	1.0000	0.5000	0.6667	
Emphasis	2	0	1	1	0	0	0	0	**8**	0	1	0	0	0	0	0.6154	0.3810	0.4706	
Continuer	5	0	0	0	0	0	1	0	0	**6**	0	0	0	0	0	0.5000	0.3529	0.4138	
Reject	3	0	0	0	0	1	0	0	0	0	**2**	0	3	0	0	0.2222	0.1429	0.1739	
yAnswer	0	0	0	0	0	2	1	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.7500	0.2727	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.3333	0.4000	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

Figure C.30: Experiment Run 50: Emoticons Assigned “EMO” Tag

Figures C.31 through C.40 show the results of corresponding experiment runs with emoticons segregated into two categories based on type.

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**301**	4	4	11	10	4	6	6	13	11	6	3	2	2	1	0.7839	0.8853	0.8315	85.41%
System	1	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9821	0.9892	
Greet	14	0	**118**	2	0	0	0	1	0	0	0	1	0	0	0	0.8676	0.9593	0.9112	
Emotion	2	0	0	**97**	0	0	0	0	1	1	0	0	0	0	0	0.9604	0.8584	0.9065	
ynQuestion	4	1	1	0	**43**	6	0	0	1	0	1	0	0	0	0	0.7544	0.7288	0.7414	
whQuestion	1	0	0	0	6	**47**	0	0	1	0	0	0	0	0	0	0.8545	0.8246	0.8393	
Accept	6	0	0	1	0	0	**12**	0	0	0	0	2	0	0	0	0.5714	0.5714	0.5714	
Bye	1	0	0	1	0	0	0	**18**	1	0	0	0	0	0	0	0.8571	0.7200	0.7826	
Emphasis	2	0	0	1	0	0	1	0	**5**	0	0	0	0	0	0	0.5556	0.2174	0.3125	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	1	0	0.5714	0.2353	0.3333	
Reject	6	0	0	0	0	0	1	0	1	1	**3**	0	1	0	0	0.2308	0.3000	0.2609	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**5**	0	0	0	0.8333	0.4167	0.5556	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.4000	0.5714	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**297**	3	5	13	14	4	7	5	13	11	8	3	2	2	1	0.7655	0.8735	0.8159	84.59%
System	1	**277**	0	2	0	0	0	0	0	0	0	0	0	0	0	0.9893	0.9893	0.9893	
Greet	13	0	**117**	1	0	1	0	1	0	0	0	0	0	0	0	0.8797	0.9512	0.9141	
Emotion	1	0	0	**96**	0	0	0	0	1	1	0	0	0	0	0	0.9697	0.8496	0.9057	
ynQuestion	11	0	1	0	**38**	4	0	0	1	0	0	0	0	0	0	0.6909	0.6441	0.6667	
whQuestion	0	0	0	0	6	**48**	0	0	0	0	0	0	0	0	0	0.8889	0.8421	0.8649	
Accept	5	0	0	0	0	0	**11**	0	0	0	0	4	0	0	0	0.5500	0.5238	0.5366	
Bye	1	0	0	0	0	0	0	**19**	1	0	0	0	0	0	0	0.9048	0.7600	0.8261	
Emphasis	3	0	0	1	0	0	1	0	**5**	0	0	0	0	0	0	0.5000	0.2174	0.3030	
Continuer	2	0	0	0	1	0	0	0	1	**4**	0	1	0	1	0	0.4000	0.2353	0.2963	
Reject	5	0	0	0	0	0	0	0	1	1	**2**	0	0	0	0	0.2222	0.2000	0.2105	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**4**	0	0	0	0.8000	0.3333	0.4706	
nAnswer	0	0	0	0	0	0	1	0	0	0	0	0	**3**	0	0	0.7500	0.6000	0.6667	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	#DIV/0!	undef	undef	
Other	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	0.5000	0.5000	

Figure C.31: Experiment Run 5: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	3	5	15	11	6	9	4	10	7	14	8	3	6	0	0.7363	0.8650	0.7955	83.81%
System	0	**249**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9960	0.9881	0.9920	
Greet	12	0	**136**	2	0	1	0	0	0	1	0	0	0	0	0	0.8947	0.9444	0.9189	
Emotion	3	0	1	**93**	0	0	0	0	0	0	0	0	0	0	0	0.9588	0.8378	0.8942	
ynQuestion	6	0	1	0	**36**	4	0	0	0	0	0	1	0	0	0	0.7500	0.7059	0.7273	
whQuestion	3	0	0	1	1	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.7755	0.8261	
Accept	9	0	0	0	0	0	**17**	0	0	0	0	3	0	0	0	0.5862	0.6296	0.6071	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.7500	0.8571	
Emphasis	4	0	0	0	2	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.3125	0.3571	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	1.0000	0.2353	0.3810	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1.0000	0.2500	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**279**	4	5	15	11	4	8	4	13	7	12	9	2	6	1	0.7342	0.8558	0.7904	83.33%
System	2	**248**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9880	0.9841	0.9861	
Greet	10	0	**136**	2	0	0	1	0	0	1	0	0	0	0	0	0.9067	0.9444	0.9252	
Emotion	3	0	1	**92**	0	0	0	0	0	0	0	0	0	0	1	0.9485	0.8364	0.8889	
ynQuestion	6	0	1	0	**35**	2	0	0	0	0	0	1	0	0	0	0.7778	0.6863	0.7292	
whQuestion	4	0	0	1	3	**43**	0	0	0	0	0	0	0	0	0	0.8431	0.8776	0.8600	
Accept	9	0	0	0	0	0	**16**	0	0	0	0	2	0	0	0	0.5926	0.5926	0.5926	
Bye	1	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0.9231	0.7500	0.8276	
Emphasis	4	0	0	0	1	0	0	0	**2**	0	0	1	0	0	0	0.2500	0.1250	0.1667	
Continuer	6	0	0	0	0	0	1	0	0	**3**	0	0	0	0	0	0.3000	0.2727	0.2857	
Reject	1	0	0	0	1	0	0	0	1	0	**5**	0	3	0	0	0.4545	0.2778	0.3448	
yAnswer	0	0	0	0	0	0	1	0	0	0	0	**4**	0	0	0	0.8000	0.2353	0.3636	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3750	0.4615	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**2**	0.6667	0.5000	0.5714	

Figure C.32: Experiment Run 10: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**265**	6	8	21	11	5	5	4	8	9	11	3	4	0	0	0.7361	0.8833	0.8030	83.58%
System	1	**235**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9874	0.9711	0.9792	
Greet	8	0	**119**	1	1	0	0	2	0	0	0	0	0	0	0	0.9084	0.9084	0.9084	
Emotion	3	0	0	**90**	0	0	1	0	0	0	0	0	0	0	0	0.9574	0.7826	0.8612	
ynQuestion	9	1	1	0	**42**	3	0	0	0	0	0	0	0	0	0	0.7500	0.6885	0.7179	
whQuestion	2	0	0	1	5	**40**	0	0	1	0	0	0	0	0	0	0.8163	0.8333	0.8247	
Accept	6	0	0	1	0	0	**6**	0	0	1	1	0	0	0	0	0.4000	0.3750	0.3871	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.6667	0.8000	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4500	0.5455	
Continuer	2	0	0	0	2	0	0	0	1	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	1	0	0	0	0	0	1	0	1	0	**5**	0	1	0	0	0.5556	0.2778	0.3704	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**6**	0	0	0	0.6000	0.6667	0.6316	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**4**	0	0	0.6667	0.4444	0.5333	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	1.0000	1.0000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**256**	4	5	22	16	6	4	5	11	9	9	4	4	0	0	0.7211	0.8533	0.7817	82.39%
System	2	**238**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9835	0.9835	0.9835	
Greet	9	0	**122**	0	1	0	0	1	0	0	0	0	0	0	0	0.9173	0.9313	0.9242	
Emotion	2	0	0	**90**	0	0	1	0	0	0	0	0	0	0	0	0.9677	0.7826	0.8654	
ynQuestion	12	0	1	0	**36**	1	0	0	0	0	0	0	0	0	0	0.7200	0.5902	0.6486	
whQuestion	3	0	0	1	7	**40**	0	0	1	0	0	1	0	0	0	0.7547	0.8333	0.7921	
Accept	7	0	0	0	0	0	**7**	0	0	1	1	2	0	0	0	0.3889	0.4375	0.4118	
Bye	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	1.0000	0.6667	0.8000	
Emphasis	2	0	1	1	0	0	0	0	**7**	0	0	0	0	0	0	0.6364	0.3500	0.4516	
Continuer	4	0	0	0	1	0	0	0	0	**6**	0	0	0	0	0	0.5455	0.3529	0.4286	
Reject	2	0	0	0	0	1	1	0	1	0	**7**	0	1	0	0	0.5385	0.3889	0.4516	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2222	0.2667	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**4**	0	0	0.6667	0.4444	0.5333	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	undef	undef	
Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	**1**	0.5000	1.0000	0.6667	

Figure C.33: Experiment Run 15: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**282**	4	8	17	10	4	8	5	13	9	6	2	1	0	1	0.7622	0.8952	0.8234	84.71%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	11	0	**131**	4	0	1	0	1	0	1	1	0	0	0	0	0.8733	0.8973	0.8851	
Emotion	1	0	1	**81**	0	0	2	0	0	0	0	0	0	0	0	0.9529	0.7714	0.8526	
ynQuestion	2	0	2	0	**37**	1	0	0	1	0	0	0	0	0	0	0.8605	0.7400	0.7957	
whQuestion	2	0	0	0	3	**38**	0	0	0	0	0	0	0	0	0	0.8837	0.8636	0.8736	
Accept	5	0	0	1	0	0	**15**	0	1	0	3	0	0	0	0	0.6000	0.5556	0.5769	
Bye	1	0	1	0	0	0	1	**13**	0	0	0	0	0	0	0	0.8125	0.6842	0.7429	
Emphasis	2	0	2	1	0	0	0	0	**11**	0	1	0	0	0	0	0.6471	0.4074	0.5000	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	1	0	0	1	0	0	1	0	0	0	0	**6**	0	0	0	0.6667	0.6000	0.6316	
nAnswer	0	0	0	0	0	0	0	0	0	1	2	0	**1**	0	0	0.2500	0.5000	0.3333	
Clarify	1	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.8000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**276**	4	8	21	8	4	8	4	15	10	5	3	1	0	3	0.7459	0.8762	0.8058	83.17%
System	2	**256**	1	0	0	0	0	0	0	0	0	0	0	0	0	0.9884	0.9846	0.9865	
Greet	7	0	**131**	2	0	1	0	0	0	1	1	0	0	0	0	0.9161	0.8973	0.9066	
Emotion	1	0	2	**79**	0	0	1	0	0	0	0	0	0	0	0	0.9518	0.7524	0.8404	
ynQuestion	6	0	1	0	**33**	1	0	0	1	0	0	0	0	0	1	0.7674	0.6600	0.7097	
whQuestion	3	0	0	1	9	**37**	0	0	0	0	0	0	0	0	0	0.7400	0.8409	0.7872	
Accept	7	0	0	0	0	0	**17**	0	1	0	3	1	0	0	0	0.5862	0.6296	0.6071	
Bye	0	0	1	0	0	0	0	**15**	0	0	0	0	0	0	0	0.9375	0.7895	0.8571	
Emphasis	3	0	2	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6000	0.3333	0.4286	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	1	0	0	0	0.6667	0.2667	0.3810	
Reject	4	0	0	0	0	0	0	0	1	0	**2**	0	0	0	0	0.2857	0.1333	0.1818	
yAnswer	2	0	0	1	0	1	1	0	0	0	0	**4**	0	0	0	0.4444	0.4000	0.4211	
nAnswer	3	0	0	0	0	0	0	0	0	0	4	0	**1**	0	0	0.1250	0.5000	0.2000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	1	0	**0**	0	0.0000	undef	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.2000	0.3333	

Figure C.34: Experiment Run 20: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**250**	4	6	18	8	1	7	9	6	8	11	0	4	0	4	0.7440	0.8224	0.7813	82.66%
System	3	**275**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9892	0.9857	0.9874	
Greet	17	0	**116**	1	1	0	0	0	3	1	0	0	0	0	0	0.8345	0.9063	0.8689	
Emotion	6	0	0	**106**	0	0	2	0	0	0	0	0	0	0	0	0.9298	0.8346	0.8797	
ynQuestion	7	0	2	0	**38**	2	0	0	0	0	0	0	0	0	0	0.7755	0.7308	0.7525	
whQuestion	4	0	1	0	4	**39**	0	1	0	0	0	0	0	0	0	0.7959	0.9070	0.8478	
Accept	7	0	0	1	0	0	**5**	0	0	0	0	3	0	0	0	0.3125	0.2632	0.2857	
Emphasis	3	0	2	1	0	0	1	**7**	1	0	0	0	0	0	0	0.4667	0.3889	0.4242	
Bye	2	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0.8750	0.5600	0.6829	
Continuer	2	0	0	0	0	0	0	1	0	**4**	0	1	0	0	0	0.5000	0.2857	0.3636	
Reject	2	0	0	0	0	1	0	0	0	0	**4**	0	2	0	0	0.4444	0.2667	0.3333	
yAnswer	0	0	1	0	1	0	4	0	0	0	0	**3**	0	0	0	0.3333	0.4286	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	1	0	0	**2**	1	0	0.5000	0.2500	0.3333	
Other	0	0	0	0	0	0	0	0	1	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**260**	2	7	21	9	1	7	11	7	9	9	2	5	0	4	0.7345	0.8553	0.7903	83.24%
System	2	**276**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9928	0.9892	0.9910	
Greet	16	0	**116**	0	1	1	0	0	2	1	0	0	0	0	0	0.8467	0.9063	0.8755	
Emotion	3	0	0	**105**	0	0	2	0	1	0	0	0	0	0	0	0.9459	0.8268	0.8824	
ynQuestion	9	0	2	0	**37**	2	0	0	0	0	0	0	0	0	0	0.7400	0.7115	0.7255	
whQuestion	2	0	1	0	3	**38**	0	1	0	0	0	0	0	0	0	0.8444	0.8837	0.8636	
Accept	6	0	0	0	1	0	**5**	0	0	0	0	4	0	0	0	0.3125	0.2632	0.2857	
Emphasis	4	0	2	1	0	0	1	**6**	0	0	0	0	0	0	0	0.4286	0.3333	0.3750	
Bye	0	0	0	0	0	0	0	0	**15**	0	0	0	0	0	0	1.0000	0.6000	0.7500	
Continuer	1	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.8000	0.2857	0.4211	
Reject	0	0	0	0	0	1	0	0	0	0	**4**	0	1	0	0	0.6667	0.2667	0.3810	
yAnswer	0	0	0	0	1	0	4	0	0	0	0	**1**	0	0	0	0.1667	0.1429	0.1538	
nAnswer	0	1	0	0	0	0	0	0	0	0	2	0	**2**	1	0	0.3333	0.2500	0.2857	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.35: Experiment Run 25: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**288**	3	7	20	9	6	16	7	7	12	13	1	1	0	1	0.7366	0.9057	0.8124	83.64%
System	1	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9891	0.9927	
Greet	12	0	**141**	1	0	2	1	0	1	1	0	0	0	0	0	0.8868	0.9400	0.9126	
Emotion	3	0	0	**72**	0	0	0	0	0	0	1	0	0	0	0	0.9474	0.7423	0.8324	
ynQuestion	4	0	1	0	**34**	6	0	0	0	1	0	0	0	0	0	0.7391	0.6667	0.7010	
whQuestion	0	0	0	0	7	**45**	0	0	0	0	0	1	0	0	0	0.8491	0.7627	0.8036	
Accept	3	0	0	2	1	0	**7**	0	0	0	2	1	0	0	0	0.4375	0.2593	0.3256	
Emphasis	4	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.3000	0.3000	0.3000	
Bye	0	0	1	0	0	0	1	0	**14**	0	0	0	0	0	0	0.8750	0.6087	0.7179	
Continuer	1	0	0	0	0	0	0	0	0	**5**	0	1	0	0	0	0.7143	0.2500	0.3704	
Reject	2	0	0	1	0	0	0	0	0	0	**3**	1	2	0	0	0.3333	0.1579	0.2143	
yAnswer	0	0	0	0	0	0	2	0	0	0	0	**3**	0	0	0	0.6000	0.3750	0.4615	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	1	0.6667	0.4000	0.5000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	1.0000	1.0000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	1.0000	0.6667	0.8000	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Emphasis	Bye	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**282**	3	9	20	11	8	17	7	6	9	11	4	1	1	4	0.7176	0.8868	0.7932	82.24%
System	1	**273**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9964	0.9891	0.9927	
Greet	8	0	**138**	1	0	1	1	0	1	1	0	0	0	0	0	0.9139	0.9200	0.9169	
Emotion	7	0	0	**73**	0	0	0	0	0	0	0	0	0	0	0	0.9125	0.7526	0.8249	
ynQuestion	6	0	1	0	**29**	4	0	0	0	1	1	0	0	0	0	0.6905	0.5686	0.6237	
whQuestion	1	0	0	0	10	**46**	0	0	0	1	0	1	1	0	0	0.7667	0.7797	0.7731	
Accept	5	0	0	1	1	0	**7**	0	0	0	2	1	0	0	0	0.4118	0.2593	0.3182	
Emphasis	4	0	0	1	0	0	0	**3**	1	1	0	0	0	0	0	0.3000	0.3000	0.3000	
Bye	1	0	2	0	0	0	0	0	**15**	0	0	0	0	0	0	0.8333	0.6522	0.7317	
Continuer	2	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0.7500	0.3000	0.4286	
Reject	1	0	0	1	0	0	0	0	0	0	**3**	0	1	0	0	0.5000	0.1579	0.2400	
yAnswer	0	0	0	0	0	0	2	0	0	1	0	**2**	0	0	0	0.4000	0.2500	0.3077	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**2**	0	1	0.4000	0.4000	0.4000	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	undef	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.1667	0.2857	

Figure C.36: Experiment Run 30: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**263**	4	0	20	8	7	17	5	9	9	12	2	4	4	0	0.7225	0.8946	0.7994	83.72%
System	2	**298**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9933	0.9868	0.9900	
Greet	8	0	**114**	0	0	0	0	0	2	2	1	0	0	0	1	0.8906	0.9421	0.9157	
Emotion	4	0	4	**84**	0	0	0	0	0	0	0	0	0	0	0	0.9130	0.7778	0.8400	
ynQuestion	6	0	0	0	**39**	4	0	0	0	1	0	1	0	0	0	0.7647	0.7647	0.7647	
whQuestion	1	0	0	0	1	**52**	0	0	0	0	0	0	0	0	0	0.9630	0.8254	0.8889	
Accept	7	0	0	0	1	0	**8**	0	1	0	2	3	0	0	0	0.3636	0.2963	0.3265	
Bye	0	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.9333	0.7368	0.8235	
Emphasis	0	0	2	4	1	0	0	0	**6**	0	0	1	0	0	0	0.4286	0.3000	0.3529	
Continuer	0	0	0	0	0	0	1	0	1	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**3**	0	3	0	0	0.3333	0.1667	0.2222	
yAnswer	1	0	0	0	0	0	1	0	0	0	0	**3**	0	0	0	0.6000	0.2727	0.3750	
nAnswer	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	1.0000	0.3000	0.4615	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**268**	3	1	19	7	9	15	5	11	8	10	2	4	3	1	0.7322	0.9116	0.8121	84.28%
System	0	**299**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9967	0.9901	0.9934	
Greet	3	0	**115**	2	0	1	0	0	0	2	1	0	0	0	0	0.9274	0.9504	0.9388	
Emotion	4	0	2	**82**	0	0	0	0	2	0	0	0	0	0	0	0.9111	0.7593	0.8283	
ynQuestion	6	0	0	0	**38**	3	0	0	0	2	0	1	0	0	0	0.7600	0.7451	0.7525	
whQuestion	1	0	0	0	4	**50**	0	0	0	0	0	0	0	0	0	0.9091	0.7937	0.8475	
Accept	5	0	0	0	1	0	**11**	0	1	0	2	3	0	0	0	0.4783	0.4074	0.4400	
Bye	1	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0	0.8750	0.7368	0.8000	
Emphasis	0	0	2	4	0	0	0	0	**5**	0	0	1	0	0	0	0.4167	0.2500	0.3125	
Continuer	1	0	0	0	0	0	1	0	0	**7**	0	1	0	0	0	0.7000	0.3684	0.4828	
Reject	1	0	0	0	1	0	0	0	1	0	**4**	0	3	0	0	0.4000	0.2222	0.2857	
yAnswer	2	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0.6000	0.2727	0.3750	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.6000	0.3000	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.2500	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	1.0000	0.5000	0.6667	

Figure C.37: Experiment Run 35: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Accuracy
Statement	**279**	3	7	10	15	3	10	4	8	7	16	4	1	0	6	0.7480	0.8532	0.7971	83.66%
System	3	**264**	2	0	0	0	0	0	0	0	0	0	0	0	0	0.9814	0.9814	0.9814	
Greet	16	0	**131**	1	0	0	0	0	0	0	1	0	0	0	0	0.8792	0.9034	0.8912	
Emotion	6	0	1	**99**	0	0	0	0	1	0	0	0	0	1	1	0.9083	0.8839	0.8959	
ynQuestion	4	2	1	0	**44**	1	0	0	1	0	0	0	0	0	0	0.8302	0.7097	0.7652	
whQuestion	1	0	1	0	2	**44**	0	0	1	0	0	1	0	0	0	0.8800	0.8980	0.8889	
Accept	6	0	0	0	1	0	**13**	0	1	1	1	2	0	0	0	0.5200	0.5200	0.5200	
Bye	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0	1.0000	0.7778	0.8750	
Emphasis	3	0	2	2	0	0	0	0	**6**	0	0	0	0	0	0	0.4615	0.3158	0.3750	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	3	0	0	0	0	0	0	0	1	0	**2**	0	1	0	0	0.2857	0.0952	0.1429	
yAnswer	1	0	0	0	0	1	2	0	0	0	0	**5**	0	0	0	0.5556	0.4167	0.4762	
nAnswer	0	0	0	0	0	0	0	0	0	1	1	0	**0**	0	0	0.0000	0.0000	undef	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	3	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Other	Clarify	Precision	Recall	F-score	Overall Acc
Statement	**276**	2	5	11	14	4	10	4	11	8	15	5	0	0	6	0.7439	0.8440	0.7908	82.92%
System	2	**265**	3	1	0	0	0	0	0	0	0	0	0	0	0	0.9779	0.9851	0.9815	
Greet	16	0	**131**	0	0	0	1	0	0	0	1	0	0	0	0	0.8792	0.9034	0.8912	
Emotion	3	0	1	**98**	1	0	0	0	1	0	0	0	0	1	1	0.9245	0.8750	0.8991	
ynQuestion	10	1	3	0	**42**	2	0	0	1	0	0	0	0	0	0	0.7119	0.6774	0.6942	
whQuestion	0	1	0	0	4	**42**	0	0	0	0	0	1	0	0	0	0.8750	0.8571	0.8660	
Accept	11	0	0	1	1	0	**12**	0	1	1	1	2	0	0	0	0.4000	0.4800	0.4364	
Bye	1	0	0	0	0	0	1	**14**	0	0	0	0	0	0	0	0.8750	0.7778	0.8235	
Emphasis	2	0	2	1	0	0	0	0	**5**	0	0	0	0	0	0	0.5000	0.2632	0.3448	
Continuer	2	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0.6667	0.3077	0.4211	
Reject	1	0	0	0	0	0	0	0	0	0	**3**	0	1	0	0	0.6000	0.1429	0.2308	
yAnswer	0	0	0	0	0	1	1	0	0	0	0	**4**	0	0	0	0.6667	0.3333	0.4444	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**1**	0	0	0.3333	0.5000	0.4000	
Other	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	1.0000	0.5000	0.6667	
Clarify	2	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0.0000	0.0000	undef	

Figure C.38: Experiment Run 40: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**285**	4	7	19	8	6	5	4	7	11	11	3	2	1	0	0.7641	0.8559	0.8074	83.77%
System	4	**270**	0	1	0	0	0	0	0	0	0	0	0	0	0	0.9818	0.9854	0.9836	
Greet	11	0	**109**	2	0	1	0	0	0	1	0	0	0	0	0	0.8790	0.8790	0.8790	
Emotion	4	0	2	**94**	0	0	0	0	1	0	0	0	0	0	0	0.9307	0.7899	0.8545	
ynQuestion	10	0	1	0	**35**	2	0	0	0	0	0	0	0	1	0	0.7143	0.7609	0.7368	
whQuestion	0	0	3	0	2	**48**	0	0	0	1	0	0	0	0	0	0.8889	0.8421	0.8649	
Accept	8	0	0	0	0	0	**14**	0	0	0	0	2	1	0	0	0.5600	0.6087	0.5833	
Bye	1	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	0.9412	0.7273	0.8205	
Emphasis	2	0	2	2	0	0	0	0	**8**	1	1	0	0	0	0	0.5000	0.4706	0.4848	
Continuer	4	0	0	0	1	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	1	0	0	0	0	0	0	0	1	0	**1**	0	0	0	0	0.3333	0.0714	0.1176	
yAnswer	0	0	0	1	0	0	3	0	0	1	0	**3**	0	0	0	0.3750	0.3750	0.3750	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	2	0	0	0	0	0	0	1	0	0	0	0	0	**1**	0	0.2500	0.3333	0.2857	
Other	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**4**	0.8000	1.0000	0.8889	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**292**	4	7	21	9	3	4	6	9	10	12	4	2	2	2	0.7545	0.8769	0.8111	83.49%
System	2	**270**	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9926	0.9854	0.9890	
Greet	8	0	**110**	2	0	1	0	0	0	1	0	0	0	0	0	0.9016	0.8871	0.8943	
Emotion	3	0	2	**90**	0	0	0	0	0	0	0	0	0	0	0	0.9474	0.7563	0.8411	
ynQuestion	11	0	1	0	**29**	1	0	0	0	1	0	0	0	1	0	0.6591	0.6304	0.6444	
whQuestion	1	0	2	0	8	**52**	0	0	0	1	0	0	0	0	0	0.8125	0.9123	0.8595	
Accept	8	0	0	0	0	0	**15**	0	0	0	0	2	0	0	0	0.6000	0.6522	0.6250	
Bye	0	0	0	0	0	0	0	**16**	0	0	0	0	0	0	0	1.0000	0.7273	0.8421	
Emphasis	2	0	2	3	0	0	0	0	**7**	1	1	0	0	0	0	0.4375	0.4118	0.4242	
Continuer	4	0	0	1	0	0	1	0	0	**4**	0	0	0	0	0	0.4000	0.2105	0.2759	
Reject	1	0	0	0	0	0	0	0	1	0	**0**	0	1	0	0	0.0000	0.0000	undef	
yAnswer	0	0	0	0	0	0	3	0	0	1	0	**2**	0	0	0	0.3333	0.2500	0.2857	
nAnswer	1	0	0	0	0	0	0	0	0	0	1	0	**6**	0	0	0.7500	0.6667	0.7059	
Clarify	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	#DIV/0!	0.0000	undef	
Other	0	0	0	2	0	0	0	0	0	0	0	0	0	0	**2**	0.5000	0.5000	0.5000	

Figure C.39: Experiment Run 45: Emoticons Assigned “EMO” or “EMO2” Tags

Using Actual POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Accuracy
Statement	**266**	4	7	13	11	3	6	7	9	11	7	4	4	1	2	0.7493	0.8721	0.8061	83.09%
System	0	**238**	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000	0.9835	0.9917	
Greet	10	0	**131**	0	0	0	0	0	0	1	0	1	0	0	1	0.9097	0.9161	0.9129	
Emotion	3	0	1	**92**	0	0	1	0	1	0	0	0	1	0	0	0.9293	0.8288	0.8762	
ynQuestion	2	0	1	0	**29**	6	0	0	1	0	0	0	1	1	0	0.7073	0.6744	0.6905	
whQuestion	1	0	2	0	3	**57**	0	0	0	0	0	0	0	0	0	0.9048	0.8382	0.8702	
Accept	8	0	0	1	0	0	**5**	0	0	0	2	4	0	0	0	0.2500	0.3333	0.2857	
Bye	0	0	0	1	0	0	0	**10**	0	0	0	0	0	0	0	0.9091	0.5556	0.6897	
Emphasis	2	0	1	1	0	0	0	0	**9**	0	0	0	0	0	0	0.6923	0.4286	0.5294	
Continuer	4	0	0	0	0	0	1	0	0	**5**	0	0	0	0	0	0.5000	0.2941	0.3704	
Reject	6	0	0	0	0	1	0	1	0	0	**3**	0	2	0	0	0.2308	0.2143	0.2222	
yAnswer	0	0	0	0	0	1	2	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	2	0	**3**	0	0	0.6000	0.2727	0.3750	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0.5000	0.3333	0.4000	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

Using Cheap POS tags

	Statement	System	Greet	Emotion	ynQuestion	whQuestion	Accept	Bye	Emphasis	Continuer	Reject	yAnswer	nAnswer	Clarify	Other	Precision	Recall	F-score	Overall Acc
Statement	**269**	5	7	16	12	2	5	8	10	9	8	4	4	2	3	0.7390	0.8820	0.8042	82.60%
System	0	**237**	0	0	0	0	0	0	0	1	0	0	0	0	0	0.9958	0.9793	0.9875	
Greet	7	0	**131**	0	0	1	0	0	0	1	0	0	0	0	0	0.9357	0.9161	0.9258	
Emotion	2	0	1	**90**	0	0	1	0	1	0	0	0	1	0	0	0.9375	0.8108	0.8696	
ynQuestion	4	0	2	0	**24**	4	0	0	1	0	0	0	0	1	0	0.6667	0.5581	0.6076	
whQuestion	1	0	1	0	7	**58**	0	1	0	0	0	0	0	0	0	0.8529	0.8529	0.8529	
Accept	9	0	0	1	0	0	**7**	0	0	0	2	5	0	0	0	0.2917	0.4667	0.3590	
Bye	0	0	0	0	0	0	0	**9**	0	0	0	0	0	0	0	1.0000	0.5000	0.6667	
Emphasis	2	0	1	1	0	0	0	0	**8**	0	1	0	0	0	0	0.6154	0.3810	0.4706	
Continuer	5	0	0	0	0	0	1	0	0	**6**	0	0	0	0	0	0.5000	0.3529	0.4138	
Reject	3	0	0	0	0	1	0	0	0	0	**2**	0	3	0	0	0.2222	0.1429	0.1739	
yAnswer	0	0	0	0	0	2	1	0	1	0	0	**3**	0	0	0	0.4286	0.2500	0.3158	
nAnswer	0	0	0	0	0	0	0	0	0	0	1	0	**3**	0	0	0.7500	0.2727	0.4000	
Clarify	1	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0.0000	0.0000	#DIV/0!	
Other	2	0	0	3	0	0	0	0	0	0	0	0	0	0	**3**	0.3750	0.5000	0.4286	

Figure C.40: Experiment Run 50: Emoticons Assigned “EMO” or “EMO2” Tags

LIST OF REFERENCES

[1] P. Adams, "Conversational thread extraction and topic detection in text-based chat," Master's thesis, Naval Postgraduate School, Monterey, CA, 2008.

[2] B. Eovito, "An assessment of joint chat requirements from current usage patterns," Master's thesis, Naval Postgraduate School, Monterey, CA, 2006.

[3] S. Herring, *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*. John Benjamins Publishing Co, 1996.

[4] T. Kucukyilmaz, B. B. Cambazoglu, C. Aykanat, and F. Can, "Chat mining: Predicting user and message attributes in computer-mediated communication," *Information Processing & Management*, vol. 44, no. 4, pp. 1448–1466, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/B6VC8-4S02DDT-1/2/a5f27b8d3a925f8f2281e2e0618c71a8

[5] J. Lin, "Automatic author profiling of online chat logs," Master's thesis, Naval Postgraduate School, Monterey, CA, 2007.

[6] E. N. Forsyth, "Improving automated lexical and discourse analysis of online chat dialog," Master's thesis, Naval Postgraduate School, Monterey, CA, 2007.

[7] B. Santorini, "Part-of-speech tagging guidelines for the Penn treebank project," *University of Pennsylvania, 3rd Revision, 2nd Printing*, 1990.

[8] T. Wu, F. M. Khan, T. A. Fisher, L. A. Shuler, and W. M. Pottenger, "Posting act tagging using transformation-based learning," *The Proceedings of the Workshop on Foundations of Data Mining and Discovery.*, December 2002.

[9] E. Alpaydin, *Introduction to machine learning*. Cambridge, MA: The MIT Press, 2004.

[10] D. Jurafsky and J. H. Martin, *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Second edition*. Upper Saddle River, New Jersey: Pearson Education, Inc, 2009.

[11] S. Fugate, B. Gordon, and M. Snider, "Performing part of speech tagging on chat corpora and improving text-to-speech of chat dialog," *UNKNOWN*, 2009.

[12] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Ess-Dykema, and M. Meteer, "Dialogue act modeling for automatic tagging and recognition of conversational speech," *Computational linguistics*, vol. 26, no. 3, pp. 339–373, 2000.

[13] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python*. Sebastopol, CA: O'Reilly & Associates, Inc., 2009.

[14] S. Russell and P. Norvig, *Artificial intelligence: A modern approach*, 2nd ed. Pearson Education, Inc., 2003.

[15] I. Witten and T. Bell, "The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression," *IEEE Transactions on Information Theory*, vol. 37, p. 1085, 1991.

[16] C. Manning and H. Schütze, *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press, 1999.

[17] Wikipedia, "(15 july 2010). support vector machine," http://en.wikipedia.org/wiki/Support_vector_machine, July 2010, (accessed 17 July 2010).

[18] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc., 2008.

[19] F. Jelinek, *Statistical methods for speech recognition*. Massachusetts Institute of Technology, 1997.

[20] T. Mitchell, *Machine learning*. McGraw-Hill, 1997.

[21] J. Pearl, "Probabilistic reasoning in intelligent systems: Networks of plausible inference," *Morgan Kauffman, San Francisco*, 1988.

[22] R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, 1986.

[23] L. Adam, B. Berger, and V. Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[24] A. Ratnaparkhi, "A simple introduction to maximum entropy models for natural language processing," *IRCS Report*, pp. 97–08, 1997.

[25] Z. Harris, “Distributional structure,” *Word*, vol. 10, no. 23, pp. 146–162, 1954.

[26] F. Yang, G. Tür, and E. Shriberg, “Exploiting dialogue act tagging and prosodic information for action item identification,” in *ICASSP*. IEEE, 2008, pp. 4941–4944. [Online]. Available: http://dx.doi.org/10.1109/ICASSP.2008.4518766

[27] M. Walker and R. Passonneau, “DATE: A dialogue act tagging scheme for evaluation of spoken dialogue systems,” Oct. 23 2001. [Online]. Available: http://citeseer.ist.psu.edu/462074.html;http://www.research.att.com/~walker/dtag6.pdf

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Dr. Craig Martell
   Naval Postgraduate School
   Monterey, California

4. J.R. Hitt
   Naval Postgraduate School
   Monterey, California