Prediction of Dialogue Acts on the Basis of the Previous Act

Sergio R. Coria coria@uxmcc2.iimas.unam.mx
Luis A. Pineda luis@leibniz.iimas.unam.mx
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS)
Universidad Nacional Autónoma de México (UNAM)
Ciudad Universitaria, Coyoacán, México, D.F.

Resumen: En este trabajo se evalúa empíricamente el reconocimiento automático de actos de diálogo. Se usan datos provenientes de un corpus de diálogos con habla espontánea. En cada diálogo dos hablantes colaboran en el diseño de cocinas usando herramientas C.A.D.; uno de ellos desempeña el rol del Sistema y el otro el del Usuario. Los actos de diálogo se etiquetan con DIME-DAMSL, esquema que considera dos planos de expresión: obligaciones y common ground. La evaluación se realiza probando modelos clasificadores creados con algoritmos de aprendizaje máquina: uno para obligaciones y otro para common ground. El principal dato predictor analizado es el acto de diálogo correspondiente al enunciado inmediato anterior. Se pondera también la contribución de información adicional, como la entonación, etiquetada con INTSINT, la modalidad del enunciado, el rol del hablante y el tipo de acto de diálogo del plano complementario. Una aplicación práctica sería en sistemas de administración de diálogo.
Palabras clave: Diálogos prácticos, acto de diálogo, DIME-DAMSL, aprendizaje máquina, entonación, INTSINT, corpus de diálogo, árbol de clasificación y regresión

Abstract: In this paper the automatic recognition of dialogue acts is evaluated on an empirical basis. Data from a dialogue corpus with spontaneous speech are used. In each dialogue two speakers collaborate to design a kitchen using a C.A.D. software tool; one of them plays the System's role and the other plays the User's role. Dialogue acts are annotated with DIME-DAMSL, a scheme considering two expression planes: obligations and common ground. The evaluation is performed by testing classification models created with Machine Learning algorithms: one model for obligations and other for common ground. The mainly analyzed predictor data is the dialogue act corresponding to the immediately previous utterance. The contribution of other information sources is also evaluated, such as intonation, annotated with INTSINT, utterance mood, speaker role and dialogue act type of the complementary expression plane. A practical application can be the implementation of dialogue management systems.
Keywords: Practical dialogues, dialogue act, DIME-DAMSL, machine learning, intonation, INTSINT, dialogue corpus, classification and regression tree

Introduction

Automatic recognition of dialogue acts has been addressed in previous work, such as (Shriberg *et al.*, 1998) and the VERBMOBIL Project (Wahlster, 1993); it is a relevant issue because it provides speech recognition and dialogue management systems with additional information, which tends to improve their accuracy and efficiency. These two pieces of work have used intonational and lexical information to perform the dialogue act recognition for English and German languages, respectively. Another relevant reference is (Garrido, 1996), where the relation between intonation and utterance mood in Spanish is addressed.

In (Coria and Pineda, 2006) dialogue act in Spanish is addressed from an intonational view and also considering some other non-prosodic features; these experimental settings are immediate predecessors of the present work.

Machine learning algorithms, such as classification trees and neural networks, in addition to language models and polygrams are commonly used to analyze the phenomenon and to find out the most contributing features for the implementation of recognition or prediction models. This work uses a classification tree algorithm to evaluate the contribution of the previous dialogue act to the prediction task, assuming as baseline a recognition setting where the previous act is not used as one of the predictors.

A key issue in dialogue act recognition is the annotation of dialogue acts. The present work adopts the DIME-DAMSL scheme for this annotation.

1 Dialogue acts and the DIME-DAMSL scheme

1.1 Speech acts and dialogue acts

Searle's theory on speech acts states that the production or emission of an utterance-instance under certain conditions constitutes a speech act, and speech acts are the basic or minimal units of linguistic communication. The dialogue act is an adaptation of the this notion and involves a speech act in the context of a dialogue (Bunt, 1994) or an act with internal structure specifically related to its dialogue function, as assumed in (Allen and Core, 1997), or a combination of the speech act and the semantic force of an utterance (Bunt, 1995). The present work is based on Allen and Core's view.

1.2 DAMSL scheme

Allen and Core define a tag set and a series of tagging principles in order to produce a computational scheme for the annotation of dialogue acts in a particular class of dialogues: the so-called *practical dialogues*, where the interlocutors collaborate to achieve a common goal and do not need to use a too complex language because the conversation is simpler than the general conversation.

The DAMSL scheme defines four tag sets for utterance annotation, as follows: *communicative status*, *information level*, *forward-looking* and *backward-looking functions*. One of the main purposes of the *communicative status* is to specify if an utterance is intelligible or not; the *information level* describes the general subject of the utterance, e.g. task, task-management, communication management.

The forward looking functions resemble diverse categories defined in the traditional speech acts theory; e.g. action directives, commitments or affirms in DAMSL resemble directives, commisives or representatives, respectively, in Searle's scheme.

The backward-looking functions specify how an utterance is related to the ones preceding it in the dialogue; e.g. to accept a proposal, to confirm understanding of a previous utterance, to answer a question.

1.3 DIME-DAMSL scheme

As DAMSL scheme did not suffice to obtain a high enough inter-annotator agreement, it was not reliable enough to set machine-learning experiments, which require consistent information. A source of low agreement in DAMSL is the lack of a higher level structure to constraint the possible label(s) an utterance can be assigned to; i.e. the scope of DAMSL scheme is restricted to analyze single utterances without considering the context within the dialogue where previous or following utterances occur. This allows a broad space to select and combine labels but, on the other hand, there is a high risk that inter-annotator agreement for dialogue act types is low because of the influence of subjectivity.

Evolving from DAMSL, DIME-DAMSL adopts its tag set and its dimensions and extends them by defining three additional notions, as follows. 1) two expression planes: the obligations and the common ground, 2) transaction structure and 3) charge and credit contributions of dialogue acts in balanced transactions.

The obligations and the common ground planes are parallel structures along which dialogue acts flow. A dialogue act might contribute to any (or both) of the two planes.

In DIME-DAMSL the obligations plane is construed by dialogue acts that generate a responsibility either on the speaker himself or on the listener to perform an action, either verbal or non-verbal; e.g. the obligation to provide some piece of information or to perform a non-verbal action. Dialogue acts that mainly contribute to the obligations plane are: commit, offer (when it is accepted by the interlocutor), action directive and information request. For instance, in utterances from dialogues of the DIME corpus, *okay* is a commit (in certain contexts)*; can you move the stove to the left?* is an action directive, and *where do you want me to put it?* is an information request.

The common ground is the set of dialogue acts that add, reinforce and repair the shared knowledge and beliefs of the interlocutors and preserve and repair the communication flow. This notion differs from Clark and Schaefer's grounding model. DIME-DAMSL defines two sub-planes in the common ground: *agreement* and *understanding*; agreement is the set of dialogue acts that add knowledge or beliefs to be shared on the grounding of the dialogue participants; understanding is defined by acts that keep, reinforce or recreate the communication channel. Dialogue acts that mainly contribute to the agreement sub-plane are: *open option* (e.g. *these are the cupboards we have*), *affirm* (e.g. *because I need a cabinet*), *hold* (e.g. *do you want me to move this cabinet to here?*), *accept* (e.g. *yes*), *reject* (e.g. *no, there is no design problem*), *accept part*, *reject part* and *maybe*. Dialogue acts on the understanding sub-plane are *acknowledgment* (e.g. *yeah*, *yes, okay*, etc.), *repeat-or-rephrase* (e.g. *do you want me to put this stove here?*), and *backchannel* (e.g. *mhum*, *okay*, *yes*, etc.).

Charges and credits are the basic mechanism underlying the interaction between pairs of dialogue acts along each of the two expression planes. A charge generated by a dialogue act introduces an imbalance requesting for satisfaction, and a credit is the item balancing that charge. Instances of balanced pairs are, on the obligations plane, action directive, a charge, which can be balanced with a graphical action; on the agreement plane a charge introduced by an open option can be balanced with an accept; on the understanding plane an affirm creates a charge that can be satisfied with an acknowledgment, etc. These and other additional pairs guide a charge-credit annotation to identify and annotate the most prominent dialogue acts of the utterance; this annotation of dialogue acts is called *Preliminary DIME-DAMSL* and supports the completion of the dialogue act tagging in a subsequent stage, the so-called *Detailed DIME-DAMSL*, where the annotation is added with other labels if necessary.

A transaction is defined by a set of consecutive charge-credit pairs intending a sub-goal within a dialogue. A transaction presents two general phases: intention specification and intention satisfaction. In turn, the intention specification is divided into two subphases: intention specification and intention interpretation and the intention satisfaction is divided into intention satisfaction and action interpretation.

2 *The DIME Corpus*

The DIME Corpus (Pineda, 2007) is the empirical information source to perform the experiments; it is a collection of 26 human-to-human dialogues with their corresponding video and audio recordings and their annotations on a series of levels. It was created to analyze phonetic, phonologic and dialogue phenomena in Mexican Spanish. Speakers are approximately 15 individuals, males and females, most of them from Mexico City with ages between 22 and 30 y/o.

In each dialogue two speakers collaborate to design a kitchen using a C.A.D. software; one of them plays the *System's* role and the other plays the *User's* role. The System is always the same speaker in all dialogues. The speakers perform a task that consists in placing pieces of furniture in a virtual kitchen as specified by a drawing on a piece of paper.

Every User interacts with the System using the C.A.D. tool. The User commands the System to design the virtual kitchen. There is no written script, so the language spoken in the dialogue is spontaneous.

2.1 Annotation levels

The DIME corpus is segmented into utterances and annotated on these levels: orthographic transcription (transliteration), allophones, phonemes, phonetic syllables (considering the possible presence of re-syllabication), words, break indices from Sp-Tobi (Beckman *et al.*, 2002), parts of speech (P.O.S.), discourse markers, speech repairs, intonation and utterance mood. The MexBet phonetic alphabet (Cuétara, 2004) is used to annotate allophones, phonemes, phonetic syllables and words.

2.1.1 Intonational annotation

Intonation is annotated with INTSINT (Hirst, Di Cristo and Espesser, 2000), implemented in the M.E.S. tool (Motif

Environment for Speech). A stylized contour of the fundamental frequency is automatically obtained and its inflection points are detected, saving their respective frequency (*Hz*) and timestamp. A perceptive verification is performed by a human annotator in order to assure that the stylized contour is perceptively similar to the original speech signal; the inflection points can be relocated on the frequency or time axis by the annotator. Every inflection point is then automatically annotated with the INTSINT tag set according to the relative location of the point regarding its predecessor and its successor. The tag set is construed of 3 absolute tones: *T* (*top*, the absolute highest), *B* (*bottom*, the absolute lowest), and *M* (*medium*, the frequency average); and 5 iterative tones: *H* (*higher*, a local maximal), *L* (*lower*, a local minimal), *U* (*up-step*, a point on an ascending region), *D* (*down-step*, a point on a descending region), *S* (*same*, a point at the same height than its predecessor). Absolute tones can occur only once along an intonational contour; i.e. *T*, *B* and *M* appear usually one single time in the intonational annotation of an utterance. On the other hand, iterative tones can appear an arbitrary number of times.

The original INTSINT tags and timestamps produced with M.E.S. are transformed into tag concatenations without timestamps in order to generate simple strings. This representation without time information provides with a higher level abstraction and allows compare intonational contours from different speakers without requiring a normalization process, as it is required when using a numerical representation. This way, the initial or final regions of a contour can be represented by sequences of the first or the last INTSINT tags of a string.

2.1.2 Utterance mood annotation

Utterance mood, i.e. *interrogative*, *declarative*, *imperative*, etc. is annotated as specified by a series of formalized conventions; some of which are as follows:

The human annotator reads the orthographical transcription and listens to the audio file, focusing on the final region of the utterance.

The tag set is: *dec* (declarative), *imp* (imperative), *int* (interrogative) and *other*. The *other* label includes any other mood that does not fit into the first three categories. It is also used in any of the following cases: the end of the utterance is too noisy, the end presents a too long silence whose duration is greater than the one of a pause, the utterance does not contain lexical information but instead a sound such as breathing, laughing, lip-clicks, etc.

As one single annotator performs this tagging, annotation agreement is not computed.

A machine-learning algorithm is used to create a model for automatic annotation of utterance mood by using the manual tagging as target data. The automatic annotation is later used as one of the inputs for dialogue act recognition because this would be the case in a real-world application.

3 Experimental settings and information features

The setting is implemented as a machine learning experiment, selecting a subset of the features as targets and others as predictors. Table 1 presents a data dictionary of the features involved in the prediction models for obligations and common ground dialogue acts. Its right-most column specifies if a feature is used as either predictor (*P*) or target (*T*); the *T/P* value specifies that the feature is used as target in a particular model and as predictor in other. Lexical information is not used in the predictor feature set. The *last_2* feature is based on the toneme notion (Navarro-Tomas, 1974).

Two recognition models are produced: one for obligations and other for common ground. The previous dialogue act refers to both *obligations_minus1* and *commgr_minus1* features; i.e. both features are evaluated as predictors for obligations and also for common ground.

The machine learning algorithm to generate the models is J48 (Witten and Frank, 2000); it creates classification and regression trees using an approach similar to CART (Breiman *et al.*, 1983). J48 is implemented in WEKA (Witten and Frank, 2000), a free software tool.

The dataset for the experiment contains features corresponding to 1,043 utterances in 12 dialogues from the DIME corpus.

Baselines to evaluate the results are determined by an experimental setting where the previous dialogue act is not used as one of the predictors. These are: optimal predicted

Feature	Description	Why it is Used	P or T
first_1	The first INTSINT label of an utterance	The initial region of the intonational contour contributes to utterance mood recognition; each of the three features is evaluated	P
first_2	The first two INTSINT labels of an utterance		
first_3	The first three INTSINT labels of an utterance		
last_2	The last 2 INTSINT labels of an utterance	Preliminary experiments show that it is highly contributive to utterance mood recognition because it contains the utterance toneme	P
optimal_pred_mood	Utterance mood (e.g. declarative, interrogative, imperative) is obtained by an automatic recognition task prior to dialogue act recognition. Its predictors are: speaker role, utterance duration and the last 2 and the first 1, 2 and 3 INTSINT tags of the intonational contour.	Particular utterance moods are related to dialogue act types. An automatically recognized mood instead of the manually annotated is used because this is more similar to a real-world application	T/P
utt_duration	Utterance duration in milliseconds; it is *not* normalized	Preliminary experiments show that it might contribute to the recognition of dialogue act type	P
speaker_role	Role of the speaker in the dialogue, either *System* or *User*	Statistical analyses show that *speaker_role* is correlated to dialogue act; e.g. *System* and *commit*, *User* and *action directive*	P
obligations	Manually annotated tag for dialogue act on the obligations plane of an utterance	It is used as target data in the obligations recognition model and as one of the predictors for the common ground model	T/P
obligations_minus1	Dialogue act tag (manually annotated) of obligations in the utterance *n-1*, where *n* is the utterance whose dialogue act is the target	Its contribution as one of the predictors for dialogue act is evaluated	P
commgr	Manually annotated tag for dialogue act on the common ground plane of an utterance; agreement and understanding tags are concatenated as one single feature	It is used as target in the common ground recognition model and as one of the predictors in the obligations model	T/P
commgr_minus1	Dialogue act tag (manually annotated) of common ground in the utterance *n-1*, where *n* is the utterance whose dialogue act is the target	Its contribution as one of the predictors for dialogue act is evaluated	P

Table 1. Data dictionary of the features involved in the prediction models

mood, utterance duration (in milliseconds) and speaker role; besides, the obligations model uses common ground dialogue act and the common ground model uses the obligation dialogue act. Table 2 presents the baseline values, where accuracy is the percent of correctly classified instances and kappa, introduced by (Siegel and Castellan, 1988) and (Carletta, 1996), is a consistency measurement for manual (or automatic) tagging tasks. Number of labels, instances to be annotated and annotators determine a default agreement value that might artificially increase the actual inter-annotator agreement (or the model accuracy), so the default agreement value is computed and substracted.

Kappa in Table 4 and in the other machine-learning models is automatically computed by WEKA. Kappa of manual annotations, except of utterance mood, is computed by using Excel-style worksheets. Utterance mood was first manually annotated by one only human annotator and then automatic recognition models were produced using the manual tagging data as target.

	Acc. (%)	Kappa
Obligations	66.2500	0.58120
Comm. Ground	68.4564	0.55510

Table 2. Baseline values of recognition without the previous act

Dialogue act annotation was formatted and processed in order to manage utterances with more than one tag on any expression plane; e.g. if the tagging contains *affirm* and *accept*, involving that the utterance simultaneously affirms and accepts, then it is concatenated as *affirm_accept*. Other instances are: *info-request_graph-action* or *hold_repeat-rephrase*.

4 *Results and evaluation*

Two classification trees were produced: one for obligations, containing 155 rules and one for common ground, containing 151 rules. Each tree was generated and tested by the 10-fold cross validation method. The complete rule sets are available on demand.

Results in Table 3 show that accuracy and kappa of obligations recognition when using the previous dialogue act as one of the predictors are greater than their baselines: the improvement is +5.658 in accuracy and +0.0791 in kappa. Regarding common ground recognition, there is a marginal decreasing in accuracy (-0.1918) and a marginal improvement in kappa (+0.0409).

	Acc. (%)	Kappa
Obligations	71.9080	0.6603
Comm. Ground	68.2646	0.5960

Table 3. Accuracies and kappas of recognition models

Confidence and *support* values were computed for every *if-then* rule in the two trees. Confidence is computed as (*a*-*b*)/*a*, and support as *a*/*n*, where *a* is the number of cases where the rule premise occurs, *b* is the number of non-satisfactory cases and *n* is the total number of instances in the data set, i.e. 1,043 utterances. Tables 4 and 5 present the 5 rules with highest supports in each model.

In the rules, the *no-tag* value represents that an utterance does not have a tag associated to a dialogue act feature, e.g. rule 1 in Table 4, where the utterance expresses a dialogue act on the obligations but not on the common ground. Features that do not contribute to the classification task are not present in the rules because they are automatically discarded by J48.

In the obligations plane model, the most important feature for dialogue act classification is the complementary dialogue act, i. e. *commgr*.

Rule ID	Rule	a	b	Confidence	Support
1	IF commgr=*no-tag* AND commgr_minus1=*accept* AND utt_duration<=5792, THEN *info-request*	90	52	42.2	8.6
2	IF commgr=*graph-action* AND obligations_minus1=*commit*, THEN *info-request_graph-action*	72	1	98.6	6.9
3	IF commgr=*accept* AND speaker_role=*system* AND obligations_minus1=*action-dir*, THEN *commit*	71	19	73.2	6.8
4	IF commgr=*hold_repeat-rephr*, THEN *info-request*	54	1	98.1	5.2
5	IF commgr=*accept* AND speaker_role=*user* AND commgr_minus1=*graph-action*, THEN *answer*	51	0	100.0	4.9

Table 4. The five rules with highest support for obligations prediction

Rule ID	Rule	a	b	Confidence	Support
1	IF obligations=*commit*, THEN *accept*	112	3	97.3	10.7
2	IF obligations=*info-request* AND speaker_role=*system*, THEN *hold_repeat-rephr*	99	47	52.5	9.5
3	IF obligations=*info-request_graph-action*, THEN *graph-action*	98	2	98.0	9.4
4	IF obligations=*answer* AND commgr_minus1=*graph-action*, THEN *accept*	56	5	91.1	5.4
5	IF obligations=*answer* AND commgr_minus1=*hold_repeat-rephr*, THEN *accept*	48	7	85.4	4.6

Table 5. The five rules with highest support for common ground prediction

Table 6 presents the features ranking according to their presence in the rule set. Features with higher percents are associated to a higher contribution to the classification task because they have a higher discriminative capability.

Feature	% of Rules
commgr	100.0
commgr_minus1	51.0
obligations_minus1	29.0
speaker_role	26.5
first_3	17.4
utt_duration	9.0
first_2	5.2
optimal_pred_mood	2.6

Table 6. Presence of features in the obligations model rules

In the common ground model, also the complementary dialogue act (i.e. *obligations*) is the most contributing feature, as can be seen in Table 7. *Optimal_pred_mood* is not a contributing feature in this model.

Recognition rate per class is evaluated by three ratios: *recall*, *precision* and *F measure*. *Recall* is the number of cases actually belonging to a class divided by the number of cases of that class recognized by the model; *precision* is the number of cases of a class recognized by the model divided by the number of cases actually belonging to it. *F measure* is computed as 2x((*Precision* x *Recall*)/(*Precision* + *Recall*)). *F measure* is satisfactory if it is greater than or equal to 0.8. In the obligations acts model, classes with satisfactory F measures are: *info-request_graph-action*, *info-request_graph-action_answer*, *answer*, *commit* and *offer*. In the common ground model, these are: *graph-action* and *offer_conv-open*.

Feature	% of Rules
obligations	100.0
commgr_minus1	91.4
first_3	27.8
speaker_role	22.5
obligations_minus1	11.9
utt_duration	9.9
first_2	7.9
last_2	2.0

Table 7. Presence of features in the common ground model rules

5 *Conclusions*

The dialogue act from the previous utterance as one of the predictors is useful to improve the accuracy (+5.6 percent points) in the obligations recognition. The recognition of common ground dialogue acts is not benefited from this setting.

An automatic recognition process might be implemented by taking advantage of a two-steps recognition, where the dialogue act from one of the two expression planes can be recognized by a lexical-based algorithm and then this dialogue act can be used as one of the inputs for the recognition of the dialogue act on the complementary plane by a classification tree; i.e. to use obligations as one of the inputs for common ground or vice versa.

A model for automatic recognition of dialogue acts is useful to implement dialogue management systems by providing information that complements the speech recognition processes.

Acknowledgments

The authors thank the anonymous reviewers of this paper. Sergio Coria also thanks Varinia Estrada for annotations and valuable comments, and CONACyT and DGEP-UNAM for support to this work.

References

Allen, J. and M. Core. 1997. *Draft of DAMSL: Dialog Act Markup in Several Layers*. Informe técnico, The Multiparty Discourse Group. University of Rochester, Rochester, USA, October.

Beckman, M.E., M. Diaz-Campos, J. Tevis-McGory, and T.A. Morgan. 2002. *Intonation across Spanish, in the Tones and Break Indices framework*. Probus 14, 9-36. Walter de Gruyter.

Breiman, L., J.H. Friedman, R.A. Olshen and C.J. Stone. 1983. *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth & Brooks, USA.

Bunt, H. 1994. Context and Dialogue Control. *THINK Quarterly*.

Bunt, H. 1995. *Dynamic interpretation and dialogue theory*. The structure of multimodal dialogue, ed. by M. M. Taylor, F. Neel, and D. G. Bouwhuis. Amsterdam. John Benjamins

Carletta, Jean. 1996. Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics, 22(2):249-254.

Coria, S. and L. Pineda. 2006. Predicting Dialogue Acts from Prosodic Information. In *Proceedings of the Seventh International Conference on Intelligent Text Processing and Computational Linguistics, CICLing* (Mexico City), February.

Cuétara, J. 2004. *Fonética de la ciudad de México. Aportaciones desde las tecnologías del habla. Tesis para obtener el título de Maestro en Lingüística Hispánica.* Maestría en Lingüística Hispánica, Posgrado en Lingüística, Universidad Nacional Autónoma de México.

Garrido, J.M. 1996. *Modelling Spanish Intonation for Text-to-Speech Applications*. *Doctoral Dissertation*. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona, Spain.

Hirst, D., A. Di Cristo and R. Espesser. 2000. Levels of representation and levels of analysis for the description of intonation systems. In *M. Horne (ed) Prosody: Theory and Experiment* (Kluwer, Dordrecht).

Navarro-Tomás, T. 1974. *Manual de entonación española*. New York: Hispanic Institute, 2ª edición corregida, 1948 .- México: Colección Málaga, 3ª edición, 1966. - Madrid: Guadarrama (Punto Omega, 175), 4ª edición, 1974.

Pineda, L. 2007. *The DIME Corpus*. Department of Computer Science, Institute of Applied Mathematics and Systems. National Autonomous University of Mexico. http://leibniz.iimas.unam.mx/~luis/DIME/CORPUS-DIME.html

Pineda, L., V. Estrada and S. Coria. 2006. The Obligations and Common Ground Structure of Task Oriented Conversations. In *Proceedings of X Iberoamerican Artificial Intelligence Conference, Iberamia*, Ribeirao Preto, Brazil, October.

Shriberg, E., R. Bates, A. Stolcke, P. Taylor, D. Jurafsky, K. Ries, N. Coccaro, R. Martin, M. Meteer, and C. Van EssDykema. 1998. Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech? *Language and Speech* 41(3-4), Special Issue on Prosody and Conversation, 439-487, USA.

Siegel, S. and N.J. Castellan, Jr. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, second edition, 1988.

Wahlster, W. 1993. VERBMOBIL: Translation of Spontaneous Face-to-Face Dialogs. In *Proceedings of the 3rd EUROSPEECH*, pp. 29-38, Berlin, Germany.

Witten, I. and E. Frank. 2000. *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan-Kauffman Publishers. San Francisco, CA. USA: 89-97.