Dialogue Acts and Emotions in Multimodal Dyadic Conversations

Costanza Navarretta
Centre for Language Technology
Department of Nordic Studies and Linguistics
University of Copenhagen
Denmark
Email: costanza@hum.ku.dk

***Abstract*—This paper addresses the relation between dialogue acts and emotions in a Danish multimodal annotated corpus of first encounters. Dialogue acts are semantic generalizations of the communicative functions of speech and gestures. Certain emotion types have been found to be strongly related to feedback in previous studies, and therefore we wanted to investigate the relation between emotions and dialogue acts in this corpus further. Our analysis of the most frequently occurring dialogue acts and the co-occurring emotions in the corpus confirms that there is a strong relation between some dialogue act types and specific emotion types and the relation is not only limited to the feedback function. Moreover, the study confirms previous work indicating that the emotions expressed in dialogues are also strictly related to the communicative setting. Two speech segment representations and the co-occurring emotion labels are used as features in machine learning experiments in which various classifiers were trained to identify the 15 most frequent dialogue acts in the data. The results of the experiments show that using the two speech segment representations as training data gives state-of-the art results for dialogue act classification relying on speech segments information. Adding information about emotions improves classification when the classifiers are Logistic Regression and Multilayer Perceptron.**

I. Introduction

This paper investigates the automatic classification of dialogue acts using natural language processing methods applied to the transcriptions of Danish dyadic conversations and information about the emotions expressed by the speakers.

While communicating and exchanging information, humans show their emotions and attitudinal states [1] consciously or unconsciously via their voice, body behavior and especially facial expressions. The identification of the emotions of the participants in conversations does not only contribute to understand how speech should be interpreted in e.g. cases in which the emotions indicate that the speaker is making a joke or is ironic and therefore the content should not been understood literally, but they also show what is the attitude of the conversation participants towards the communicative setting, the content of the conversation, or the other conversation participants. Knowing the attitudinal state of a person and reacting to it in an appropriate way is extremely important in every day life, but it is also essential in advanced HCI and HRI systems, which must react to humans in an empathic way. Attitudes are part of the cognitive state of people and thus should be integrated in infocommunicative systems as those proposed in e.g. [2], [3], [4].

Dialogue acts are segments of speech, which can be considered as a communicative unit. They are an operationalized version, adapted to dialogues, of the Speech Act Theory proposed by [5] and then further developed by [6]. According to the theory, words do not only transmit information, but they can also carry out actions. Usually, a dialogue act corresponds to an utterance, including syntactic clauses, but it can also correspond to one or more words that have a communicative function. This is for example the case for English words such as *yeah*, *okay*, and *all right*, which often have a backchannelling function signaling to the speaker that the addressee is following what has been said. Dialogue acts also describe speech related phenomena, such as self corrections and unfinished utterances.

The past decades, a number of dialogue act taxonomies have been constructed for modeling and implementing the structure of dialogues in dialogue systems. These taxonomies reflect the specific domains addressed by the various projects and systems, such as the HCRC MapTask [7], the Verbmobil project [8], and the AMI project [9]. Even though dialogue acts have traditionally been used in spoken systems, they also apply to gestures, that is to body behavior in general, comprising e.g. facial expressions, head movements and hand gestures. One example is that of head nods and shakes having a backchannelling function. These gestures can occur alone or together with spoken backchannelling. Another example is that of iconic hand gestures which provide information about proprieties of objects, being these concrete objects or events. As other communicative gestures they can occur alone or they can co-occur with speech. Consequently, dialogue acts taxonomies are a central component in both unimodal and multimodal dialogue systems.

Since emotions are expressed by people continuously while they are talking, in this paper we aim to see whether there is a relation between emotions and the semantics expressed by dialogue acts and test whether the annotations of the participants in Danish dyadic first encounters influence the classification of dialogue acts.

The paper contains the following parts. A short discussion of background studies and related work is in section II and a description of the used data is in section III. In section IV, the classification experiments are presented and evaluated, while in section V there is a discussion of these results. Finally in section VI, we conclude and present future work.

II. Background and Related Studies

On the basis of the many domain specific dialogue acts classifications developed in various research projects, [10], [11] define a general dialogue act annotation scheme, known as the ISO 24617 dialogue acts standard. The aims behind this standard is to facilitate the interoperability between the different dialogue act annotation schemes. The ISO 24617 standard supports the multi-functionality of speech segments by classifying dialogue acts over six dimensions that are not mutually exclusive. The six dialogue dimensions accounted for by the standard are the following: a) general purpose dialogue acts, such as questions, answers, and information giving, b) interaction structuring dialogue acts, such as exchanging greetings, apologies and thanking, c) feedback-related acts, which include backchannelling and self feedback, d) turn management acts, e.g. turn take and turn keep e) own communication management, that is self corrections and retractions, and f) time management which comprises stalling and pausing. The ISO 24617 annotation scheme has also been applied for re-annotating dialogue corpora, which had been coded previously according to different annotation schemes [12].

[13] describe the annotation of dialogue acts in a multimodal Danish corpus of first encounters following the ISO 24617 standard, and they analyze how the function of feedback is expressed in the dialogue act standard and in the existing multimodal annotations of the corpus. These multimodal annotations followed the MUMIN annotation scheme proposed by [14]. [15] describes an extended version of the dialogue act annotations of the corpus, accounting for pauses and abandoned utterances and analyses the strong relation between speech pauses and dialogue acts annotations. These dialogue act annotations are used in the present study.

Numerous researchers have addressed the automatic annotation of dialogue acts. The most recent approaches train classifiers on the dialogue act annotations of human-human dialogue corpora e.g. [16], [17], [18] and human-agent interactions [19]. Speech features, such as the length of the utterances, lexical knowledge and syntax are used as features in the classification process. [19] report an accuracy of 77.34 on the classification of 13 dialogue acts in English annotated dialogues between humans and a robot. In their experiments, the authors use syntactic and lexical features as well as the larger context of the dialogue acts. This context consists of three turns preceding each dialogue act. The baseline accuracy obtained without including the context was 65.73.

Researchers have proposed different ways of describing emotions. The most common strategies are the categorical approach based on pre-defined emotion labels [1], [20], [14], and the dimensional approach in which emotions are accounted for via their position in unary or multi-dimensional space [21], [22]. Some studies follow a mixed approach [23], [24] combining emotion labels and dimensional values. There is no common agreement on the number and type of emotion labels or dimensions, but most annotation projects have relied on a restricted number of labels and dimensions. A strong relation between emotions and the communicative function of feedback expressed by facial expressions was found in the data we address in this study by [25] and [26].

[27] analyze the relation between emotions and dialogue acts in the DiaCoSk corpus which consists of 46 minutes of dialogues taken from Slovak TV talk shows. The corpus was annotated with 14 dialogue acts and 32 emotions, comprising a neutral emotion. The analysis of the annotations show that questions, information and clarification requests are the dialogue acts which are most often related to emotions. Moreover, statement related dialogue acts are the dialogue acts that are most often co-occurring with emotions in the talk shows. Interest and Surprise are mostly related to interrogative dialogue acts, while Apprehension and Fear are often conveyed through Action-request dialogue acts. Finally, [27] measure the distance between the feature vectors consist of lexical and prosodic parameters of observations and patterns. The observations and the patterns with the minimal distance are marked as the best ones. The results of these experiments were promising, even if the data used was very limited in size.

In the present work, we also analyze the relation between dialogue acts and emotions in another language and in another communicative setting of that in [27], and we apply classifiers to the annotated data to test whether emotions can improve the automatic classification of dialogue acts based on speech features. The dialogue acts and the emotions are annotated with different annotation schemes in the Danish corpus and in the Slovak corpus, but some of the most common label used are similar.

III. The annotated data

The data we use in our study are the annotations of dialogue acts and emotions in twelve Danish dyadic conversations [28], one of the NOMCO comparable first encounters corpora . The first encounters between young males and females were collected in Denmark, Estonia, Finland and Sweden in a Nordic funded project [29]. All corpora were transcribed and multimodally annotated with features describing the shape and communicative functions of gestures according to a common annotation scheme [30] implementing the MUMIN annotation framework [14][1] .

Emotions and attitudinal states expressed by facial expressions were annotated in the Danish encounters using an open list of emotion labels and a simplification of [22]' s three dimensional emotion labels as proposed by [23]. 28 emotion labels were used and the inter-coder agreement results in terms of Cohen's kappa [31] was of 0.61 for the emotion label annotations and ranged between 0.73 and 0.88 for the dimensional values. The final annotations were agreed upon by three coders. In the corpus, 1027 facial expressions (71% of the 1449 facial expressions) expressed an emotion, according to the annotators [26]. Neutral facial expressions were annotated as not having an expression (value=NONE). The corpus was annotated with 37 dialogue act labels, and contains 4517 dialogue acts [15].

In the present work, we extracted all the dialogue act annotations in the Danish first encounters and the speech tokens related to them. Some of the dialogue acts labels only occur few times. For this reason, we sorted out dialogue acts

[1]The MUMIN annotation schemes for the ANVIL and ELAN multimodal annotation systems are distributed under the CLARIN-DK infrastructure (https://repository.clarin.dk/repository/xmlui/), with the permanent identifier https: http://hdl.handle.net/20.500.12115/43.

TABLE I. EXAMPLE OF THE DATA: TWO ROWS

Utterance	Emotion	Dialogue Act
og jeg h'ar været til det 'en gang (and I have been to it once) +_ +_ tidligere i d'ag (earlier today)	selfconfident	Inform
j'a +_ jamen det er f'int yes well this is fine	NULL	AlloFeedbackGive

TABLE II. THE FREQUENCY OF DIALOGUE ACT LABELS

Dialog Act	Occurrences
Inform	1196
AlloFeedbackGive	979
Pausing	702
Retraction	374
Inform-Answer	301
AutoFeedback	163
Confirm	127
Check-Question	114
Propos-Question	105
AlloFeedbackAgree	96
AlloFeedbackElicit	84
Stalling	50
Disconfirm	44
Choice-Question	32

that occurred less then 30 times in the data. Successively, we extracted all the emotion annotations from the corpus and found with a python script the emotions that co-occurred temporally with the dialogue acts. Our final data set consists therefore of speech segments/utterances and their corresponding dialogue act labels as well as the emotions that were shown by the speaker.

Two rows of the data are shown in Table I as an example. In the transcriptions of speech, an hyphen indicates that the following syllable is stressed, while +_ indicates a silent speech pause. The resulting data consisted of 4479 entries classified with 15 dialogues act labels. 4102 of these entries (92%) co-occurred with an emotion. The dialogue acts in these reduced data set and their frequencies are in Table II, while the emotions which co-occurred with the most frequently occurring dialogue acts are in Table III. Table II shows that the most frequent dialogue acts in this corpus are Inform, AlloFeedbackGive (backchannelling) and Pausing. Pausing is a phenomenon common to speech, while the large number of Inform and AlloFeedbackGive is strongly related to the type of dialogues, first encounters, during which the participants exchange information about themselves and continuously provide feedback to the other participant. The two participants meet for the first time and want to give a good impression [15]. Furthermore, the participants in the first encounters are standing and facing each other, and it is therefore natural that they often nod and smile. The most common emotions which co-occur with the dialogue acts segments are Friendly, Amused and Uncertain and they are also expected in the context of first encounters. What it is surprising is that only 7 out of the 27 emotions annotated in the corpus co-occur with the 15

TABLE III. THE EMOTIONS CO-OCCURRING WITH THE DIALOGUE ACTS

Friendly	2295
Amused	559
Uncertain	519
NONE	377
Uneasy	256
Embarrassed	192
SelfConfident	153
Confident	128

most frequent dialogue acts, and therefore we do not expect that they will have a large influence on the classification of these dialogue acts. However, it is interesting that the Friendly and Amused emotions often co-occur with feedback dialogue acts, and especially AlloFeedbackGive (Backchannelling). Moreover, Confident and SelfConfident are often related to the Inform, Inform-Answer and Confirm dialogue acts, while Embarassed, Uncertain and Uneasy co-occur with question-related dialogue acts and AutoFeedback, which labels cases in which the speaker provides feedback to her own speech. In the future, the relation between the less frequently occurring dialogue acts and emotions should be investigated.

It is not possible to compare directly these results with the results presented in [27] since, as noticed in section II, the two corpora are annotated with different labels of emotions and dialogue acts. However, it is clear that also in the Danish corpus, some dialogue acts co-occur more frequently with some emotions, and the emotions identified in the two corpora are related to the type of dialogues. Therefore, there are both similarity and differences between the findings in the two corpora. Similarities are the following: the emotion *Confident* often co-occurs with the dialogue act *Inform* and *Interest* co-occurs with inform-requests acts in both corpora. Dissimilarities related to the type of dialogues are that the emotions *Fear* and *Anger* occur quite often in the Slovak corpus of talk shows in relation to action-requests, while they do not occur at all in the Danish first encounters because the participants only speak about themselves and do not compete with each other in this corpus.

IV. CLASSIFICATION OF DIALOGUE ACTS

Since many emotions were found to be related to the communicative function of feedback facial expressions and head nods in the Danish first encounters [25], we wanted to determine whether they also influence positively the F1 scores of dialogue act classification, even if the analysis of the data indicates that few emotion types often co-occur with the most frequently occurring dialogue acts, *Inform* and *AlloFeedbackGive*.

Before applying classification algorithms to the data, some pre-processing was made. Differing from the most common approaches to dialogue act classification, we do not use syntactic features and features related to the number of words and characters in a speech segment, but we apply transformations to the speech sequences, which are frequently used in information retrieval and in other natural language processing applications. More specifically, we first applied a bag of words

(BOW) transformation to the speech segments, providing a frequency-based representation for each word, and then a Term Frequency-Inverse Document Frequency (TF*IDF) transformation. TF*IDF is a technique usually used for determining the most characteristic words in a document with respect to other documents in a corpus. Finally, we converted the nominal emotion labels to dummy variables taking binary values and concatenated them to the two speech segment representations.

The classifiers were then trained on the following datasets: a) BOW, b) BOW plus emotion variables, c) TF*IDF and d) TF*IDF plus emotion variables. The scikit-learn python3 package was used for the classification experiments, in which we trained and tested the following four classifiers: support vector machine, multinomial Naive Bayes, logistic regression and multilayer perceptron. A majority and a random classifier, which takes into account the frequency of occurrences of the various classes, were used as baselines. The data was divided in three parts: a training set (60% of the data), a test set (20% of the data) and an evaluation set (20%). The results obtained by the various classifiers on the four data sets are in Table IV. The first column of the Table shows the data set, the second column gives the classifier, while the third, fourth and fifth column indicate the Precision (P), the Recall (R) and the weighted F1-score, respectively. The table shows that all

TABLE IV. RESULTS OF CLASSIFICATION

Data	Algorithm	P	R	F1
	Majority	0.06	0.25	0.1
	Random	0.06	0.25	0.09
BOW	NaiveBayes	0.62	0.66	0.58
BOW+Emo	NaiveBayes	0.56	0.66	0.57
TFIDF	NaiveBaies	0.55	0.65	0.55
TFIDF+Emo	NaiveBayes	0.52	0.65	0.54
BOW	SVM	0.66	0.7	0.65
BOW+Emo	SVM	0.62	0.68	0.62
TFIDF	SVM	0.65	0.69	0.63
TFIDF+Emo	SVM	0.62	0.66	0.59
BOW	Logist	0.64	0.69	0.65
BOW+Emo	Logist	0.62	0.67	0.63
TFIDF	Logist	0.63	0.68	0.63
TFIDF+Emo	Logist	0.64	0.69	0.65
BOW	MLP	0.62	0.65	0.63
BOW+Emo	MLP	0.61	0.64	0.61
TFIDF	MLP	0.64	0.67	0.65
TFIDF+Emo	MLP	0.66	0.68	**0.67**

classifiers perform significantly better than both the majority and random classifier, which nearly return the same values. The best results in terms of F1-score are achieved by the Multilayer Perceptron when the TF*IDF model and emotion information are used. The Multilayer Perceptron on this data set gives an F1-score of 0.67. This must be compared to the 0.1 F1-score of the best baseline. Contrary to our expectation, the emotion annotations only in some cases improve classification and when there is an improvement, this is not large. Moreover, the TF*IDF model performs better than BOW with some classifiers, and worse with others. The Multilayer Perceptron was run with the default parameters, but vi tested both the *adam* and the *sgd* solver and the *tahn* and *relu* activation, and we got the best results (those reported) with the *sgd* solver and the *tahn* activation.

The results obtained training classifiers on Danish speech data are in line with the results obtained on a larger corpus of English speech dialogues when using only features extracted from speech transcriptions without including the three preceding dialogue acts [19]. It must be noted, however, that we used BOW and TF*IDF models while the English study used syntactic and lexical features, as well as word and character length.

The analysis of the confusion matrices for the best performing algorithms shows that the classes that are identified most correctly are the three most frequent ones, that is *Pausing*, *Inform* and *AlloFeedbackGive*. The most frequent error is that *Inform-Answer* is classified as *Inform*. This is not surprising and the error could be solved considering the preceding context as proposed by [19] or joining the two classes. Another frequent misclassification is that of *AutoFeeeback* which is confused with *AlloFeedbackGive*. Also this error is not surprising since the two dialogue acts are often expressed linguistically in the same way.

V. DISCUSSION

In this paper, we have analyzed the occurrences of attitudinal states and dialogue acts in a multimodally annotated Danish corpus of first encounters, which was used in previous studies for analyzing many discourse-related phenomena. For example, emotion annotations were found to be strictly correlated to the communicative feedback function both in speech and gestures [25]. This study is similar to the research in [27] in which dialogue acts and emotions in a Slovak corpus of talk shows were analyzed. Our study confirms that certain dialogue acts are often related to specific emotions as noticed by [27], but they also show that this relation strongly depends on the communicative setting. In fact, even if there are some similarities between the most common emotions co-occurring with *Inform* and question-related dialogue acts, some of the emotions identified in the Slovak talk shows did not occur at all in the Danish corpus of first encounters, even when they co-occurred with corresponding dialogue acts. This was for example the case of the emotion *Fear*, which was common in the Slovak corpus in connection with action-requests, while it did not occur in the Danish corpus. Also some of the dialogue acts annotated in the Slovak televised corpus are not relevant to the spontaneous Danish corpus of first encounters.

The emotions co-occurring with the most frequent dialogue acts in thefirst encounters as well as the dialogue acts identified in them reflect the communicative type of the dialogues and the physical setting in which the encounters took place. The most common emotional reactions are related to the various semantic contexts described by the dialogue acts. For example, the participants were friendly while providing feedback to the speakers, and they showed amusement when the interlocutor or they self made a joke. The participants were also confident and self-confident when they provided information and were insecure when they misunderstood something said by the other participant.

We also found that the most common dialogue acts in the corpus co-occur with a restricted number of frequently

occurring emotions, while the large majority of the remaining emotion labels only co-occur with discourse act labels that appear less than 30 times in the corpus and therefore are not included in this study. In the future, it should be investigated whether there is a strong relation between some of these emotion labels and the more rare dialogue acts.

In the classification experiments aimed to identify correctly dialogue acts from speech segments and emotions, we tested a number of classifiers on two representations of speech utterances that are often used in information retrieval and text mining: bag of words and TF*IDF. These representations of words have also been found to be useful in other natural language processing applications and in image processing. In the latter case, the transformations are applied to visual features instead than to words. We then added emotion labels to these representations and found that classifiers performed as well as state-of-the-art on speech data when no larger contextual information is used. However, the results with respect to the contribution of emotions are not convincing, since in only two cases the emotion features contributed to classification and the improvement is not large. This is probably due to the fact that only few emotion types co-occurred with the most frequent dialogue acts. Moreover, the data set is skewed also with respect to dialogue act types. The manual analysis of the results of the best performing classifier also shows that the best recognized dialogue acts are the three most frequent ones.

VI. Conclusions and Future Work

In this paper, we have presented an analysis of the most frequent dialogue acts, which are a generalization of the content of dialogues, and the co-occurring emotions expressed by the speakers in a Danish corpus of first encounters. The analysis shows that the same emotions often co-occur with the same dialogue acts. This confirms what was found in a study of annotated televised Slovak talk shows by [27]. Even if the Danish and Slovak corpus were annotated according to different annotation frameworks, most labels can be easily compared. The two studies show both similarities and differences between the two corpora which are due to the different communicative settings and dialogue types (first encounters vs. talk shows). This also confirms that the emotions shown in dialogues are often related to the conversation type and the communicative settings as proposed in [25].

In the second part of the paper, we have described classification experiments in which bag of words and TF*IDF representations of the transcriptions of the first encounters were used in order to automatically classify the dialogue acts that occurred at least 30 times in the dialogues. The obtained results are state-of-the-art compared with the results obtained in classification experiments of dialogue acts in English interactions between humans and a robot when speech features were used [19]. This is the case even if the information used in the two studies are different since in the English study lexical and syntactic features were used.

We also tested whether adding information about the emotions showed by the speakers could improve the classification of dialogue acts and we found that this is the case with the best performing classifiers, logistic regression and multilayer perceptron. However, the improvement with the extended data set is not as large as expected. This might be due to the fact that the same emotion types co-occur with the most frequent dialogue acts, which are also those more correctly identified by the classifiers.

A lot of work can be done in the future to improve the automatic classification of dialogue acts. Morphosyntactic features and a syntactic representation can be automatically added to the speech segments. The use of stop-words can be tested and data sets features such as the length of the words, and word2vec representations can be added. Moreover, all multimodal annotations present in the corpus could be used as training features, and we could investigate optimal set ups for the various classifier. Finally, following [19], the preceding context of dialogue acts could be included to improve the performance of classifiers.

References

[1] P. Ekman and W. V. Friesen, "The repertoire of non-verbal behaviour categories, origins usage and coding," *Semiotica*, vol. 1, pp. 49–98, 1969.

[2] P. Baranyi, A. Csapo, and G. Sallai, *Cognitive Infocommunications (CogInfoCom)*. Springer International Publishing, 2015.

[3] P. Baranyi, "Special issue on cognitive infocommunications theory and applications – guest editorial," *Infocommunications Journal*, vol. XII, no. 1, p. 1, March 2020.

[4] C. Vogel and A. Esposito, ""interaction analysis and cognitive infocommunications," *Infocommunications Journal*, vol. XII, no. 1, pp. 2–9, March 2020.

[5] J. Austin, *How To Do Things With Words*. Cambridge, MA: Harvard University Press, 1962.

[6] J. Searle, *Speech Acts*. Cambridge University Press, 1969.

[7] A. Anderson, M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert, "The hcrc map task corpus," *Language and Speech*, vol. 34, pp. 351–366, 1991.

[8] J. Alexandersson, B. Buschbeck-Wolf, T. Fujinami, M. Kipp, S. Koch, E. Maier, N. Reithinger, B. Schmitz, and M. Siegel, "Dialogue acts in verbmobil-2," DFKI, Tech. Rep., 1997.

[9] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything ami meeting corpus." *Language Resources and Evaluation Journal*, vol. 41, no. 2, pp. 181–190, 2007.

[10] H. Bunt, J. Alexandersson, J. Carletta, J.-W. Choe, A. C. Fang, K. Hasida, K. Lee, V. Petukhova, A. Popescu-Belis, L. Romary, C. Soria, and D. Traum, "Towards and iso standard for dialogue act annotation," in *Proceedings 7th international conference on language resources and evaluation (LREC 2010)*, 2010, pp. 2548–2555.

[11] H. Bunt, V. Petukhova, and A. Fang, "Revisiting the ISO standard for dialogue act annotation," in *Proceedings 13th joint ISO-ACL workshop on interoperable semantic annotation (ISA-13)*, Montpellier, France, 2017, pp. 37–50.

[12] H. Bunt, V. Petukhova, A. Malchanau, A. Fang, and K. Wijnhoven, "The DialogBank: dialogues with interoperable annotations," *Language Resources and Evaluation*, vol. 53, no. 2, pp. 213–249, 2019.

[13] C. Navarretta and P. Paggio, "Dialogue act annotation in a multimodal corpus of first encounter dialogues," in *Proceedings of The 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 634–643. [Online]. Available: https://www.aclweb.org/anthology/2020.lrec-1.80

[14] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, "The mumin coding scheme for the annotation of feedback, turn management and sequencing," *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special Issue of the International Journal of Language Resources and Evaluation*, vol. 41, no. 3–4, pp. 273–287, 2007.

[15] C. Navarretta, "Speech pauses and dialogue acts," in *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, IEEE, Ed., 2020, pp. 560–565.

[16] D. Verbree, R. Rienks, and D. Heylen, "Dialogue-act tagging using smart feature selection; results on multiple corpora," in *IEEE Spoken Language Technology Workshop*. IEEE, 2006, p. 4 pages.

[17] D. Milajevs and M. Purver, "Investigating the contribution of distributional semantic information for dialogue act classification," in *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality*. ACL, 2014, pp. 4a–47.

[18] D. Amanova, V. Petukhova, and D. Klakow, "Creating annotated dialogue resources: Cross-domain dialogue act classification," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, C. et al., Ed. Paris, France: European Language Resources Association (ELRA), may 2016.

[19] A. Ahmadvand, J. I. Choi, and E. Agichtein, "Contextual dialogue act classification for open-domain conversational agents," ser. SIGIR'19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3331184.3331375

[20] A. Ortony, G. L. Clore, and A. Collins, *The cognitive structure of emotions*. Cambridge University Press, MA, 1988.

[21] W. Wundt, *Grundzüge der physiologischen Psychologie*. Leipzig: Engelmann, 1905.

[22] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273–294, 1977.

[23] M. Kipp and J.-C. Martin, "Gesture and emotion: Can basic gestural form features discriminate emotions?" in *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII-09)*, 2009. IEEE Press, 2009.

[24] C. Navarretta, "Annotating and analyzing emotions in a corpus of first encounters," in *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications*, IEEE, Ed., Kosice, Slovakia, December 2012, pp. 433–438.

[25] C. Navarretta, "Predicting speech overlaps from speech tokens and co-occurring body behaviours in dyadic conversations," in *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI 2013)*, Sydney, Australia, December 2013, pp. 157–163.

[26] C. Navarretta, "Predicting emotions in facial expressions from the annotations in naturally occurring first encounters," *Knowledge-Based Systems*, vol. 71, pp. 34–40, 2014.

[27] S. Ondas, L. Mackova, and D. Hladek, "Emotion analysis in diacosk dialog corpus," in *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2016, pp. 000 151–000 156.

[28] P. Paggio and C. Navarretta, "The Danish NOMCO Corpus of Multimodal Interaction in First Acquaintance Conversations," *Language Resources and Evaluation*, vol. 51, pp. 463–494, 2017.

[29] P. Paggio, E. Ahlsén, J. Allwood, K. Jokinen, and C. Navarretta, "The NOMCO multimodal Nordic resource - goals and characteristics," in *Proceedings of LREC 2010*, Malta, May 17-23 2010, pp. 2968–2973.

[30] C. Navarretta, E. Ahlsén, J. Allwood, K. Jokinen, and P. Paggio, "Feedback in nordic first-encounters: a comparative study," in *Proceedings of LREC 2012*, Istanbul Turkey, May 2012, pp. 2494–2499.

[31] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.