

MICHEL GÉNÉREUX

TOWARDS STOCHASTIC MODELING OF
SPOKEN DIALOGUE SYSTEMS

Dissertation zur Erlangung des
Doktorgrades der Philosophie aus dem
Fachgebiet *Sprachwissenschaft* eingereicht an
der Geistes- und Kulturwissenschaftlichen
Fakultät der Universität Wien

Wien 2004

ABSTRACT

*Towards Stochastic Modeling of
Spoken Dialogue Systems*

Michel Génèreux

Department of Linguistics

University of Vienna

2004

This thesis describes a statistical framework for Natural Language Understanding (NLU) and Generation (NLG) in broad domain human-machine speech interaction. Class-based n-grams models and a single set of speech acts are put to contribution from Understanding to Generation, while contextual example-based parsing provides the additional necessary depth of analysis for the difficult and crucial task of understanding human speech in dialogues.

In Understanding, the input is first channeled through a class-based n-grams model for unit selection in contextual parsing. Classification is made according to a pair of speech act and topic. In contextual parsing, an empirical method for mapping speech input to shallow semantic representation is used. Semantic parsing is realized through a bottom-up type parsing paradigm where the operators are based on semantic concepts, obtained from a lexicon. A statistically trained model specializes the parser, by guiding the runtime beam-like search of possible parses. The semantic representation is a logical form equivalent to a Discourse Representation Structure (DRS), in which basic discourse phenomena are represented. In Generation, a n-grams model is used to produce a random set of utterances for a given class, among which an arbitration module makes a selection, based on linguistic likelihood, length and attribute replacement. The key aspect of Dialogue Management (DM) is initiative taking based on the appropriate selection of a speech act.

The overall system is suitable for fast prototyping of broad conversational speech domain, while the robust composite model, n-grams with contextual parsing, combines efficiently the robustness of statistical modeling and the correctness of linguistic analysis.

ACKNOWLEDGEMENTS

I am grateful to the many people who have contributed to this thesis, and I wish to thank them for making my time in the *Sprachwissenschaft* graduate program at the University of Vienna a great learning experience.

First I would like to thank both my supervisors who helped me to navigate through the shady circles of multidisciplinary work; Prof. Wolfgang U. Dressler, my first supervisor, whose guidance was essential in shaping the linguistic backbone of the thesis and my second supervisor, Prof. Harald Trost, who got me started in my first job in Computational Linguistics and whose comments smoothed away technical flaws in my work and helped clarify the thesis. I am also grateful to them for facilitating my integration into a new environment when I came to Austria. Special thanks to the ÖFAI in Vienna, where I gained much experience in the field, exchanged ideas with colleagues and drew most of my inspiration for this thesis. I wish to thank NIS-Lab in Denmark for letting me use some of the conversational data we collected while working on the NICE project. Thanks also to Ashling for correcting my English mistakes, but I am the only responsible for any remaining errors.

This thesis has been lugged around Europe, and I am grateful to the many friendly Austrian, Danish and French people around me who put up with my moods and encouraged me patiently. My folks in Canada deserve a special mention for their support.

Contents

List of Tables	12
List of Figures	14
List of DRSs	15
List of Algorithms	17
1 Introduction	19
1.1 The Proposed Model	20
1.2 Contribution of this Thesis	22
1.3 Overview of the Thesis	24
2 Background and Starting Points	27
2.1 Background	27
2.1.1 Speech Recognition	27
2.1.2 The Domain	30
2.1.3 Talking with HCA	31
2.1.4 A few Principles	31
2.1.5 Data Collection	34
2.1.6 Data Transcription	36
2.1.7 State of the Art	36
2.2 Starting Points	38

2.2.1	The Arbitrary of Linguistics	39
2.2.2	A Model of Language	40
2.2.3	Sentence Level	42
2.2.4	Discourse Level	48
2.2.5	A Formal Representation for Discourse	53
3	Using Discourse Representation Structures	59
3.1	Introduction	59
3.2	The Need for DRT	60
3.3	Overview of DRT	61
3.3.1	Discourse Representation Structure	61
3.3.2	Discourse Referents	62
3.3.3	Conditions	62
3.3.4	Accessibility	63
3.3.5	How to go from a Sentence to its DRS	64
3.4	Pronoun Resolution	66
3.4.1	Refining the Resolution Process: Focusing	67
3.5	Presupposition Resolution	70
3.5.1	Carrying out the Resolution	71
3.6	DRT and Stochastic Parsing	73
4	Speech Act and Topic Classification	77
4.1	Topic Definition and Speech Act Selection	77
4.1.1	Topics and Topicalization	77
4.1.2	Speech Acts	78
4.2	General Observations	84
4.3	Reviewing two Systems	86
4.3.1	The Boosting Approach	86
4.3.2	Word Sequence Tagging	88
4.4	N-gram Based Text Categorization	91

5	An Example-Based Semantic Parser	93
5.1	Introduction and Motivation	93
5.2	Parser Architecture	94
5.3	Background	96
5.3.1	Semantic Parsing	96
5.3.2	Statistical Parsing	98
5.3.3	First Order Logic	99
5.4	Previous Work	100
5.4.1	Semantic Tagging	100
5.4.2	Machine Learning Approach	102
5.4.3	A Fully Statistical Approach	103
5.4.4	Other Systems	105
5.5	Overview of the Parsing Process	106
5.5.1	The SHIFT-REDUCE Parser	106
5.5.2	The Input String	107
5.5.3	The Parse Stack	108
5.5.4	The SHIFT Action	108
5.5.5	The INTRODUCE Action	109
5.5.6	The DROP Action	109
5.5.7	The COREF Action	109
5.5.8	The SCOPE Action	110
5.5.9	A Parse State (no statistics)	110
5.5.10	A Final State (no statistics)	110
5.5.11	Semantic Lexicon	111
5.5.12	Our Variant of the Shift-Reduce Parser	111
5.5.13	Topic Extraction	112
5.6	Training	113
5.6.1	Training File	113
5.6.2	File used by the Specialized Parser: the Statistical File	114
5.7	Statistical Parsing	115
5.7.1	Search Space	115
5.7.2	Measure of Similarity between Sets	115

5.7.3	Parametrizing the Model	118
5.7.4	A Full Parse with Statistics	120
6	Dialogue Management	127
6.1	Taking Initiative	128
6.2	Computing the Speech Acts	129
6.3	Building a Semantic Frame	130
6.4	Discussion	131
7	Natural Language Generation	133
7.1	Introduction	133
7.2	N-gram Language Models	134
7.3	Generating Utterances	135
7.4	Arbitration	137
8	Results	141
8.1	Integration	141
8.2	Topic Classification	142
8.3	Natural Language Understanding	144
8.4	Dialogue Management and Natural Language Generation	145
9	Conclusion	147
9.1	Further Work	149
A	Translating DRSS to FOL	159
B	Data Collection	161
B.1	Recording	161
B.2	Transcription Guidelines	161
B.3	Transcription - Annotation of Topics (NLU), Speech Acts (NLU,DM,NLG) and POS (NLG)	164
B.4	XML Transcription Produced by Transcriber	165

C Topic and Speech Act Spotting	173
C.1 Full List of Topics	173
C.2 Using the Topic and Speech Act Spotter	176
D NLU: Browsing Newspapers (Window:1, FOL, German)	179
D.1 Semantic Lexicon	179
D.2 Training File	182
D.3 Parsing	184
E NLG: Closing a Conversation	189
E.1 Training Data	189
E.2 Language Model	189
E.3 Generation	193
Bibliography	195