Trade-off between incrementality and accuracy
Seminar on Incremental Processing

Benjamin Weitz

June 30, 2011
Outline

1. Introduction
2. Baumann et al.
4. Summary
Introduction

Incremental Dialogue Systems:
Introduction

Incremental Dialogue Systems:
- fast(er)
Introduction

Incremental Dialogue Systems:

- fast(er)
- sometimes wrong decisions
Introduction

Incremental Dialogue Systems:

- fast(er)
- sometimes wrong decisions $\Rightarrow$ revisions
Introduction

Incremental Dialogue Systems:

- fast(er)
- sometimes wrong decisions $\Rightarrow$ revisions
- components depend on each other
Introduction

Incremental Dialogue Systems:
- fast(er)
- sometimes wrong decisions $\Rightarrow$ revisions
- components depend on each other

Idea: decrease incrementality a bit to reduce revisions
Introduction

Incremental Dialogue Systems:

- fast(er)
- sometimes wrong decisions ⇒ revisions
- components depend on each other

Idea: decrease incrementality a bit to reduce revisions

- Baumann et. al: Automatic Speech Recognition
Introduction

Incremental Dialogue Systems:

- fast(er)
- sometimes wrong decisions $\rightarrow$ revisions
- components depend on each other

Idea: decrease incrementality a bit to reduce revisions

- Baumann et. al: Automatic Speech Recognition
- Kato et. al: Parsing
Baumann et al.
Define measures to evaluate incremental ASR-Systems
Baumann et al.

- Define measures to evaluate incremental ASR-Systems
- Evaluate an existing system with these measures
Baumann et al.

- Define measures to evaluate incremental ASR-Systems
- Evaluate an existing system with these measures
- Use the measures to improve ASR
Define measures to evaluate incremental ASR-Systems

Evaluate an existing system with these measures

Use the measures to improve ASR

Conclusions
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Baumann et al.</th>
<th>Kato et al.</th>
<th>Summary</th>
</tr>
</thead>
</table>

**Baumann et al.**

- Define measures to evaluate incremental ASR-Systems
- Evaluate an existing system with these measures
- Use the measures to improve ASR
- Conclusions
Definitions

Hypothesis at time $t$:

$$hyp_t = w_{hyp_t}$$
Hypothesis at time $t$:

$$hyp_t = w_{hyp_t}$$

How to evaluate the quality of $hyp_t$?

- use actually spoken input as gold standard
- use final hypothesis of ASR as gold standard
Definitions

Hypothesis at time $t$:

$$hyp_t = w_{hyp_t}$$

How to evaluate the quality of $hyp_t$?

- use actually spoken input as gold standard
- use final hypothesis of ASR as gold standard
Definitions

Hypothesis at time $t$:

$$hyp_t = w_{hyp_t}$$

How to evaluate the quality of $hyp_t$?

- use actually spoken input as gold standard
- use final hypothesis of ASR as gold standard

Why?

- more meaningful:
  - relates partial hypothesis to what can be expected from ASR
  - correct interpretation might never be recognized
Relative Correctness

\( w_{gold} \): final, non incremental hypothesis
Relative Correctness

$w_{\text{gold}}$: final, non incremental hypothesis

Relative Correctness

$w_{\text{hyp}_t}$ is relatively correct, iff

$w_{\text{hyp}_t} = w_{\text{gold}_t}$
Relative Correctness

\( w_{\text{gold}} \): final, non incremental hypothesis

Relative Correctness

\( w_{\text{hyp}_t} \) is relatively correct, iff

\[ w_{\text{hyp}_t} = w_{\text{gold}_t} \]
Relative Correctness

$w_{\text{gold}}$: final, non incremental hypothesis

$w_{\text{hyp}_t}$ is relatively correct, iFF

$w_{\text{hyp}_t} = w_{\text{gold}_t}$
Relative Correctness

$w_{\text{gold}}$: final, non incremental hypothesis

$w_{\text{hyp}_t}$ is relatively correct, iff

$w_{\text{hyp}_t} = w_{\text{gold}_t}$
**Relative Correctness**

$W_{\text{gold}}$: final, non incremental hypothesis

$W_{\text{hyp}_t}$ is relatively correct, iff

$W_{\text{hyp}_t} = W_{\text{gold}_t}$
Relative Correctness

\( w_{\text{gold}} \): final, non incremental hypothesis

**Relative Correctness**

\( w_{\text{hyp}_t} \) is relatively correct, iff

\[ w_{\text{hyp}_t} = w_{\text{gold}_t} \]

**Prefix Correctness**

\( w_{\text{hyp}_t} \) is prefix-correct, iff

\( w_{\text{hyp}_t} \) is a prefix of \( w_{\text{gold}_t} \)
Relative Correctness

\( w_{\text{gold}} \): final, non incremental hypothesis

Relative Correctness

\( w_{\text{hyp}_t} \) is relatively correct, iff \( w_{\text{hyp}_t} = w_{\text{gold}_t} \)

Prefix Correctness

\( w_{\text{hyp}_t} \) is prefix-correct, iff \( w_{\text{hyp}_t} \) is a prefix of \( w_{\text{gold}_t} \)
Three ways for $hyp_{t+1}$ to differ from $hyp_t$:

- extension
- revokation
- revision
Edit Overhead

Three ways for $hyp_{t+1}$ to differ from $hyp_t$:

- extension
- revokation
- revision

Two types of edit messages:

- add message: $\oplus$
- revoke message: $\ominus$
Edit Overhead

Three ways for $h_{t+1}$ to differ from $h_t$:

- extension: $\oplus$
- revokation
- revision

Two types of edit messages:

- add message: $\ominus$
- revoke message: $\ominus$
Edit Overhead

Three ways for $hyp_{t+1}$ to differ from $hyp_t$:

- extension: $\oplus$
- revokation: $\ominus$
- revision

Two types of edit messages:

- $add$ message: $\oplus$
- $revoke$ message: $\ominus$
Three ways for $hyp_{t+1}$ to differ from $hyp_t$:

- extension: $\oplus$
- revokation: $\ominus$
- revision: $\ominus$, $\oplus$

Two types of *edit messages*:

- *add* message: $\oplus$
- *revoke* message: $\ominus$
Edit Overhead

$w_{gold}$: sil eins zwei drei ...

time: 0 1 2 3 4 5 6 7 8 9 10 11 12

$w_{hyp_1}$: sil

$w_{hyp_2}$: sil

$w_{hyp_3}$: sil

$w_{hyp_4}$: sil an $\oplus$(an)

$w_{hyp_5}$: sil ein $\oplus$(an), $\oplus$(ein)

$w_{hyp_6}$: sil eins $\oplus$(ein), $\oplus$(eins)

$w_{hyp_7}$: sil eins zwei $\oplus$(zwei)

$w_{hyp_8}$: sil eins zwar $\oplus$(zwei), $\oplus$(zwar)

$w_{hyp_9}$: sil eins zwei $\oplus$(zwar), $\oplus$(zwei)

$w_{hyp_{10}}$: sil eins zwei $\oplus$(zwei)

$w_{hyp_{11}}$: sil eins zwei sil

$w_{hyp_{12}}$: sil eins zwei drei $\oplus$(drei)

$WFC_{zwei}$ $WFF_{zwei}$
Edit Overhead

$w_{gold}$: sil eins zwei drei ...

$w_{hyp_1}$: sil...

$w_{hyp_2}$: sil...

$w_{hyp_3}$: sil...

$w_{hyp_4}$: sil an...

$w_{hyp_5}$: sil ein...

$w_{hyp_6}$: sil eins...

$w_{hyp_7}$: sil eins zwei...

$w_{hyp_8}$: sil eins zwar...

$w_{hyp_9}$: sil eins zwei...

$w_{hyp_{10}}$: sil eins zwei...

$w_{hyp_{11}}$: sil eins zwei sil...

$w_{hyp_{12}}$: sil eins zwei drei...

$W_{FC_{zwei}}$ $W_{FF_{zwei}}$

$\oplus$(an)

$\oplus$(an), $\oplus$(ein)

$\oplus$(ein), $\oplus$(eins)

$\oplus$(zwei)

$\oplus$(zwei), $\oplus$(zwar)

$\oplus$(zwei), $\oplus$(zwei)

$\oplus$(zwei), $\oplus$(zwar)

$\oplus$(drei)
Edit Overhead

Perfect ASR-System:
1 extension for each word
**Edit Overhead**

Perfect ASR-System:
1 extension for each word

Edit Overhead (EO)
rate of spurious edits
Edit Overhead

Perfect ASR-System:
1 extension for each word

Edit Overhead (EO)
rate of spurious edits

11 edits
Edit Overhead

Perfect ASR-System:
1 extension for each word

Edit Overhead (EO)
rate of spurious edits

11 edits, 3 words
Edit Overhead

Perfect ASR-System:
1 extension for each word

Edit Overhead (EO)
rate of spurious edits

11 edits, 3 words
⇒ 8 unnecessary edits
**Edit Overhead**

**Perfect ASR-System:**
1 extension for each word

**Edit Overhead (EO)**
rate of spurious edits

11 edits, 3 words
⇒ 8 unnecessary edits
⇒ $EO = \frac{8}{11}$
Timing Measures

**Word First Correct Response (WFC)**

The first time a word appears in the correct position
### Timing Measures

**Word First Correct Response (WFC)**

The first time a word appears in the correct position
**Introduction**

Baumann et al.

Kato et al.

**Summary**

---

## Timing Measures

### Word First Correct Response (WFC)

The first time a word appears in the correct position

$$WFC_{\text{zwei}} = 7$$
**Timing Measures**

**Word First Correct Response (WFC)**

The first time a word appears in the correct position

**Word First Final Response (WFF)**

The time a hypothesis remains stable / doesn’t change anymore
Timing Measures

**Word First Correct Response (WFC)**

The first time a word appears in the correct position

**Word First Final Response (WFF)**

The time a hypothesis remains stable / doesn’t change anymore
Introduction

Baumann et al.

Kato et al.

Summary

Timing Measures

**Word First Correct Response (WFC)**

The first time a word appears in the correct position

**Word First Final Response (WFF)**

The time a hypothesis remains stable / doesn’t change anymore

\[ WFF_{\text{zwei}} = 9 \]
Timing Measures

**Word First Correct Response (WFC)**

The first time a word appears in the correct position

**Word First Final Response (WFF)**

The time a hypothesis remains stable / doesn’t change anymore

**Correction Time (CT)**

\[ CT = WFF - WFC \]
Timing Measures

**Word First Correct Response (WFC)**

The first time a word appears in the correct position.

**Word First Final Response (WFF)**

The time a hypothesis remains stable / doesn’t change anymore.

**Correction Time (CT)**

\[ CT = WFF - WFC \]
## Timing Measures

### Word First Correct Response (WFC)

The first time a word appears in the correct position

### Word First Final Response (WFF)

The time a hypothesis remains stable / doesn’t change anymore

### Correction Time (CT)

\[ CT = WFF - WFC \]

\[ CT_{\text{zwei}} = 9 - 7 = 2 \]
Baumann et al.

- Define measures to evaluate incremental ASR-Systems
- Evaluate an existing system with these measures
- Use the measures to improve ASR
- Conclusions
Setup and Data

- continuous speech framework *Sphinx-4*
- acoustic model
  - German
  - instructions in a puzzle domain
- trigram language model
- test data
  - 85 recordings
  - two speakers
  - sentence similar to training sentences
# Measurements

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SER</td>
<td>68.2%</td>
</tr>
<tr>
<td>WER</td>
<td>18.8%</td>
</tr>
<tr>
<td>r-correct</td>
<td>30.9%</td>
</tr>
<tr>
<td>p-correct</td>
<td>53.1%</td>
</tr>
<tr>
<td>Edit overhead</td>
<td>90.5%</td>
</tr>
<tr>
<td>Mean word duration</td>
<td>0.378 s</td>
</tr>
<tr>
<td>WFC mean</td>
<td>0.276 s</td>
</tr>
<tr>
<td>WFF mean</td>
<td>0.004 s</td>
</tr>
<tr>
<td>Immediately correct</td>
<td>58.6%</td>
</tr>
</tbody>
</table>
## Measurements

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SER</td>
<td>68.2%</td>
</tr>
<tr>
<td>WER</td>
<td>18.8%</td>
</tr>
<tr>
<td>r-correct</td>
<td>30.9%</td>
</tr>
<tr>
<td>p-correct</td>
<td>53.1%</td>
</tr>
<tr>
<td>edit overhead</td>
<td>90.5%</td>
</tr>
<tr>
<td>mean word duration</td>
<td>0.378 s mean: 0.276 s</td>
</tr>
<tr>
<td>WFC</td>
<td></td>
</tr>
<tr>
<td>WFF</td>
<td>mean: 0.004 s</td>
</tr>
<tr>
<td>immediately correct</td>
<td>58.6%</td>
</tr>
</tbody>
</table>

- rather low correctness
## Measurements

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SER</td>
<td>68.2%</td>
</tr>
<tr>
<td>WER</td>
<td>18.8%</td>
</tr>
<tr>
<td>r-correct</td>
<td>30.9%</td>
</tr>
<tr>
<td>p-correct</td>
<td>53.1%</td>
</tr>
<tr>
<td><strong>edit overhead</strong></td>
<td><strong>90.5%</strong></td>
</tr>
<tr>
<td>mean word duration</td>
<td>0.378 s</td>
</tr>
<tr>
<td>WFC mean</td>
<td>0.276 s</td>
</tr>
<tr>
<td>WFF mean</td>
<td>0.004 s</td>
</tr>
<tr>
<td>immediately correct</td>
<td>58.6%</td>
</tr>
</tbody>
</table>

- rather low correctness
- very high edit overhead!
## Measurements

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SER</td>
<td>68.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WER</td>
<td>18.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r-correct</td>
<td>30.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-correct</td>
<td>53.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edit overhead</td>
<td>90.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean word duration</td>
<td>0.378 s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WFC mean:</td>
<td>0.276 s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WFF mean:</td>
<td>0.004 s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>immediately correct</td>
<td>58.6%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- rather low correctness
- very high edit overhead!
- information available after \( \frac{3}{4} \) have been spoken
# Measurements

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SER</td>
<td>68.2%</td>
</tr>
<tr>
<td>WER</td>
<td>18.8%</td>
</tr>
<tr>
<td>r-correct</td>
<td>30.9%</td>
</tr>
<tr>
<td>p-correct</td>
<td>53.1%</td>
</tr>
<tr>
<td>edit overhead</td>
<td>90.5%</td>
</tr>
<tr>
<td>mean word duration</td>
<td>0.378 s</td>
</tr>
<tr>
<td>WFC mean</td>
<td>0.276 s</td>
</tr>
<tr>
<td>WFF mean</td>
<td>0.004 s</td>
</tr>
<tr>
<td>immediately correct</td>
<td>58.6%</td>
</tr>
</tbody>
</table>

- rather low correctness
- very high edit overhead!
- information available after $\frac{3}{4}$ have been spoken
- word becomes final when it has ended
## Measurements

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SER</td>
<td>68.2%</td>
</tr>
<tr>
<td>WER</td>
<td>18.8%</td>
</tr>
<tr>
<td>r-correct</td>
<td>30.9%</td>
</tr>
<tr>
<td>p-correct</td>
<td>53.1%</td>
</tr>
<tr>
<td>edit overhead</td>
<td>90.5%</td>
</tr>
<tr>
<td>mean word duration</td>
<td>0.378 s</td>
</tr>
<tr>
<td>WFC mean:</td>
<td>0.276 s</td>
</tr>
<tr>
<td>WFF mean:</td>
<td>0.004 s</td>
</tr>
<tr>
<td>immediately correct</td>
<td>58.6%</td>
</tr>
</tbody>
</table>

- rather low correctness
- very high edit overhead!
- information available after $\frac{3}{4}$ have been spoken
- word becomes final when it has ended
58.6% immediately correct
Measures of correct words as a function of correction time.

- 58.6% immediately correct
- 90% correct after a correction time of 320 ms
Measurements

58.6% immediately correct

90% correct after a correction time of 320 ms

95% correct after a correction time of 550 ms
Independency of the measures?

Are the measures independent of specific settings?

- vary LM/AM-weight
- vary audio quality by adding white noise
Independency of the measures?

Are the measures independent of specific settings?
- vary LM/AM-weight
- vary audio quality by adding white noise
Baumann et al.

- Define measures to evaluate incremental ASR-Systems
- Evaluate an existing system with these measures
- Use the measures to improve ASR
- Conclusions
Which improvements?

main goal: improve edit overhead
- reduce amount of wrong hypotheses
- still as quick as possible
allow larger right context of size $\Delta$:

- at time $t$: take into account output of ASR until $t - \Delta$ only
- $hyp_{t-\Delta}$ has a lookahead up to $t$
Right Context

allow larger right context of size $\Delta$:

- at time $t$: take into account output of ASR until $t - \Delta$ only
- $hyp_{t-\Delta}$ has a lookahead up to $t$

$\Rightarrow$ reduction of edit overhead
Right Context

allow larger right context of size $\Delta$:

- at time $t$: take into account output of ASR until $t - \Delta$ only
- $hyp_{t-\Delta}$ has a lookahead up to $t$

$\Rightarrow$ reduction of edit overhead

$\Rightarrow$ hypothesis lags behind the gold standard

- WFC increases by $\Delta$
- effects on correctness, because $w_{gold_t}$ may contain more words
Right Context

allow larger right context of size $\Delta$:

- at time $t$: take into account output of ASR until $t - \Delta$ only
- $hyp_{t-\Delta}$ has a lookahead up to $t$

$\Rightarrow$ reduction of edit overhead

$\Rightarrow$ hypothesis lags behind the gold standard

- WFC increases by $\Delta$
- effects on correctness, because $w_{gold_t}$ may contain more words

Fair R-Correctness

$w_{hyp_t}$ is fairly r-correct, iff $w_{hyp_{t-\Delta}} = w_{gold_{t-\Delta}}$
Right Context

correctness and EO improve with more right context
Right Context

correctness and EO improve with more right context

timing measures increase with larger right context
correctness and EO improve with more right context

timing measures increase with larger right context

percentage of immediately correct hypotheses increases:
90% @ 580 ms
98% @ 1060 ms
edit message must be result of $N$ consecutive hypotheses before commitment
Message Smoothing

edit message must be result of $N$ consecutive hypotheses before commitment

$N = 2$: 
Message Smoothing

Edit message must be result of $N$ consecutive hypotheses before commitment

$N = 2$:

- *an, ein* and *zwar* would never be committed
Message Smoothing

edit message must be result of \( N \) consecutive hypotheses before commitment

\( N = 2 \):

- *an*, *ein* and *zwar* would never be committed
- only 3 edit messages
Message Smoothing

edit overhead falls rapidly:
50% after 110 ms
10% after 320 ms
Message Smoothing

edit overhead falls rapidly:
50% after 110 ms
10% after 320 ms
decreasing (strict and fair)
r-correctness
Message Smoothing

edit overhead falls rapidly:
50% after 110 ms
10% after 320 ms
decreasing (strict and fair)
r-correctness
increasing timing measures
Define measures to evaluate incremental ASR-Systems

Evaluate an existing system with these measures

Use the measures to improve ASR

Conclusions
goal: improve edit overhead
Conclusions

goal: improve edit overhead

**Right Context**

- improvements with larger delays, increasing correctness

**Message Smoothing**

- improvements with shorter delays, decreasing correctness
Conclusions

**goal:** improve edit overhead

**Right Context**
improvements with larger delays, increasing correctness

**Message Smoothing**
improvements with shorter delays, decreasing correctness

could be combined to yield a good effect
### Kato et. al

<table>
<thead>
<tr>
<th>Introduction</th>
<th>Baumann et al.</th>
<th>Kato et al.</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### PIT AG: incremental-parsing-oriented tree adjoining grammar

### PIT AG: probabilistic PIT AG

### Validity of Partial Parse Trees

### Experimental Results and Conclusions
Kato et. al

- ITAG: incremental-parsing-oriented tree adjoining grammar
Kato et al.

- ITAG: incremental-parsing-oriented tree adjoining grammar
- PITAG: probabilistic ITAG
Kato et. al

- ITAG: incremental-parsing-oriented tree adjoining grammar
- PITAG: probabilistic ITAG
- Validity of Partial Parse Trees
Kato et. al

- ITAG: incremental-parsing-oriented tree adjoining grammar
- PITAG: probabilistic ITAG
- Validity of Partial Parse Trees
- Experimental Results and Conclusions
Kato et. al

- ITAG: incremental-parsing-oriented tree adjoining grammar
- PITAG: probabilistic ITAG
- Validity of Partial Parse Trees
- Experimental Results and Conclusions
Elementary Trees

Initial Trees

must be *leftmost-expanded*:
Elementary Trees

Initial Trees

must be leftmost-expanded:

1. \([t]_X\)
   
   \(t\): terminal symbol
   
   \(X\): nonterminal symbol
Elementary Trees

**Initial Trees**

must be *leftmost-expanded*:

1. \([t]_X\)  \(X\)

   - \(t\): terminal symbol
   - \(X\): nonterminal symbol
Elementary Trees

Initial Trees

must be *leftmost-expanded*:

1. $[t]_X$
   - $t$: terminal symbol
   - $X$: nonterminal symbol

2. $[\sigma X_1 \cdots X_k]_X$
   - $\sigma$: leftmost expanded tree
   - $X, X_1, \ldots, X_k$: nonterminal symbols
Elementary Trees

Initial Trees

must be *leftmost-expanded*:

1. \([t]_X\)
   - \(t\): terminal symbol
   - \(X\): nonterminal symbol

2. \([\sigma X_1 \cdots X_k]_X\)
   - \(\sigma\): leftmost expanded tree
   - \(X, X_1, \ldots, X_k\): nonterminal symbols
Elementary Trees

Initial Trees

must be *leftmost-expanded*:

1. \([t]_X\)
   - \(t\): terminal symbol
   - \(X\): nonterminal symbol

2. \([\sigma X_1 \cdots X_k]_X\)
   - \(\sigma\): leftmost expanded tree
   - \(X, X_1, \ldots, X_k\): nonterminal symbols

Auxiliary Trees

\([X^* \sigma X_1 \cdots X_k]_X\)

- \(\sigma\): leftmost expanded tree
- \(X, X_1, \ldots, X_k\): nonterminal symbols
Elementary Trees

Initial Trees

must be *leftmost-expanded*:

1. \([t]_X\)
   - \(t\): terminal symbol
   - \(X\): nonterminal symbol

2. \([\sigma X_1 \cdots X_k]_X\)
   - \(\sigma\): leftmost expanded tree
   - \(X, X_1, \ldots, X_k\): nonterminal symbols

Auxiliary Trees

\([X^* \sigma X_1 \cdots X_k]_X\)
- \(\sigma\): leftmost expanded tree
- \(X, X_1, \ldots, X_k\): nonterminal symbols
Elementary Trees

Initial trees:

$\alpha_1$: $S$
- NP
- PRP
- I

$\alpha_2$: VP
- VB
- NP
- found

$\alpha_3$: VP
- VB
- NP
- ADJP
- found

$\alpha_4$: VP
- VB
- NP
- found

$\alpha_5$: NP
- DT
- NN
- a

$\alpha_6$: NP
- DT
- JJ
- NN
- a

$\alpha_7$: NN
- "dime"

$\alpha_8$: NP
- DT
- NN
- the

$\alpha_9$: NP
- DT
- JJ
- NN
- the

$\alpha_{10}$: NN
- "wood"
Elementary Trees

Auxiliary trees:

\[ \beta_1 \quad \begin{array}{c} NP \\ NP^* \\ \text{IN} \\ \text{in} \end{array} \quad \text{PP} \quad \text{IN} \quad \text{NP} \]

\[ \beta_2 \quad \begin{array}{c} VP \\ VP^* \\ \text{IN} \\ \text{in} \end{array} \quad \text{PP} \quad \text{IN} \quad \text{NP} \]
Operations

Substitution

replaces a leftmost nonterminal leaf of a partial parse tree \( \sigma \) with an initial tree \( \alpha \) having the same nonterminal symbol at its root

\( s_\alpha: \) substituting \( \alpha \)

\( s_\alpha(\sigma): \) result of applying \( s_\alpha \) to \( \sigma \)
Operations

Substitution

replaces a leftmost nonterminal leaf of a partial parse tree $\sigma$ with an initial tree $\alpha$ having the same nonterminal symbol at its root

$s_\alpha$: substituting $\alpha$

$s_\alpha(\sigma)$: result of applying $s_\alpha$ to $\sigma$

Example

```
      S
     / \  \
    NP  VP
   /   \
  PRP   \
```

$\frac{29}{50}$
**Operations**

### Substitution

replaces a leftmost nonterminal leaf of a partial parse tree $\sigma$ with an initial tree $\alpha$ having the same nonterminal symbol at its root:

$s_\alpha$: substituting $\alpha$

$s_\alpha(\sigma)$: result of applying $s_\alpha$ to $\sigma$

---

**Example**

```
S
  /  \\
NP  VP
  /  \\
PRP  VP
       /  \\
      VB  NP
            /  \\
           found
```
**Substitution**

replaces a leftmost nonterminal leaf of a partial parse tree $\sigma$ with an initial tree $\alpha$ having the same nonterminal symbol at its root

$s_\alpha$: substituting $\alpha$

$s_\alpha(\sigma)$: result of applying $s_\alpha$ to $\sigma$

**Example**

```
S
  NP  VP
  |    |
PRP  found
```

```
VP
  VB  NP
```

```
S
  NP  VP
  |    |
PRP  found
```

```
NP  VP
  |    |
PRP  VB  NP
  |    |
found
```
Op erations

Ad junction

splits a partial parse tree $\sigma$ at a nonterminal node having no nonterminal leaf and inserts an auxiliary tree $\beta$ having the same nonterminal symbol at its root

$a_\beta$: adjoining $\beta$

$a_\beta(\sigma)$: result of applying $a_\beta$ to $\sigma$
**Adjunction**

splits a partial parse tree $\sigma$ at a nonterminal node having no nonterminal leaf and inserts an auxiliary tree $\beta$ having the same nonterminal symbol at its root

$a_\beta$: adjoining $\beta$

$a_\beta(\sigma)$: result of applying $a_\beta$ to $\sigma$

**Example**

```
S
  NP
    PRP
      I
VP
    found a dime
```
Adjunction

splits a partial parse tree $\sigma$ at a nonterminal node having no nonterminal leaf and inserts an auxiliary tree $\beta$ having the same nonterminal symbol at its root

$a_\beta$: adjoining $\beta$

$a_\beta(\sigma)$: result of applying $a_\beta$ to $\sigma$

Example

![Diagram of parse trees]
Adjunction

splits a partial parse tree $\sigma$ at a nonterminal node having no nonterminal leaf and inserts an auxiliary tree $\beta$ having the same nonterminal symbol at its root

$a_\beta$: adjoining $\beta$

$a_\beta(\sigma)$: result of applying $a_\beta$ to $\sigma$

Example
at $i$-th word $w_i$: 
at $i$-th word $w_i$: 

- combine (substitution, adjunction) elementary trees for $w_i$ with partial parse trees for $w_1 \cdots w_{i-1}$
at $i$-th word $w_i$:

- combine (substitution, adjunction) elementary trees for $w_i$ with partial parse trees for $w_1 \cdots w_{i-1}$

$\Rightarrow$ partial parse trees for $w_1 \cdots w_i$
Example

$\alpha_1$ S \\
NP VP \\
PRP I \\

$\alpha_2$ VP \\
VB NP \\
found \\

$\alpha_3$ VP \\
VB NP ADJP \\
found \\

$\alpha_4$ VP \\
VB \\
found

S
Example
Example

I found

\( \alpha_1 \quad S \quad \alpha_2 \quad VP \quad \alpha_3 \quad VP \quad \alpha_4 \quad VP \)

\( NP \quad VP \quad VB \quad NP \quad VB \quad NP \quad ADJP \quad VB \quad found \quad found \quad found \)
Example

I found

I found

I found

I found

Found

Found

Found

Found

Found
Example

I found

\[
\begin{align*}
\alpha_1 &: S \\
&: NP \quad VP \\
&: PRP \quad VB \quad NP \\
&: PRP \quad I \quad found \\
\alpha_2 &: VP \\
&: VB \quad NP \\
&: I \quad found \\
\alpha_3 &: VP \\
&: VB \quad NP \quad ADJP \\
&: I \quad found \\
\alpha_4 &: VP \\
&: VB \quad found
\end{align*}
\]
## Example

<table>
<thead>
<tr>
<th>word</th>
<th>#</th>
<th>partial parse tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>s</td>
<td>$[[I_{prp}npvp]_s$</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>$[[I_{prp}np[found]_{vb}npvp]_s$</td>
</tr>
<tr>
<td>found</td>
<td>3</td>
<td>$[[I_{prp}np[found]_{vb}npadjp]_vp]_s$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$[[I_{prp}np[found]_{vb}npadjp]_vp]_s$</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$[[I_{prp}np[found]_{vb}vp]_s$</td>
</tr>
<tr>
<td>a</td>
<td>6</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}nn]npvp]_s$</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}jj nn]npvp]_s$</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}nn]npadjp]_vp]_s$</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}jj nn]npadjp]_vp]_s$</td>
</tr>
<tr>
<td>dime</td>
<td>10</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnpvp]_s$</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnpadjp]_vp]_s$</td>
</tr>
<tr>
<td>in</td>
<td>12</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[vp]_s$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[in]npvp]_s$</td>
</tr>
<tr>
<td>the</td>
<td>14</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[vp]_s$</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[vp]_s$</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[vp]_s$</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[vp]_s$</td>
</tr>
<tr>
<td>wood</td>
<td>18</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[vp]_s$</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>$[[I_{prp}np[found]<em>{vb}[a</em>{dt}dime]nnnp[vp]_s$</td>
</tr>
</tbody>
</table>
### Example

<table>
<thead>
<tr>
<th>word</th>
<th>#</th>
<th>partial parse tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2</td>
<td>$[[I_{prp}]_{np}vp]_s$</td>
</tr>
<tr>
<td>found</td>
<td>3</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}np_{vp}]_s$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}np_{adjp}]_{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}vp]_s$</td>
</tr>
<tr>
<td>a</td>
<td>6</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}nn]_{np} _vp]_s$</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}jj nn]_{np} _vp]_s$</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}nn]_{np} <em>adjp]</em>{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}jj nn]_{np} <em>adjp]</em>{vp} _s$</td>
</tr>
<tr>
<td>dime</td>
<td>10</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{adjp} _{vp} _s$</td>
</tr>
<tr>
<td>in</td>
<td>12</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{in} _{in} _{np} _{pp} _{np} _{vp} _s$</td>
</tr>
<tr>
<td>the</td>
<td>14</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{in} _{in} _{the} _{dt} _{nn} _{np} _{pp} _{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{pp} _{np} _{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{in} _{in} _{the} _{dt} _{jj} _{nn} _{np} _{np} _{vp} _s$</td>
</tr>
<tr>
<td>wood</td>
<td>18</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{vp} _s$</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>$[[I_{prp}]<em>{np}[found]</em>{vb}[a_{dt}dime]_{nn} <em>np]</em>{in} _{in} _{the} _{dt} _{wood} _{nn} _{np} _{pp} _{np} _{vp} _s$</td>
</tr>
</tbody>
</table>

constructing parse trees of initial fragments for every word input possible!
How to get the elementary trees?
Tree Construction

How to get the elementary trees?

- Extraction from a treebank!
## Kato et. al

- **ITAG**: incremental-parsing-oriented tree adjoining grammar
- **PITAG**: probabilistic ITAG
- Validity of Partial Parse Trees
- Experimental Results and Conclusions
probabilities for events combining an elementary tree and another tree by substitution and adjunction
Probabilities

probabilities for events combining an elementary tree and another tree by substitution and adjunction

\( \alpha \): initial tree with root symbol X

**Probability of substituting** \( \alpha \)

\[
P(s_\alpha) = \frac{|s_\alpha| \text{ in the treebank}}{|\text{substitutions using other initial trees with root } X| \text{ in the treebank}}
\]
Probabilities

probabilities for events combining an elementary tree and another tree by substitution and adjunction

\( \alpha \): initial tree with root symbol \( X \)

**Probability of substituting \( \alpha \)**

\[ P(s_\alpha) = \frac{|s_\alpha| \text{ in the treebank}}{|\text{substitutions using other initial trees with root } X| \text{ in the treebank}} \]

\( \beta \): auxiliary tree with root symbol \( X \)

**Probability of adjoining \( \beta \)**

\[ P(a_\beta) = \frac{|a_\beta| \text{ in the treebank}}{|X| \text{ in the treebank}} \]
Probabilities

probability of a parse-tree:

- product of probabilities of operations used for construction
probability of a parse-tree:

- product of probabilities of operations used for construction

\[
S
\]

<table>
<thead>
<tr>
<th>operation</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{\alpha_1}$</td>
<td>1.0</td>
</tr>
<tr>
<td>$s_{\alpha_2}$</td>
<td>0.7</td>
</tr>
<tr>
<td>$s_{\alpha_5}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$s_{\alpha_7}$</td>
<td>0.5</td>
</tr>
</tbody>
</table>
probability of a parse-tree:

- product of probabilities of operations used for construction

<table>
<thead>
<tr>
<th>operation</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{\alpha_1}$</td>
<td>1.0</td>
</tr>
<tr>
<td>$s_{\alpha_2}$</td>
<td>0.7</td>
</tr>
<tr>
<td>$s_{\alpha_5}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$s_{\alpha_7}$</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Probabilities

probability of a parse-tree:

- product of probabilities of operations used for construction

<table>
<thead>
<tr>
<th>operation</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{\alpha_1}$</td>
<td>1.0</td>
</tr>
<tr>
<td>$s_{\alpha_2}$</td>
<td>0.7</td>
</tr>
<tr>
<td>$s_{\alpha_5}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$s_{\alpha_7}$</td>
<td>0.5</td>
</tr>
</tbody>
</table>

$1.0 \times 0.7$
## Probabilities

Probability of a parse-tree:
- Product of probabilities of operations used for construction

$$1.0 \times 0.7 \times 0.3$$

<table>
<thead>
<tr>
<th>operation</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{\alpha_1}$</td>
<td>1.0</td>
</tr>
<tr>
<td>$s_{\alpha_2}$</td>
<td>0.7</td>
</tr>
<tr>
<td>$s_{\alpha_5}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$s_{\alpha_7}$</td>
<td>0.5</td>
</tr>
</tbody>
</table>

```
S
  NP
    PRP
    VB
      VP
    NP
      DT
      NN
        a
```

**Example**: `found a NN`
Probabilities

probability of a parse-tree:

- product of probabilities of operations used for construction

<table>
<thead>
<tr>
<th>operation</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{\alpha_1}$</td>
<td>1.0</td>
</tr>
<tr>
<td>$s_{\alpha_2}$</td>
<td>0.7</td>
</tr>
<tr>
<td>$s_{\alpha_5}$</td>
<td>0.3</td>
</tr>
<tr>
<td>$s_{\alpha_7}$</td>
<td>0.5</td>
</tr>
</tbody>
</table>

$1.0 \times 0.7 \times 0.3 \times 0.5$
probabilities

probability of a parse-tree:

- product of probabilities of operations used for construction

\[
1.0 \times 0.7 \times 0.3 \times 0.5 = 0.105
\]
Parsing with PITAG

Improve efficiency by:

- discard tree with lower probability when there are parse trees with same possible operations
- only keep n-best partial parse-trees
Kato et. al

- ITAG: incremental-parsing-oriented tree adjoining grammar
- PITAG: probabilistic ITAG
- Validity of Partial Parse Trees
- Experimental Results and Conclusions
Definitions

σ, τ: partial parse trees

Subsumption

σ subsumes τ, iff τ can be constructed by applying substitution- and adjunction-operations to σ and every parse tree subsumes itself
Definitions

$\sigma, \tau$: partial parse trees

**Subsumption**

$\sigma$ subsumes $\tau$, iff $\tau$ can be constructed by applying substitution- and adjunction-operations to $\sigma$

**Validity**

a partial parse tree is valid for an input sentence, iff it subsumes the correct parse tree for the input sentence
Example

S
   /\  \\
  /   \ \\
 NP   VP
    /\  \\
   /   \
 PRP I

VP*
   /\  \\
  /   \
 VB   NP
    /\  \\
   /   \
 found a dime

PP
in the wood
Example

```
S
  NP
  PRP
  |
  VP
  |
  VP*
     |
     VB
     |
     found
     |
     NP
     |
in the wood

S
  NP
  PRP
  |
  VP
  |
  VB
  |
  found
  |
  NP
  |
a dime
```
Example
Example
Evaluating Validity

validity for a partial parse tree depends on the rest of the sentence
Evaluating Validity

validity for a partial parse tree depends on the rest of the sentence
⇒ dynamically varies for every input word
validity for a partial parse tree depends on the rest of the sentence
⇒ dynamically varies for every input word

\( \sigma \): partial parse tree for initial fragment \( w_1 \cdots w_i \) \((i \leq j)\)

**Conditional Validity of partial parse tree** \( \sigma \)

\[
V(\sigma|w_1 \cdots w_j) = \frac{\sum \text{(Probabilities of partial parse trees at } w_j \text{ subsumed by } \sigma)}{\sum \text{(Probabilities of partial parse trees constructed at } w_j)}
\]
validity for a partial parse tree depends on the rest of the sentence
⇒ dynamically varies for every input word

\( \sigma \): partial parse tree for initial fragment \( w_1 \cdots w_i (i \leq j) \)

**Conditional Validity of partial parse tree \( \sigma \)**

\[
V(\sigma | w_1 \cdots w_j) = \frac{\sum (\text{Probabilities of partial parse trees at } w_j \text{ subsumed by } \sigma)}{\sum (\text{Probabilities of partial parse trees constructed at } w_j)}
\]

example: later
delay output until validity is high enough:
Output Partial Parse Trees

delay output until validity is high enough:

**Parse Tree to be returned**

Parse Tree with the longest initial fragment whose validity is greater than threshold $\theta$

$\theta$: threshold between 0 and 1
## Example

<table>
<thead>
<tr>
<th>word</th>
<th>#</th>
<th>partial parse tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>s</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>[[[I]_{prp}_np_vp]_s</td>
</tr>
<tr>
<td>found</td>
<td>3</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_np_vp]_s</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_np_adjp_vp]_s</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_vp]_s</td>
</tr>
<tr>
<td>a</td>
<td>6</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_nn_np_vp]_s</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_jj_nn_np_vp]_s</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_nn_np_adjp_vp]_s</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_jj_nn_np_adjp_vp]_s</td>
</tr>
<tr>
<td>dime</td>
<td>10</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_vp]_s</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_adjp_vp]_s</td>
</tr>
<tr>
<td>in</td>
<td>12</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_vp_[in_in_np_pp_vp]_s</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_[in_in_np_pp_np_vp]_s</td>
</tr>
<tr>
<td>the</td>
<td>14</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_vp_[in_in_[the]_dt_nn_np_pp_vp]_s</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_vp_[in_in_[the]_dt_jj_nn_np_pp_vp]_s</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_[in_in_[the]_dt_nn_np_pp_vp]_s</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_[in_in_[the]_dt_jj_nn_np_pp_vp]_s</td>
</tr>
<tr>
<td>wood</td>
<td>18</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_vp_[in_in_[the]_dt_[wood]_nn_np_pp_vp]_s</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>[[[I]<em>{prp}_np_[found]</em>{vb}_[a]_{dt}_[dime]_nn_np_[in_in_[the]_dt_[wood]_nn_np_pp_np_vp]_s</td>
</tr>
</tbody>
</table>
Example

I found a dime in the wood

#1 → #2 → #3 → #6 → #10 → #12 → #14 → #18

#4 → #8 → #11

#5

subsumption relation
Example

I found a dime in the wood

<table>
<thead>
<tr>
<th>Parse Tree</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>0.7</td>
</tr>
<tr>
<td>#4</td>
<td>0.1</td>
</tr>
<tr>
<td>#5</td>
<td>0.2</td>
</tr>
<tr>
<td>#6</td>
<td>0.21</td>
</tr>
<tr>
<td>#7</td>
<td>0.14</td>
</tr>
<tr>
<td>#8</td>
<td>0.03</td>
</tr>
<tr>
<td>#9</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Example

I found a dime in the wood

Parse Tree

<table>
<thead>
<tr>
<th>Parse Tree</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>0.7</td>
</tr>
<tr>
<td>#4</td>
<td>0.1</td>
</tr>
<tr>
<td>#5</td>
<td>0.2</td>
</tr>
<tr>
<td>#6</td>
<td>0.21</td>
</tr>
<tr>
<td>#7</td>
<td>0.14</td>
</tr>
<tr>
<td>#8</td>
<td>0.03</td>
</tr>
<tr>
<td>#9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\[ \theta = 0.8 \]
Example

I found a dime in the wood

\[ \theta = 0.8 \]

\[ V(\sigma|w_1 \cdots w_j) = \frac{\sum (\text{Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum (\text{Probabilities of partial parse trees constructed at } w_j)} \]
Example

I found a dime in the wood

<table>
<thead>
<tr>
<th>Parse Tree</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>0.7</td>
</tr>
<tr>
<td>#4</td>
<td>0.1</td>
</tr>
<tr>
<td>#5</td>
<td>0.2</td>
</tr>
<tr>
<td>#6</td>
<td>0.21</td>
</tr>
<tr>
<td>#7</td>
<td>0.14</td>
</tr>
<tr>
<td>#8</td>
<td>0.03</td>
</tr>
<tr>
<td>#9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\[ \theta = 0.8 \]

\[ V(\sigma | w_1 \cdots w_j) = \frac{\sum \text{(Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum \text{(Probabilities of partial parse trees constructed at } w_j)} \]

When will #3 (valid) be output?
Example

V(σ|w₁ ⋯ wₗ) = \frac{\sum(\text{Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum(\text{Probabilities of partial parse trees constructed at } w_j)}

When will #3 (valid) be output? “I found”: θ = 0.8
Example

I found a dime in the wood

Parse Tree | Probability
---|---
#3 | 0.7
#4 | 0.1
#5 | 0.2
#6 | 0.21
#7 | 0.14
#8 | 0.03
#9 | 0.02

$\theta = 0.8$

$$V(\sigma|w_1 \cdots w_j) = \frac{\sum \text{(Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum \text{(Probabilities of partial parse trees constructed at } w_j)}$$

When will #3 (valid) be output? “I found”:

$$V(#3, \text{I found}) = \frac{P(#3)}{P(#3) + P(#4) + P(#5)}$$
Example

I found a dime in the wood

\[
V(\sigma|w_1 \cdots w_j) = \frac{\sum \text{(Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum \text{(Probabilities of partial parse trees constructed at } w_j)}
\]

When will #3 (valid) be output? “I found”:

\[
V(#3, \text{I found}) = \frac{P(#3)}{P(#3) + P(#4) + P(#5)}
\]

\[
\theta = 0.8
\]

<table>
<thead>
<tr>
<th>Parse Tree</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>0.7</td>
</tr>
<tr>
<td>#4</td>
<td>0.1</td>
</tr>
<tr>
<td>#5</td>
<td>0.2</td>
</tr>
<tr>
<td>#6</td>
<td>0.21</td>
</tr>
<tr>
<td>#7</td>
<td>0.14</td>
</tr>
<tr>
<td>#8</td>
<td>0.03</td>
</tr>
<tr>
<td>#9</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Example

\[ \theta = 0.8 \]

When will \#3 (valid) be output? “I found”:

\[ V(\#3, \text{I found}) = \frac{P(\#3)}{P(\#3) + P(\#4) + P(\#5)} = \frac{0.7}{0.7 + 0.1 + 0.2} = 0.7 \]
Example

Introduction

Baumann et al.

Kato et al.

Summary

Example

I found a dime in the wood

Parse Tree

<table>
<thead>
<tr>
<th>Parse Tree</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>0.7</td>
</tr>
<tr>
<td>#4</td>
<td>0.1</td>
</tr>
<tr>
<td>#5</td>
<td>0.2</td>
</tr>
<tr>
<td>#6</td>
<td>0.21</td>
</tr>
<tr>
<td>#7</td>
<td>0.14</td>
</tr>
<tr>
<td>#8</td>
<td>0.03</td>
</tr>
<tr>
<td>#9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

When will #3 (valid) be output? “I found”:

\[ V(#3, \text{I found}) = \frac{P(#3)}{P(#3) + P(#4) + P(#5)} = \frac{0.7}{0.7 + 0.1 + 0.2} = 0.7 < \theta \]

\[ \theta = 0.8 \]
Example

I found a dime in the wood

\[
\begin{align*}
\#1 & \rightarrow \#2 \rightarrow \#3 \rightarrow \#6 \rightarrow \#10 \rightarrow \#12 \rightarrow \#14 \rightarrow \#18 \\
& \quad \rightarrow \#7 \rightarrow \#15 \\
& \quad \rightarrow \#4 \rightarrow \#8 \rightarrow \#11 \rightarrow \#13 \rightarrow \#16 \rightarrow \#19 \\
& \quad \rightarrow \#9 \\
& \quad \rightarrow \#5 \\
\end{align*}
\]

\begin{align*}
\text{Parse Tree} & \quad \text{Probability} \\
\#3 & \quad 0.7 \\
\#4 & \quad 0.1 \\
\#5 & \quad 0.2 \\
\#6 & \quad 0.21 \\
\#7 & \quad 0.14 \\
\#8 & \quad 0.03 \\
\#9 & \quad 0.02 \\
\end{align*}

\[\theta = 0.8\]

\[V(\sigma|w_1 \cdots w_j) = \frac{\sum \text{(Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum \text{(Probabilities of partial parse trees constructed at } w_j)}\]

When will \#3 (valid) be output? “I found a”: 43 / 50
### Example

**Introduction**

Baumann et al.

Kato et al.

**Summary**

---

#### Parse Tree

<table>
<thead>
<tr>
<th>Parse Tree</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>0.7</td>
</tr>
<tr>
<td>#4</td>
<td>0.1</td>
</tr>
<tr>
<td>#5</td>
<td>0.2</td>
</tr>
<tr>
<td>#6</td>
<td>0.21</td>
</tr>
<tr>
<td>#7</td>
<td>0.14</td>
</tr>
<tr>
<td>#8</td>
<td>0.03</td>
</tr>
<tr>
<td>#9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\[ \theta = 0.8 \]

**V(σ|w₁ \cdots wₗ) = \frac{\sum (\text{Probabilities of parse trees at } w_j \text{ subsumed by } σ)}{\sum (\text{Probabilities of partial parse trees constructed at } w_j)} \]

When will #3 (valid) be output? “I found a”:

\[ V(#3, \text{ I found a}) = \frac{P(#6) + P(#7)}{P(#6) + P(#7) + P(#8) + P(#9)} \]

---
Example

When will #3 (valid) be output? “I found a”:

\[ V(\#3, \text{I found a}) = \frac{P(\#6) + P(\#7)}{P(\#6) + P(\#7) + P(\#8) + P(\#9)} \]

\[ \theta = 0.8 \]
Example

I found a dime in the wood

 Parse Tree | Probability
-------------|-------------
 #3          | 0.7         
 #4          | 0.1         
 #5          | 0.2         
 #6          | 0.21        
 #7          | 0.14        
 #8          | 0.03        
 #9          | 0.02        

\[ \theta = 0.8 \]

\[
V(\sigma|w_1 \cdots w_j) = \frac{\sum \text{(Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum \text{(Probabilities of partial parse trees constructed at } w_j)}
\]

When will #3 (valid) be output? “I found a”:

\[
V(#3, \text{ I found a}) = \frac{P(#6) + P(#7)}{P(#6) + P(#7) + P(#8) + P(#9)} = 0.875
\]
Example

I found a dime in the wood

Parse Tree

<table>
<thead>
<tr>
<th>Parse Tree</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>#3</td>
<td>0.7</td>
</tr>
<tr>
<td>#4</td>
<td>0.1</td>
</tr>
<tr>
<td>#5</td>
<td>0.2</td>
</tr>
<tr>
<td>#6</td>
<td>0.21</td>
</tr>
<tr>
<td>#7</td>
<td>0.14</td>
</tr>
<tr>
<td>#8</td>
<td>0.03</td>
</tr>
<tr>
<td>#9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

$\theta = 0.8$

\[ V(\sigma|w_1 \cdots w_j) = \frac{\sum (\text{Probabilities of parse trees at } w_j \text{ subsumed by } \sigma)}{\sum (\text{Probabilities of partial parse trees constructed at } w_j)} \]

When will #3 (valid) be output? “I found a”:

\[ V(#3, \text{I found a}) = \frac{P(#6) + P(#7)}{P(#6) + P(#7) + P(#8) + P(#9)} = 0.875 \geq \theta \]
### Kato et. al

- **ITAG**: incremental-parsing-oriented tree adjoining grammar
- **PITAG**: probabilistic ITAG
- **Validity of Partial Parse Trees**
- **Experimental Results and Conclusions**
Experimental Setup

- parser implemented in Lisp
- input: POS-Sequences
- elementary trees extracted from Penn Treebank
- only used correctly parsed sentences
Degree of delay at j-th word

\[ D(j, s) = j - |\text{output parse tree at } w_j| \]

\( s = w_1 \cdots w_n \): Input sentence
**Measures**

**Degree of delay at j-th word**

\[ D(j, s) = j - |\text{output parse tree at } w_j| \]

\[ s = w_1 \cdots w_n: \text{Input sentence} \]

**Precision**

percentage of valid partial parse trees in the output
Results and Conclusions

<table>
<thead>
<tr>
<th>precision(%)</th>
<th>$D_{\text{max}}$</th>
<th>$D_{\text{ave}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 1.0$</td>
<td>100.0</td>
<td>11.9</td>
</tr>
<tr>
<td>$\theta = 0.9$</td>
<td>97.3</td>
<td>7.5</td>
</tr>
<tr>
<td>$\theta = 0.8$</td>
<td>95.4</td>
<td>6.4</td>
</tr>
<tr>
<td>$\theta = 0.7$</td>
<td>92.5</td>
<td>5.5</td>
</tr>
<tr>
<td>$\theta = 0.6$</td>
<td>88.4</td>
<td>4.5</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>83.0</td>
<td>3.4</td>
</tr>
<tr>
<td>baseline</td>
<td>73.6</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Higher threshold $\Rightarrow$ higher precision, greater delays $\Rightarrow$ trade-off between precision and delay.
Results and Conclusions

higher threshold

<table>
<thead>
<tr>
<th>precision(%)</th>
<th>$D_{max}$</th>
<th>$D_{ave}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 1.0$</td>
<td>100.0</td>
<td>11.9</td>
</tr>
<tr>
<td>$\theta = 0.9$</td>
<td>97.3</td>
<td>7.5</td>
</tr>
<tr>
<td>$\theta = 0.8$</td>
<td>95.4</td>
<td>6.4</td>
</tr>
<tr>
<td>$\theta = 0.7$</td>
<td>92.5</td>
<td>5.5</td>
</tr>
<tr>
<td>$\theta = 0.6$</td>
<td>88.4</td>
<td>4.5</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>83.0</td>
<td>3.4</td>
</tr>
<tr>
<td>baseline</td>
<td>73.6</td>
<td>0.0</td>
</tr>
</tbody>
</table>

(higher threshold)
Results and Conclusions

higher threshold
⇒ higher precision, greater delays
higher threshold
⇒ higher precision, greater delays
⇒ trade-off between precision and delay

\[
\begin{array}{cccc}
\theta & \text{precision(\%)} & D_{max} & D_{ave} \\
1.0 & 100.0 & 11.9 & 6.4 \\
0.9 & 97.3 & 7.5 & 2.9 \\
0.8 & 95.4 & 6.4 & 2.2 \\
0.7 & 92.5 & 5.5 & 1.8 \\
0.6 & 88.4 & 4.5 & 1.3 \\
0.5 & 83.0 & 3.4 & 0.9 \\
\text{baseline} & 73.6 & 0.0 & 0.0 \\
\end{array}
\]
**Summary**

**Baumann et. al**

use right context and message smoothing to reduce/avoid wrong hypotheses in incremental ASR

**Kato et al.**

...
Summary

Baumann et al.
use right context and message smoothing to reduce/avoid wrong hypotheses in incremental ASR

Kato et al.
use PITAG to reduce/avoid wrong hypotheses in incremental parsing
Summary

**Baumann et. al**
use right context and message smoothing to reduce/avoid wrong hypotheses in incremental ASR

**Kato et. al**
use PITAG to reduce/avoid wrong hypotheses in incremental parsing

delay of output
Baumann et al. use right context and message smoothing to reduce/avoid wrong hypotheses in incremental ASR.

Kato et al. use PITAG to reduce/avoid wrong hypotheses in incremental parsing.

delay of output

⇒ Trade-off between incrementality/speed and output quality (revisions)
Thank you!

Thanks for your attention!
References
