**Summary**

- **Motivation**
  - Domain language may drift over time so that ensuring language coverage in dialog systems can be a challenge (Furnas et al., 1987).
  - The mismatch between training data and current input increases recognition errors and misunderstanding.
  - Detect-and-Learn strategy requires human effort and takes more time to adapt the vocabulary and LM.

- **Approach: Expect-and-Learn**
  - Automatically acquiring potential out-of-vocabulary (OOV) words by leveraging different types of words relatedness.

- **Result**
  - Both recognition and semantic parsing accuracy can be improved after acquiring potential OOVs.

**OOV Learning Method**

- **Detect-and-Learn (Qin et al., 2011; 2012):**
  - Discover OOV words during the conversation
  - Example: S: "I heard something like SELF, can you repeat it?" U: "It’s SELFIE."
  - Drawbacks
    - Limited number of new words
    - Required human efforts to correct spellings and pronunciations

- **Expect-and-Learn (proposed):**
  - Use semantic relatedness to automatically enrich the vocabulary and language model beforehand

- **Advantages**
  - Large amount of potentially useful new words can be learned
  - No human involved

**Expect-and-Learn Procedure**

- **Idea: learn new words related to the current domain represented by in-vocabulary words (IVs)**
  1. From the IV with the highest frequency v*, one unseen word w* is extracted from the resource according to:
     - Local relatedness (Algo1): w* is mostly related to v*
     - Global relatedness (Algo2): w* is mostly related to the complete IV set
  2. Repeat until the size of vocabulary satisfies a threshold

- **Language Model Expansion**
  - Use Kneser-Ney smoothing to estimate the unigram for the newly learned OOVs.

**Relatedness Resources**

1. **Linguistically semantic relatedness**
   - Defined by linguistics, e.g., WordNet (WN), Paraphrase Database (PPDB) (Ganitkevitch et al., 2013)

2. **Data-driven semantic relatedness**
   - Distributional semantics, e.g., continuous bag-of-word embeddings (CBOW) (Mikolov et al., 2013)

**Experimental Results**

- **Dataset:** Wall Street Journal
  - Acoustic model: WSJ GMM-HMM semi continuous
  - Pronunciation: CMU Dictionary + Logios Lexicon Tool
- **OOV Coverage Evaluation**
  - How much OOV tokens in test set can be covered by using different relatedness resources.

**Recognition and Understanding Performance**

**Only Domain Specific Models**

<table>
<thead>
<tr>
<th>Learning Strategy</th>
<th>Vocab Size</th>
<th>OOV Rate (%)</th>
<th>Recog. WER (%)</th>
<th>SLU F1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>2854</td>
<td>22.6</td>
<td>49.9</td>
<td>57.0</td>
</tr>
<tr>
<td>Algol1</td>
<td>5394</td>
<td>11.7</td>
<td>41.6</td>
<td>65.4</td>
</tr>
<tr>
<td>Algol2</td>
<td>5394</td>
<td>11.6</td>
<td>42.0</td>
<td>65.1</td>
</tr>
<tr>
<td>Oracle</td>
<td>4254</td>
<td>0.0</td>
<td>23.5</td>
<td>80.9</td>
</tr>
</tbody>
</table>

**Domain + Generic Models**

<table>
<thead>
<tr>
<th>Learning Strategy</th>
<th>Vocab Size</th>
<th>OOV Rate (%)</th>
<th>Recog. WER (%)</th>
<th>SLU F1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>20175</td>
<td>3.6</td>
<td>21.7</td>
<td>82.2</td>
</tr>
<tr>
<td>Algol1</td>
<td>22599</td>
<td>3.0</td>
<td>20.3</td>
<td>83.2</td>
</tr>
<tr>
<td>Algol2</td>
<td>22599</td>
<td>3.0</td>
<td>20.4</td>
<td>83.2</td>
</tr>
<tr>
<td>Oracle</td>
<td>20431</td>
<td>0.0</td>
<td>15.1</td>
<td>87.1</td>
</tr>
</tbody>
</table>

**Conclusion**

- **Speech recognition and language understanding performance can be improved through an OOV expect-and-learn procedure.**
- A limited domain vocabulary can be utilized to effectively acquire OOVs by the word relatedness theory through web knowledge bases.
- With data-driven semantic relatedness, both the global and local learning procedures are able to successfully harvest more than 50% of OOVs, leading to better recognition and understanding performance.
- This work demonstrates that
  - OOV learning may benefit dialog system
  - the proposed expect-and-learn strategy outperforms the traditional detect-and-learn in both higher effectiveness and no human involvement.