A Methodology for Dialogue Act Modeling from Part-of-Speech Annotations of Practical Dialogues in Mexican Spanish

Sergio R. Coria[1] and Luis A. Pineda[2]

[1] Software Engineering Group, University of the Sierra Sur, Calle Guillermo Rojas Mijangos S/N, Col. Ciudad Universitaria
70800 Miahuatlan de P. Diaz, Oax., Mexico
coria@unsis.edu.mx
[2] Department of Computer Science, National Autonomous University of Mexico, Circuito Escolar S/N, Ciudad Universitaria
04510 Mexico, D.F.
luis@leibniz.iimas.unam.mx

Abstract. This article proposes a methodology to model dialogue acts (DA) from part-of-speech (POS) annotations in practical dialogues in Mexican Spanish. An example using empirical data is presented to illustrate this methodology. DA and POS annotations were obtained from DIME, a spoken dialogue corpus in which speakers perform tasks in a virtual environment on a collaborative basis. POS structure is a key notion in this methodology. It is the POS tagging sequence for particular instances of utterances. The models are produced as probability tables of pairs of DA annotations and POS structures. Results show that a number of POS structures are more frequently used by speakers to communicate particular DA types. These patterns can be exploited to enhance systems for automatic recognition of DA.

Keywords: spoken dialogue, dialogue act modeling, DAMSL, DIME-DAMSL, part of speech.

1 Introduction

Practical dialogue, also known as task-oriented conversation, is that in which speakers cooperate to achieve a common goal. This type of conversation is simpler than general conversation because the number and complexity of its dialogue act (DA) types are less. In addition, it is the typical conversation to be performed in automatic systems for dialogue management. An advantage for computational purposes is that its analysis and modeling can be better controlled on experimental conditions.

Special Issue in Advances in Pattern Recognition
Research in Computing Science 44, 2009, pp. 69-76

The relation between DA type and POS in Spanish has been poorly investigated from an empirical, computational, view. Therefore, this research aims to propose a methodology to characterize its patterns in spoken dialogue corpora. Other motivation for this methodology is to improve the performance of dialogue management systems by taking advantage of a number of information sources from speech in addition to lexical content.

This paper is organized as follows: section 2 describes a previous research in the area. Section 3 comments theoretical and computational foundations of DA from the views of DAMSL and DIME-DAMSL annotation schemes. Section 4 addresses part of speech (POS) annotation. Section 5 describes a methodology for modeling DA from POS on an empirical, statistical, basis; an example using data from the DIME corpus is explained. Section 6 presents an overview of the DIME corpus and the data sample for the example. Section 7 presents the example results. Finally, section 8 discusses the results and suggests future research.

2 Previous Research in the Area

Instances of previous research with a similar approach are: [1], [2], [3], [4] and [5]. Most of them address the problem for English specifically. Specific research for Spanish is scarce, so this is a strong motivation for addressing the issue. [1] and others address the problem on subsets of DA types only, *e.g.* [3] studies the specific domain of business-appointment scheduling. Others use high complexity methods, such as Hidden Markov Models. [2] uses statistical language models produced from word transcription or automatic recognition. [4] presents a supervised adaptation method for dialog act tagging, evaluating model adaptation for dialog act tagging by using out-of-domain data or models. [5] explores the tasks of dialogue act segmentation and classification by employing simple lexical and prosodic knowledge sources. [6] addresses the problem from the perspective of intonational information in Spanish and it uses the same empirical resource that the present research uses. Unlike most of previous work, the present research proposes a methodology that creates models as probability tables on POS sequences and DA annotation pairs.

3 Dialogue Act

Searle's speech act is the production or emission of an utterance-instance under certain conditions, and speech acts are the basic or minimum units of linguistic communication. On this basis, DA is the characterization of a speech act within the context of a task-oriented conversation. From a computational view, DA need to be represented so that they can be analyzed within experimental conditions. Therefore, a number of DA

representation schemes, construed by annotation tag sets and annotation rules, have been proposed. A widely used scheme is DAMSL [7]. This research uses DIME-DAMSL [8], a scheme that is based on DAMSL and that provides with high inter-annotator consistency. DIME-DAMSL identifies two communication planes in DA: *obligations* and *common ground*. Tables 1 and 2 present the tag set of DIME-DAMSL scheme. A compound tag involves that an utterance communicates two or more DA simultaneously.

Table 1. Simple and compound tags for annotation of obligations DA (from DIME-DAMSL).

info-request	*answer*	*no-tag*	*action-dir*
commit	*info-request answer*	*action-dir answer*	*offer*
info-request offer	*action-dir offer*		

Obligations DA are those in which an obligation to perform an action or to provide some piece of information is generated either on the listener or on the speaker.

Table 2. Simple and compound tags for annotation of common ground DA (from DIME-DAMSL).

no-tag	*accept*	*affirm*	*hold repeat-rephr*
open-option	*accept-part*	*reaffirm*	*hold*
reject	*affirm accept*	*ack*	*NUS(non-understanding signal)*
offer conv-open	*repeat-rephr*	*reject-part*	*conv-close*
affirm-reject	*open-option accept*	*offer*	*accept hold repeat-rephr*
affirm maybe	*affirm hold*	*offer accept*	*affirm conv-close*
affirm accept exclam	*affirm accept-part exclam*	*affirm correct*	*affirm perform conv-close*
hold NUS	*open-option reject*	*perform*	*reaffirm complement*
reaffirm hold	*hold in task-management*	*other*	

Common ground DA are those in which shared knowledge or beliefs are established or re-established (*agreement* subplane). Also, these DA allow the dialogue participants to manage the communication channel (*understanding* subplane).

Table 3. Tags for POS annotation (reproduced from [9]).

Tag	Description	Tag	Description	Tag	Description	Tag	Description
A	Adjective	*AD*	Demonstrative adjective	*C*	Conjunction	*N*	Noun
P	Pronoun	*PC*	Clitic pronoun	*PI*	Interrogative pronoun	*PR*	Relative pronoun
R	Adverb	*RA*	Acceptation adverb	*RI*	Interrogative adverb	*RN*	Negation adverb
RR	Relative adverb	*S*	Preposition	*TD*	Definite article	*TI*	Indefinite article
V	Verb	*VAM*	Modal auxiliary verb	*VC*	Verb with clitic		

4 Part of Speech (POS)

Part of speech (POS) is the characterization of a word according to the function that it performs in an utterance. A POS annotation process is based on a POS tag set and annotation rules that guide a human or automatic annotator to assign a label to a particular

instance of a word in an utterance. A number of POS tag sets and annotation rules exist. This research uses [9], which is reproduced in Table 3.

Table 4. POS structures and their frequencies in the dataset.

POS structure	Frequency	% (threshold=0.6%)	Accum. %
RA	170	20.0	20.0
R_V_R	115	13.5	33.5
R	20	2.4	35.8
PD	14	1.6	37.5
RN	13	1.5	39.0
RI_V_C_PC_V	9	1.1	40.1
V_TI_N	7	0.8	40.9
RI_V_R	6	0.7	41.6
Other structures (each <= 0.6%)	497	58.4	100.0
Total	851		

5 Methodology

The methodology proposed by this research to modeling DA from POS annotations is based on correlation analyses of data sets from both DA and POS annotations in a spoken dialogue corpus. Models are construed on a Pareto analysis basis and probability tables and they can also be represented as classification trees.

POS structure is the basic notion in this methodology. POS structure is the complete sequence of POS labels that are assigned to words in a particular utterance. The whole set of POS structures and the probabilities of DA types for each POS structure in a corpus are a simplified statistical language model of DA types obtained from that specific corpus. Once POS annotations of the corpus are available, the basic steps to perform are:

Step 1. Creation of POS structures from POS label sequences: this involves concatenating the POS labels of each utterance by inserting a character such as _ (*underscore*) or any other similar to produce one single string. For instance, a POS sequence as *R, V, R* is transformed into *R_V_R*.

Step 2. Pareto analysis of POS structures: its purpose is to identify the most frequent POS structures (see Table 4). This is needed to determine a threshold for discarding POS structures with lowest frequencies. These structures might be useless for DA modeling because they are associated to syntactic structures that are scarcely used by speakers. The threshold can be determined by an empirical fine-tuning for specific applications. Every POS structure with a relative frequency less than the threshold cannot be recognized by the model; however, this behavior is useful because the recognition task can be focused on the most frequent structures.

Step 3. Statistical analyses of DA for each of the most frequent POS structures for obligations and for common ground tags (see Tables 5 and 6). Two separated analyses (*i.e.* one for obligations and one for common ground) are needed because DIME-DAMSL scheme assigns both an obligations and a common ground tag for every single utterance.

In a real-world application, DA recognition can be supported by mapping a particular POS structure to its most probable DA types on obligations and common ground. Therefore, statistical analysis of these patterns is the core of a model. A threshold for relative frequencies of DA tags that are associated to every POS structure is convenient to disregard DA types that are not statistically significant in the dataset.

Table 5. Obligations tags of the most frequent POS structures.

POS structure	% of dataset (threshold=0.6)	Obligations tag	% of POS structure (threshold=30.0)
RA	20.0	*answer*	52.4
		Other	47.6
R_V_R	13.5	*info-request*	82.6
		Other	17.4
R	2.4	*no-tag*	40.0
		answer	35.0
		Other	25.0
PD	1.6	*info-request*	71.4
		Other	28.6
RN	1.5	*answer*	53.8
		no-tag	30.8
		Other	15.4
RI V C PC V	1.1	*info-request*	100.0
V_TI_N	0.8	*no-tag*	42.9
		Other	57.1
RI V R	0.7	*info-request*	100.0
Other structures (each <= 0.6%)	58.4		

Step 4. Statistical analyses of the most frequent POS structures for each DA type. Like analyses in step 3, two analyses (one for obligations and one for common ground tags) are produced. The purpose is to identify the POS structures that occur most frequently for each DA type. Usually, only a lower number of pairs of POS structure and DA type with high percent for the pair exist in a corpus. The identification of these patterns is useful to validate results from steps 2 and 3. See Table 7 for obligations. Also, a table for common ground is available at the WWW[1].

Step 5. Implementation of two DA recognition models (one for obligations and one for common ground). Models can be implemented in a real-world dialogue management system considering the following stages: 1) word-level automatic speech recognition that parses words into POS, 2) POS tagging can then be transformed into POS structures, and 3) the DA tag with highest probability is assigned for the POS structure using the tables previously computed. One tag is assigned for obligations and one for common ground. Thresholds defined in steps 2, 3 and 4 involve that the recognition task is focused on the most frequent DA types and POS structures.

This methodology is illustrated by an example on a practical dialogue corpus that is explained below.

[1] http://www.unsis.edu.mx/~coria/mwpr_2009/table_7_A.pdf

Table 6. Common ground tags of the most frequent POS structures.

POS structure	% of dataset (threshold=0.6)	Common ground tag	% of POS structure (threshold=30.0)
RA	20.0	*accept*	94.1
		Other	5.9
R_V_R	13.5	*no-tag*	86.1
		Other	6.2
R	2.4	*accept*	40.0
		Other	60.0
PD	1.6	*hold_repeat-rephr*	64.3
		Other	35.7
RN	1.5	*reject*	46.2
		Other	53.8
RI_V_C_PC_V	1.1	*no-tag*	55.6
		accept	33.3
		Other	11.1
V_TI_N	0.8	*no-tag*	42.9
		Other	57.1
RI V R	0.7	****	****
Other structures (each <= 0.6%)	58.4		

6 Empirical Resource

The DIME Corpus [10] is the empirical resource used in this research. An empirical approach is preferable for DA modeling because it increases the model efficiency and allows creating probabilistic, instead of deterministic, models. The DIME corpus contains 26 screen videos and audios of two-person practical dialogues. In each dialogue one person plays the role of the computer *System* and his partner acts as the system *User*. Videos contain the graphical status of a virtual scenario and graphical actions performed by individuals on virtual pieces of furniture as they interact in a CAD (computer aided design) software. The common goal for participants in every dialogue is to arrange pieces of furniture in a virtual kitchen while satisfying a number of design specifications and constraints. The corpus also contains orthographic, phonetic, POS and DA transcriptions, among others. A sample containing 851 utterances from 12 dialogues is used for this research. The sample includes the orthographic, POS and DA annotations of utterances.

7 Example Results

Results from an example to illustrate the methodology using data from the DIME corpus are presented in Tables 4 to 7.

Data in Table 5 are ordered by POS structure percent (second column). Percents of DA tags are ordered for each POS structure. A useful alternate ordering can be by percent of

majority class of POS structure; *i.e.* the pairs *RI_V_C_PC_V* with *info-request* and *RI_V_R* with *info-request* would be the top 2, each with 100.0%, then *R_V_R* with *info-request* with 82.6% would be third, etc. This way, the most significant patterns are easily identified. Table 6 is similar, but its values for *RI_V_R* do not satisfy the threshold. The complete models, *i.e.* with no thresholds, are available at the WWW[2].

Table 7. Most frequent POS structures for obligations DA types.

Obligations tag (single or compound)	Frequency	%	Most frequent POS structure	% of obligs. tag (threshold=35.0)
info-request	262	30.8	*R_V_R*	36.3
			Other	63.7
answer	213	25.0	*RA*	41.8
			Other	58.2
commit	34	4.0	*RA*	91.2
			Other	8.8
info-request_answer	16	1.9	*R_V_R*	43.8
			Other	56.2

8 Discussion and Future Research

A methodology for DA modeling from POS annotations has been presented. POS structure, *i.e.* the sequence of POS tags from a particular utterance, is an important notion in this methodology. Results from an empirical analysis show that a number of POS structures are more frequently used by speakers to communicate specific DA types in practical dialogues. Some examples of the most frequent pairs are: *RA* for *commits*, *R_V_R* for *info-requests*, *RN* for *reject-parts*, etc.

Defining thresholds for the most frequent DA types and/or POS structures is a solution to identify the most significant patterns. Therefore, this approach can focus the recognition task on the DA types and/or the POS structures that have highest frequencies in a training corpus.

The main contributions and significance of this research are: 1) results suggest that a number of DA types are uttered using specific subsets of POS structures more frequently than other structures and these patterns could be exploited to improve DA managements systems, 2) the methodology is simple and it can be used on other languages for theoretical and practical purposes, 3) although patterns that are discovered from a specific corpus cannot be used to improve general-purpose DA management systems, the methodology can be useful on a domain-specific basis. Implementations for real-world systems should use other additional sources from speech, such as intonation, cue words, etc. to obtain higher accuracy rates, since only using POS structure of utterances does not suffice.

[2] See obligations and common ground models at http://www.unsis.edu.mx/~coria/mwpr_2009/

Future research can address applying the methodology to implement and evaluate an automatic recognition model for DA using standard evaluation metrics. Also, comparing results to similar systems is needed. Implementation can be based on probability tables for the pair patterns. In a later implementation, other models can include POS structure information along with intonation and other sources. The models can be implemented in decision trees, such as those created with classification and regression algorithms.

Acknowledgments. The authors thank the DIME Project team for their effort in producing POS and DA annotations of the DIME Corpus.

References

1. Jurafsky, D., Shriberg, E., Fox, B., Curl, T., 1998. Lexical, prosodic, and syntactic cues for dialog acts. In: Procs. of the ACL/COLING Workshop on Discourse Relations and Discourse Markers, Montreal, Canada, August 1998, pp. 114--120.
2. Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., Van EssDykema, C., 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? USA, Language and Speech 41(3–4) 439-487 (Special Issue on Prosody and Conversation).
3. Jekat, S., Klein, A., Maier E., Maleck, I., Mast, M., Quantz, J. J., 1995. Dialogue Acts in VERBMOBIL, VM-Report 65.
4. Tur, G., Guz, U., Hakkani-Tür, D., 2006. Model adaptation for dialog act tagging. In: Proceedings of SLT 2006, 1st biannual IEEE/ACL Workshop on Spoken Language Technologies, Aruba, December 2006.
5. Ang, J., Liu, Y., Shriberg, E., 2005. Automatic dialog act segmentation and classification in multiparty meetings. In: Proceedings of ICASSP.
6. Coria, S.R., Pineda, L.A. An Analysis of Prosodic Information for the Recognition of Dialogue acts in a Multimodal Corpus in Mexican Spanish. Computer Speech and Language, Vol. 23 (2009), pp. 277--310, Elsevier.
7. Allen, J.F., Core, M., 1997. Draft of DAMSL: Dialogue Act Markup in Several Layers. Technical Report, The Multiparty Discourse Group. University of Rochester, Rochester, USA.
8. Pineda, L.A., Estrada, V.M., Coria, S.R., 2006. The obligations and common ground structures of task oriented conversations. In: Proceedings of the Fourth Workshop in Information and Language Technology TIL-2006, in IBERAMIA 06, Brazil.
9. Moreno, I., Pineda, L., Speech Repairs in the DIME Corpus, Research in Computing Science, Vol. 20, pp. 63 – 74, 2006.
10. Villaseñor, L., Massé, A., Pineda, L. The DIME Corpus, Memorias 3er. Encuentro Internacional de Ciencias de la Computación ENC01, Tomo II, C. Zozaya, M. Mejía, P. Noriega y A. Sánchez (eds.), SMCC, Aguascalientes, Ags. México, Septiembre, 2001.