Cascaded Finite-State Parsing

Steven Abney
University of Tübingen
Cascaded Finite-State Parsing

- Grammar divided in strata/levels (chunks & clauses)
- Pipeline of finite-state recognizers/transducers

the woman in the lab coat thought you were sleeping
Themes

• Robustness
  – Local decisions, not global optimization
    * No closed-world assumption with consequent fragility
  – Easy-first parsing
    * High-precision (low-entropy) decisions
    * Acceptable error rates even without search

• Speed
  – Deterministic, not exhaustive search

• Reorganizing the decision space
  – Islands of certainty
  – Containment of ambiguity
  – Divorcing control structure from parse structure
• Bootstrapping
Why Chunking Helps with Valencies

- Fewer input units
- Smaller domain
  \[ S [\text{NX the woman}] [\text{PX in the lab coat}] [\text{VX thought}] \]
  \[ S [\text{NX you}] [\text{VX were sleeping}] \]
- Statistics filter out noise
  - Adjuncts—*don’t appear specially with V*
  - Noun args—*don’t appear reliably with V*
Trading Off Speed and Accuracy

- No search → faster, less accurate
- Sometimes faster & more accurate

\[
S \rightarrow b A B : p_{S1} \\
| c B C : p_{S2} \\
| d D A : q_S = 1 - (p_{S1} + p_{S2})
\]

\[
A \rightarrow a : p_A \\
| a a : q_A = 1 - p_A
\]

\[
B \rightarrow a : p_B \\
| a a : q_B = 1 - p_B
\]

\[
C \rightarrow a : p_C \\
| a a : q_C = 1 - p_C
\]

\[
D \rightarrow a : p_{D1} \\
| a a : p_{D2} \\
| a a a : q_D = 1 - (p_{D1} + p_{D2})
\]
Tree Set

S
 b A B

c B C

d D A

A B
 a a

A B
 a a a

A B
 a a a a

B C
 a a

B C
 a a a

B C
 a a a a

D A
 a a

D A
 a a a

D A
 a a a a
Maximum Likelihood Estimate

- Corpus $C$ = one of each longest-match tree
- Likelihood

$$L(C; p) = [p_{S_1} p_{A} p_B \cdot p_{S_1} q_A p_B \cdot p_{S_1} q_A q_B] \cdot [p_{S_2} p_{B} p_C \cdot p_{S_2} q_B p_C \cdot p_{S_2} q_B q_C] \cdot [q_S p_D p_A \cdot q_S p_D p_A \cdot q_S q_D p_A \cdot q_S q_D q_A]$$
- Maximize $\frac{\partial}{\partial p_i} \ln L(C; p)$

  - Result: probabilities are relative counts
  - E.g.: $p_{S_1} = \frac{\#(S \rightarrow baB)}{\#(S \rightarrow)} = \frac{3}{10}$
  - E.g.: $p_A = \frac{\#(A \rightarrow a)}{\#(A \rightarrow)} = \frac{4}{7}$
Going Astray

\[
\begin{align*}
S & \rightarrow b \ A \ B : 3/10 \quad | \quad c \ B \ C : 3/10 \quad | \quad d \ D \ A : 4/10 \\
A & \rightarrow a : 4/7 \quad | \quad a \ a : 3/7 \\
B & \rightarrow a : 3/6 \quad | \quad a \ a : 3/6 \\
C & \rightarrow a : 2/3 \quad | \quad a \ a : 1/3 \\
D & \rightarrow a : 1/4 \quad | \quad a \ a : 1/4 \quad | \quad a \ a \ a : 1/2
\end{align*}
\]

\[
\begin{align*}
\text{t}_1 &= \begin{array}{c}
S \\
\quad b \\
\quad A \quad B \\
\quad a \quad a \quad a
\end{array} \\
\text{t}_2 &= \begin{array}{c}
S \\
\quad b \\
\quad A \quad B \\
\quad a \quad a \quad a
\end{array}
\end{align*}
\]

\[
p(t_1) = \left( \frac{3}{10} \right) \left( \frac{4}{7} \right) \left( \frac{1}{2} \right) \\
p(t_2) = \left( \frac{3}{10} \right) \left( \frac{3}{7} \right) \left( \frac{1}{2} \right)
\]

\[
p(t_1|baaa) = \frac{4}{7} \\
p(t_2|baaa) = \frac{3}{7}
\]
Finite-State Cascade

- Finite-State Cascade

\[
L_3 \quad \text{---------------------------} \quad S \quad \text{---------------------------} \quad S \\
T_3
\]

\[
L_2 \quad \text{--------} \quad NP \quad PP \quad VP \quad NP \quad VP
T_2
\]

\[
L_1 \quad \text{--------} \quad NP \quad P \quad NP \quad VP \quad NP \quad VP
T_1
\]

\[
L_0 \quad \text{--------} \quad D \quad N \quad P \quad D \quad N \quad N \quad V-\text{tns} \quad Pron \quad Aux \quad V-ing
\]

*the woman in the lab coat thought you were sleeping*

- Regular-Expression Grammar

\[
L_1 : \begin{cases} 
\text{NP} \rightarrow \text{D? N}\ast \text{N} \\
\text{VP} \rightarrow \text{V-\text{tns}} \mid \text{Aux V-ing} 
\end{cases}
\]

\[
L_2 : \{ \text{PP} \rightarrow \text{P NP} \}
\]

\[
L_3 : \{ \text{S PP}\ast \text{NP PP}\ast \text{VP PP}\ast \}
\]
• At level $L_k$: take union, determinize
Longest Match

the woman in the lab coat thought
Grammar

• Tagfixes \((word, tag) \rightarrow tag\)

• Measure Phrase \(Dates and times, numerals, predeterminers, dollar amounts\)

• Chunks \(Noun, adjective, adverb, verb, infinitive chunks\)

• N-Mess \(Unassembled noun chunk pieces\)

• NP: Possessors

• NG: Coordinated NP’s

• PP

• RC \(Center embedding\)

• C0 \(Subj+Pred: bleed Clause\)

• Clause
Measure Phrase Level

```plaintext
date → month cd (cma cd cma?)?;
cdqlx → as much as | more than | about | only | well? over;
cdx → (cdql | cdqlx) cd | (cdql | cdqlx)? cd+ cd;
doll → cdqlx? dol cd;
ci-st → nnp+ cma place cma?;
xm → cdx h=(units | tunits)
    | (cdql | cdqlx)? dt-a h=(unit | tunit)
    ;
timex → dtp h=tunit;
tadvx → ( cdx h=(nns | units | tunits)
       | (cdql | cdqlx)? dt-a h=(nn | unit | tunit)
        )
    ago
    ;
person → tt? i* fn (i | nnp)*
       | tt (i | nnp)*
       ;
name → (nnp | nnps | i)+;
```

Noun Chunk

PROPER = place | person | name | ci-st | doll;
COMMON = nn | nns | month | unit | units | tunit | tunits;
N = PROPER | COMMON | date;
ADVHD = rb | cdql | then | well;
ADV = ADVHD | rbr | more | rbs | ql;
VADVP = ADV* (ADVHD | only);
JX = ADV? (jj | jjr | jjs);
JXC = JX (cma JX)* (cc | cma) JX;
ADJ = JX | JXC | mx;
PTC = (ADV | rbr | more | rbs)? (vbn | vbg);
DET = dt | dtp | prp$ | (cdql | cdqlx)? (dt-a | dt-q | dtp-q);
NUM = cd | cdx;
MX = mx | units | tunits;

nx → DET? NUM? (ADJ | PTC)* (ADJ | N)* h=COMMON cd?
   | DET NUM? (ADJ | PTC)* h=PROPER
   | DET h=(jjr | jj)
   | cdql? h=dtp-q
   | h=( prp | cd | dtp | cd | dtp | qq | ex
      | name | person | date | doll | ci-st | rbr | rbs
   )
;
Verb Chunk

VB-TNS = vb | vbp | vbz | vbd;
DO-TNS = do | doz | dod;
HV-TNS = hv | hvz | hvd;
BE-TNS = be | bem | bez | bedz | bed | ber | bedr;
MODAL = md | doz | do | dod;
VP-PASS = VADVP? (vbn | vbd | hvn | ben);
VP-PROG = VADVP? (vbg | hvg | beg (VP-PASS | ax)?);
VP-PERF = VADVP? (vbn | hvn | ben (VP-PROG | VP-PASS | ax)?);
VP-INF = VADVP? (vb | do | hv VP-PERF? | be (VP-PROG | VP-PASS | ax)?);

vx → VADVP? ( md VP-INF?
                  | DO-TNS VP-INF?
                  | VB-TNS
                  | HV-TNS VP-PERF?
                  | BE-TNS (VP-PROG | VP-PASS | ax)?
              )

; inf → VADVP? to VADVP? VP-INF;
Minor Chunks

\[
\text{perx} \rightarrow ( (cd | cdx | dt-a) (\text{COMMON} | \text{doll}) \mid \text{MX} ) \mid (dt-a | \text{per}) \text{COMMON};
\]

\[
\text{rx} \rightarrow \text{ADV+ ADV} \mid \text{by then} \mid \text{MX ago} ;
\]

\[
\text{ax} \rightarrow (\text{ADV} | \text{rx})* \text{jj} \mid \text{MX JX} \mid \text{VADVP?} (\text{vbn} | \text{vbg}) ;
\]
Larger Noun Phrases

\[
\begin{align*}
\text{DET} & = \text{dt} \mid \text{dtp} \mid \text{prp}\$ \text{cdql}\? (\text{dt-q} \mid \text{dtp-q}) ; \\
\text{NOM} & = \text{nx} \mid \text{mx} \mid \text{cdx} \mid \text{place} \mid \text{person} \mid \text{name} \mid \text{ci-st} \mid \text{doll} \mid \text{date} ; \\
\text{NP} & = (\text{NOM} \mid \text{np} \mid \text{nmsg}) (\text{of} (\text{NOM} \mid \text{np} \mid \text{nmsg}))^{*} ; \\
\text{CONNECT} & = \text{cma} \mid \text{cc} \mid \text{cma cc} ;
\end{align*}
\]

- **nmsg:**
  \[
  \text{nmsg} \rightarrow \text{DET} \text{ax}^{*} \text{ h=NOM}? ;
  \]

- **np:**
  \[
  \text{np} \rightarrow \text{only?} (\text{NOM} \mid \text{nmsg}) \text{ pos h=NOM}? ;
  \]

- **ng:**
  \[
  \text{ng} \rightarrow \text{h=NOM} (\text{of} \text{NOM})^{+} \\
  \mid \text{h=NOM} (\text{of} \text{NOM})^{*} \text{ cc} \text{ NP} \\
  \mid \text{h=NOM} (\text{of} \text{NOM})^{*} (\text{CONNECT} \text{ NP})^{+} \text{ cma? cc} \text{ NP} \\
  ;
  \]

Prepositional and Larger Verbal Phrases

\[
\text{PREP} = \text{in} | \text{by} | \text{to} | \text{of} | \text{than} | \text{as} | \text{cdqlx} | \text{according to} | \text{because of};
\]
\[
\text{NOM} = \text{ng} | \text{nmess} | \text{np} | \text{nx} | \text{mx} | \text{cdx} | \text{place} | \text{person} | \text{name} | \text{ci-st} | \text{doll}
\]
\[
\text{PRED-TAIL} = \text{NOM? (pp | pp-comp | rx)*};
\]
\[
\begin{align*}
\text{pp} & \rightarrow \text{prep h=NOM}; \\
\text{pp-comp} & \rightarrow \text{p-comp h=NOM}; \\
\text{infp} & \rightarrow \text{h=inf PRED-TAIL}; \\
\text{vbnp} & \rightarrow \text{h=(vbn | vnx) (pp | pp-comp | rx)*}; \\
\text{vbgp} & \rightarrow \text{h=(vbg | vx) PRED-TAIL}; \\
\text{cc-vp} & \rightarrow \text{cc h=vp PRED-TAIL};
\end{align*}
\]
Clauses

NOM = ng | nmess | np | nx | mx | cdx | place | person | name | ci-st | doll
WH = wdt | wp;
SUBJ-TAIL = pp* ( cma? (src | orc | infp | vbnp | vbgp) cma?
            | cma NOM cma
            ) ?
PRED-TAIL = NOM? (pp | pp-comp | rx)*;

• rc:
  src → WH h=vp NOM? (pp | rx)*;
  orc → WH (ng | np | nx) pp* h=vp (pp | rx)*;

• c0:
  c0 → NOM SUBJ-TAIL h=vp;
  subc0 → (pp-comp | (comp | because) NOM) SUBJ-TAIL h=vp;

• s:
  c → h=c0 PRED-TAIL;
  subc → h=subc0 PRED-TAIL;
• Features

• Internal phrases

\[
\text{Subj} \rightarrow [\text{NP } num = D? \ num = [\text{NBAR } A* N* \ num = N ] ] V
\]

• Take \(\epsilon\)-closure of left automaton \(\rightarrow\) same recognizer as before
At Compile Time

• Transducer output = Action
  – ] Create node, set end position
  – [ Set start position, attach children
  – ftr= Assign feature

• Sort actions
  – Right bracket, left bracket, feature assignments
  – Inside to outside

\[
\text{Subj} \rightarrow [\text{NP } num=1 \text{ D? } num=2 \ [\text{Nbar A* N* } num=3 \text{ N } ]_{\text{Nbar}}]_{\text{Nbar}}
\]

\[
5 \quad 6 \quad 7 \quad 2 \quad 3 \quad 1
\]
Sorted Actions

Subj → [NP \(num=1\) D? \(num=2\) [Nbar A* N* \(num=3\) N ]Nbar

\[
\begin{array}{cccccc}
5 & 6 & 7 & 2 & 3 & 1
\end{array}
\]

\[\begin{array}{c}
\text{1}Nbar \\
\text{1}Nbar \\
\text{num}=3 \\
\text{1}NP \\
\text{1}NP \\
\text{num}=1 \\
\text{num}=2
\end{array}\]
Run Transducer
Set Positions for Actions

- If same action appears at multiple positions, take last
Execute Actions
Wrap Up

- Restart at end of NP, *before* $V$
- No unification failures—bit operations
- Disambiguation—postpone actions
- Divorce control from structure
Evaluation

• Utility

“Chunks as defined by this grammar are good because they improve the performance of our IR/MT/... system”

“Trees as defined by this treebank are good because ...”

“Selectional restrictions as defined in this stylebook are good because they improve performance at resolving attachment ambiguities”

– Of grammar/treebank/stylebook/specification/theory
– Custom statistical model vs. Intuitive classes—greater utility for specific application

• Accuracy

– At meeting specification
– Upper bound: interjudge agreement
Sample from Corpus Positions

<table>
<thead>
<tr>
<th>corp.tag:</th>
<th>smp:</th>
<th>smp.conc:</th>
</tr>
</thead>
<tbody>
<tr>
<td>In in an dt</td>
<td>82900</td>
<td>em directly comparable, each index i says Krys Spain, research speciali</td>
</tr>
<tr>
<td>Oct. month 19 cd</td>
<td>102153</td>
<td>x lower for the next fiscal year</td>
</tr>
<tr>
<td>review nn of of</td>
<td>133333</td>
<td>imitated at $193, according to House</td>
</tr>
<tr>
<td>“ nil The dt Misanthrope nn” nil at in Chicago nnp ’s pos Goodman nnp Theatre nnp</td>
<td>63092</td>
<td>only about 8,000 are written for NEC</td>
</tr>
<tr>
<td></td>
<td>121520</td>
<td>it stopped advertising its name</td>
</tr>
<tr>
<td></td>
<td>60127</td>
<td>estimated at $6,0127</td>
</tr>
<tr>
<td></td>
<td>3805</td>
<td>male sterile and herbicide</td>
</tr>
<tr>
<td></td>
<td>147276</td>
<td>ry in share prices. // The dollar f</td>
</tr>
<tr>
<td></td>
<td>66527</td>
<td>o the world that now is the time to g</td>
</tr>
<tr>
<td></td>
<td>76930</td>
<td>h defensive issues as food, tobacco</td>
</tr>
<tr>
<td></td>
<td>113209</td>
<td>lle, Ky. ) – David R. Jackson, for</td>
</tr>
<tr>
<td></td>
<td>21391</td>
<td>ireproofing concern said the transact</td>
</tr>
<tr>
<td></td>
<td>121081</td>
<td>rationships rather than complex finan</td>
</tr>
<tr>
<td></td>
<td>137109</td>
<td>n was being “ conservative ” in his</td>
</tr>
<tr>
<td></td>
<td>25825</td>
<td>% less full fees via Nikko Securities</td>
</tr>
</tbody>
</table>
Scoring

test file          = s3.tst
std file           = s3.std
writing diff file  = s3.diff

sample size:       N = 1000
answers in common: X = 921
nonzeros in test:  t = 390
nonzeros in std:   s = 394
nonzeros in common:x = 343

per-word accuracy: \( \frac{X}{N} = \frac{921}{1000} = 92.1 \pm 1.7 \% \)
precision:         \( \frac{x}{t} = \frac{343}{390} = 87.9 \pm 3.2 \% \)
recall:            \( \frac{x}{s} = \frac{343}{394} = 87.1 \pm 3.3 \% \)
**Speed**

- Factor 1: fast machine
- Factor 2: determinism
- Factor 3: write C

<table>
<thead>
<tr>
<th>Program</th>
<th>depth</th>
<th>sw</th>
<th>hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fidditch3</td>
<td>parse</td>
<td>C</td>
<td>SGI</td>
</tr>
<tr>
<td>Cass2</td>
<td>chunk</td>
<td>C</td>
<td>SparcELC</td>
</tr>
<tr>
<td>Copsy</td>
<td>np</td>
<td>Pascal</td>
<td>BS2000</td>
</tr>
<tr>
<td>Cass2</td>
<td>clause</td>
<td>C</td>
<td>SparcELC</td>
</tr>
<tr>
<td>CG</td>
<td>dep</td>
<td></td>
<td>Sparc10</td>
</tr>
<tr>
<td>Fidditch3</td>
<td>parse</td>
<td>C</td>
<td>Sun4</td>
</tr>
<tr>
<td>Pos</td>
<td>tag</td>
<td></td>
<td>Sun4</td>
</tr>
<tr>
<td>Fidditch2</td>
<td>parse</td>
<td>Lisp</td>
<td>Sun4</td>
</tr>
<tr>
<td>Cass1</td>
<td>chunk</td>
<td>Lisp</td>
<td>Sun4</td>
</tr>
<tr>
<td>Clarit</td>
<td>np</td>
<td>Lisp</td>
<td>Sun4</td>
</tr>
<tr>
<td>Fastus</td>
<td>chunk</td>
<td>Lisp</td>
<td>Sparc2</td>
</tr>
<tr>
<td>Cass1</td>
<td>chunk</td>
<td>Lisp</td>
<td>UX400S</td>
</tr>
<tr>
<td>Scisor</td>
<td>skim</td>
<td></td>
<td>UX400S</td>
</tr>
<tr>
<td>Fidditch1</td>
<td>parse</td>
<td>Lisp</td>
<td>Sym-36xx</td>
</tr>
<tr>
<td>McDonald</td>
<td>parse</td>
<td></td>
<td>MacII</td>
</tr>
<tr>
<td>Chupa</td>
<td>parse</td>
<td>Lisp</td>
<td>UX400S</td>
</tr>
<tr>
<td>Traditional</td>
<td>parse</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

- Finite-State Cascade, recognize then parse
- Speed: determinism
- Robustness
  - Local decisions
  - Easy-first
  - Islands of certainty
  - Containment of ambiguity
  - Divorcing control from structure