IJCLA VOL. 1, NO. 1-2, JAN-DEC 2010, PP. 9-26
RECEIVED 23/11/09 ACCEPTED 16/01/10 FINAL 11/03/10

Exploring the Lexical Semantics of Dialogue Acts

NICOLE NOVIELLI[1] AND CARLO STRAPPARAVA[2]

[1] *Università degli Studi di Bari, Italy*
[2] *FBK-irst, Italy*

ABSTRACT

People proceed in their conversations through a series of dialogue acts to yield some specific communicative intention. In this paper, we study the task of automatic labeling dialogues with the proper dialogue acts, relying on empirical methods and simply exploiting lexical semantics of the utterances. In particular, we present some experiments in both a supervised and an unsupervised framework on an English and an Italian corpus of dialogue transcriptions. In the experiments we consider the settings of dealing with or without additional information from the dialogue structure. The evaluation displays good results, regardless of the used language. We conclude the paper exploring the relation between the communicative goal of an utterance and its affective content.

1 INTRODUCTION

When engaged in dialogues, people ask for information, agree with their partner, state some facts and express opinions. They proceed in their conversations through a series of dialogue acts to yield some particular communicative intention.

Dialogue Acts (DA) have been well studied in linguistics [1,2] and attracted computational linguistics research for a long time [3,4]. There is a large number of application domains that can benefit from the automatic extraction of the underlying structure of dialogues: dialogue systems for human-computer interaction, conversational agents for monitoring and supporting human-human conversations forums and chat logs analysis for opinion mining, affective state recognition by mean of dialogue pattern analysis, automatic meeting summarization and so on. This kind of applications requires a deep understanding of the conversational structure and dynamic evolution of the dialogue: at every step of the interaction the system should be able to understand who is telling what to whom. With the advent of the Web, a large amount of material about natural language interactions (e.g. blogs, chats, conversation transcripts) has become available, raising the attractiveness of empirical methods of analyses on this field.

In this paper, we study the task of automatic labeling dialogues with the proper speech acts. We define a method for DA recognition by relying on empirical methods that simply exploit lexical semantics of the sentences. Even if prosody and intonation surely play a role (e.g. [5,6]), nonetheless language and words are what the speaker uses to convey the communicative message and are just what we have at disposal when we consider texts found on the Web.

We present some experiments in a supervised and unsupervised framework on both an English and an Italian corpus of dialogue transcriptions. In particular we consider the classification of dialogue acts with and without taking into account dialogue contextual features. We achieved good results in all settings, independently from the used language. Finally, we explore the relation between the communicative goal of an utterance and its affective content, using a technique [7] for checking the emotional load in a text.

The paper is organized as follows. Section 2 gives a brief sketch of the NLP background on Dialogue Act recognition. In Section 3 we introduce the English and Italian corpora of dialogues, their characteristics, DA labeling and preprocessing. Then, Section 4 explains the supervised and unsupervised settings, showing the experimental results obtained on the two corpora and providing detailed results and error analysis. In Section 5 we presents the results considering also dialogue contextual features. Section 6 describes the preliminary results of a qualitative study about the relation between the dialogue acts and their affective load. Finally, in Section 7 we conclude the paper with a brief discussion and some directions for future work.

2 BACKGROUND

A DA can be identified with the communicative goal of a given utterance [1]. Researchers use different labels and definitions to address the com-

Table 1. An excerpt from the Switchboard corpus

Speaker	**Dialogue Act**	**Utterance**
A	OPENING	*Hello Ann.*
B	OPENING	Hello Chuck.
A	STATEMENT	*Uh, the other day, I attended a conference here at Utah State University on recycling*
A	STATEMENT	*and, uh, I was kind of interested to hear cause they had some people from the EPA and lots of different places, and, uh, there is going to be a real problem on solid waste.*
B	OPINION	Uh, I didn't think that was a new revelation.
A	AGREE /ACCEPT	*Well, it's not too new.*
B	INFO-REQUEST	So what is the EPA recommending now?

municative goal of a sentence: Searle [2] talks about *speech act*; Schegloff [8] and Sacks [9] refer to the concept of *adjacency pair part*; Power [10] adopts the definition of *game move*; Cohen and Levesque [11] focus more on the role speech acts play in interagent communication.

Traditionally, the NLP community has employed DA definitions with the drawback of being domain or application oriented. In the recent years some efforts have been made towards unifying the DA annotation [4]. In the present study we refer to a domain-independent framework for DA annotation, the DAMSL architecture (Dialogue Act Markup in Several Layers) by Core and Allen [3].

Recently, the problem of DA recognition has been addressed with promising results. Stolcke et al. [5] achieve an accuracy of around 70% and 65% respectively on transcribed and recognized words by combining a discourse grammar, formalized in terms of Hidden Markov Models, with evidences about lexicon and prosody. Reithinger and Klesen's approach [12] employs a bayesian approach achieving 74.7% of correctly classified labels. A partially supervised framework by Venkataraman et al. [13] has also been explored, using five broad classes of DA and obtaining an accuracy of about 79%. Regardless of the model they use (discourse grammars, models based on word sequences or on the acoustic features or a combination of all these) the mentioned studies are developed in a supervised framework. Rather than improving the performance of supervised frameworks, our main goal is to explore the use of an unsupervised methodology.

3 DATA SETS

Table 2. The set of labels employed for Dialogue Act

Label	**Description and Examples**	**Italian**	**English**
INFO-REQUEST	Utterances that are pragmatically, semantically, and syntactically questions - *'What did you do when your kids were growing up?'*	34%	7%
STATEMENT	Descriptive, narrative, personal statements - *'I usually eat a lot of fruit'*	37%	57%
S-OPINION	Directed opinion statements - *'I think he deserves it.'*	6%	20%
AGREE-ACCEPT	Acceptance of a proposal, plan or opinion - *'That's right'*	5%	9%
REJECT	Disagreement with a proposal, plan, or opinion - *'I'm sorry no'*	7%	.3%
OPENING	Dialogue opening or self-introduction - *'Hello, my name is Imma'*	2%	.2%
CLOSING	Dialogue closing (e.g. farewell and wishes) - *'It's been nice talking to you.'*	2%	2%
KIND-ATT	Kind attitude (e.g. thanking and apology) - *'Thank you.'*	9%	.1%
GEN-ANS	Generic answers to an Info-Request - *'Yes'*, *'No'*, *'I don't know'*	4%	4%
total cases		1448	131,265

In the experiments described in this paper we exploit two corpora, both annotated with Dialogue Acts labels. We aim at developing a recognition methodology as much general as possible, so we selected corpora that differ in the content and in the used language: the Switchboard corpus [14] of English telephone conversations about general interest topics, and an Italian corpus of dialogues in the healthy-eating domain [15].

The Switchboard corpus is a collection of transcripts of English human-human telephone conversations [14] involving couples of randomly selected strangers: they were asked to select one general interest topic and to talk informally about it. Full transcripts of these dialogues are distributed by the Linguistic Data Consortium. A part of this corpus is annotated [16] with DA labels (overall 1155 conversations, for a total of 205,000 utterances and 1.4 million words) [3]. Tables 1 shows a short sample fragment of dialogue from this corpus.

The Italian corpus had been collected in the scope of some previous research about Human-ECA (Embodied Conversational Agent) interaction: to collect these data a Wizard of Oz tool was employed [15] in which the application domain and the ECA's appearance may be settled at the beginning of simulation. During the interaction, the ECA played the role of an artificial therapist and the users were free to interact with it in natural language, without any particular constraint. This corpus is about healthy eating and contains overall 60 dialogues, 1448 users' utterances and 15,500 words.

Labelling. The two corpora are annotated in order to capture the communicative intention of each dialogue move. Defining a DA markup language is out of the scope of the present study, hence we employed the original annotation of the two corpora [17,16], which is consistent, in both cases, with the Dialogue Act Markup in Several Layers (DAMSL) scheme [3]. In particular the Switchboard corpus employs the SWBD-DAMSL revision [16].[4]

Table 2 shows the set of labels employed for the purpose of this study, with definitions and examples: it maintains the DAMSL main characteristic of being domain-independent and it is also consistent with the original semantics of the SWBD-DAMSL markup language employed in the Switchboard annotation. As shown in Table 3, the SWBD-DAMSL had been automatically converted into the categories included in our markup language. Also we did not consider the utterances formed only by nonverbal material (e.g. laughter). The DA label distribution and the total number of cases (utterances) considered in the two data sets are reported in Table 2.

[3] `ftp.ldc.upenn.edu/pub/ldc/public_data/swb1_dialogact_annot.tar.gz`

[4] The SWBD-DAMSL modifies the original DAMSL framework by further specifying some categories or by adding extra (mainly prosodic) features, which were not originally included in the scheme.

Table 3. The Dialogue Act set of labels with their mapping with the SWBD-DAMSL correspondent categories

Label	**SWBD-DAMSL**
INFO-REQ	*Yes-No question* (*qy*), *Wh-Question* (qw), *Declarative Yes-No-Question* (qyˆd), *Declarative Wh-Question* (qwˆd), *Alternative ('or') question* (qr) *and OR-clause* (qrr) , *Open-Question* (qo), *Declarative* (ˆd) *and Tag questions* (ˆg)
STATEMENT	*Statement-non-opinion* (sd)
S-OPINION	*Statement-opinion* (sv)
AGREE-ACC	*Agreement /accept* (aa)
REJECT	*Agreeement /reject* (ar)
OPENING	*Conventional-opening* (fp)
CLOSING	*Conventional-closing* (fc)
KIND-ATT	*Thanking* (ft) *and Apology* (fa)
GEN-ANS	*Yes answers* (ny), *No answers* (nn), *Affirmative non-yes answers* (na) *Negative non-no answers* (ng)

Data preprocessing. To reduce the data sparseness, we used a POS-tagger and morphological analyzer [18] for preprocessing the corpora and we used lemmata instead of tokens. No feature selection was performed, keeping also stopwords. In addition, we augment the features of each sentence with a set of linguistic markers, defined according to the semantics of the DA categories. We hypothesize, in fact, these features could play an important role in defining the linguistic profile of each DA. The addition of these markers is performed automatically, by just exploiting the output of the POS-tagger and of the morphological analyzer, according to the following rules:

– **WH-QTN**, used whenever an interrogative determiner is found, according to the output of the POS-tagger (e.g. 'when' does not play an interrogative role when tagged as conjunction);
– **ASK-IF**, used whenever an utterance presents some cues of the pattern 'Yes/No' question. ASK-IF and WH-QTN markers are supposed to be relevant for the recognition of the INFO-REQUEST category;
– **I-PERS**, used for all declarative utterance whenever a verb is in the first person form, singular or plural (relevant for the STATEMENT);
– **COND**, used when a conditional form is detected.
– **SUPER**, used for superlative adjectives;

- **AGR-EX**, used whenever an agreement expression (e.g. 'You are right', 'I agree') is detected (relevant for AGREE-ACCEPT);
- **NAME**, used whenever a proper name follows a self-introduction expression (e.g. 'My name is') (relevant for the OPENING);
- **OR-CLAUSE**, used when the utterance is an or-clause, i.e. it starts with the conjunction 'or' (should be helpful for the characterization of the INFO-REQUEST);
- **VB**, used only for the Italian, it is when a dialectal form of agreement is detected.

4 Minimally Supervised Dialogue Act Recognition

It is not always easy to have large training material at disposal, partly because of manual labeling effort and moreover because often it is not possible to find it. Schematically, our unsupervised methodology consists of the following steps: (i) building a semantic similarity space in which words, set of words, text fragments can be represented homogeneously, (ii) finding seeds that properly represent dialogue acts and considering their representations in the similarity space, and (iii) checking the similarity of the utterances.

To get a similarity space with the required characteristics, we used Latent Semantic Analysis (LSA). LSA is a corpus-based measure of semantic similarity proposed by Landauer [19]. In LSA, term co-occurrences in a corpus are captured by means of a dimensionality reduction operated by a singular value decomposition (SVD) on the term-by-document matrix $\mathbf{T}$ representing the corpus.

SVD is a well-known operation in linear algebra, which can be applied to any rectangular matrix in order to find correlations among its rows and columns. In our case, SVD decomposes the term-by-document matrix $\mathbf{T}$ into three matrices $\mathbf{T} = \mathbf{U}\mathbf{\Sigma_k}\mathbf{V}^T$ where $\mathbf{\Sigma_k}$ is the diagonal $k \times k$ matrix containing the k singular values of $\mathbf{T}$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$, and $\mathbf{U}$ and $\mathbf{V}$ are column-orthogonal matrices. When the three matrices are multiplied together the original term-by-document matrix is recomposed. Typically we can choose $k' \ll k$ obtaining the approximation $\mathbf{T} \simeq \mathbf{U}\mathbf{\Sigma_{k'}}\mathbf{V}^T$.

LSA can be viewed as a way to overcome some of the drawbacks of the standard vector space model (sparseness and high dimensionality). In fact, the LSA similarity is computed in a lower dimensional space, in which second-order relations among terms and texts are exploited. The similarity in the resulting vector space is then measured with the standard cosine similarity. Note also that LSA yields a vector space model that allows for a *homogeneous* representation (and hence comparison) of words, sentences, and texts. For representing a word set or a sentence in the LSA space we use the *pseudo-document* representation technique, as described by Berry [20]. In practice, each text segment is represented in the LSA space by summing up the normalized LSA vectors of all the constituent words, using also a *tf.idf* weighting scheme [21].

Table 4. The complete sets of seeds for the unsupervised experiment

Label	**Seeds**
INFO-REQ	WH-QTN, ‘?’, ASK-IF
STATEMENT	I-PERS, I
S-OPINION	Verbs which directly express opinion or evaluation (guess, think, suppose)
AGREE-ACC	AGR-EX, yep, yeah, absolutely, correct
REJECT	Verbs which directly express disagreement (disagree, refute)
OPENING	Expressions of greetings (hi, hello), words and markers related to self-introduction formula (name, NAME)
CLOSING	Interjections/exclamations ending discourse (alright, okey, ‘!’), Expressions of thanking (thank) and farewell (bye, bye-bye, goodnight)
KIND-ATT	Lexicon which directly expresses wishes (wish), apologies (apologize), thanking (thank) and sorry-for (sorry, excuse)
GEN-ANS	no, yes, uh-huh, nope

The methodology is unsupervised[5] as we do not exploit any ‘labeled’ training material. For the experiments reported in this paper, we run the SVD using 400 dimensions (i.e. k') respectively on the English and Italian corpus, without any DA label information. Starting from a set of seeds (words) representing the communicative acts, we build the corresponding vectors in the LSA space and then we compare the utterances to find the communicative act with the highest similarity.

[5] Or minimally supervised, since providing hand-specified seeds can be regarded as a minimal sort of supervision.

Table 5. Evaluation of the supervised and unsupervised methods on the two corpora

	Italian						English					
	SVM			LSA			SVM			LSA		
Label	prec	rec	F1	prec	rec	F1	prec	rec	F1	prec	rec	F1
INFO-REQ	.92	.99	.95	.96	.88	.92	.92	.84	.88	.93	.70	.80
STATEMENT	.85	.68	.69	.76	.66	.71	.79	.92	.85	.70	.95	.81
S-OPINION	.28	.42	.33	.24	.42	.30	.66	.44	.53	.41	.07	.12
AGREE-ACC	.50	.80	.62	.56	.50	.53	.69	.74	.71	.68	.63	.65
REJECT	-	-	-	.09	.25	.13	-	-	-	.01	.01	.01
OPENING	.60	1.00	.75	.55	1.00	.71	.96	.55	.70	.20	.43	.27
CLOSING	.67	.40	.50	.25	.40	.31	.83	.59	.69	.76	.34	.47
KIND-ATT	.82	.53	.64	.43	.18	.25	.85	.34	.49	.09	.47	.15
GEN-ANS	.20	.63	.30	.27	.38	.32	.56	.25	.35	.54	.33	.41
micro	.71	.71	.71	.66	.66	.66	.77	.77	.77	.68	.68	.68

Table 4 shows the complete sets of seeds used for building the vector of each DA. We defined seeds by only considering the communicative goal and the specific semantics of every single DA, just avoiding the overlapping between seed groups as much as possible. We wanted to design an approach which is as general as possible, so we did not consider domain words that would have made easier the classification in the specific corpora. The seeds are the same for both languages, which is coherent with our goal of defining a language-independent method. There are only a few exceptions: in Italian it is not necessary to specify the pronoun when formulating a sentence so we did not include the 'I' equivalent pronoun in the seeds for the STATEMENT label; the VB linguistic marker is used only for the Italian and is included in the seeds for the S-OPINION vector.

An upper-bound performance is provided by running experiment in a supervised framework. We used Support Vector Machines [22], in particular SVM-light package [23] under its standard configuration. We randomly split the two corpora into 80/20 training/test partitions. SVMs have been used in a large range of problems, including text classification, image recognition tasks, bioinformatics and medical applications, and they are regarded as the state-of-the-art in supervised learning. To allow comparison, the performance is measured on the same test set partition for both the unsupervised and supervised experiments.

4.1 *Experimental Results and Discussion*

We evaluated the performance of our method in terms of precision, recall and F1-measure (see Table 5) according to the DA labels given by annotators in the datasets. As baselines we can consider (i) most-frequent label assignment (respectively 37% for Italian, 57% for English) for the supervised setting, and (ii) random DA selection (11%) for the unsupervised one.

We got .71 and .77 of F1 respectively for the Italian and the English corpus in the supervised condition, and .66 and .68 respecitvely in the unsupervised one. The performance is quite satisfying and is comparable to the state of the art in the domain. In particular, the unsupervised technique is significantly above the baseline, for both the Italian and the English corpus experiments. We note that the methodology is independent from the language and the domain: the Italian corpus is a collection of dialogue about a very restricted domain (advice-giving dialogue about healthy-eating) while in the Switchboard corpus the conversations revolve around general topics chosen by the two interlocutors. Moreover, in the unsupervised setting we use the same seed definitions. Secondly, it is independent on the differences in the linguistic style due to the specific interaction scenario and input modality. Finally, the performance is not affected by the difference in size of the two data sets.

Error analysis. After conducting an error analysis, we noted that many utterances are misclassified as STATEMENT. One possible reason is that statements usually are quite long and there is a high chance that some linguistic markers that characterize other dialogue acts are present in those sentences too. On the other hand, looking at the corpora we observed that many utterances that appear to be linguistically consistent with the typical structure of statements have been annotated differently, according to the actual communicative role they play. The following is an example of a statement-like utterance (by speaker B) that has been annotated differently because of its context (speaker A's move):

A: 'In fact, it's easier for me to say, uh, the types of music that I don't like are opera and, uh, screaming heavy metal.' STATEMENT
B: 'The opera, yeah, it's right on track.' AGREE-ACCEPT

For similar reasons, we observed some misclassification of S-OPINION as STATEMENT. The only significative difference between the two labels seems to be the wider usage of 'slanted' and affectively loaded lexicon when conveying an opinion. Another source of confounding is the misclassification of the OPENING as INFO-REQUEST. The reason is not clear yet, since the misclassified openings are not question-like. Eventually, there is some confusion among the backchannel labels (GEN-ANS, AGREE-ACC and REJECT) due to the inherent ambiguity of common words like *yes, no, yeah, ok*.

Recognition of such cases could be improved (i) by enabling the classifiers to consider not only the lexical semantics of the given utterance but also the knowledge about a wider context window (e.g. the previous *n* utterances), (ii) by enriching the data preprocessing (e.g. by exploiting information about lexicon polarity and subjectivity parameters).

5 Exploiting Contextual Features

The findings in Section 4.1 highlight the role played by the context in determining the actual communicative goal of a given dialogue turn: manual annotation of utterances is shown to depend not only on the linguistic realization itself. On the contrary, the knowledge about the dialogue history constitutes a bias for human annotators.

This is consistent with Levinson's theory of conversational analysis. Both local and global contextual information contribute in defining the communicative intention of a dialogue turn [24]. In this perspective, top-down expectation about the next likely dialogue act and bottom-up information (i.e. the actual words used in the utterance or its acoustic and prosodic parameters) should be combined to achieve better performance in automatic DA prediction.

Stolcke et al. [5] propose an approach that combines HMM discourse modeling with consideration of linguistic and acoustic features extracted from the dialogue turn. Poesio and Mikheev [25] exploit the hierarchical structure of discourse, described in terms of game structure, to improve DA classification in spoken interaction. Reithinger and Klesen [12] employ a Bayesian approach to build a probabilistic dialogue act classifier based on textual input.

In this section we present some experiments that exploit knowledge about dialogue history. In our approach, each utterance is enriched with contextual information (i.e. the preceding DA labels) in form of either 'bag_of_words' or 'n-grams'. We explore the supervised learning framework, using SVM, under five different experimental settings. Then, we propose a bootstrap approach for the unsupervised setting. In order to allow comparison with the results in Section 4 we refer, for both languages, to the same train/test partitions employed in our previous experiments.

Supervised. We have tested the role played by the context in DA recognition, experimenting with: (i) the number of turn (one vs. two turns) considered in extracting contextual features (i.e. DA labels) based on the dialogue history of a given turn and (ii) the approach used for representing the knowledge about the context, i.e. Bag_of_Words style (BoW) vs. n-grams.

Data preprocessing involves enriching both, the train and test sets, with contextual information, as shown in Table 6. When building the context for a given utterance we only consider the label included in our DA annotation language (see Table 2). In fact, our markup language does not allow mapping of SWBD-DAMSL labels such as ‘non verbal turn’ or ‘abandoned turn’. According to our goal of defining a method which simply exploits textual information, we consider all cases originally annotated with such labels as a lack of knowledge about the context.

Table 6. Enriching the data set with contextual features

natural language input:		
(a1)	STATEMENT	‘I don’t feel comfortable about leaving my kids in a big day care center’
(b1)	INFO-REQ	‘Worried that they’re not going to get enough attention?’
(a2)	GEN-ANS	‘Yeah’
correspondent dataset item for the utterance a2:		
BoW	STATEMENT:1 INFO-REQUEST:1 yeah:1	
Bigram	STATEMENT&INFO-REQUEST:1 yeah:1	

Table 7 (a) shows the results in terms of precision, recall and F1-measure. As comparison, we also report the global performance when no context features are used in the supervised setting. For both the Italian and English corpora, bigrams seem to best capture the dialogue structure. In particular, using a BoW style seems to even lower the performance with respect to the setting in which no information about the context is exploited. Neither combining bigrams with Bag_of_Words nor using higher-order n-gram improve the performance.

Table 7. Overall performance of the different approaches for exploiting contextual information in the supervised setting (a) and bootstrap on the unsupervised method (b)

	English		
Experimental Setting	prec	rec	F1
no context	*.77*	*.77*	*.77*
1 turn of context	.49	.49	.49
BoW (2 turns)	.76	.76	.76
Bigrams (2 turns)	.83	.83	.83
BoW + Bigrams (2 turns)	.83	.83	.83
	Italian		
no context	*.71*	*.71*	*.71*
Bigrams (2 turns)	.82	.82	.82

(a)

	English		
Experimental Setting	prec	rec	F1
no context	*.68*	*.68*	*.68*
Bigrams (2 turns)	.70	.70	.70
	Italian		
no context	*.66*	*.66*	*.66*
Bigrams (2 turns)	.72	.72	.72

(b)

Unsupervised. According to the results in the previous section, we decided to investigate the use of bigrams in the unsupervised learning condition using a bootstrap approach. Our bootstrap procedure is composed by the following steps: (i) annotating the English and Italian corpora using the unsupervised approach described in Section 4; (ii) using the result of this unsupervised annotation for extracting knowledge about contextual information for each utterance: each item in the data sets is then enriched with the appropriate bigram, as shown in Table 6; (iii) training an SVM classifier on the bootstrap data enriched with bigrams. Then performance is evaluated on the test sets (see Table 7 (b)) according to the actual label given by human annotators.

6 Affective Load of Dialogue Acts

Sensing emotions from text is a particularly appealing task of natural language processing [26,27]: the automatic recognition of affective states is becoming a fundamental issue in several domains such as human-computer interaction or sentiment analysis for opinion mining. Recently there have been several attempts to integrate emotional intelligence into user interfaces [28,29,15]. A first attempt to exploit affective information in dialogue act disambiguation has been made by Bosma and André [30], with promising results. In their study, the recognition of emotions is based on sensory inputs which evaluate physiological user input.

In this section we present some preliminary results of a qualitative study aimed at investigating the affective load of DAs. To the best of our knowledge, this is the first attempt to study the relation between the communicative act of an utterance and its affective load by applying lexical similarity techniques to textual input.

We calculate the affective load of each DA label using the methodology described in [7]. The idea underlying the method is the distinction between *direct* and *indirect* affective words. For direct affective words, authors refer to the WordNet Affect [31] lexicon, an extension of the WordNet database [32] which employs six basic emotion labels (anger, digust, fear, joy, sadness, surprise) to annotate WordNet synsets. LSA is then used to learn, in an unsupervised setting, a vector space from the British National Corpus[6]. As said before, LSA has the advantage of allowing homogeneous representation and comparison of words, text fragments or entire documents, using the pseudo-document technique exploited in Section 4. In the LSA space, each emotion label can be represented in various way. In particular, we employ the 'LSA Emotion Synset' setting, in which the synsets of direct emotion words are considered. The affective load of a given utterance is calculated in terms its lexical similarity with respect to one of the six emotion labels. The overall affective load of a sentence is then calculated as the average of its similarity with each emotion label.

Results are shown in Table 8 (a) and confirm our preliminary hypothesis (see error analysis in Section 4.1) about the use of slanted lexicon in opinions. In fact, S-OPINION is the DA category with the highest affective load. Opinions are immediately followed by KIND-ATT due to the high frequency of politeness formulas in such utterances (see Table 8 (b)).

7 CONCLUSIONS AND FUTURE WORK

The long-term goal of our research is to define an unsupervised method for Dialogue Acts recognition. The techniques employed have to be independent from some important features of the corpus used such as domain, language, size, interaction scenario.

In this study we propose a method that simply exploits the lexical semantics of dialogue turns. In particular we consider DA classification with and without considering contextual features. The methodology starts

[6] `http://www.hcu.ox.ac.uk/bnc/`

Table 8. Affective load of DA labels (a) and examples of slanted lexicon (b)

Label	Affective Load
S-OPINION	.1439
KIND-ATT	.1411
STATEMENT	.1300
INFO-REQ	.1142
CLOSING	.0671
REJECT	.0644
OPENING	.0439
AGREE-ACC	.0408
GEN-ANS	.0331

(a)

S-OPINION
Gosh uh, it's getting pathetic now, absolutely pathetic.
They're just horrid, you'll have nightmares, you know.
That's no way to make a decision on some terrible problem.
They are just gems of shows. Really, fabulous in every way.
And, oh, that is so good. Delicious.
KIND-ATTITUDE
I'm sorry, I really feel strongly about this.
Sorry, now I'm probably going to upset you.
I hate to do it on this call.

(b)

with automatically enriching the corpus with additional features (linguistic markers). Then the unsupervised case consists of defining a very simple and intuitive set of seeds that profiles the specific dialogue acts, and subsequently performing a similarity analysis in a latent semantic space. The performance of the unsupervised experiment has been compared with a supervised state-of-art technique such as Support Vector Machines.

Results are quite encouraging and show that lexical knowledge plays a fundamental role in distinguishing among DA labels. Though, the analysis of misclassified cases suggested us to (i) include the consideration of knowledge about context (e.g. the previous *n* utterances) and (ii) to check the possibility of enriching the preprocessing techniques by introducing new linguistic markers (e.g. features related to the use of slanted lexicon, which seems to be relevant in distinguishing between objective statements and expressions of opinion).

Regarding the consideration of knowledge about the dialogue history, we have tested first of all the role played by contextual features in different experimental settings, achieving promising results. In particular bigrams are shown to cause a significant improvement in the DA recognition performance especially in the supervised framework. The improvement is less significant in the unsupervised learning condition, in which a bootstrap based approach is implemented. Improving the bootstrap approach for including contextual information in our unsupervised framework will be object of further investigation in our future research.

We also performed a qualitative study about the affective load of utterances. The experimental results are preliminary but show that a relation exists between the affective load and the DA of a given utterance. According to these experimental evidences, we decided to further investigate, in the next future, the possibility of considering the affective load of utterances in disambiguating DA recognition. In particular, it would be interesting to exploit the role of slanted or affective-loaded lexicon to deal with the misclassification of opinions as statements. Along this perspective, DA recognition could serve also as a basis for conversational analysis aimed at improving a fine-grained opinion mining in dialogues.

REFERENCES

1. Austin, J.: How to do Things with Words. Oxford University Press, New York (1962)
2. Searle, J.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge, London (1969)
3. Core, M., Allen, J.: Coding dialogs with the DAMSL annotation scheme. In: Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines, Cambridge, MA (1997) 28–35
4. Traum, D.: 20 questions for dialogue act taxonomies. Journal of Semantics **17** (2000) 7–30
5. Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Ess-Dykema, C.V., Ries, K., Shriberg, E., Jurafsky, D., Martin, R., Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational Linguistics **26** (2000) 339–373
6. Warnke, V., Kompe, R., Niemann, H., Nöth, E.: Integrated dialog act segmentation and classification using prosodic features and language models. In: Proceedings of 5th European Conference on Speech Communication and Technology. Volume 1., Rhodes, Greece (1997) 207–210
7. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: SAC '08: Proceedings of the 2008 ACM symposium on Applied computing, New York, NY, USA, ACM (2008) 1556–1560
8. Schegloff, E.: Sequencing in conversational openings. American Anthropologist **70** (1968) 1075–1095
9. Sacks, H., Schegloff, E., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. Language **50** (1974) 696–735

10. Power, R.: The organisation of purposeful dialogues. Linguistics **17** (1979) 107–152
11. Cohen, P.R., Levesque, H.J.: Communicative actions for artificial agents. In: in Proceedings of the First International Conference on Multi-Agent Systems, AAAI Press (1995) 65–72
12. Reithinger, N., Klesen, M.: Dialogue act classification using language models. In: In Proceedings of EuroSpeech-97. (1997) 2235–2238
13. Venkataraman, A., Liu, Y., Shriberg, E., Stolcke, A.: Does active learning help automatic dialog act tagging in meeting data? In: Proceedings of EUROSPEECH-05, Lisbon, Portugal (2005)
14. Godfrey, J., Holliman, E., McDaniel, J.: SWITCHBOARD: Telephone speech corpus for research and development. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), San Francisco, CA, IEEE (1992) 517–520
15. Clarizio, G., Mazzotta, I., Novielli, N., deRosis, F.: Social attitude towards a conversational character. In: Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, UK (2006) 2–7
16. Jurafsky, D., Shriberg, E., Biasca, D.: Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13. Technical Report 97-01, University of Colorado Institute of Cognitive Science (1997)
17. Novielli, N.: Hmm modeling of user engagement in advice-giving dialogues. Journal on Multimodal User Interfaces (2009)
18. Pianta, E., Girardi, C., Zanoli, R.: The TextPro tool suite. In: Proceedings of LREC-08, Marrakech, Morocco (2008)
19. Landauer, T.K., Foltz, P., Laham, D.: Introduction to latent semantic analysis. Discourse Processes **25** (1998)
20. Berry, M.: Large-scale sparse singular value computations. International Journal of Supercomputer Applications **6** (1992)
21. Gliozzo, A., Strapparava, C.: Domains kernels for text categorization. In: Proc. of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), University of Michigan, Ann Arbor (2005) 56–63
22. Vapnik, V.: The Nature of Statistical Learning Theory. Springer-Verlag (1995)
23. Joachims, T.: Text categorization with Support Vector Machines: learning with many relevant features. In: Proceedings of the European Conference on Machine Learning. (1998)
24. Levinson, S.C.: Pragmatics. Cambridge University Press, Cambridge; New York (1983)
25. Poesio, M., Mikheev, A.: The predictive power of game structure in dialogue act recognition: Experimental results using maximum entropy estimation. In: Proceedings of ICSLP-98, Sydney (1998)
26. Strapparava, C., Mihalcea, R.: SemEval-2007 task 14: Affective Text. In: Proceedings of the 4^{th} International Workshop on Semantic Evaluations (SemEval 2007), Prague (2007) 70–74

27. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval **2** (2008) 1–135
28. Conati, C.: Probabilistic assessment of user's emotions in educational games. Applied Artificial Intelligence **16** (2002) 555–575
29. Picard, R.W., Klein, J.: Computers that recognise and respond to user emotion: Theoretical and practical implications. Technical report, MIT Media Lab (2001)
30. Bosma, W., André, E.: Exploiting emotions to disambiguate dialogue acts. In: IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces, New York, NY, USA, ACM (2004) 85–92
31. Strapparava, C., Valitutti, A.: WordNet-Affect: an affective extension of WordNet. In: Proceedings of LREC. Volume 4. (2004) 1083–1086
32. Fellbaum, C., ed.: WordNet: An Electronic Lexical Database (Language, Speech, and Communication). The MIT Press (1998)

NICOLE NOVIELLI
UNIVERSITÀ DEGLI STUDI DI BARI,
DIPARTIMENTO DI INFORMATICA
VIA ORABONA, 4 - 70125 BARI, ITALY
E-MAIL: <NOVIELLI@DI.UNIBA.IT>

CARLO STRAPPARAVA
FBK-IRST, ISTITUTO PER LA RICERCA
SCIENTIFICA E TECNOLOGICA,
VIA SOMMARIVE, 18 - I-38050 POVO TRENTO, ITALY
E-MAIL: <STRAPPA@FBK.EU>