Automatic dialogue act recognition with syntactic features

Pavel Kral, Christophe Cerisara

Automatic Dialogue Act Recognition with Syntactic Features

Pavel Král[1, 2] · **Christophe Cerisara**[3]

Abstract This work studies the usefulness of syntactic information in the context of automatic dialogue act recognition in Czech. Several pieces of evidence are presented in this work that support our claim that syntax might bring valuable information for dialogue act recognition. In particular, a parallel is drawn with the related domain of automatic punctuation generation and a set of syntactic features derived from a deep parse tree is further proposed and successfully used in a Czech dialogue act recognition system based on Conditional Random Fields. We finally discuss the possible reasons why so few works have exploited this type of information before and propose future research directions to further progress in this area.

Keywords Dialogue Act · Language Model · Sentence Structure · Speech Act · Speech Recognition · Syntax

[1]Dept. of Computer Science & Engineering
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
Tel.: +420-377632454
Fax: +420-377632402
E-mail: pkral@kiv.zcu.cz

[2] NTIS - New Technologies for the Information Society
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic

[3] LORIA UMR 7503
BP 239 - 54506 Vandoeuvre
France
Tel.: +33-354958625
Fax: +33-383278319
E-mail: cerisara@loria.fr

1 Introduction

1.1 Definition

Modelling and automatically identifying the structure of spontaneous dialogues is very important to better interpret and understand them. The precise modelling of dialogues is still an open issue, but several specific characteristics of dialogues have already been clearly identified. *Dialogue Acts (DAs)* are one of these characteristics.

Although the term "dialogue acts" that is commonly used nowadays has been defined by Austin in [1], a number of other seminal works have proposed very similar notions, including speech acts proposed by John R. Searle in [2], conversational game moves introduced by R. J. D. Power in [3], adjacency pairs proposed by Schegloff in [4,5] or acts of communication in the plan-based approaches to understanding introduced by Litman et al. in [6–8]. The theory of the dialogue acts has been further developed by Hary Bunt in [9]. The dialogue acts represent the meaning of an utterance in the context of a dialogue, where the context is divided into several types, with both global and local views: linguistic, semantic, physical, social and cognitive. Bunt also developed a multidimensional taxonomy of the dialogue acts, while David R. Traum developed the notion of speech acts in [10] with dialogue agents. A better overview of the notion of dialogue acts can be found in [11].

In this work, the dialogue act is seen as a function of an utterance, or its part, in the dialogue. For example, the function of a question is to request some information, while an answer shall provide this information.

Table 1 illustrates the dialogue acts that may occur in a dialogue between the passenger (P) and the agent (A) in a ticket reservation task. The corresponding dialogue act labels are also shown. Each utterance is labelled with a unique dialogue act. This example is taken from our Czech corpus (see Section 5.1).

Table 1 Example of a dialogue between the passenger (P) and the agent (A) in a ticket reservation task with the English translation

Speaker	DA	Dialogue in Czech	English translation
P	Question	Kdy pojede první vlak do Prahy ?	When will the first train go to Prague ?
A	Question yes/no	Chcete rychlík ?	Do you want the express train ?
P	Statement	To je jedno.	I don't care.
A	Statement	V osm hodin.	At eight o'clock.
P	Order	Dejte mi tedy jeden lístek, prosím.	Give me one ticket, please.
A	Statement	Tady je.	Here it is.

Dialogue acts represent useful and relevant information for many applications, such as dialogue systems, machine translation, automatic speech recognition, topic tracking [12] or talking head animation. For instance, in dialogue systems, dialogue acts might be used to recognize the intention of the user and thus differentiate situations where the user is requesting some information from situations where the user is simply giving some information or backchannels. In the former case, the system has to react, while in the latter case, a system reaction may be perceived as intrusive. In the machine translation domain, recognizing dialogue acts may bring relevant cues to choose between alternative translations, as the adequate syntactic structure may depend on the user intention. Automatic recognition of dialogue acts may also be used to improve the word recognition accuracy of automatic speech recognition systems, as proposed for instance in [13], where a different language model is applied during recognition depending on the dialogue act. Finally, dialogue act recognition is a fundamental building block of any understanding system and typically completes semantic role labelling and semantic frame inference.

The usefulness of dialogue act recognition has thus been demonstrated in a number of large applicative systems, such as the *VERBMOBIL* [14], *NESPOLE* [15] and *C-STAR* [16] machine translation and dialogue systems that rely on dialogue act classification.

1.2 Objectives

The main objective of this work is to propose and investigate the usefulness of syntactic features to improve dialogue act recognition in Czech. In previous works, we have first designed a baseline dialogue act recognition system for the Czech language that was based on generative models [17]. Although reasonably good results have been obtained, this approach was limited because it only exploits the local context around any given word of the utterance. We then proposed in [18] and [19] several approaches to address this limitation and include global features in the model that represent the sentence structure. One of these approaches consists in modelling the word position in the sentence as a random variable and integrating this variable in the generative model. Intuitively, this information is important for dialogue act recognition, as for instance, the word "who" is often located at the beginning of sentences for questions and at other positions for declarative sentences. In the following, we propose a different approach to model such global information implicitly, via a conditional stochastic model. The second and most important contribution of this work concerns the design and exploitation of syntactic features for dialogue act recognition in Czech. As summarized in Section 2, only a few types of features are generally used in the literature to automatically recognize dialogue acts: lexical, Part-of-Speech (POS) tags, dialogue history and prosody. Furthermore, word sequences are most of the time modelled by statistical n-gram models, which encode the relationship between words and dialogue acts only locally. While we have already shown the importance of global informa-

tion such as word position in the utterance for dialogue act recognition, the current work goes beyond this type of information by investigating whether the conditional distribution of the target dialogue act depends on the syntactic structure of the utterance.

In the following section, we briefly review the state of the art about dialogue act recognition, with a focus on how syntactic information has already been considered for this task and for related tasks. In Section 3, we propose and describe new syntactic features. The proposed model is described in Section 4. The relative importance of each of these features is evaluated on a Czech dialogue corpus in Section 5. In the last section, we discuss these results and propose some future research directions.

2 Related Work

We will now briefly review the standard definitions of dialogue acts, the different types of models classically used for dialogue act recognition and the standard types of information used in such models. Then, we review and discuss the previous design of syntactic features for dialogue act recognition as well as in closely related domains.

Some generic sets of domain-independent dialogue acts have been proposed in the state-of-the-art and are now commonly used to create the baseline tag set for most types of applications. Hence, in [11], 42 dialogue acts classes are defined for English, based on the Discourse Annotation and Markup System of Labelling (DAMSL) tag-set [20]. The Switchboard-DAMSL tag-set [21] (SWBD-DAMSL) is an adaptation of DAMSL in the field of telephone conversations. The Meeting Recorder Dialogue Act (MRDA) tag-set [22] is another very popular tag-set, which is based on the SWBD-DAMSL taxonomy. MRDA contains 11 general dialogue act labels and 39 specific labels. Finally, Jekat [23] defines for German and Japanese 42 dialogue acts, with 18 dialogue acts at the illocutionary level, in the context of the VERBMOBIL corpus. The ISO standard 24617-2 for dialogue annotation has been published in 2012. DIT++ [1] is a recent implementation of this standard. Because of the limited size of the available corpus, as well as several other technical reasons, these tag sets are frequently reduced by merging several tags together, so that the number of final actual generic tags is often about 10. Part of such typical generic dialogue acts, also referred to as speech acts, include for instance [24] statements, questions, backchannels, commands, agreements, appreciations as well as a broad "miscellaneous" class. In addition to such generic tags, application-specific tags may be defined, such as "request booking" for a hotel booking application.

Manually annotating dialogue acts on every new corpus may be very costly and efforts have been put into developing semi-automatic methods for dialogue act tagging and discovery. Hence, the authors of [25] propose a predictive paradigm where dialogue act models are first trained on a small-size corpus and

[1] `http://dit.uvt.nl`

used afterwards to predict future sentences or dialogue acts. In a related vein, unsupervised dialogue act tagging of unlabelled text has recently raised a lot of attention [26,27], but we will limit ourselves in the following on supervised approaches.

The dialogue act recognition task is often considered jointly with the segmentation task. We assume in our work that sentence segmentation is known, because we rather prefer to concentrate on the challenge of designing relevant syntactic features for dialogue act recognition. Yet, many related works propose powerful solutions for the segmentation task as well. In particular, the work described in [28] considers the input text as a stream of words and segments and tags it incrementally with a BayesNet model with lexical, prosodic, timing and dialogue act-history features. Zimmermann et al. successfully use in [29] for joint DA segmentation and classification hidden-event language models and a maximum entropy classifier. They use word sequence and pause duration as features. The authors of [30] exploit a Switching Dynamic Bayesian Network for segmentation, cascaded with a Conditional Random Field for dialogue act classification, while [31] jointly segments and tags with a single model.

The dialogue act modelling schemes that are commonly used for dialogue act recognition are traditionally chosen from the same set of general machine learning methods used in most natural language processing tasks. These include Hidden Markov Models [11], Bayesian Networks [32], Discriminative Dynamic Bayesian Networks [33], BayesNet [28], Memory-based [34] and Transformation-based Learning [35], Decision Trees [36], Neural Networks [37], but also more advanced approaches such as Boosting [38], Latent Semantic Analysis [39], Hidden Backoff Models [40], Maximum Entropy Models [41], Conditional Random Fields [31,30] and Triangular-chain CRF [42].

Regarding features, most dialogue act recognition systems exploit both prosodic and lexical features. The dialogue history is also often used as relevant information. Some cue words and phrases can also serve as explicit indicators of dialogue structure [43]. For example, 88.4% of the trigrams "<start> do you" occur in English in *yes/no questions* [44].

Prosody is an important source of information for dialogue act recognition [24]. For instance, prosodic models may help to capture the following typical features of some dialogue acts [45]:

- a falling intonation for most statements
- a rising F0 contour for some questions (particularly for declaratives and yes/no questions)
- a continuation-rising F0 contour characterizes (prosodic) clause boundaries, which is different from the end of utterance

In [24], the duration, pause, fundamental frequency (F0), energy and speaking rate prosodic attributes are modelled by a CART-style decision trees classifier. In [46], prosody is used to segment utterances. The duration, pause,

F0-contour and energy features are used in [13] and [47]. In both [13] and [47], several features are computed based on these basic prosodic attributes, for example the max, min, mean and standard deviation of F0, the mean and standard deviation of the energy, the number of frames in utterance and the number of voiced frames. The features are computed on the whole sentence and also on the last 200 ms of each sentence. The authors conclude that the end of sentences carry the most important prosodic information for dialogue act recognition. Shriberg et al. show in [24] that it is better to use prosody for dialogue act recognition in three separate tasks, namely question detection, incomplete utterance detection and agreements detection, rather than for detecting all dialogue acts in one task.

Apart from prosodic and contextual lexical features, only a few works actually exploit syntactic relationships between words for dialogue act recognition. Some syntactic relations are captured by HMM word models, such as the widely-used n-grams [11], but these approaches only capture local syntactic relations, while we consider next global syntactic trees. Most other works thus focus on morphosyntactic tags, as demonstrated for instance in [48], where a smart compression technique for feature selection is introduced. The authors use a rich feature set with POS-tags included and obtain with a decision tree classifier an accuracy of 89.27%, 65.68% and 59.76% respectively on the ICSI, Switchboard and on a selection of the AMI corpus. But while POS-tags are indeed related to syntax, they do not encode actual syntactic relations.

A very few number of works have nevertheless proposed some specific structured syntactic features, such as for instance the subject of verb type [49]. The authors of [39,50] exploit a few global syntactic features, in particular POS-tags and the MapTask SRule annotation that indicates the main structure of the utterance, i.e., Declarative, Imperative, Inverted or Wh-question, but without obtaining a clear gain from syntax in their context, hence suggesting that further investigation is needed. Indeed, syntax is a very rich source of information and the potential impact of syntactic information highly depends on the chosen integration approach and experimental setup. We thus propose in the next section other types of syntactic features and a different model and show that syntax might indeed prove useful for dialogue act recognition in the proposed context. But let us first support our hypothesis by briefly reviewing a few other papers that also support the use of syntax for both dialogue act recognition and closely related domains.

First, as already shown, word n-grams features, with n greater than 1, do implicitly encode local syntactic relations and are used successfully in most dialogue act recognition systems. But more importantly, a recent work [51] concludes that both dialogue context and syntactic features dramatically improve dialogue act recognition, compared to words only, more precisely from an accuracy of 48.1% up to 61.9% when including context and 67.4% when further including syntactic features. They use in their experiments a Bayesian Network model and their syntactic features are the syntactic class of the predicate, the list of arguments and the presence of a negation. Although this work actually focuses on predicate-argument structures, while our main objective

is rather to exploit the full syntactic tree without taking into account any semantic-level information for now, this work supports our claim that syntactic information may prove important for dialogue act recognition. In addition, Zhou et al. employ in [52] three levels of features: 1) word level (unigram, bigram and trigram), 2) syntax level (POS-tags and chunks recognized as Base Noun Phrase (BNP)) and 3) restraint information (word position, utterance length, etc.). Syntactic and semantic relations are acquired by information extraction methods. They obtain 88% of accuracy with a SVM classifier on a Chinese corpus and 65% on the SWBD corpus.

We further investigated closely related domains that have already explored this research track in more depth. This is for instance the case of automatic classification of rhetorical relations, as reviewed in [53]. Another very close task is punctuation recovery, which aims at generating punctuation marks in raw words sequences, as typically obtained from speech recognition systems. In particular, this implies to discriminate between questions (ending with a question mark), orders (ending with an exclamation points) and statements (ending with a period), which is a task that is obviously strongly correlated to dialogue act recognition. Interestingly enough, a richer set of syntactic features have been exploited in the punctuation recovery domain than in the dialogue act recognition area. Hence, the authors of [54] design several syntactic features derived from the phrase structure trees and show that these features significantly reduce the detection errors. This is in line with our own previous conclusions published in [55] regarding the use of syntactic features for punctuation recovery, where a large improvement in performances is obtained thanks to syntactic information derived from dependency trees. Similar gains are obtained on a Chinese punctuation task [56], where including rich syntactic features, such as the word grammatical function, its ancestors and children, its head, the yield of the constituent or subtree border indicators, improve the F-measure from 52.61% up to 74.04%.

Finally, we have shown that there is an increasing amount of work that successfully exploits structural syntactic dependencies both for dialogue act recognition and in related domains such as punctuation recovery. We further believe that parsing of natural language utterances will constitute a fundamental pre-processing step of most if not all subsequent NLP modules, although it has probably not been as widely used as POS tagging for instance because of its complexity and lack of robustness to ill-formed input. However, thanks to the current progress in Bayesian approaches and feature-rich log-linear models, we expect parsing to be more and more robust to automatic speech recognition errors in the near future. Other recent reviews of the literature about dialogue act recognition are realized in [57] and [43].

3 Syntax for Automatic Dialogue Act Recognition

In our previous work [18], [58] and [19], we proposed to include in our DA recognition approach information related to the position of the words within

the sentence. In this work, we propose a different approach that derives features from the sentence parse tree and includes these features as input to a conditional random field. The parse trees are defined within the dependency framework [59] and are automatically computed on the input sentences. We evaluate this approach in two conditions, respectively when the input sentences are manually and automatically transcribed.

3.1 Features

We distinguish next two types of features, respectively the baseline and syntactic features. The baseline features are:

- words inflected form
- lemmas
- part-of-speech tags
- pronoun or adverb at the beginning of the sentence
- verb at the beginning of the sentence

The syntactic features rely on a dependency parse tree:

- dependency label
- root position in the utterance
- unexpressed subjects
- basic composite pair (subject-verb inversion)

All these features are described in details next.

3.1.1 Baseline Features

Words Inflected Form The word form is used as a baseline lexical feature in most modern lexicalized natural language processing approaches [11,44,32, 33]. In our case, sentence segmentation is known but capitalization of the first word of the sentence is removed, which decreases the total number of features in our model without impacting accuracy, thanks to the insertion of a special "start-of-utterance" word. Although word bigrams or trigrams are commonly used in other systems, we only use word unigrams because of the limited size of the training corpus. We rather compensate for this lack of local structural information by investigating global syntactic dependencies. The word forms are obtained in our experiments using both manual and automatic transcriptions of speech audio files.

Lemmas We used the *lemma* structure from the Prague Dependency Treebank (PDT) 2.0[1] [60] project, which is composed of two parts. The first part is a unique identifier of the lexical item. Usually it is the base form (e.g. infinitive for a verb) of the word, possibly followed by a digit to disambiguate different lemmas with the same base forms. The second optional part contains

additional information about the lemma, such as semantic or derivational information. Lemmas may in some circumstances bring additional information, notably by removing irrelevant variability introduced by inflected forms. This may have some importance in particular for rare words that may occur with different inflected forms but still may have some impact on the dialogue act decision process. The lemmas are obtained automatically in our experiment with a lemmatizer.

Part-of-Speech (POS) tags The part-of-speech is a word linguistic category (or more precisely lexical item), which can be defined by the syntactic or morphological behaviour of the lexical item in question. There are ten POS categories defined in the PDT [60] for the Czech language: nouns, adjectives, pronouns, numerals, verbs, adverbs, prepositions, conjunctions, particles and interjections. The part-of-speech tags are inferred automatically in our experiment with a POS tagger.

Pronoun or Adverb at the Beginning of the Sentence This boolean feature indicates whether the utterance starts with a pronoun or an adverb. It can be particularly useful for detecting *Wh-questions*, which usually start with a *Pronoun* (POS-tag "P") or an *Adverb* (POS-tag "D"), such as in: "Kdy přijdeš domů?" (When do you come home?).

Note that similar features that emphasize the importance of initial words in the sentence have already been proposed, for instance in [61, 43, 41].

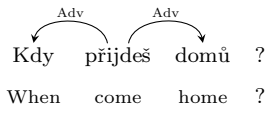

Fig. 1 Example of adverb as first word.

Verb at the Beginning of the Sentence This feature is also a boolean indicator of the presence of a verb as the first word of an utterance. It can be particularly useful for the detection of *Commands* and *Yes-no questions*, which usually start with a verb, such as in: "Jdi domů!" (Go home!) and "Půjdeš domů?" (Do you go home?).

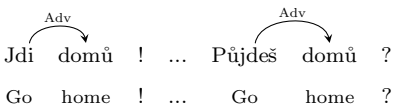

Fig. 2 Example of verb as first word.

[1] http://ufal.mff.cuni.cz/pdt2.0/

3.1.2 Syntactic Features

All syntactic features are computed from the syntactic tree obtained after automatic parsing of the target sentence: a detailed description of our automatic parser is given in Section 5.2. We have chosen to represent the syntactic relations with *dependencies*, as it is commonly done nowadays for many natural language processing tasks. Furthermore, we have chosen the Prague Dependency Treebank to train our stochastic parser and our annotation formalism thus follows the one used in the PDT.

An example of such a dependency tree is shown in Figure 3, where the words represent the nodes of the tree and the arcs the dependencies between words. Dependencies are oriented, with each arrow pointing to the dependent word of the relation. Each dependency is further labelled with the name of a syntactic relation, such as *Sb* for subject, *Pred* for predicate, *Atr* for attribute, *Obj* for object, etc.

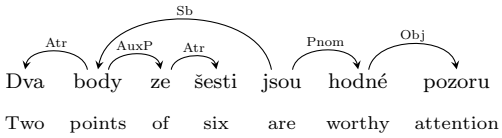

Fig. 3 Example of an instance of a Czech *statement* dialogue act (*Two points out of six are worthy of attention.*) with its parse tree.

Dependency Label The first generic feature derived from the parse tree is the label of the dependency from the target word to its head. For example, the values taken by this feature in Figure 3 are, for each word: *Atr, Sb, AuxP, Atr, Root, Pnom, Obj.*

Root Position in the Utterance In theory, every utterance is parsed into a single dependency tree. The position of the root of this tree is likely to depend on the type of sentence and dialogue act. Hence, intuitively, the root tends to be positioned in the middle of declarative sentences, as in Figure 3, while it is more often located at the start of utterances for commands/orders, such as in: "Zavři dveře!"(Close the door!).

This feature is the absolute position of the root, after normalization of the sentence length to 10 words. The normalization is realized with a standard binning technique, eventually filling empty internal bins with virtual non-root words for short sentences, so that the word at the middle of the sentence is in bin 5 and recursively in the left and right halves of the sentence.

Unexpressed Subject This feature is a boolean feature that is true if and only if a subject dependency exists for the first verb in the sentence. In-

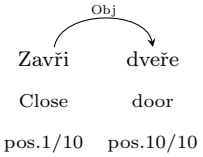

Fig. 4 Example of leftmost root position.

deed, verbs without subjects may intuitively occur more frequently in commands/orders than in declarative sentences, as illustrated in the previous example. This is however not always true, especially in the Czech language, where unexpressed subjects are quite common and thus often occur in most dialogue acts, such as in: "Šel do kina." (He went to the cinema.).

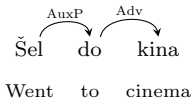

Fig. 5 Example of unexpressed subject.

Basic Composite Pair This feature is a boolean value that encodes the relative position of each pair *Subject* and *verb*. When the verb precedes the subject, this is often viewed as strong evidence in favour of detecting a question in many European languages such as English and French. However, in the Czech language, this is not always true because of two main factors:

1. Subjects may be omitted, as explained in the previous section.
2. A statement can start with a *Direct Object*, followed by a *Verb* and its *Subject*, such as in "Květiny dostala matka." (The mother got flowers).

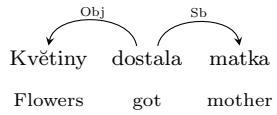

Fig. 6 Example of inverted subject.

4 Dialogue Act Model

4.1 General Principle

The general principle of our dialogue act recognition approach is to decompose the problem of tagging a complete sentence into the (easier) problems

of tagging individual words. Our basic assumption is that every single word contributes to the global dialogue act depending on its form, nature and global context. The proposed approach thus assigns a single dialogue act tag to every word and then combines all the dialogue act tags that contribute to the same sentence to infer a single dialogue act tag for this sentence. The word-tagging process is implemented with a Conditional Random Field (CRF) while the sentence-tagging process is realized with two simple combination models that are described next.

4.2 Training and Pre-processing

Only the word-level CRF model is trained. The second combination stage is realized by a non-parametric decision process and thus does not need any training.

The manually annotated dialogue act tag associated to each training utterance is first duplicated and assigned to every word of the utterance. Then, these utterances are automatically tagged with POS-tags and parsed with the Malt parser [62] to produce a dependency tree. A vector of lexical and syntactic features is then derived from this parse tree for each word of the utterance. A special word is inserted before every utterance, with a single feature that indicates the start of an utterance. This special word is given the same dialogue act tag as the other words of the sentence. Finally, all these feature vectors, along with their associated dialogue act tags are pooled together in sequence and the CRF is trained on this corpus with the classical L-BFGS algorithm.

The data pre-processing procedure described above also applies to the test corpus.

4.3 Testing and Dialogue Act Inference

During testing, both word-level and sentence-level models are involved to infer dialogue acts. In the first step, the previously trained CRF is applied on the current words sequence and outputs one dialogue act tag for every word of the sentence. Then, the sentence-level decision process converts this resulting sequence of dialogue act tags into a single dialogue act tag per sentence.

Note that an alternative, single-stage strategy may have been to use a non-stochastic global approach, for instance with a maximum entropy model and global features. However, such an approach usually exploits a bag-of-word hypothesis or otherwise implies to explicitly define sentence-global features. Although we have already used with some success a similar approach in a previous work with words position [19], we rather investigate in the current work the proposed two-stage strategy, which focuses on modelling the succession of word-level dialogue act tags.

Hidden Markov Models, Maximum-Entropy Markov models (MEMMs) and Conditional Random Fields (CRF) are amongst the most common stochastic

classifiers. We have chosen CRF because Laferty et al. have shown in [63] that CRFs avoid the label bias problem, as compared to MEMMs. Furthermore, CRFs are conditional models, and as such, make a better use of their parameters than generative models such as HMMs to model the target distribution of probability. They have also proven in recent years to be superior to most variants of HMMs in many natural language processing tasks and in particular in a punctuation generation application, which is closely related to dialogue act recognition. Hence, Favre et al. compared in [54] three sequence models: Hidden-Event Language Model (HELM), factored-HELM and Conditional Random Fields (CRF) for comma prediction. They have shown that the best results are obtained with CRFs, although CRFs may not scale easily to large databases.

We thus use CRFs to compute the conditional probability:

$$\begin{aligned} P(DA|F) &= P(c_0, c_1, \cdots, c_n | f_0, f_1, \cdots, f_n) \\ &= \prod_{i=0}^{n} P(c_i | f_i, c_1, \cdots, c_{i-1}) \end{aligned} \tag{1}$$

where $F =< f_0, f_1, \cdots, f_n >$ represents the sequence of features vectors, n is the number of words in the utterance, f_0 the initial start word and $DA =< c_0, c_1, \cdots, c_n >$ is the output sequence of dialogue acts.

4.4 Sentence-level combination and decision process

We investigate two approaches for the final decision process, which shall output a single dialogue act tag for the whole utterance: Majority voting and Naive Bayes classification.

4.4.1 Majority Voting

The final dialogue act tag is simply the tag with the highest frequency counts amongst the n tags $c_1, \cdots, c_n$. Ambiguous cases are resolved by choosing the tag with the largest posterior probability.

4.4.2 Naive Bayes Classification

In the "Naive Bayes" classifier [64], every tag c_i is assumed independent from all others given the Markov assumption. Hence, the probability over the whole utterance is given by Equation 2:

$$P(c|F) = \prod_{i=1}^{n} P(c_i = c | f_i, c_{i-1}) \tag{2}$$

where $P(c_i = c|f_i, c_{i-1})$ is the word-level posterior probability returned by the first-order CRF, when the CRF is constrained to follow the sub-optimal

path ($c_0 = c, c_1 = c, \cdots, c_n = c$). Note that this tag sequence is different from the one used in the "majority voting case", where the global optimal path returned by the CRF is used.

The resulting dialogue act is the one that maximizes the *a posteriori* probability:

$$\hat{c} = \arg\max_{c} P(c|F)$$

5 Evaluation

The proposed two-step model is evaluated on a Czech train reservation corpus and compared with a unigram model and with a baseline CRF model that only exploits lexical and morpho-syntactic features. The evaluation metric is the dialogue act recognition accuracy. In the following, we first describe the Czech corpus, then the two pieces of software that have been used to compute the morphosyntactic tags and the parse tree and we finally discuss the experimental results.

5.1 Corpus

The corpus used to validate the proposed approaches is the Czech Railways corpus that contains human-human dialogues. It was created at the University of West Bohemia mainly by members of the Department of Computer Science and Engineering in the context of a train ticket reservation dialogue expert system. The whole corpus has been recorded in laboratory conditions and contains about 12 hours of audio recordings. The audio files have been both manually and automatically transcribed. We thus evaluate our dialogue act recognition approach on both types of transcriptions, in order to further assess its robustness to speech recognition errors.

Automatic transcription has been realized with the jLASER [65] recogniser, which has been developed in our LICS [1] laboratory. It is based on a so called *hybrid* framework that combines the advantages of the hidden Markov model approach with those of artificial neural networks. We use HMMs with state emission probabilities computed from the output neuron activations of a neural network (such as the multi-layer perceptron). jLASER has been trained on 6234 sentences (about nine hours), while 2173 sentences (about three hours) pronounced by different speakers are used for testing. Because of the size of the corpus, a class-based 3-gram language model has been used.

All sentences of this "test" corpus have been manually labelled by three different annotators with the following dialogue acts: statements (S), orders (O), yes/no questions (Q[y/n]) and other questions (Q). The DA corpus structure is reported in Table 2, where the number of dialogue acts is shown in column 2. This choice of dialogue acts has been done because our DA recognition module is designed to be used with a rule-based dialogue system that only exploits

these four types of information as an input. The following dialogue act recognition experiments are realized on this labelled corpus using a cross-validation procedure, where 10% of the corpus is reserved for the test, another 10% for the development set and 80% for training of the CRF.

Table 2 Description of the Czech Railways DA corpus

DA corpus			
DA	No.	Example	English translation
S	566	Chtěl bych jet do Písku.	I would like to go to Písek.
O	125	Najdi další vlak do Plzně!	Give me the next train to Plzeň !
Q[y/n]	282	Řekl byste nám další spojení?	Do you say next connection ?
Q	1200	Jak se dostanu do Šumperka?	How can I go to Šumperk ?
Sent.	2173		

5.2 Tools

For lemmatization and POS-tagging, we use the mate-tools `http://code.google.com/p/mate-tools/`. The lemmatizer and POS tagger models are trained on 5853 sentences (94,141 words) randomly taken from the Prague Dependency Tree Bank (PDT 2.0) [60] corpus. The PDT 2.0 is a collection of Czech newspaper texts that are annotated on the three following layers: morphological (2 million words), syntactic (1.5 million words) and complex syntactic and semantic layer (0.8 million words). In this work, only the syntactic dependencies of the second layer are considered. The performance of the lemmatizer and POS tagger are evaluated on a different set of 5181 sentences (94,845 words) extracted from the same corpus. The accuracy of the lemmatizer is 81.09%, while the accuracy of our POS tagger is 99.99%. Our tag set contains 11 POS-tags as described in Table 3.

Table 3 Tag-set description

Abbreviation	Description	Abbreviation	Description
A	Adjective	P	Pronoun
C	Numeral	V	Verb
D	Adverb	R	Preposition
I	Interjection	T	Particle
J	Conjunction	Z	Punctuation
N	Noun		

[1] http://liks.fav.zcu.cz

Our dependency parser is the Malt Parser v 1.3 trained on 32,616 sentences (567,384 words) from PDT 2.0. The dependency set is thus: *Adv, AdvAtr, Apos, Atr, AtrAdv, AtrAtr, AtrObj, Atv, AtvV, AuxC, AuxG, AuxK, AuxO, AuxP, AuxR, AuxT, AuxV, AuxX, AuxY, AuxZ, Coord, ExD, Obj, ObjAtr, Pnom, Pred, Sb.* The Labelled Attachment Score (LAS) of our parser is about 66%.

Our CRF toolkit is based on the Stanford OpenNLP library [2], which has been modified in order to include syntactic features. The resulting model has about 3200 parameters.

5.3 Baseline rule-based system

Our claim in this work is that structured syntactic features, which cannot be simply derived from word forms, bring relevant information that help a classifier to discriminate between some dialogue acts, even in Czech, which is known to be a free-word order language. We actually show next that despite the theoretical linguistic constructions in Czech, which do not a priori strongly constrain the grammatical structures with regard to word orders, common usage in Czech exhibits statistical properties that are discriminative for the few dialogue acts considered here. Furthermore, we show that such statistical properties cannot be captured with simple deterministic rules, but that they must be considered instead in context within a stochastic model like the proposed CRF that is trained on real data.

To illustrate this idea, let's consider the particularly difficult case of yes-no questions vs. statement. Table 4 shows a typical example of such a case, which cannot be captured by syntactic information in Czech.

Table 4 Example of discrimination between yes/no question and statement that cannot be realized from syntactic information in Czech. Punctuation is not shown because it is not available at the output of speech recognizers and is thus not used by our system.

	English	Czech
Statement	He loves her	Miluješ ho
Yes/no question	Does he love her	Miluješ ho

However, despite such difficult theoretical constructions, we have automatically parsed our Czech speech corpus and analyzed the relative frequency of subject relations with the verb on the left of the subject (feature "Basic composite pair"): 48% of such inverted relations occur in statements, which corresponds to a pure random ratio and complies with the free-word order property of Czech, while this ratio goes up to 88% in yes/no questions, which demonstrates that such a feature is indeed informative in common usages of Czech. Nevertheless, we also show next that this observation, in itself, is not

[2] `http://incubator.apache.org/opennlp`

enough to accurately discriminate between yes/no questions and statements, and that it must be considered in context to be really useful.

In order to validate this claim, we build next a deterministic baseline model that classifies the four proposed dialogue acts using hand-crafted rules that:

- Include common lexical knowledge, such as interrogative words
- Use syntactic rules that match the proposed features described in Section 3.1.2, such as the subject-verb inversion rule just described.

The set of rules is described in Table 5. When several rules apply on the same sentence, the chosen dialogue act is decided with a majority vote. In case of equality, the winner amongst competing dialogue acts is the one with the higher prior probability on the corpus, i.e., in decreasing order: Q, S, Q[y/n], O.

The recognition accuracy of the rule-based system is shown in Table 6. We evaluate two cases: manual word transcription and automatic transcription by jLASER recognizer. Table 6 shows that errors from the speech recognizer don't play an important role for DA recognition, resulting in a decrease of accuracy of about 4%.

The highest score for class O might result from the precision of the set of rules defined for this class. Conversely, the lower score of class S may be due to the difficulty to define a specific rule for this class. We thus only used a list of key-words for S and sentences are mainly classified into this class when no rule from another class is triggered.

5.4 Experiments

Two experiments are realized next. The first one performs dialogue act recognition on manual word transcriptions and evaluates and compares the impact of the proposed lexical and syntactic features and the relative performances of both sentence-level combination models. The unigram model corresponds to a very basic baseline approach that only exploits lexical unigram probabilities. We further compare the proposed approach with more advanced baselines that are also based on a CRF but with lexical and morphosyntactic features only (*word forms*, *lemmas* and *POS-tags*). In the second experiment, the same set of models is applied on automatic word transcriptions. This allows assessing the robustness of both our parsers and feature sets to speech recognition errors.

5.4.1 Manual transcription evaluation

Table 7 shows the dialogue act recognition accuracies obtained with the different proposed models. We have computed statistical significance of the difference between two models with the McNemar test, as suggested in [66] for a similar classification task. The p-value is in general below the traditional threshold of 0.05. The p-values of some important comparisons are for instance:

Table 5 Hand-crafted rules used in the deterministic baseline system

Rule	Trigger DA	Description
1. Lexical Rules		
R1	S	Occurrence of a word in the list: "bych": conditional form, first person singular "bychom": conditional form, first person plural "jsem": *"to be"*, first person singular "jsme": *"to be"*, first person plural "potřebuji", "potřebujeme" *"I need"*, *"we need"* "chci", "chceme", "chtěl", "chtěla", "chtěli" *"I want"*, *"we want"*, *"he wanted"*, *"she wanted"*, *"they wanted"* "znát", "vědět" and "doptat" *"to know" and "to ask"*
R2	Q[y/n]	The first word in the sentence is one of: "můžu", "můžete", "můžeme" *"can you"* in singular and in plural, *"can we"* "má", "máte", "máme" *"do you have"* in singular and in plural, *"do we have"*
R3	Q	"Wh*" word or a word from the list below is at the beginning of the sentence: "jak", "jakpak" and "kolik" *"how"* and *"how many"*
2. Morpho-syntactic Rules		
R4	O	Verb with imperative form at the beginning of the sentence
R5	O	Verb at the beginning of the sentence has a suffix amongst: "ej", "me", "te" Suffix of imperative form for 2nd person singular, 1st person plural, 2nd person plural
3. Syntactic Rules		
R6	O	The first word of the sentence is the syntactic root
R7	O	A verb doesn't have any subject
R8	Q[y/n]	Subject-verb inversion
4. Default Rule		
R9	S	When no previous rules apply, by default, the statement is chosen

Table 6 Dialogue act recognition accuracy for the baseline rule-based system with manual and automatic word transcription by jLASER recognizer

Transcription type	S	O	Q[y/n]	Q	Global
Manual	67.3	95.2	92.2	87.9	**83.5**
jLASER	67.5	88.8	85.1	82.0	**79.0**

- SyNB vs. B3NB: $p < 0.001$
- SyNB vs. B4NB: $p = 0.016$
- SyNB vs. BANB: $p = 0.002$

We can first observe that the Naive Bayes combination gives in general better results than majority voting, which was expected, as Naive Bayes ex-

Table 7 Dialogue act recognition accuracy for different features/approaches with manual word transcription

		Accuracy in [%]				
	Features/Approach	S	O	Q[y/n]	Q	Global
1. Unigram						
B0	Words	93.5	77.6	96.5	89.9	**91.0**
2. Majority Voting						
B1MV	Words	87.63	76.61	81.21	99.42	92.50
B2MV	Words + Lemmas	87.63	76.61	82.27	99.42	92.82
B3MV	Words + POS-tags	87.63	72.58	81.21	99.50	92.50
B4MV	Words + Pronoun at the beginning	90.28	76.61	95.39	99.33	95.17
B5MV	Words + Verb at the beginning	87.63	79.03	95.04	99.42	94.61
BAMV	Words + All baseline features	88.87	84.68	94.68	99.42	95.21
S1MV	Words + Dependency labels	87.81	74.19	88.30	99.42	93.51
S2MV	Words + Root position	89.05	87.10	92.20	99.33	95.03
S3MV	Words + Unexpressed subject	89.22	78.23	84.75	99.25	93.55
S4MV	Words + Basic composite pair	88.34	78.23	84.75	99.33	93.37
SyMV	All features	89.93	92.74	94.68	99.42	95.95
3. Naive Bayes Classifier						
B1NB	Words	94.52	71.77	83.33	99.25	94.38
B2NB	Words + Lemmas	94.88	87.90	91.13	99.42	96.50
B3NB	Words + POS-tags	95.05	80.65	88.30	99.00	95.53
B4NB	Words + Pronoun at the beginning	96.47	83.06	94.68	99.33	97.05
B5NB	Words + Verb at the beginning	88.87	48.39	94.33	99.00	92.86
BANB	Words + All baseline features	92.05	86.29	94.68	99.5	96.18
S1NB	Words + Dependency labels	94.52	81.45	89.01	99.00	95.53
S2NB	Words + Root position	90.11	85.48	93.62	99.00	95.21
S3NB	Words + Unexpressed subject	93.11	85.48	85.82	99.08	95.03
S4NB	Words + Basic composite pair	93.11	85.48	86.52	99.08	95.12
SyNB	All features	95.23	95.16	96.81	99.33	97.70

ploits the posteriors, which are a richer source of information than just the knowledge of the winning class.

This table also shows relatively low recognition scores for the class O. This is probably due to the relatively smaller amount of training data for this class. This analysis is supported by the good recognition accuracy obtained by the baseline rule-based system for this class, which does not depend on any training corpus. The best recognition rate is for the class Q, which is both the most frequent class and which is characterized by strong cues, especially concerning the influence of the first word in the sentence (B4NB) as well as distinctive interrogative word forms (B1NB, B2NB).

The most important remark is that the combination of all proposed syntactic and baseline features significantly outperforms all baseline features, which confirms that the proposed syntactic features bring complementary information. This result supports our claim that structured syntactic information might prove useful for dialogue act recognition.

5.4.2 Automatic transcription evaluation

Table 8 shows a similar evaluation to the one in Table 7, except that the textual transcriptions are now obtained automatically with the jLASER speech recogniser. Sentence recognition accuracy is 39.8% and word recognition accuracy is 83.4%. The complete annotation process starts from these imperfect transcriptions, including: lemmatization, POS-tagging, parsing and dialogue act recognition. This experiment thus assess the robustness of the complete processing chain to speech recognition errors, in order to match as closely as possible the actual use of the proposed approach in realistic conditions.

Table 8 Dialogue act recognition accuracy for different features/approaches with automatic word transcription using jLASER speech recogniser

		Accuracy in [%]				
	Features/Approach	S	O	Q[y/n]	Q	Global
1. Unigram						
B0	Words	93.1	68.8	94.7	86.3	**88.2**
2. Majority Voting						
B1MV	Words	82.69	29.84	62.06	99.25	86.14
B2MV	Words + Lemmas	83.92	41.13	65.25	99.17	87.48
B3MV	Words + POS-tags	81.80	28.23	63.48	99.17	85.96
B4MV	Words + Pronoun at the beginning	86.04	54.03	80.50	98.75	90.52
B5MV	Words + Verb at the beginning	81.10	25.00	79.43	98.17	87.11
BAMV	Words + All baseline features	88.52	58.47	81.25	97.97	90.84
S1MV	Words + Dependency labels	80.04	33.06	70.57	99.25	86.74
S2MV	Words + Root position	83.39	50.00	79.08	98.33	89.18
S3MV	Words + Unexpressed subject	81.63	33.87	61.35	99.33	86.05
S4MV	Words + Basic composite pair	81.63	33.87	60.28	99.33	85.91
SyMV	All features	84.28	71.77	82.98	98.17	91.07
3. Naive Bayes Classifier						
B1NB	Words	88.16	29.84	79.79	99.17	89.83
B2NB	Words + Lemmas	90.81	33.87	82.27	99.33	91.16
B3NB	Words + POS-tags	89.93	30.65	80.85	99.08	90.42
B4NB	Words + Pronoun at the beginning	91.34	66.13	89.36	98.67	93.21
B5NB	Words + Verb at the beginning	82.51	29.03	82.62	97.83	87.94
BANB	Words + All baseline features	88.34	68.55	87.59	97.67	92.27
S1NB	Words + Dependency labels	87.99	33.06	81.56	99.25	90.24
S2NB	Words + Root position	84.81	69.35	83.33	88.42	91.25
S3NB	Words + Unexpressed subject	89.93	34.68	80.50	99.08	90.61
S4NB	Words + Basic composite pair	89.75	33.06	80.14	99.17	90.47
SyNB	All features	91.17	70.97	88.65	98.00	93.46

We can first observe that the impact of speech recognition errors is moderately large, but not dramatic and thus does not jeopardize the applicability of the proposed approach in real conditions. Hence, while the dialogue act classification errors increase by 30% with the unigram model, they increase by

113% with the baseline CRF B3NB, which was expected because the CRF exploits the correlation between successive words and tags, which may propagate errors amongst words. However, despite its lower robustness, the CRF model still performs better in absolute value than the unigram model. The increase in classification error of the syntactic-aware model is about 183%, which is due to the greater sensibility of the processing chain for this model. Indeed, speech recognition errors are known to have a large impact on POS-tagging and parsing performances. The derived syntactic features are thus also largely impacted by such errors. This also explains why the simple proposed baseline features, such as B4NB, are also the most robust ones.

6 Conclusions

This work extends our previous works that tended to demonstrate the importance of global structural information for dialogue act recognition by implicitly modelling local constraints with Conditional Random Fields and explicitly proposing global syntactic features derived from automatic parsing of the sentence. Regarding the efficiency of syntactic features for dialogue act recognition, we have provided a number of evidence to support our claim that syntactic information might be important for dialogue act recognition and that the main reason why they have not been widely used so far in this domain is due to (i) the difficulty to reliably parse speech and dialogues; (ii) the intrinsic complexity of the syntactic material as compared to the classical lexical and morphosyntactic tags; and (iii) the lack of robustness of parsers to speech recognition errors. This claim is based on a review of several companion works that show the importance of syntax for both dialogue act recognition and closely related domains such as punctuation generation. Second, we have proposed several simple as well as more complex syntactic features that are derived from a full deep parsing of the sentence and have shown that the use of such features indeed significantly improves the dialogue act classification performance on our Czech corpus. Finally, we have studied the robustness of the proposed system and have shown that, as expected, the most complex syntactic features are also the most sensitive to speech recognition errors.

Hence, given the evidence collected in this work, we conclude that syntax information might prove important for dialogue act recognition, as it has already been shown relevant for many other Natural Language Processing tasks. The main challenge that remains is to increase its robustness to speech recognition errors, but we expect this challenge to be soon overcome, thanks to the great progresses realized in the automatic parsing community in recent years.

Acknowledgment

This work has been partly supported by the European Regional Development Fund (ERDF), project “NTIS - New Technologies for Information Society”, European Centre of Excellence, CZ.1.05/1.1.00/02.0090. We would like also to thank Ms. Michala Beranová for some implementation work.

References

1. J. L. Austin, How to do Things with Words, Clarendon Press, Oxford, 1962.
2. J. R. Searle, Speech Acts: An essay in the philosophy of language, 1969.
3. R. J. D. Power, The organization of purposeful dialogues, Linguistics 17 107–152.
4. E. A. Schegloff, Sequencing in conversational openings, American Anthropologist 70 (1) (1968) 1075–1095.
5. H. Sacks, E. A. Schegloff, G. Jefferson, A simplest semantics for the organization of turn-taking in conversation, Language 50 (4) (1974) 696–735.
6. D. J. Litman, Plan recognition and discourse analysis: an integrated approach for understanding dialogues, Ph.D. thesis, Univ. of Rochester, Rochester, NY (1985).
7. H. A. Kautz, A formal theory of plan recognition, Tech. Rep. 215, Department of Computer Science, University of Rochester, NY (1987).
8. S. Carberry, Plan Recognition in Natural Language Dialogue, MIT Press, Cambridge, MA.
9. H. Bunt, Context and Dialogue Control, Think Quarterly 3 (1994) 19–31.
10. D. R. Traum, Speech acts for dialogue agents, in: M. Wooldridge, A. Rao (Eds.), Foundations and Theories of Rational Agents, Kluwer, Dordrecht, 1999, pp. 169–201.
11. A. Stolcke *et al.*, Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech, in: Computational Linguistics, Vol. 26, 2000, pp. 339–373.
12. P. N. Garner, S. R. Browning, R. K. Moore, R. J. Russel, A Theory of Word Frequencies and its Application to Dialogue Move Recognition, in: ICSLP'96, Vol. 3, Philadelphia, USA, 1996, pp. 1880–1883.
13. H. Wright, Automatic Utterance Type Detection Using Suprasegmental Features, in: ICSLP'98, Vol. 4, Sydney, Australia, 1998.
14. J. Alexandersson, N. Reithinger, E. Maier, Insights into the Dialogue Processing of VERBMOBIL, Tech. Rep. 191, Saarbrücken, Germany (1997).
15. A. Lavie, F. Pianesi, L. Levin, The NESPOLE! System for multilingual speech communication over the Internet, Audio, Speech, and Language Processing, IEEE Transactions on 14 (5) (2006) 1664–1673.
16. H. Blanchon, C. Boitet, Speech translation for French within the C-STAR II consortium and future perspectives, in: INTERSPEECH'00, 2000, pp. 412–417.
17. P. Král, C. Cerisara, J. Klečková, Combination of Classifiers for Automatic Recognition of Dialog Acts, in: Interspeech'2005, ISCA, Lisboa, Portugal, 2005, pp. 825–828.
18. P. Král, C. Cerisara, J. Klečková, Automatic Dialog Acts Recognition based on Sentence Structure, in: ICASSP'06, Toulouse, France, 2006, pp. 61–64.
19. P. Král, C. Cerisara, J. Klečková, Lexical Structure for Dialogue Act Recognition, Journal of Multimedia (JMM) 2 (3) (2007) 1–8.
20. J. Allen, M. Core, Draft of Damsl: Dialog Act Markup in Several Layers, in: `http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html`, 1997.
21. D. Jurafsky, E. Shriberg, D. Biasca, Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation (Coders Manual, Draft 13), Tech. Rep. 97-01, University of Colorado, Institute of Cognitive Science (1997).
22. R. Dhillon, B. S., H. Carvey, S. E., Meeting Recorder Project: Dialog Act Labeling Guide, Tech. Rep. TR-04-002, International Computer Science Institute (February 2004).
23. S. Jekat *et al.*, Dialogue Acts in VERBMOBIL, in: Verbmobil Report 65, 1995.
24. E. Shriberg, R. Bates, A. Stolcke, P. Taylor, D. Jurafsky, K. Ries, N. Coccaro, R. Martin, M. Meteer, C. Van Ess-Dykema, Language and Speech, Vol. 41 of Special Double Issue on Prosody and Conversation, 1998, Ch. Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?, pp. 439–487.
25. J. Orkin, D. Roy, Semi-automated dialogue act classification for situated social agents in games, in: Proc. of the Agents for Games And Simulations Workshop at the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Toronto, Canada, 2010.
26. S. Joty, G. Carenini, C.-Y. Lin, Unsupervised approaches for dialog act modeling of asynchronous conversations, in: Proc. IJCAI, Barcelona, Spain, 2011.

27. N. Crook, R. Granell, S. Pulman, Unsupervised classification of dialogue acts using a dirichlet process mixture model, in: Proc. of the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue (SIGDIAL), 2009, pp. 241–348.
28. V. Petukhova, H. Bunt, Incremental dialogue act understanding, in: Proc. of the 9th International Conference on Computational Semantics (IWCS-9), Oxford, 2011.
29. M. Zimmermann, A. Stolcke, E. Shriberg, Joint segmentation and Classification of Dialog Acts in Multiparty Meetings, in: ICASSP'06, Toulouse, France, 2006, pp. 581–584.
30. A. Dielmann, S. Renals, Recognition of dialogue acts in multiparty meetings using a switching dbn, IEEE trans. on Audio, Speech, and Language Processing 16 (7) (2008) 1303–1314.
31. S. Quarteroni, A. V. Ivanov, G. Riccardi, Simultaneous dialog act segmentation and classification from human-human spoken conversations, in: Proc. ICASSP, Prague, Czech Republic, 2011.
32. S. Keizer, A. R., A. Nijholt, Dialogue Act Recognition with Bayesian Networks for Dutch Dialogues, in: 3rd ACL/SIGdial Workshop on Discourse and Dialogue, Philadelphia, USA, 2002, pp. 88–94.
33. G. Ji, J. Bilmes, Dialog Act Tagging Using Graphical Models, in: Proc. ICASSP, Vol. 1, Philadelphia, USA, 2005, pp. 33–36.
34. P. Lendvai, A. van den Bosch, K. E., Machine Learning for Shallow Interpretation of User Utterances in Spoken Dialogue Systems, in: EACL-03 Workshop on Dialogue Systems: Interaction, Adaptation and Styles Management, Budapest, Hungary, 2003, pp. 69–78.
35. K. Samuel, S. Carberry, K. Vijay-Shanker, Dialogue Act Tagging with Transformation-Based Learning, in: 17th international conference on Computational linguistics, Vol. 2, Association for Computational Linguistics, Morristown, NJ, USA, Montreal, Quebec, Canada, 1998, pp. 1150–1156.
36. M. Mast *et al.*, Automatic Classification of Dialog Acts with Semantic Classification Trees and Polygrams, in: Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing, 1996, pp. 217–229.
37. L. Levin, C. Langley, A. Lavie, D. Gates, D. Wallace, K. Peterson, Domain Specific Speech Acts for Spoken Language Translation, in: 4th SIGdial Workshop on Discourse and Dialogue, Sapporo, Japan, 2003.
38. G. Tur, U. Guz, D. Hakkani-Tur, Model adaptation for dialogue act tagging, in: Proceedings of the IEEE Spoken Language Technology Workshop, 2006.
39. R. Serafin, B. Di Eugenio, LSA: Extending latent semantic analysis with features for dialogue act classification, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Spain, 2004.
40. J. Bilmes, Backoff model training using partially observed data: Application to dialog act tagging, Tech. Rep. UWEETR-2005-0008, Department of Electrical Engineering, University of Washington (Aug. 2005).
41. J. Ang, Y. Liu, E. Shriberg, Automatic dialog act segmentation and classification in multiparty meetings, in: Proc. ICASSP, Philadelphia, USA, 2005.
42. M. Jeong, G. G. Lee, Triangular-chain conditional random fields, IEEE trans. on Audio, Speech, and Language Processing 16 (7) (2008) 1287–1302.
43. N. Webb, Cue-based dialog act classification, Ph.D. thesis, Univ. of Sheffield (Mar. 2010).
44. D. Jurafsky *et al.*, Automatic Detection of Discourse Structure for Speech Recognition and Understanding, in: IEEE Workshop on Speech Recognition and Understanding, Santa Barbara, 1997.
45. R. Kompe, Prosody in Speech Understanding Systems, Springer-Verlag, 1997.
46. M. Mast, R. Kompe, S. Harbeck, A. Kiessling, H. Niemann, E. Nöth, E. G. Schukat-Talamazzini, V. Warnke., Dialog Act Classification with the Help of Prosody, in: ICSLP'96, Philadelphia, USA, 1996.
47. H. Wright, M. Poesio, S. Isard, Using High Level Dialogue Information for Dialogue Act Recognition using Prosodic Features, in: ESCA Workshop on Prosody and Dialogue, Eindhoven, Holland, 1999.
48. D. Verbree, R. Rienks, D. Heylen, Dialog-act tagging using smart feature selection; results on multiple corpora, in: The first International IEEE Workshop on Spoken Language Technology (SLT), Palm Beach, Aruba, 2006.

49. T. Andernach, A machine learning approach to the classification of dialogue utterances, Computing Research Repository, July 1996.
50. B. Di Eugenio, Z. Xie, R. Serafin, Dialogue act classification, higher order dialogue structure, and instance-based learning, Journal of Discourse and Dialogue Research 1 (2) (2010) 1–24.
51. T. Klüwer, H. Uszkoreit, F. Xu, Using syntactic and semantic based relations for dialogue act recognition, in: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10, Association for Computational Linguistics, Stroudsburg, PA, USA, 2010, pp. 570–578.
URL http://portal.acm.org/citation.cfm?id=1944566.1944631
52. K. Zhou, C. Zong, Dialog-act recognition using discourse and sentence structure information, in: Proceedings of the 2009 International Conference on Asian Language Processing, IALP '09, IEEE Computer Society, Washington, DC, USA, 2009, pp. 11–16.
53. C. Sporleder, A. Lascarides, Using automatically labelled examples to classify rhetorical relations: A critical assessment, Natural Language Engineering, 2008, 14 (3).
54. B. Favre, D. Hakkani-Tür, E. Shriberg, Syntactically-informed models for comma prediction, in: ICASSP'09, Taipei, Taiwan, 2009, pp. 4697–4700.
55. C. Cerisara, P. Král, C. Gardent, Commas recovery with syntactic features in French and in Czech, in: INTERSPEECH'11, Firenze, Italy, 2011, pp. 1413–1416.
56. Y. Guo, H. Wang, J. v. Genabith, A linguistically inspired statistical model for Chinese punctuation generation, ACM Transactions on Asian Language Information Processing 9 (2) (2010) 27.
57. J. Geertzen, Dialog act recognition and prediction, Ph.D. thesis, Univ. of Tilburg (Feb. 2009).
58. P. Král, J. Klečková, T. Pavelka, C. Cerisara, Sentence Structure for Dialog Act recognition in Czech, in: ICTTA'06, Damascus, Syria, 2006.
59. E. Hajičová, Dependency-Based Underlying-Structure Tagging of a Very Large Czech Corpus (2000) 57–78.
60. J. Hajič, A. Böhmová, E. Hajičová, B. Vidová-Hladká, The Prague Dependency Treebank: A Three-Level Annotation Scenario, in: A. Abeillé (Ed.), Treebanks: Building and Using Parsed Corpora, Amsterdam: Kluwer, 2000, pp. 103–127.
61. D. Jurafsky, J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics, second edition Edition, Prentice-Hall, 2009.
62. J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kbler, S. Marinov, E. Marsi, MaltParser: A language-independent system for data-driven dependency parsing, Natural Language Engineering 13 (2) (2007) 95–135.
63. J. D. Lafferty, A. McCallum, F. C. N. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in: Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001, pp. 282–289.
URL http://portal.acm.org/citation.cfm?id=645530.655813
64. S. Grau, E. Sanchis, M. J. Castro, D. Vilar, Dialogue Act Classification using a Bayesian Approach, in: 9th International Conference Speech and Computer (SPECOM'2004), Saint-Petersburg, Russia, 2004, pp. 495–499.
65. T. Pavelka, K. Ekštein, JLASER: An automatic speech recognizer written in Java, in: XII International Conference Speech and Computer (SPECOM'2007), Moscow, Russia, 2007, pp. 165–169.
66. L. Gillick, S. Cox, Some statistical issues in the comparison of speech recognition algorithms, in: ICASSP'1989, 1989, pp. 532–535.