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 Généreux
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. Wolfgand 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 NISLab 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