# Frequency Effects in the Processing of Unbounded Dependencies

Marten van Schijndel, William Schuler, and Peter Culicover
Department of Linguistics
The Ohio State University

July 24, 2014

# Background

## Unbounded Dependency

A non-local dependency that potentially spans an unbounded # of lexemes.

e.g. That's {the ball} John kicked ___.
e.g. That's {the ball} Mary said John kicked ___.

This is hard because:

- Filler must be remembered
- Where is the gap?
  - Maybe people use subcategorization bias?

# Background

## Subcategorization Bias

The preference for a lexical item to take a particular type of argument

The girl realized $\begin{cases} \{\text{the house was on fire}\}. \\ \{\text{her potential}\}. \end{cases}$

*Realized* prefers a sentential complement over a noun phrase

# Background

Several studies have investigated subcat usage...

- Mitchell (1987)
- Pickering et al. (2000)
- van Gompel & Pickering (2001)
- Pickering & Traxler (2003)

# Background

## Pickering & Traxler (2003)

(1) That's the plane that the pilot landed carefully behind in the fog at the airport.
(2) That's the truck that the pilot landed carefully behind in the fog at the airport.

# Background

## Pickering & Traxler (2003)

(1) That's the plane that the pilot landed carefully behind in the fog at the airport.
(2) That's the truck that the pilot landed carefully behind in the fog at the airport.

Readers slow down at *landed* in (2)

Suggests they try to link *truck* as the object of *landed* despite:

- *landed* biased for PP complement (e.g. "landed on the ground")
    - 40% PP complement
    - 25% NP complement

∴ Readers initially adopt a transitive interpretation despite subcat bias

# Background

Several studies have investigated subcat usage...

- Mitchell (1987)
- Pickering et al. (2000)
- van Gompel & Pickering (2001)
- Pickering & Traxler (2003)

Suggest subcat information actually isn't used immediately for unbounded dependency processing.

Finding supports multistage models of sentence processing

- Garden Path Model (Frazier, 1987)
- Construal (Frazier & Clifton, 1996)

# Background

More recent studies have revisited this claim

- Staub et al. (2006)
- Staub (2007)
- Arai & Keller (2013)

Provide indirect evidence that the previous set of results may be driven by frequency effects of larger syntactic structures

Present Contribution:
The current work provides direct evidence in support of this hypothesis

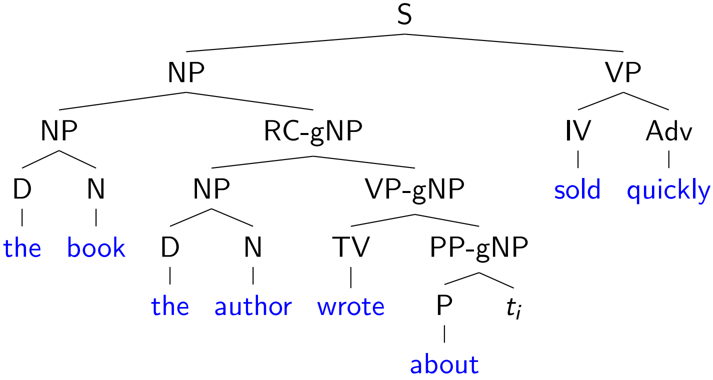
# Conclusion Teaser

## Alternative Explanation

Unbounded dependencies more often go to arguments than modifiers.

# Model

Probabilistic context-free grammar (PCFG) for unbounded dependencies:



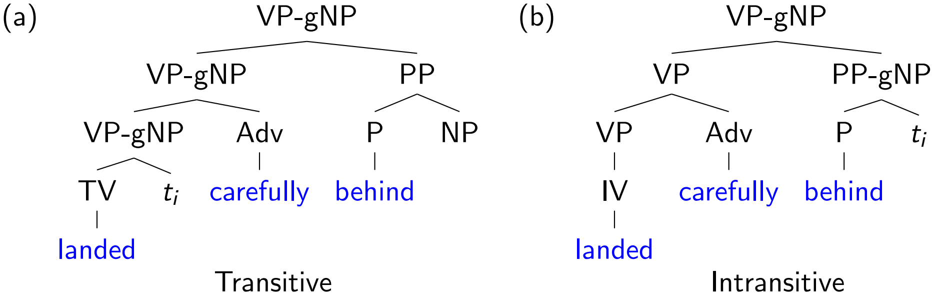

$$P(NP \rightarrow D\ N) = P(D\ N \mid NP) = 0.66$$

$$P(NP \rightarrow NP\ RC\text{-}gNP) = P(NP\ RC\text{-}gNP \mid NP) = 0.33$$

# Model

### Pickering & Traxler (2003)

(1) That's the plane that the pilot landed carefully behind in the fog at the airport.
(2) That's the truck that the pilot landed carefully behind in the fog at the airport.

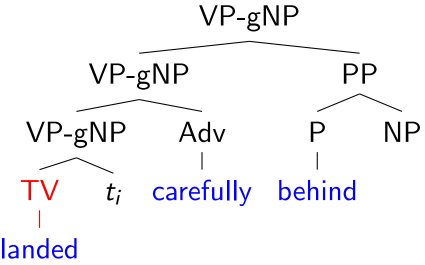


# Model

## Pickering & Traxler (2003)

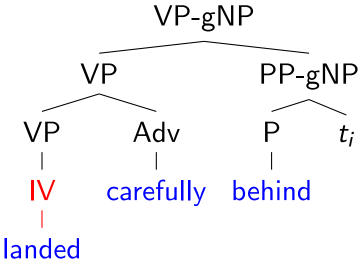
(1) That's the plane that the pilot landed carefully behind in the fog at the airport.
(2) That's the truck that the pilot landed carefully behind in the fog at the airport.

(a)




Transitive

(b)



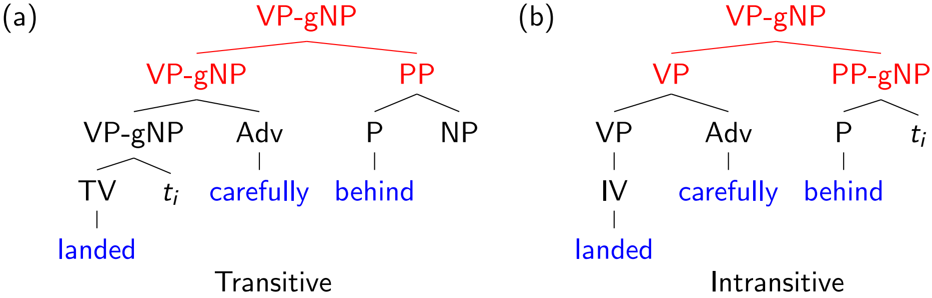

Intransitive

# Model

## Pickering & Traxler (2003)

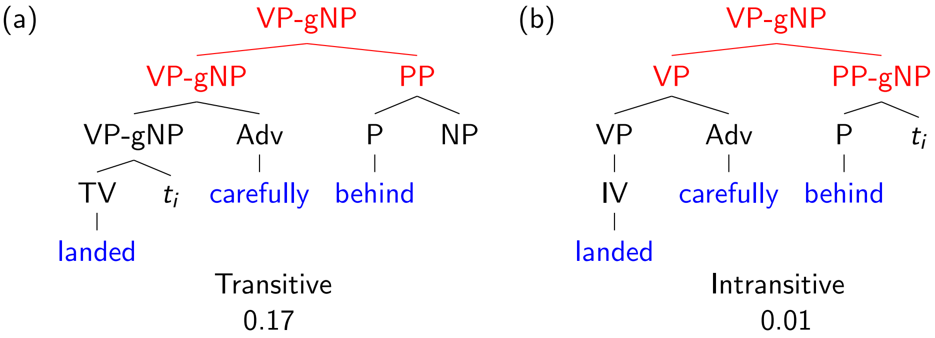
(1) That's the plane that the pilot landed carefully behind in the fog at the airport.

(2) That's the truck that the pilot landed carefully behind in the fog at the airport.



# Model

How probable is each subtree?
Wall Street Journal (WSJ) section of the Penn Treebank:



# Model

What is the probability of each interpretation?
P(syntactic configuration)·P(generating the verb from that tree)

$$\text{P(Transitive)} = \text{P}(\text{VP-gNP} \rightarrow \text{VP-gNP} \;\; \text{PP}) \cdot \text{P}(\textit{verb} \mid \text{TV}) \tag{1}$$

$$\text{P(Intransitive)} = \text{P}(\text{VP-gNP} \rightarrow \text{VP} \;\; \text{PP-gNP}) \cdot \text{P}(\textit{verb} \mid \text{IV}) \tag{2}$$

# Model

What is the probability of each interpretation?

P(syntactic configuration) · ~~P(generating the verb from that tree)~~

P(subcat bias)/P(preterminal prior)

$$\text{P(Transitive)} = \text{P(VP-gNP} \rightarrow \text{VP-gNP PP)} \cdot \text{P(}\textit{verb} \mid \text{TV)} \qquad (1)$$

$$\text{P(Intransitive)} = \text{P(VP-gNP} \rightarrow \text{VP PP-gNP)} \cdot \text{P(}\textit{verb} \mid \text{IV)} \qquad (2)$$

# Model

What is the probability of each interpretation?
P(syntactic configuration)·P(subcat bias)/P(preterminal prior)

$$\begin{aligned} \text{P(Transitive)} &= \text{P(VP-gNP}\rightarrow\text{VP-gNP PP)} \cdot \text{P}(\textit{verb} \mid \text{TV}) && (1)\\ &\propto \text{P(VP-gNP}\rightarrow\text{VP-gNP PP)} \cdot \frac{\text{P(TV} \mid \textit{verb})}{\text{P(TV)}} \\ \text{P(Intransitive)} &= \text{P(VP-gNP}\rightarrow\text{VP PP-gNP)} \cdot \text{P}(\textit{verb} \mid \text{IV}) && (2)\\ &\propto \text{P(VP-gNP}\rightarrow\text{VP PP-gNP)} \cdot \frac{\text{P(IV} \mid \textit{verb})}{\text{P(IV)}} \end{aligned}$$

# Model

What are the preterminal priors?
Relative prior probability from the WSJ:

$$P(\text{TV}) : P(\text{IV}) = 2.6 : 1$$

# Model

What is the probability of each interpretation?
P(syntactic configuration)·P(subcat bias)/P(preterminal prior)

$$
\begin{aligned}
\text{P(Transitive)} &\propto \text{P(VP-gNP}\rightarrow\text{VP-gNP} \;\; \text{PP)} \cdot \frac{\text{P(TV} \mid \textit{verb}\text{)}}{\text{P(TV)}} \\
&= 0.17 \cdot \frac{\text{P(TV} \mid \textit{verb}\text{)}}{2.6} = 0.065 \cdot \text{P(TV} \mid \textit{verb}\text{)} \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
\text{P(Intransitive)} &\propto \text{P(VP-gNP}\rightarrow\text{VP} \;\; \text{PP-gNP)} \cdot \frac{\text{P(IV} \mid \textit{verb}\text{)}}{\text{P(IV)}} \\
&= 0.01 \cdot \frac{\text{P(IV} \mid \textit{verb}\text{)}}{1.0} = 0.01 \cdot \text{P(IV} \mid \textit{verb}\text{)} \qquad (2)
\end{aligned}
$$

Pickering & Traxler (2003) experimentally determined subcat biases for a wide variety of verbs

# Evaluation

## Pickering & Traxler (2003)

(1) That's the plane that the pilot landed carefully behind in the fog at the airport.
(2) That's the truck that the pilot landed carefully behind in the fog at the airport.

Using Pickering & Traxler's (2003) subcat bias data:

$$P(\text{Transitive} \mid \text{landed}) \propto 0.17 \cdot \frac{0.25}{2.6} = 0.016$$
$$P(\text{Intransitive} \mid \text{landed}) \propto 0.01 \cdot \frac{0.40}{1.0} = 0.004$$

Transitive interpretation is 300% more likely!

# Evaluation

| | Bias | | | P(Interpretation) | | |
|---|---|---|---|---|---|---|
| Verb | Trans | Intrans | Str | Trans | Intrans | Str |
| spoke | 0 | **55** | 100 | 0 | **0.55** | 100 |
| worried | 0 | **50** | 100 | 0 | **0.50** | 100 |
| pointed | 10 | **90** | 90 | 0.65 | **0.90** | 58 |
| fished | 5 | **45** | 90 | 0.33 | **0.45** | 58 |
| argued | 11 | **64** | 85 | **0.72** | 0.64 | 53 |
| searched | 15 | **75** | 83 | **0.98** | 0.75 | 57 |
| communicated | 10 | **50** | 83 | **0.65** | 0.50 | 57 |
| shouted | 10 | **50** | 83 | **0.65** | 0.50 | 57 |
| swore | 6 | **17** | 74 | **0.39** | 0.17 | 70 |
| travelled | 20 | **40** | 67 | **1.31** | 0.40 | 77 |
| landed | 25 | **40** | 62 | **1.63** | 0.40 | 80 |
| raced | 35 | **55** | 61 | **2.29** | 0.55 | 81 |
| blabbed | 30 | **45** | 60 | **1.96** | 0.45 | 81 |
| preached | 30 | **45** | 60 | **1.96** | 0.45 | 81 |

Verbal bias percentages (Pickering & Traxler 2003)

# Evaluation: Unaccusatives

Obligatorily intransitive verbs (e.g. erupt) do not cause such a slow down

(Staub 2007)

Current model explains this via 0 transitive bias:

$$\mathsf{P(Transitive)} \propto 0.17 \cdot \frac{0.0}{2.6} = 0.0$$

# Criticism

## WSJ won't generalize

- Subcat biases determined experimentally by P&T (2003)
- Nguyen et al.'s (2012) results suggest WSJ unbounded dependency distribution may generalize
- The current work accounts for a variety of findings. . .
    - Pickering & Traxler (2003), etc. "lack" of subcat usage
    - Staub et al. (2006) heavy-NP shift processing heuristics
    - Staub (2007) unaccusative subcat usage

# Conclusions

- Unbounded dependencies more often go to arguments than modifiers
- Previous studies of subcat bias confounded by syntactic frequency
  - Replication possible when subcat biases taken into immediate account
  - Weakens support for multistage models of sentence processing
- Shows the need to account for frequency at multiple levels of processing; not simply in terms of lexical bias

# Questions?

Thanks to:

- Matthew Traxler
- Shari Speer
- OSU Syntactic Processing Seminar
- OSU Linguistics Targeted Investment for Excellence (2012-2013)

# Bibliography I


Arai, M. and Keller, F. (2013).
The use of verb-specific information for prediction in sentence processing.
*Language and Cognitive Processes*, 28(4):525–560.

Frazier, L. (1987).
Sentence processing: A tutorial review.
In Coltheart, M., editor, *Attention and Performance 12: The Psychology of Reading*, pages 559–586. Erlbaum, Hillsdale, NJ.

Frazier, L. and Clifton, Jr, C. (1996).
*Construal*.
MIT Press, Cambridge, MA.

# Bibliography II


Mitchell, D. C. (1987).
Lexical guidance in human parsing: Locus and processing characteristics.
In Coltheart, M., editor, *Attention and performance XII: The Psychology of Reading*, pages 601–618. Erlbaum, Hillsdale, NJ.

Nguyen, L., van Schijndel, M., and Schuler, W. (2012).
Accurate unbounded dependency recovery using generalized categorial grammars.
In *Proceedings of the 24th International Conference on Computational Linguistics (COLING '12)*, pages 2125–2140, Mumbai, India.

Pickering, M. J. and Traxler, M. J. (2003).
Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies.
*Language and Cognitive Processes*, 18(4):469–503.

# Bibliography III


Pickering, M. J., Traxler, M. J., and Crocker, M. W. (2000).
Ambiguity resolution in sentence processing: Evidence against frequency-based accounts.
*Journal of Memory and Language*, 43:447–475.

Staub, A. (2007).
The parser doesn't ignore intransitivity, after all.
*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3):550–569.

Staub, A., Clifton, C., and Frazier, L. (2006).
Heavy NP shift is the parser's last resort: Evidence from eye movements.
*Journal of Memory and Language*, 54:389–406.

# Bibliography IV

van Gompel, R. P. G. and Pickering, M. J. (2001).
Lexical guidance in sentence processing: A note on Adams, Clifton, and Mitchell (1998).
*Psychonomic Bulletin and Review*, 8:851–857.