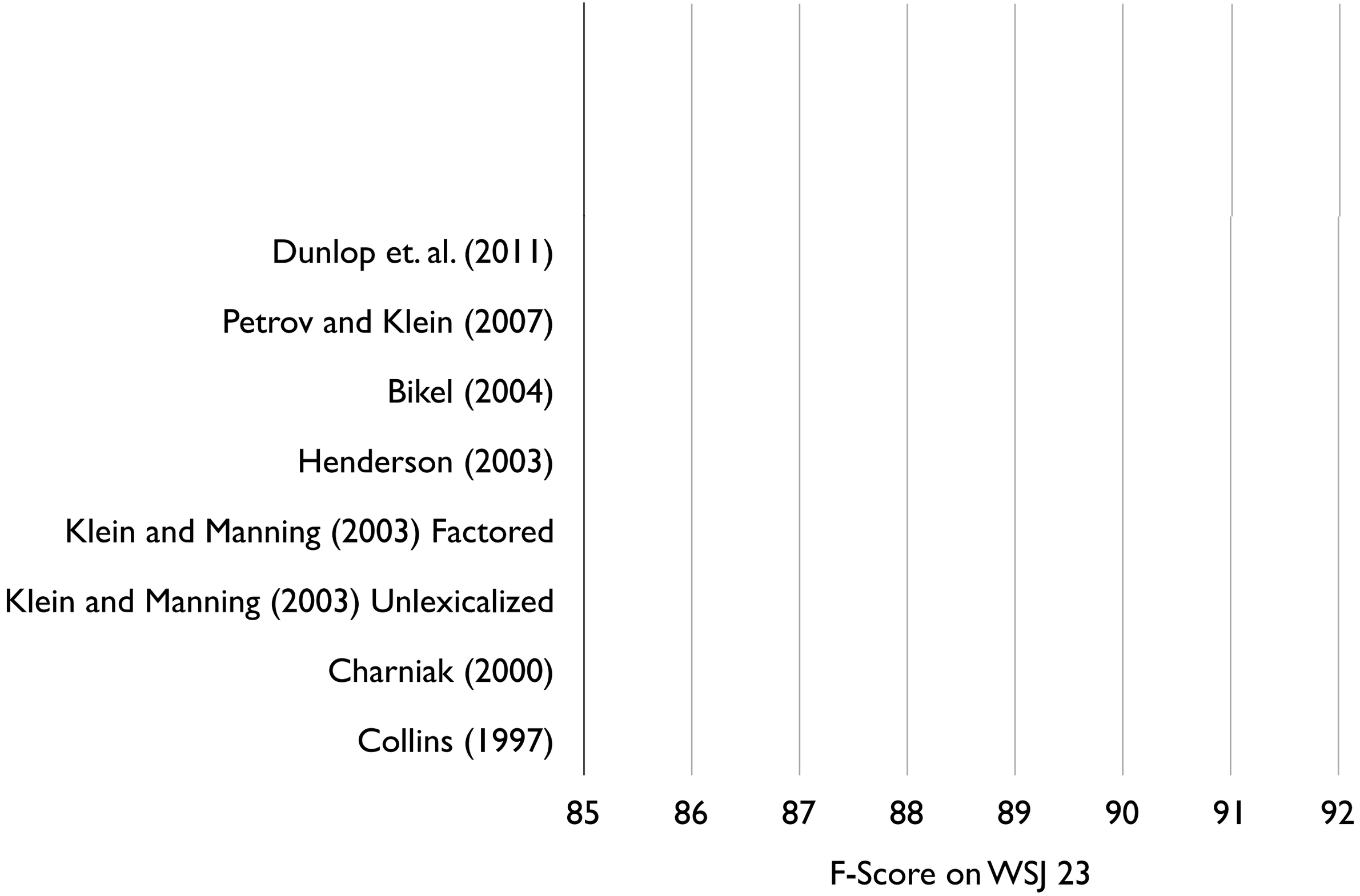
Parser Showdown
at the
Wall Street Corral

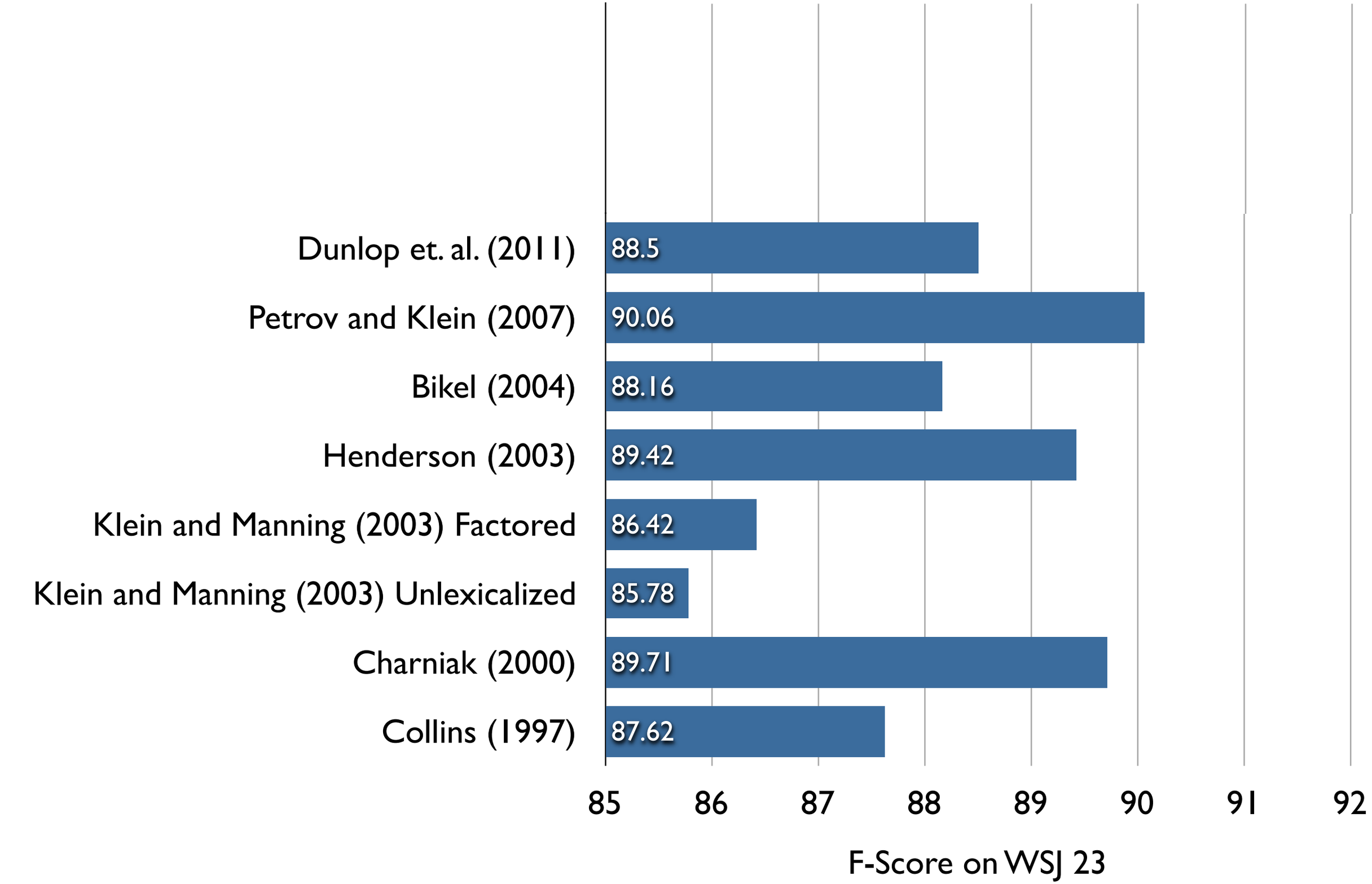
Jonathan K. Kummerfeld†, David Hall†, James R. Curran‡,
and Dan Klein†

UC Berkeley† and U Sydney‡

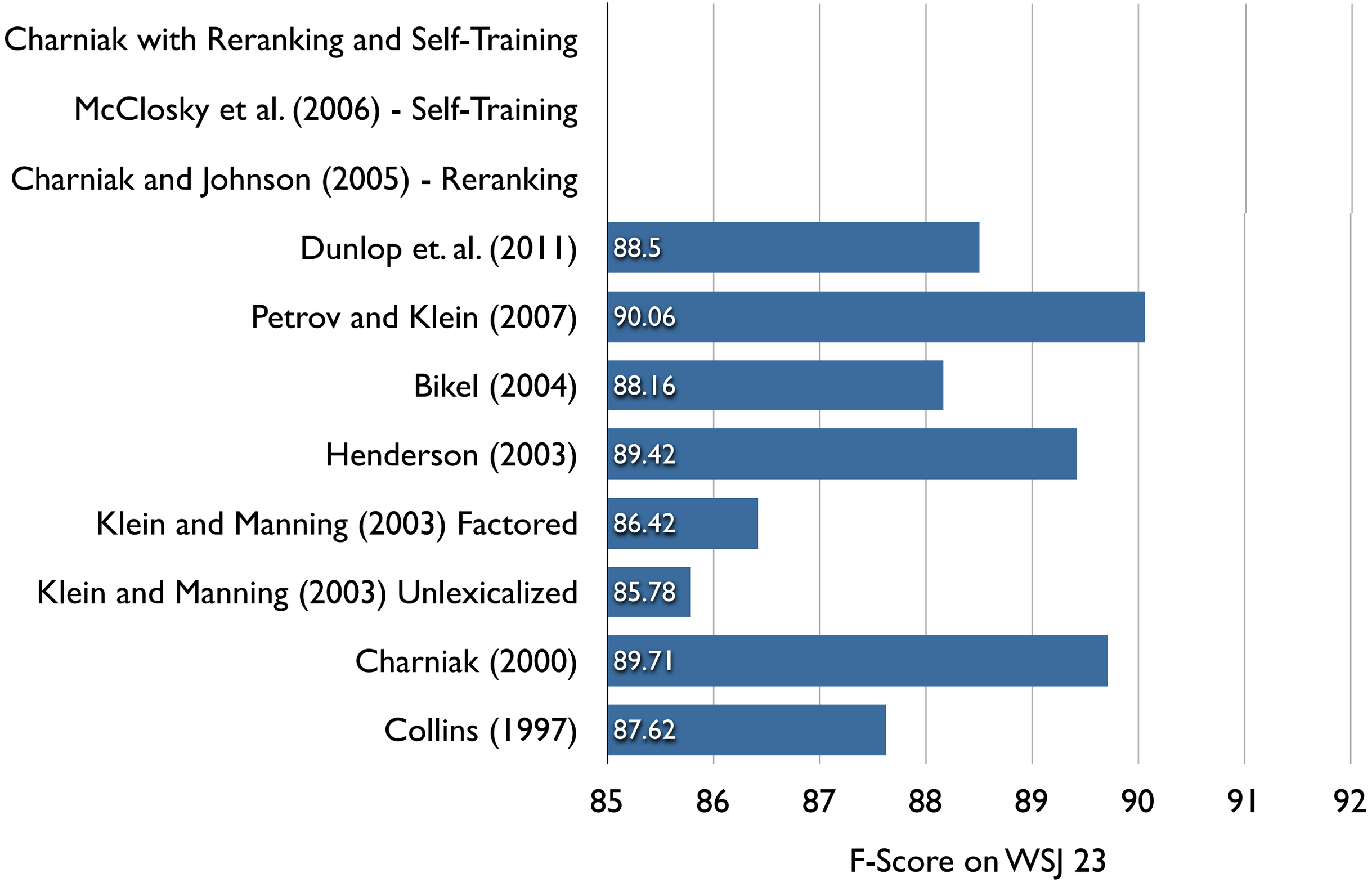
Parser F-Score



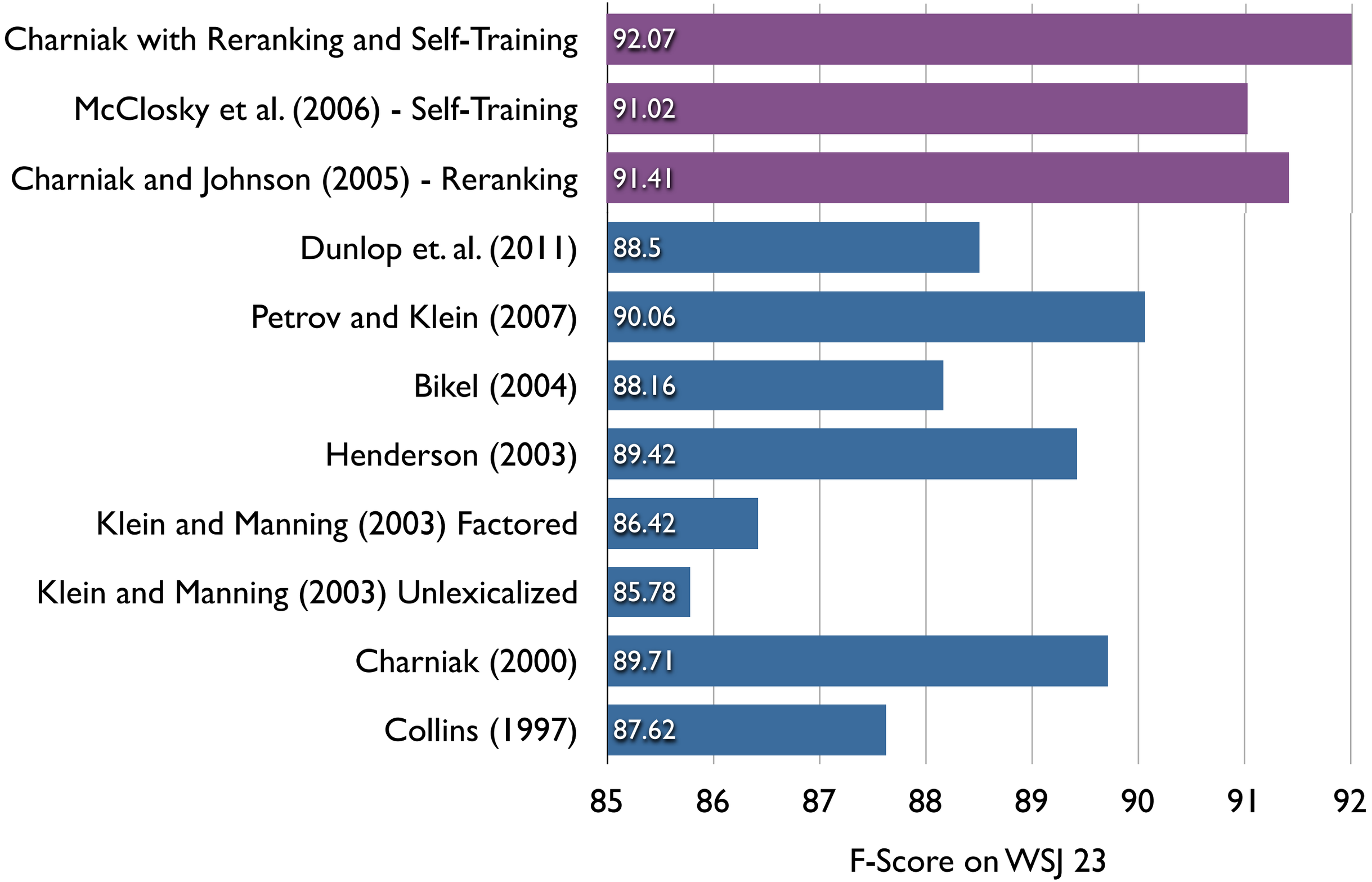
Parser F-Score



Parser F-Score



Parser F-Score

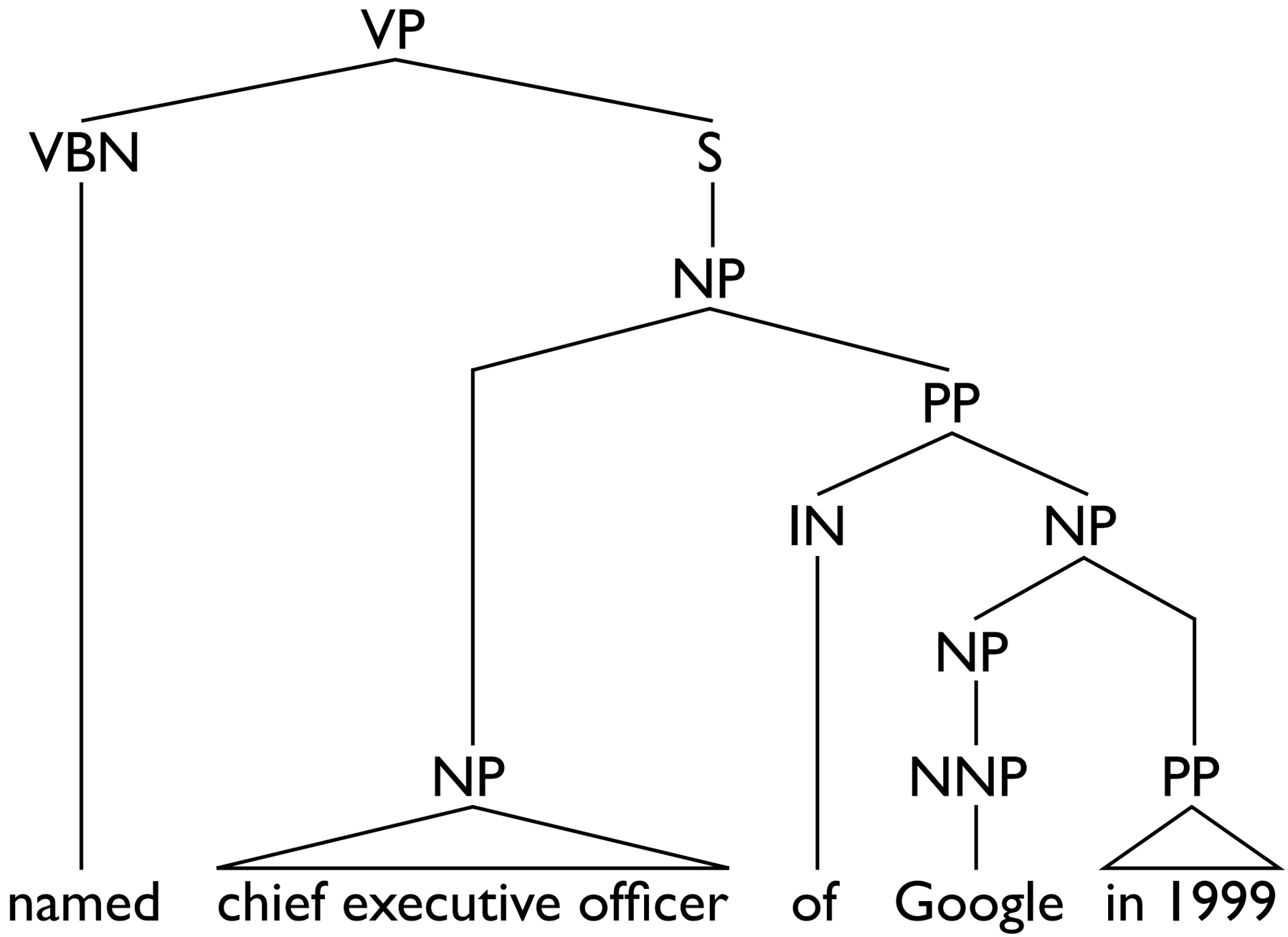


Surface vs Underlying Errors

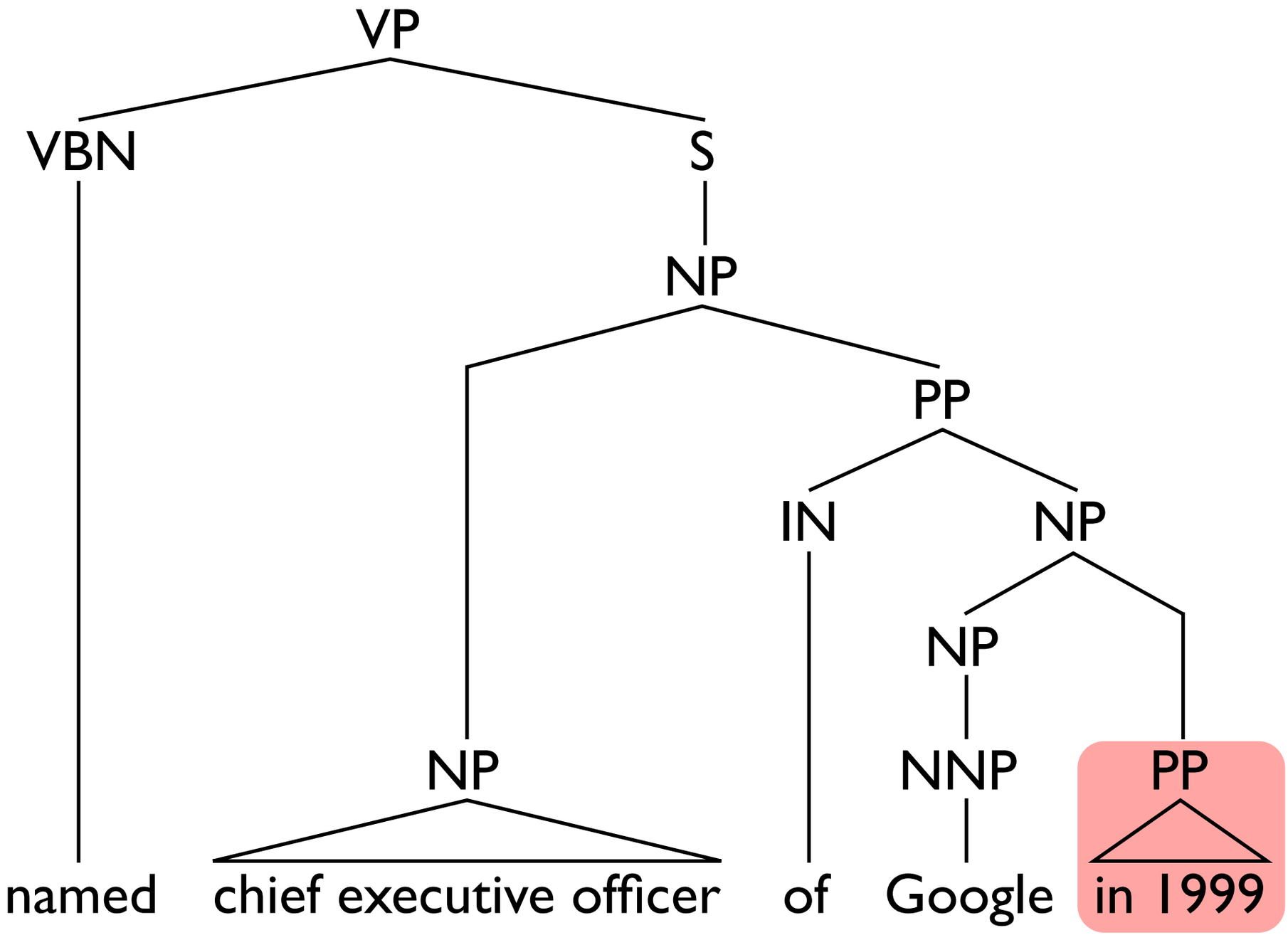
Surface vs Underlying Errors

named chief executive officer of Google in 1999

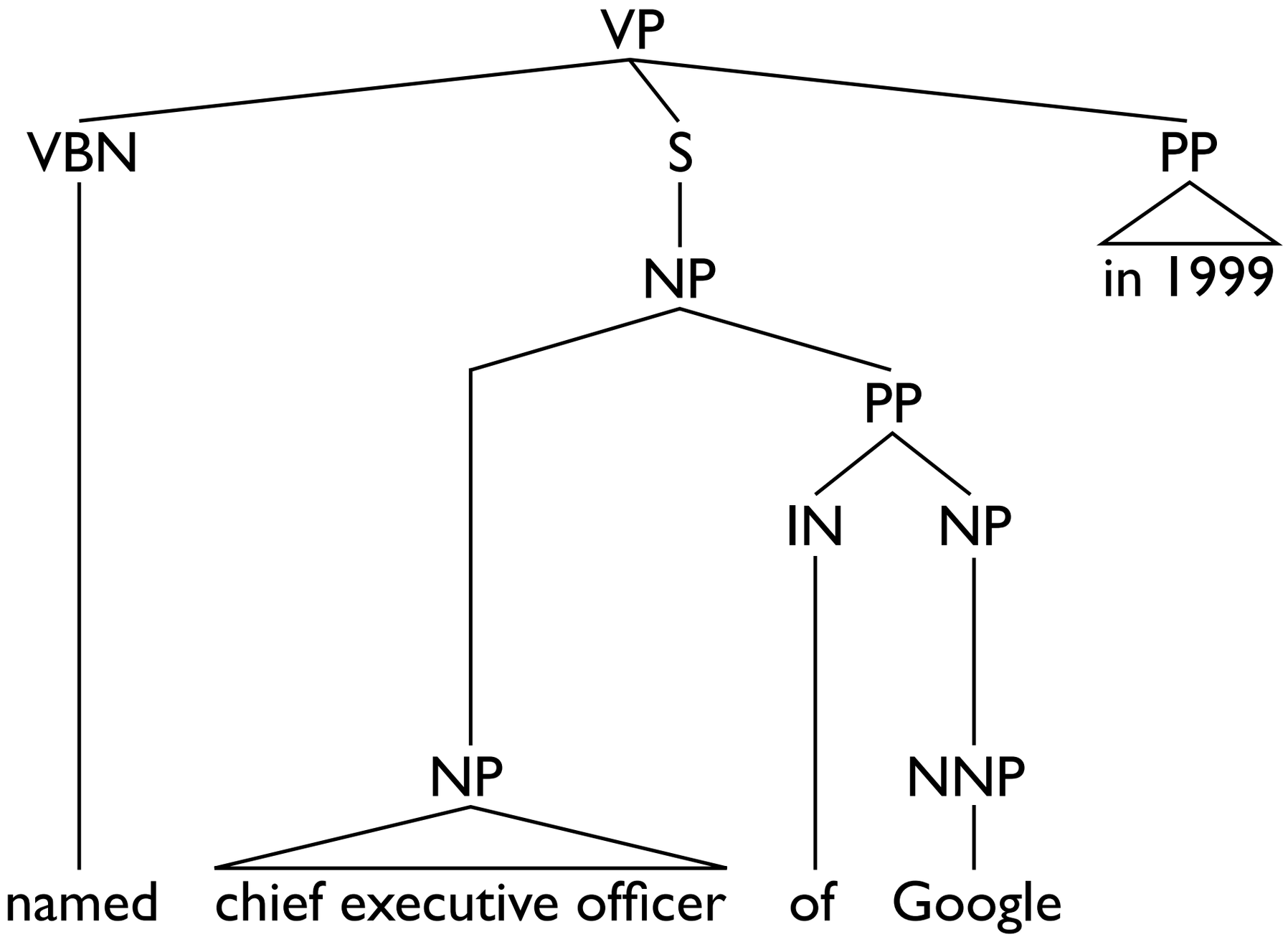
Surface vs Underlying Errors



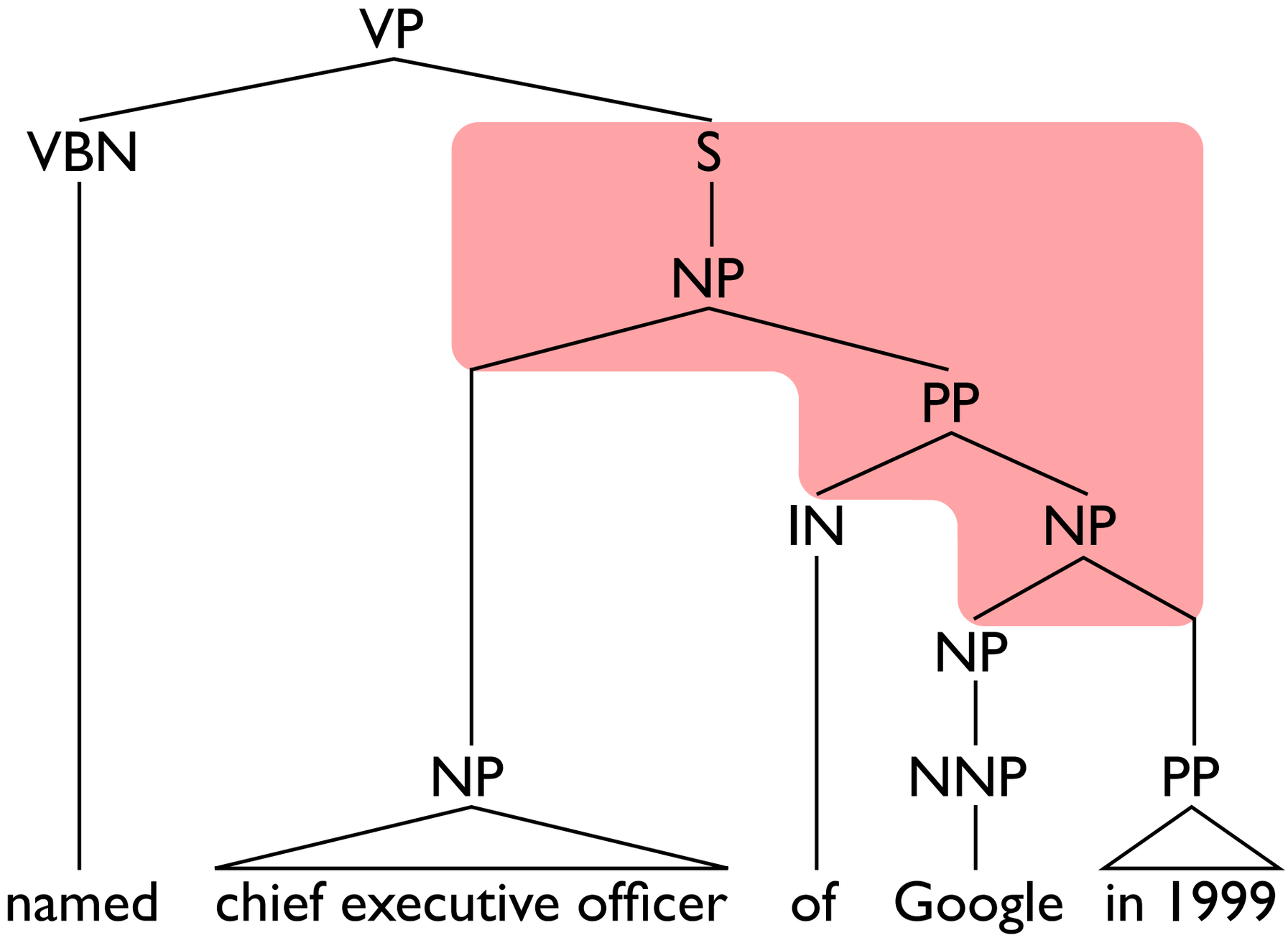
Surface vs Underlying Errors



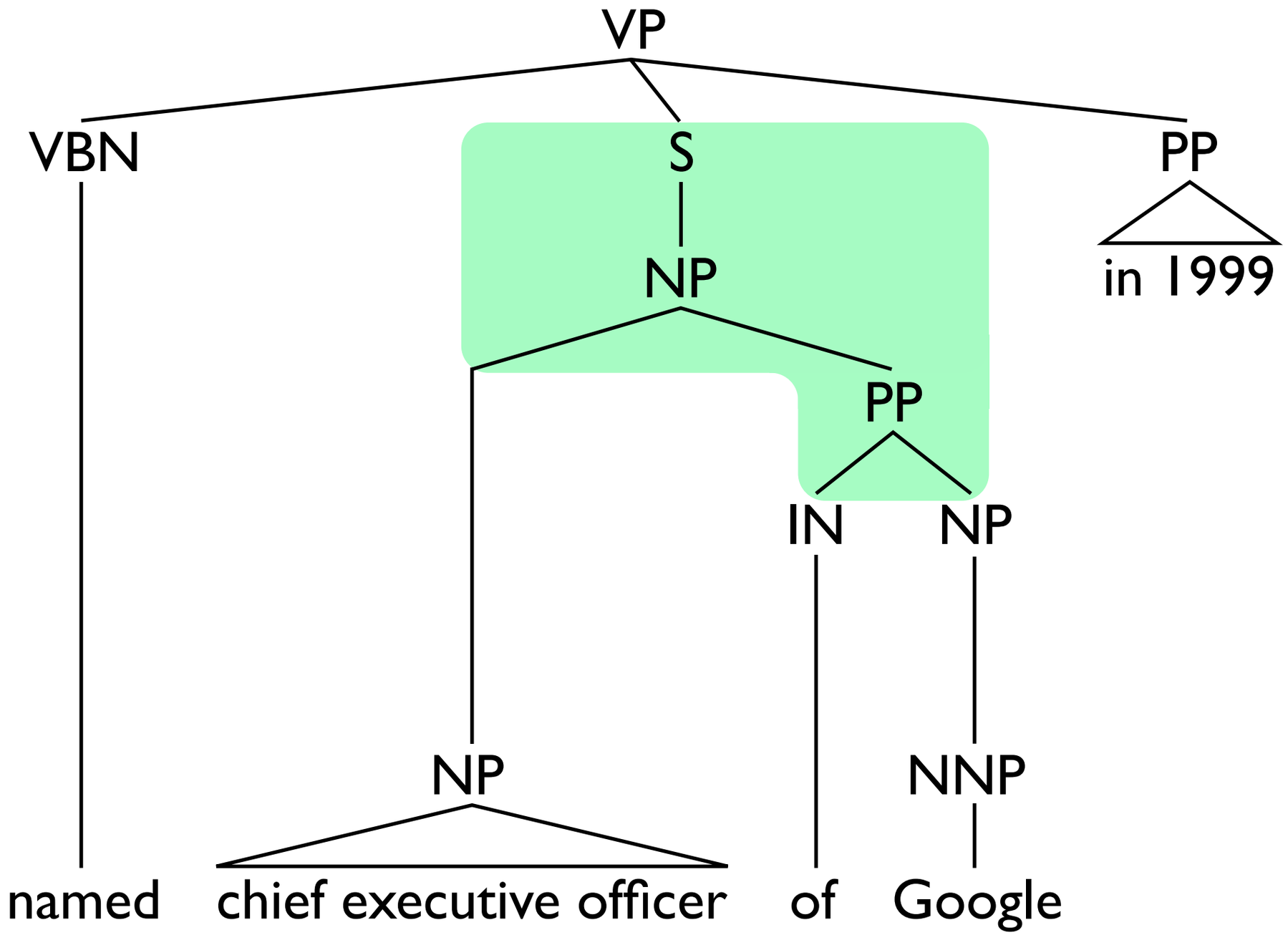
Surface vs Underlying Errors



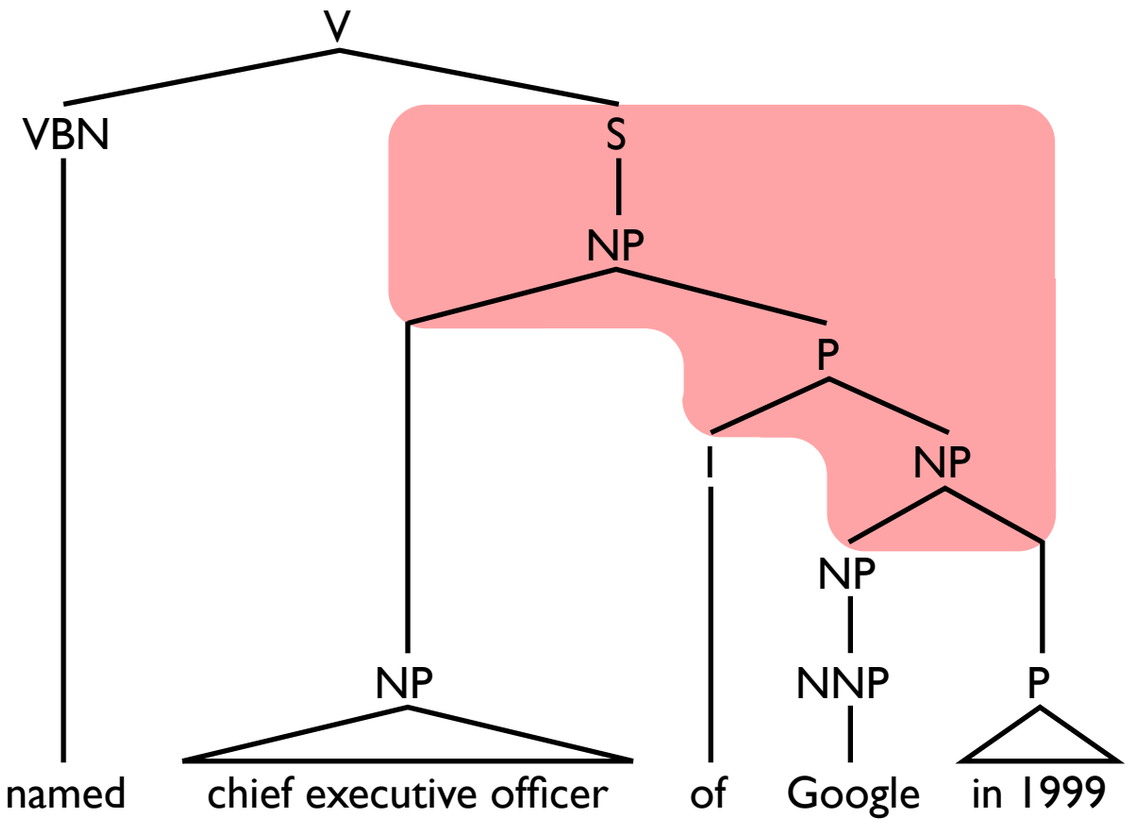
Surface vs Underlying Errors

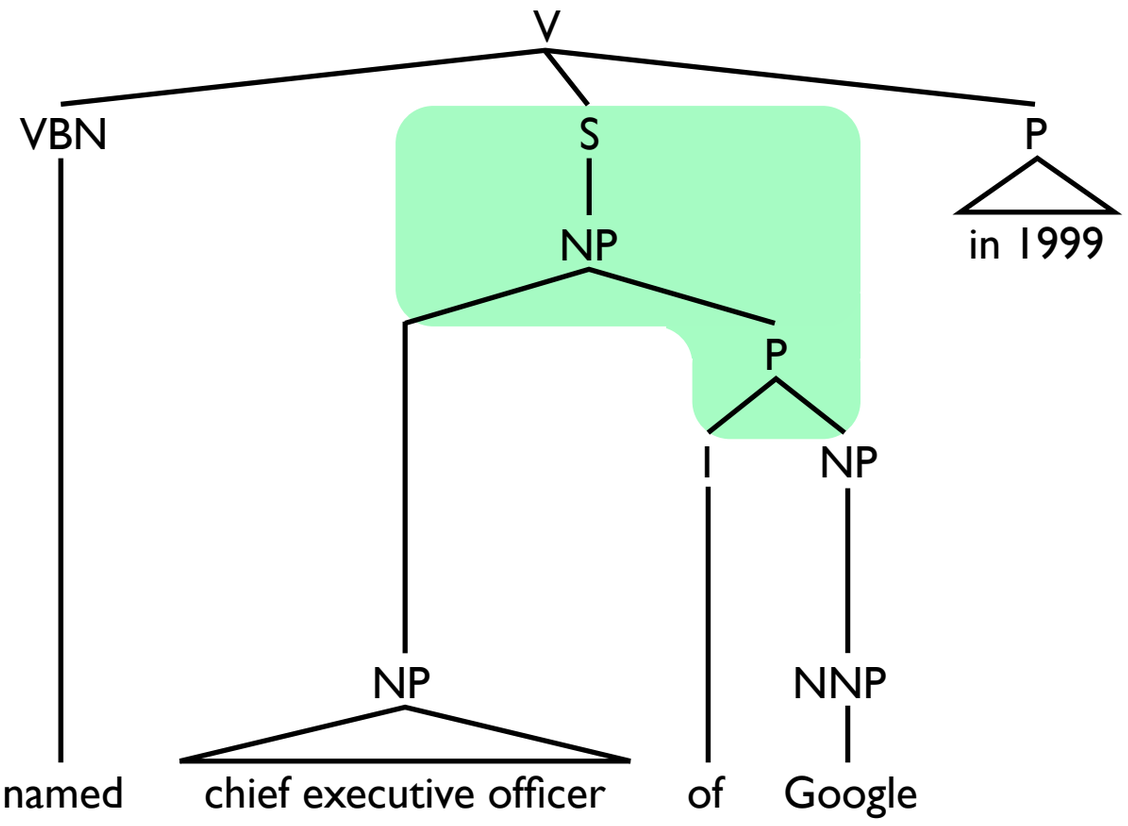


Surface vs Underlying Errors



Surface vs Underlying Errors

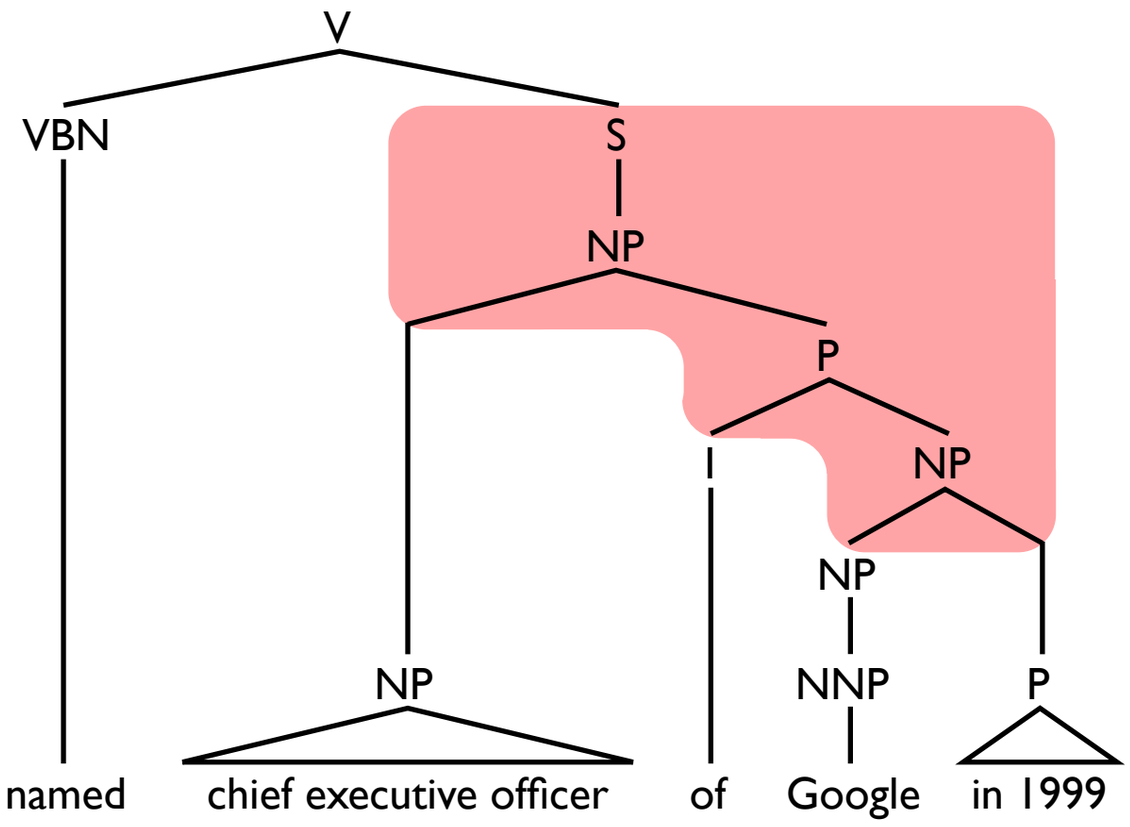


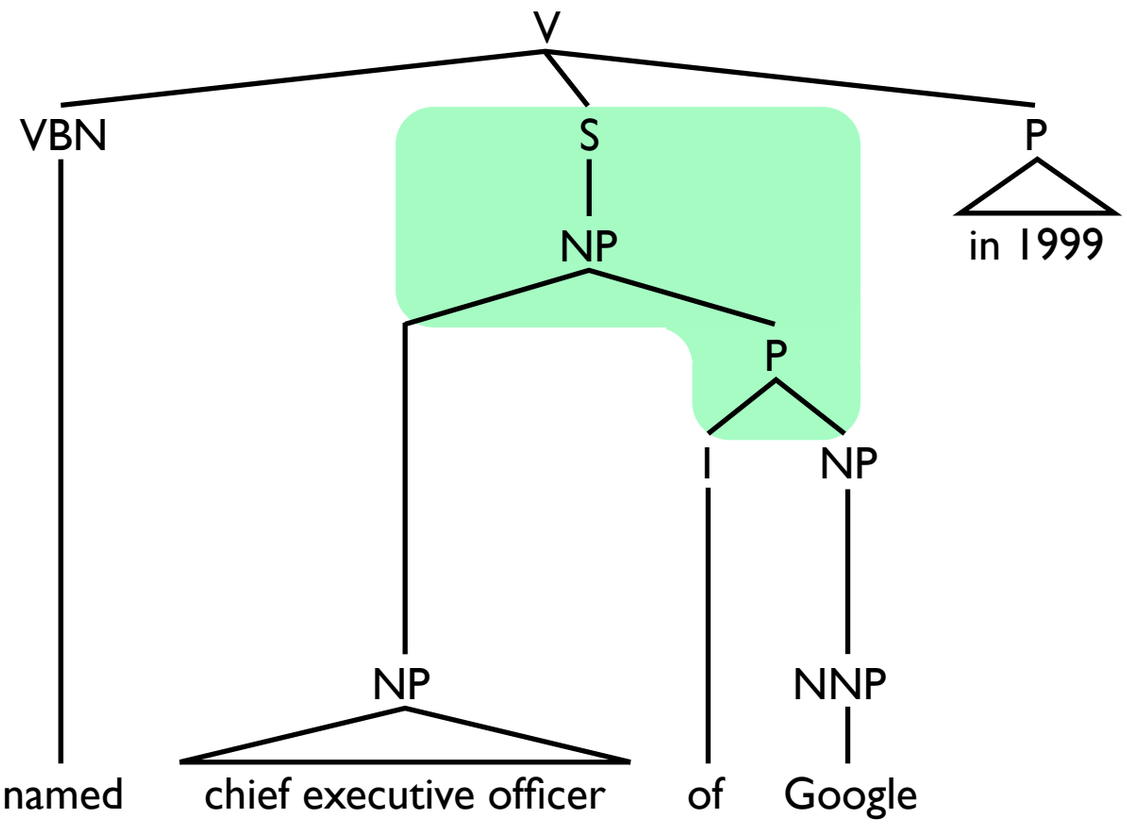


Surface vs Underlying Errors

Surface Errors: Extra S, NP, PP, NP

Missing S, NP, PP



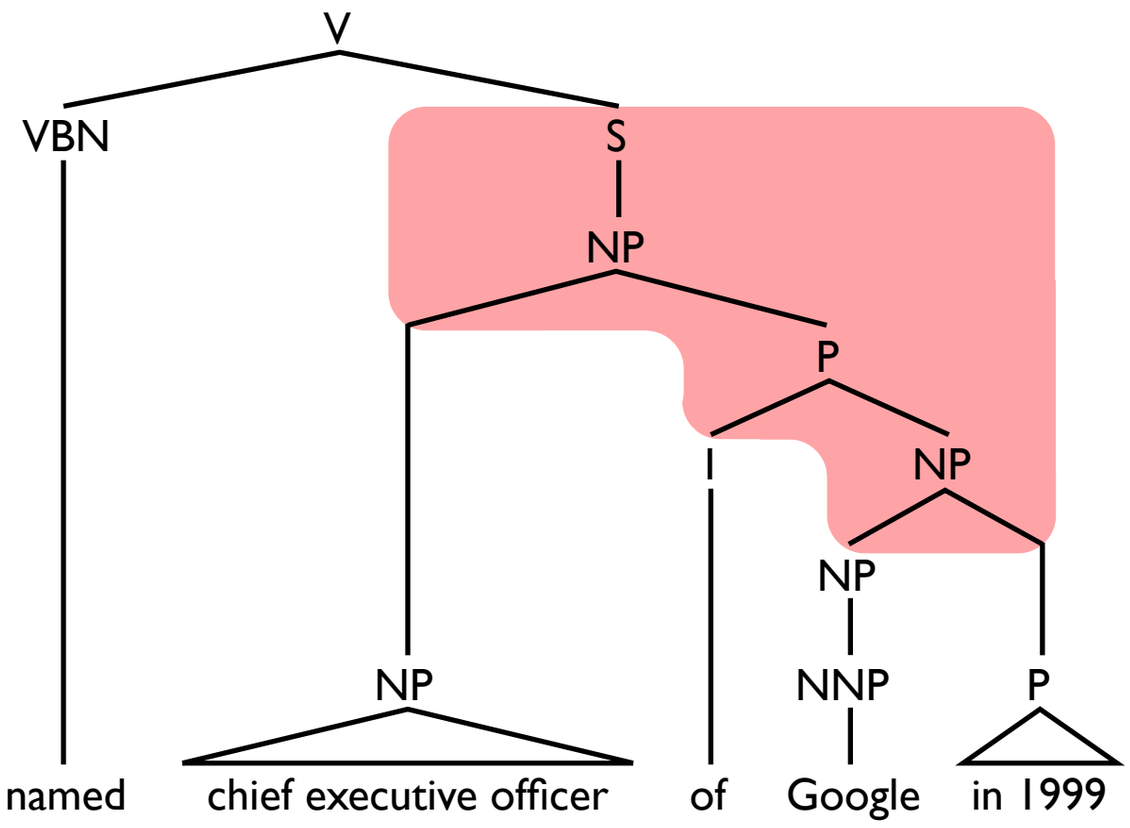


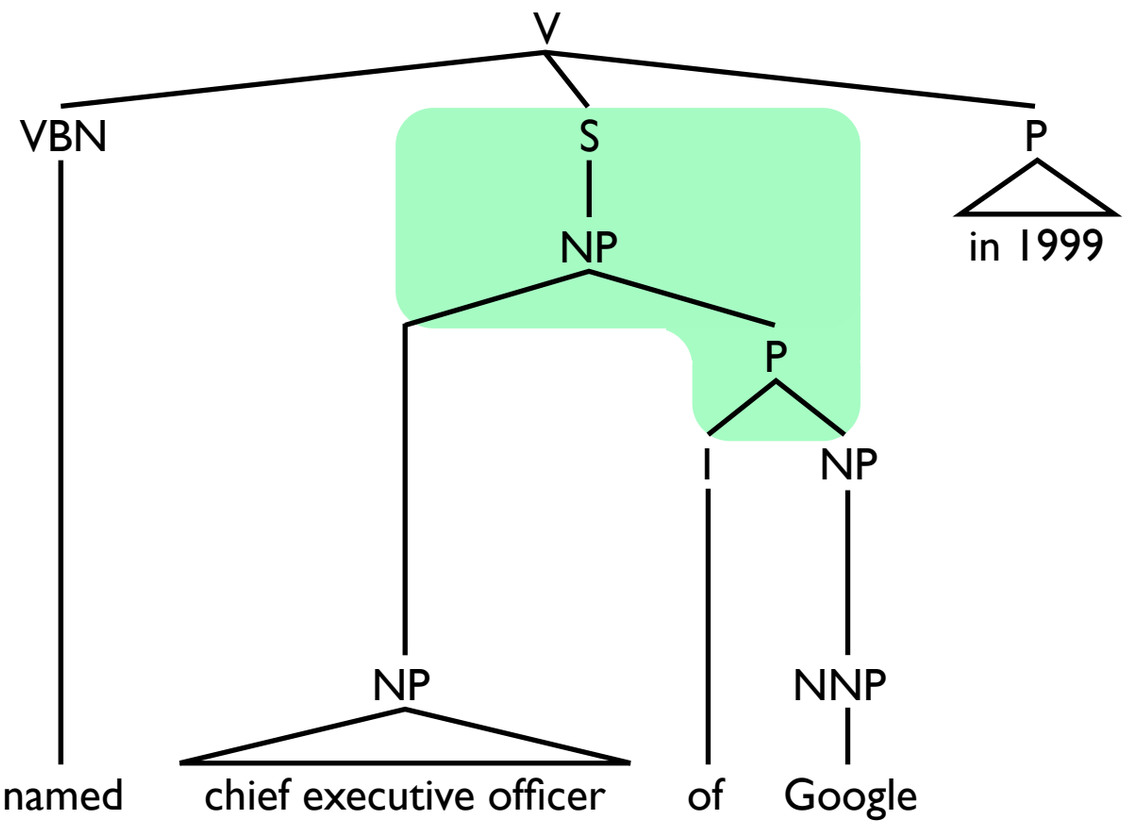
Surface vs Underlying Errors

Surface Errors: Extra S, NP, PP, NP

Missing S, NP, PP

Underlying Error: PP Attachment

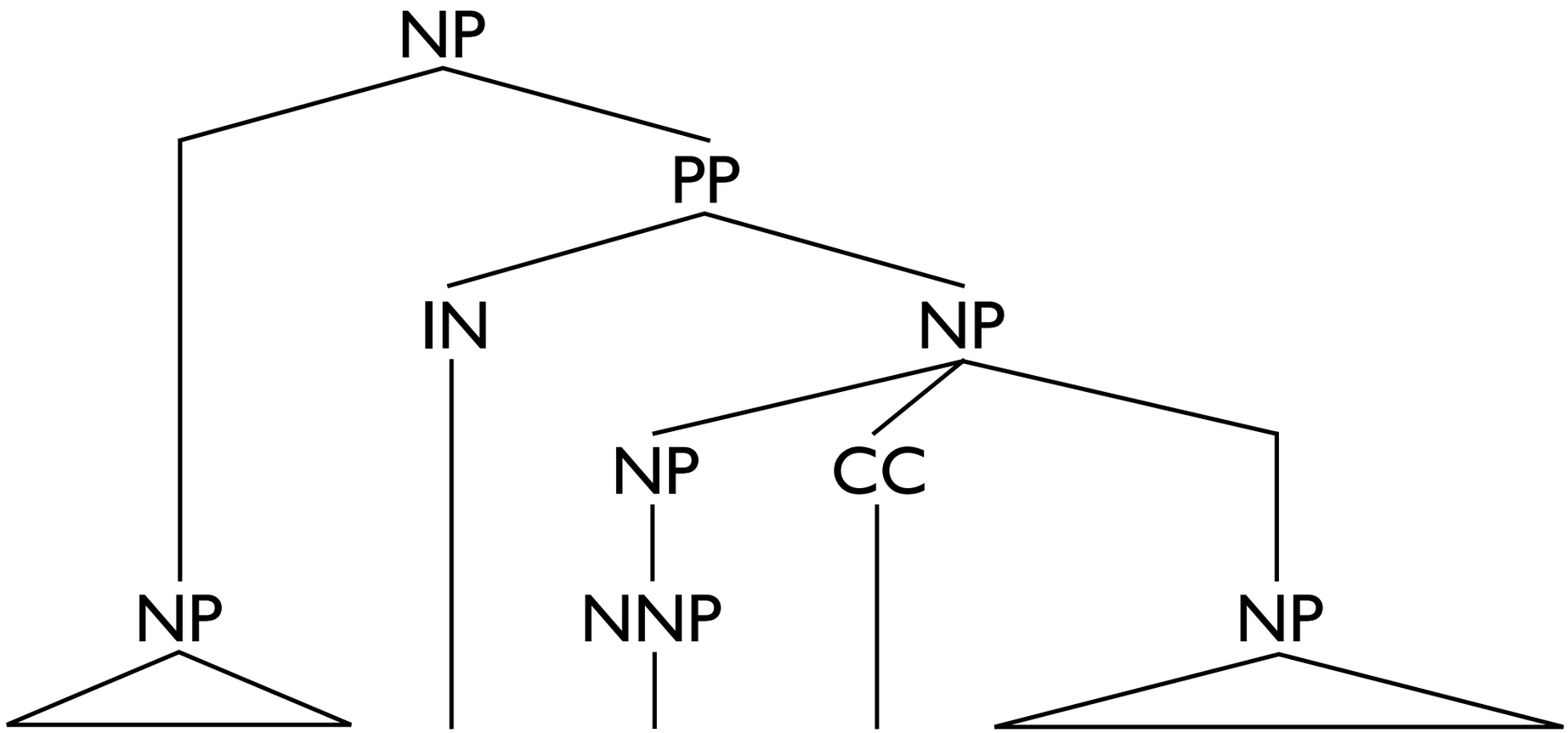




Surface vs Underlying Errors

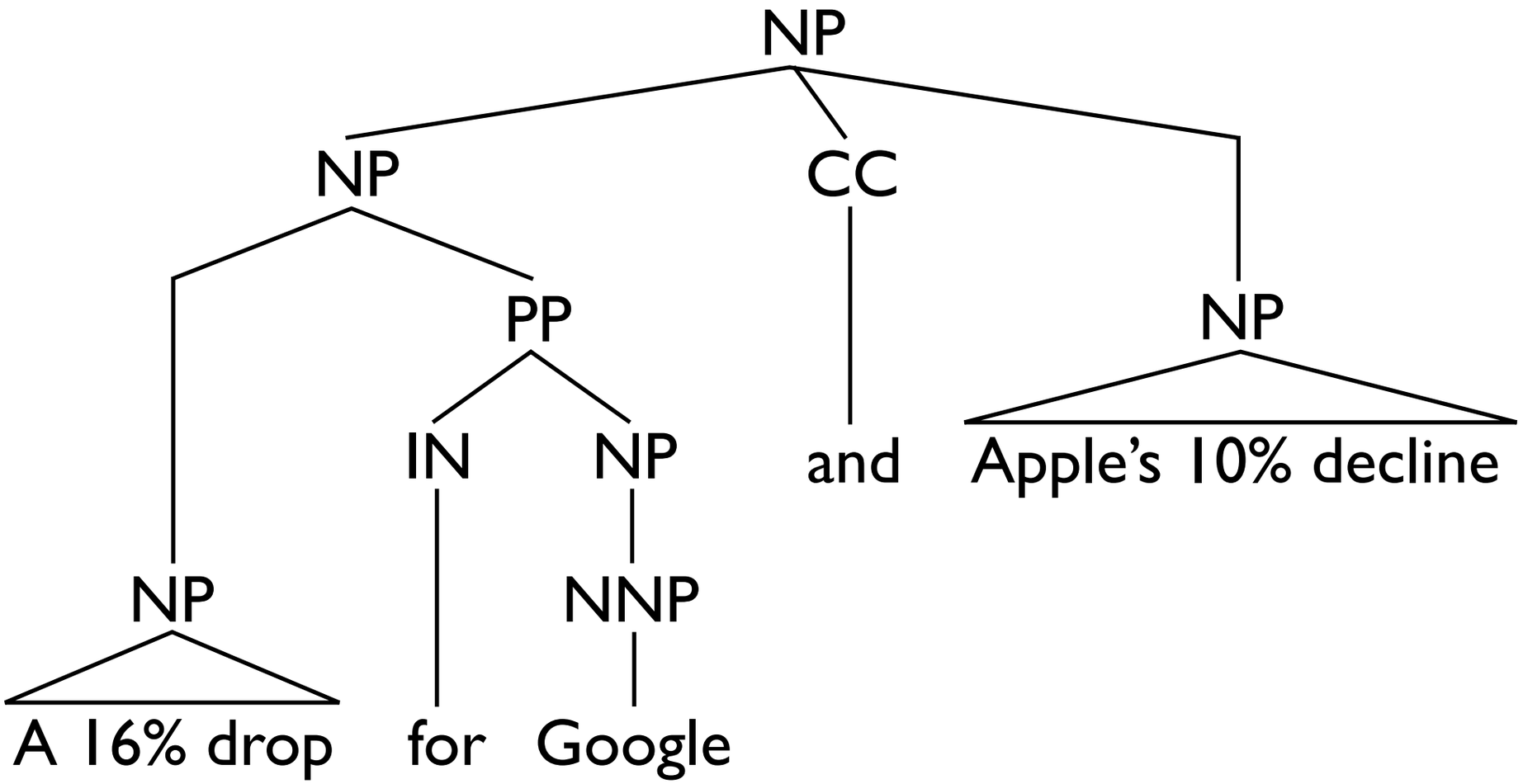
A 16% drop for Google and Apple's 10% decline

Surface vs Underlying Errors

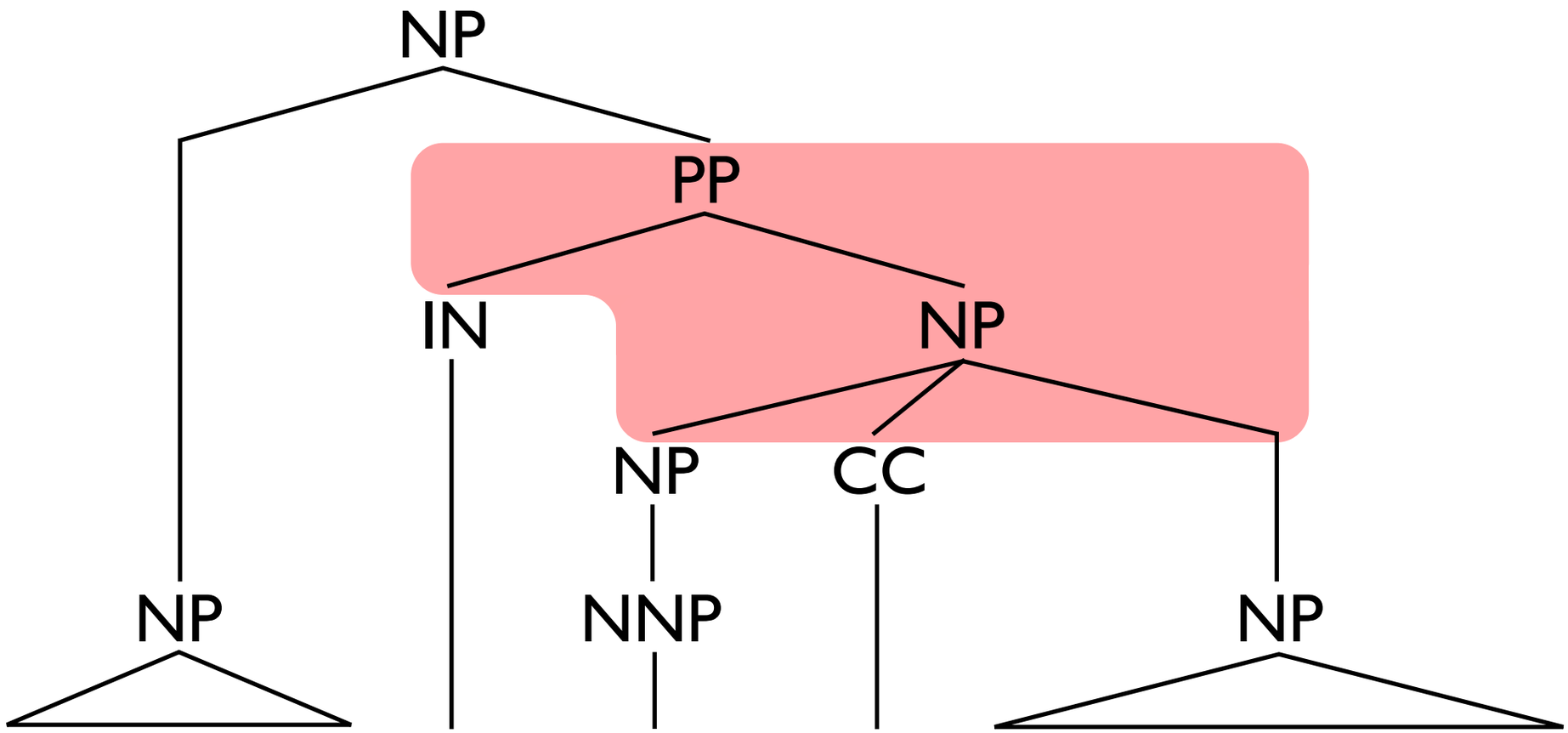


A 16% drop for Google and Apple's 10% decline

Surface vs Underlying Errors

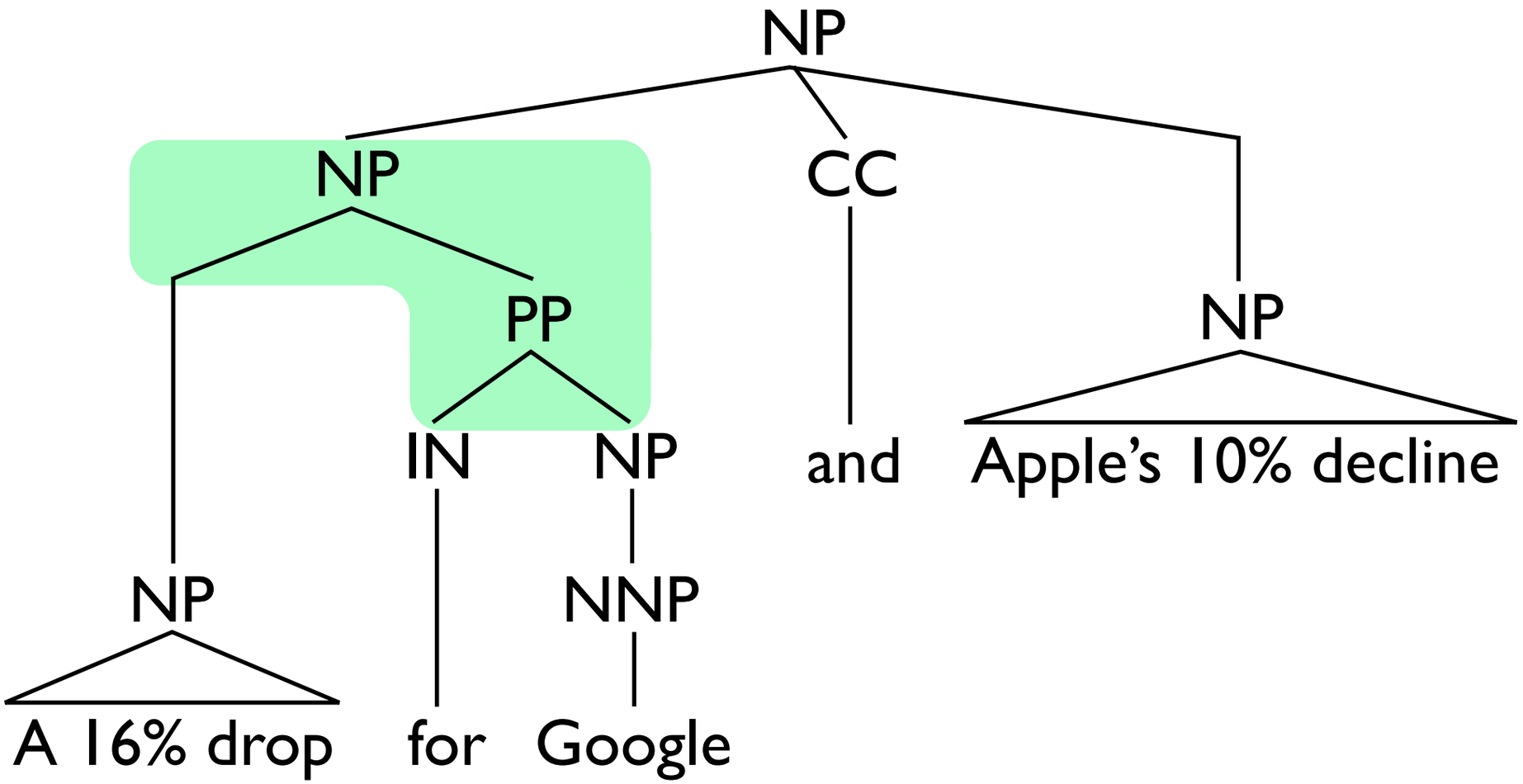


Surface vs Underlying Errors

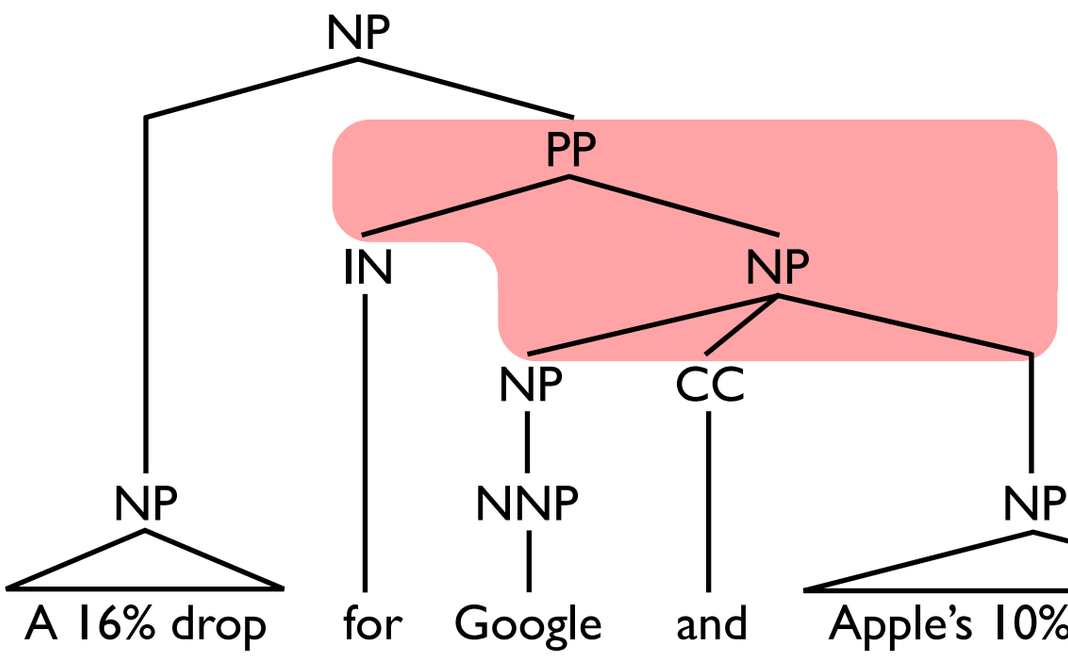


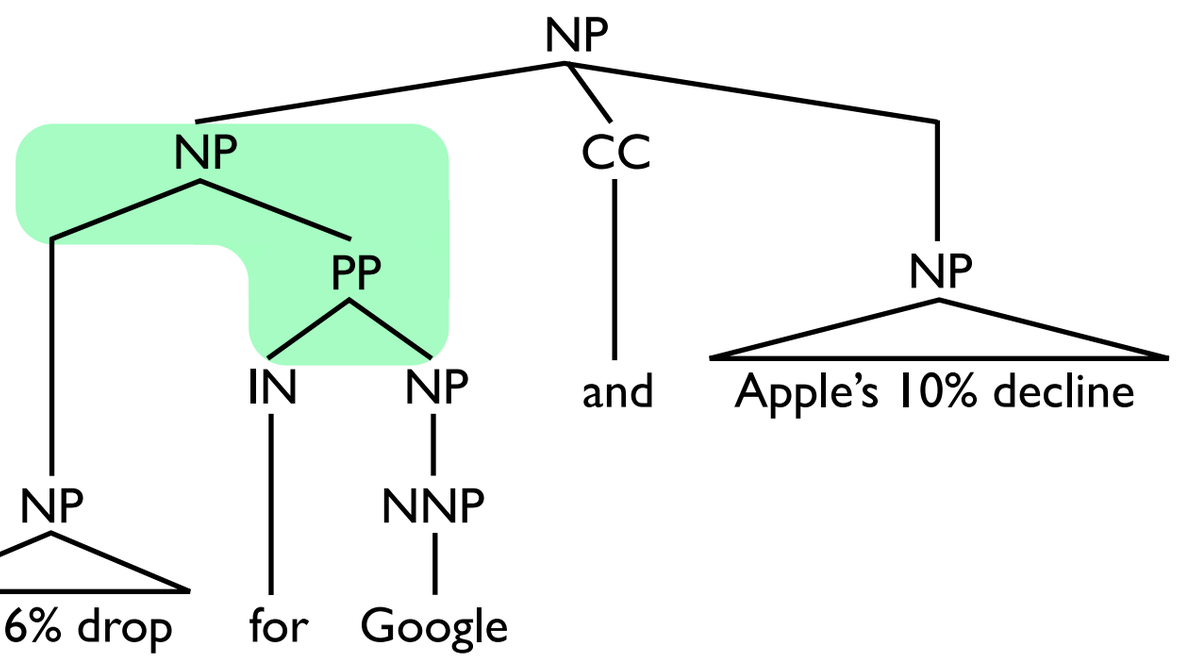
A 16% drop for Google and Apple's 10% decline

Surface vs Underlying Errors



Surface vs Underlying Errors

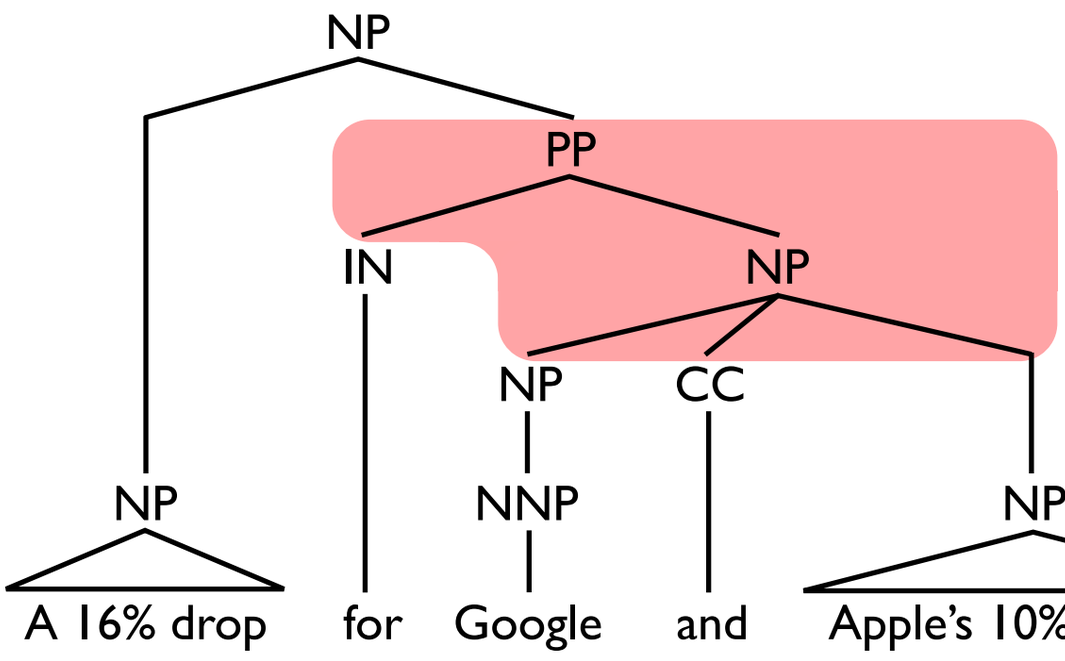


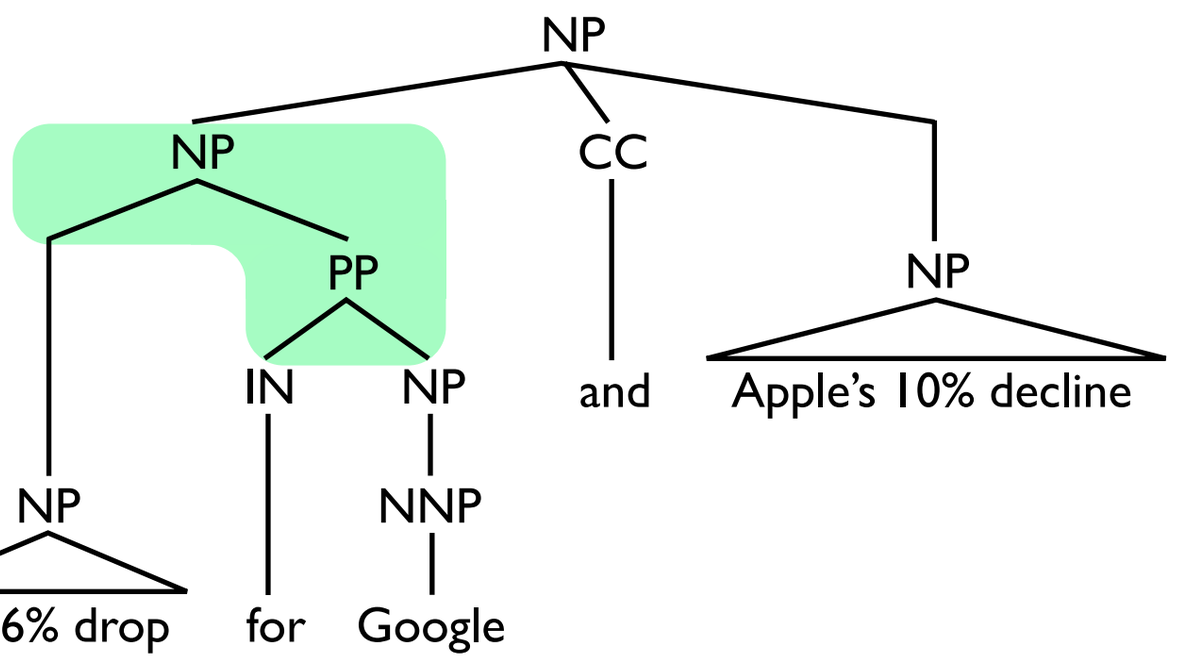


Surface vs Underlying Errors

Surface Errors: Extra PP, NP

Missing NP, PP



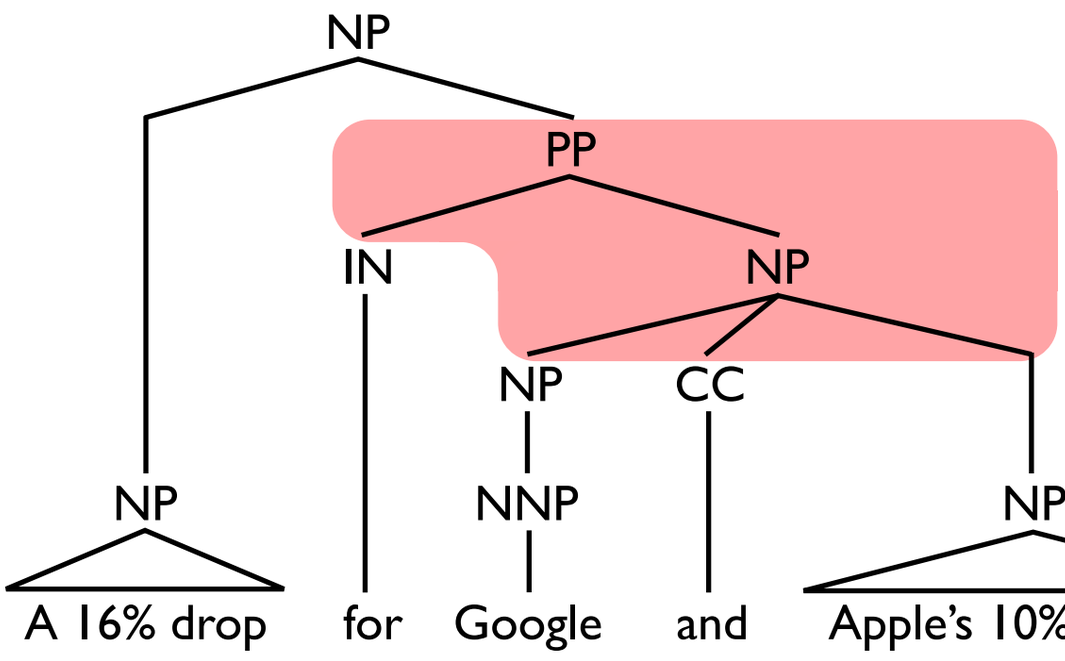


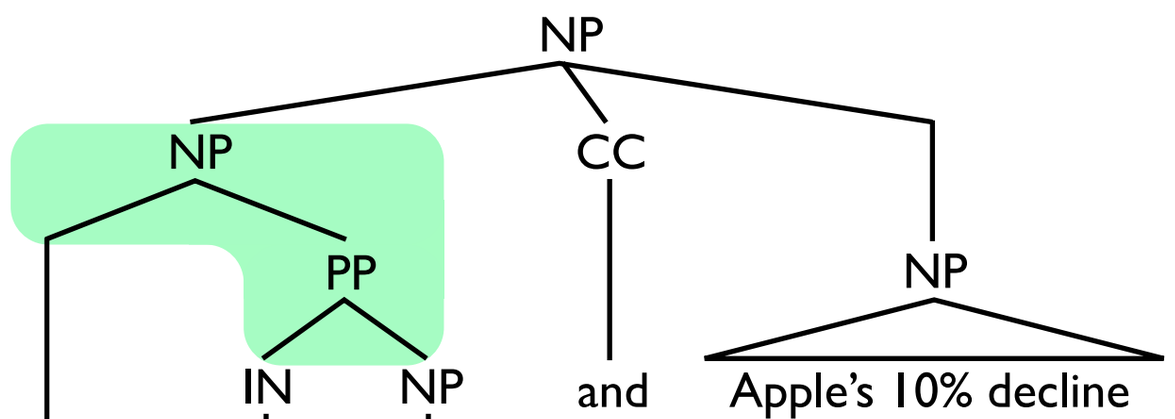
Surface vs Underlying Errors

Surface Errors: Extra PP, NP

Missing NP, PP

Underlying Error: Coordination

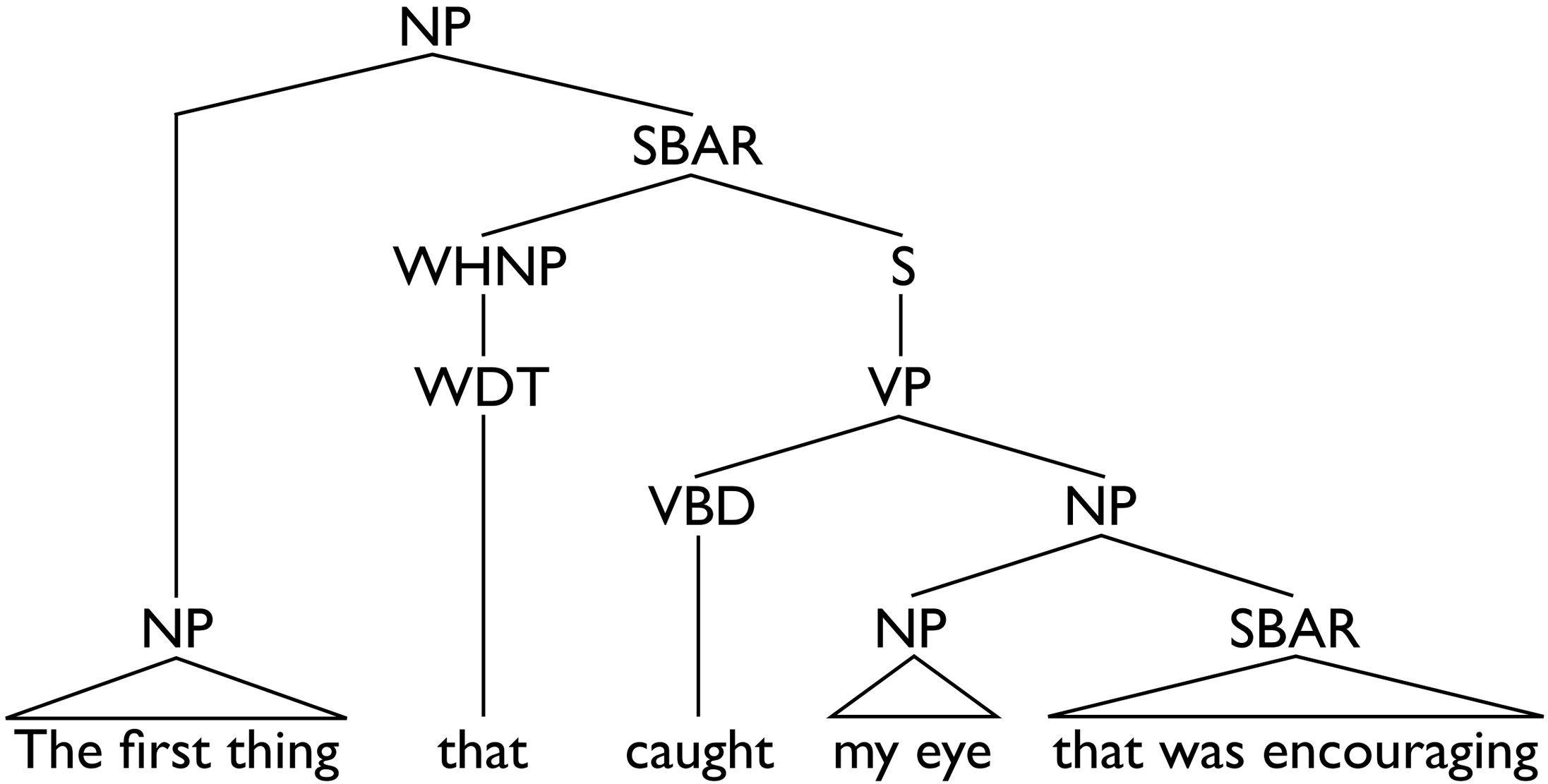




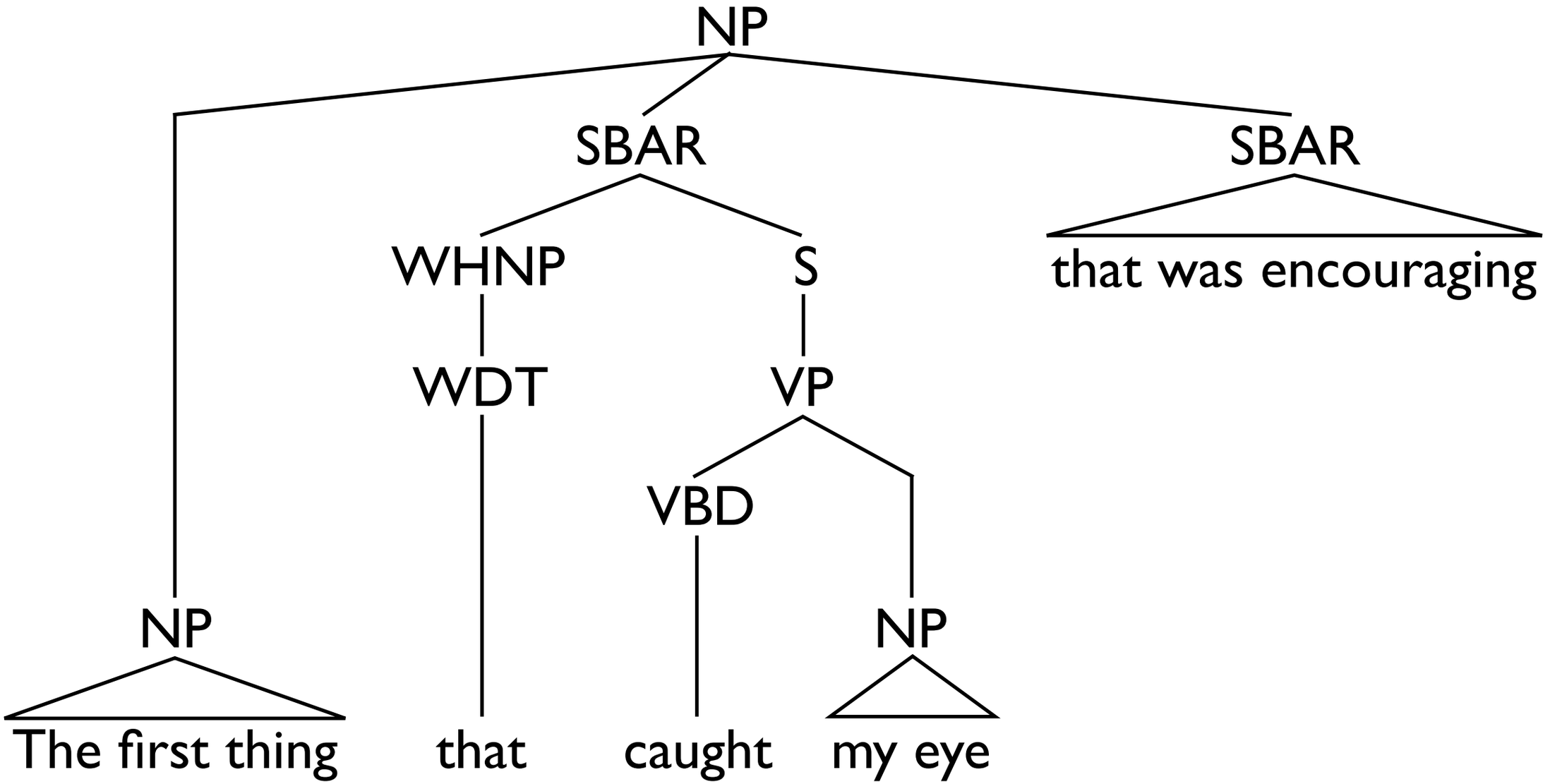
Surface vs Underlying Errors

The first thing that caught my eye that was encouraging

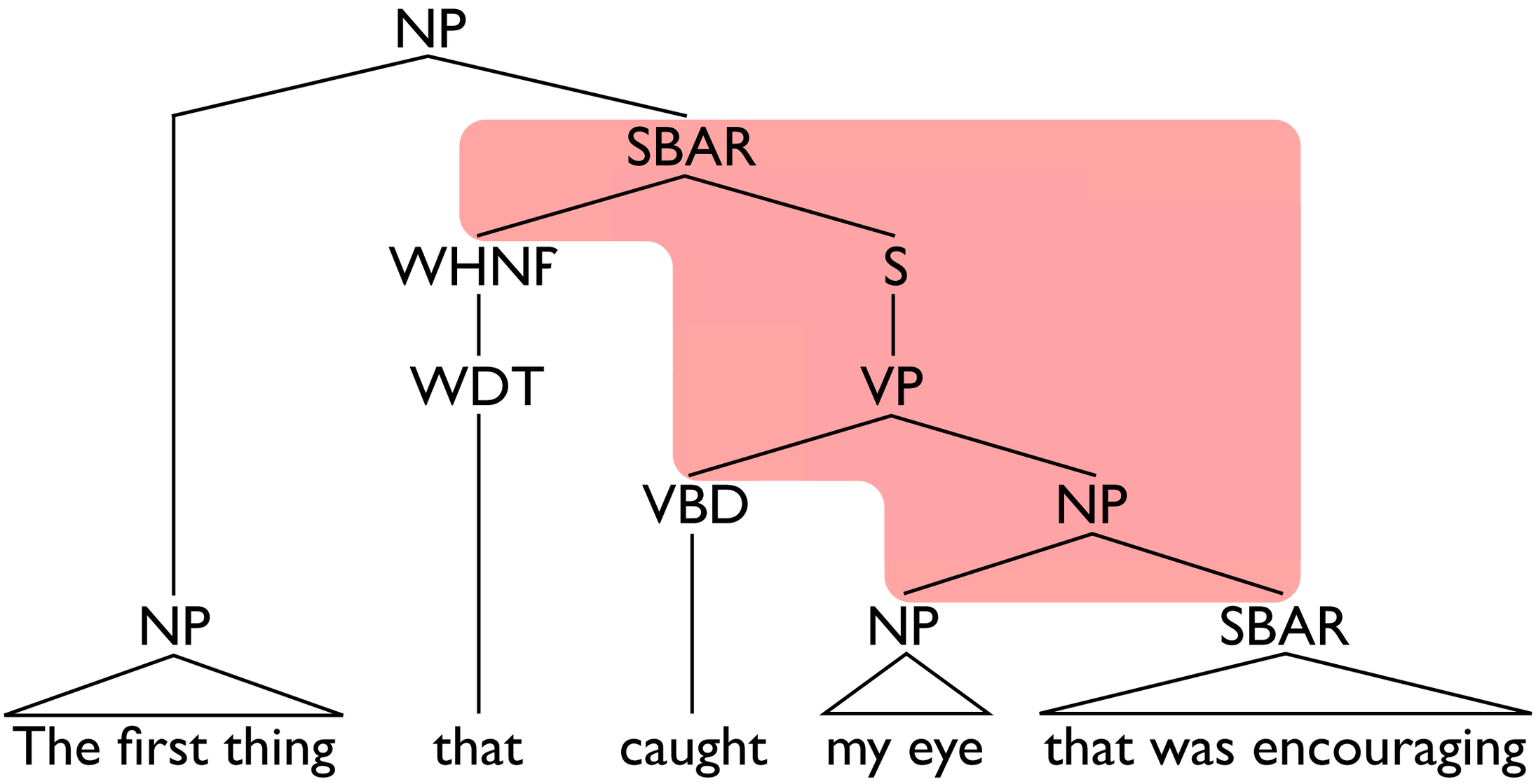
Surface vs Underlying Errors



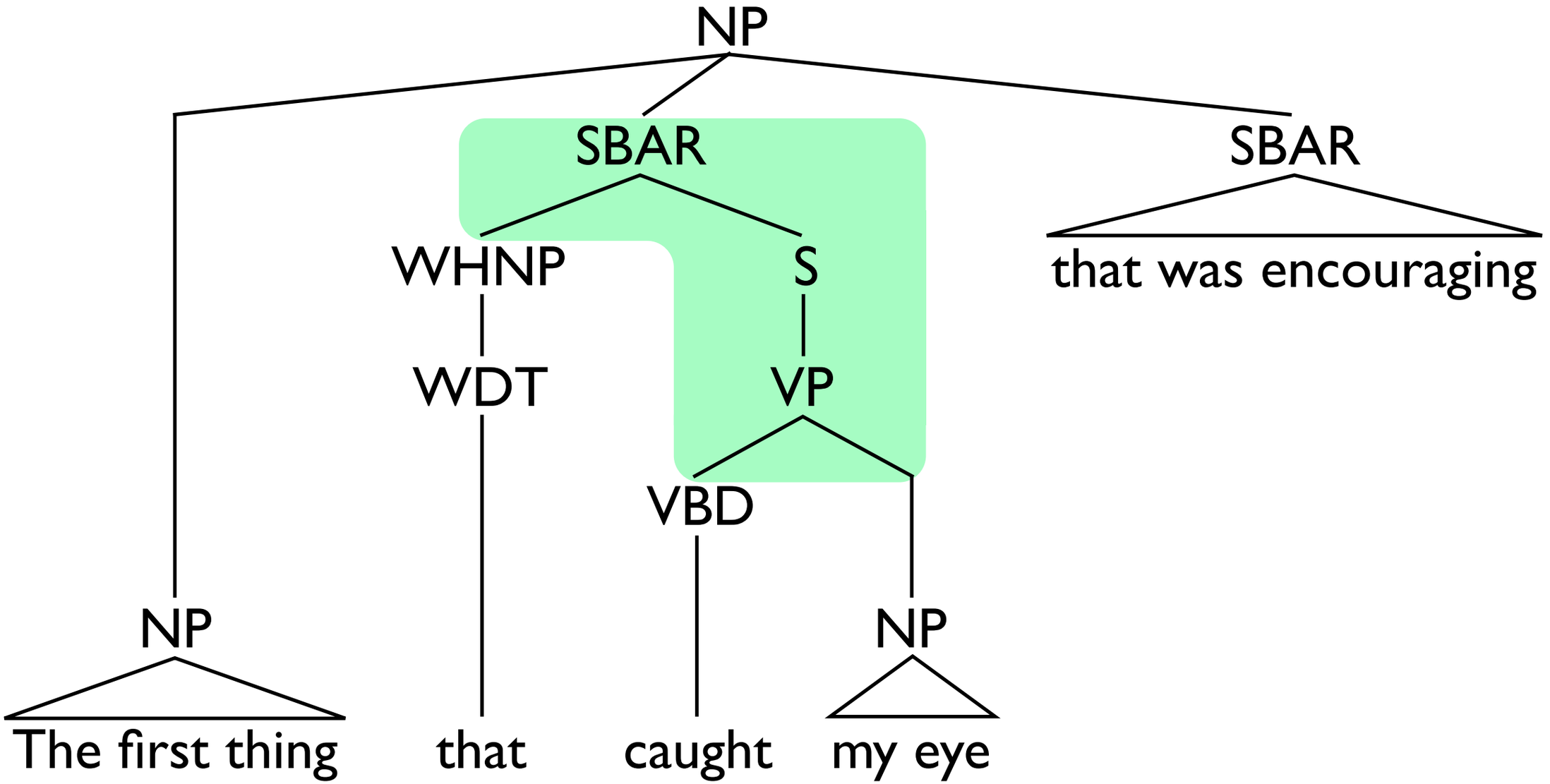
Surface vs Underlying Errors



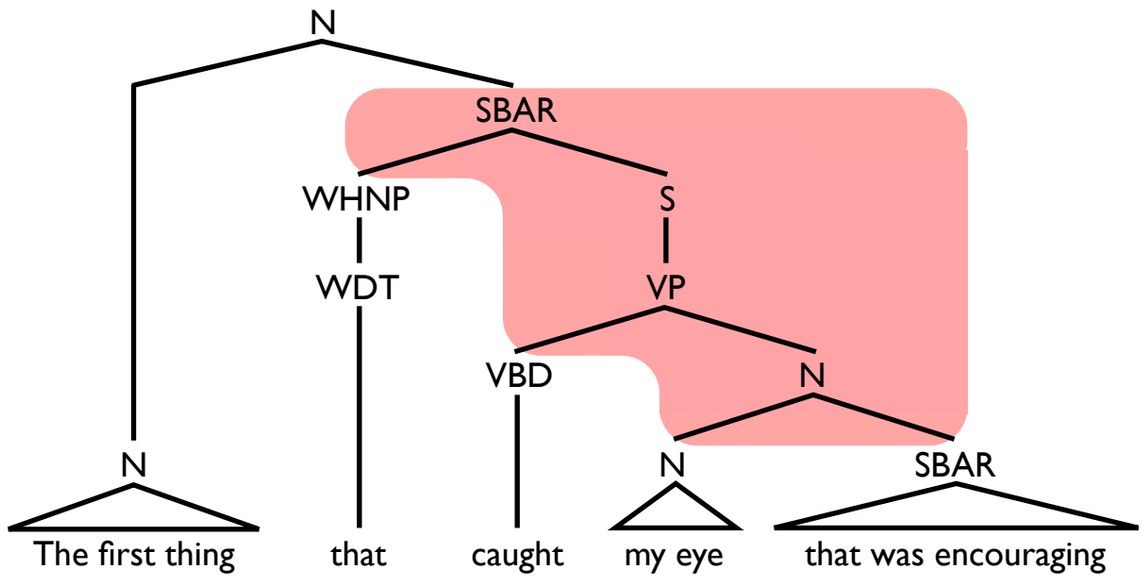
Surface vs Underlying Errors

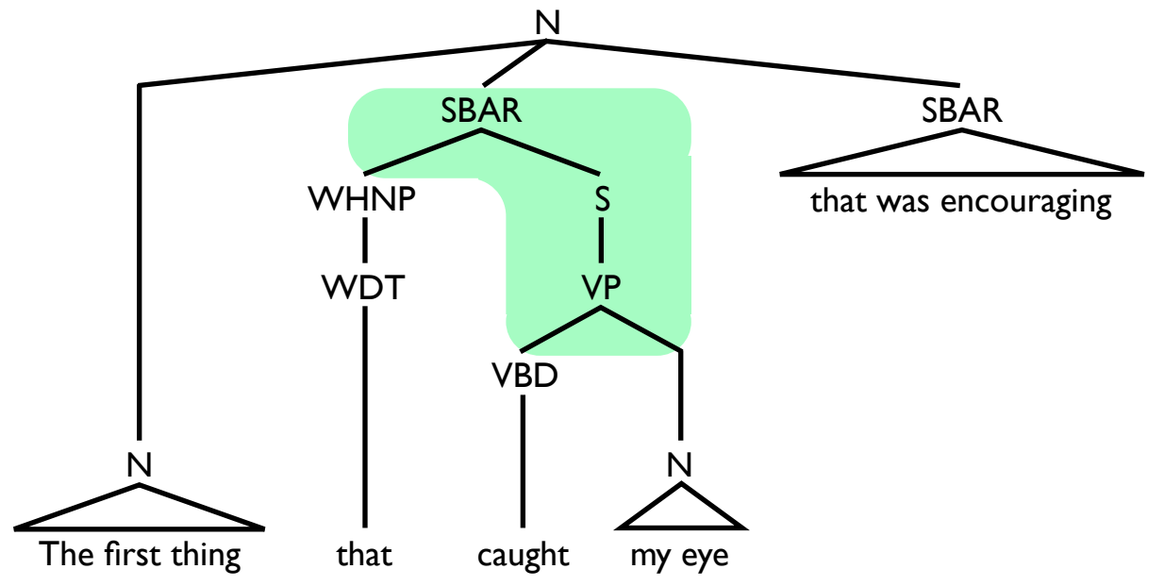


Surface vs Underlying Errors



Surface vs Underlying Errors

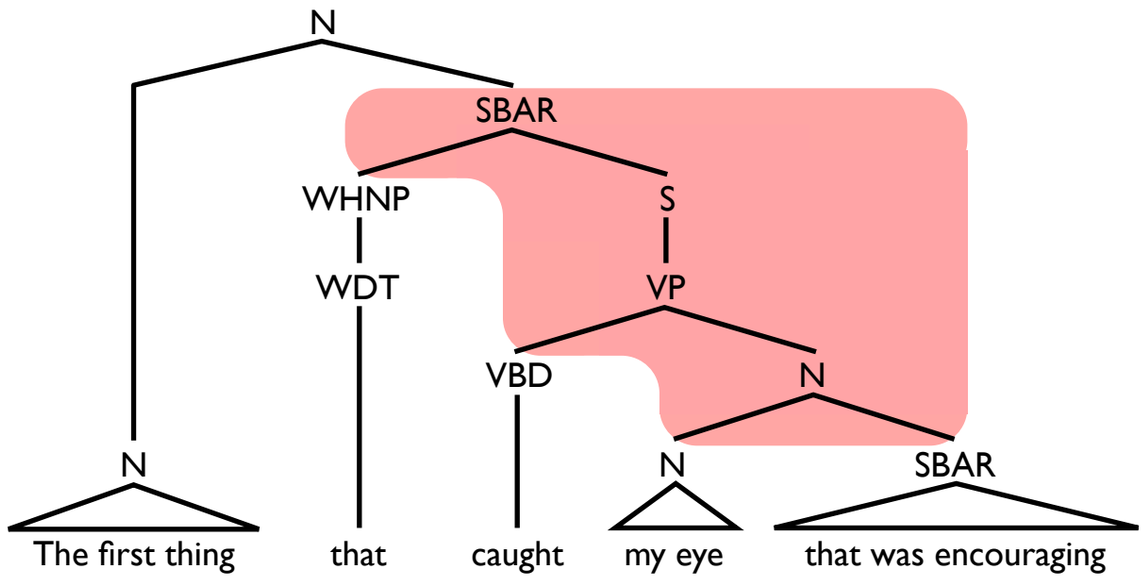


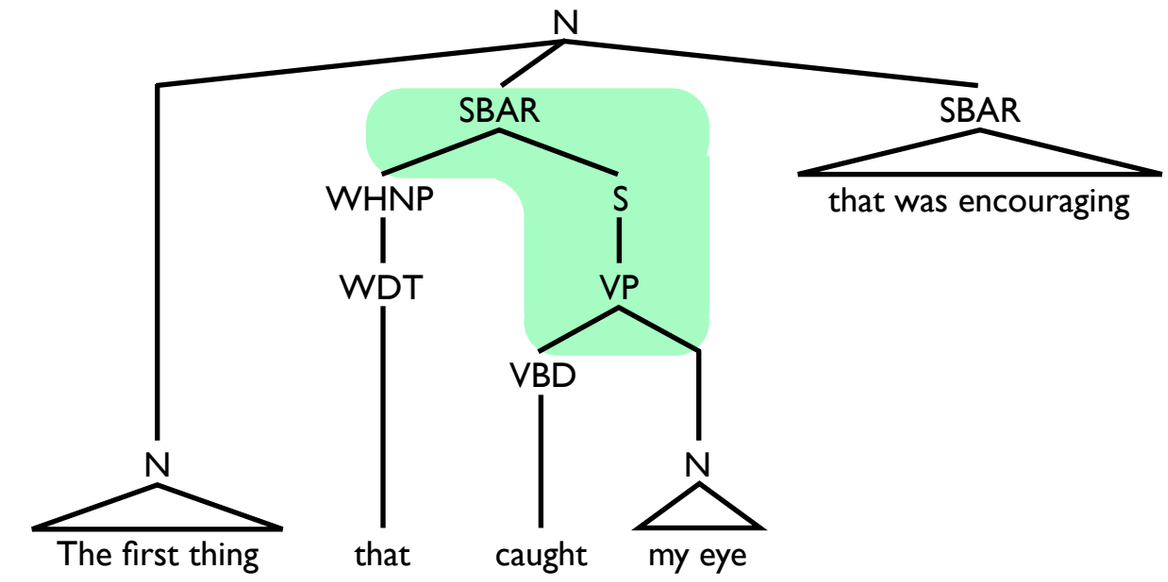


Surface vs Underlying Errors

Surface Errors: Extra SBAR, S, VP, NP

Missing SBAR, S, VP



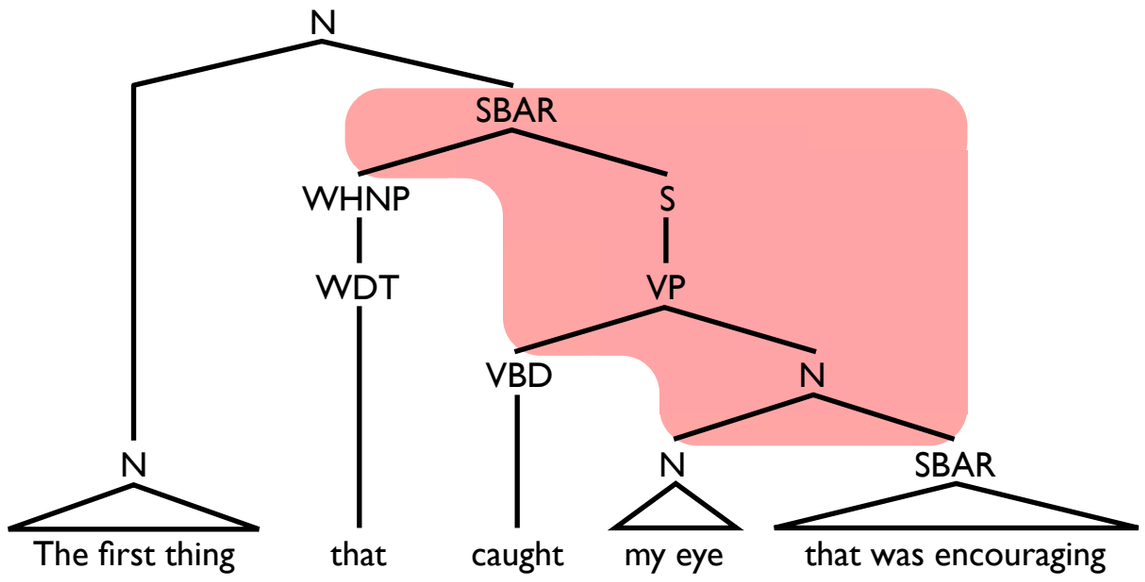


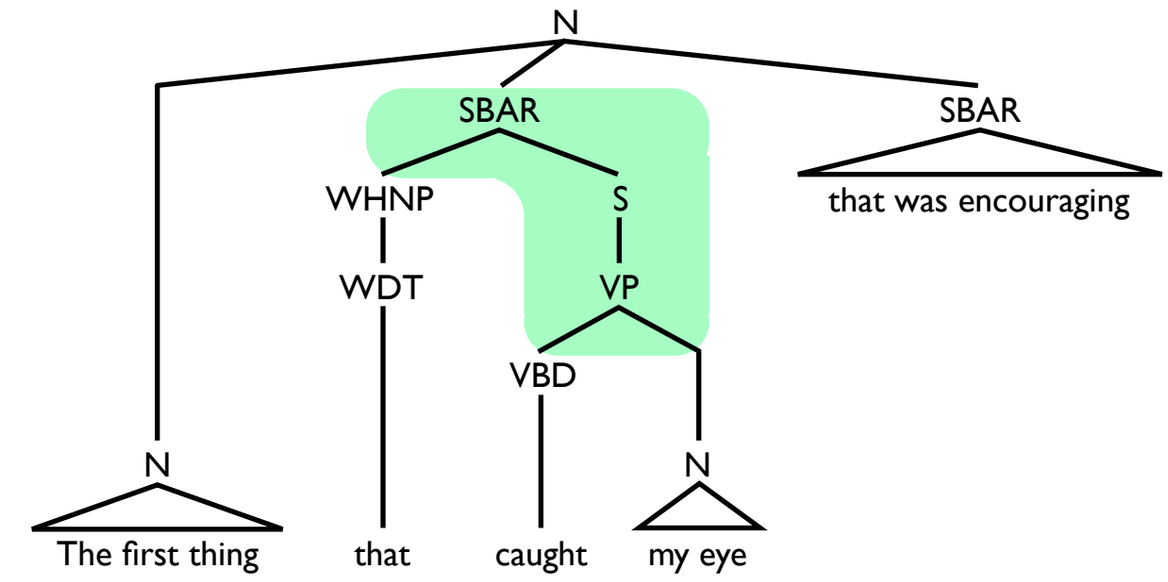
Surface vs Underlying Errors

Surface Errors: Extra SBAR, S, VP, NP

Missing SBAR, S, VP

Underlying Error: Clause Attachment

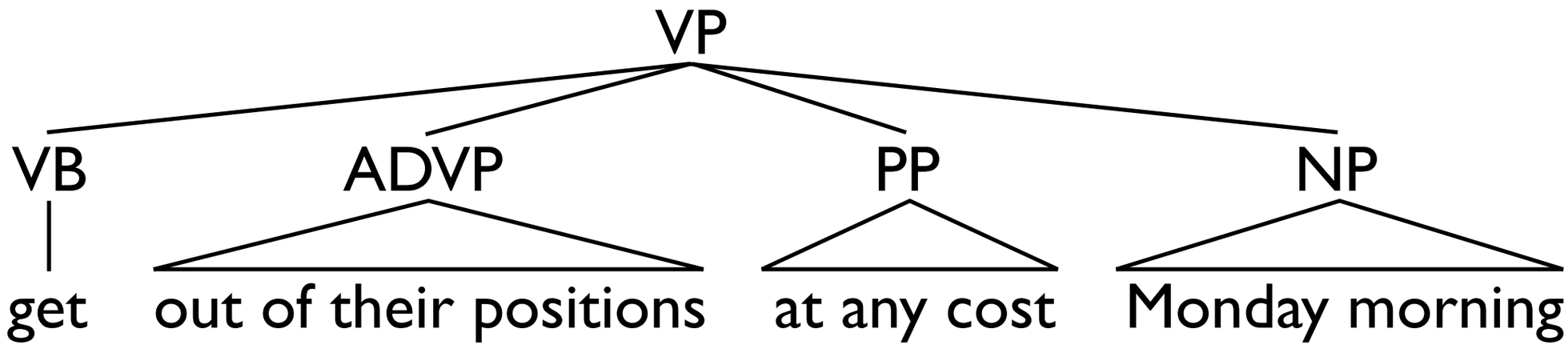




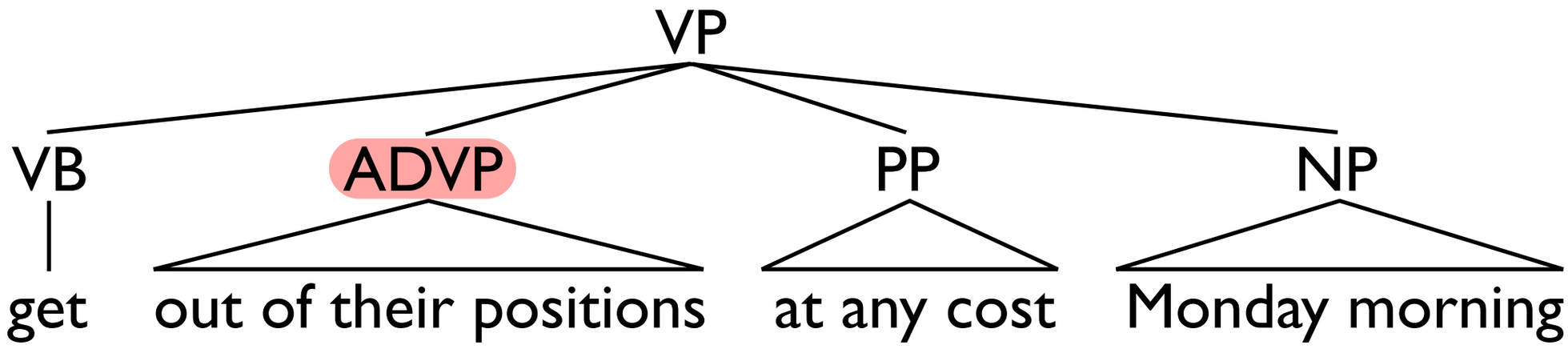
Surface vs Underlying Errors

get out of their positions at any cost Monday morning

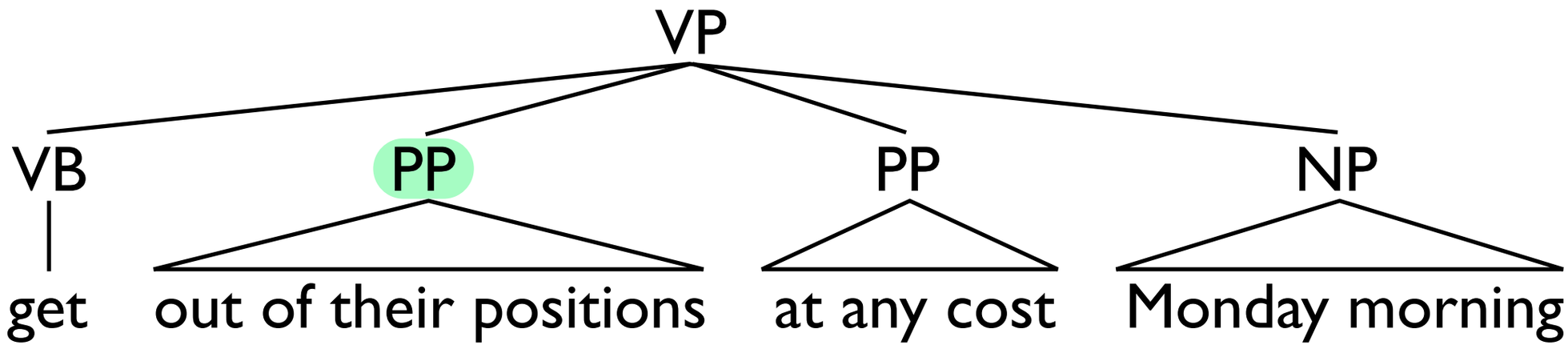
Surface vs Underlying Errors



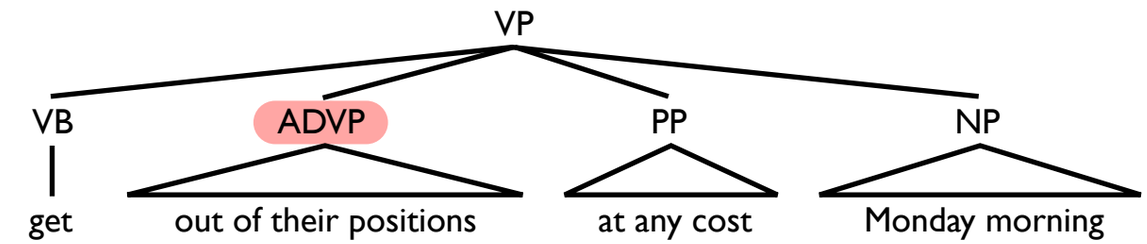
Surface vs Underlying Errors

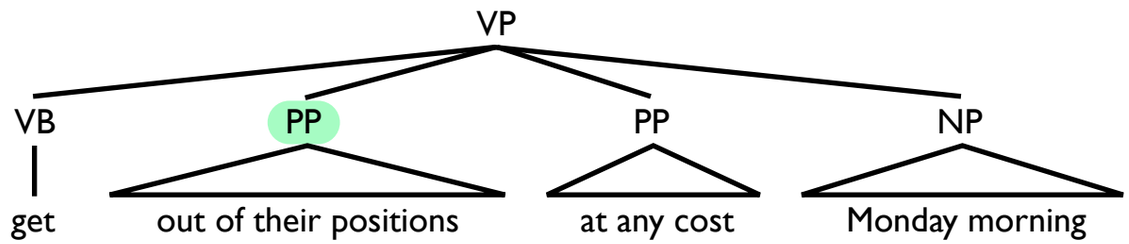


Surface vs Underlying Errors



Surface vs Underlying Errors

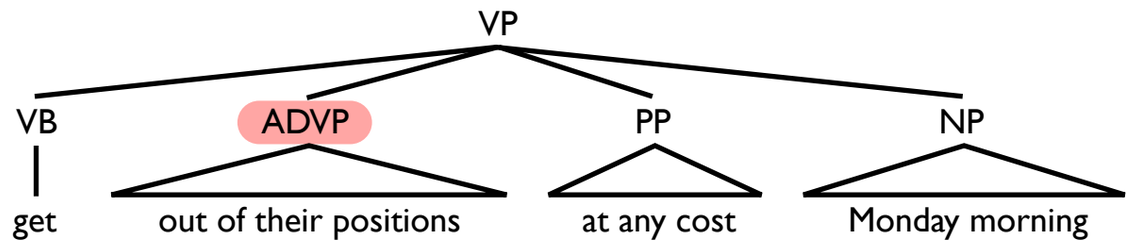


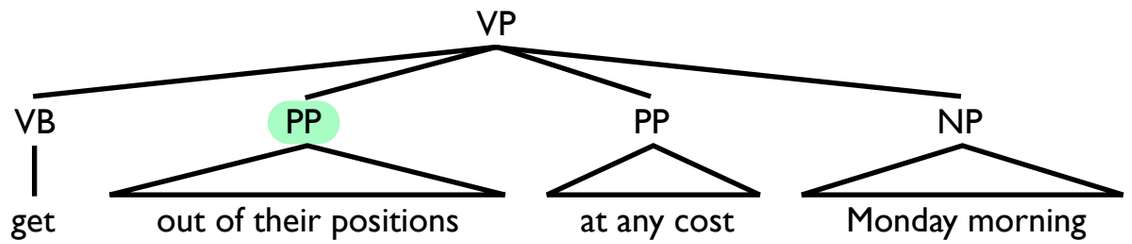


Surface vs Underlying Errors

Surface Errors: Extra ADVP

Missing PP



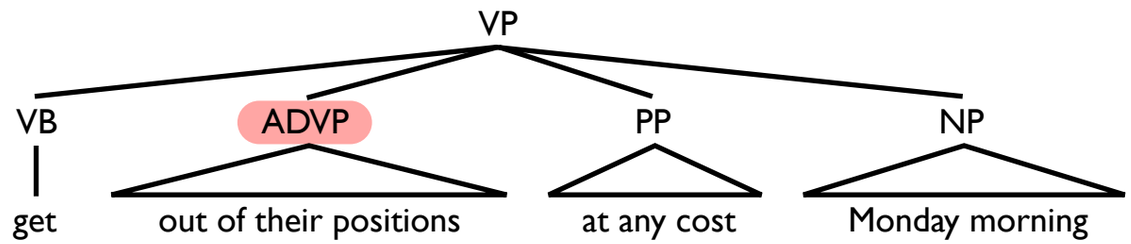


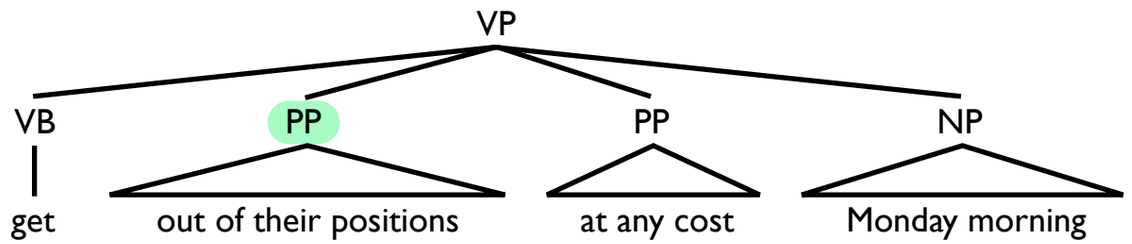
Surface vs Underlying Errors

Surface Errors: Extra ADVP

Missing PP

Underlying Error: Different Label





Underlying Errors

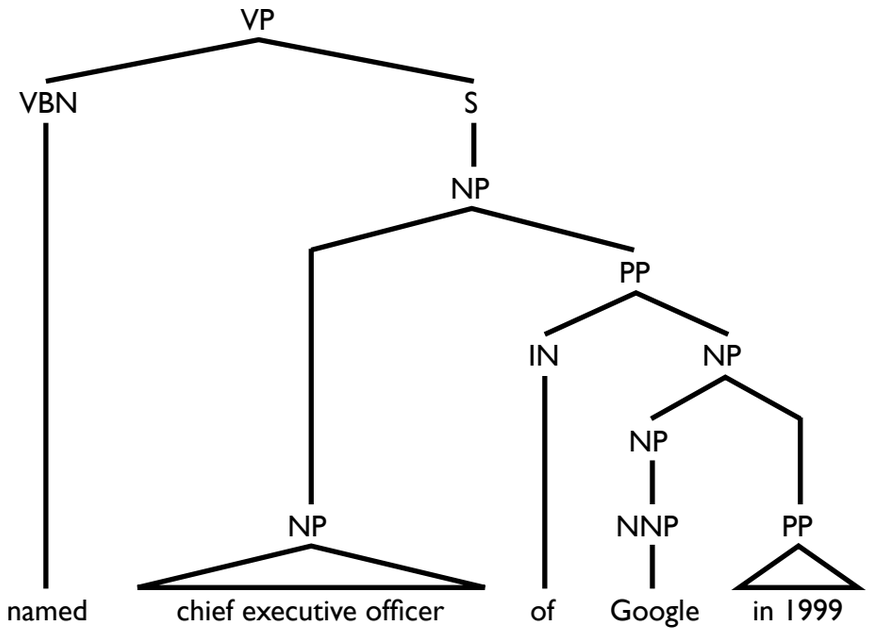
PP Attachment
Clause Attachment
Different Label
Modifier Attachment
NP Attachment
Coordination
Single Word Phrases
Unary
NP Internal Structure
Other

Underlying Errors

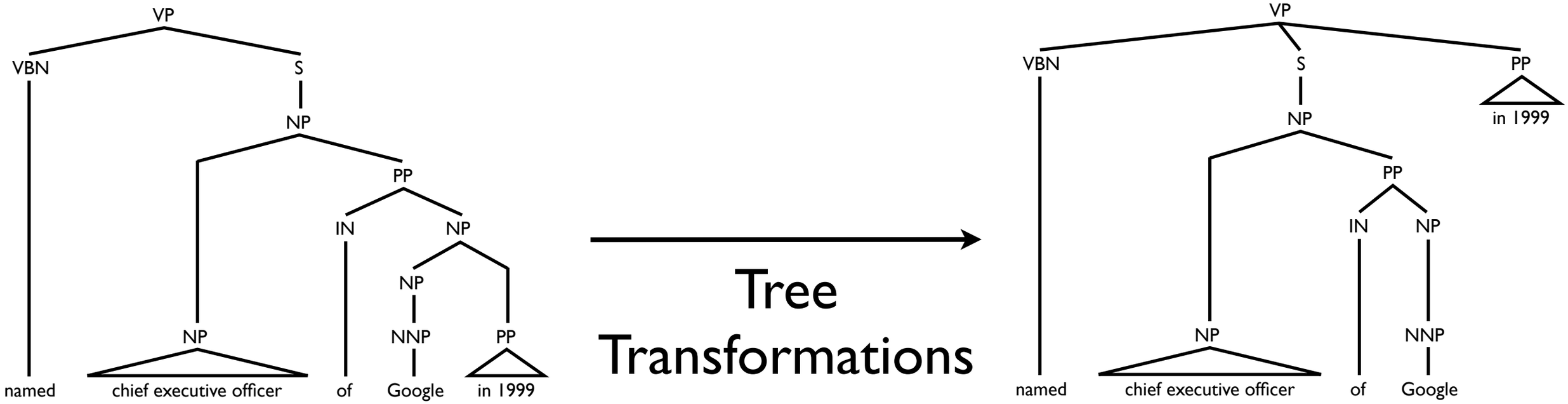
PP Attachment
Clause Attachment
Different Label
Modifier Attachment
NP Attachment
Coordination
Single Word Phrases
Unary
NP Internal Structure
Other

Error Identification

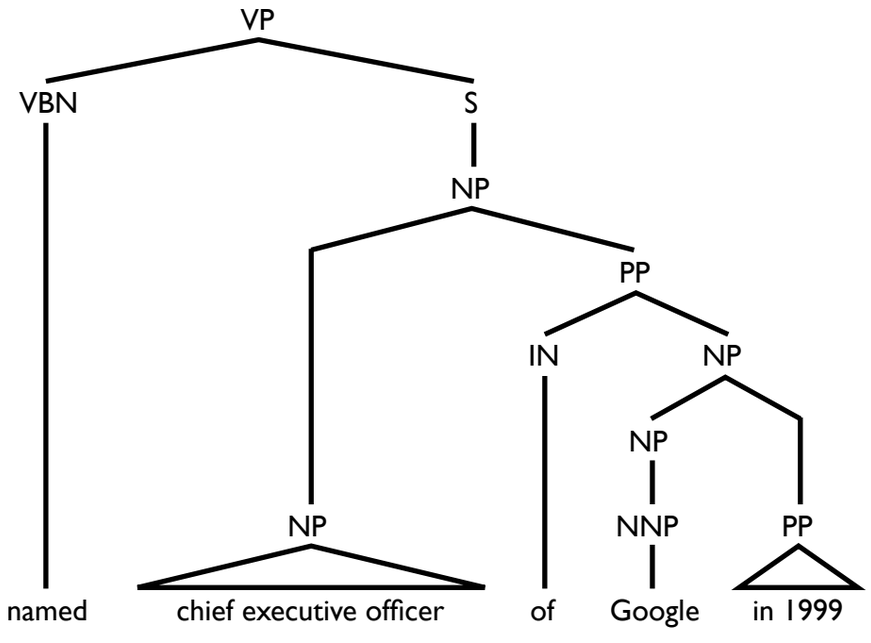
Error Identification



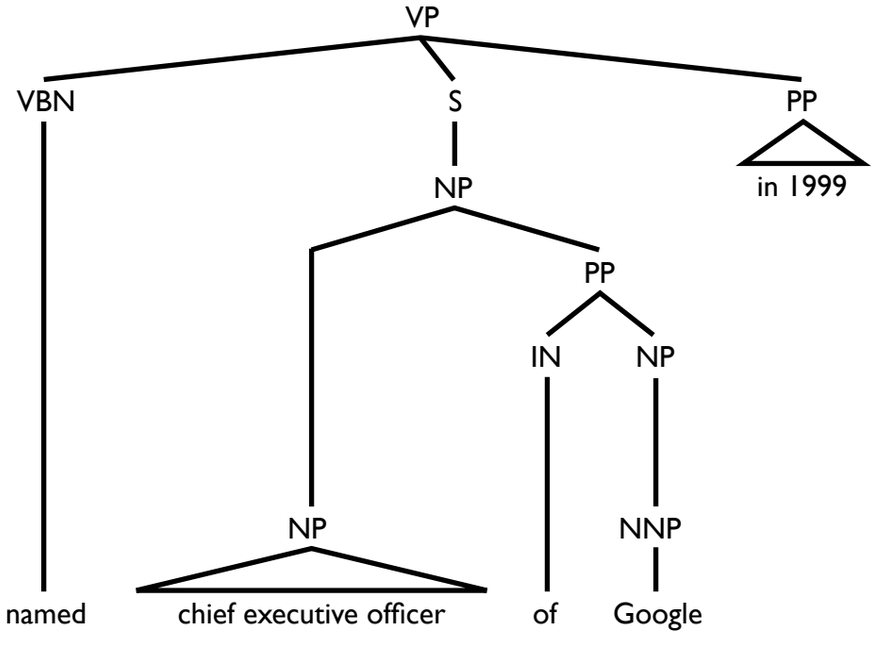
Error Identification



Error Identification

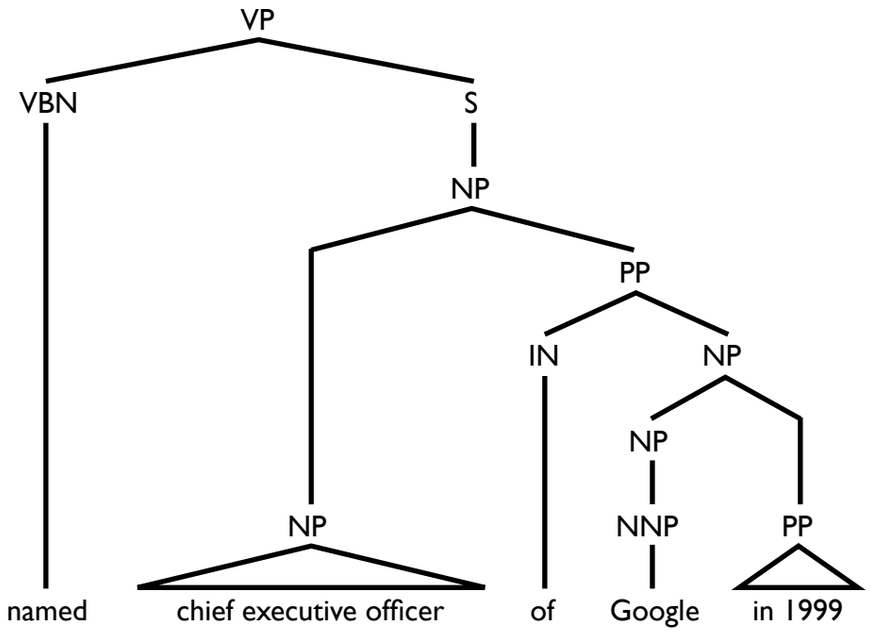


Tree Transformations

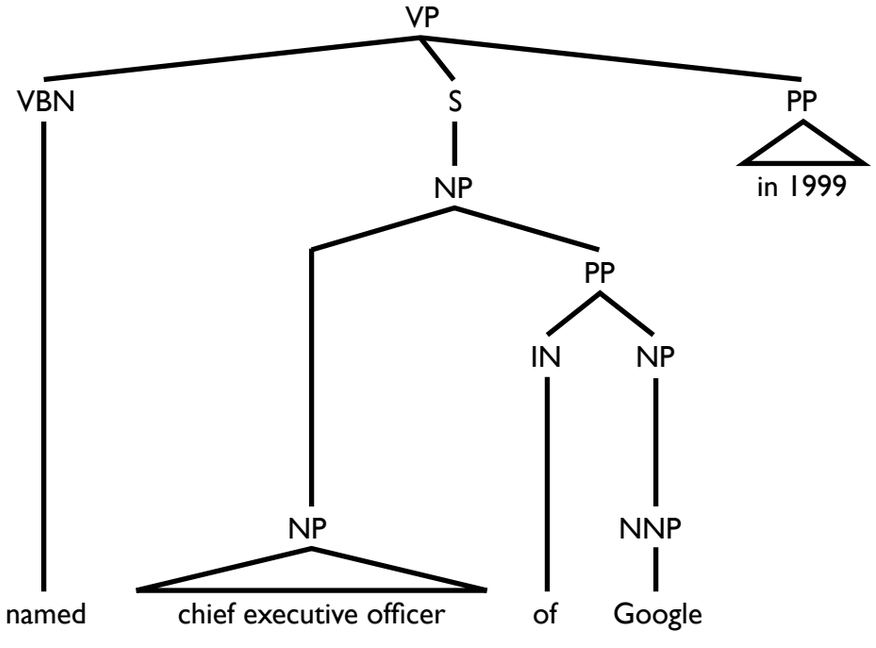


List of Transformation Steps

Error Identification



Tree Transformations

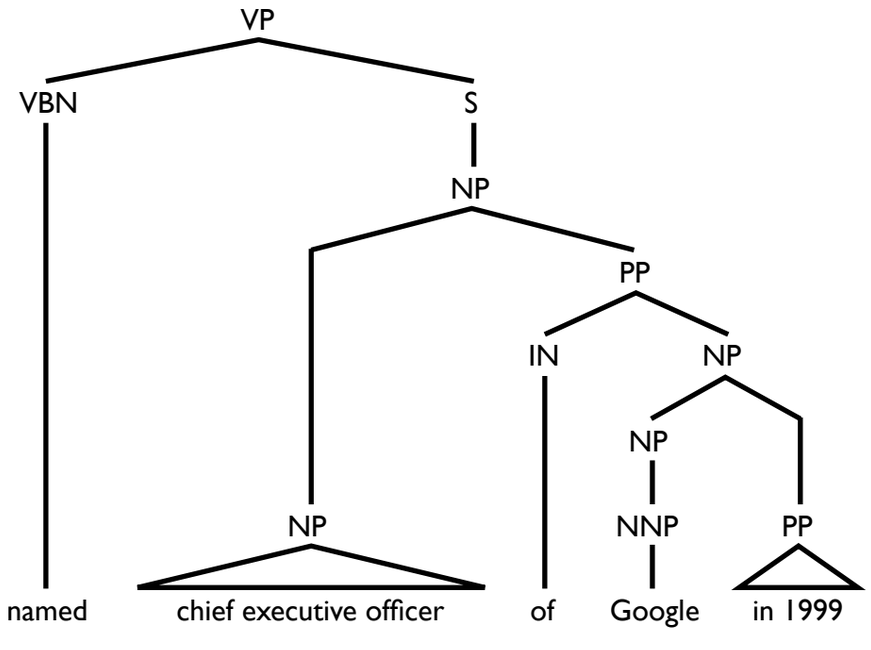


List of Transformation Steps

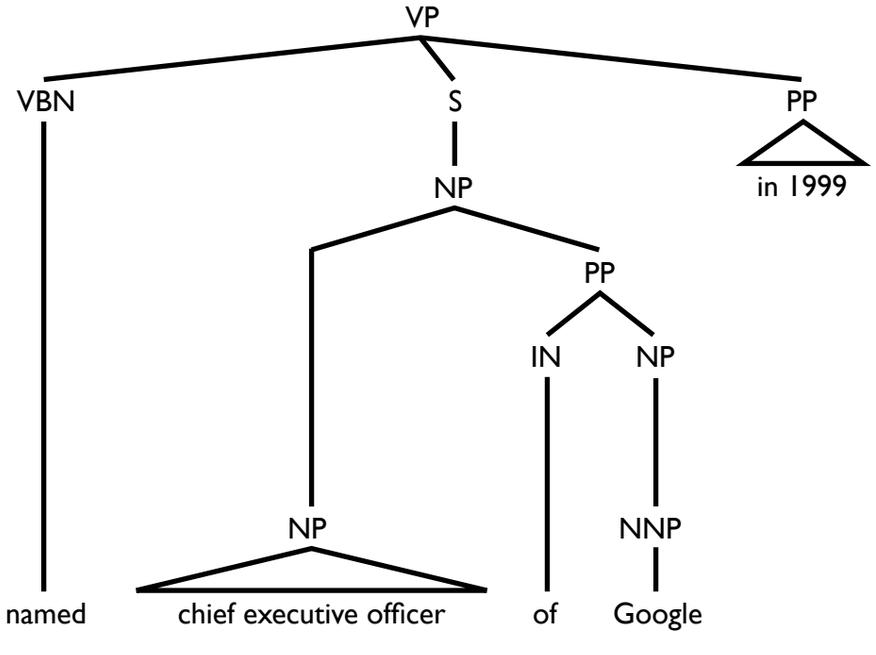
Categorise

List of Errors

Error Identification



Tree Transformations



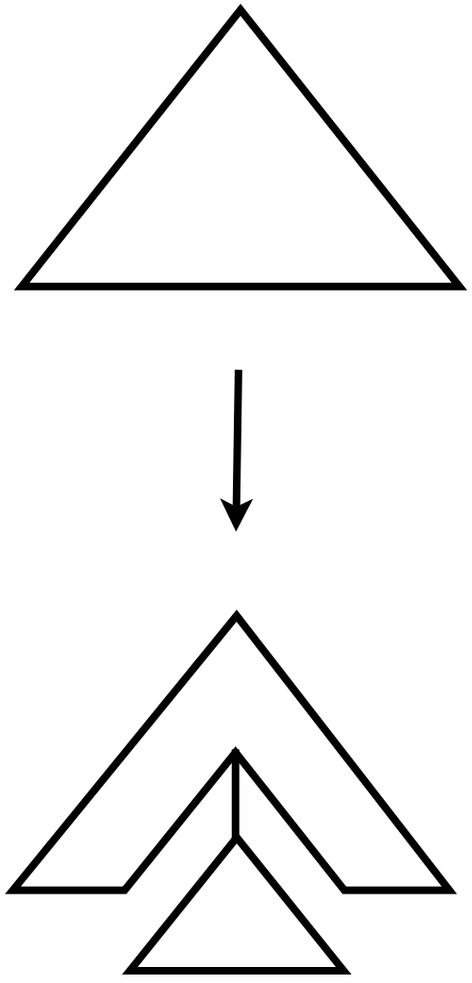
List of Transformation Steps

Categorise

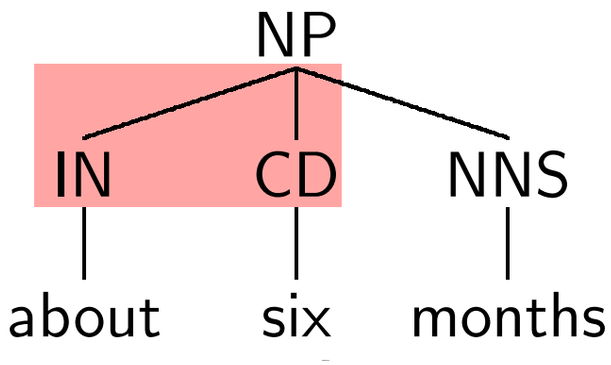
List of Errors

Tree Transformations

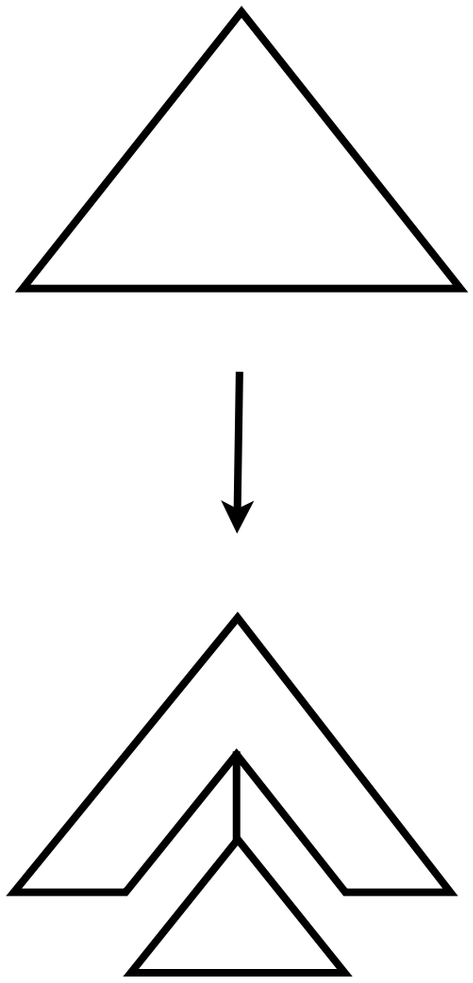
Tree Transformations



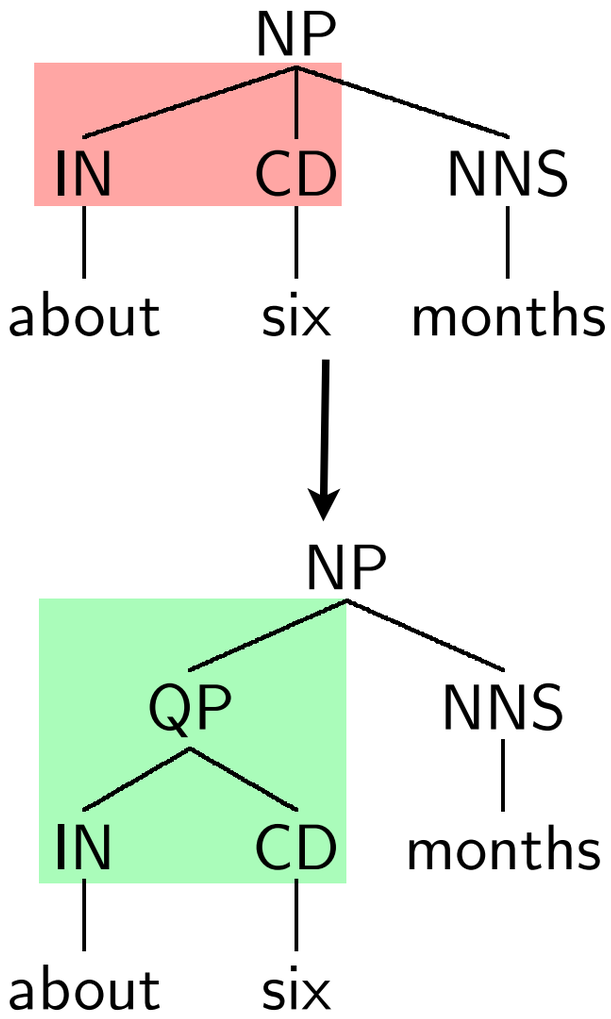
Create Node



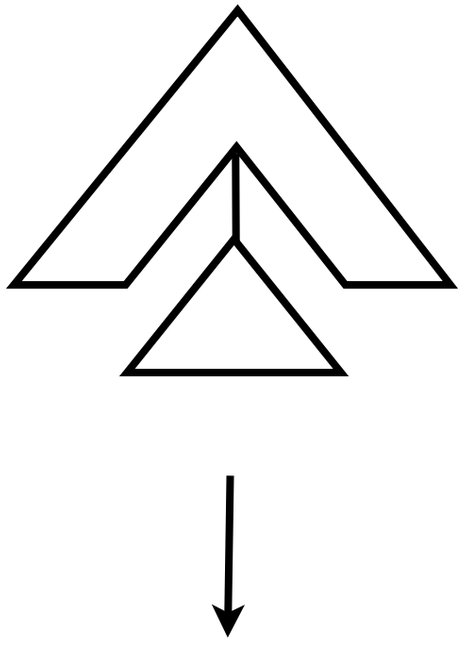
Tree Transformations



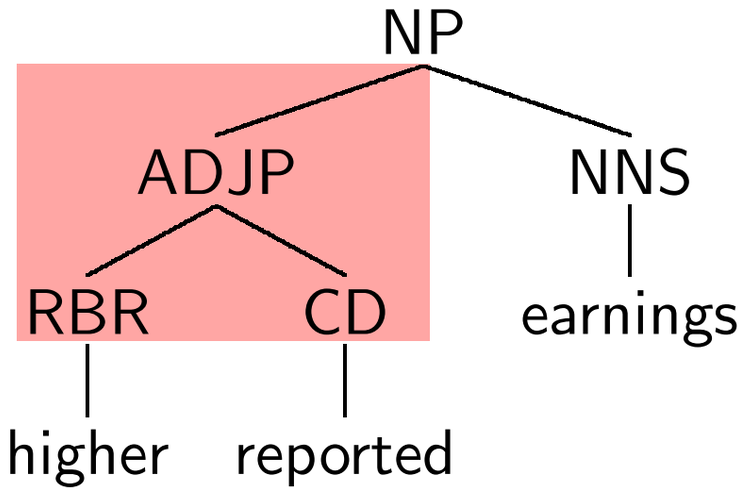
Create Node



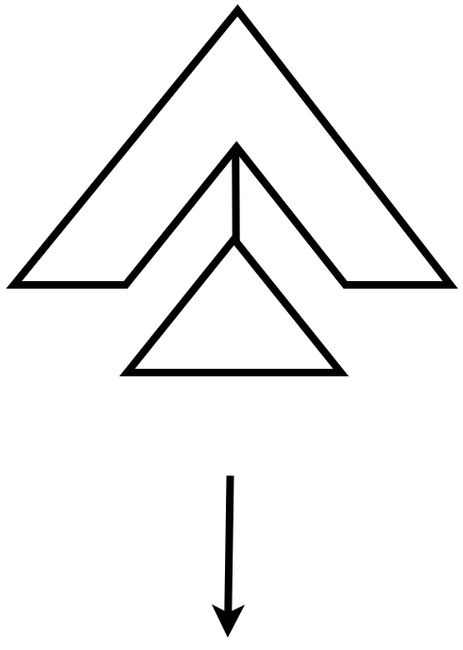
Tree Transformations

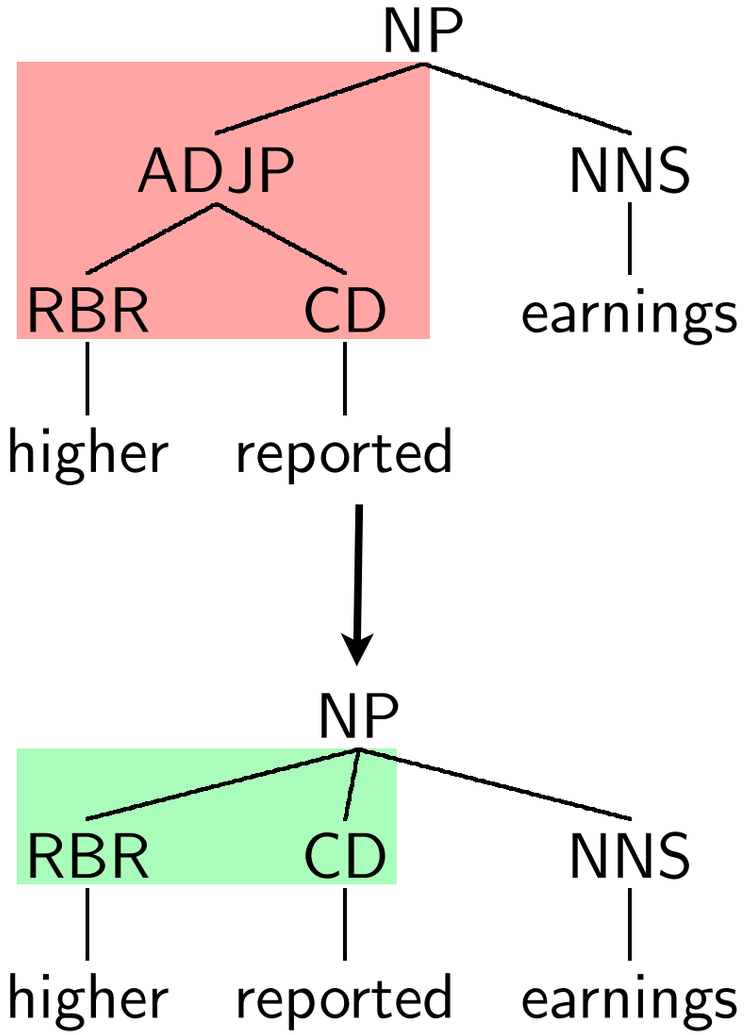


Delete Node



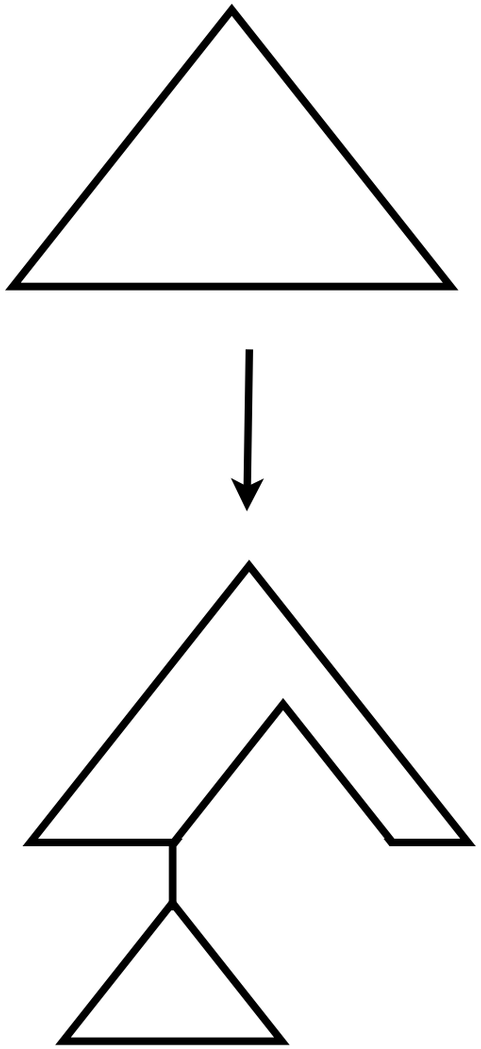
Tree Transformations



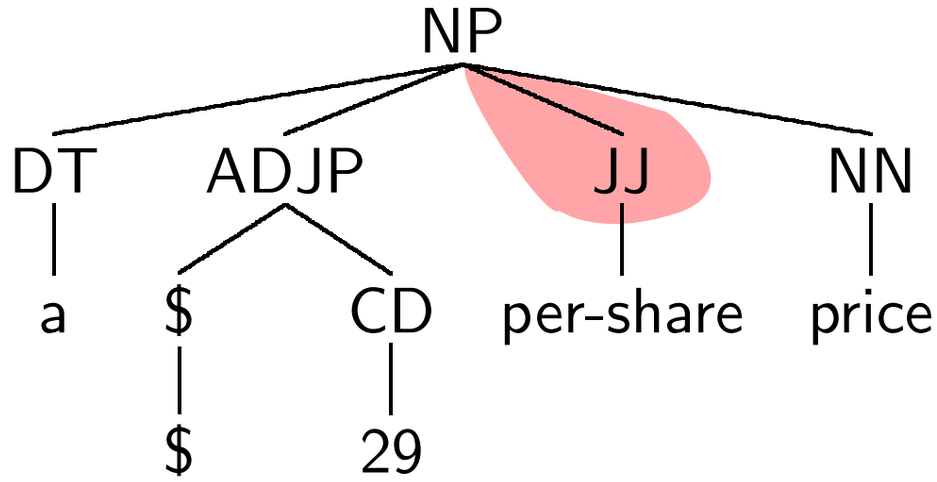


Delete Node

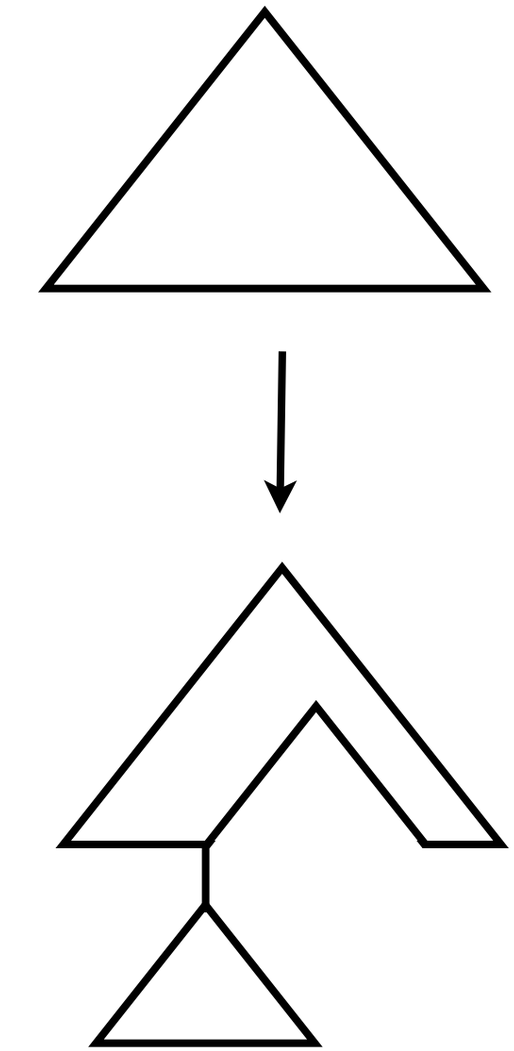
Tree Transformations



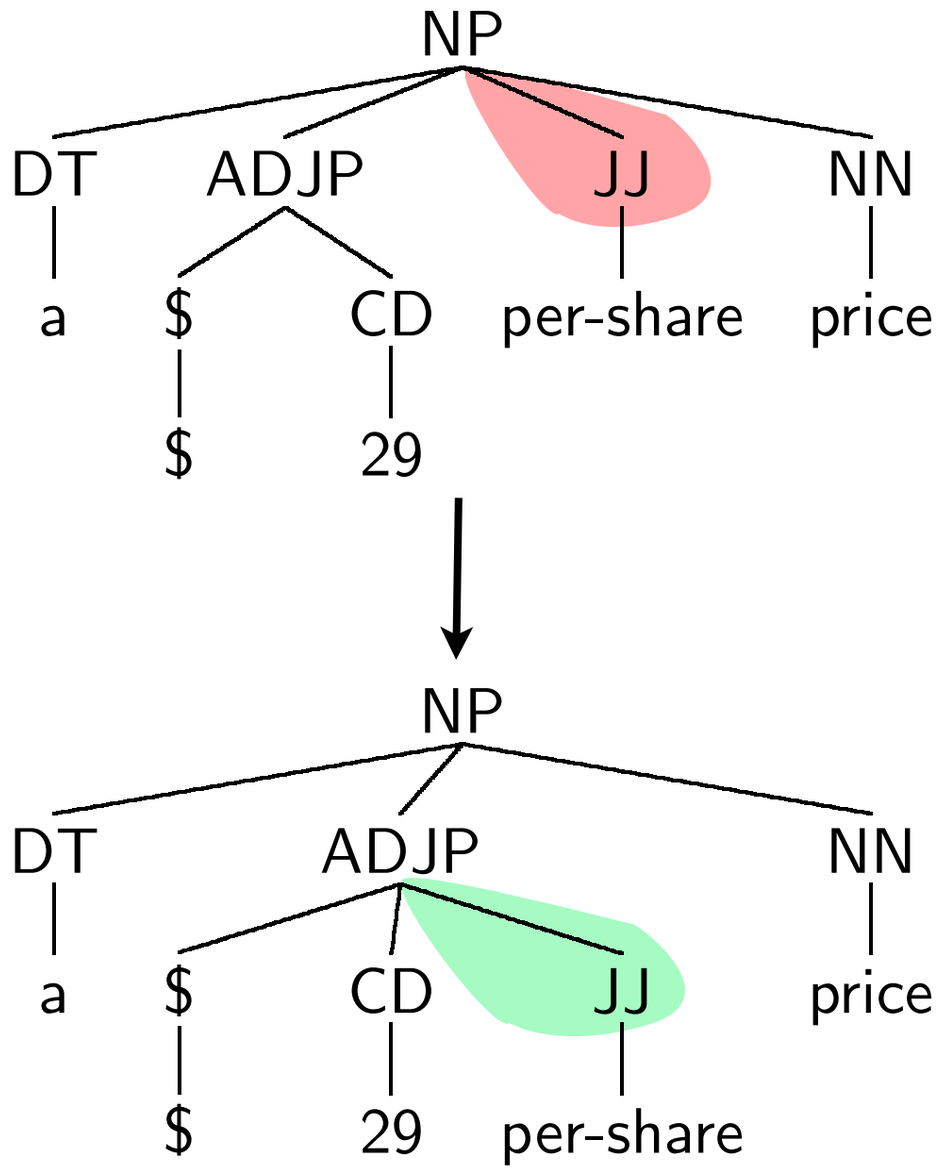
Move Node



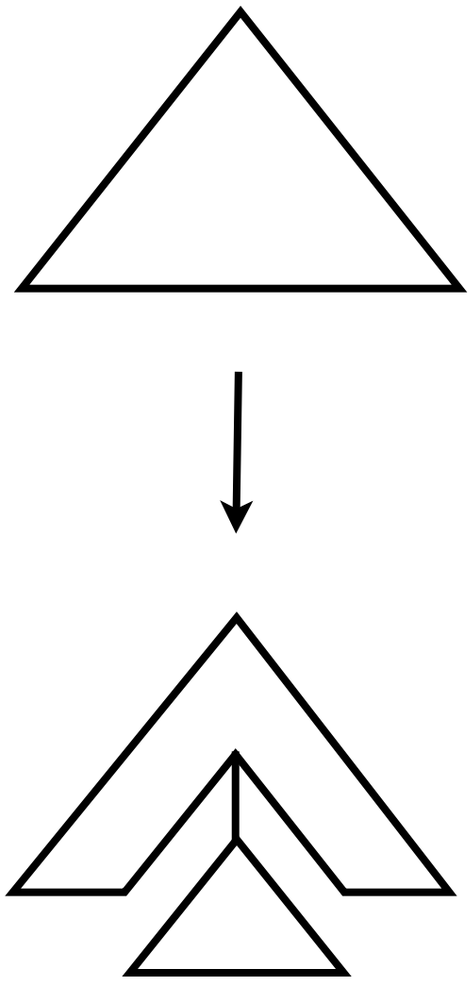
Tree Transformations



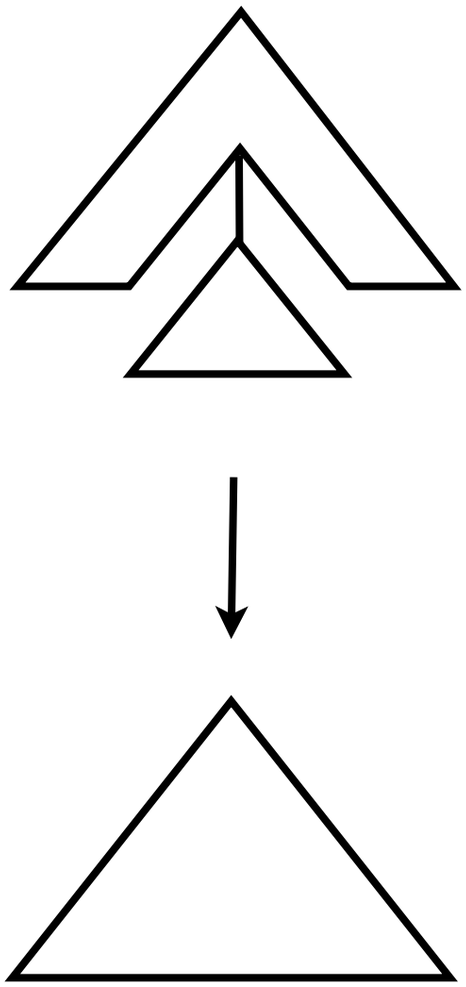
Move Node



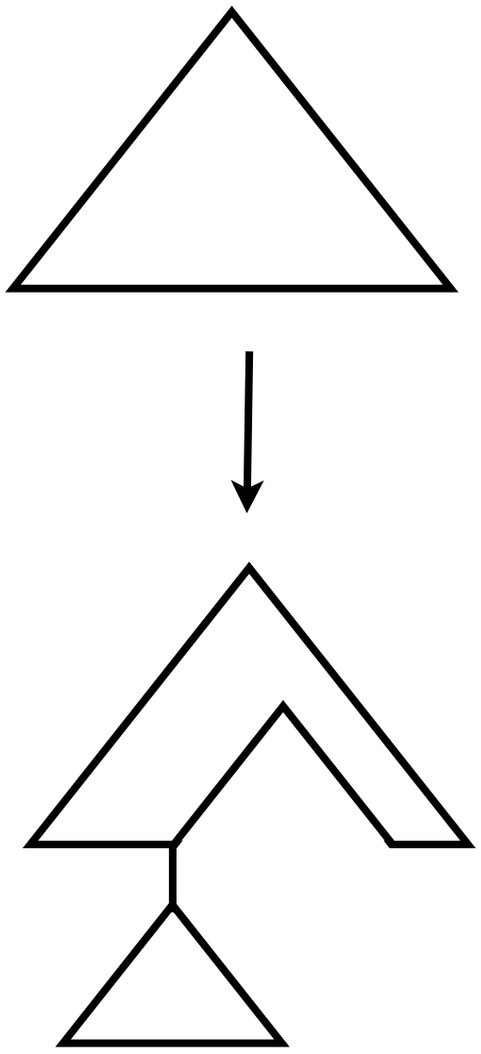
Tree Transformations



Create Node

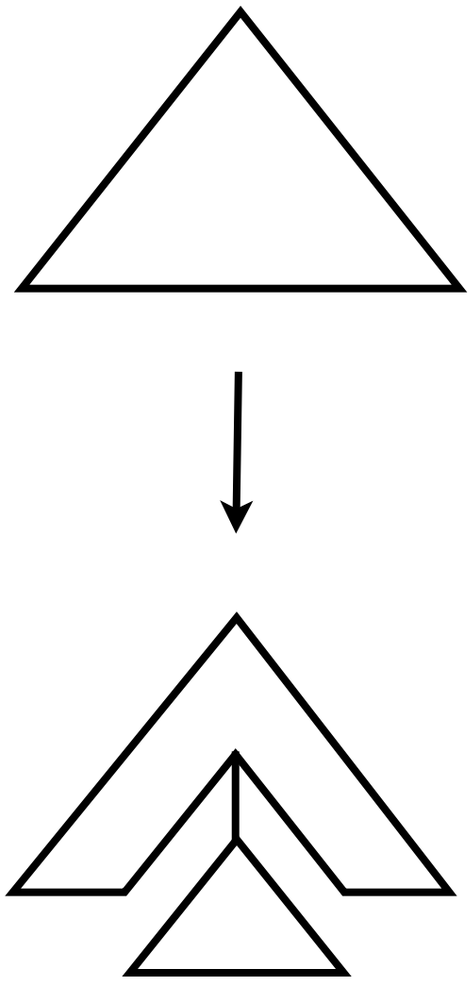


Delete Node

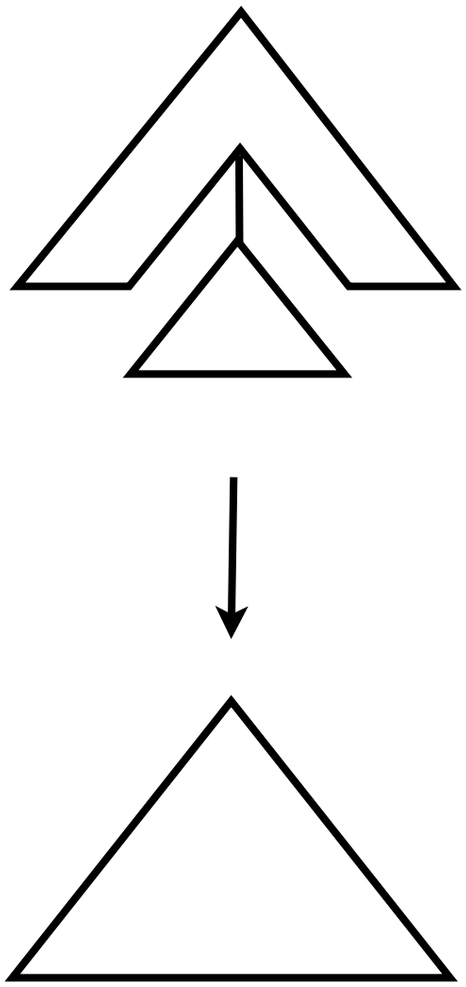


Move Node

Tree Transformations

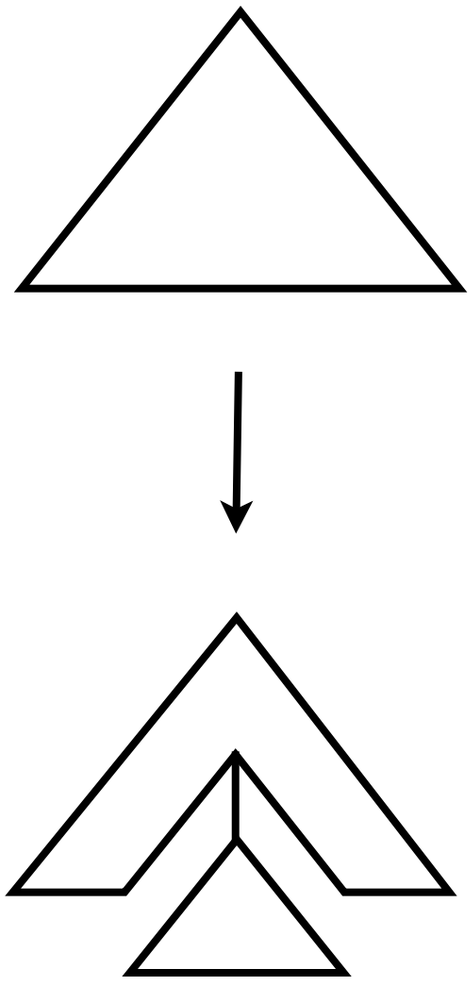


Create Node

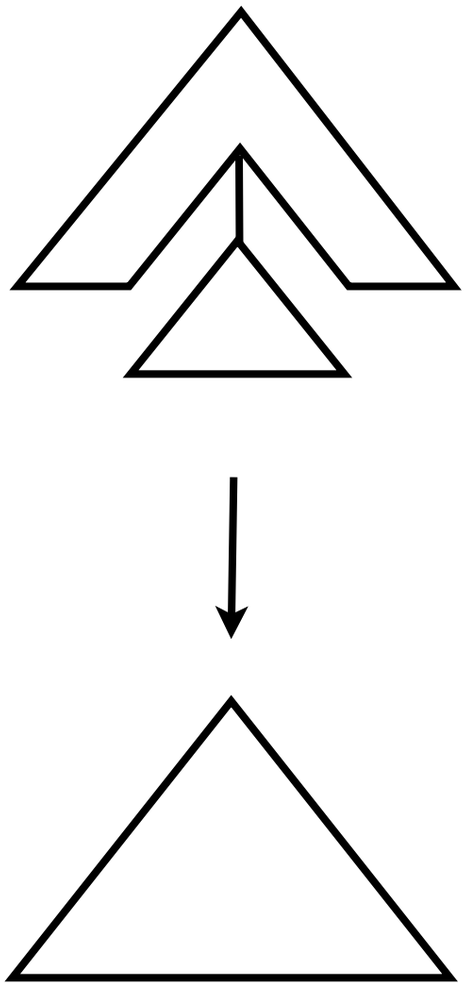


Delete Node

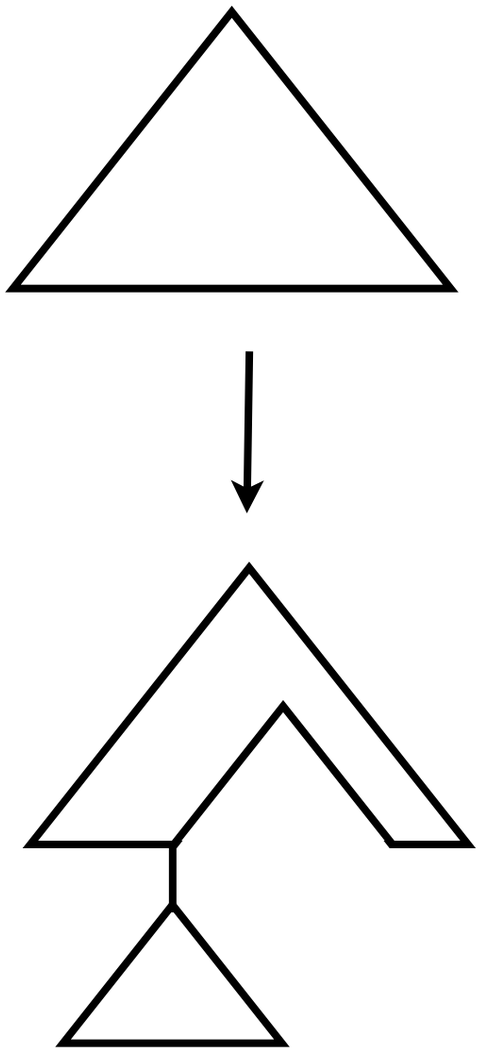
Tree Transformations



Create Node



Delete Node

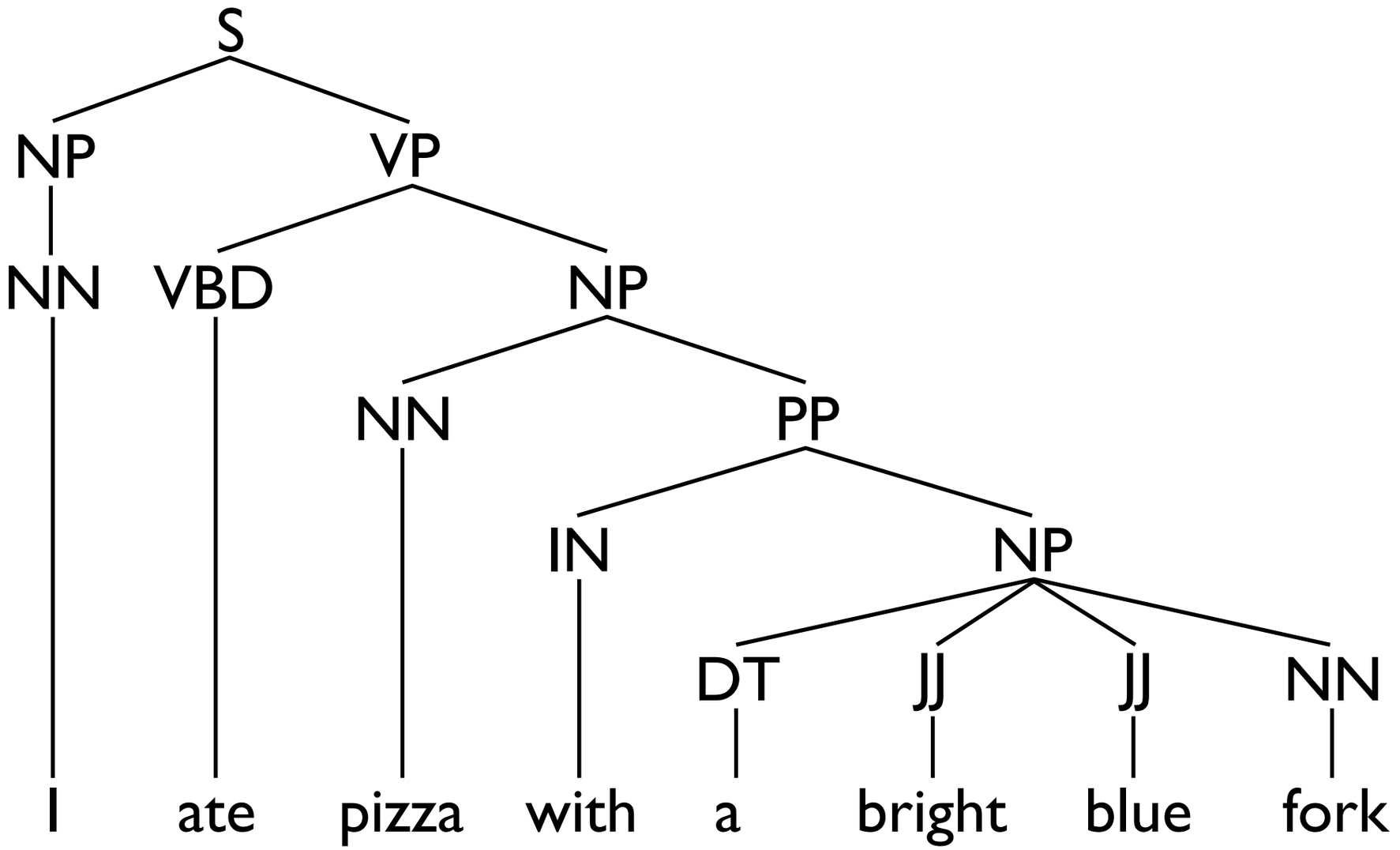


Move Node

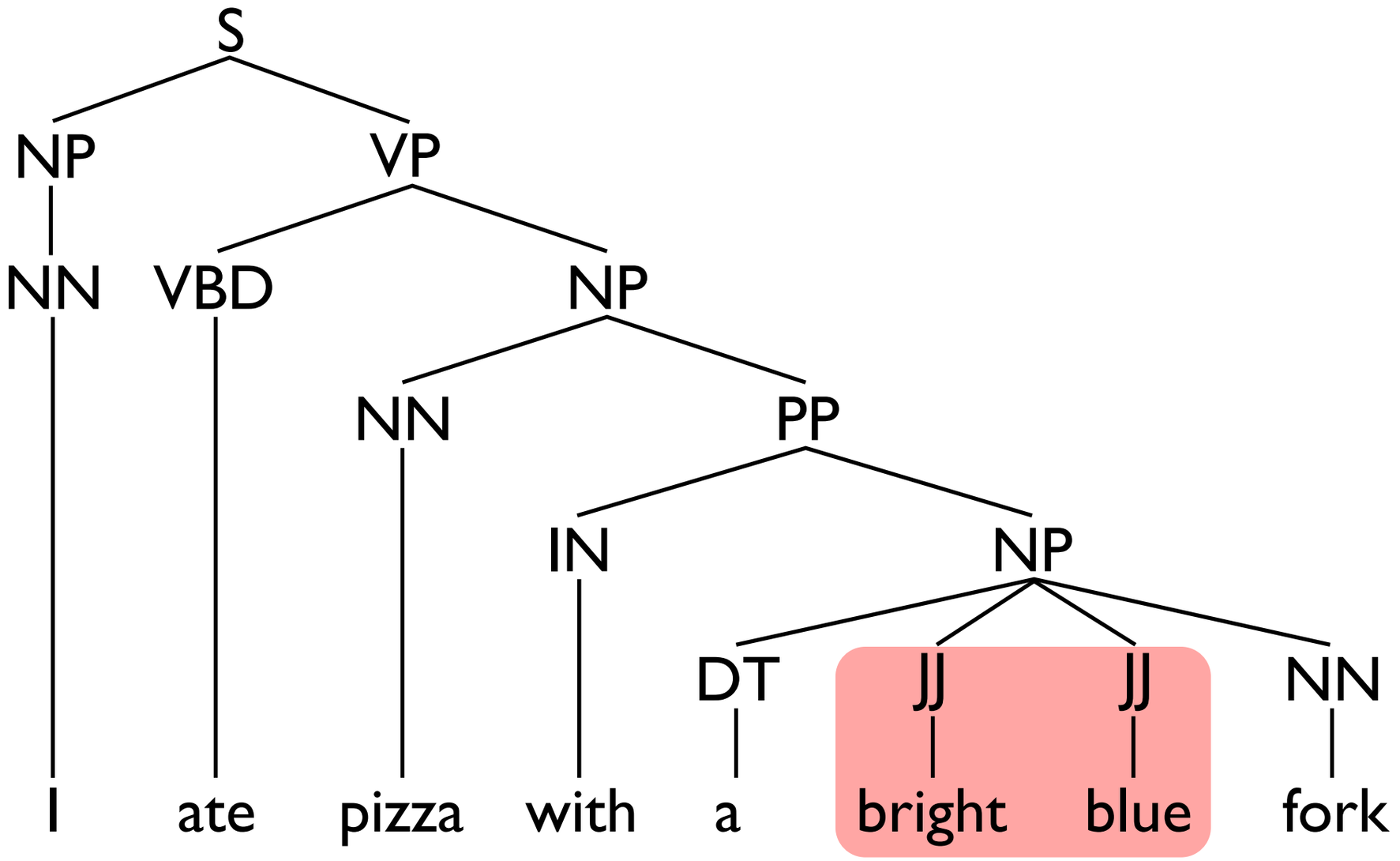
Bottom-Up Greedy Search

I ate pizza with a bright blue fork

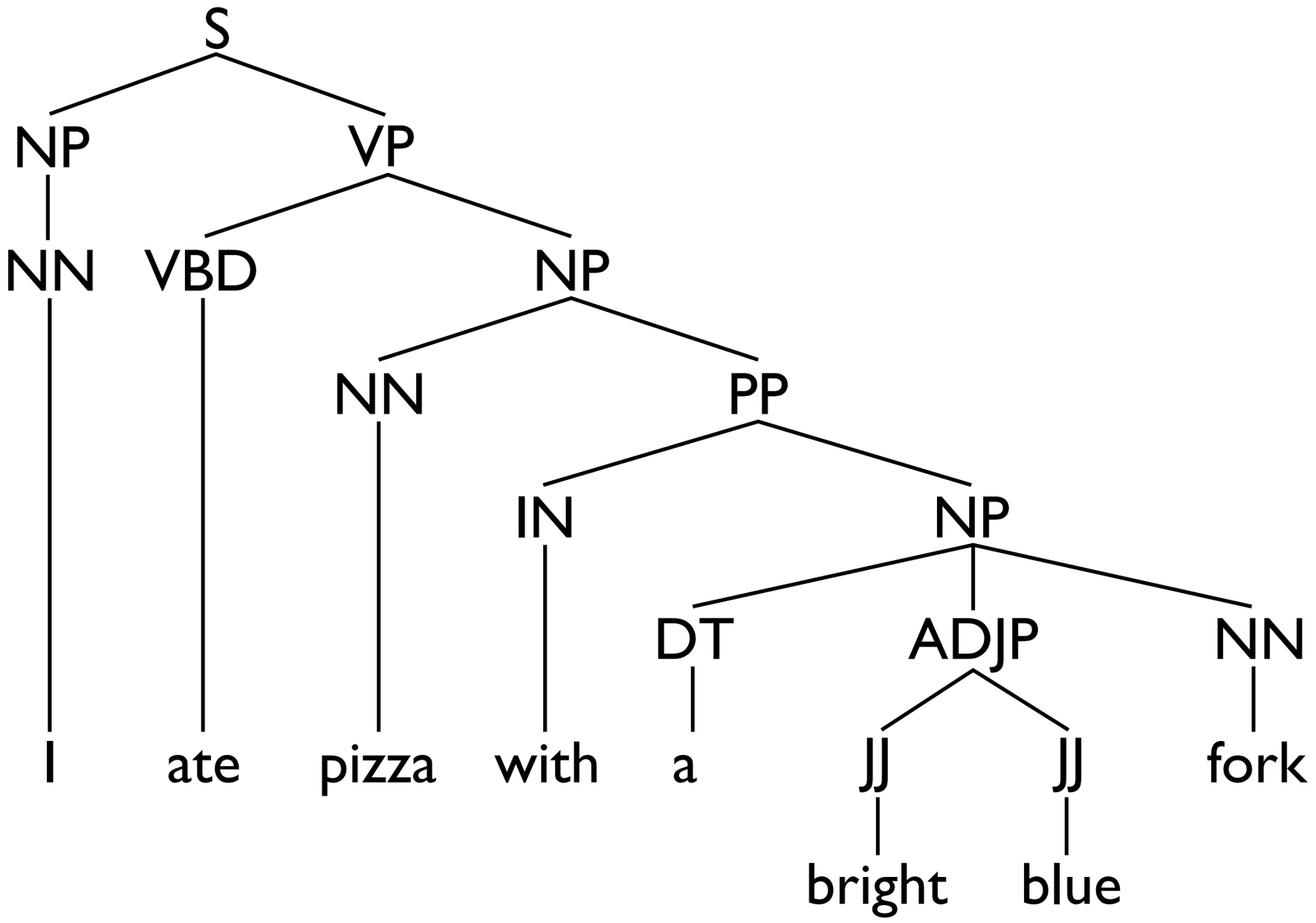
Bottom-Up Greedy Search



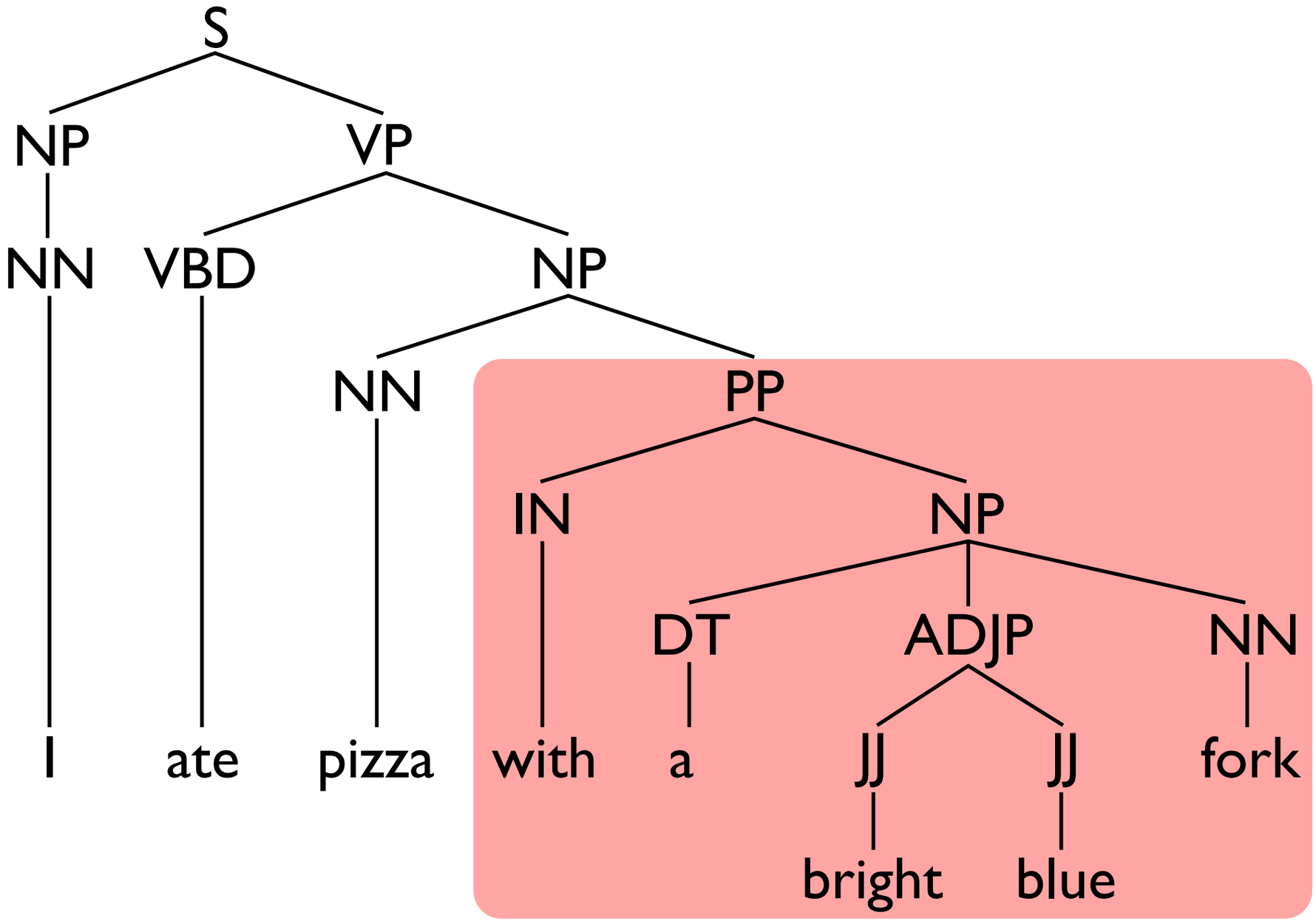
Bottom-Up Greedy Search



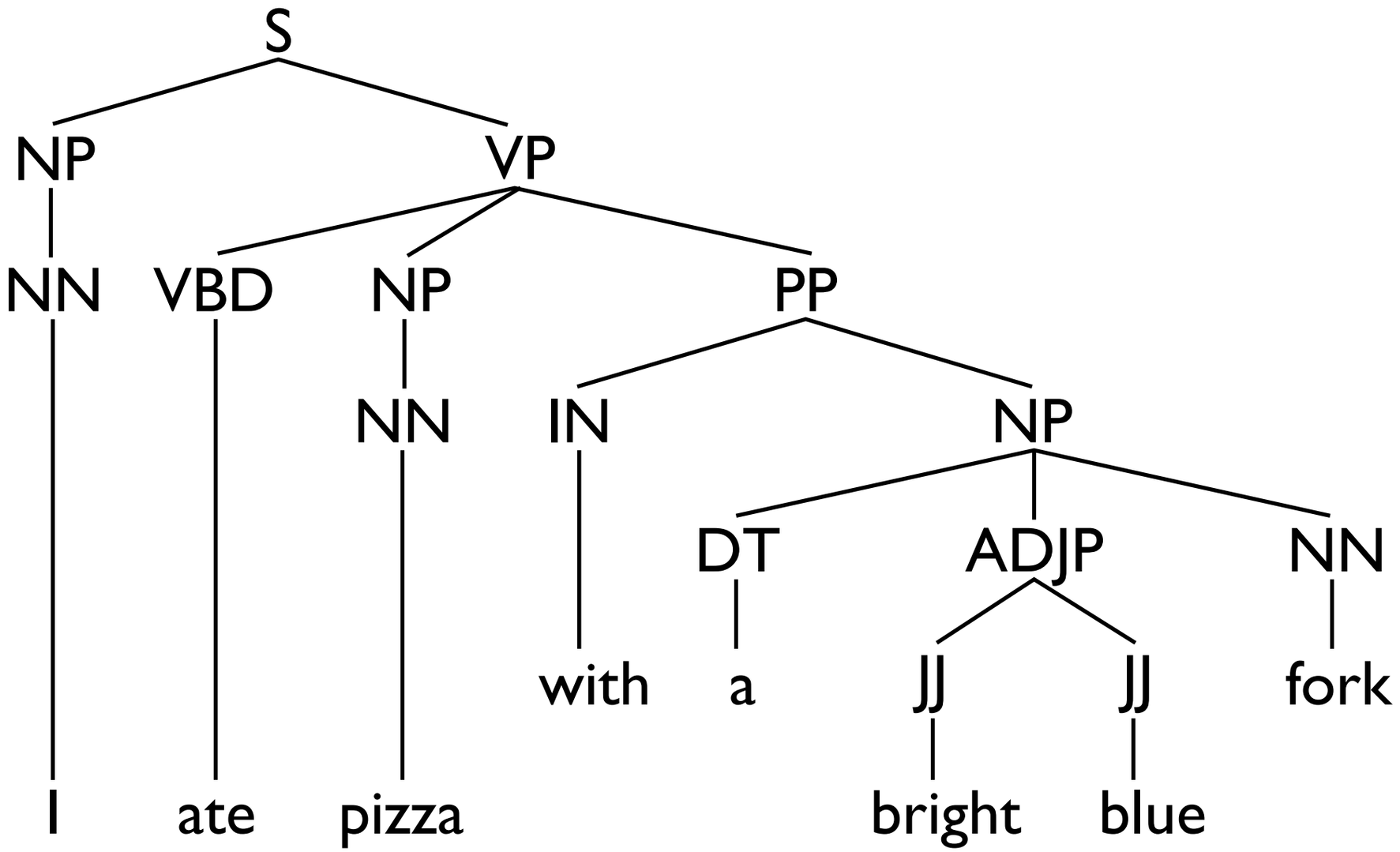
Bottom-Up Greedy Search



Bottom-Up Greedy Search



Bottom-Up Greedy Search



Templates

Templates

named chief executive officer of Google in 1999

Templates

named chief executive officer of Google in 1999

Templates

named chief executive officer of Google in 1999

Templates

named chief executive officer of Google in 1999

- Extra bracket

Templates

named chief executive officer of Google in 1999

- Extra bracket

Templates

named chief executive officer of Google in 1999

- Extra bracket
- Set of extra brackets ending at the same point

Templates

named chief executive officer of Google in 1999

- Extra bracket
- Set of extra brackets ending at the same point

Templates

named chief executive officer of Google in 1999

- Extra bracket
- Set of extra brackets ending at the same point
- Set of missing brackets that cross the extra bracket

Templates

named chief executive officer of Google in 1999

- Extra bracket
- Set of extra brackets ending at the same point
- Set of missing brackets that cross the extra bracket
- No other bracket errors present

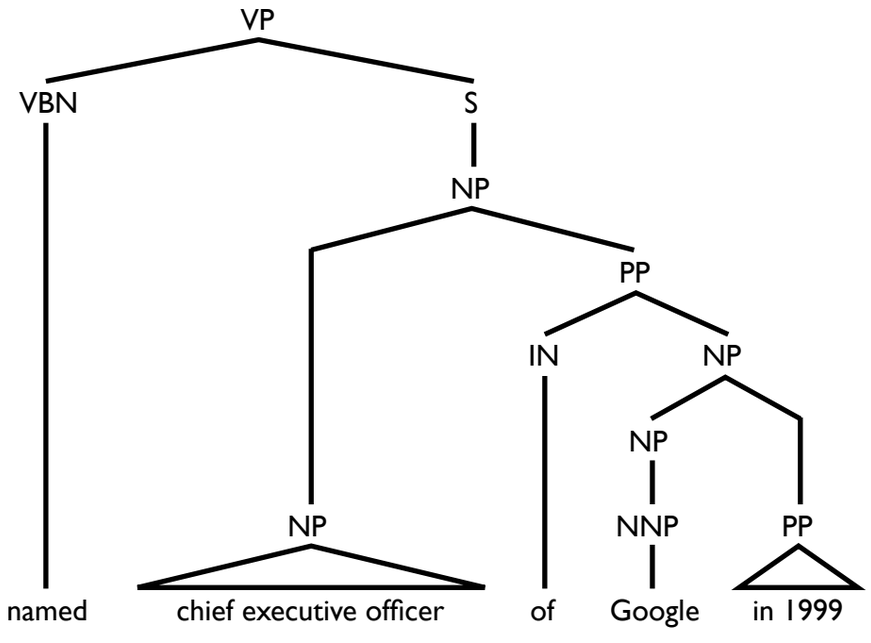
Templates

named chief executive officer of Google in 1999

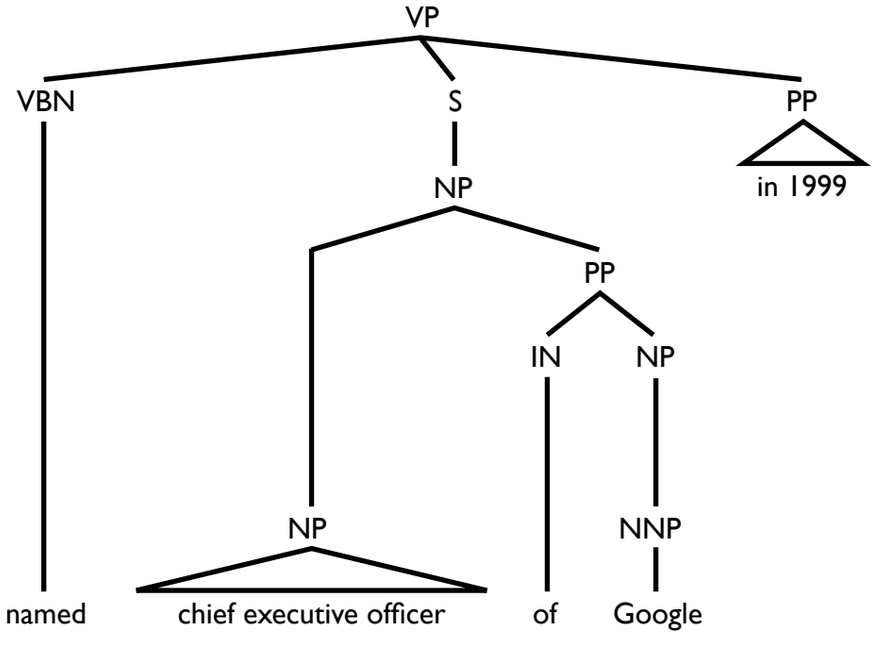
- Extra bracket
- Set of extra brackets ending at the same point
- Set of missing brackets that cross the extra bracket
- No other bracket errors present

Attachment too low, move bracket up

Error Identification



Tree Transformations

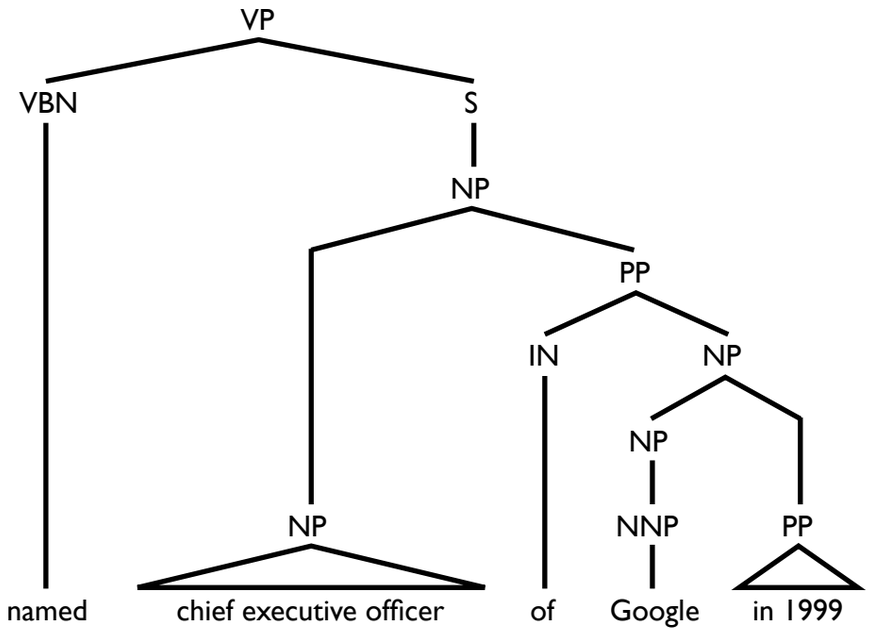


List of Transformation Steps

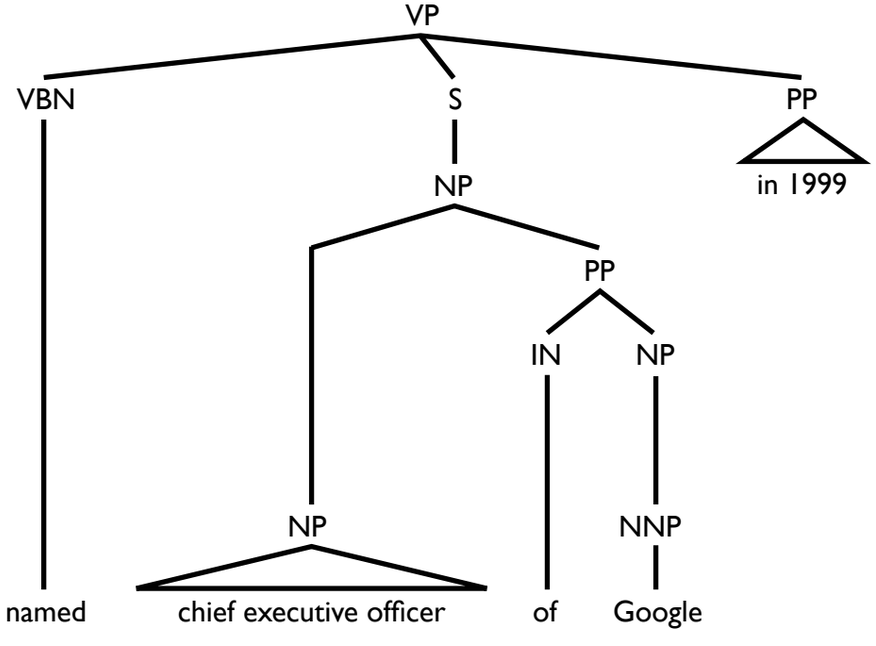
Categorise

List of Errors

Error Identification



Tree Transformations

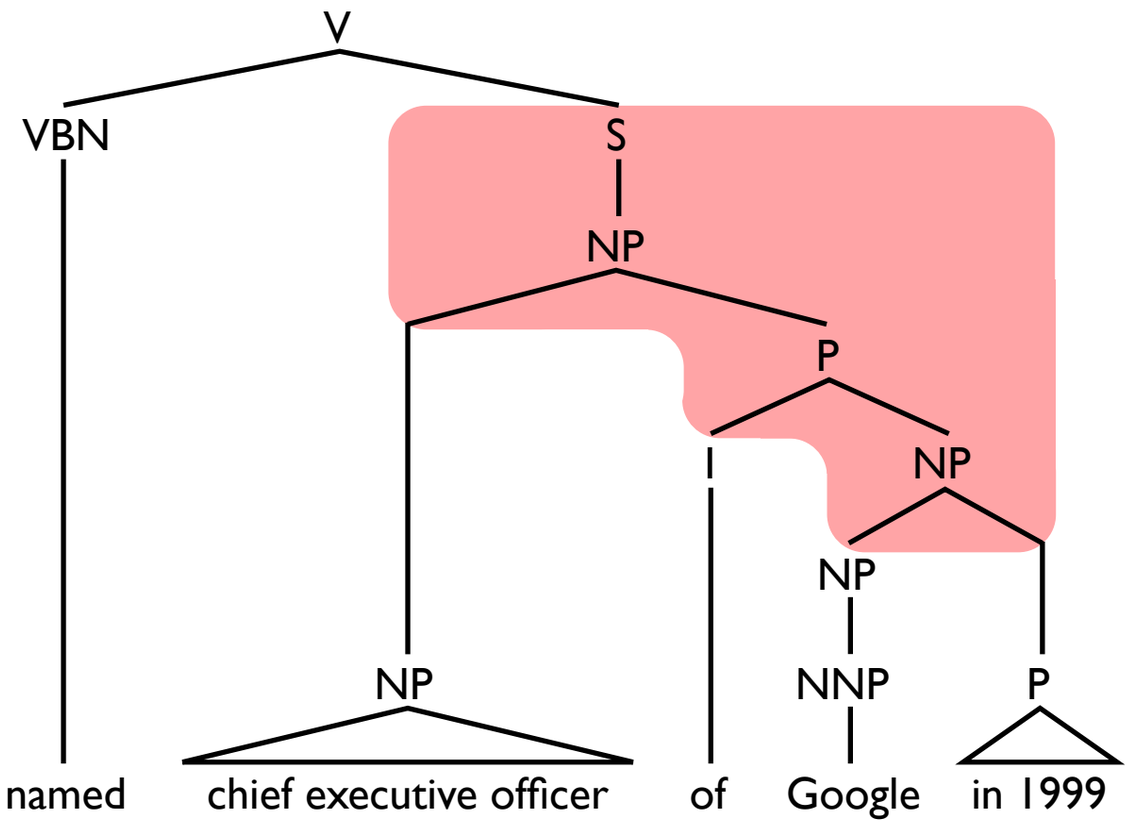


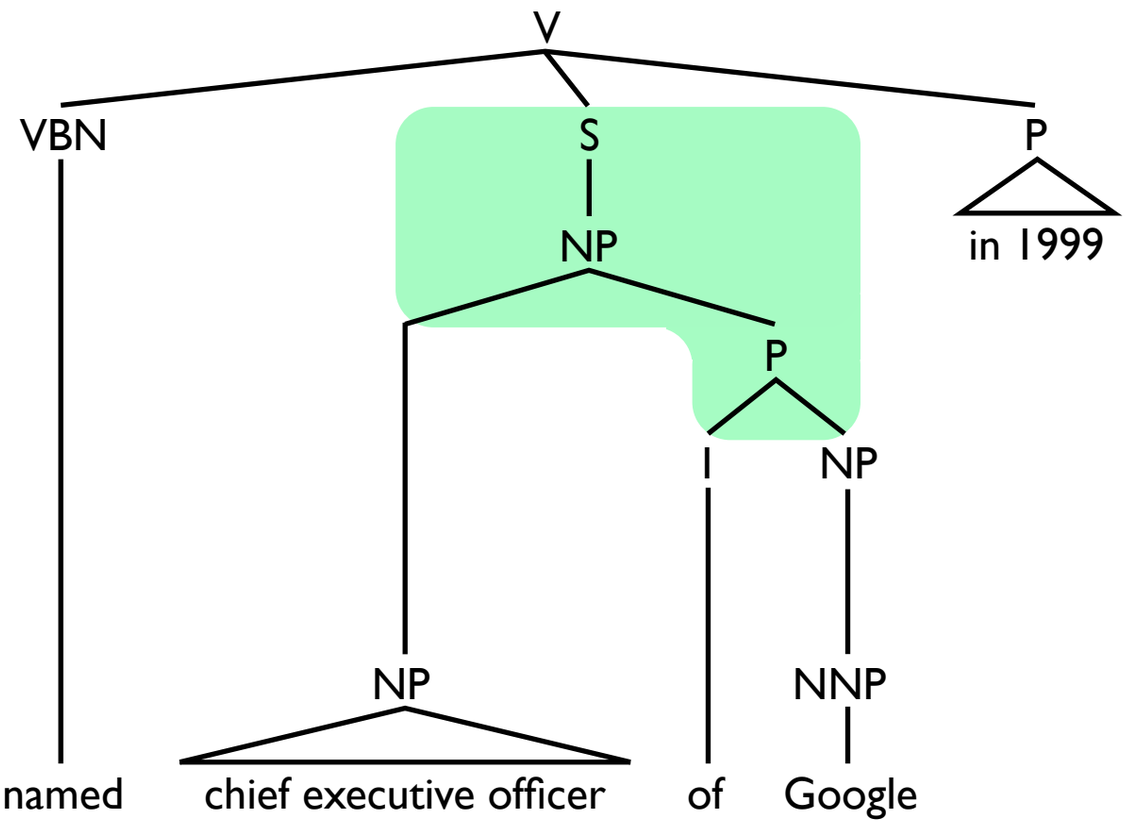
List of Transformation Steps

Categorise

List of Errors

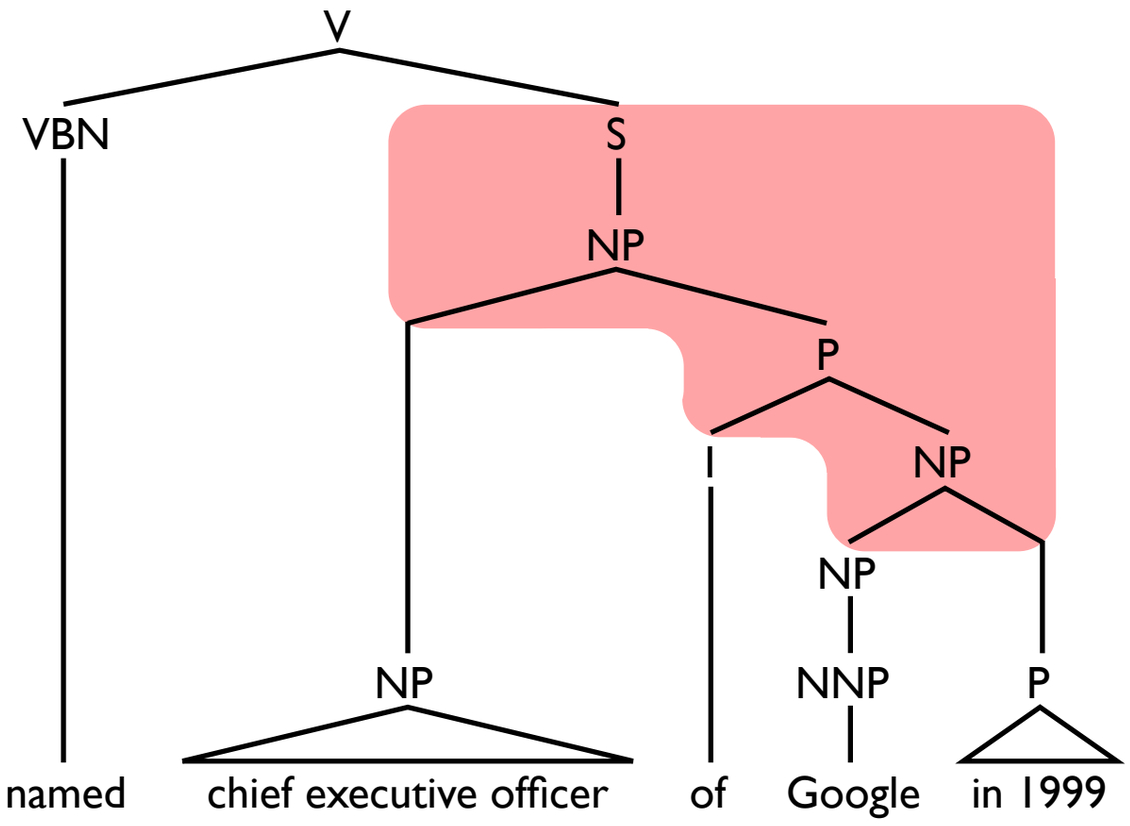
Categorise

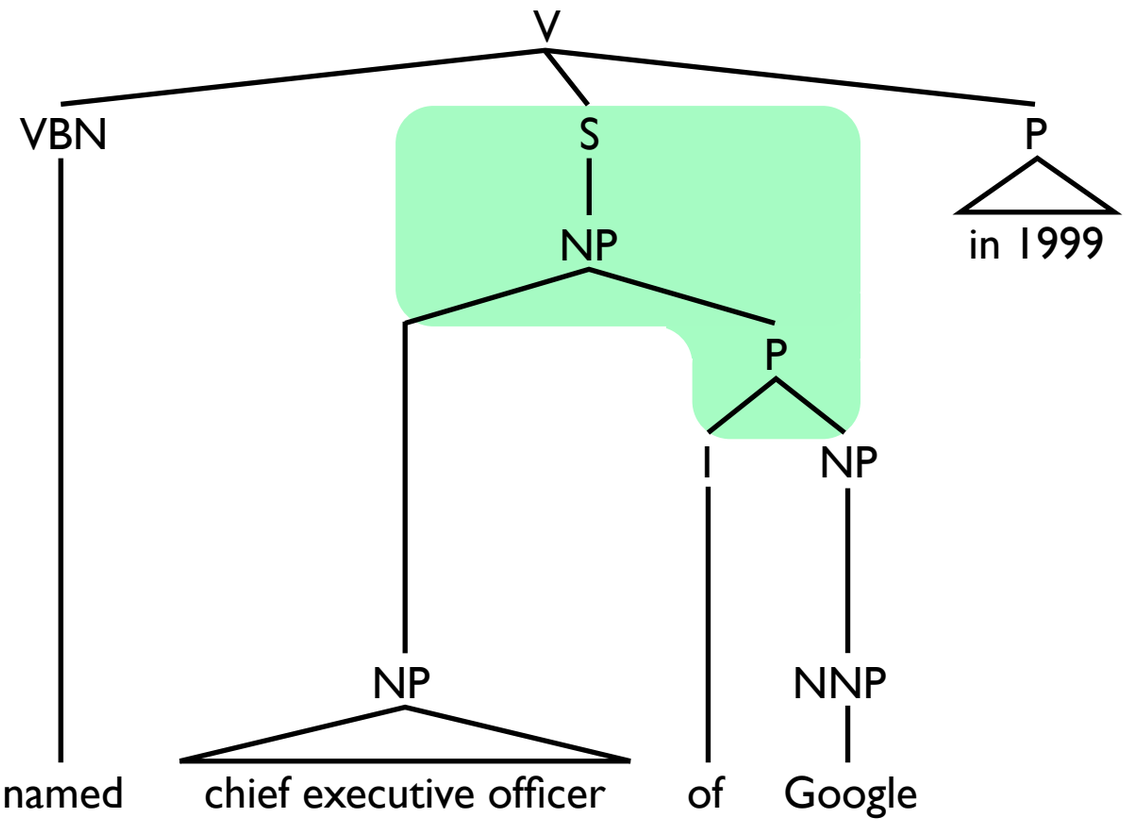

V
VBN
S
NP
P
NP
I
NP
NP
NNP
P
named
chief executive officer
of
Google
in 1999


V
VBN
S
P
NP
in 1999
P
I
NP
NP
NNP
named
chief executive officer
of
Google

Categorise

Transformation: Move PP, delete NP

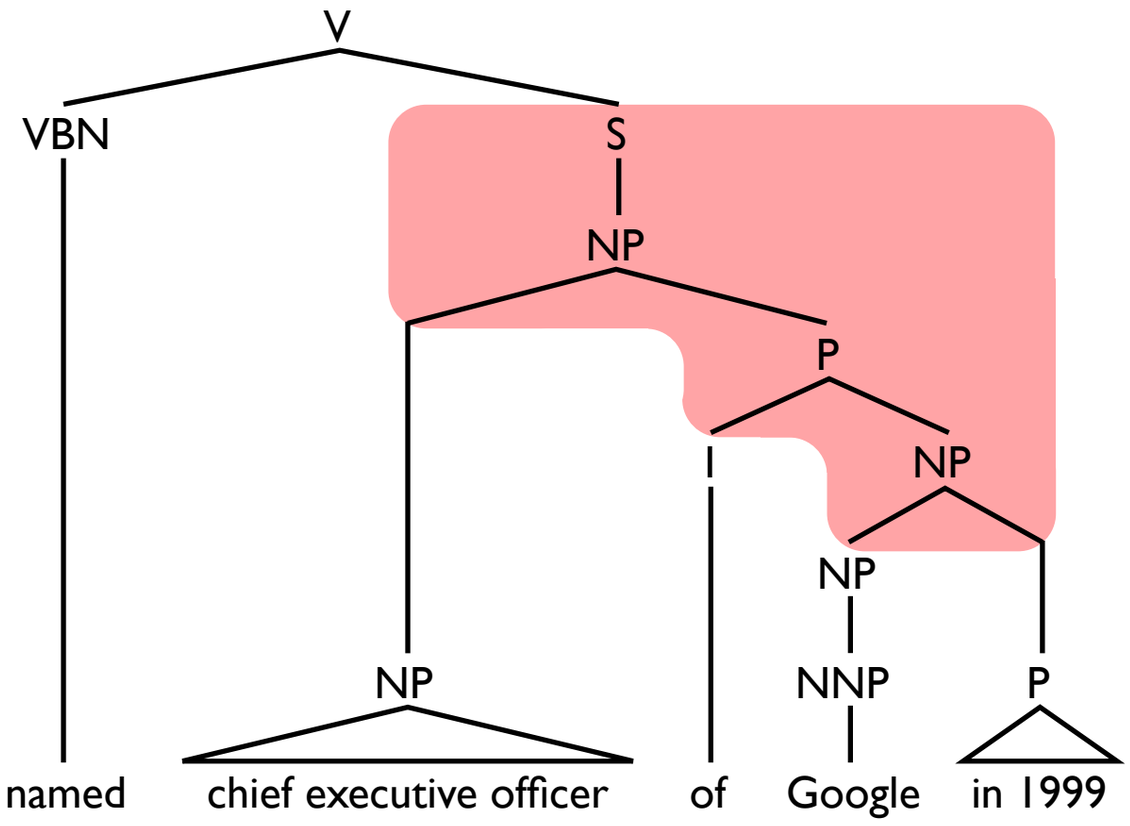


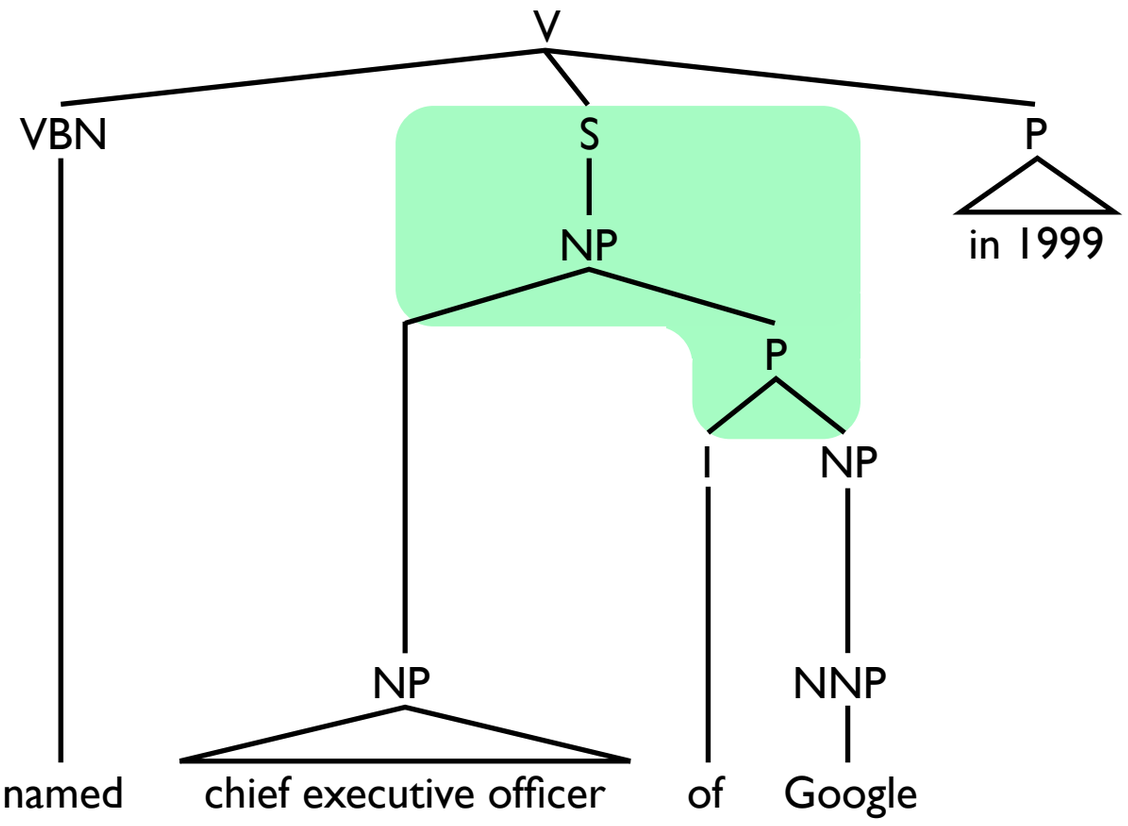


Categorise

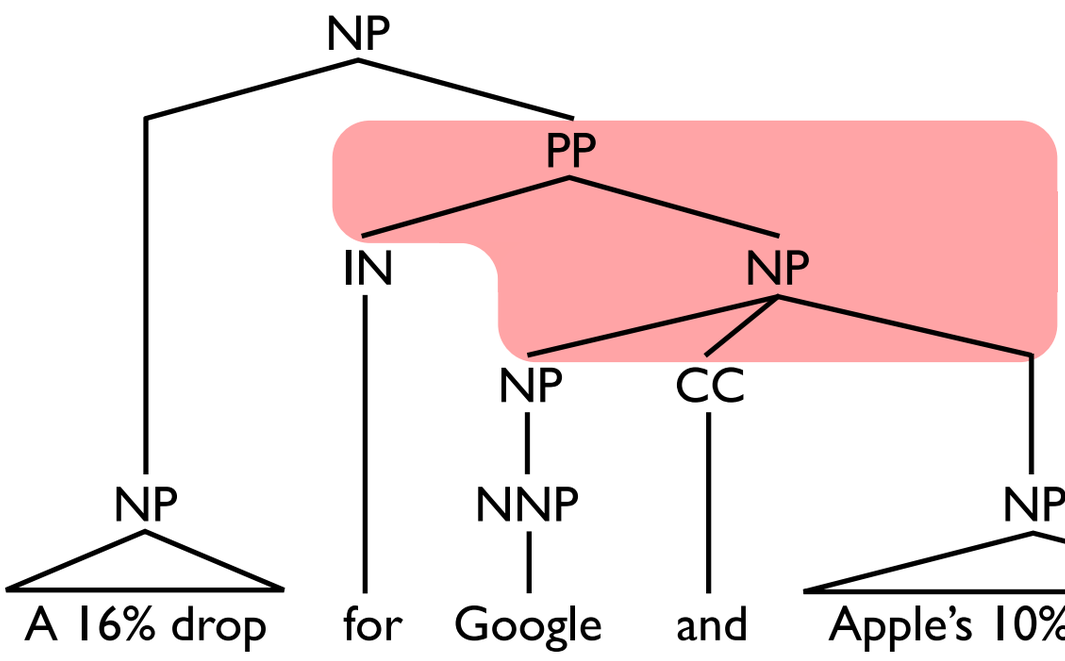
Transformation: Move PP, delete NP

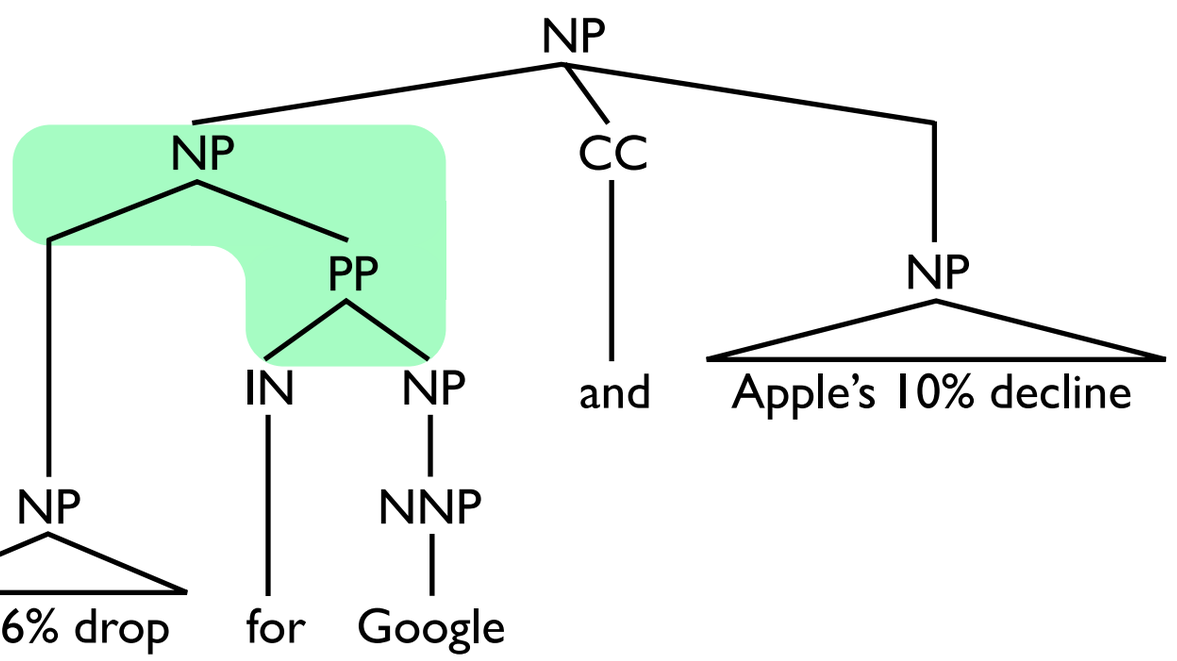
Underlying Error: PP Attachment





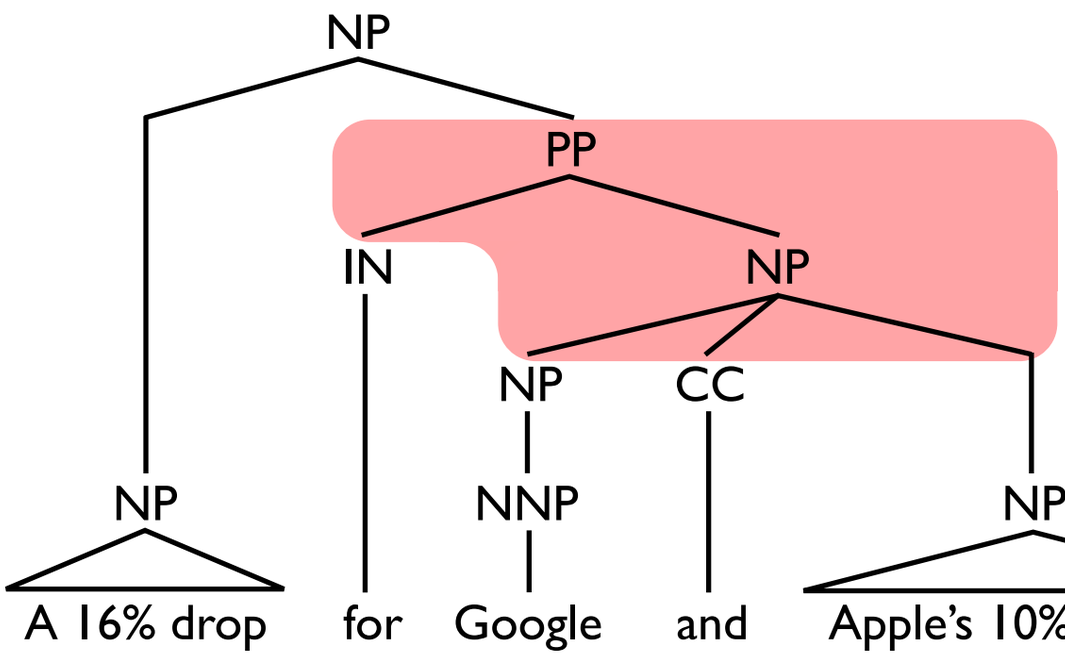
Categorise

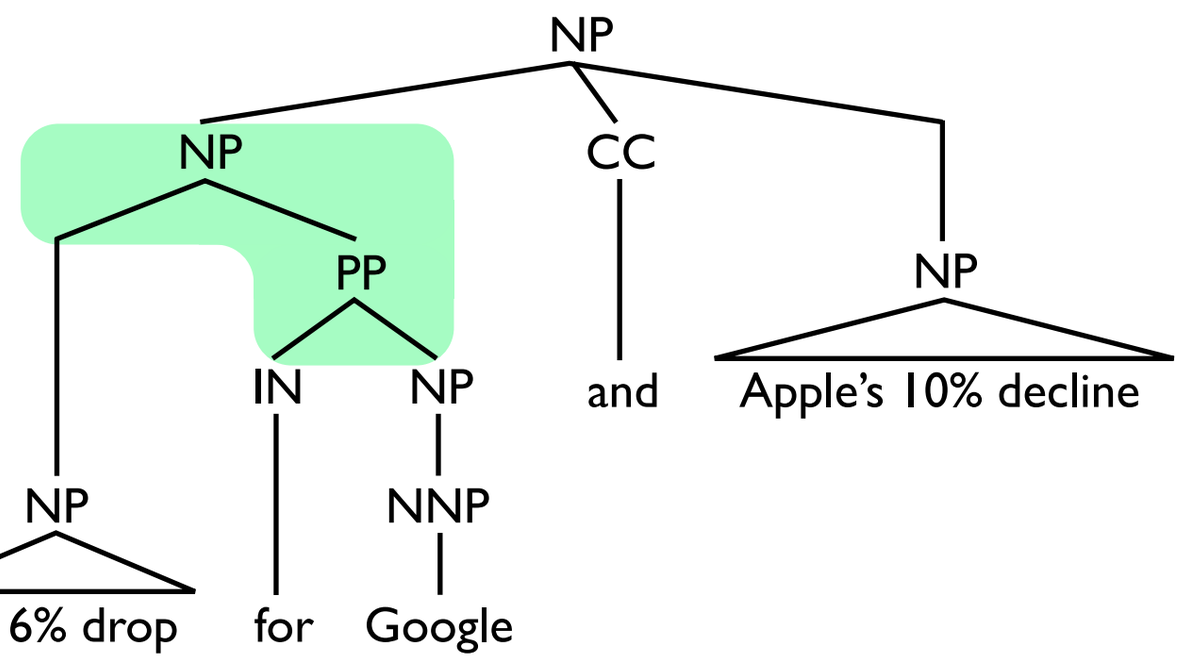

NP
PP
IN
NP
NP
CC
NP
NNP
NP
A 16% drop
for
Google
and
Apple's 10% decline


NP
NP
CC
NP
PP
IN
NP
and
Apple's 10% decline
NP
NNP
A 16% drop
for
Google

Categorise

Transformation: Move CC, move NP, delete NP, create NP

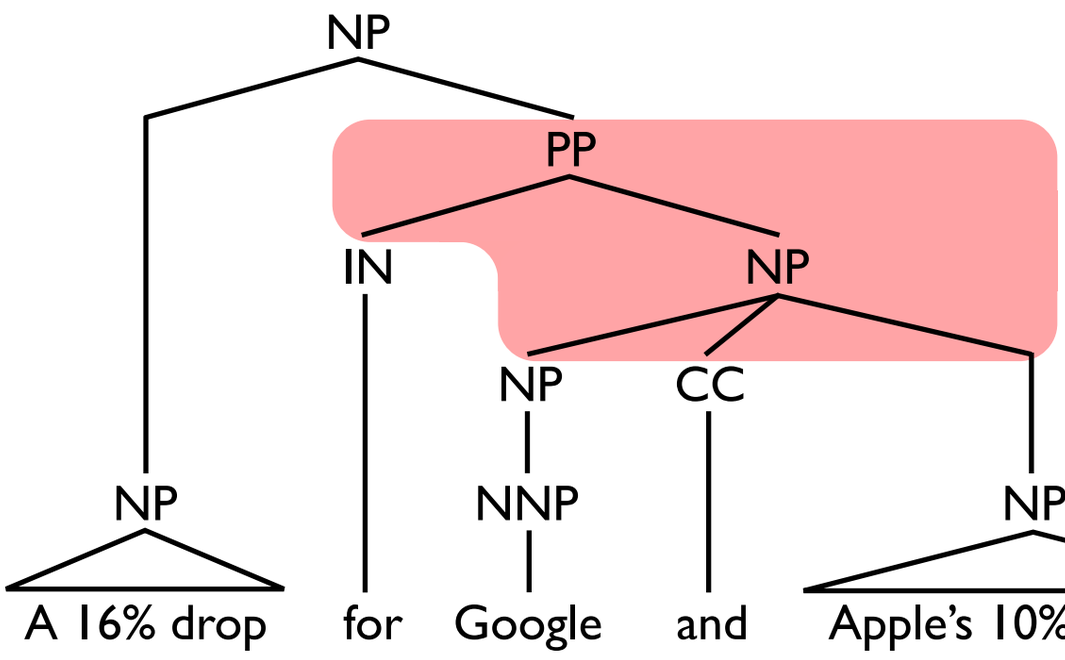


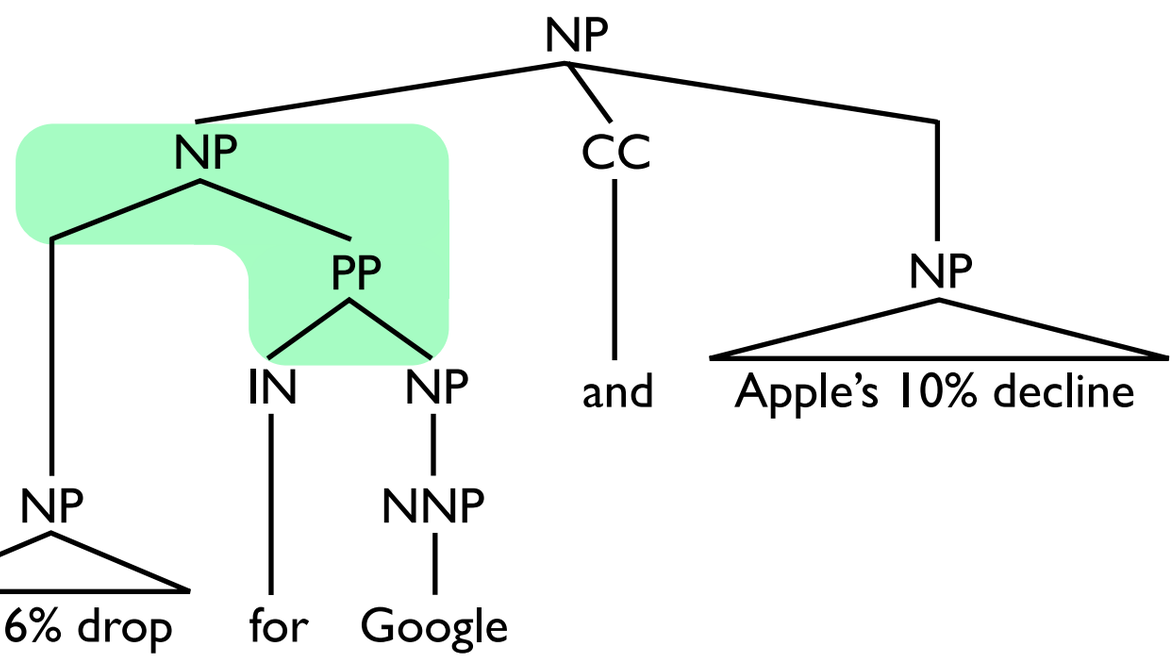


Categorise

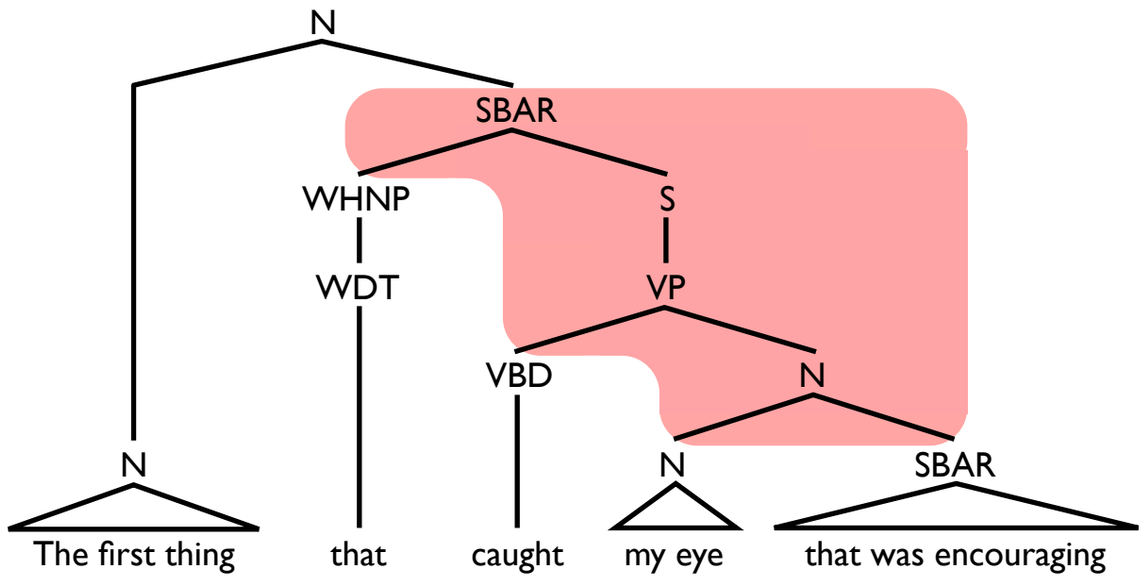
Transformation: Move CC, move NP, delete NP, create NP

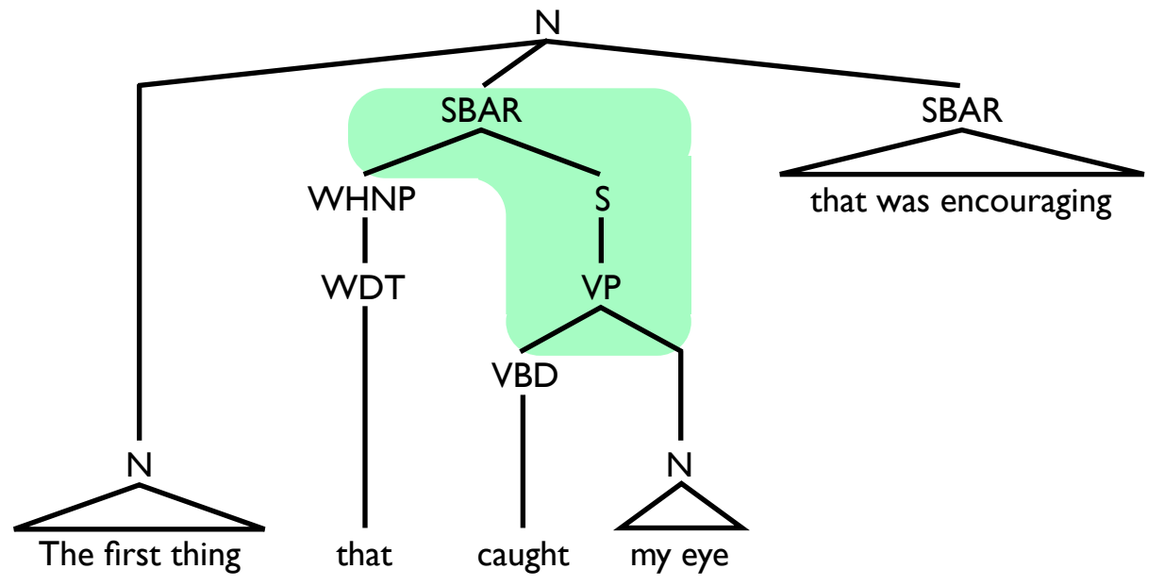
Underlying Error: Coordination





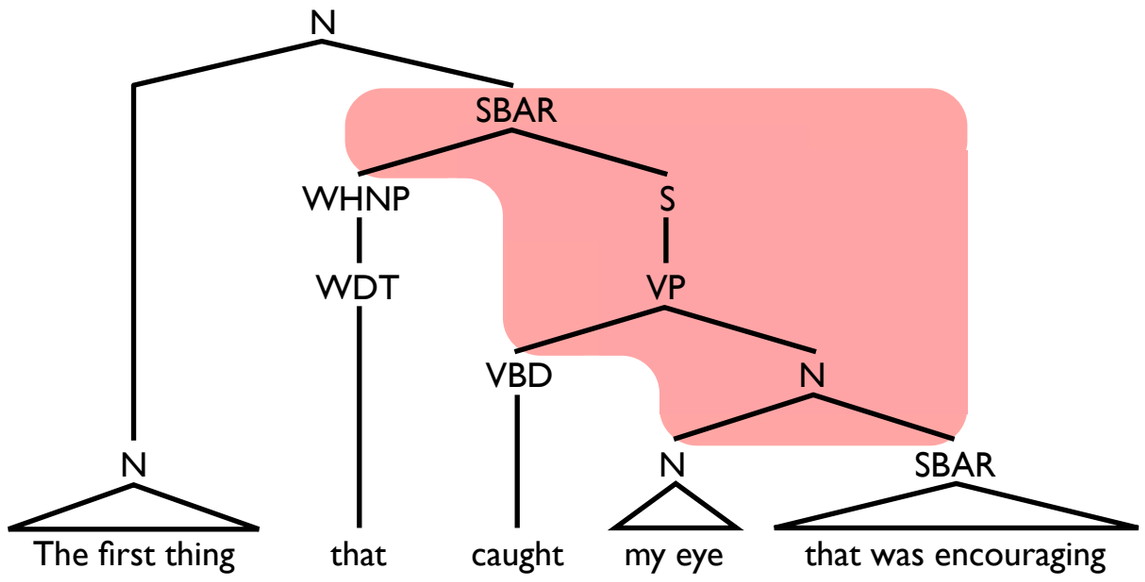
Categorise

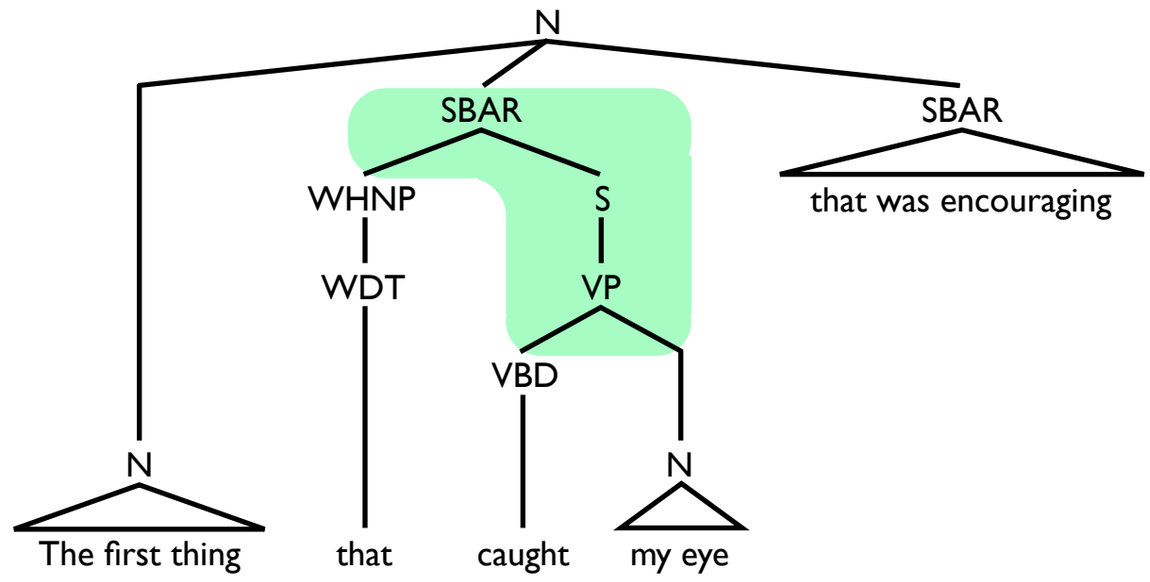

N
SBAR
WHNP
S
WDT
VP
VBD
N
N
SBAR
The first thing
that
caught
my eye
that was encouraging


N
SBAR
SBAR
WHNP
S
that was encouraging
WDT
VP
VBD
N
N
The first thing
that
caught
my eye

Categorise

Transformation: Move SBAR, delete NP

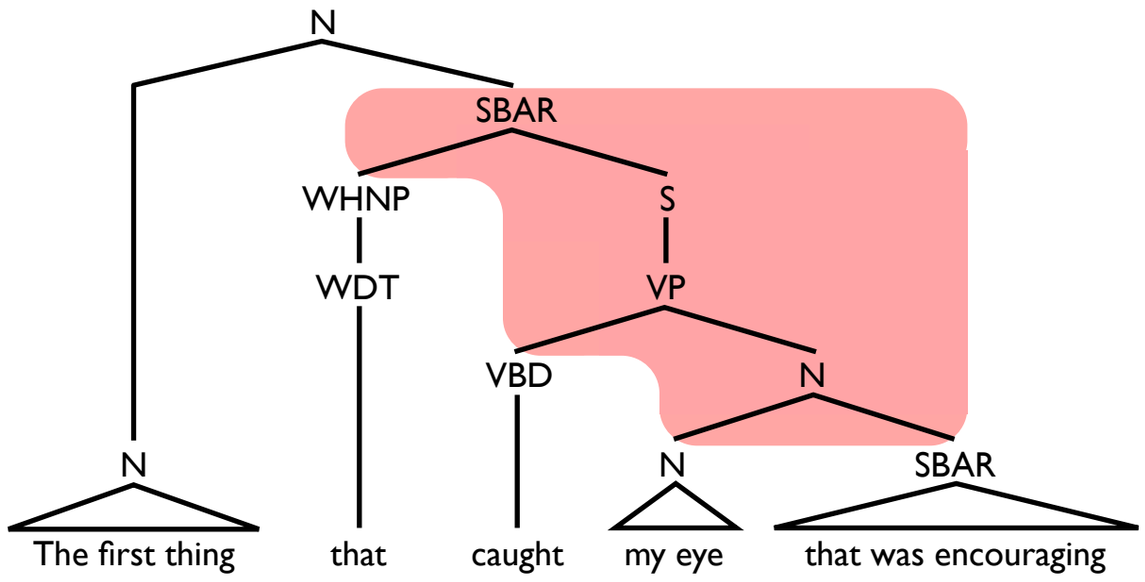


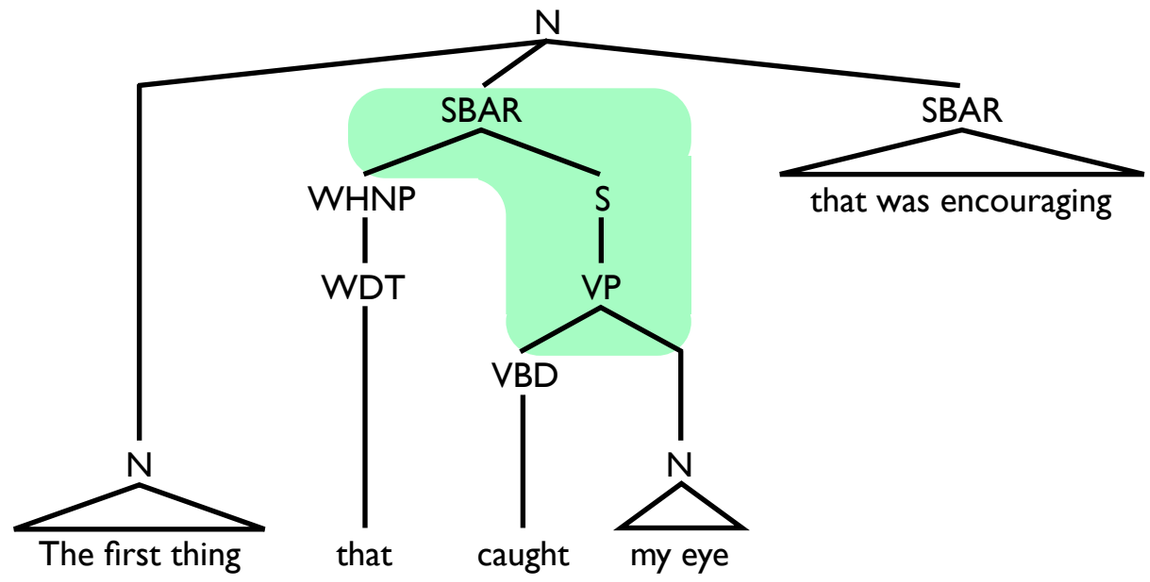


Categorise

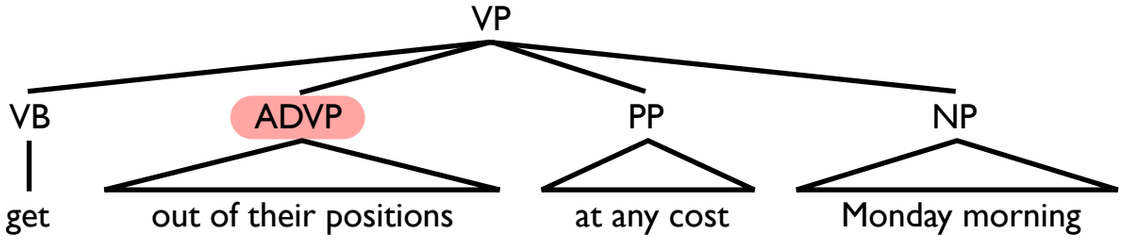
Transformation: Move SBAR, delete NP

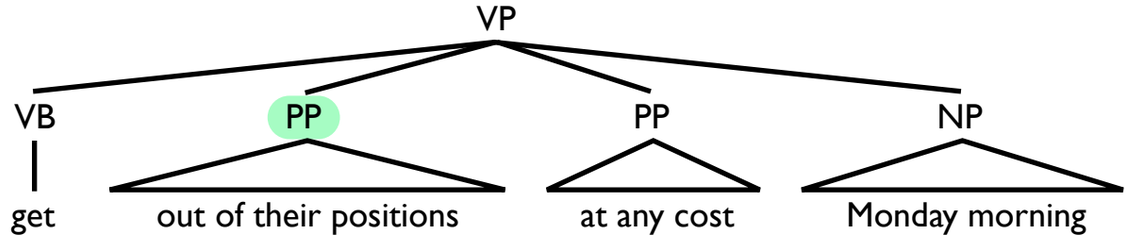
Underlying Error: Clause Attachment





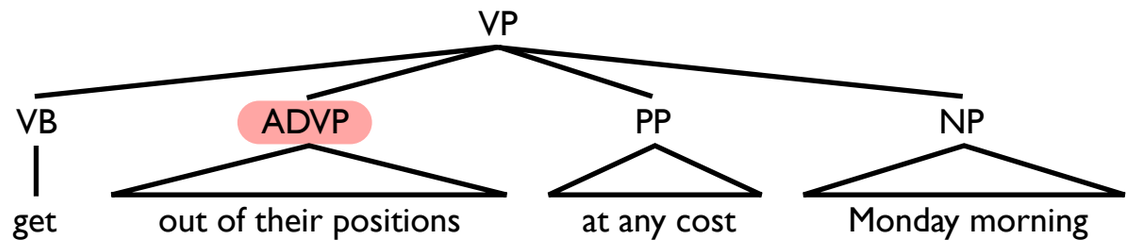
Categorise

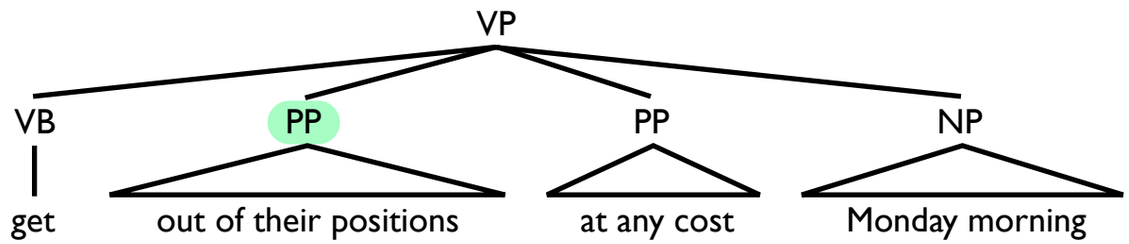

VP
VB
ADVP
PP
NP
get
out of their positions
at any cost
Monday morning


VP
VB
PP
PP
NP
get
out of their positions
at any cost
Monday morning

Categorise

Transformation: Delete ADVP, create PP

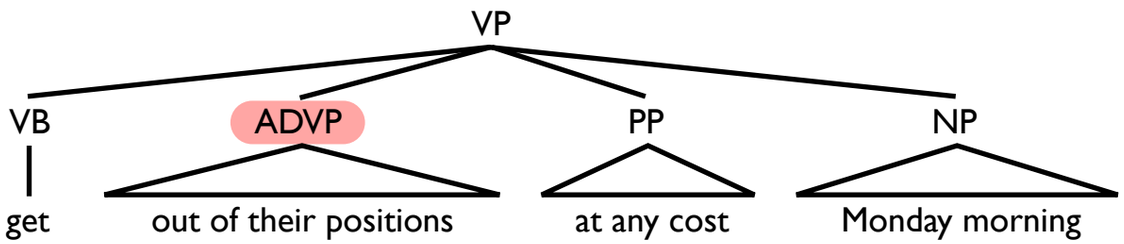


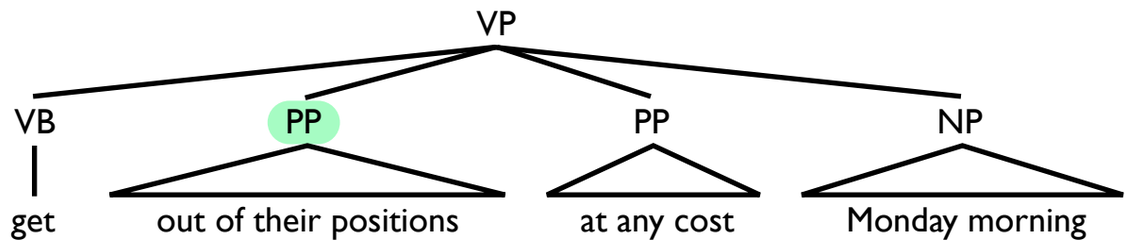


Categorise

Transformation: Delete ADVP, create PP

Underlying Error: Different Label





Related Work

Related Work

Intrinsic Evaluation

- Collins (2003) — Studied surface errors for a single parser
- Hara et al. (2007), Yu et al. (2011) — Manually analysed underlying errors

Related Work

Intrinsic Evaluation

- Collins (2003) — Studied surface errors for a single parser
- Hara et al. (2007), Yu et al. (2011) — Manually analysed underlying errors

Dependencies

- McDonald and Nivre (2007) — Investigated the impact of various factors
- Nivre et al. (2010) — Designed data to test specific types of constructions

WSJ Top Errors

WSJ Top Errors

Parser	F-score	PP Attach	Clause Attach	Diff Label	Co-ord

Min errors:

Max errors:

WSJ Top Errors

Parser	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Charniak++	92.07				
Berkeley	90.06				
Collins	87.62				
Stanford Fact.	86.42				
Stanford Unlex.	85.78				

WSJ Top Errors

Parser	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Min Errors		0.60			
Charniak++	92.07				
Berkeley	90.06				
Collins	87.62				
Stanford Fact.	86.42				
Stanford Unlex.	85.78				
Max Errors		1.12			

Min errors:

Max errors:

WSJ Top Errors

Parser	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Min Errors		0.60	0.38	0.31	0.23
Charniak++	92.07				
Berkeley	90.06				
Collins	87.62				
Stanford Fact.	86.42				
Stanford Unlex.	85.78				
Max Errors		1.12	0.61	0.51	0.40

WSJ Top Errors

Parser	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Min Errors		0.60	0.38	0.31	0.23
Charniak++	92.07				
Berkeley	90.06				
Collins	87.62				
Stanford Fact.	86.42				
Stanford Unlex.	85.78				
Max Errors		1.12	0.61	0.51	0.40

WSJ Top Errors

Parser	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Min Errors		0.60	0.38	0.31	0.23
Charniak++	92.07				
Berkeley	90.06				
Collins	87.62				
Stanford Fact.	86.42				
Stanford Unlex.	85.78				
Max Errors		1.12	0.61	0.51	0.40

Reranking

Reranking

System	K	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Charniak++	1	91.02				
Max Errors			0.66	0.43	0.33	0.26

Reranking

System	K	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Charniak++	50	92.07				
Charniak++	1	91.02				
Max Errors			0.66	0.43	0.33	0.26

Reranking

System	K	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Min Errors			0.16	0.09	0.14	0.07
Oracle	50	97.18				
Charniak++	50	92.07				
Charniak++	1	91.02				
Max Errors			0.66	0.43	0.33	0.26

Out of Domain

Out of Domain

Corpus	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Best		0.022	0.016	0.013	0.010
WSJ 23	92.07				
Popular	85.91				
Mystery	83.95				
Adventure	85.20				
Humour	83.60				
G-Web Blogs	84.15				
G-Web Email	81.18				
Worst		0.040	0.035	0.053	0.023

Out of Domain

Corpus	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Best		0.022	0.016	0.013	0.010
WSJ 23	92.07				
Popular	85.91				
Mystery	83.95				
Adventure	85.20				
Humour	83.60				
G-Web Blogs	84.15				
G-Web Email	81.18				
Worst		0.040	0.035	0.053	0.023

Min errors:

Max errors:

Out of Domain

Corpus	F-score	PP Attach	Clause Attach	Diff Label	Co-ord
Best		0.022	0.016	0.013	0.010
WSJ 23	92.07				
Popular	85.91				
Mystery	83.95				
Adventure	85.20				
Humour	83.60				
G-Web Blogs	84.15				
G-Web Email	81.18				
Worst		0.040	0.035	0.053	0.023

Conclusion

Conclusion

- An argument for more detailed analysis

Conclusion

- An argument for more detailed analysis
- A new tool for investigating parser behaviour: nlp.cs.berkeley.edu/software

Conclusion

- An argument for more detailed analysis
- A new tool for investigating parser behaviour: nlp.cs.berkeley.edu/software
- A deep and broad examination of parsers

Conclusion

- An argument for more detailed analysis
- A new tool for investigating parser behaviour: nlp.cs.berkeley.edu/software
- A deep and broad examination of parsers

Thank you!