Master-Praktikum: Parsing linear context-free rewriting systems

Thomas Ruprecht

April 30, 2018

Overview

- Weighted LCFRS
- Range instantiation
- Weighted deduction systems
- Parsing with deduction systems
  - Overview
  - CYK parser
  - Naive parser
  - Active parser
- Implementation results
- Conclusion

Weighted LCFRS [Vijay-Shanker, Weir, and Joshi 1987; Denkinger 2016]

- string LCFRS as case of MCFG
- rules of form $A \to \langle a x_{1,1}, c x_{1,2} \rangle (A)$
  - composition representation implies function using string substitution
  - linear and non-deleting composition
- rule weights $p_G : P \to W$
  - semiring of weights
  - weight of a derivation: product of rule weights

$$p_G\left(\begin{array}{c} A \to f(A_1, \dots, A_k) \\ \begin{array}{ccc} A_1 & \cdots & A_k \\ \triangle & & \triangle \end{array} \end{array}\right) = p_G(A \to f(A_1, \dots, A_k)) \cdot \prod_{i=1}^{k} p_G\left(\begin{array}{c} A_i \\ \triangle \end{array}\right)$$

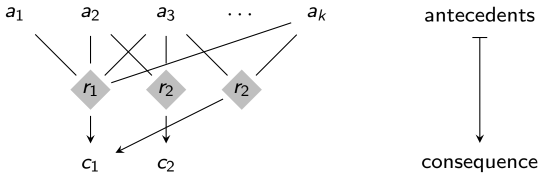
  - weight of a derived word: sum of derivations yielding word

Ranges and instantiations

- range: pair of indices limiting subword, e.g. ${}_{0}\text{aa}_{2}\text{bccd} \to (0, 2)$
- range concatenation: $(i, j)(j, k) = (i, k)$
- range vector: sequence of non-overlapping ranges
- instantiated composition: replace terminal sequences with ranges in word
  - $\langle a x_{1,1}, c x_{1,2} \rangle_{aabccd} = \{ \langle (0, 1) x_{1,1}, (3, 4) x_{1,2} \rangle, \langle (0, 1) x_{1,1}, (4, 5) x_{1,2} \rangle, \dots \}$
  - is ambiguous
  - implies partial function using range concatenation

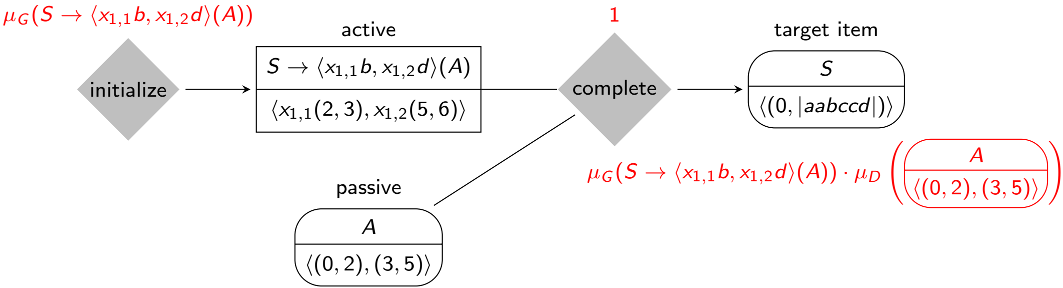
Weighted deduction [Shieber, Schabes, and Pereira 1995]

- rule application



- weighted rules $\mu_D(r)$
- weighted items $\mu_{D,\mathcal{C}}(c_1) = \mu_D(r_1) \cdot \mu_{D,\mathcal{A}}(a_1) \cdot \ldots \cdot \mu_{D,\mathcal{A}}(a_k)$
  $+ \mu_D(r_2) \cdot \mu_{D,\mathcal{A}}(a_3) \cdot \mu_{D,\mathcal{A}}(a_3)(a_k)$
- enumerate items using Knuth's algorithm [Nederhof 2003]
- heuristics [Angelov and Ljunglöf 2014]

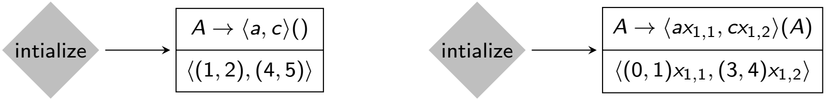
Deductive parsing: common behavior



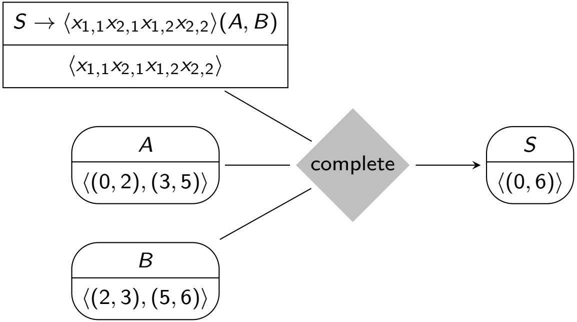
- parsing items
- weight paid on initialization
- completion is free
- heuristic: approx. cost for target item via inide/outside weights [Angelov and Ljunglöf 2014]

CYK parsing [Seki et al. 1991]

- *initialize*



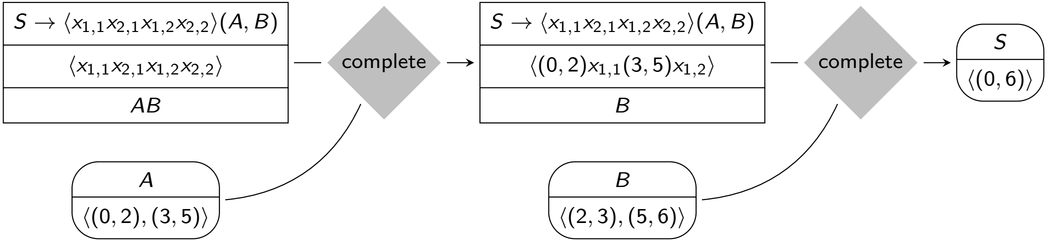
- *complete*



Naive-active parsing [Burden and Ljunglöf 2005]

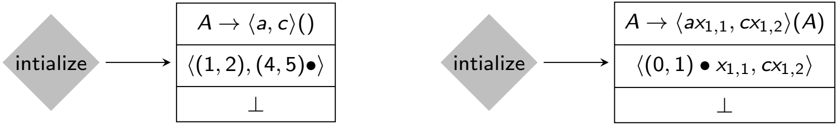
- *initialize*

- *complete*

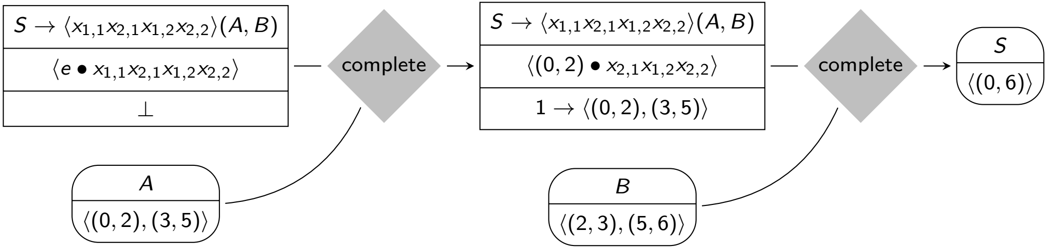


Active parsing [Burden and Ljunglöf 2005]

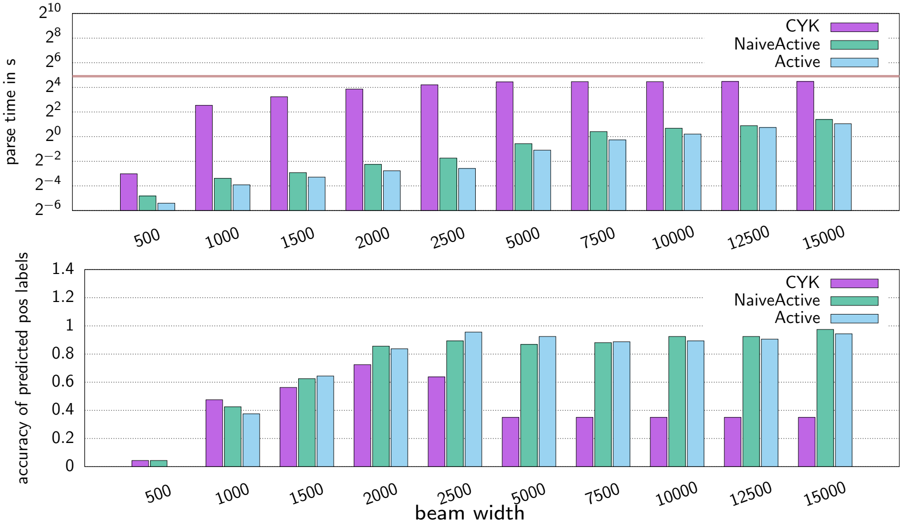
- *initialize*



- *complete*

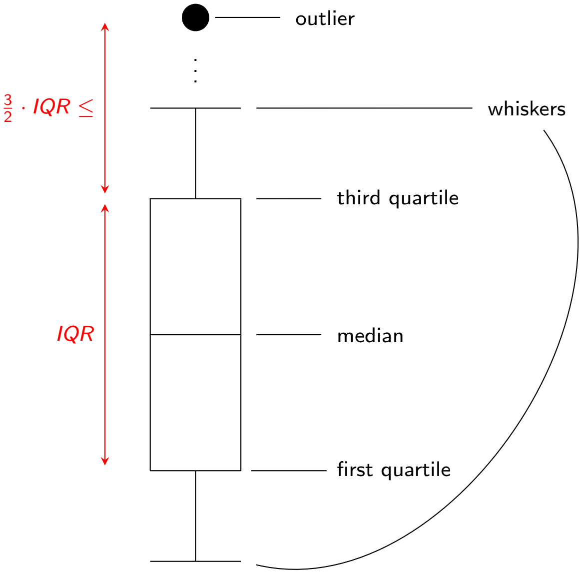


Results I – Beam search

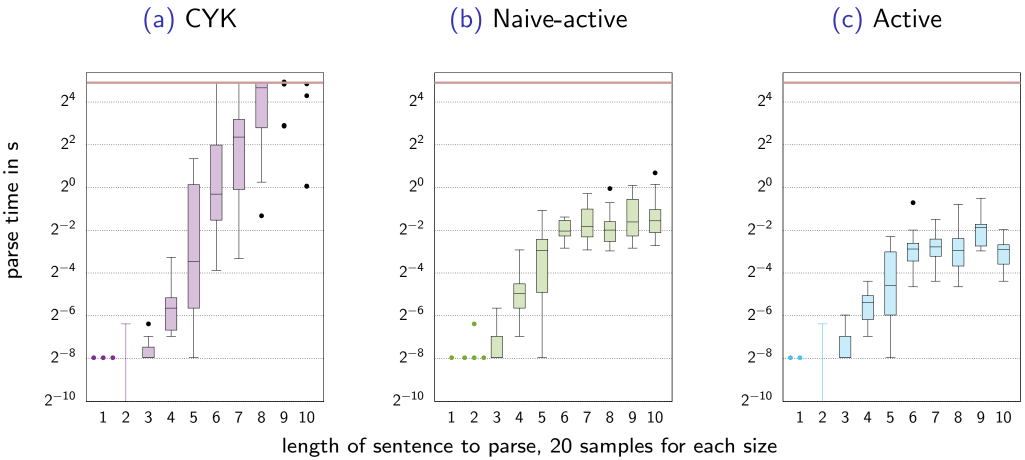


- average over 20 sentences with 8 tokens
- grammar extracted from entire NeGra corpus

Short interlude: box plots

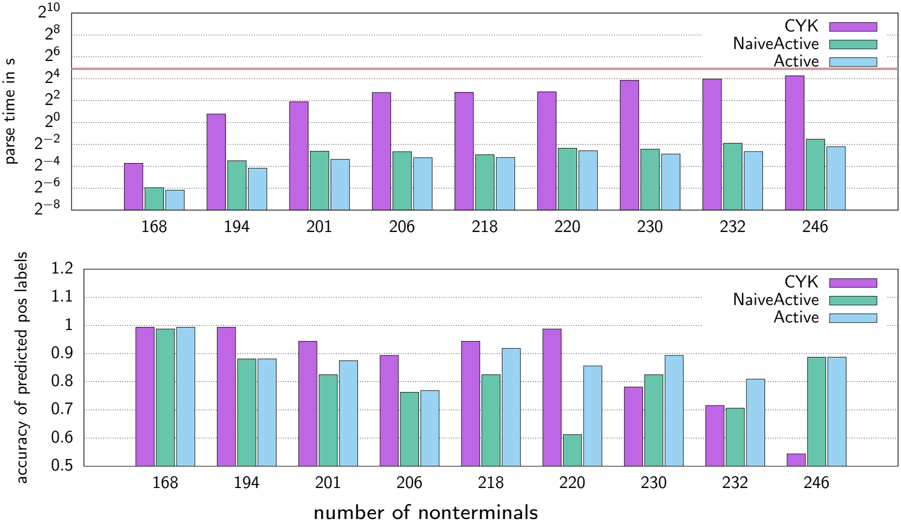


Results II – parse time for fixed sentence lengths



- average over 20 sentences for each length
- grammar from entire NeGra corpus
- beam with 2500

Results III – Grammar size



- average over 20 sentences with 8 tokens
- 9 random sized grammars from NeGra and one from entire corpus
- beam width 2500

Conclusions

- implemented
  - deduction systems for parsers,
  - Knuth's algorithm,
  - three parsers

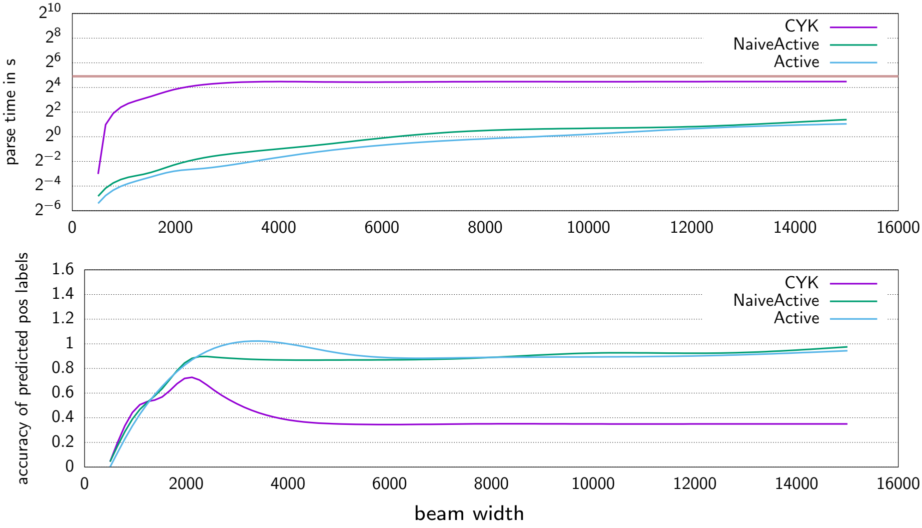
  in Vanda
- used implementation to parse German sentences in NeGra
- best performance using active parser
- CYK parser performed significantly worse
- possible future work:
  - implement incremental parser
  - extend implementation for (P)MCFG

References

- Angelov, Krasimir and Peter Ljunglöf (2014). "Fast Statistical Parsing with Parallel Multiple Context-Free Grammars". In: *Proc. of the 14th Conference of the European Chapter of ACL*.
- Burden, Håkan and Peter Ljunglöf (2005). "Parsing Linear Context-free Rewriting Systems". In: *Proc. of the Ninth International Workshop on Parsing Technology*. ACL.
- Denkinger, Tobias (2016). "An Automata Characterisation for Multiple Context-Free Languages". In: *Proc. of DLT 2016*. Springer Berlin Heidelberg.
- Nederhof, Mark-Jan (2003). "Weighted deductive parsing and Knuth's algorithm". In: *Computational Linguistics* 29.1.
- Seki, Hiroyuki et al. (1991). "On multiple context-free grammars". In: *Theoretical Computer Science* 88.2.
- Shieber, Stuart M., Yves Schabes, and Fernando C. N. Pereira (1995). "Principles and Implementation of Deductive Parsing". In: *J. Log. Program.* 24.1&2.
- Vijay-Shanker, Krishnamurti, David J Weir, and Aravind K Joshi (1987). "Characterizing structural descriptions produced by various grammatical formalisms". In: *Proc. of the 25th annual meeting on ACL*. ACL.

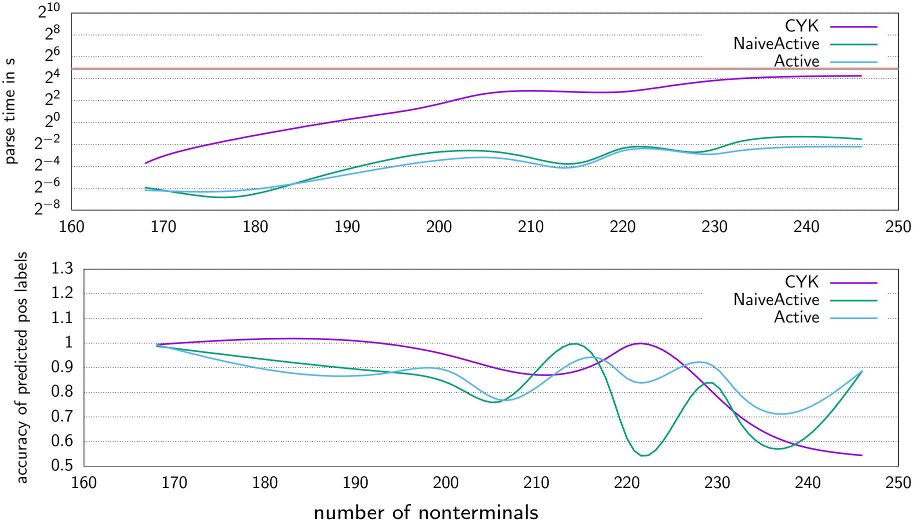
BACKUP

Results I – Beam search



- average over 20 sentences with 8 tokens
- grammar extracted from entire NeGra corpus

Results III – Grammar size



- average over 20 sentences with 8 tokens
- 9 random sized grammars from NeGra and one from entire corpus
- beam width 2500

Knuth's Algorithm

```
knuth(D = (I, R)):
  Chart ← ∅

  Agenda ← {(r(ε), (r, ε)) | r ∈ R}

  while Agenda ≠ ∅:
    (trigger, backtrace), Agenda ← pop(Agenda)

    if not trigger ∈ dom(Chart):
      Chart ← Chart{trigger/backtrace}
      Agenda ← Agenda ∪{(r(α), (r, α)) | r ∈ R, α ∈ dom(Chart)* : trigger in α}
    else:
      Chart ← Chart{trigger/backtrace}

  return Chart
```

CYK parsing

- two types rules
  - initial active items of nonterminals

$$\frac{}{[A \to f(A_1, \ldots, A_k), \phi]} \forall A \to f(A_1, \ldots, A_k) \in P, \phi \in f_w$$

  - complete all nonterminals at once using passive items

$$\frac{[A \to f(A_1, \ldots, A_k), \phi]\,(A_1, \rho_1) \ldots (A_k, \rho_k)}{(A, \phi(\rho_1, \ldots, \rho_k))}$$

- weights
  - $\mu_{D_{CYK}(G,w)}([A \to f(A_1, \ldots, A_k), \phi]) = p_G(A \to f(A_1, \ldots, A_k))$
  - $\alpha_{D_{CYK}(G,w)}([A \to f(A_1, \ldots, A_k), \phi]) = \alpha_{D_{CYK}(G,w)}((A, \psi)) = \alpha_G(A)$
  - $\beta_{D_{CYK}(G,w)}([A \to f(A_1, \ldots, A_k), \phi]) = \bigodot_{i=1}^{k} \beta_G(A_i)$

CYK parsing - example

$$
\begin{aligned}
&\emptyset \\
\vdash_{D_{CYK}(G,w)} &\left\{ \begin{bmatrix} A \rightarrow \langle \varepsilon, \varepsilon \rangle \\ \langle \varepsilon, \varepsilon \rangle \end{bmatrix}, \begin{bmatrix} A \rightarrow \langle a x_{1,1}, c x_{1,2} \rangle (A) \\ \langle (0,1) x_{1,1}, (2,3) x_{1,2} \rangle \end{bmatrix}, \begin{bmatrix} B \rightarrow \langle \varepsilon, \varepsilon \rangle \\ \langle \varepsilon, \varepsilon \rangle \end{bmatrix} \right. \\
&\left. , \begin{bmatrix} B \rightarrow \langle b x_{1,1}, d x_{1,2} \rangle (B) \\ \langle (1,2) x_{1,1}, (3,4) x_{1,2} \rangle \end{bmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle \end{bmatrix}, \dots \right\} \\
\vdash_{D_{CYK}(G,w)} &\left\{ \dots, \begin{pmatrix} A \\ \langle \varepsilon, \varepsilon \rangle \end{pmatrix}, \begin{pmatrix} B \\ \langle \varepsilon, \varepsilon \rangle \end{pmatrix}, \dots \right\} \\
\vdash_{D_{CYK}(G,w)} &\left\{ \dots, \begin{pmatrix} A \\ \langle (0,1), (2,3) \rangle \end{pmatrix}, \begin{pmatrix} B \\ \langle (1,2), (3,4) \rangle \end{pmatrix}, \begin{pmatrix} S \\ \langle \varepsilon, \varepsilon \rangle \end{pmatrix}, \dots \right\} \\
\vdash_{D_{CYK}(G,w)} &\left\{ \dots, \begin{pmatrix} S \\ \langle (0,4) \rangle \end{pmatrix}, \dots \right\}
\end{aligned}
$$

CYK parsing (optimized) - example

$$
\begin{aligned}
&\emptyset \\
\vdash_{D_{CYK}(G,w)} &\left\{ \begin{pmatrix} A \\ \langle e_1, e_2 \rangle \end{pmatrix}, \begin{bmatrix} A \rightarrow \langle a x_{1,1}, c x_{1,2} \rangle (A) \\ \langle (0,1) x_{1,1}, (2,3) x_{1,2} \rangle \end{bmatrix}, \begin{pmatrix} B \\ \langle e_3, e_4 \rangle \end{pmatrix} \right. \\
&\left. , \begin{bmatrix} B \rightarrow \langle b x_{1,1}, d x_{1,2} \rangle (B) \\ \langle (1,2) x_{1,1}, (3,4) x_{1,2} \rangle \end{bmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle \end{bmatrix}, \dots \right\} \\
\vdash_{D_{CYK}(G,w)} &\left\{ \dots, \begin{pmatrix} A \\ \langle (0,1), (2,3) \rangle \end{pmatrix}, \begin{pmatrix} B \\ \langle (1,2), (3,4) \rangle \end{pmatrix}, \begin{pmatrix} S \\ \langle e_5 \rangle \end{pmatrix}, \dots \right\} \\
\vdash_{D_{CYK}(G,w)} &\left\{ \dots, \begin{pmatrix} S \\ \langle (0,4) \rangle \end{pmatrix}, \dots \right\}
\end{aligned}
$$

Naive parsing

- three types rules
  - initial active items of nonterminals

$$\frac{}{[A \to f(A_1, \dots, A_k), \phi, A_1 \dots A_k]} \forall A \to f(A_1, \dots, A_k) \in P, \forall \phi \in f_w$$

  - complete one nonterminal using a passive item

$$\frac{[A \to f(A_1, \dots, A_k), \phi, A_i \dots A_k] \; (A_i, \rho_i)}{[A \to f(A_1, \dots, A_k), \phi(\rho_i), A_{i+1} \dots A_k]}$$

  - convert a completed active item

$$\frac{[A \to f(A_1, \dots, A_k), \rho, \varepsilon]}{(A, \rho)}$$

- weights
  - $\mu_{D_{CYK}(G,w)}([A \to f(A_1, \dots, A_k), \phi, A_i \dots A_k]) = \mu_G(A \to f(A_1, \dots, A_k)) \odot \bigodot_{j=1}^{i-1} \mu_{D_{CYK}(G,w)}((A_j, \rho_j))$ if the item was previously completed with (A_j, ρ_j) for each $j \in [i]$
  - $\alpha_{D_{CYK}(G,w)}([A \to f(A_1, \dots, A_k), \phi, A_i \dots A_k]) = \alpha_{D_{CYK}(G,w)}((A, \rho) = \alpha_G(A)$
  - $\beta_{D_{CYK}(G,w)}([A \to f(A_1, \dots, A_k), \phi, A_i \dots A_k]) = \bigodot_{j=i}^{k} \beta_G(A_j)$

Naive parsing – example

$$\emptyset$$

$$\vdash_{D_{NA}(G,w)} \left\{ \begin{bmatrix} A \to \langle \varepsilon, \varepsilon \rangle () \\ \langle e_1, e_2 \rangle \\ \epsilon \end{bmatrix}, \begin{bmatrix} A \to \langle a x_{1,1}, c x_{1,2} \rangle (A) \\ \langle (0,1) x_{1,1}, (2,3) x_{1,2} \rangle \\ A \end{bmatrix}, \begin{bmatrix} B \to \langle \varepsilon, \varepsilon \rangle () \\ \langle e_3, e_4 \rangle \\ \epsilon \end{bmatrix} \right.$$

$$\left. , \begin{bmatrix} B \to \langle b x_{1,1}, d x_{1,2} \rangle (B) \\ \langle (1,2) x_{1,1}, (3,4) x_{1,2} \rangle \\ B \end{bmatrix}, \begin{bmatrix} S \to \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle \\ AB \end{bmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \dots, \begin{pmatrix} A \\ \langle e_1, e_2 \rangle \end{pmatrix}, \begin{pmatrix} B \\ \langle e_3, e_4 \rangle \end{pmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \begin{bmatrix} A \to \langle a x_{1,1}, c x_{1,2} \rangle (A) \\ \langle (0,1), (2,3) \rangle \\ \varepsilon \end{bmatrix}, \begin{bmatrix} B \to \langle b x_{1,1}, d x_{1,2} \rangle (B) \\ \langle (1,2), (3,4) \rangle \\ \varepsilon \end{bmatrix}, \begin{bmatrix} S \to \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle x_{2,1} x_{2,2} \rangle \\ B \end{bmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \dots, \begin{pmatrix} A \\ \langle (0,1), (2,3) \rangle \end{pmatrix}, \begin{pmatrix} B \\ \langle (1,2), (3,4) \rangle \end{pmatrix}, \begin{pmatrix} S \\ \langle e_5 \rangle \end{pmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \dots, \begin{bmatrix} S \to \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle (0,1) x_{2,1} (2,3) x_{2,2} \rangle \\ B \end{bmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \dots, \begin{pmatrix} S \\ \langle (0,4) \rangle \end{pmatrix}, \dots \right\}$$

Naive parsing (optimized) - example

$$
\begin{aligned}
&\emptyset \\
\vdash_{D_{NA}(G,w)} &\left\{ \begin{pmatrix} A \\ \langle e_1, e_2 \rangle \end{pmatrix}, \begin{bmatrix} A \rightarrow \langle a x_{1,1}, c x_{1,2} \rangle (A) \\ \langle (0,1) x_{1,1}, (2,3) x_{1,2} \rangle \\ A \end{bmatrix}, \begin{pmatrix} B \\ \langle e_3, e_4 \rangle \end{pmatrix} \right. \\
&\left. , \begin{bmatrix} B \rightarrow \langle b x_{1,1}, d x_{1,2} \rangle (B) \\ \langle (1,2) x_{1,1}, (3,4) x_{1,2} \rangle \\ B \end{bmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle \\ AB \end{bmatrix}, \dots \right\} \\
\vdash_{D_{NA}(G,w)} &\left\{ \dots \begin{pmatrix} A \\ \langle (0,1), (2,3) \rangle \end{pmatrix}, \begin{pmatrix} B \\ \langle (1,2), (3,4) \rangle \end{pmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle x_{2,1} x_{2,2} \rangle \\ B \end{bmatrix}, \dots \right\} \\
\vdash_{D_{NA}(G,w)} &\left\{ \dots, \begin{pmatrix} S \\ \langle e_5 \rangle \end{pmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle (0,1) x_{2,1} (2,3) x_{2,2} \rangle \\ B \end{bmatrix}, \dots \right\} \\
\vdash_{D_{NA}(G,w)} &\left\{ \dots, \begin{pmatrix} S \\ \langle (0,4) \rangle \end{pmatrix}, \dots \right\}
\end{aligned}
$$

Active parsing I

- rules
  - initial active items of nonterminals

$$\frac{}{\left[A \to f(A_1,\dots,A_k), \langle e \bullet f \rangle, \bot\right]} \forall A \to f(A_1,\dots,A_k) \in P$$

  - read terminal

$$\frac{\left[A \to f(A_1,\dots,A_k), \langle \psi, r \bullet \sigma\phi \rangle, \Gamma\right]}{\left[A \to f(A_1,\dots,A_k), \langle \psi, (r \cdot r_\sigma) \bullet \phi \rangle, \Gamma\right]}$$

  - read component of non-completed nonterminal

$$\frac{\left[A \to f(A_1,\dots,A_k), \langle \psi, r \bullet x_{i,j}\phi \rangle, \Gamma\right] \, (A_i, \rho_i)}{\left[A \to f(A_1,\dots,A_k), \langle \psi, (r \cdot \rho_{i,j}) \bullet \phi \rangle, \Gamma[i/\rho]\right]}$$

  - read component of known nonterminal

$$\frac{\left[A \to f(A_1,\dots,A_k), \langle \psi, r \bullet x_{i,j}\phi \rangle, \Gamma\right]}{\left[A \to f(A_1,\dots,A_k), \langle \psi, (r \cdot \Gamma(i)_j) \bullet \phi \rangle, \Gamma\right]}$$

  - convert a completed active item

$$\frac{\left[A \to f(A_1,\dots,A_k), \langle \psi \bullet \varepsilon \rangle, \Gamma\right]}{(A, \psi,)}$$

Active parsing II

- weights
  - $\mu_{D_{ACT}(G,w)}([A \to f(A_1,\ldots,A_k), \langle\psi \bullet \phi\rangle, \Gamma]) = \mu_G(A \to f(A_1,\ldots,A_k)) \odot \bigodot_{i \in \mathrm{dom}(\Gamma)} \mu_{D_{ACT}(G,w)}((A_i, \Gamma(i)))$
  - $\alpha_{D_{ACT}(G,w)}([A \to f(A_1,\ldots,A_k), \langle\psi \bullet \phi\rangle, \Gamma]) = \alpha_{D_{ACT}(G,w)}((A,\rho)) = \alpha_G(A)$
  - $\beta_{D_{ACT}(G,w)}([A(A_1,\ldots,A_k) \to f, \langle\psi \bullet \phi\rangle, \Gamma]) = \bigodot_{i\in[k]:\ i \notin \mathrm{dom}(\Gamma)} \beta_G(A_i)$

Active parsing (optimized) – example

$$\emptyset$$

$$\vdash_{D_{NA}(G,w)} \left\{ \begin{pmatrix} A \\ \langle e_1, e_2 \rangle \end{pmatrix}, \begin{bmatrix} A \rightarrow \langle a x_{1,1}, c x_{1,2} \rangle (A) \\ \langle (0,1) \bullet x_{1,1}, c x_{1,2} \rangle \\ \bot \end{bmatrix}, \begin{pmatrix} B \\ \langle e_3, e_4 \rangle \end{pmatrix} \right.$$

$$\left. , \begin{bmatrix} B \rightarrow \langle b x_{1,1}, d x_{1,2} \rangle (B) \\ \langle (1,2) \bullet x_{1,1}, d x_{1,2} \rangle \\ \bot \end{bmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle e \bullet x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle \\ \bot \end{bmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \dots \begin{pmatrix} A \\ \langle (0,1), (2,3) \rangle \end{pmatrix}, \begin{pmatrix} B \\ \langle (1,2), (3,4) \rangle \end{pmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle e \bullet x_{2,1} x_{1,2} x_{2,2} \rangle \\ 1 \rightarrow \langle e_1, e_2 \rangle \end{bmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \dots, \begin{pmatrix} S \\ \langle e_5 \rangle \end{pmatrix}, \begin{bmatrix} S \rightarrow \langle x_{1,1} x_{2,1} x_{1,2} x_{2,2} \rangle (A, B) \\ \langle (0,1) \bullet x_{2,1} x_{1,2} x_{2,2} \rangle \\ 1 \rightarrow \langle (0,1), (2,3) \rangle \end{bmatrix}, \dots \right\}$$

$$\vdash_{D_{NA}(G,w)} \left\{ \dots, \begin{pmatrix} S \\ \langle (0,4) \rangle \end{pmatrix}, \dots \right\}$$