Proceedings of the Twenty-Ninth International
Florida Artificial Intelligence Research Society Conference

Supervised Speech Act Classification of Messages in German Online Discussions

Berken Bayat
Fraunhofer FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin, Germany
Email: berken.bayat
@fokus.fraunhofer.de

Christopher Krauss
Fraunhofer FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin, Germany
Email: christopher.krauss
@fokus.fraunhofer.de

Agathe Merceron
Beuth Hochschule fuer Technik
Luxemburger Strasse 10
13353 Berlin, Germany
Email: merceron
@beuth-hochschule.de

Stefan Arbanowski
Fraunhofer FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin, Germany
Email: stefan.arbanowski
@fokus.fraunhofer.de

Abstract

University lectures often offer online discussion forums for students to discuss and solve issues with other students and instructors. Correlating the participation of a student in a discussion forum to his performance in the course is subject of current research. Therefore, to qualify the different parts a student plays in a discussion, be it asking or answering a question, is sought in this paper. In current analysis of online discussion forums, such parts are annotated by hand. Thereby, identifying corresponding roles manually is a costly task, which requires the work of more than one person to annotate and approve the chosen roles. The desired step to a better understanding of student online discussion forums is the automated annotation of student roles. A student's role is determined by classifying the student's message into different speech act categories.

This paper introduces a supervised speech act classification method for messages in German discussion forums that aims at solving the problem of manually detecting speech acts in online discussion for further discourse analysis. A comparative evaluation shows the significant improvements of the new classifier and its appropriateness for the German language.

Introduction

The theory of speech acts was introduced by Austin and Searle to describe the performed act by a speaker's utterance (Searle 1969). Supervised machine learning algorithms presuppose the existence of labeled training data to build a model, which is used to classify unlabeled input data. A raw data set of German discussion messages is provided by Merceron (Merceron 2014). The data is obtained from a university online course, and is consisted of the messages and context details about the message posters. The raw corpus is annotated by two annotators accordingly to the speech acts described in (Merceron 2014). Before the annotation task, the data is cleaned up and pre-processed. The annotated data set will be used to train a model for automatic speech act classification. The first classification method is an adaptation of the approach in (Kim, Li, and Kim 2010), which is applied on the same domain as ours (messages in university course forums), and yielded significantly better results than comparable approaches. A support vector machine (SVM) using several lexical and contextual features is trained to build a model. One feature used by them are cue phrases, which are phrases marked by the annotators, because they are relevant to the particular speech act of the message. However, marking cue phrases manually is an expensive and time-consuming task as every part of a message, which could be considered relevant to the speech act, has to be annotated manually. For practical reasons and for simplicity, annotators assigned a speech act category to the sentences they considered most relevant for the message's speech act. Consequently, n-grams of common but not relevant phrases are excluded. An example would be the obligatory addressing of fellow students and instructors at the beginning of a message. Further common natural language processing (NLP) tasks for text classification such as stemming and stop word removal were applied on the data. This work is part of a German research project that aims at recommending learning objects by analyzing the students learning behavior with the help of an interactive learning companion application (Krauss 2016).

The evaluation will firstly show how suitable the classifier is for the German language with the effect of the addition of a new category, and secondly how the classifier performs with the addition of two new features. The speech act of the next message and n-grams of part-of-speech-tag (POS) stem pairs are added to the feature set of a second classifier. The speech acts are categorized in the six not mutually exclusive categories: question, answer, issue, two acknowledgment classes and reference. The categories are borrowed by (Kim, Li, and Kim 2010) and (Merceron 2014), that stem from their experience on the analysis of online education platforms. For this reason, a binary SVM classifier for each speech act category is applied on the corpus, where every category is trained against all others. Support vector machines have proven to perform substantially better and robust in comparison to other text classification methods. Furthermore, the discussed results will give an insight which features are the most relevant for the speech act classification and what other features could be considered to gain better results.

The remainder of this paper is organized as follows. The next section discusses related works on speech act classi-

fication and the used features by other researchers. In the third section the corpus data, its source, the pre-processing, data cleansing, and annotation task are described. The forth section describes the features used for the SVM. In the fifth section, the speech act classifiers are evaluated by comparing the classifiers and discussion the results. The last section gives a conclusion and some suggestions for future works.

Related Work

Different approaches are devoted to the classification of messages into domain dependent categories, focusing on different features. Here some of the features and categories used by others to classify messages are presented. The here introduced features are not bound to a domain and can be used in every arbitrary context. Furthermore, the approaches are applied on data sets, which feature similar structures and properties as our data.

Contextual features are widely used to improve the performance of the classification. Sequential information such as the speech acts of the previous utterances are used in (Choi, Cho, and Seo 1999), (Kim, Li, and Kim 2010), and (Samei et al. 2014). Further (Carvalho and Cohen 2005) also includes the speech act of the subsequent utterance. Additionally, (Kim, Li, and Kim 2010) and (Samei et al. 2014) consider the position of the current message in a discussion thread, and the speaker of the previous message and a change of speaker from the previous to the current message. (Samei et al. 2014) and (Carvalho and Cohen 2005) conclude that even though not all categories benefit from the addition of contextual features, the performance of some categories is improved by the addition.

The addition of n-grams to the feature set improves the speech act classification as shown by (Carvalho and Cohen 2006) and (Qadir and Riloff 2011). (Carvalho and Cohen 2006) generated a new feature set by extracting all n-grams of a length 1 to 5 after pre-processing the messages, which led to a drop of 26.4% in error rate and improvements of more than 30% for some speech acts compared to their previous approach. According to them, the ranking of n-grams based on Information Gain score revealed an impressive agreement with the linguistic intuition behind the email speech acts. (Ravi and Kim 2007) use n-grams to classify messages in online discussions into answers and questions. The classifiers have an accuracy of 88% and 73% for question and answer respectively, when using uni-, bi-, tri- and quadro-grams. In a subsequent work (Kim, Li, and Kim 2010) reduced the included n-grams to uni-, bi- and trigrams.

Due to data sparseness (Novielli and Strapparava 2009) considered lemmata in the format lemma#POS instead of tokens to reduce sparsness in their approach of an unsupervised method to label dialogues with proper speech acts. Consequently, a step in the pre-processing task was using a part-of-speech tagger to add the POS features and a morphological analyzer to determine the lemmata. Their results for the comparison of supervised and unsupervised methods are both quite promising. (Moldovan, Rus, and Graesser 2011) include the manually annotated part-of-speech tags of the first 2 to 6 words of a chat message as features. (Jeong, Lin, and Lee 2009) present a semi-supervised method for automatic speech act recognition in email and forums. They include part-of-speech n-grams to the feature set to minimize the number of lexical features. For comparison, they create a supervised classifier, which achieved an F-score of 0.84. (Král and Cerisara 2014) include part-of-speech acts in their approach for dialog act recognition in Czech. (Qadir and Riloff 2011) made the same decision to add part-of-speech tags to their feature set.

The approach by (Kim, Li, and Kim 2010) extends the number of categories and features used in their previous work. Cue phrases and their positions are manually annotated and included in the feature set. Cue phrases are phrases seen as the relevant part of a message for a class. The position of the cues are included, because a cue in the beginning sentences and the end sentences can indicate different speech acts. Instead of extracting n-grams of all sentences in a message, only the n-grams from each annotated cue phrase are extracted. Webb and Ferguson investigate in (Webb and Ferguson 2010) the usage of cue phrases in dialogue act classification and present a method to extract cue phrases automatically. They demonstrate that the usage of the automatically discovered cue phrases are sufficiently useful features for the text classification task.

(Kim, Li, and Kim 2010) employed their own speech act categories based on their experience of student interactions in online discussions. The challenge of classifying messages with incoherent and noisy data remains. On the contrary, (Qadir and Riloff 2011) relies on Searle's original taxonomy for speech acts and do not create domain-specific categories. (Rus et al. 2012) propose a data-driven approach that infers the intrinsic speech act categories from the data based on similarities of the dialogue utterances according to some model. Most of the researchers are either using the DAMSL tagset by (Core and Allen 1997) or introducing their own categories for speech act recognition. This approach uses the speech act categories introduced by (Kim, Li, and Kim 2010) and extended by one category by (Merceron 2014), as our data is in the same domain as theirs.

Corpus Data

The raw corpus contains messages posted by students and instructors in the discussion forum of a Java programming course, taught online in a university. The data was obtained by (Merceron 2014), who used it to connect the analysis of speech acts and the analysis of performance of students. In sum, the corpus contains 182 threads and 694 posts by students and instructors posted on ten separate forums from 2010 to 2013. The names of posters are either removed or replaced by placeholders to anonymize their identities. The messages posted by the students are very unstructured and noisy. Grammatical and syntactical rules are violated, punctuation missing or overused, smiley's or other internet jargon used and sentences interrupted by code fragments. Students and instructors post not working code or code solutions, exceptions and URL links to other material. For the raw corpus to be useful for the classification, some pre-processing steps have to be taken.

Student9:	
orig.:	[...] Allerdings wird keines gezeichnet, es erscheint nur das weiße Fenster. Quelltext: @CODE [...] Wie kann dies denn stattdessen lauten? Ich hoffe, Ihr könnt mir helfen.
trans.:	[...] However none is drawn, there is only a white window. Source: @CODE [...] How could it be called instead? I hope you can help me.
Student4:	
orig.:	[...] soweit ich das überblicken kann, musst du noch ”extends Frame” machen. Dann sollte es gehen. [...]
trans.:	[...] as far as I can grasp it, you have to add ”extends Frame”. Then it should work. [...]
Mentor1:	
orig.:	[...] sehr gute Frage. [...] Von Applet nach JFrame (jetzt mehr benutzt als Frame) empfehle ich die folgende Übersetzung: @CODE
trans.:	[...] very good question. [...] from Applet to JFrame (nowadays more used than Frame) I would suggest following translation: @CODE
Student9:	
orig.:	Danke für die Antworten. [...]
trans.:	Thanks for the answers. [...]

Table 1: Excerpt of a linear discussion from the corpus data with the anonymized author names

Table 1 shows an example of such a linear discussion on the forum going from top to bottom. A student (*Student9*) starts a new thread by stating an issue and question. A second student (*Student4*) follows by providing an answer. An instructor (*Mentor1*) gives an, to the previous answer unrelated, answer to the question. The threads ends with an acknowledgment by the original poster to the other posters.

Data Cleansing & Pre-processing

To reduce noise and increase structure of the messages some manual steps were performed without affecting a poster's intention. The messages contain code fragments, file names and extensions, exceptions, console commands, and URLs, which were replaced manually by placeholders. Code fragments were replaced by @CODE placeholder, file names by @FILE, file extension names by @EXTENSION, exceptions by @EXCEPTION, console commands by @CONSOLE and URLs by @URL.

Marking cue phrases manually is an expensive and time consuming-task. Every part of a message, which could be considered relevant to the speech act, has to be annotated manually. For practical reasons and for simplicity, annotators assigned a speech act category to the sentences they considered most relevant for the message's speech act. Thus, whole sentences are regarded as cue phrases, instead of parts of a sentence. A sentence splitter splits the messages into sentences. An excerpt of such sentences and their categories is shown in Table 3. Consequently, n-grams of common but not relevant phrases are excluded. Further common NLP tasks for text classification such as stemming and stop word removal were applied on the data.

Category	#	kappa
ques	233	0.81
ans	258	0.82
iss	85	0.78
pos-ack	124	0.86
neg-ack	23	0.83
ref	94	0.74

Table 2: Number of annotations and their kappa values for each speech act category

Category	Sentence	
ans	*orig.:*	Die @CODE Methode wird in der Klasse @FILE sein.
	trans.:	The @CODE method will be in class @FILE.
ques	*orig.:*	Also sollen sie keine @CODE -Methode haben?
	trans.:	So they shoudln't have any @CODE-method?
iss	*orig.:*	Allerdings wird keines gezeichnet, es erscheint nur das weiße Fenster.
	trans.:	However none is drawn, there is only a white window.
pos-ack	*orig.:*	Danke für die Antworten.
	trans.:	Thanks for the answers.
neg-ack	*orig.:*	Ich bin mir nicht sicher, ob ich es verstanden habe!
	trans.:	I am not sure if I understood it!
ref	*orig.:*	Falls Sie generische Typen (wie im @CODE) vertiefen wollen, hier die Seite des Tutorials: @URL
	trans.:	If you want to immerse in generic types (as in @CODE), here is the site of the tutorial: @URL

Table 3: Excerpt of sentences from the corpus and the corresponding speech act category

Annotation Task

Each post was annotated by two annotators assigning posts and sentences to the following predefined categories:

- **ques:** A question about a problem, including question about previous messages
- **ans:** A simple or complex answer, suggestion or advice to a previous question
- **iss:** Report misunderstanding, unclear concepts or issues in solving problems
- **pos-ack:** An acknowledgement, compliment or support in response to a previous message
- **neg-ack:** A correction, objection or complaint to/on a previous message

- **ref:** A hint or suggestion related to the subject and not answering any previous message

No hints or suggestions what to consider during the annotation were given. The decision to assign a category to a post depended merely on the annotator's interpretation. Table 2 shows the kappa values for each speech act category the number of posts marked as the corresponding speech act. The numbers exceed the number of posts as one post can have multiple speech acts. The lower kappa value for *reference* is attributed to the fact, that some answers contain references to other solutions and the annotators could not agree if it is part of the answer or not. The same applies for the *issue* category, where annotators were not sure if the post is about a general problem or an issue in solving a problem. According to (Artstein and Poesio 2008), who compiled the interpretations of the kappa value of different researchers, the values for >0.8 are in the range of good reliability. The reliability of the values <0.8 is debatable, but as the values are still over 0.7 and the difference to 0.8 for *issues* is only 0.02, the values are seen as acceptable. The final data set includes only annotations, where both annotators agreed on the speech act. For example when a message is annotated with the speech acts *ques*, *iss* by one annotator and *iss* by the other, then the message would have the speech act *iss* in the final data set. A message does not have any speech act when both annotators disagree on the speech acts.

Features

(Kim, Li, and Kim 2010) used six different types of features for the speech act classifier. These six features are extended by two additional features for further analysis. The first feature are the uni-, bi- and tri-grams of cue phrases and their position within a post. They are added as n-grams of tokens and n-grams of token position pairs to the feature set. N-grams of any sequence length could be included into the feature set, but (Moldovan, Rus, and Graesser 2011) have shown that adding longer sequences as tri-grams decreases the performance. The second feature is the position of the message itself in the thread as a numeric value beginning with 1. The third feature is the speech act of the previous message to that the current message is replying. If the current message was posted by the same author as the previous message is included, additionally to the role of a poster, in this case student or mentor, to the feature set. The sixth feature is the arrangement of messages depending by their length into short(1-5 words), medium(6-30 words), and long($>$30 words) messages. The numerical values 1, 2, and 3 simply represent short, medium, and long messages. The two additional features are uni-, bi- and tri-grams, where the elements are *stem_pos* pairs and the speech act of the next message replying to the current message.

Evaluation

For the evaluation first the approach by (Kim, Li, and Kim 2010) is applied to our data set and secondly, compared to the addition of new features. Due to the small sample size, the chosen validation technique is repeated random subsampling, which splits the data repeatedly into a training set and test set of ratio 3:1. For small sample sizes, repeated random sub-sampling is the preferred method, because the folds in k-fold cross-validation would be too small and therefore highly inaccurate. As the data is split randomly and thus one category could predominate one set, the average of 100 runs is calculated.

SA Category	precision	recall	F-score
ans	0.78	0.85	0.81
ques	0.83	0.78	0.81
iss	0.38	0.02	0.04
pos-ack	0.87	0.56	0.68
neg-ack	0	0	0
ref	0.81	0.21	0.32
Overall	0.81	0.6	0.69

Table 4: Results of the classifier using the feature set of (Kim, Li, and Kim 2010) classifier

SA Cat.	feature	weight
ans	prev_sa_question	1.93
	prev_sa_issue	1.7
	du<>	0.78
ques	?<>	5.12
	jemand<>	1.08
	frag<>	0.87
iss	ich<>	0.86
	probl<>	0.67
	nicht<>	0.53
pos-ack	dank<> & cue-position	1.86
	viel<>dank<>	1.01
	!<>	0.9
neg-ack	gefund<> & cue-position	0.29
	nicht<>	0.18
	den<>fehl<>gefund<> & cue-position	0.18
ref	@url<>	0.73
	poster_role_mentor	0.45
	hier<>	0.45

Table 5: Some of the top features for each speech act category and their weights for the feature set by (Kim, Li, and Kim 2010)

At first the results of the classifier using our feature set, shown in Table 4, are compared to the results of using the feature set of (Kim, Li, and Kim 2010). The number of positive cases for the categories *issue* and *reference* is too low to get any useful performance results for both. Compared to the results of (Kim, Li, and Kim 2010), the results for the categories *question*, *answer*, and *positive-acknowledgment* are not that far off with a drop of 13% and only 4% for *question* and *answer* and even an increase of 14% for *positive-acknowledgment*. Even though 13% seems like a huge drop, the classifier of (Kim, Li, and Kim 2010) does perform exceptionally well for the *question* category. In comparison a score of 0.81 is still a very good result for the classifier. (Kim, Li, and Kim 2010) had the same problem of small number of examples for *negative-acknowledgment*, making it not measurable. Similarly, the number of examples for *is-*

SA Category	POS			Next SA			POS + Next SA		
	prec.	**rec.**	**F**	**prec.**	**rec.**	**F**	**prec.**	**rec.**	**F**
ans	0.78	0.85	0.82	0.8	0.85	0.82	0.81	0.86	0.83
ques	0.86	0.83	0.84	0.85	0.77	0.8	0.88	0.81	0.84
iss	0.48	0.05	0.09	0.67	0.06	0.1	0.66	0.09	0.15
pos-ack	0.84	0.56	0.67	0.83	0.63	0.71	0.81	0.61	0.7
neg-ack	0	0	0	0	0	0	0	0	0
ref	0.82	0.23	0.35	0.84	0.6	0.7	0.86	0.6	0.7
Overall	0.81	0.62	0.70	0.82	0.65	0.73	0.83	0.67	0.74

Table 6: Results for the extended classifier with n-grams of the cue phrases

sue is too low to have meaningful results. The highest performances are reached for the *answer* and *question* categories and the lowest for the *reference* category. However, as the number of positives cases is low for *reference*, it is not predictable if the low performance is due to the classifier or the number of posts. The performance for the new category *reference* is low, but the low recall value indicates that the number of negative cases is too high compared to positive cases. A high precision at least suggests that the low amount of documents classified as *reference* were classified correctly. As the other categories perform relatively well, and there is no linguistic relation between *issue* and *reference*, it can be stated that the addition of a new category does not affect the overall performance of the classifier.

In Table 5 some of the top features and their weights are listed. The weights give a measure on the distance of a message from the class separating hyperplane in the feature space. The <> separates the tokens of the n-gram and indicates if it is an uni-, bi-, or tri-gram. All categories with a good performance have at least two features with a weight >1. One weight of the feature for *question* even being >5. The features even match with the linguistic intention behind the categories. Specifically the top features for *question* being the uni-gram ?<>, for *answer* the previous messages having speech acts of *question* or *issue*, and the poster thanking for *positive acknowledgment*.

The addition of the stem-POS n-grams to the feature set had only a marginal effect on the performance as shown in Table 6. The overall performance was only increased by 1%, where the biggest improvement for a category is 3%. Contrary, did the addition of the next speech act feature without the POS-tag n-grams increase the performance of the classifier significantly for the *reference* category, nearly doubling the F-score. The overall performance is increased by 4%. Other categories are only slightly affected, with the performance of *positive acknowledgment* increasing by 3%. The combination of the next speech act feature and the stem-POS pair n-grams feature did not have any effect on the overall performance of the classifier. The *reference* category has not improved any further by the combination, but the performance of *answer* and *question* categories increased by 1% and 4% respectively.

N-gram feature are predominate in the feature set, as shown in Table 7. Only the features of the *answer* category contain primary non n-gram features. Other features like the post position do not play any significant role for the classification. Beforehand, when passing through the annotated corpus, the assumption was that *reference* features like the poster role being a mentor and URLs in the post, will weigh much more. The top features stayed in all approaches the same, only the weights in the weight vector changed. The best performance was reached by extending the classifier by the next speech act and stem-POS n-gram features.

SA Cat.	feature	weight
ans	prev_sa_question	1.7
	prev_sa_issue	1.4
	du#PPER<>	0.71
	next_sa_answer	0.7
ques	?#$<>	4.24
	jemand#PIS<>	0.85
	next_sa_answer	0.68
	frag#NN<>	0.54
iss	ich#PPER<>	0.88
	probl#NN<>	0.74
	nicht#PTKNEG<>	0.49
pos-ack	dank#NN<> &	
	cue_position	1.41
	!#$<>	1.01
	fur#APPR<>	0.84
neg-ack	next_sa_neg-ack	0.29
	gefund#VVPP<> &	
	cue_position	0.26
	nicht#PTKNEG<>	0.24
ref	poster_role_mentor	0.48
	next_sa_reference	0.47
	hier#ADV<>	0.38
	@url#NN<>	0.34

Table 7: Some of the top features for each speech act category and their weights for the classifier with POS and next speech act features

Conclusion

In this paper the speech act classifier proposed by (Kim, Li, and Kim 2010) was implemented and evaluated for messages in German discussion forums. For this purpose, an annotated corpus for the training of the classifier was created. The reached kappa values indicate high inter-annotator agreement. Cue phrases were annotated manually, as they improve the performance of the classifier. However, the task

of manually marking cue phrases is expensive and time costly. For future works, the cue phrases could be marked automatically using the approach presented in (Webb and Ferguson 2010). The speech act classifier using the feature set of (Kim, Li, and Kim 2010) reached an overall accuracy of 0.69, with the classifier for the *answer* and *question* categories both reaching an accuracy of 0.81. The addition of two new features increased the overall performance slightly with an overall accuracy of 0.74. Especially the accuracy for the *reference* category could be significantly improved. Nevertheless, it has to be considered that the number of examples in the corpus is low. Therefore, the extended classifier should be applied on a bigger corpus to get results that are more meaningful. The approach of (Kim, Li, and Kim 2010) is applicable on German messages and the number of categories adaptable to the used data.

Acknowledgments

The author would like to especially thank Manfred Hauswirth for his supervision, as well as Kim Mensing for her assistance. This work is sponsored by the German Federal Ministry of Education and Research.

References

Artstein, R., and Poesio, M. 2008. Inter-coder agreement for computational linguistics. *Comput. Linguist.* 34(4):555–596.

Carvalho, V. R., and Cohen, W. W. 2005. On the collective classification of email "speech acts". In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, 345–352. New York, NY, USA: ACM.

Carvalho, V. R., and Cohen, W. W. 2006. Improving "email speech acts" analysis via n-gram selection. In *Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech*, ACTS '09, 35–41. Stroudsburg, PA, USA: Association for Computational Linguistics.

Choi, W. S.; Cho, J.-M.; and Seo, J. 1999. Analysis system of speech acts and discourse structures using maximum entropy model. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, 230–237. Stroudsburg, PA, USA: Association for Computational Linguistics.

Core, M. G., and Allen, J. F. 1997. Coding dialogs with the damsl annotation scheme.

Jeong, M.; Lin, C.-Y.; and Lee, G. G. 2009. Semi-supervised speech act recognition in emails and forums. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, 1250–1259. Stroudsburg, PA, USA: Association for Computational Linguistics.

Kim, J.; Li, J.; and Kim, T. 2010. Towards identifying unresolved discussions in student online forums. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, IUNLPBEA '10, 84–91. Stroudsburg, PA, USA: Association for Computational Linguistics.

Král, P., and Cerisara, C. 2014. Automatic dialogue act recognition with syntactic features. *Lang. Resour. Eval.* 48(3):419–441.

Krauss, C. 2016. Smart learning: Time-dependent context-aware learning object recommendations. In *Proceedings of The 29th AAAI International Florida AI Research Society Conference*, FLAIRS-29. Key Largo, USA: AAAI.

Merceron, A. 2014. Connecting analysis of speech acts and performance analysis - an initial study. In *Proceedings of the Workshops at the LAK 2014 Conference co-located with 4th International Conference on Learning Analytics and Knowledge (LAK 2014), Indianapolis, Indiana, USA, March 24-28, 2014.*

Moldovan, C.; Rus, V.; and Graesser, A. C. 2011. Automated speech act classification for online chat. In Visa, S.; Inoue, A.; and Ralescu, A. L., eds., *MAICS*, volume 710 of *CEUR Workshop Proceedings*, 23–29. CEUR-WS.org.

Novielli, N., and Strapparava, C. 2009. Towards unsupervised recognition of dialogue acts. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium*, SRWS '09, 84–89. Stroudsburg, PA, USA: Association for Computational Linguistics.

Qadir, A., and Riloff, E. 2011. Classifying sentences as speech acts in message board posts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, 748–758. Stroudsburg, PA, USA: Association for Computational Linguistics.

Ravi, S., and Kim, J. 2007. Profiling student interactions in threaded discussions with speech act classifiers. In *Proceedings of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, 357–364. Amsterdam, The Netherlands, The Netherlands: IOS Press.

Rus, V.; Graesser, A. C.; Moldovan, C.; and Niraula, N. B. 2012. Automatic discovery of speech act categories in educational games. In *Proceedings of the 5th International Conference on Educational Data Mining, Chania, Greece, June 19-21, 2012*, 25–32.

Samei, B.; Li, H.; Keshtkar, F.; Rus, V.; and Graesser, A. 2014. Context-based speech act classification in intelligent tutoring systems. In Trausan-Matu, S.; Boyer, K.; Crosby, M.; and Panourgia, K., eds., *Intelligent Tutoring Systems*, volume 8474 of *Lecture Notes in Computer Science*. Springer International Publishing. 236–241.

Searle, J. R. 1969. Speech act theory. *Cambridge: Cambridge UP*.

Webb, N., and Ferguson, M. 2010. Automatic extraction of cue phrases for cross-corpus dialogue act classification. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, 1310–1317. Stroudsburg, PA, USA: Association for Computational Linguistics.