Speech is a natural channel for human-computer interaction in robotics and consumer applications. Natural language understanding pipelines that start with speech can have trouble recovering from speech recognition errors. Black-box automatic speech recognition (ASR) systems, built for general purpose use, are unable to take advantage of in-domain language models that could otherwise ameliorate these errors. In this work, we present a method for re-ranking black-box ASR hypotheses using an in-domain language model and semantic parser trained for a particular task. Our re-ranking method significantly improves both transcription accuracy and semantic understanding over a state-of-the-art ASR’s vanilla output.

### Approach

We re-rank the n-best hypothesis list from an ASR system by interpolating scores from an in-domain semantic parser and language model.

\[
h^* = \arg \max \limits_{h \in H} (S(h)) \\
S(h) = (1 - \alpha) \cdot S_{\text{lm}}(h) + \alpha \cdot S_{\text{sem}}(h)
\]

### Semantic Parsing

Used a Combinatory Categorical Grammar (CCG) based probabilistic CKY parser

<table>
<thead>
<tr>
<th>Surface Form</th>
<th>CCG Category</th>
<th>Semantic Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>walk</td>
<td>S/PP</td>
<td>( \lambda x . (\text{walk}(x)) )</td>
</tr>
<tr>
<td>to</td>
<td>PP/NP</td>
<td>( \lambda x . (\text{to}(x)) )</td>
</tr>
<tr>
<td>John</td>
<td>N</td>
<td>john</td>
</tr>
</tbody>
</table>

### Language Modeling

Used a trigram back-off language model with Witten-Bell discounting

\[
P(w_n | w_1, \ldots, w_{n-1}) = P(w_n | w_{n-2}, w_{n-1}) \\
P(w_1, \ldots, w_n) = \prod_{i=1}^{N} P(w_i | w_{i-2}, w_{i-1})
\]

### Results

Tested our methodology using the Google Speech API

- Requested 10 hypotheses per utterance.
- Gave parser budget of 10 seconds per hypothesis.

Measured system performance over 5 different conditions:

- **Oracle**: Best achievable performance from re-ranking.
- **ASR**: System performance without re-ranking.
- **SemP**: Re-ranking using solely semantic parser scores.
- **LM**: Re-ranking using solely language model scores.
- **Both**: Re-ranking using interpolated semantic parser and language model scores.

Evaluated system performance on 3 metrics:

- **Word error rate (WER)**: Computes number of insertions, deletions, and substitutions in hypothesis in order to measure transcription accuracy.
- **Semantic form accuracy (ACC)**: Checks for a one-to-one match between hypothesis logical form and correct logical form.
- **Semantic form F1**: Measures harmonic mean of recall and precision of the predicates in the hypothesis semantic form.

### Acknowledgements

This work is supported by an NSF EAGER grant (IIS-1548567), and NSF NRI grant (IIS-1637736), and a National Science Foundation Graduate Research Fellowship to the second author.