Robust Parsing for HPSG

Woodley Packard

November 24th, 2015

N.B.: Much of this is joint work with Dan Flickinger.
Head Driven Phrase Structure Grammar

HPSG Pollard and Sag [1994]: simultaneously
Head Driven Phrase Structure Grammar

HPSG Pollard and Sag [1994]: simultaneously

1. linguistically **precise**
Head Driven Phrase Structure Grammar

HPSG Pollard and Sag [1994]: simultaneously

1. linguistically precise
2. grounded in rich semantics
Head Driven Phrase Structure Grammar

HPSG Pollard and Sag [1994]: simultaneously

1. linguistically **precise**
2. grounded in rich semantics
3. bidirectional
Head Driven Phrase Structure Grammar

HPSG Pollard and Sag [1994]: simultaneously

1. linguistically **precise**
2. grounded in rich semantics
3. bidirectional
4. computationally tractable.
Head Driven Phrase Structure Grammar

HPSG Pollard and Sag [1994]: simultaneously

1. linguistically **precise**
2. grounded in rich semantics
3. bidirectional
4. computationally tractable.

All are good qualities, but...
Achilles Heel?

“It’s a precision grammar, not a recall grammar.”

Glenn Slayden
“It’s a *precision* grammar, not a *recall* grammar.”

Glenn Slayden

ERG Flickinger [2000, 2011] on edited English: 92-96%
Achilles Heel?

“‘It’s a precision grammar, not a recall grammar.’”

Glenn Slayden

ERG Flickinger [2000, 2011] on edited English: 92-96%

Unedited English or edited un-English (e.g. spontaneous speech, Twitter, headlines): lower
In Defense of Precision

Imperfect coverage — sad
In Defense of Precision

Imperfect coverage — sad
Precision — not sad
In Defense of Precision

Imperfect coverage — sad
Precision — not sad

1. Well-formed generation
In Defense of Precision

Imperfect coverage — sad
Precision — not sad

1. Well-formed generation
2. Dramatic reduction in ambiguity
In Defense of Precision

Imperfect coverage — sad
Precision — not sad

1. Well-formed generation
2. Dramatic reduction in ambiguity
3. Application in language tutoring
In Defense of Precision

Imperfect coverage — sad
Precision — **not sad**

1. Well-formed generation
2. Dramatic reduction in ambiguity
3. Application in language tutoring

...but *sometimes* it would be nice to be able to *relax* precision
Robust Parsing

robustness:
The ability to produce an analysis even when the input utterance is not well-formed
Robust Parsing

Robustness:
The ability to produce an analysis even when the input utterance is not well-formed
  ▶ errors in the grammar
robustness:
The ability to produce an analysis even when the input utterance is not well-formed

- errors in the grammar
- errors in the utterance
Robust Parsing

robustness:
The ability to produce an analysis even when the input utterance is not well-formed

- errors in the grammar
- errors in the utterance
- playful language
- etc.
Degrees of Robustness

Regardless of how ill-formed the input is?
Degrees of Robustness

Regardless of how ill-formed the input is?

- A longed-haired cat with it’s eyes closed.
Degrees of Robustness

Regardless of how ill-formed the input is?

- A longed-haired cat with its eyes closed.
- A woman laying across two men sitting on a sofa.
Degrees of Robustness

Regardless of how ill-formed the input is?

- A longed-haired cat with it’s eyes closed.
- A women laying across two men sitting on a sofa.
- Two sheep on to of a cliff looking at the camera.
Degrees of Robustness

Regardless of how ill-formed the input is?

- A long-haired cat with its eyes closed.
- A women laying across two men sitting on a sofa.
- Two sheep on top of a cliff looking at the camera.
- What the what?! ?: Voice of Charlie Brown arrested, charged. ?
Degrees of Robustness

Regardless of how ill-formed the input is?

- A longed-haired cat with it’s eyes closed.
- A women laying across two men sitting on a sofa.
- Two sheep on to of a cliff looking at the camera.
- What the what?! ?: Voice of Charlie Brown arrested, charged. ?
- The the slowly dog slept.
Degrees of Robustness

Regardless of how ill-formed the input is?

- A longed-haired cat with it’s eyes closed.
- A women laying across two men sitting on a sofa.
- Two sheep on to of a cliff looking at the camera.
- What the what?! ?: Voice of Charlie Brown arrested, charged. ?
- The the slowly dog slept.
- 000fe10: 2d36 2d32 3031 3320 3134 3a32 383a 3234
Introduction

Robust Parsing

Methodology for Evaluation
Practical Techniques
Evaluation
Let’s pretend.

Suppose we have a robust parser. How can we tell if it’s good or bad?
Let’s pretend.

Suppose we have a robust parser. How can we tell if it’s good or bad?

- Blunt metrics: coverage and speed on real data?
Let’s pretend.

Suppose we have a robust parser. How can we tell if it’s good or bad?

- Blunt metrics: coverage and speed on real data?
- Extrinsic evaluation: helpful on some downstream task?
Let’s pretend.

Suppose we have a robust parser. How can we tell if it’s good or bad?

- Blunt metrics: coverage and speed on real data?
- Extrinsic evaluation: helpful on some downstream task?
- Intrinsic evaluation: get the right analysis?
Coverage and Speed

Let’s use two types of text:

- Wall Street Journal PTB Section 22
- Twitter News
Extrinsic Evaluation

Robot control (SemEval 2014 Task 6)
  ▶ Translating simple (but noisy) English instructions into machine-readable commands.
  ▶ Metric: exact match on the machine-readable command
Intrinsic Evaluation

Difficult.

- Typical HPSG parser evaluation: gold treebank on *grammatical inputs*
- No infrastructure to treebank *ungrammatical inputs*, so no gold data.
Intrinsic Evaluation

Difficult.

- Typical HPSG parser evaluation: gold treebank on *grammatical inputs*
- No infrastructure to treebank *ungrammatical inputs*, so no gold data.

But—small corpus of 45 painstakingly constructed gold MRSes for unparseable WSJ inputs.
Introduction

Robust Parsing
Methodology for Evaluation
Practical Techniques
Evaluation
Mini-review of HPSG mechanics

- Grammar: lexemes and rules, which are feature structures
- Signs: lexemes, and recursive combinations by rules
- Unification: determines whether a rule can apply or not
- ... and when a spanning sign is sentential
- Semantics: contained in the feature structure of sentential sign

```
SB-HD_MC_C
  ├── SP-HD_N_C
  │    ├── DET
  │    │    └── the
  │    └── N_SG_ILR
  │         └── N
  │             └── dog
  └── W_PERIOD_PLR
      └── V_PST_OLR
          └── V
              └── slept.
```
Two basic ways to increase robustness
Two basic ways to increase robustness

1. Add rules (or lexemes) to admit more analyses
Two basic ways to increase robustness

1. Add rules (or lexemes) to admit more analyses
2. Relax constraints so existing rules apply more freely
Approach 1: Add more rules

Two variants, mainly investigated by Dan Flickinger:
Approach 1: Add more rules

Two variants, mainly investigated by Dan Flickinger:

- *Bridging* — any sequence of signs can be sentential
Approach 1: Add more rules

Two variants, mainly investigated by Dan Flickinger:

- Bridging — any sequence of signs can be sentential
- Pacman — certain parts of the input can be ignored
1a: Bridging

Leave the core grammar alone,
1a: Bridging

Leave the core grammar alone,
...but let any sequence of legal signs constitute a sentential analysis.
Leave the core grammar alone, ... but let any sequence of legal signs constitute a sentential analysis. “Self-help” — with a few new “grammar” rules, can use pre-existing parsers.
Promote anything to “bridge”:

\[
\begin{align*}
\text{bridge\_unary\_phrase} &\quad \text{canonical\_synsem} \\
\text{SS} &\quad \text{LOC.CAT} &\quad \text{HEAD} &\quad \text{bridge} \\
\text{LEX} &\quad \text{VAL} &\quad \text{MOD} &\quad \langle \rangle \\
\text{INFLECTD} &\quad \text{na} &\quad \text{LEX} \\
\text{ARGS} &\quad \text{sign} &\quad \text{SS...HEAD} &\quad \text{non\_frag} \\
&\quad \text{INFLECTD} &\quad \text{INFLECTD} &\quad + \\
\end{align*}
\]
Posit “bridge” lexical entries to fill lexical gaps:

\[
\begin{array}{c}
\text{INFLECTD} \\
\text{LEX} \\
\text{LOC.CAT} \\
\text{HEAD} \\
\text{VAL} \\
\text{SS}
\end{array}
\begin{array}{c}
na \\
- \\
\text{canonical_synsem} \\
\text{bridge} \\
\text{sat} \\
\end{array}
\]
Combine two bridges (strictly right-branching):

```
bridge_unary_phrase

SS

INFLECTD

ARGS

canonical_synsem

LOC.CAT

HEAD

VAL

bridge

MOD

sat

INFLECTD

ARGS

SS...HEAD

bridge

INFLECTD

na

SS...HEAD

bridge
```
Bridging root condition:

\[
\begin{bmatrix}
bridge\_\text{root} \\
SS\ldots\text{HEAD} & bridge
\end{bmatrix}
\]
Bridging: The dog slept the cat.
Bridging: I believe the quickly dog slept.
Let (some) signs “gobble up” bits of input that would otherwise be indigestible.
1b: Pacman

Let (some) signs “gobble up” bits of input that would otherwise be indigestible.

Frequently, problems are surfacey, e.g. a word usage or punctuation choice that the grammar doesn’t know about.

Hypothesis: a lot of the time we can do pretty well by just pretending that word (or phrase) isn’t there.

*I believe the quickly dog slept.*
Pacman verb rule schema: take one bite to the right, of anything you want.
Pacman noun rule schema: take one bite to the left, of anything you want.

```
pacman_n_phrase

[SS

  LOCAL.CAT [1 [HEAD basic_noun]]

--BRDG +

[SS.--BRDG -]

ARGS

  SS

    [SS

      LOCAL.CAT [1]]

    [SS

      --BRDG -]

] ]
```
Pacman: The dog slept the cat.
Pacman: I believe the quickly dog slept.
### Review (adding rules)

<table>
<thead>
<tr>
<th></th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridging</td>
<td>Very robust</td>
<td>Extreme ambiguity</td>
</tr>
<tr>
<td></td>
<td>Provenance clear</td>
<td>Poor semantics</td>
</tr>
<tr>
<td></td>
<td>Packability</td>
<td>Intractable unpacking</td>
</tr>
<tr>
<td>Pacman</td>
<td>Problems localized</td>
<td>Less robust</td>
</tr>
<tr>
<td></td>
<td>OK semantics</td>
<td>Extreme ambiguity</td>
</tr>
</tbody>
</table>

Another universal drawback: optimal results would require (currently nearly nonexistent) training data for disambiguating the new “ambiguities.”
Approach 2: Relax constraints

Basic idea: let things unify even when they aren’t supposed to.
Approach 2: Relax constraints

Basic idea: let things unify even when they aren’t supposed to.

- Formal problem: what is the result of unifying two incompatible feature structures?
Approach 2: Relax constraints

Basic idea: let things unify even when they aren’t supposed to.

▶ Formal problem: what is the result of unifying two incompatible feature structures?
▶ ... partial answer: the failure is sometimes localized.
Approach 2: Relax constraints

Basic idea: let things unify even when they aren’t supposed to.

- Formal problem: what is the result of unifying two incompatible feature structures?
- ... partial answer: the failure is sometimes localized.
- Practical problem: what does parsing mean, with no hard constraints at all?
Approach 2: Relax constraints

Basic idea: let things unify even when they aren’t supposed to.

▶ Formal problem: what is the result of unifying two incompatible feature structures?
▶ ... partial answer: the failure is sometimes localized.
▶ Practical problem: what does parsing mean, with no hard constraints at all?
▶ ... partial answer: statistics!
Yi Zhang’s *jigsaw*

Induce a PCFG from...
Yi Zhang’s *jigsaw*

Induce a PCFG from...

- Gold treebanks (100k sentences)
Yi Zhang’s jigsaw

Induce a PCFG from...

- Gold treebanks (100k sentences)
- WikiWoods (50m sentences)
Induce a PCFG from...

- Gold treebanks (100k sentences)
- WikiWoods (50m sentences)
- Level of robustness depends on amount of training data and fanciness of PCFG labels
Yi Zhang’s *jigsaw*

Induce a PCFG from...
- Gold treebanks (100k sentences)
- WikiWoods (50m sentences)
- Level of robustness depends on amount of training data and fanciness of PCFG labels

Output: a derivation tree
Induce a PCFG from...

- Gold treebanks (100k sentences)
- WikiWoods (50m sentences)
- Level of robustness depends on amount of training data and fanciness of PCFG labels

Output: a derivation tree...that usually doesn’t unify successfully.
Robust Unification

Ideas of Stephan Oepen et al., unpublished(?)
Robust Unification

Ideas of Stephan Oepen et al., unpublished(?)

- When two types at the same path have no GLB, unification normally fails
Robust Unification

Ideas of Stephan Oepen et al., unpublished(?)

▶ When two types at the same path have no GLB, unification normally fails

▶ Instead, pretend they do have a GLB – pick something (typically one or the other input type) and keep going
Robust Unification

Ideas of Stephan Oepen et al., unpublished(?)

- When two types at the same path have no GLB, unification normally fails
- Instead, pretend they do have a GLB – pick something (typically one or the other input type) and keep going
- Allow cyclic feature structures by fiat
Robust Unification

Ideas of Stephan Oepen et al., unpublished(?)

- When two types at the same path have no GLB, unification normally fails
- Instead, pretend they do have a GLB – pick something (typically one or the other input type) and keep going
- Allow cyclic feature structures by fiat

Output: a feature structure
Robust Unification

Ideas of Stephan Oepen et al., unpublished(?)

- When two types at the same path have no GLB, unification normally fails
- Instead, pretend they do have a GLB – pick something (typically one or the other input type) and keep going
- Allow cyclic feature structures by fiat

Output: a feature structure ... that may not be formally well-formed
Robust Unification

Ideas of Stephan Oepen et al., unpublished(?)

- When two types at the same path have no GLB, unification normally fails
- Instead, pretend they do have a GLB – pick something (typically one or the other input type) and keep going
- Allow cyclic feature structures by fiat

Output: a feature structure . . . that may not be formally well-formed . . . but generally contains a comprehensible MRS
Putting those together . . .
Putting those together . . .

- get an MRS for any input in the scope of the PCFG
Putting those together . . .

- get an MRS for any input in the scope of the PCFG
- easy to pick PCFG settings leading to 100% coverage*
Putting those together . . .

- get an MRS for any input in the scope of the PCFG
- easy to pick PCFG settings leading to 100% coverage*
- . . . harder to get good quality and coverage simultaneously
Problems:

- “research system” – pieces lying around a workbench that 1 or 2 people know how to assemble
jigsaw + robust unification (continued)

Problems:

- “research system” – pieces lying around a workbench that 1 or 2 people know how to assemble
- Previous investigators have moved on
jigsaw + robust unification (continued)

Problems:

- “research system” – pieces lying around a workbench that 1 or 2 people know how to assemble
- Previous investigators have moved on
- Result: nobody can use it.
csaw

Remedy(?): csaw
Remedy(?): *csaw*

- Integrated reimplementation of the parts that proved successful
Remedy(?)：csaw

- Integrated reimplementatio of the parts that proved successful
- Single tool that deals with preprocessing, PCFG parsing, robust unification, and MRS readout
Remedy(?): csaw

- Integrated reimplemention of the parts that proved successful
- Single tool that deals with preprocessing, PCFG parsing, robust unification, and MRS readout
- Sentence in, MRS out
Remedy(?) : csaw

- Integrated reimplementation of the parts that proved successful
- Single tool that deals with preprocessing, PCFG parsing, robust unification, and MRS readout
- Sentence in, MRS out
- But before I brag too much: not really released yet.
Remedy(?): csaw

- Integrated reimplementation of the parts that proved successful
- Single tool that deals with preprocessing, PCFG parsing, robust unification, and MRS readout
- Sentence in, MRS out
- But before I brag too much: not really released yet.
- But for the brave:
  http://sweaglesw.org/linguistics/csaw/
- ...or svn co
  http://sweaglesw.org/svn/csaw/trunk csaw
csaw: The dog slept the cat.
csaw: I believe the quickly dog slept.
## Reflections on “self-help” vs PCFG

<table>
<thead>
<tr>
<th></th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-help</strong></td>
<td>Tight control</td>
<td>Sometimes intractable</td>
</tr>
<tr>
<td></td>
<td>Easily deployed</td>
<td>Marginal semantics</td>
</tr>
<tr>
<td><strong>PCFG</strong></td>
<td>Flexible</td>
<td>Can be slow</td>
</tr>
<tr>
<td></td>
<td>Can be accurate</td>
<td>Requires training data</td>
</tr>
</tbody>
</table>
Robust Parsing for HPSG

Woodley Packard

Introduction

Robust Parsing
Methodology for Evaluation
Practical Techniques
Evaluation

References
Evaluation: The Suspects

- 1214 — stock ERG with no robustness measures
- bridge — bridging with no ubertagging
- bridge+ut — with --ubertag=0.01
- pacnv — pacman with no ubertagging
- pacnv+ut — with --ubertag=0.01
- csaw-tb — csaw trained on the ERG’s manually curated treebanks (roughly 100k sentences)
- csaw-ww — csaw trained on automatic ERG parses of Wikipedia (roughly 50m sentences)
### Intrinsic Evaluation: The Corpora

<table>
<thead>
<tr>
<th>Items</th>
<th>Avg. Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSJ22</td>
<td>1700</td>
</tr>
<tr>
<td>Twitter</td>
<td>1500</td>
</tr>
<tr>
<td>Robot</td>
<td>909</td>
</tr>
<tr>
<td>Alchemy</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>20.1</td>
</tr>
<tr>
<td></td>
<td>10.0</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>29.0</td>
</tr>
</tbody>
</table>
## Coverage (overall)

<table>
<thead>
<tr>
<th></th>
<th>wsj22</th>
<th>twitter</th>
<th>alchemy45</th>
<th>robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1214</td>
<td>94.35</td>
<td>76.73</td>
<td>17.78</td>
<td>92.08</td>
</tr>
<tr>
<td>bridge</td>
<td>94.06</td>
<td>99.20</td>
<td>66.67</td>
<td>96.92</td>
</tr>
<tr>
<td>bridge+ut</td>
<td>98.24</td>
<td>99.00</td>
<td>93.33</td>
<td>96.92</td>
</tr>
<tr>
<td>pacnv</td>
<td>57.35</td>
<td>86.47</td>
<td>44.44</td>
<td>96.04</td>
</tr>
<tr>
<td>pacnv+ut</td>
<td>90.88</td>
<td>86.20</td>
<td>53.33</td>
<td>96.59</td>
</tr>
<tr>
<td>csaw-tb</td>
<td>99.12</td>
<td>92.73</td>
<td>97.78</td>
<td>96.92</td>
</tr>
<tr>
<td>csaw-ww</td>
<td>99.94</td>
<td>95.67</td>
<td>100.00</td>
<td>96.81</td>
</tr>
</tbody>
</table>
## Coverage (1214 gaps)

<table>
<thead>
<tr>
<th></th>
<th>wsj22</th>
<th>twitter</th>
<th>alchemy45</th>
<th>robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1214</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>bridge</td>
<td>72.92</td>
<td>97.71</td>
<td>72.97</td>
<td>61.11</td>
</tr>
<tr>
<td>bridge+ut</td>
<td>82.29</td>
<td>96.56</td>
<td>91.89</td>
<td>61.11</td>
</tr>
<tr>
<td>pacnv</td>
<td>26.04</td>
<td>66.76</td>
<td>51.35</td>
<td>58.33</td>
</tr>
<tr>
<td>pacnv+ut</td>
<td>43.75</td>
<td>52.72</td>
<td>54.05</td>
<td>56.94</td>
</tr>
<tr>
<td>csaw-tb</td>
<td>93.75</td>
<td>77.65</td>
<td>97.30</td>
<td>61.11</td>
</tr>
<tr>
<td>csaw-ww</td>
<td>98.96</td>
<td>81.66</td>
<td>100.00</td>
<td>59.72</td>
</tr>
</tbody>
</table>
## Speed (overall)

<table>
<thead>
<tr>
<th></th>
<th>wsj22</th>
<th>twitter</th>
<th>alchemy45</th>
<th>robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1214</td>
<td>9.6</td>
<td>62.9</td>
<td>1.4</td>
<td>199.2</td>
</tr>
<tr>
<td>bridge</td>
<td>3.9</td>
<td>24.4</td>
<td>1.5</td>
<td>83.7</td>
</tr>
<tr>
<td>bridge+ut</td>
<td>23.8</td>
<td>182.5</td>
<td>5.9</td>
<td>242.2</td>
</tr>
<tr>
<td>pacnv</td>
<td>1.0</td>
<td>4.5</td>
<td>0.9</td>
<td>10.6</td>
</tr>
<tr>
<td>pacnv+ut</td>
<td>3.6</td>
<td>36.8</td>
<td>2.1</td>
<td>89.1</td>
</tr>
<tr>
<td>csaw-tb</td>
<td>50.0</td>
<td>137.4</td>
<td>29.9</td>
<td>128.8</td>
</tr>
<tr>
<td>csaw-ww</td>
<td>0.6</td>
<td>3.2</td>
<td>0.2</td>
<td>2.9</td>
</tr>
</tbody>
</table>
## Speed (1214 gaps)

<table>
<thead>
<tr>
<th></th>
<th>wsj22</th>
<th>twitter</th>
<th>alchemy45</th>
<th>robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1214</td>
<td>2.6</td>
<td>359.6</td>
<td>2.0</td>
<td>393.1</td>
</tr>
<tr>
<td>bridge</td>
<td>2.2</td>
<td>75.8</td>
<td>2.2</td>
<td>157.5</td>
</tr>
<tr>
<td>bridge+ut</td>
<td>5.9</td>
<td>260.6</td>
<td>6.0</td>
<td>463.5</td>
</tr>
<tr>
<td>pacnv</td>
<td>1.0</td>
<td>8.7</td>
<td>1.0</td>
<td>16.0</td>
</tr>
<tr>
<td>pacnv+ut</td>
<td>2.4</td>
<td>116.7</td>
<td>2.5</td>
<td>178.3</td>
</tr>
<tr>
<td>csaw-tb</td>
<td>34.6</td>
<td>151.0</td>
<td>31.7</td>
<td>205.6</td>
</tr>
<tr>
<td>csaw-ww</td>
<td>0.3</td>
<td>3.4</td>
<td>0.3</td>
<td>5.9</td>
</tr>
</tbody>
</table>
Alchemy: comparing to 45 manually curated gold MRSes for out-of-coverage items.

<table>
<thead>
<tr>
<th>System</th>
<th>F₁</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1214</td>
<td>14.28</td>
<td>96.58</td>
<td>14.20</td>
</tr>
<tr>
<td>bridge</td>
<td>44.99</td>
<td>82.91</td>
<td>41.56</td>
</tr>
<tr>
<td>bridge+ut</td>
<td>70.08</td>
<td>81.39</td>
<td>66.24</td>
</tr>
<tr>
<td>pacnv</td>
<td>30.61</td>
<td>86.12</td>
<td>30.75</td>
</tr>
<tr>
<td>pacnv+ut</td>
<td>39.87</td>
<td>86.55</td>
<td>40.01</td>
</tr>
<tr>
<td>csaw-tb</td>
<td>68.82</td>
<td>73.58</td>
<td>66.77</td>
</tr>
<tr>
<td>csaw-ww</td>
<td>76.97</td>
<td>77.54</td>
<td>76.63</td>
</tr>
</tbody>
</table>
Accuracy (disregarding parse failures)

Alchemy: comparing to 45 manually curated gold MRSes for out-of-coverage items.

<table>
<thead>
<tr>
<th>System</th>
<th>F₁</th>
<th>P</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1214</td>
<td>80.32</td>
<td>80.78</td>
<td>79.89</td>
</tr>
<tr>
<td>bridge</td>
<td>67.49</td>
<td>74.36</td>
<td>62.34</td>
</tr>
<tr>
<td>bridge+ut</td>
<td>75.09</td>
<td>80.06</td>
<td>70.97</td>
</tr>
<tr>
<td>pacnv</td>
<td>68.88</td>
<td>68.77</td>
<td>69.19</td>
</tr>
<tr>
<td>pacnv+ut</td>
<td>74.75</td>
<td>74.78</td>
<td>75.02</td>
</tr>
<tr>
<td>csaw-tb</td>
<td>70.39</td>
<td>72.98</td>
<td>68.28</td>
</tr>
<tr>
<td>csaw-ww</td>
<td>76.97</td>
<td>77.54</td>
<td>76.63</td>
</tr>
</tbody>
</table>
Extrinsic Evaluation: Robot Control Language

SemEval 2014 Task 6 Dukes [2014], Packard [2014]
Convert an English command into RCL, using (part of) the winning system from the shared task.

<table>
<thead>
<tr>
<th>Parser</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1214</td>
<td>55.8</td>
</tr>
<tr>
<td>bridge</td>
<td>42.7</td>
</tr>
<tr>
<td>bridge+ut</td>
<td>54.5</td>
</tr>
<tr>
<td>pacnv</td>
<td>37.2</td>
</tr>
<tr>
<td>pacnv+ut</td>
<td>51.6</td>
</tr>
<tr>
<td>csaw-tb</td>
<td>17.9</td>
</tr>
<tr>
<td>csaw-ww</td>
<td>47.2</td>
</tr>
</tbody>
</table>
Thank You!

References:


