Transition-Based Parsing

Joakim Nivre

Uppsala University
Department of Linguistics and Philology
joakim.nivre@lingfil.uu.se
1. Transition Systems
2. Greedy Classifier-Based Parsing
3. Beam Search and Structured Learning
4. Non-Projective Parsing
Transition Systems

\[ S = (C, T, c_s, C_t) \]

1. \( C \) is a set of configurations
2. \( T \) is a set of transitions, each \( t : C \rightarrow C \)
3. \( c_s \) is an initialization function, \( c_s(x) \in C \)
4. \( C_t \subseteq C \) is a set of terminal configurations
\[ c = (\Sigma, B, A) \]

1. \( \Sigma \) is a list of nodes in \( V_x \), known as the \textit{stack}
2. \( B \) is a list of nodes in \( V_x \), known as the \textit{buffer}
3. \( A \) is a set of dependency arcs in \( V_x \times L \times V_x \)
Transition Systems

\[ C_{0,m} = (c_0, c_1, \ldots, c_m) \]

1. \( c_0 = c_s(x) \),
2. \( c_m \in C_t \),
3. for every \( i (1 \leq i \leq m) \), \( c_i = t(c_{i-1}) \) for some \( t \in T \)

\[ G_{c_m} = (V_x, A_{c_m}) \]
Initialization: \[ c_s(x = x_1, \ldots, x_n) = ([0], [1, \ldots, n], \emptyset) \]

Terminal: \[ C_t = \{ c \in C | c = ([0], [], A) \} \]

Transitions:

- \[(\sigma, [i|\beta], A) \Rightarrow ([\sigma|i], \beta, A)\] (Shift)
- \[([\sigma|i|j], B, A) \Rightarrow ([\sigma|j], B, A \cup \{(j, l, i)\})\] (Left-Arc\(_l\))
- \[([\sigma|i|j], B, A) \Rightarrow ([\sigma|i], B, A \cup \{(i, l, j)\})\] (Right-Arc\(_l\))

\(^1\) Permitted only if \(i \neq 0\).
[ROOT]_\Sigma \quad [\text{Economic, news, had, little, effect, on, financial, markets, .}]_B

\textbf{ROOT} \quad \text{Economic news had little effect on financial markets .}
[ROOT, Economic]_\Sigma [news, had, little, effect, on, financial, markets, .]_B

ROOT Economic news had little effect on financial markets .
[ROOT, Economic, news]_{\Sigma} [had, little, effect, on, financial, markets, .]_{B}

**ROOT** Economic news had little effect on financial markets .
[ROOT, news]_\sum_ \left[ \text{had, little, effect, on, financial, markets, .} \right]_B
[ROOT, news, had]_\Sigma \ [little, effect, on, financial, markets, .]_B

\hspace{1cm}

\begin{align*}
\text{ATT} \\
\text{ROOT} \quad \text{Economic news had little effect on financial markets .}
\end{align*}
[ROOT, had]_\Sigma [little, effect, on, financial, markets, .]_B

ROOT  Economic news had little effect on financial markets .
[ROOT, had, little]_\Sigma \quad [\text{effect, on, financial, markets, .}]_B

\begin{align*}
\text{ATT} & \quad \text{SBJ} \\
\text{ROOT} & \quad \text{Economic} \quad \text{news} \quad \text{had} \quad \text{little effect on financial markets .}
\end{align*}
[ROOT, had, little, effect]_Σ [on, financial, markets, .]_Β

ROOT Economic news had little effect on financial markets .
[\text{ROOT}, \text{had}, \text{effect}]_\Sigma \quad [\text{on, financial, markets, .}]_B

\textbf{ROOT} \quad \text{Economic news had little effect on financial markets .}
Transition Systems

\[
[\text{ROOT, had, effect, on}]_\Sigma \quad [\text{financial, markets, .}]_B
\]

\[
\text{ROOT} \quad \text{Economic news had little effect on financial markets .}
\]
[ROOT, had, effect, on, financial]_Σ [markets, .]_B

ROOT Economic news had little effect on financial markets .
Transition Systems

[ROOT, had, effect, on, financial, markets]Σ [.]B

ROOT  Economic news had little effect on financial markets .
[ROOT, had, effect, on, markets] \Sigma \ [.]_B

ROOT  Economic news had little effect on financial markets .
[ROOT, had, effect, on]_{\Sigma} \quad [.]_{B}

\[ \text{ROOT} \quad \text{Economic} \quad \text{news} \quad \text{had} \quad \text{little} \quad \text{effect} \quad \text{on} \quad \text{financial} \quad \text{markets} \]
Transition Systems

[ROOT, had, effect]_Σ [.]_B

ROOT  Economic news had little effect on financial markets .
[ROOT, had]_{\Sigma} \ [.]_{B}

ROOT \quad Economic \quad news \quad had \quad little \quad effect \quad on \quad financial \quad markets
[ROOT, had, .]Σ [ ]B

ROOT  Economic  news  had  little  effect  on  financial  markets .
[ROOT, had] \Sigma \ [ ] \_B

ROOT Economic news had little effect on financial markets.
\[ \text{ROOT} \Sigma \ [\ ]_B \]

**Diagram:**
- **ROOT:** Economic news had little effect on financial markets.
Transition Systems

Initialization: \( c_s(x = x_1, \ldots, x_n) = ([0], [1, \ldots, n], \emptyset) \)

Terminal: \( C_t = \{ c \in C | c = (\Sigma, [], A) \} \)

Transitions:

1. \((\sigma, [i|\beta], A) \Rightarrow ([\sigma|i], \beta, A)\) (Shift)
2. \(([\sigma|i], [j|\beta], A) \Rightarrow (\sigma, [j|\beta], A \cup \{(j, l, i)\})^1\) (Left-Arc\(_l\))
3. \(([\sigma|i], [j|\beta], A) \Rightarrow ([\sigma|i|j], \beta, A \cup \{(i, l, j)\})\) (Right-Arc\(_l\))
4. \(([\sigma|i], B, A) \Rightarrow (\sigma, B, A)^2\) (Reduce)

1 Permitted only if \( i \neq 0 \) and there are no \( k, l' \) such that \((k, l', i) \in A\).
2 Permitted only if there are \( k, l' \) such that \((k, l', i) \in A\).
[ROOT]Σ [Economic, news, had, little, effect, on, financial, markets, .]_B

ROOT Economic news had little effect on financial markets .
[ROOT, Economic]_{\Sigma} [news, had, little, effect, on, financial, markets, .]_B

**ROOT** Economic news had little effect on financial markets .
\[
\text{[ROOT]} \Sigma \quad \text{[news, had, little, effect, on, financial, markets, .]}_B
\]

\[
\text{ROOT} \quad \text{Economic news had little effect on financial markets .}
\]
\[ \text{ROOT, news}]_\Sigma [\text{had, little, effect, on, financial, markets, .}]_B \]

\begin{align*}
\text{ATT} \\
\text{ROOT} \quad \text{Economic} \quad \text{news} \quad \text{had} \quad \text{little} \quad \text{effect} \quad \text{on} \quad \text{financial} \quad \text{markets} \quad .
\end{align*}
Transition Systems

\[[\text{ROOT}]^{\Sigma} \quad \text{[had, little, effect, on, financial, markets, .]}_{B}\]

\[\text{ROOT} \quad \text{Economic news had little effect on financial markets .}\]
[ROOT, had]\Sigma [little, effect, on, financial, markets, .]_B
Transition-Based Parsing 9(19)

[ROOT, had, little]_\Sigma [effect, on, financial, markets, .]_B

ROOT  Economic news had little effect on financial markets .
[ROOT, had]_\Sigma [effect, on, financial, markets, .]_B
[ROOT, had, effect]_Σ [on, financial, markets, .]_B

 Transition Systems
[ROOT, had, effect, on]_Σ [financial, markets, .]_B

ROOT  Economic news had little effect on financial markets .
[ROOT, had, effect, on, financial]_Σ [markets, .]_B
\[[\text{ROOT}, \text{had, effect, on}]_\Sigma\ [\text{markets, .}]_B\]
$\text{[ROOT, had, effect, on, markets]} \sum \ [.]_B$

Transition Systems
[ROOT, had, effect, on]_Σ [.]_B
[ROOT, had, effect]_\Sigma [.]_B

ROOT Economic news had little effect on financial markets.
[ROOT, had] \Sigma \ [.]_B

ROOT: Economic news had little effect on financial markets.
\[ \text{ROOT, had, .}] \Sigma [ ]_B \]
Parse($x = (w_0, w_1, \ldots, w_n)$)
1 $c \leftarrow c_s(x)$
2 while $c \not\in C_t$
3 \hspace{1em} $t^* \leftarrow \text{argmax}_t \text{Score}(c, t)$
4 \hspace{1em} $c \leftarrow t^*(c)$
5 return $G_c$
Score$(c, t) = \sum_{k=1}^{K} f_k(c, t) \cdot w_k$
<table>
<thead>
<tr>
<th><strong>Unigrams</strong></th>
<th><strong>Bigrams</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Sigma_0.\text{pos}$</td>
<td>$\Sigma_0.\text{pos}$, $B_0.\text{pos}$</td>
</tr>
<tr>
<td>$\Sigma_1.\text{pos}$</td>
<td>$\Sigma_0.\text{pos}$, $\Sigma_0.\text{lab}$</td>
</tr>
<tr>
<td>$B_0.\text{pos}$</td>
<td>$B_0.\text{pos}$, $\text{ldep}(B_0).\text{lab}$</td>
</tr>
<tr>
<td>$B_1.\text{pos}$</td>
<td></td>
</tr>
<tr>
<td>$B_2.\text{pos}$</td>
<td></td>
</tr>
<tr>
<td>$B_3.\text{pos}$</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_0.\text{lab}$</td>
<td></td>
</tr>
<tr>
<td>$\text{ldep}(\Sigma_0).\text{lab}$</td>
<td></td>
</tr>
<tr>
<td>$\text{rdep}(\Sigma_0).\text{lab}$</td>
<td></td>
</tr>
<tr>
<td>$\text{ldep}(B_0).\text{lab}$</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_0.\text{word}$</td>
<td></td>
</tr>
<tr>
<td>$B_0.\text{word}$</td>
<td></td>
</tr>
<tr>
<td>$B_1.\text{word}$</td>
<td></td>
</tr>
<tr>
<td>$\text{hd}(\Sigma_0).\text{word}$</td>
<td></td>
</tr>
<tr>
<td><strong>Trigrams</strong></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_1.\text{pos}$, $\Sigma_0.\text{pos}$, $B_0.\text{pos}$</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_0.\text{pos}$, $B_0.\text{pos}$, $B_1.\text{pos}$</td>
<td></td>
</tr>
<tr>
<td>$B_0.\text{pos}$, $B_1.\text{pos}$, $B_2.\text{pos}$</td>
<td></td>
</tr>
<tr>
<td>$B_1.\text{pos}$, $B_2.\text{pos}$, $B_3.\text{pos}$</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_0.\text{pos}$, $\text{ldep}(\Sigma_0).\text{lab}$, $\text{rdep}(\Sigma_0).\text{lab}$</td>
<td></td>
</tr>
</tbody>
</table>
Beam Search and Structured Learning

\[ \text{Score}(c_0, m, x) = \sum_{i=0}^{m-1} \text{Score}(c_i, t_i) \quad \text{[where } t_i(c_i) = c_{i+1}] \]
Parser($x = (w_0, w_1, \ldots, w_n)$)
1. $\text{Beam} \leftarrow \{c_s(x), 0.0\}$
2. while $\exists (c, s) \in \text{Beam} : c \not\in C_t$
3. $\text{NewBeam} \leftarrow \emptyset$
4. for every $(c, s) \in \text{Beam}$
5. for every $t \in T$
6. $\text{NewBeam} \leftarrow \text{NewBeam} \cup \{t(c), s + \text{Score}(c, t)\}$
7. $\text{Beam} \rightarrow \text{QBest}(\text{NewBeam})$
8. return $\leftarrow \text{1Best}(\text{Beam})$
Training data: \( \mathcal{T} = \{(x^i, c^i_{0,m})\}_{i=1}^{\lvert \mathcal{T} \rvert} \)

1. \( w \leftarrow 0 \)
2. \textbf{for} \( n : 1..N \)
3. \textbf{for} \( i : 1..\lvert \mathcal{T} \rvert \)
4. \( c^*_0,m \leftarrow \text{Parse}(x^i, w) \)
5. \textbf{if} \( c^*_0,m \neq c^i_{0,m} \)
6. \( w \leftarrow \text{Update}(w, c^*_0,m, c^i_{0,m}) \)
7. \textbf{return} \( w \)
Update($w, c^*_0, m, c^i_0, m$)

1 \textbf{for} $k : 1..K$
2 \quad \textbf{for} $i : 0..m - 1$
3 \quad \quad w_k \leftarrow w_k - f_k(c_i, t_i)$
4 \quad \textbf{for} $i : 0..m - 1$
5 \quad \quad w_k \leftarrow w_k + f_k(c_i, t_i)$
Initialization: 
\[ c_s(x = x_1, \ldots, x_n) = ([0], [1, \ldots, n], \emptyset) \]

Terminal: 
\[ C_t = \{ c \in C | c = ([0], [], A) \} \]

Transitions:
- (\[ \sigma, [i|\beta], A \] \Rightarrow ([\sigma|i], \beta, A) \) (Shift)
- ([\sigma|i], B, A) \Rightarrow ([\sigma], B, A \cup \{(j, l, i)\})^1 (Left-Arc_1)
- ([\sigma|i], B, A) \Rightarrow ([\sigma|i], B, A \cup \{(i, l, j)\}) (Right-Arc_1)
- ([\sigma|i]|k|j], B, A) \Rightarrow ([\sigma], B, A \cup \{(j, l, i)\})^1 (Left-Arc2_1)
- ([\sigma|i]|k|j], B, A) \Rightarrow ([\sigma|i], B, A \cup \{(i, l, j)\}) (Right-Arc2_1)
- ([\sigma|i]|k_1|k_2|j], B, A) \Rightarrow ([\sigma], B, A \cup \{(j, l, i)\})^1 (Left-Arc3_1)
- ([\sigma|i]|k_1|k_2|j], B, A) \Rightarrow ([\sigma|i], B, A \cup \{(i, l, j)\}) (Right-Arc3_1)

^1 Permitted only if \( i \neq 0 \).
Non-Projective Parsing

Initialization: \( c_s(x = x_1, \ldots, x_n) = ([0], [1, \ldots, n], \emptyset) \)

Terminal: \( C_t = \{ c \in C | c = ([0], [], A) \} \)

Transitions: 
- \((\sigma, [i|\beta], A) \Rightarrow ([\sigma|i], \beta, A)\) \hspace{1cm} \text{(Shift)}
- \(([\sigma|i]|j], B, A) \Rightarrow ([\sigma|j], B, A \cup \{(j, l, i)\})\) \hspace{1cm} \text{(Left-Arc)}
- \(([\sigma|i]|j], B, A) \Rightarrow ([\sigma|i], B, A \cup \{(i, l, j)\})\) \hspace{1cm} \text{(Right-Arc)}
- \(([\sigma|i]|j], \beta, A) \Rightarrow ([\sigma|j], [i|\beta], A)\) \hspace{1cm} \text{(Swap)}

1 Permitted only if \( i \neq 0 \).
2 Permitted only if \( i \neq 0 \) and \( i < j \).
A hearing is scheduled on the issue today.