

CLASSIFICATION OF DIALOGUE ACTS IN URDU MULTI-PARTY DISCOURSE

Samira Shaikh¹, Tomek Strzalkowski^{1,2} and Nick Webb¹

¹State University of New York, University at Albany, New York, U.S.A.

²Polish Academy of Sciences, Warsaw, Poland

Keywords: Dialogue-Act Tagging, Urdu, Classification, Multi-Party Discourse.

Abstract: Classification of dialogue acts constitutes an integral part of various natural language processing applications. In this paper, we present an application of this task to Urdu language online multi-party discourse. With language specific modifications to established techniques such as permutation of word order in detected n-grams and variation of n-gram location, we developed an approach that is novel to this language. Preliminary performance results when compared to baseline are very encouraging for this approach.

1 INTRODUCTION

Urdu, belonging to the family of Indo-European languages, has an estimated 487 million speakers worldwide, next in line after English. In our context, Urdu is the modern variant, more generally referred to as Hindi/Urdu. The need arises for the development of robust natural language tools aimed at understanding and investigation of the language.

Social interaction in an increasingly online world also provides a rich resource for research. The dynamics of small group interaction have been well studied for spoken and face-to-face conversation. However, for a reduced-cue environment such as online chat in a virtual chat room, these dynamics are obtained distinctly, and require explicit linguistic devices to convey social and cultural nuances. Indeed, how are social behaviours exhibited and conveyed when the only medium of communication is language?

Our overall objective is to develop computational models of how certain social phenomena are manifested in language through the choice of linguistic, semantic and conversational forms by discourse participants. The social behaviors we are currently studying include, among others, Topic Control, Task Control, Disagreement and Involvement. These are, in turn, utilized to predict higher-level social phenomena such as leadership and group cohesion. Dialogue act tagging forms an

essential component of our prototype system. Using dialogue acts to model the functional aspect of an utterance in discourse, we can arrive at determinations of socio-linguistic behaviors by the participants. For example, we posit that an equal amount of agreement and disagreement between all participants of a discourse, points towards a more cohesive group; as opposed to a discourse characterized by an inordinate amount of disagreement or agreement. It is essential that the data corpus used contain the discourse features we are interested in modeling, which led us to collect our own data. Another requirement was that the discourse participants be native speakers of the target language, so that natural and spontaneous discourse may be obtained. We also developed a hierarchy of dialogue acts that are tuned significantly towards dialogue pragmatics and eschew syntactic variations.

This paper pertains to Urdu online chat conversations; we selected Roman Urdu, as this is the preferred form of writing used in most Urdu chat rooms. We use a cue-phrase based method, using n-grams as features and enhance it by adding a word order alteration feature specifically targeting the Urdu grammar structure. Classification of dialogue acts in the Urdu language is a novel task that has not been hitherto addressed. While our approach is preliminary, we are quite encouraged by the performance.

2 RELATED WORK

We model our cue-based approach in line with the work by researchers in the field. Fraser (1990) showed that discourse markers are “part of the grammar of a language”. Grosz and Sidner (1986) proposed ways in which discourse may be segmented into constituent structures. Several researchers (Heeman et al., 1998; Hirschberg and Litman, 1993; Marcu, 1997; Reichman, 1985; Schiffrin, 1987; Warner, 1985; Zukerman and Pearl, 1986) identified and selected cue phrases in dialogue that are generally useful for dialogue processing. Samuel, Carberry and Vijay-Shanker (1999) used n -grams in utterances to automatically detect cue phrases. Webb, Hepple and Wilks (2005) had a similar approach of identifying cue phrases using intra-utterance features and determining the n most likely dialogue acts for an utterance. All of these works point to learning features that are specific to the language under consideration. A morphologically rich language such as Urdu requires techniques that both exploit and overcome its structure.

Work in dialogue act classification in languages such as Urdu is still nascent. In fact, we have been able to discover no prior work towards dialogue act classification in Urdu discourse. Somewhat related is the use of n -grams in Urdu for authorship attribution in Urdu poetry (Raza et al., 2009a). Word segmentation in Urdu is an issue that affects machine-learning algorithms (Durrani and Hussain, 2010). However, the use of Roman Urdu in our corpus mitigates this issue.

3 DATA AND ANNOTATION

Our initial focus has been on on-line chat dialogues. Chat data, although plentiful on-line, its adaptation for research purposes present a number of challenges. On the one hand there are users’ privacy issues, and their complete anonymity on the other. Furthermore, most data that may be obtained from public chat-rooms is of limited value for the type of modeling tasks we are interested in due to its high-level of noise, lack of focus, and rapidly shifting, chaotic nature, which makes any longitudinal studies virtually impossible. To derive complex models of conversational behavior, we needed the interaction to be reasonably focused on a task and/or social objectives within a group.

Few data collections exist covering multiparty dialogue, and even fewer with on-line chat. Moreover, the few collections that exist were built primarily for the purpose of training dialogue act tagging and similar linguistic phenomena; few if any of these corpora are suitable for deriving pragmatic models of conversation, including socio-linguistic phenomena. Existing resources include a multi-person meeting corpus ICSI-MRDA (Janin et al., 2003) and the AMI Meeting Corpus (Carletta, 2007), which contains 100 hours of meetings captured using synchronized recording devices. Still, all of these resources look at spoken language rather than on-line chat. Some corpora exist such as the NPS Internet chat corpus (Forsyth and Martell, 2007), which has been hand-anonymized and labeled with part-of-speech tags and dialogue act labels. The StrikeCom corpus (Twitchell et al., 2007) consists of 32 multi-person chat dialogues between players of a strategic game, where in 50% of the dialogues one participant has been asked to behave ‘deceptively’. These are resources in the English language; some of the corpora that exist in Urdu are aimed towards tasks such as part of speech tagging and lexicon building (Hussain, 2008; Raza et al., 2009b; Ijaz and Hussain, 2007). Few, if any of these corpora are suitable for deriving pragmatic models of conversation, including socio-linguistic phenomena. It is thus more typical that those interested in the study of Internet chat compile their own corpus on an as needed basis, e.g., Khan et al. (2002), Kim et al. (2007).

We designed a series of experiments in which recruited subjects were invited to participate in a series of on-line chat sessions in a specially designed secure chat-room. The experiments were carefully designed around topics, tasks, and games for the participants to engage in so that appropriate types of behavior, e.g., disagreement, power play, persuasion, etc. may emerge spontaneously. Discussions were centered on a range of topics that included issues relevant to native speakers of Urdu, such as the “Value of the Burka in Modern-Day Women” and “Politics of Pakistan under Prime Minister Zardari” as well as task-oriented topics such as choosing the best candidate for a given job from an array of resumes. These experiments and the resulting corpus have been described in a separate publication. We assembled a corpus of 20 hours of Urdu chat, consisting of 40,000+ words, 6000+ turns and 9 different participants, over the course of fourteen 90-minute chat sessions.

Figure 1 shows a fragment of one Urdu dialogue, where 6 participants in the chat session discuss the

selection of a candidate for given job description. Note the use of short sentences, lack of punctuation and capitalization and typically improper grammar. In addition, emoticons (e.g. ☺, ☹), misspellings, and abbreviations are also common.

5. **MM:** *mujhe carla ka resume pasand hai ☺*
(**MM:** I like resume of carla ☺)
6. **MM:** *experience ke hisab se*
(**MM:** based on experience)
7. **RI:** *mujhe bhi*
(**RI:** me too)
8. **SA:** *ji carla ka tek hai*
(**SA:** yes, (resume) of carla is fine)

Figure 1: A fragment of Urdu conversation.

We have annotated 5 Urdu dialogues (2000+ turns) in total which were used for training our Urdu modules. There is need of more training data, and we intend to collect and annotate more Urdu dialogues. All annotation was done using a specially designed annotation tool for the purpose, by two trained annotators who are native speakers of the language. The inter-annotator agreement for dialogue acts is 0.82 alpha (Krippendorf, 2005), which is adequate for the training we need to accomplish.

4 DIALOGUE ACT HEIRARCHY

The functional or dialogic aspect of an utterance has to do with its role or purpose in conversation. Statements, questions, answers, offers, acceptances and rejections, as well as expressions of thanks are all examples of such functions in a dialogue. Our objective is to capture how an utterance functions in dialogue, which may or may not be directly related to its form. For example, the utterance “Can you close the window?” may function as a question or as a directive, depending upon the context in which it is used.

We developed a hierarchy of 15 dialogue acts adapted to better capture significant social nuances within conversation. Syntactic distinctions between categories, e.g., *wh-questions* vs. *yes-no questions*, etc are avoided. The tagset we adopted is based on DAMSL (Allen and Core, 1997) and SWBD (Jurafsky, Shriberg and Biasca, 1997), but compressed to 15 tags tuned significantly towards

dialogue pragmatics and away from more surface characteristics of utterances.

Our classification of dialogue acts is bi-fold (Figure 2). At the Information-Level (Figure 3), we seek to identify the purpose of an utterance in relation to the task given to the participants.

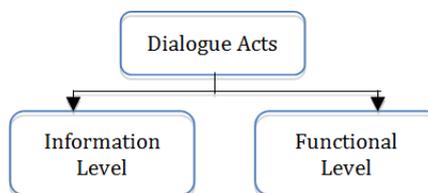


Figure 2: Dialogue act levels.

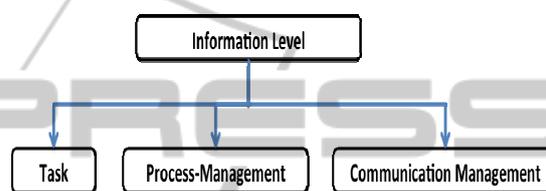


Figure 3: Information-Level dialogue act categories.

At the Functional-Level, we classify Dialogue Acts into three hierarchical categories (see Table 1 below): (a) Statements-and-Responses, (b) Questions-and-Directives, and (c) Conversational-Norms. Each of these categories consists of several top-level tags and may also contain specialized tags under these. This makes a total of 15 dialogue acts at the Functional-Level. In addition, there are 3 dialogue acts at the Information-Level.

The Assertion-Opinion category contains four specialized tags under it, A.1.1 Response-Answer, A.1.2 Response-Non-Answer, A.1.3 Agree-Accept and A.1.4 Disagree-Reject. For an utterance, a specialized tag is preferably applied wherever pertinent. For example, the utterance “*mein aapse sehmat hu us baat par*” (I agree with you on that) functions as an assertion, as well as an agreement; and is assigned the tag Agree-Accept rather than Assertion-Opinion. A full description of these dialogue acts is beyond the scope of this paper, and will be the subject of a future publication. It is important to note that the annotation and categories have been developed to support the objectives of our project and do not necessarily conform to other similar classification systems used in the past.

Each utterance in dialogue is assigned two dialogue acts, one at the Information-Level and one at the Functional-Level. Figure 4 shows the annotation applied to the conversation fragment in Figure 1.

Table 1: Functional-Level dialogue act categories.

Category	Top Level
A. Statements and Responses	A.1 Assertion-Opinion
	-A.1.1 Response-Answer
	-A.1.2 Response-Non-Answer
	-A.1.3 Agree-Accept
	-A.1.4 Disagree-Reject
	A.2 Offer-Commit
	A.3 Acknowledge
	A.4 Signal-Non-Understanding
B. Questions and Directives	B.1 Information-Request
	B.2 Confirmation-Request
	B.3 Action-Directive
C. Conversational Norms	C.1 Conventional-Opening
	C.2 Conventional-Closing
	C.3 Other-Conventional-Phrase
	C.4 Correct-Misspelling

5.	MM: <i>mujhe carla ka resume pasand hai ☺</i> (MM: I like resume of carla ☺) DA Info-Level: Task; Func-Level: Assertion-Opinion
6.	MM: <i>experience ke hisab se</i> (MM: based on experience) DA Info-Level: Task; Func-Level: Assertion-Opinion
7.	RI: <i>mujhe bhi</i> (RI: me too) DA Info-Level: Task; Func-Level: Agree-Accept
8.	SA: <i>ji carla ka tek hai</i> (SA: yes, (resume) of carla is fine) DA Info-Level: Task; Func-Level: Agree-Accept

Figure 4: Annotation of conversation fragment showing Information-Level and Functional-Level tags.

5 LEARNING CUE PHRASES

We use annotated data to learn cue phrases in each dialogue act category. We generate n-grams of varying length from the utterances, discarding stop words, emoticons (e.g. ☺, ☹ etc) and some prepositions. This process generated ~11500 n-grams. The n-grams of length no greater than 3 are saved and ranked in order of their frequency and length. The threshold has been determined experimentally and varies with the dialogue act under consideration. Some dialogue acts appear more frequently in data and generate higher number of n-grams, than those that are infrequent. Frequency

and n-gram length are generally inversely proportional to each other. However, n-grams of length greater or equal to 2 are preferable, due to their high accuracy and predictive nature and thus we use a lower threshold for longer length n-grams. We use the most frequent n-grams that appear in utterances tagged with a specific dialogue act and the most predictive ones. Frequency values fluctuate significantly. For example, for the Conventional-Opening category of dialogue act in Conversational-Norms, the n-grams are stable and highly predictive. To give an idea about the spread of frequency, the most frequent Information-Level dialogue act tag assigned in our annotated corpus is Task. The frequency of the most frequent n-gram learned for this tag is 392. The least applied dialogue act tag in our corpus is Signal-Non-Understanding (at the Functional-Level). The frequency of the most frequent n-gram for this tag is 2. Note the frequency distribution of learned n-grams shown in Figure 5, which follows Zipf’s law (1949) with a long tail of the curve. To get the best performance, we select the most frequent n-grams from the head of the curve, and the highly predictive (i.e. greater length) yet less frequent n-grams from the tail of the curve. Very low frequency unigrams are not selected.

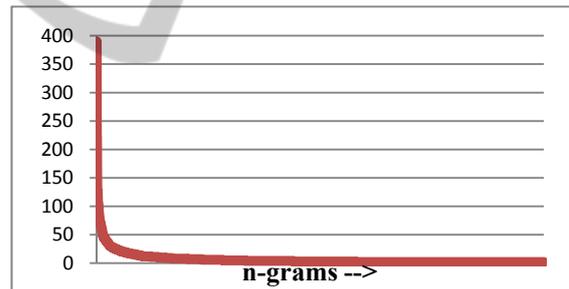


Figure 5: Frequency of learned n-grams.

We are currently using absolute frequency counts of n-grams in our determinations, we may replace them with normalized counts or percentages in the next prototype.

Some dialogue act classification systems (Stolcke et al., 2000; Samuel, Carberry and Vijay-Shanker, 1999), place <start> and <end> tags, to determine the position in utterance where the n-gram should occur. While this provides a salient handle over the utterance in the English language, Urdu grammar is not restrictive on word order, and using such a mechanism presents a challenge.

To illustrate, consider the sentences below:

1. *mein aapse sehmata hu us baat par*
(I agree with you on that)
2. *us baat par mein aapse sehmata hu*

(On that I agree with you)

Both are valid utterances in Urdu grammar, wherein the tri-gram “*mein aapse sehmat*” (I agree with you) occurs in different positions. This non-restrictive word order voids the use of markers. Accordingly, we do not utilize the <start> and <end> markers for the learned n-grams. They may occur at any position in an utterance.

Another modification, made specifically for Urdu is adding new n-grams by changing the word order in the learned n-grams. This is a modification to overcome the lack of training data. The post-positions applied as suffixes to Urdu words, are a parallel to English prepositions.

To that end, the two phrases below:

1. *chalo karte hain*
(come let's do)
2. *karte hain chalo*
(let's do come)

are both likely to occur in an utterance.

We learn additional n-grams by deriving permutations of existing n-grams. These then add to the frequency count of the original n-gram, although there are ways in which this frequency assignment can be refined. Adding to the frequency of the original n-gram, instead of treating them as separate instances is practical, since there may be permutations that occur very rarely in text. “*hain karte chalo*” (one other permutation of the above sentences) does not usually occur in the type of colloquial or informal text we are looking at; it may occur in highly stylized forms of text such as Urdu poetry and as such does not warrant treatment as separate a n-gram.

Notably, both modifications described above may notionally be applied for a similar task to any language with a similar grammar and post-positional suffixes. (Turkish and Japanese are examples).

The n-grams that have a frequency above a certain threshold act as cue phrases for that dialogue act. Using the above mechanism to extract cue phrases, we use a method where these cue phrases act as features for machine-learning algorithms. Other researchers (Samuel et al., 1999) have also used this method of passing their cue phrases as a feature to a machine learning method. If the extracted cues are reliable in identifying dialogue acts, then a classifier that uses these cues directly should perform reasonably well.

Table 2 shows a few n-grams that have been learned for the dialogue act category Action-Directive with their English translations. Note that a phrase in Urdu may have different meanings in

English, depending on the context. A total of ~180 n-grams were selected as cue phrases for the various dialogue acts in our corpus.

Table 2: Examples of n-grams learned for the Action-Directive dialogue act, their English translations and frequencies.

Urdu n-gram	English	Frequency
kar sakte ho	will you do/ you will do/ you may do	12
karo	will do	31
aap log	you people/ you guys/ you	7
chalo karte hain	let's do/ come let's do/ let's do come	4
karoge	will you do/ you will do	18
padh lo	read/ you read/ read it	3

6 EVALUATION AND RESULTS

Using the most frequently applied dialogue act tags in the corpus, we can derive a baseline result. We simply assign the Task tag for the Information-Level and the Assertion-Opinion tag for the Functional-Level on the test data set and compute performance accuracy. This serves as the baseline to compare results against. The results are shown in the first column in Table 3.

We use Weka (www.cs.waikato.ac.nz/ml/weka/) machine-learning software to run our classification algorithms and use the NaiveBayesMultinomial classifier given in the software kit. Results are shown in the third column in Table 3, using 10-fold cross validation on the data corpus. To assess the improvement in performance by the addition of cue phrases, we first ran the classifier using simply the entire utterance as a feature (results shown in the second column of Table 3 below). Performance at the Information-Level is much better, as there are 3 classes (categories) of dialogue acts, as opposed to the Information-Level where there are 15 categories. And we note that using the entire utterance as a feature to predict dialogue acts performs poorly. This is to be expected, due to the noise presented by the extraneous words in the utterance. However, it

does perform better than the baseline.

Table 3: Accuracy of dialogue act classifier using our approach vs. a simple baseline.

	Baseline	NaiveBayes, utterance as feature	NaiveBayes, cue phrases as features
Info-Level	57.2%	60.3%	88.0%
Func-Level	29.6%	37.4%	75.1%

In the above Table 3, the cue phrases that act as features do not utilize the modification of frequency counts by using permuted n-grams technique we described earlier. Even so, there is a significant gain over the baseline, which indicates that the selected cue phrases are highly predictive indicators for dialogue acts in our corpus. Table 4 shows the performance after the addition of the n-gram permutation module. There is a solid albeit small increase in performance accuracy. This is likely due to the small size of training data corpus. A big percentage of the cue-phrases we have used are unigram words, whose frequencies are unaffected by this modification. However, the 3% of the total trigrams from our selected n-grams show an increase in frequency counts. This may push some of the n-grams above the threshold and result in their selection as cue phrases.

As an example, the trigrams:

1. *kaise hai aap?*
(how are you?)
2. *aap kaise hai?*
(you how are?)

both valid sentences and both occur as indicators for the Conventional-Opening tag. Since they are essentially permutations of each other, we can add their frequency counts. This results in the selection of this very accurate trigram as a cue-phrase for the Conventional-Opening tag, where it was not previously chosen.

Table 4: Accuracy of dialogue act classifier using permutation of n-grams modification.

	Without permutation of n-grams modification	With permutation of n-grams modification
Info-Level	88.0%	89.4%
Func-Level	75.1%	77.3%

Our goal was to establish the strength of

cue-phrases as features for this task. Using n-grams is a natural choice since they provide an understanding of the characteristics of the underlying utterance. Our method overcomes the challenges presented by the highly irregular structure of language used in virtual chat rooms by filtering out noise present in superfluous words, emoticons and stop words and using only the crucial words that are also highly predictive to act as cue-phrases. While the cue-phrases are significant intra-utterance features, we plan to expand feature set by adding inter-utterance features as well. This serves to add the context of conversation to the classifier. We have also annotated the corpus for communication links which indicate who is speaking to whom and whether the utterance is addressed to a subset of speakers, a response to a prior utterance or a continuation of the speaker’s own prior utterance. This can provide additional evidence for the classifier.

7 CONCLUSIONS

We presented an approach to the dialogue act classification task in Urdu language. This is an application novel for this language. We also described certain modifications designed to address Urdu grammar. Lack of sufficient data for training and testing is an issue. However, current performance results are encouraging and provide insight towards future modifications. One enhancement would be to improve the selection of cue phrases, using additional indicators that complement the frequency counts we currently use. We hope to test our algorithms on a significantly larger data set to further validate the hypotheses and mechanisms. Our contention is that this is a very promising first attempt at the dialogue act classification task in a language and grammar previously uncharted for this task.

REFERENCES

Allen, J. M. Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers. www.cs.rochester.edu/research/cisd/resources/damsl/

Carletta, J. 2007. Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus. *Language Resources and Evaluation Journal* 41(2): 181-190

Durrani, N., Hussain, S. 2010. Urdu Word Segmentation. In the 11th Annual Conference of the North American

- Chapter of the Association for Computational Linguistics (NAACL HLT 2010), Los Angeles, US, 2010
- Eric N. Forsyth and Craig H. Martell. 2007. Lexical and Discourse Analysis of Online Chat Dialog. *First IEEE International Conference on Semantic Computing (ICSC 2007)*, pp. 19-26.
- Fraser, B. 1990. An Approach to Discourse Markers. *Journal of Pragmatics*. 14:383-395
- Grosz, B. and C. Sidner., 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*. 12 (3):175-204.
- Heeman, P., D. Byron, and J. Allen. 1998. Identifying Discourse Markers in Spoken Dialog. In *Applying Machine Learning to Discourse Processing: Papers from the 1998 American Association for Artificial Intelligence Spring Symposium*. 44-51. Stanford, California.
- Hirschberg, J. and D. Litman. 1993. Empirical Studies on the Disambiguation of Cue Phrases. *Computational Linguistics*. 19(3):501-530.
- Hussain, S. 2008. Resources for Urdu Language Processing. In *Proceedings of the 6th Workshop on Asian Language Resources*. IJCNLP'08, IIT Hyderabad, India.
- Ijaz, M and Hussain, S. 2007. Corpus Based Urdu Lexicon Development, In *Proceedings of Conference on Language Technology (CLT07)*, University of Peshawar, Pakistan.
- Janin, A., Baron, D., Edwards D., Gelbart D., Morgan N., Peskin B., Pfau T., Shriberg E., Stolcke A., Wooters C. 2003. The ICSI Meeting Corpus. In *Proc. ICASSP*. Hong Kong.
- Jurafsky, Dan, Elizabeth Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual. <http://stripe.colorado.edu/~jurafsky/manual.august1.html>
- Khan F. M., T. A. Fisher, L. Shuler, T. Wu, and W. M. Pottenger. 2002. Mining chatroom conversations for Social and Semantic Interactions. Technical Report LU-CSE-02-011, Lehigh University.
- Kim, Jihie., Shaw, Erin., Chern, Grace. and Donghui Feng. 2007. An Intelligent Discussion-Bot for Guiding Student Interactions in Threaded Discussions. AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants.
- Krippendorff, K. 2005. Computing Krippendorff's alpha-reliability. Technical Report. University of Pennsylvania. PA. <http://www.asc.upenn.edu/usr/krippendorff/webreliability2.pdf>
- Marcu, D. 1997. *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. Ph.D. thesis. University of Toronto, Toronto, Canada. Tech Report #CSRG-371.
- Raza, A. A., Athar, A., Nadeem, S. 2009a. N-GRAM Based Authorship Attribution in Urdu Poetry. In the Proceedings of the Conference on Language and Technology 2009 (CLT09), FAST NU, Lahore, Pakistan, 22-24 Jan 2009
- Raza, A.A., Hussain, S., Sarfraz, H., Ullah, I., Sarfraz, Z. 2009b. Design and development of phonetically rich Urdu Speech Corpus. In *Proceedings of O-COCOSDA'09*. School of Information Science and Engineering of Xinjiang University, Urumqi, China
- Reichman, R. 1985. *Getting Computers to Talk Like You and Me: Discourse Context, Focus, and Semantics*. MIT Press, Cambridge, Massachusetts.
- Samuel, K.; Carberry, S.; and Vijay-Shanker, K. 1999. Automatically selecting useful phrases for dialogue act tagging. In *Proceedings of the Fourth Conference of the Pacific Association for Computational Linguistics, Waterloo, Ontario, Canada*.
- Schiffrin, D. 1987. *Discourse Markers*. Cambridge University Press, London, England.
- Stolcke A., K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Van Ess-Dykema, & M. Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics*. 26(3), 339-373.
- Twitchell, Douglas P., Jay F. Nunamaker Jr., and Judee K. Burgoon. 2004. Using Speech Act Profiling for Deception Detection. *Intelligence and Security Informatics*. LNCS, Vol. 3073
- Warner, R. 1985. *Discourse Connectives in English*. Garland Publications, New York, New York.
- Webb, N., M. Hepple and Y. Wilks. 2005. Dialogue Act Classification using Intra-Utterance Features. In *Proceedings of the AAAI Workshop on Spoken Language Understanding*, Pittsburgh, USA.
- Zipf, G. 1949. *Human Behavior and the Principle of Least Effort*. Addison-Wesley.
- Zukerman, I. and J. Pearl. 1986. Comprehension-Driven Generation of Meta-Technical Utterances in Math Tutoring. In *Proceedings of the Sixth National Conference of the American Association for Artificial Intelligence*. Philadelphia, Pennsylvania.