Uncertain input and noisy-channel sentence comprehension

Roger Levy
UC San Diego

LSA Institute 2011
26 July 2011
Garden Paths: What we don’t understand so well

• In previous days we covered “ideal-observer” models of syntactic comprehension
• Today we’ll add nuance to this picture
  • We don’t ever see or hear words, only percepts
  • Comprehension is not entirely passive, it involves action
• I’ll motivate the modeling work by looking at some puzzles for the “ideal-observer” view of the comprehender
• This will lead us to new models and empirical results
A bit of review

- Let’s take your impressions with a variety of garden-path sentences

The horse raced past the barn fell

Since Jay always jogs a mile and a half seems like a very short distance to him.
Puzzle 1: global inference

- Try to read & comprehend this sentence:
  \textit{While Mary bathed the baby spat up in the bed.}

- And now let’s do a little math:
  \[8095 - 5107 + 4043 = 7031\]

- Question: if the sentence was true, must it follow…
  
  \begin{itemize}
    \item …that \textit{Mary bathed the baby}???
    \item …that \textit{the baby spat up in the bed}???
  \end{itemize}

- Readers tend to answer yes to both questions!!!

- They’re garden-pathed at first…

- …and then recover wrongly into some hybrid meaning

- \textit{Major problem for rational sentence processing theories: inferences incompatible with the complete sentence}

  \textit{(Christianson et al., 2001)}
Puzzle 2: incremental inference

- Try to understand this sentence:

(a) *The coach smiled at the player tossed the frisbee.*

...and contrast this with:

(b) *The coach smiled at the player thrown the frisbee.*

(c) *The coach smiled at the player who was thrown the frisbee.*

(d) *The coach smiled at the player who was tossed the frisbee.*

- Readers boggle at “tossed” in (a), but not in (b-d)

Tabor et al. (2004, JML)
Why is *tossed/thrown* interesting?

- As with classic garden-paths, part-of-speech ambiguity leads to misinterpretation
  - *The horse* raced past the barn…fell
    - verb?
    - participle?

- But now context “should” rule out the garden path:
  - *The coach smiled at the player* tossed…
    - verb?
    - participle?

- *Another challenge for rational models: failure to condition on relevant context*
Rational sentence comprehension

Online sentence comprehension is hard, due to:
- Ambiguity
- Uncertainty
- Attentional/memory limitations
- Environmental noise

But lots of information sources are available to help with the task

Therefore, it would be *rational* for people to use *all the information available, as soon as possible*

Leading models: fully incremental parsing via (generative) probabilistic grammars
Uncertain input in language comprehension

• State of the art models for ambiguity resolution ≈ probabilistic incremental parsing

• Simplifying assumption:
  • Input is *clean* and *perfectly-formed*
  • No uncertainty about input is admitted

• Intuitively seems patently wrong…
  • We sometimes *misread* things
  • We can also *proofread*

• Leads to two questions:

  1. What might a model of sentence comprehension under uncertain input look like?

  2. What interesting consequences might such a model have?
Today: a first-cut answer

- **First**: a simple noisy-channel model of rational sentence comprehension under uncertain input
- **Then**: we’ll solve the two psycholinguistic puzzles
  1. *global inference*
  2. *incremental inference*
- We use probabilistic context-free grammars (PCFGs) and weighted finite-state automata (WFSAs) to instantiate the model
- In each case, input uncertainty solves the puzzle

1. What might a model of sentence comprehension under uncertain input look like?
2. What interesting consequences might such a model have?
The noisy-channel model

- Say we use a weighted generative grammar $G$ to parse a sentence $w$. We get a posterior over structures $T$:

$$P_G(T|w) = \frac{P(T, w)}{P(w)} \propto P(T, w)$$

- Bayes’ Rule

- If we don’t observe a sentence but only a noisy input $I$:

$$P_G(T|I) \propto \sum_w P(I|T, w)P(w|T)P(T)$$

- Posterior over possible sentences:

$$P_G(w|I) \propto \sum_T P(I|T, w)P(w|T)P(T)$$

Levy (2008, EMNLP)
The noisy-channel model (II)

- This much is familiar from the parsing of speech (Hall & Johnson, 2003, 2004; Johnson & Charniak, 2004)
- Alternative scenario: we know the true sentence $w^*$ but not observed input $I$ (e.g., the study of reading)
- Expected inferences of the comprehender marginalize over the input $I$:

$$P(w|w^*) = \int_I P_C(w|I, w^*) P_T(I|w^*) \, dI$$

$$\propto Q(w, w^*)$$
Interlude

• Now see slides on weighted finite-state automata and weighted PCFGs...
Representing noisy input

- How can we represent the type of noisy input generated by a word sequence?
- *Probabilistic finite-state automata* (pFSAs; Mohri, 1997) are a good model

\[
vocab = a, b, c, d, e, f
\]

"Word 1 is a or b, and I have no info about Word 2"
Probabilistic finite-state automata

- A probabilistic finite-state automaton (PFSA) is:
  - A finite set $q_0, q_1, \ldots, q_n$ of states; $q_0$ is the start state
  - A finite set $V$ of input symbols
  - A set of transitions $<x, q_i \rightarrow q_j>$ where $x$ is in $V$
  - A probability function $P(<x, q_i \rightarrow q_j>)$ such that
    \[
    \sum_{q_j} P(<x, q_i \rightarrow q_j>) = 1
    \]
Probabilistic Linguistic Knowledge

- A generative probabilistic grammar determines beliefs about which strings are likely to be seen
  - Probabilistic Context-Free Grammars (PCFGs; Booth, 1969)
  - Probabilistic Minimalist Grammars (Hale, 2006)
  - Probabilistic Finite-State Grammars (Mohri, 1997; Crocker & Brants 2000)

In position 1, \{a,b,c,d\} equally likely; but in position 2:
- \{a,b\} are usually followed by e, occasionally by f
- \{c,d\} are usually followed by f, occasionally by e
Combining grammar & uncertain input

- Bayes’ Rule says that the evidence and the prior should be combined (multiplied)
- For probabilistic grammars, this combination is the formal operation of weighted intersection

[Diagram]

Grammar affects beliefs about the future

BELIEF
Revising beliefs about the past

• When we’re uncertain about the future, grammar + partial input can affect beliefs about what will happen
• With uncertainty of the past, grammar + future input can affect beliefs about what has already happened
The noisy-channel model (FINAL)

\[ P(w|w^*) \propto P_C(w)Q(w, w^*) \]

- For \( Q(w, w^*) \): a WFSA based on Levenshtein distance between words (\( K_{LD} \)):

Result of \( K_{LD} \) applied to \( w^* = \text{a cat sat} \):

\( \text{Cost(a cat sat)} = 0 \)

\( \text{Cost(sat a sat cat)} = 8 \)
Puzzle 1: recap

While Mary bathed the baby spat up in the bed.

• Readers tend to answer “yes” to both:
  1. Did Mary bathe the baby?
  2. Did the baby spit up in the bed?
What does our uncertain-input theory say?

• In near-neighbor sentences Mary does bathe the baby:

  (a) While Mary bathed the baby *it* spat up in the bed.

  (b) While Mary bathed *it* the baby spat up in the bed.

• (a-b) are “near” \( w^* \) in Levenshtein-distance space

• Our theory may then explain this result...

• ...if the comprehender’s grammar can push them into inferring structures more like (a-b)
Testing the intuition of our theory

\[ P(w \mid w^*) \propto P_C(w)Q(w, w^*) \]

- The Levenshtein-distance kernel \( K_{LD} \) gives us \( Q(w, w^*) \)
- A small PCFG can give us \( P_C(w) \)
- Recall that \( K_{LD} \) is a WFSA
  - So \( P(w \mid w^*) \) is a weighted intersection of \( K_{LD} \) with \( P_C \)
- **Metric of interest**: % of 100-best parses (Huang & Chiang, 2005) in which “Mary really does hunt the baby”:

  While Mary bathed the baby *[pronoun]* spat up in the bed.
  
  While Mary bathed *[pronoun]* the baby spat up in the bed
Noisy-channel inference with probabilistic grammars

- Yesterday I covered the probabilistic Earley algorithm
- The chart contains a grammar! (Lang, 1988)

NP_{[0,2]} \rightarrow \text{Det}_{[0,1]} \text{ N}_{[1,2]}
\text{Det}_{[0,1]} \rightarrow \text{the}_{[0,1]}
\text{N}_{[1,2]} \rightarrow \text{dog}_{[1,2]}
Intersection of a PCFG and a wFSA

• Generalizes from incomplete sentences to arbitrary wFSAs

• PCFGs are closed under intersection with wFSAs

1. For every rule \( X \rightarrow Y Z \) with prob. \( p \) and state triple \((q_i, q_j, q_k)\), construct rule \( X_{[i,k]} \rightarrow Y_{[i,j]} Z_{[j,k]} \) with weight \( p \)

2. For every transition \( <x, q_i \rightarrow q_j> \) with weight \( w \), construct rule \( x_{[i,j]} \rightarrow x \) with weight \( w \)

3. Normalize!

(Bar-Hillel et al., 1964; Nederhof & Satta, 2003)
Testing the intuition: results

- **GardenPath**: While Mary bathed the baby spat up in the bed
- **Comma**: While Mary bathed, baby spat up in the bed
- **Transitive**: While Mary bathed the girl the baby spat up in the bed

Model & human misinterpretations match
Puzzle 2: recap

• This sentence…

(a) *The coach smiled at the player tossed the frisbee.*

…is harder than these sentences…

(b) *The coach smiled at the player thrown the frisbee.*

(c) *The coach smiled at the player who was thrown the frisbee.*

(d) *The coach smiled at the player who was tossed the frisbee.*

• And the difficulty is localized at *tossed*
Incremental inference under uncertain input

- Near-neighbors make the “incorrect” analysis “correct”:

  The coach smiled at the player **tossed** the frisbee

  \[
  \begin{array}{ccc}
  \text{(that?)} & \text{(and?)} & \text{(and?)} \\
  \text{(who?)} & \text{(as?)} & \text{(that?)} \\
  \end{array}
  \]

  Any of these changes makes **tossed** a main verb!!!

- Hypothesis: the boggle at “tossed” involves **what the comprehender wonders whether she might have seen**
The core of the intuition

• Grammar & input come together to determine two possible “paths” through the partial sentence: (line thickness ≈ probability)

• \textit{tossed} is more likely to happen along the bottom path
  • This creates a large shift in belief in the \textit{tossed} condition

• \textit{thrown} is very unlikely to happen along the bottom path
  • As a result, there is no corresponding shift in belief
Incremental inference under uncertain input

- Traditionally, the input to a sentence-processing model has been a *sequence of words*
- But really, input to sentence processor should be more like the output of a word-recognition system

The coach smiled at the player *tossed* the frisbee

- That means that the possibility of *misreading/mishearing* words must be accounted for
- On this hypothesis, local-coherence effects are about *what the comprehender wonders whether she might have seen*
Inference through a noisy channel

- So how can we model sentence comprehension when the input is still noisy?
- A generative probabilistic grammatical model makes inference over uncertain input possible
  - This is the **noisy channel** from NLP/speech recognition
  - Inference involves Bayes’ Rule

\[
P(\text{words}|\text{input, grammar}) \propto P(\text{input}|\text{words, grammar}) P(\text{words}|\text{grammar}) P(\text{words}|\text{grammar}) \propto P(\text{input}|\text{words}) P(\text{words}|\text{grammar}) \quad \text{[by assumption]}
\]

**Prior:** Comprehender’s knowledge of language

**Evidence:** Noisy input probability, dependent only on the “words” generating the input
How does this relate to local-coherence effects?

Here’s an oversimplified noisy-input representation of the offending sentence:

*The coach smiled at the player tossed the frisbee.*
Here’s a hand-written finite-state grammar of reduced relative clauses
Ingredients for the model

\[ P(w|w^*) \propto P_C(w)Q(w, w^*) \]

- \( Q(w, w^*) \) comes from \( K_{LD} \) (with minor changes)

- \( P_C(w) \) comes from a probabilistic grammar (this time finite-state)

- We need one more ingredient:
  - a quantified signal of the alarm induced by word \( w_i \) about changes in beliefs about the past
Quantifying alarm about the past

- **Relative Entropy** (KL-divergence) is a natural metric of change in a probability distrib. (Levy, 2008; Itti & Baldi, 2005)
- Our distribution of interest is *probabilities over the previous words in the sentence*
- Call this distribution $P_i(w_{[0,j]})$
- The change induced by $w_i$ is the **error identification signal** $EIS_i$, defined as

$$D \left( P_i \left( w_{[0,i]} \right) \right) \left\| P_{i-1} \left( w_{[0,i]} \right) \right)$$

- new distribution
- old distribution
Error identification signal: example

- Measuring change in beliefs about the past:

  \( \{a, b\} \{?\} \quad \text{No change: } EIS_2 = 0 \quad \{a, b\} \{f, e\} \)

\[
\sum_{w_1} P_2(w_1) \log \frac{P_2(w_1)}{P_1(w_1)} = \begin{cases} 
0.28 (1 - 1.82) + 0.72 (1 - 0.48) = 0.14 
\end{cases}
\]
Results on local-coherence sentences

- Locally coherent: *The coach smiled at the player tossed the frisbee*
- Locally incoherent: *The coach smiled at the player thrown the frisbee*

(All sentences of Tabor et al. 2004 with lexical coverage in model)
Novel applications of the model

• Theoretical recap:
  • Comprehension inferences involve trade-offs between *uncertain perception* and *prior grammatical expectations*
  • We saw how model may account for two existing results

• Novel prediction 1:
  • Uncertain-input effects should be *dependent on the perceptual neighborhood* of the sentence

• Novel prediction 2:
  • With strongly enough biased grammatical expectations, *comprehenders may be pushed into “hallucinating” garden paths* where the input itself doesn’t license them

• Novel modeling application:
  • Comprehension as *action*: where to move the eyes during reading?
Prediction 1: neighborhood manipulation

- Uncertain-input effects should be dependent on the perceptual neighborhood of the sentence
- Resulting novel prediction: changing neighborhood of the context can affect EIS & thus comprehension behavior

The coach smiled at the player tossed the frisbee

The coach smiled toward the player tossed the frisbee

- Substituting toward for at should reduce the EIS
- In free reading, we should see less tendency to regress from tossed when the EIS is small

(Levy, Bicknell, Slattery, & Rayner, 2009, PNAS)
Model predictions

(0.00, 0.05, 0.10, 0.15, 0.20)

EIS

Noise level (low=noisy)

at...tossed

toward...tossed

toward...thrown

at...thrown

(The coach smiled at/toward the player tossed/thrown the frisbee)
Experimental design

- In a free-reading eye-tracking study, we crossed *at*/*toward* with *tossed*/*thrown*:

  The coach smiled *at* the player *tossed* the frisbee
  The coach smiled *at* the player *thrown* the frisbee
  The coach smiled *toward* the player *tossed* the frisbee
  The coach smiled *toward* the player *thrown* the frisbee

- Prediction: interaction between preposition & ambiguity in some subset of:
  - Early-measure RTs at critical region *tossed*/*thrown*
  - First-pass regressions out of critical region
  - Go-past time for critical region
  - Regressions into *at*/*toward*
Experimental results

The coach smiled at the player...
Question-answering accuracy

- 16 of 24 questions were about the RRC, equally divided:
  - *The coach smiled at the player tossed a frisbee by the opposing team*
    - Did the player toss/throw a frisbee? [NO]
    - Did someone toss/throw the player a frisbee? [YES]
    - Did the player toss the opposing team a frisbee? [NO]
    - Did the opposing team toss the player a frisbee? [YES]

**Significant main effect of question type** ($p_z < 0.01$)

**Significant interaction of question type w/ ambiguity**

($p_z < 0.05$)
What this result tells us

- Readers must have residual uncertainty about word identity
  - Word misidentification alone won’t get this result in a fully incremental model:

    The coach smiled toward the player. **thrown**
    The coach smiled at the player. **thrown**
    The coach smiled as the player. **thrown**

    The coach smiled toward the player. **tossed**
    The coach smiled at the player. **tossed**
    The coach smiled as the player. **tossed**

  - Also, readers respond to changes in uncertainty in a sensible way

  Should be about equally hard
  Should be easier, if anything
Prediction 2: hallucinated garden paths

• Try reading the sentence below:

While the soldiers marched, toward the tank lurched an injured enemy combatant.

• There’s a garden-path clause in this sentence…

• …but it’s interrupted by a comma.

• Readers are ordinarily very good at using commas to guide syntactic analysis:

While the man hunted, the deer ran into the woods

While Mary was mending the sock fell off her lap

• “With a comma after mending there would be no syntactic garden path left to be studied.” (Fodor, 2002)

• We’ll see that the story is slightly more complicated.

(Levy, 2011, ACL)
Prediction 2: hallucinated garden paths

While the soldiers marched, toward the tank lurched an injured enemy combatant.

- This sentence is comprised of an initial intransitive subordinate clause…
- …and then a main clause with *locative inversion* (Bolinger, Bresnan, 1994).
  (c.f. *an injured enemy combatant lurched toward the tank*).
- Crucially, the main clause’s initial PP would make a great dependent of the subordinate verb…
- …but doing that *would require the comma to be ignored*.
- Inferences through …*tank* should thus involve a tradeoff between perceptual input and prior expectations.
While the soldiers marched…

\[ P(w_i|\text{Context}) = \sum_{\text{Path}} P(w_i|\text{Path, Context}) P(\text{Path}|\text{Context}) \]

• Inferences as probabilistic paths through the sentence:
  • Perceptual cost of ignoring the comma
  • Unlikeliness of main-clause continuation after comma
  • Likeliness of postverbal continuation without comma
• These inferences together make \textit{soared} very surprising!
Prediction 2: Hallucinated garden paths

- Methodology: word-by-word self-paced reading

As the soldiers marched, toward the tank lurched.

- Readers aren’t allowed to backtrack
- So the comma is visually *gone* by the time the inverted main clause appears
- Simple test of whether beliefs about previous input can be revised
As the soldiers marched, toward the tank lurched...

As the soldiers marched into the bunker, toward the tank lurched...

As the soldiers marched, the tank lurched toward...

As the soldiers marched into the bunker, the tank lurched toward....
Results: word-by-word reading times

Processing boggle occurs exactly where predicted.

As the soldiers marched, into the bunker, toward the tank, lurched toward an enemy combatant.
The way forward

- Broader future goal: develop eye-movement control model integrating the insights discussed thus far:
  - Probabilistic linguistic knowledge
  - Uncertain input representations
  - Principles of adaptive, rational action
- *Reinforcement learning* is an attractive tool for this

*(Bicknell & Levy, 2010)*
A rational reader

- Very simple framework:
  - Start with prior expectations for text (linguistic knowledge)
  - Move eyes to get perceptual input
  - Update beliefs about text as visual arrives (Bayes’ Rule)
- Add to that:
  - Set of actions the reader can take in discrete time
  - A behavior policy: how the model decides between actions

*(Bicknell & Levy, 2010; Bicknell, 2011)*
A first-cut behavior policy

- Actions: *keep fixating; move the eyes; or stop reading*
- Simple behavior policy with two parameters: $\alpha$ and $\beta$
- Define *confidence* in a character position as the probability of the most likely character

> From the closet, she pulled out a *acket for the upcoming game

*Confidence=0.59*

- Move left to right, bringing up confidence in each character position until it reaches $\alpha$
- If confidence in a previous character position drops below $\beta$, regress to it
- Finish reading when you’re confident in everything

$$P(\text{jacket})=0.38$$
$$P(\text{racket})=0.59$$
$$P(\text{packet})=0.02$$
...
(Non)-regressive policies

- Non-regressive policies have $\beta = 0$
- Hypothesis: non-regressive policies strictly dominated
- Test: estimate speed and accuracy of various policies on reading the Schilling et al. (1998) corpus.

Non-regressive policies always beaten by some regressive policy
Goal-based adaptation

• Open frontier: modeling the adaptation of eye movements to specific reader goals

• We set a *reward function*: relative value $\gamma$ of speed (finish reading in $T$ timesteps) versus accuracy (guess correct sentence with probability $L$)

• PEGASUS simplex-based optimization (Ng & Jordan, 2000)

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.025</td>
<td>0.1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

• The method works, and gives intuitive results

(Bicknell & Levy, 2010)
Empirical match with human reading

• Benchmark measures in eye-movement modeling:

Other models (E-Z Reader, SWIFT) get these too, but stipulate rel’nship between word properties & “processing rate”

We derive these relationships from simple principles of noisy-channel perception and rational action

\[ \gamma = 0.05 \]
Open questions

- Effect of \textit{word length} on \textit{whether a word is fixated} is sensible
- But effect of \textit{word length} on \textit{how long a word is fixated} is weird

We think that this is because our model's (a) lexicon is too sparse; and (b) representation of word length knowledge (veridical) is too optimistic.
Reinforcement learning: summary

- Beginnings of a framework exploring interplay of:
  - Probabilistic linguistic knowledge
  - Uncertainty in input representations
  - Adaptive, goal-driven eye movement control
- We have some initial validation of viability of the modeling framework
- **But**, true payoff remains to be seen in future:
  - More expressive policy families
  - Quantitative comparison with human eye movement patterns
The Probabilistic grammars used

<table>
<thead>
<tr>
<th>Puzzle 1</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROOT → S PUNCT.</td>
<td>0.0</td>
</tr>
<tr>
<td>S → SBAR S</td>
<td>6.3</td>
</tr>
<tr>
<td>S → SBAR PUNCT_S</td>
<td>4.6</td>
</tr>
<tr>
<td>PUNCT_S → , S</td>
<td>0.0</td>
</tr>
<tr>
<td>S → NP VP</td>
<td>0.1</td>
</tr>
<tr>
<td>SBAR → IN S</td>
<td>0.0</td>
</tr>
<tr>
<td>NP → DT NN</td>
<td>1.9</td>
</tr>
<tr>
<td>NP → NNS</td>
<td>4.4</td>
</tr>
<tr>
<td>NP → NNP</td>
<td>3.3</td>
</tr>
<tr>
<td>NP → DT NNS</td>
<td>4.5</td>
</tr>
<tr>
<td>NP → PRP</td>
<td>1.3</td>
</tr>
<tr>
<td>NP → NN</td>
<td>3.1</td>
</tr>
<tr>
<td>VP → VBD RB</td>
<td>9.7</td>
</tr>
<tr>
<td>VP → VBD PP</td>
<td>2.2</td>
</tr>
<tr>
<td>VP → VBD NP</td>
<td>1.2</td>
</tr>
<tr>
<td>VP → VBD RP</td>
<td>8.3</td>
</tr>
<tr>
<td>VP → VBD</td>
<td>2.0</td>
</tr>
<tr>
<td>VP → VBD JJ</td>
<td>3.4</td>
</tr>
<tr>
<td>PP → IN NP</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Puzzle 2</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROOT → S</td>
<td>0.00</td>
</tr>
<tr>
<td>S → S-base CC S-base</td>
<td>7.3</td>
</tr>
<tr>
<td>S → S-base</td>
<td>0.01</td>
</tr>
<tr>
<td>S-base → NP-base VP</td>
<td>0</td>
</tr>
<tr>
<td>NP → NP-base RC</td>
<td>4.1</td>
</tr>
<tr>
<td>NP → NP-base</td>
<td>0.5</td>
</tr>
<tr>
<td>NP → NP-base PP</td>
<td>2.0</td>
</tr>
<tr>
<td>NP-base → DT N N</td>
<td>4.7</td>
</tr>
<tr>
<td>NP-base → DT N</td>
<td>1.9</td>
</tr>
<tr>
<td>NP-base → DT JJ N</td>
<td>3.8</td>
</tr>
<tr>
<td>NP-base → PRP</td>
<td>1.0</td>
</tr>
<tr>
<td>NP-base → NNP</td>
<td>3.1</td>
</tr>
<tr>
<td>VP/NP → V NP</td>
<td>4.0</td>
</tr>
<tr>
<td>VP/NP → V</td>
<td>0.1</td>
</tr>
<tr>
<td>VP → V PP</td>
<td>2.0</td>
</tr>
<tr>
<td>VP → V NP</td>
<td>0.7</td>
</tr>
<tr>
<td>VP → V</td>
<td>2.9</td>
</tr>
<tr>
<td>RC → WP S/NP</td>
<td>0.5</td>
</tr>
<tr>
<td>RC → VP-pass/NP</td>
<td>2.0</td>
</tr>
<tr>
<td>RC → WP FinCop VP-pass/NP</td>
<td>4.9</td>
</tr>
<tr>
<td>PP → IN NP</td>
<td>0</td>
</tr>
<tr>
<td>S/NP → VP</td>
<td>0.7</td>
</tr>
<tr>
<td>S/NP → NP-base VP/NP</td>
<td>1.3</td>
</tr>
<tr>
<td>VP-pass/NP → VBN NP</td>
<td>2.2</td>
</tr>
<tr>
<td>VP-pass/NP → VBN</td>
<td>0.4</td>
</tr>
</tbody>
</table>