Joint Satisfaction of Syntactic and Pragmatic Constraints Improves Incremental Spoken Language Understanding

Andreas Peldszus, Okko Buß, Timo Baumann & David Schlangen

Abstract
We present a model of semantic processing of spoken language that is robust against ill-formed input, such as can be expected from automatic speech recognisers, respects both syntactic and pragmatic constraints in the computation of most likely interpretations, uses a principled, expressive semantic representation formalism (RMR) with a well-defined model theory, and works continuously (producing meanings representations on a word-by-word basis, rather than only for full utterances) and incrementally (computing only the additional contribution by the new word, rather than re-computing for the whole utterance-so-far).

We show that the joint satisfaction of syntactic and pragmatic constraints improves the performance of the NLU component (around 10% absolute, over a syntax-only baseline).

1. Introduction

Motivation for Incremental Processing in SDSSs
- psycholinguistic plausibility
- descriptive adequacy (backchannel feedback, intervening corrections, clarifications, completions etc.)
- computational benefits (by making lower-level processing results earlier available to higher-level modules and by providing higher-level feedback to lower-level modules)

Aim
Investigate the potential advantage of pragmatic feedback to the parser, specifically of reranking syntactic readings according to their success in real-world reference.

2. The Model

Parser
- Monotonic growth of fully connected syntactic structures. Left factorisation delays syntactic decisions.
- Robust lexical operations (insertions, deletions and repairs) for ASR input.

Semantic Construction
- Combination is linearised in a top-down-left-to-right fashion (Instead of bottom-up).
- No need to define semantics of open projected nodes, no re-interpretation of existing parts of the tree.
- Robust Minimal Recursion Semantics (Copekaste, 2007) as semantic formalism, adapted to the chosen combination order.
- Semantic construction synchronised with syntactic expansion: monotonically, continuously and incrementally.

Reference Feedback
- Test each representation for its semantic plausibility, depending on whether definite NPs refer uniquely, ambiguously or fail to refer to pieces in the corresponding world state.
- Degrade unparsable readings (lower probability in the next parsing step).

The model is implemented in the InproTK (Schlangen et al., 2010).

3. Experiment Data

- Corpus of Pentomin puzzle games (see Fernández & Schlangen, 2007), collected in a WOZ-study with 284 games of 20 participants.
- Users instruct the wizard to manipulate different puzzle pieces in order to reach a specified goal state.
- Subset of 1026 utterances with reliable semantics (extracted from the wizard’s next action) & without pronouns referring to pieces (no discourse model) in gold transcript and in ASR output version.

Natural spontaneous speech representative for interactions of such a domain
- "das blaue ahm das teill links oben nebem dem winkel" the blue-one ahm the piece left top next-to the angle
- "genau und jetzt nochmal drehen" right and now again rotate
- "lischen unten" delete bottom

Grammar
- Small handcrafed core grammar (30 rules), easy to engineer
- Weights set according to intuition

4. Evaluation

Baselines & Settings
Evaluation of semantic accuracy by comparing the extracted "gold" semantic alignment with the resolved reference of...
- Just Syntax (JS): the single-best derivation of syntax only
- External Filtering (EF): of the 5 best syntax-only derivations, the best referring one
- Syntax/Pragmatic Interaction (SPI): the single-best reference-feedback derivation
- Combined Interaction & Filtering (CIF): of the 5 best reference-feedback derivations, the best referring one

Incremental Score Metric
A measure of how the resolved reference matches over time, with increasingly stronger influence of later changes

5. Results

With reference feedback:
- Less mismatches (-1)
- More partial matches (+)
- More unique matches (+)
- 12.3% abs. improvement for strict accuracy and 19.3% for relaxed accuracy (allowing partial matches) for transcript SPI over JS
- Effect not only at the final word, but within the utterance (incremental score)
- Similar but slightly smaller improvements for transcript EF over SPI
- Similar but slightly smaller improvements for ASR input
- No further improvement by additional filtering (SPI vs CIF)
- Low baseline due to the complexity of authentic data (see examples)