Title: Novel Methods for Natural Language Generation
in Spoken Dialogue Systems

Author: Ondřej Dušek

Department: Institute of Formal and Applied Linguistics

Supervisor: Ing. Mgr. Filip Jurčíček, Ph.D.,
Institute of Formal and Applied Linguistics

Abstract:

This thesis explores novel approaches to natural language generation (NLG) in spoken dialogue systems (i.e., generating system responses to be presented the user), aiming at simplifying adaptivity of NLG in three respects: domain portability, language portability, and user-adaptive outputs.

Our generators improve over state-of-the-art in all of them: First, our generators, which are based on statistical methods (A* search with perceptron ranking and sequence-to-sequence recurrent neural network architectures), can be trained on data without fine-grained semantic alignments, thus simplifying the process of retraining the generator for a new domain in comparison to previous approaches. Second, we enhance the neural-network-based generator so that it takes preceding dialogue context into account (i.e., user's way of speaking), thus producing user-adaptive outputs. Third, we evaluate several extensions to the neural-network-based generator designed for producing output in morphologically rich languages, showing improvements in Czech generation.

In addition, we compare different NLG architectures (a traditional two-step pipeline with separate sentence planning and surface realization steps and a joint, end-to-end approach), and we collect and make freely available two novel training datasets for NLG.

Keywords: natural language generation, spoken dialogue systems,
adaptivity, dialogue entrainment, multilingualism