Psycholinguistics: Sentence Processing III

9.59; 24.905

February 24, 2005

Ted Gibson

Today's lecture

1. Information sources used in sentence comprehension.

2. Modularity of information use in sentence comprehension? The sentence processor is probably not modular.

3. Syntactic information use in sentence processing: Locality of syntactic integrations: The dependency locality theory.

What sources of information do people use in processing sentences?

- Syntactic structure
- Word frequency
- Plausibility

  (1) The dog bit the man.

  (2) The man bit the dog.
- Discourse context
- Syntactic complexity
- Intonational information

Information that is used in sentence comprehension

3. Plausibility of the resulting linguistic expression, in the world

Unambiguous examples:
The dog bit the boy. vs. The boy bit the dog.

Ambiguity: (Trueswell, Tanenhaus & Garnsey, 1994)
The defendant examined by the lawyer turned out to be unreliable.
The evidence examined by the lawyer turned out to be unreliable.

Information that is used in sentence comprehension

4. Context (Crain & Steedman, 1985; Altmann & Steedman, 1988; Tanenhaus et al., 1995)

Ambiguity:

There were two defendants, one of whom the lawyer ignored entirely, and the other of whom the lawyer interrogated for two hours.

The defendant examined by the lawyer turned out to be unreliable.

Monitoring visual eye-movements while listening to spoken instructions
“Put the frog on the napkin into the box.”

Photo removed for copyright reasons.

Monitoring visual eye-movements while listening to spoken instructions

"Put the frog on the napkin into the box."

Two frog context: No looks to the incorrect target (the second napkin)

Photo removed for copyright reasons.

One frog context: Many looks to the incorrect target (the second napkin)

Syntactic information use in sentence processing: The Dependency Locality Theory (DLT, Gibson, 1998, 2000)

Resources are required for two aspects of language comprehension:

(a) Integration: connecting the current word into the structure built thus far;

(b) Storage: Predicting categories to complete the current structure.

More on this later in the lecture!

Open question: The modularity of information use in language processing

The time course according to which different information sources become available:

Syntactic information first?

Lexical information first?

All information sources available simultaneously?

Two kinds of modularity

- **Modularity of information**: Different information sources may be computed using separate systems. E.g., syntactic information may be computed using a separate system from plausibility or contextual information

- **Modularity of the time course of information use**: Some information may become available before other information. In particular, syntactic information may be available before other kinds of information (Frazier, 1978).

An early hypothesis regarding ambiguity resolution: The "garden-path theory": Minimal Attachment and Late Closure

Frazier's (1978) hypotheses:

(**Note: These are early hypotheses: All are highly debatable now!)**

1. The sentence processor is **serial**, retaining exactly one representation at each parser state.

2. The sentence processor is **modular**, using syntactic information before it uses other information in resolving ambiguity.

3. The particular syntactic ambiguity resolution heuristics that the parser uses are **Minimal Attachment** and **Late Closure**.

Syntactic ambiguity resolution heuristics

- Early heuristics: Minimal Attachment and Late Closure
- These are now superceded by the dependency locality theory (DLT): Syntactic Storage and Syntactic Integration

  More on these later in this class.

Syntactic ambiguity resolution heuristics

Minimal Attachment: Attach incoming material into the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language.

Argument and specifier attachments: all nodes are already present when attachments are being considered (under X-bar).

Modifier attachments: Need to construct additional nodes (under X-bar). Thus argument attachments are generally preferred over modifier attachments

Examples of Minimal Attachment preferences

PP attachment:

I put the candy **on the table** into my mouth.

CP attachment:

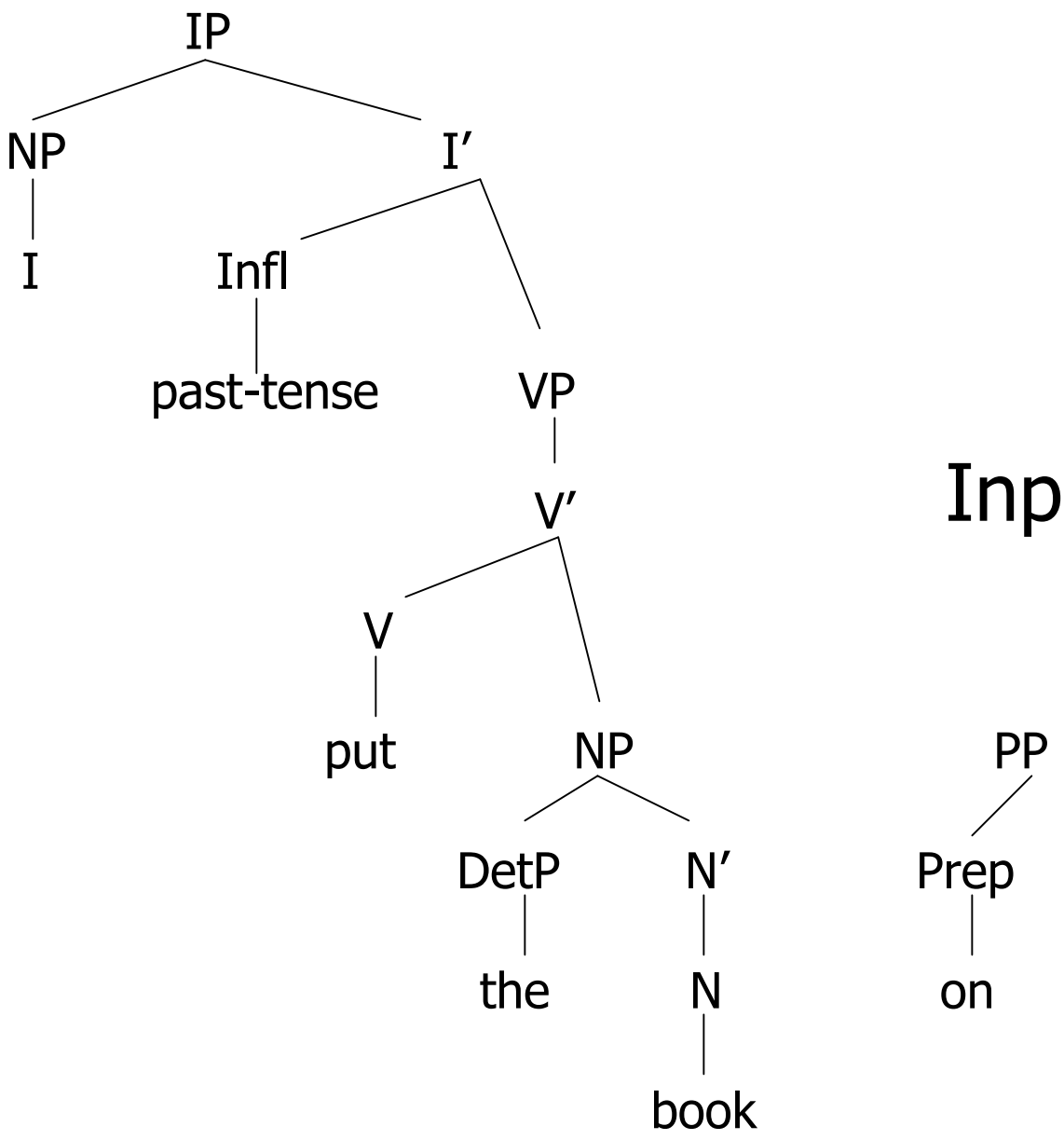
The psychologist convinced the patient **that he was having trouble with** to leave.

Main verb (MV) / Reduced relative (RR):

The dog **walked** to the park chewed the bone.

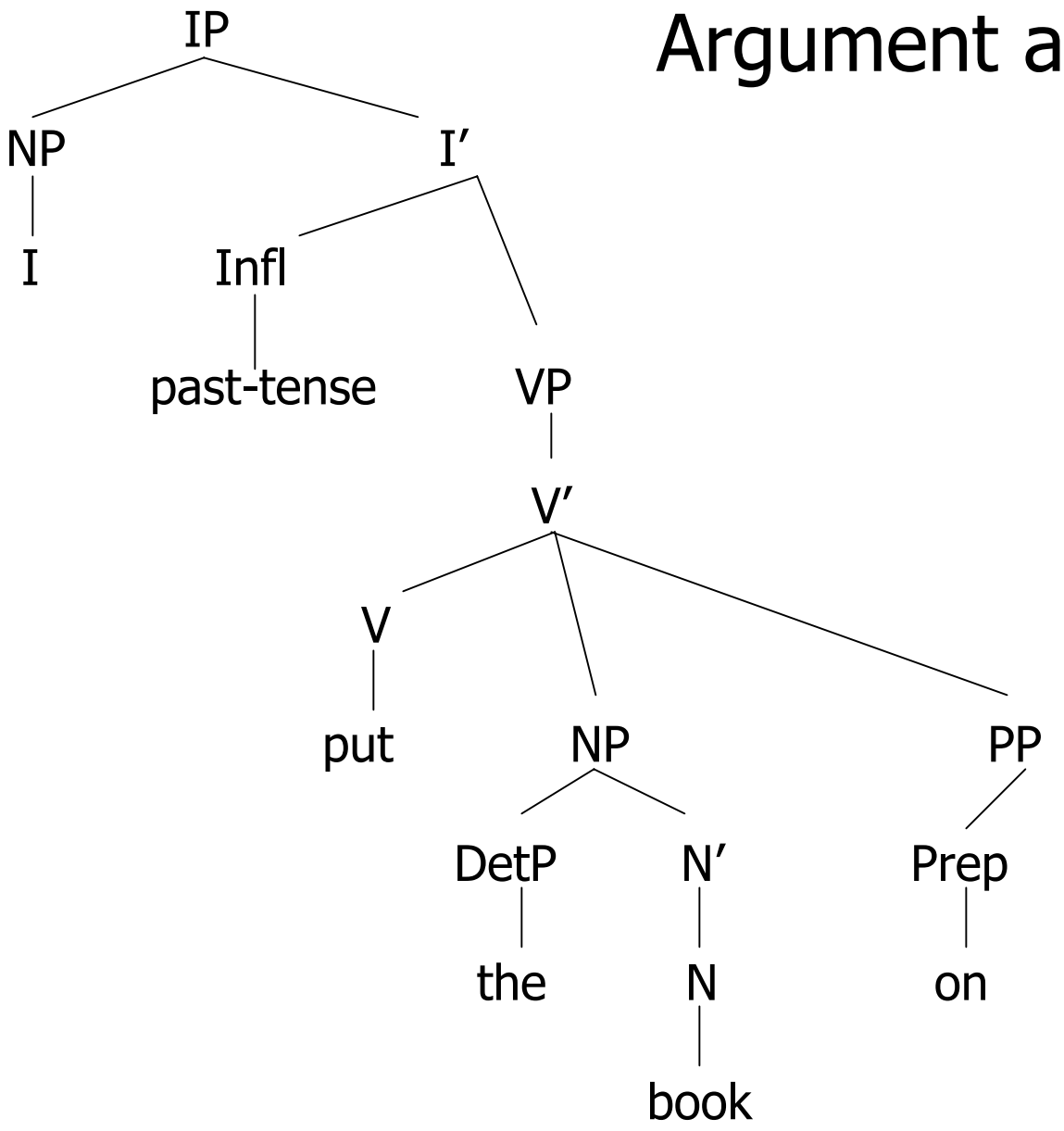
The horse **raced** past the barn fell.

?# The defendant **examined** by the lawyer turned out to be unreliable.

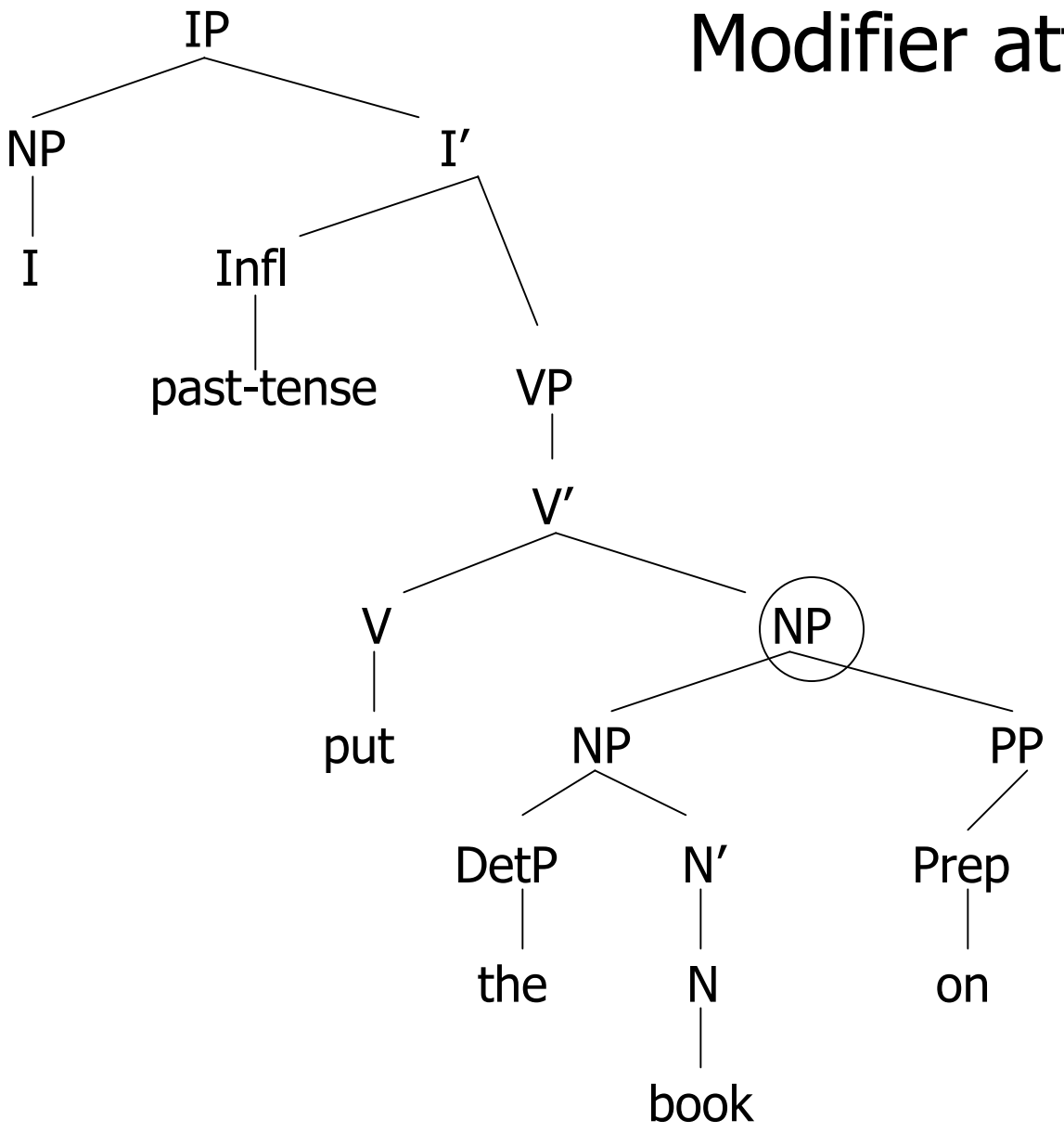


Input word: “on”

Argument attachment of “on”

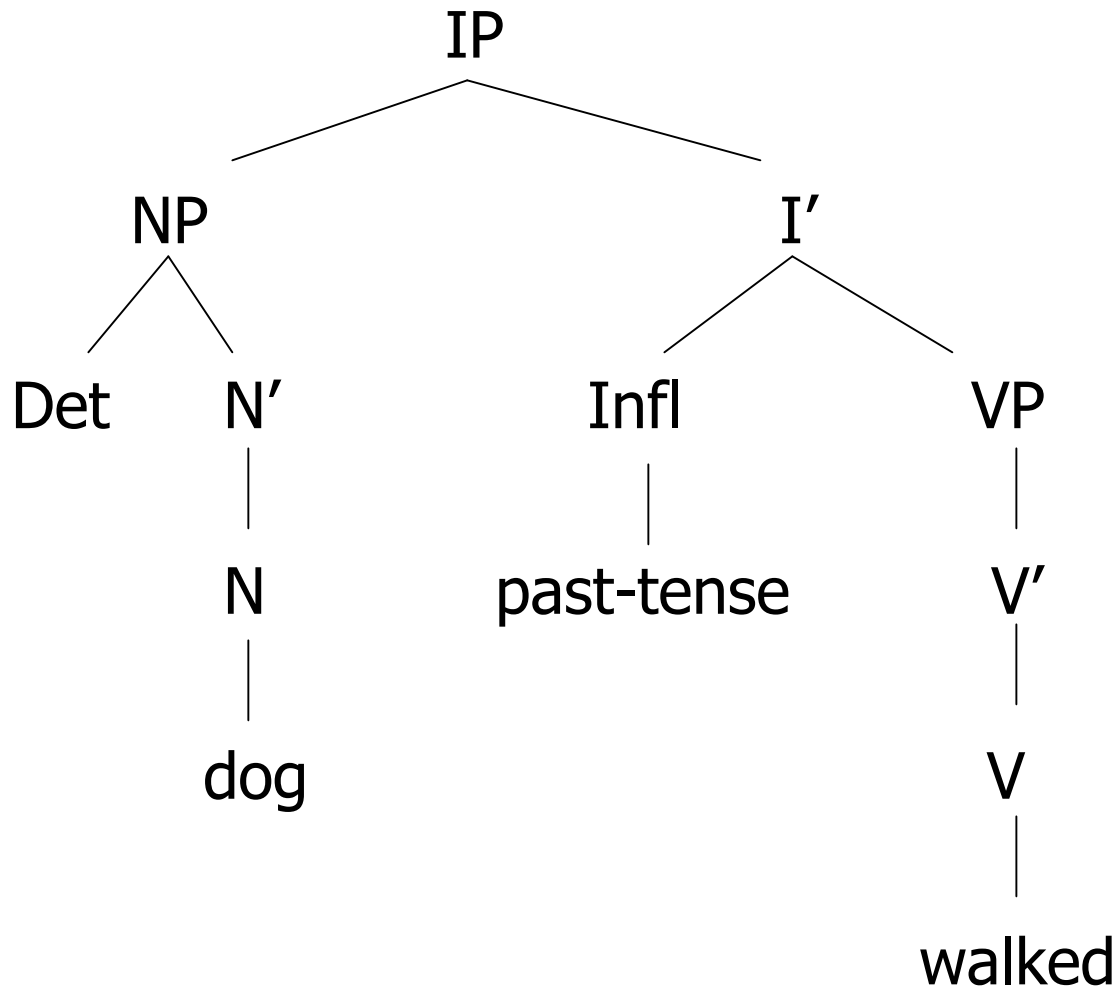


Modifier attachment of "on"

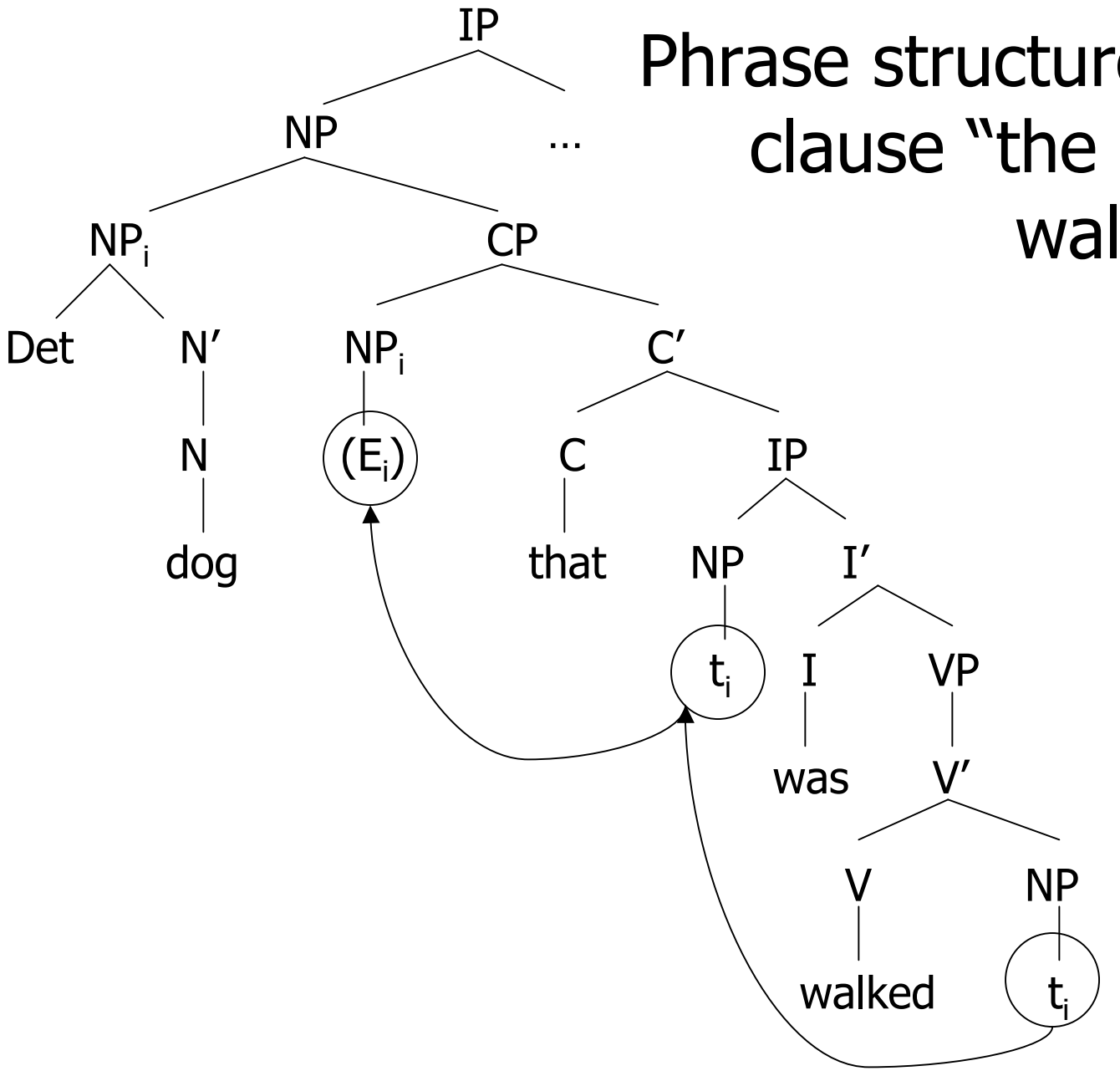


Extra node in the tree!

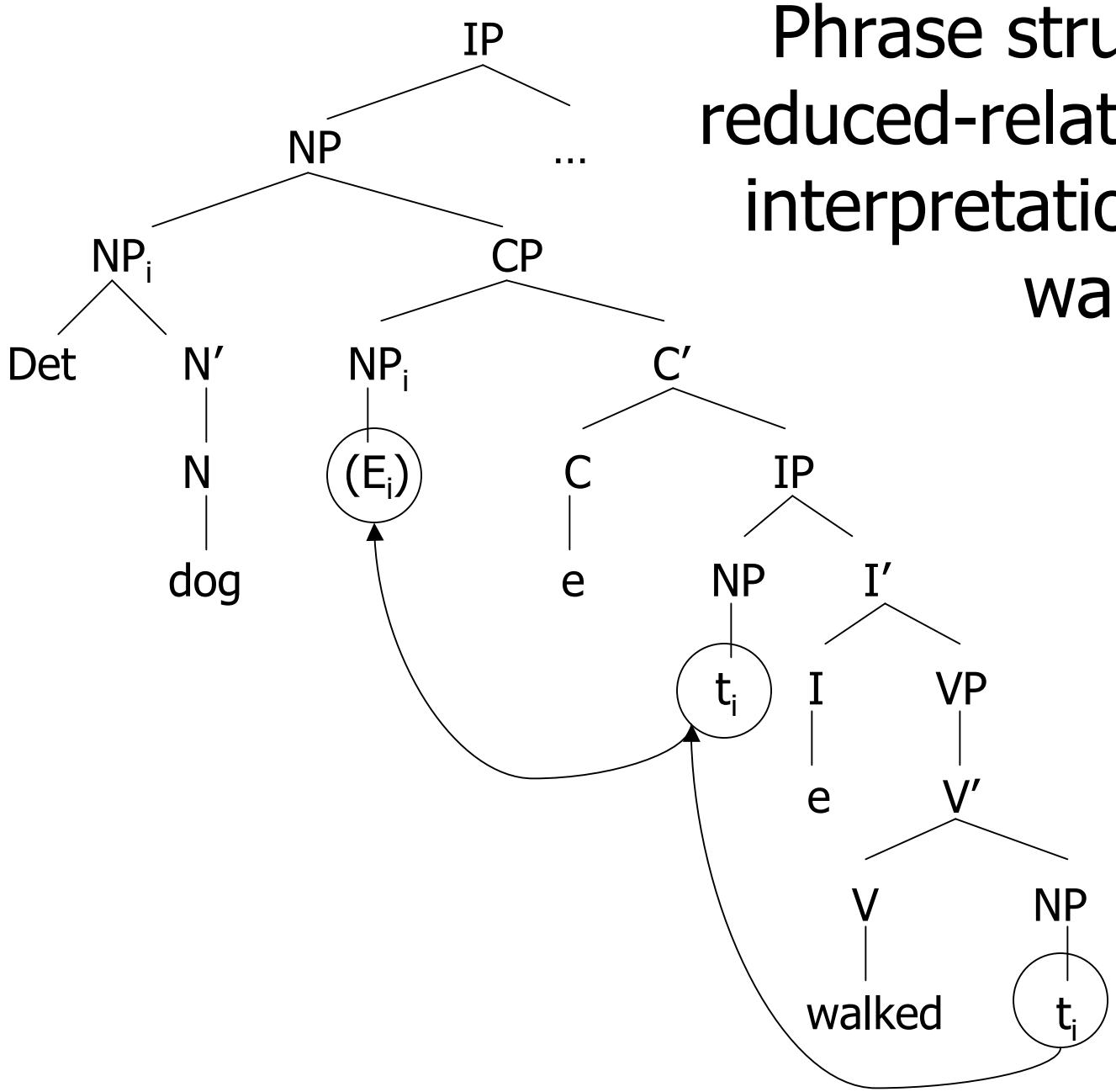
Phrase structure for the main-verb (MV) interpretation of "the dog walked"



Phrase structure for the relative clause "the dog that was walked"



Phrase structure for the reduced-relative clause (RR) interpretation of "the dog walked"



The non-modularity of language processing

Research question: Does syntactic structure processing take place before other levels of sentence processing?

Framed in terms of modularity: is syntactic processing modular, so that it is insulated from other levels of analysis, such as real-world plausibility?

Ferreira & Clifton (1986)

Eye-tracking investigation of MV/RR, manipulating the plausibility of the initial NP as agent of the MV:

The (evidence / defendant) examined by the lawyer turned out to be unreliable.

Main clause is syntactically preferred, but this interpretation is implausible for "the evidence".

Unambiguous controls:

The (evidence / defendant) that was examined by the lawyer turned out to be unreliable

Ferreira & Clifton (1986)

The evidence examined by the lawyer turned out to be unreliable.

Modularity predictions:

1. slow for syntactically ambiguous item at “examined”: The parser will notice that the structure that it has selected is implausible.
2. slow at "by the lawyer": syntactic reanalysis.

Non-modularity predictions: no difference between ambiguous and unambiguous controls in any region.

Ferreira & Clifton (1986)

Results: First pass times (msec/character)

	examined	by the lawyer
Animate ambig.	33.3	40.4
Animate unambig	31.9	30.7
Inanimate ambig	37.7	38.4
Inanimate unambig	30.1	30.3

These results support the modularity theory.

Trueswell, Tanenhaus & Garnsey, 1994

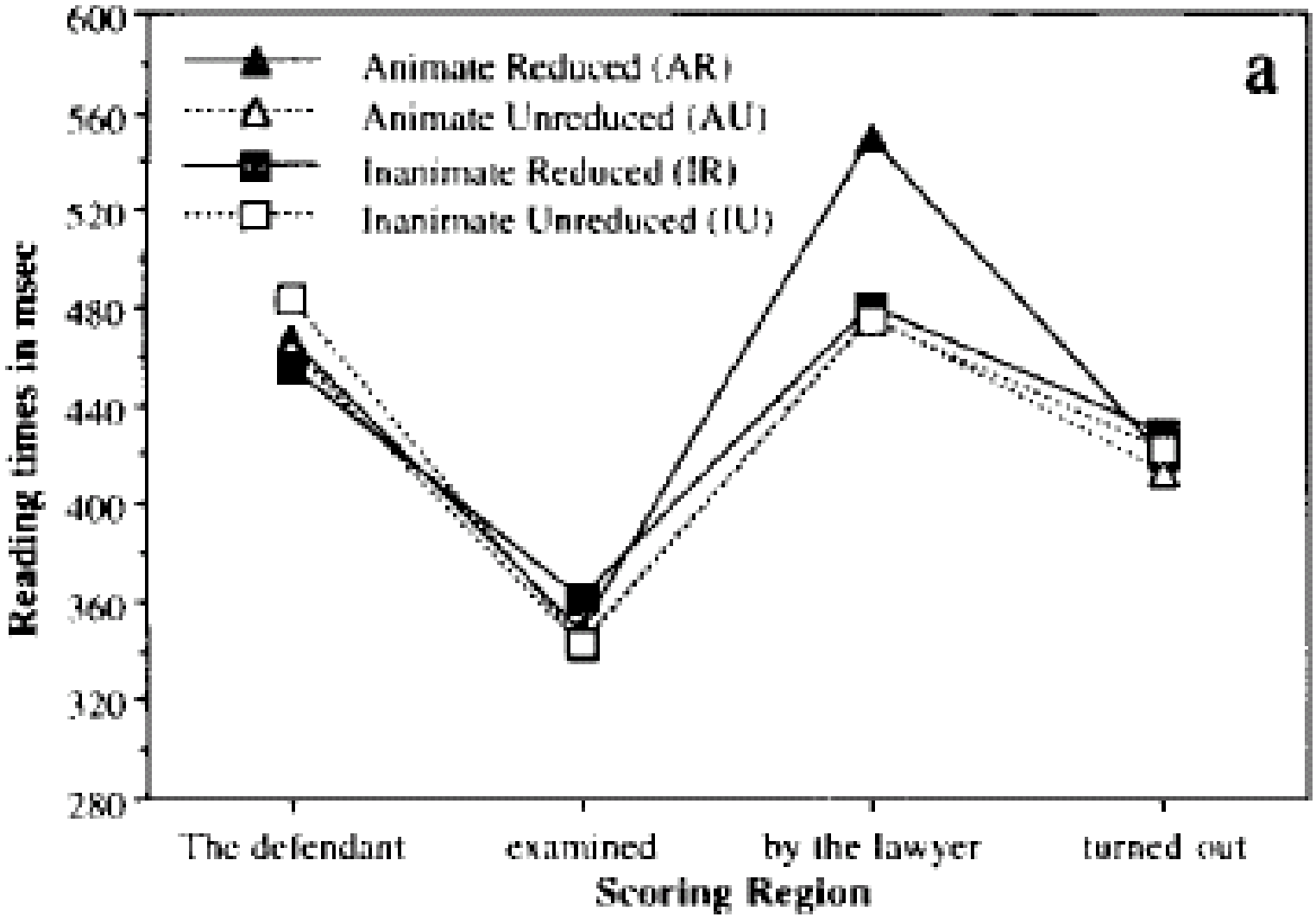
Problems in Ferreira & Clifton's items:

Half (8/16) of the inanimate items weren't implausible agents:

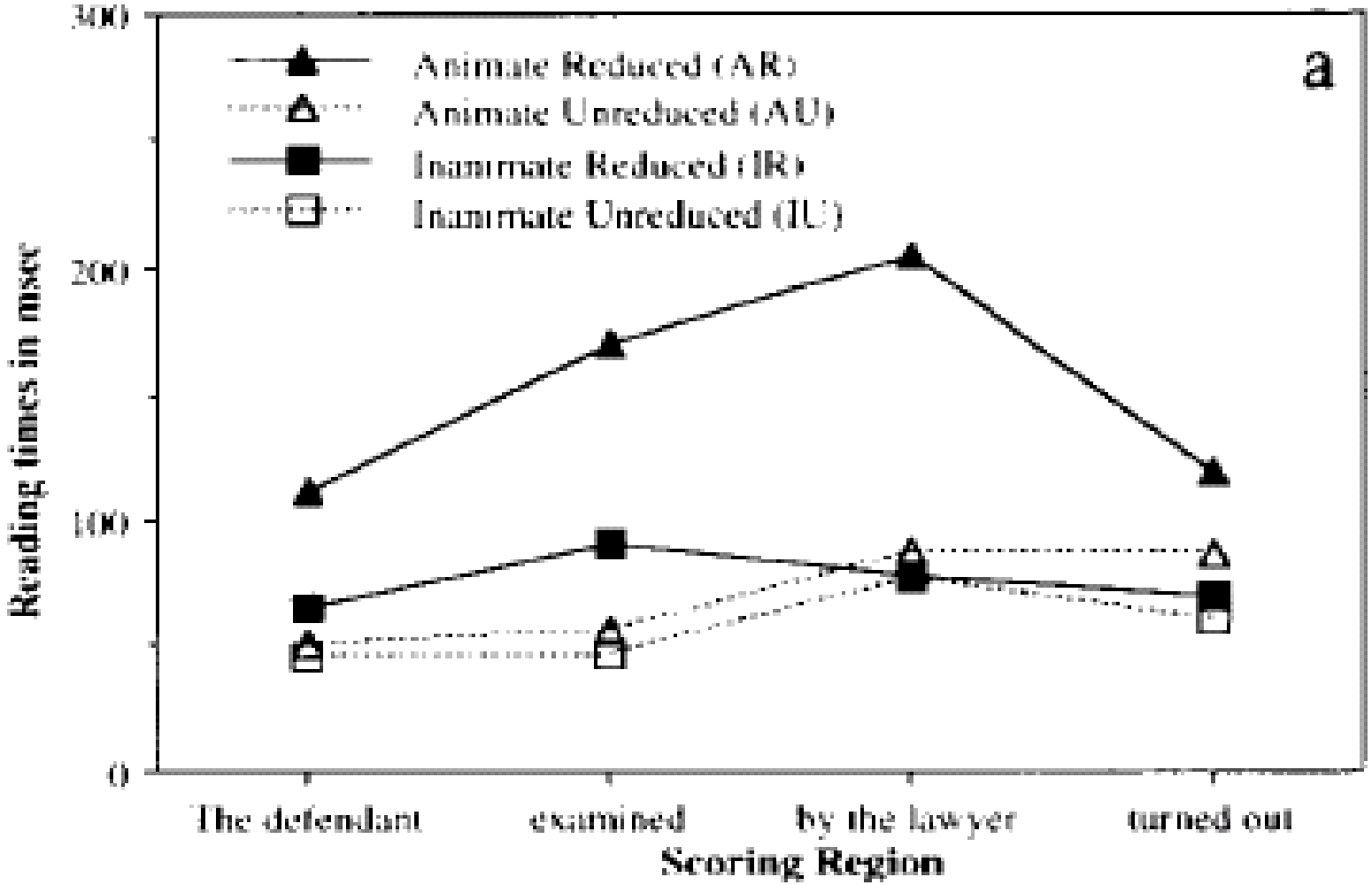
The car towed by the truck …

(cf. The car towed the trailer.)

Trueswell, Tanenhaus, and Garnsey (1994): Re-do experiment with better items.
Mean first pass times



Trueswell, Tanenhaus, and Garnsey (1994): mean second-pass times



Trueswell, Tanenhaus & Garnsey, 1994

Conclusion: Plausibility and lexical frequency are used as soon as can be measured in resolving ambiguity in on-line sentence processing.

This is evidence against the modularity hypothesis.

More evidence against the modularity hypothesis:

Visual world evidence of Tanenhaus et al. (1995): the presence of multiple elements (e.g. frogs) in the context biased people to follow the modifier reading.

Syntactic information use in sentence processing: The Dependency Locality Theory (DLT, Gibson, 1998, 2000)

Resources are required for two aspects of language comprehension:

(a) Integration: connecting the current word into the structure built thus far;

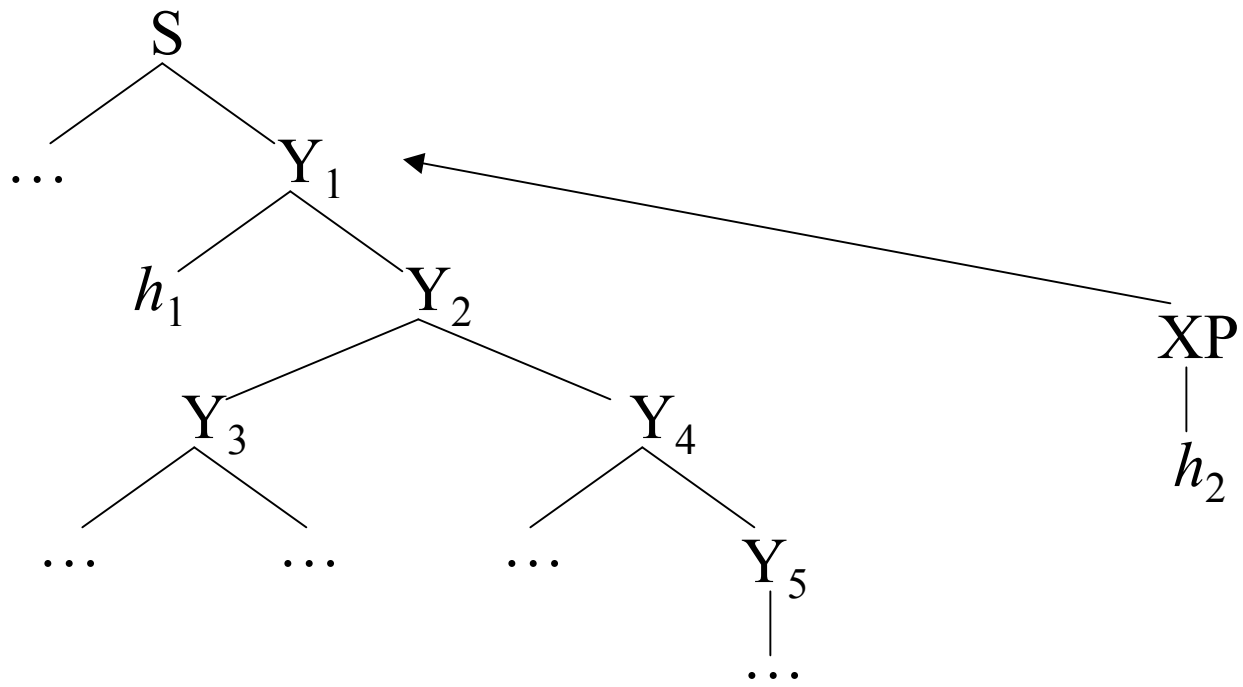
(b) Storage: Predicting categories to complete the current structure.

Syntactic resource hypotheses

- Frazier (1978): Minimal Attachment & Late Closure:
  - Ambiguity resolution only
  - Assumed to be modular: Applying before other sources of information use

- Gibson (1998; 2000): Syntactic storage and integration
  - Apply in both ambiguous and unambiguous sentences
  - Assumed to be non-modular: Interact immediately with other sources of information (but this is not a crucial part of the theory).

Integration complexity depends on the *distance* or *locality* between the head and dependent being integrated.



Computational motivation: Integrating h_2 to h_1 involves re-accessing h_1:

- h_1's activation may have **decayed** for all the integrations that have taken place since it was last highly activated.
- **Interference** of intervening elements: similar to h_1 and/or h_2

Locality effects in ambiguous structures

Right Association (Kimball, 1973); Late Closure (Frazier, 1979, 1987); Recency (Gibson, 1991)

(1) The bartender told the detective that the suspect
left the country yesterday.

Yesterday is preferred as modifying *left* rather than *told*

Experimental evidence: Frazier & Rayner, 1982;
Gibson et al., 1996; Altmann et al., 1998; Pearlmutter & Gibson, 2001.

Local attachment preference

Cartoon removed for copyright reasons.

Initial Integration Distance Hypothesis

Gibson (1998); Warren & Gibson (2002, *Cognition*)

The difficulty of integrating a new word h_2 to h_1 is proportional to the number of discourse objects and events (**nouns** and **verbs**, roughly) which were introduced since h_1 was last processed.

Locality effects in unambiguous structures

Gibson (1998, *Cognition*); Grodner & Gibson (in press, *Cognitive Science)*

Experiment 1:

Object-extracted relative clause:

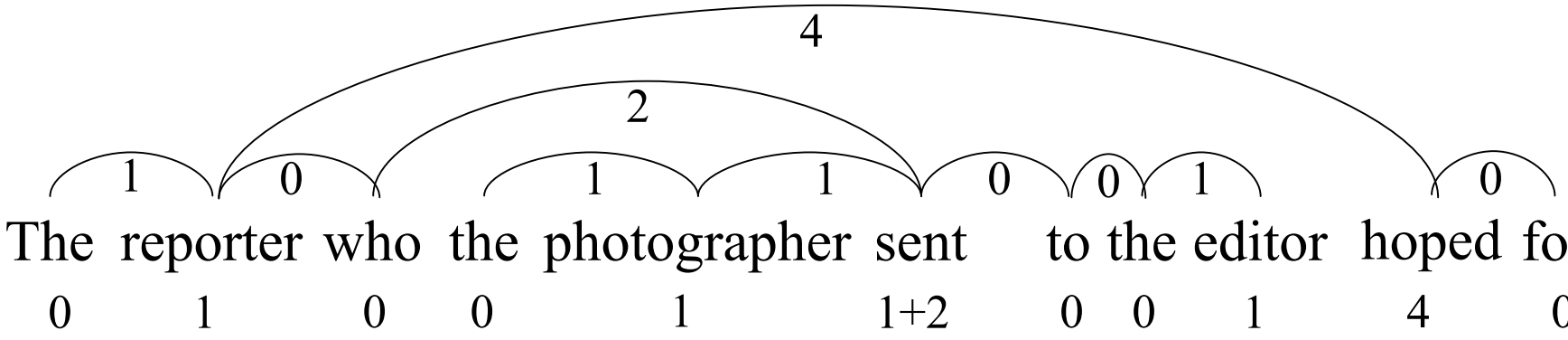
The reporter who the photographer sent to the editor hoped for a good story.

Subject-extracted relative clause:

The reporter who sent the photographer to the editor hoped for a good story.

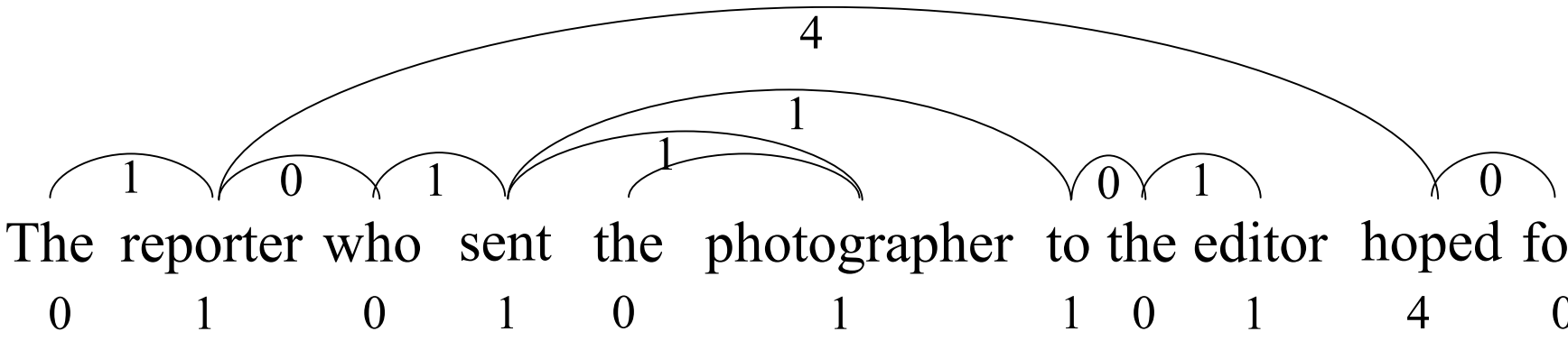
Locality effects in unambiguous structures

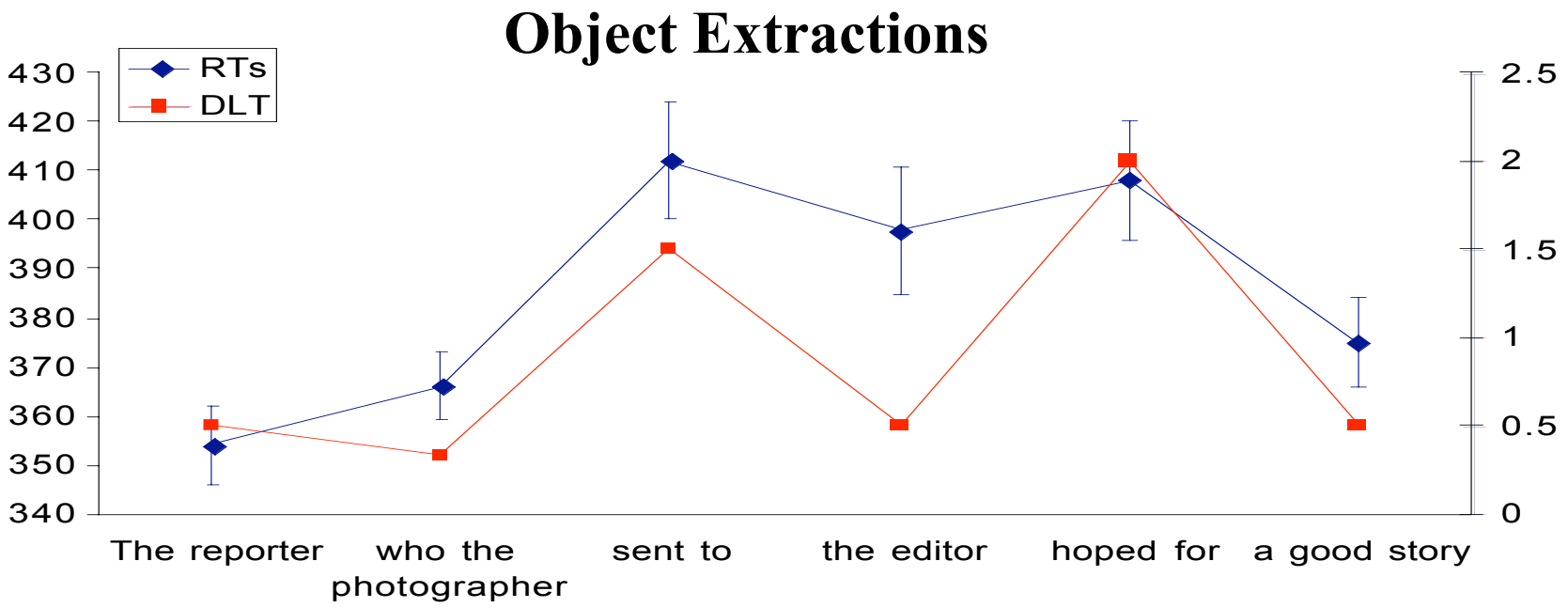
Object-extracted relative clause:
The reporter who the photographer sent to the editor hoped for a story.

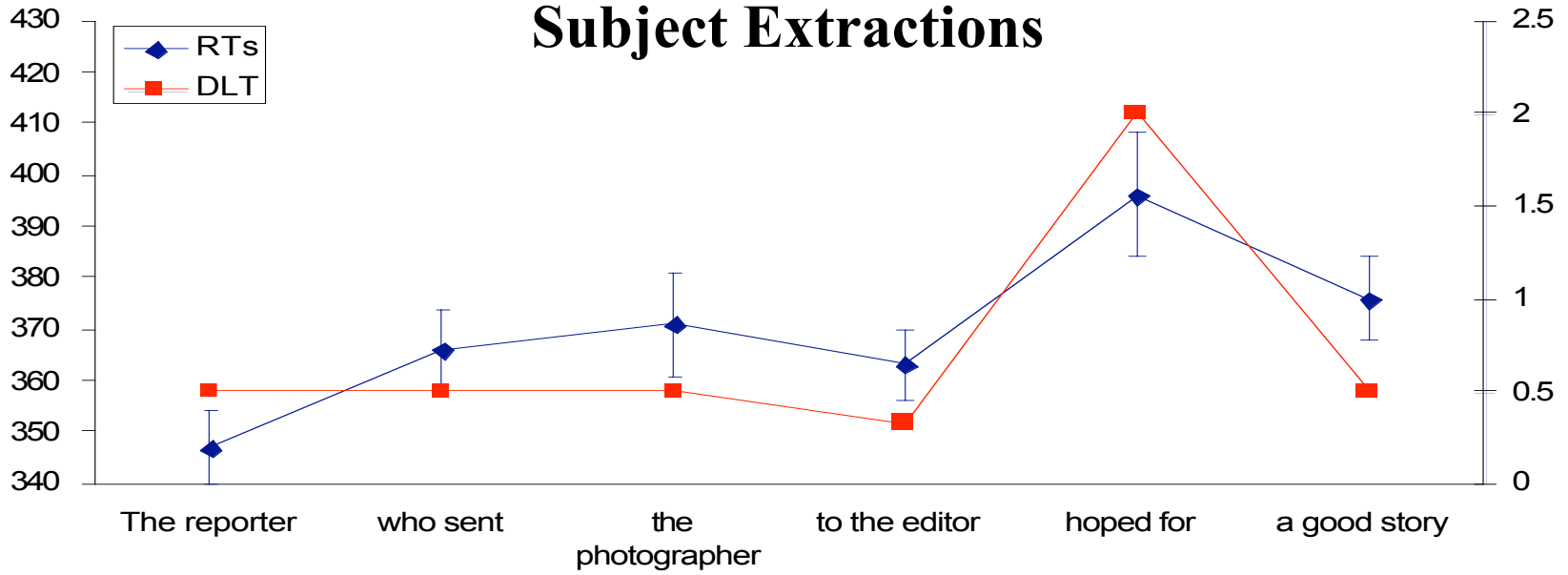


Locality effects in unambiguous structures

Subject-extracted relative clause:
The reporter who sent the photographer to the editor hoped for a story.



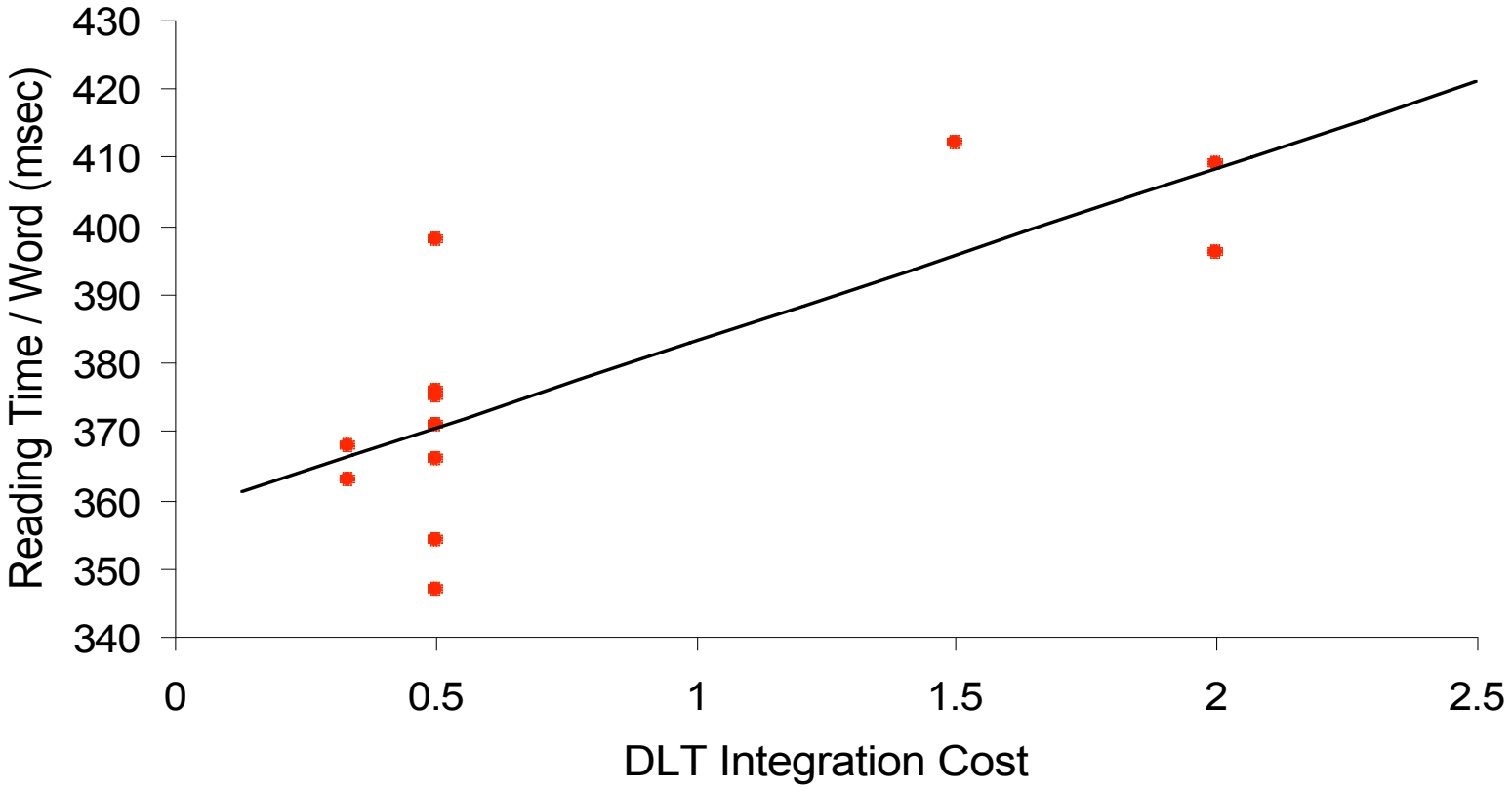

Object Extractions
RTs
DLT
430
420
410
400
390
380
370
360
350
340
2.5
2
1.5
1
0.5
0
The reporter
who the photographer
sent to
the editor
hoped for
a good story


Subject Extractions
RTs
DLT
430
420
410
400
390
380
370
360
350
340
2.5
2
1.5
1
0.5
0
The reporter
who sent
the photographer
to the editor
hoped for
a good story

Experiment 1: DLT vs. RTs

(Grodner & Gibson, in press, *Cognitive Science*)

Linear model: $r^2 = .582$, $p < .001$



Experiment 2 Materials

Matrix – Unmodified Subject

The nurse supervised the administrator while ...

0 1 1 0 1 1

Matrix – PP Modified Subject

The nurse from the clinic supervised the administrator while ...

0 1 0 0 1 2 0 1 1

Matrix – RC Modified Subject

The nurse who was from the clinic supervised the administrator while ...

0 1 0 1 0 0 1 3 0 1 1

Embedded – Unmodified Subject

The administrator who the nurse supervised scolded the medic while...

0 1 0 0 1 3 3 0 1 1

Embedded – PP Modified Subject

The administrator who the nurse from the clinic supervised scolded the medic…

0 1 0 0 1 0 0 1 5 4 0 1

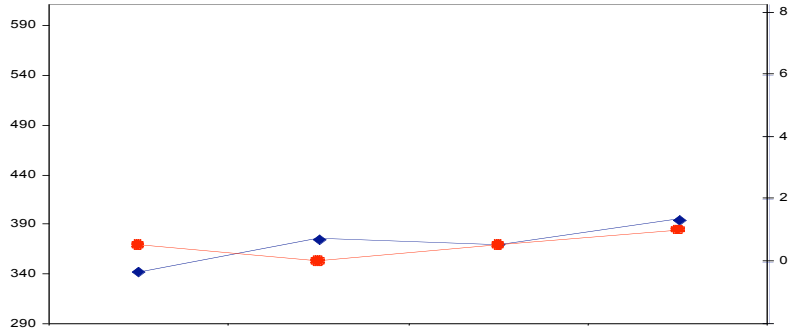
Embedded – RC Modified Subject

The administrator who the nurse who was from the clinic supervised scolded the medic…

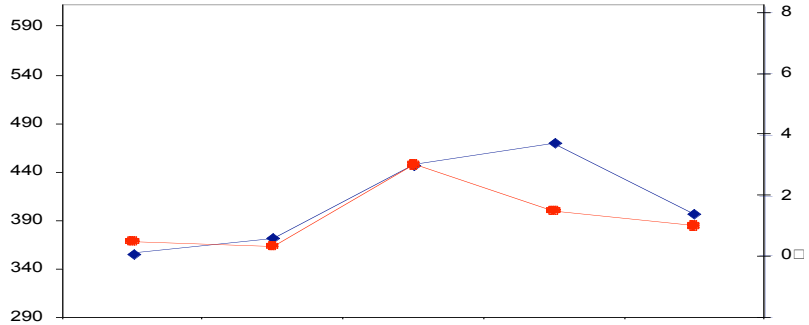
0 1 0 0 1 0 1 0 0 1 7 5 0 1

Experiment 2: DLT vs. RTs by Regions

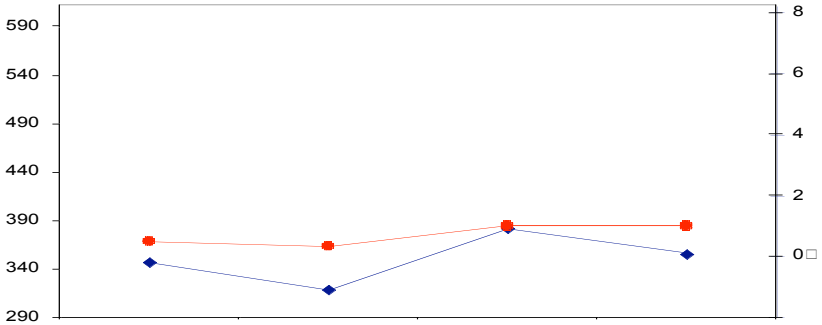
Matrix – Unmodified Subject



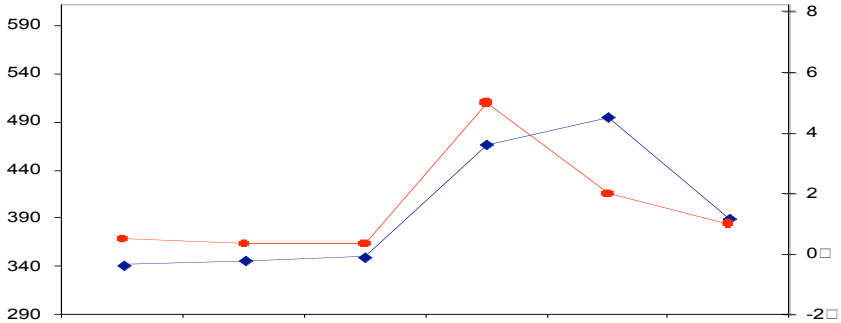
Embedded – Unmodified Subject



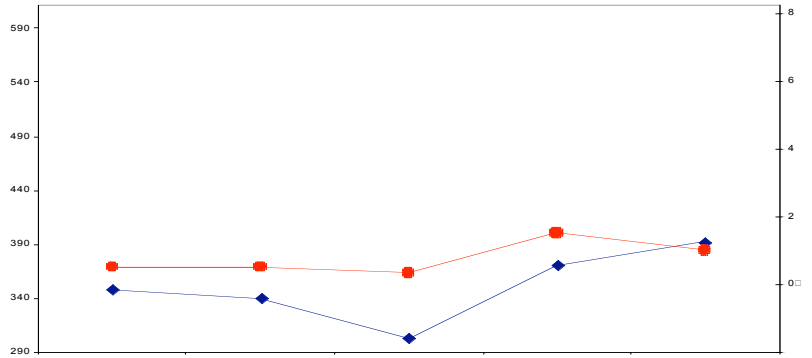
Matrix – PP Modified Subject



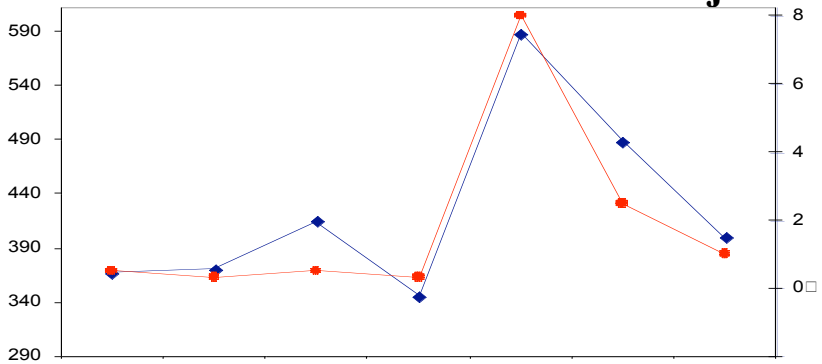
Embedded – PP Modified Subject



Matrix – RC Modified Subject

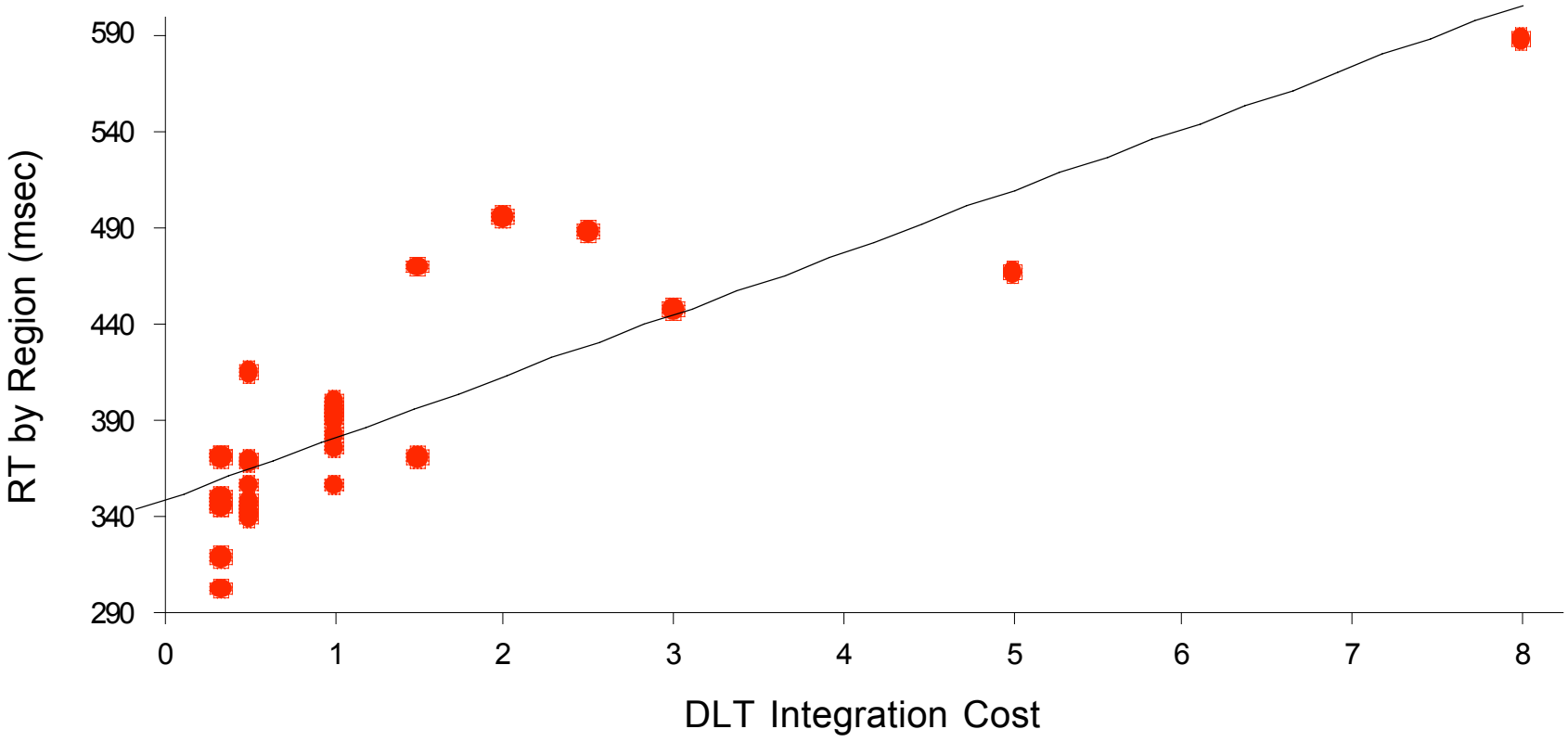


Embedded – RC Modified Subject



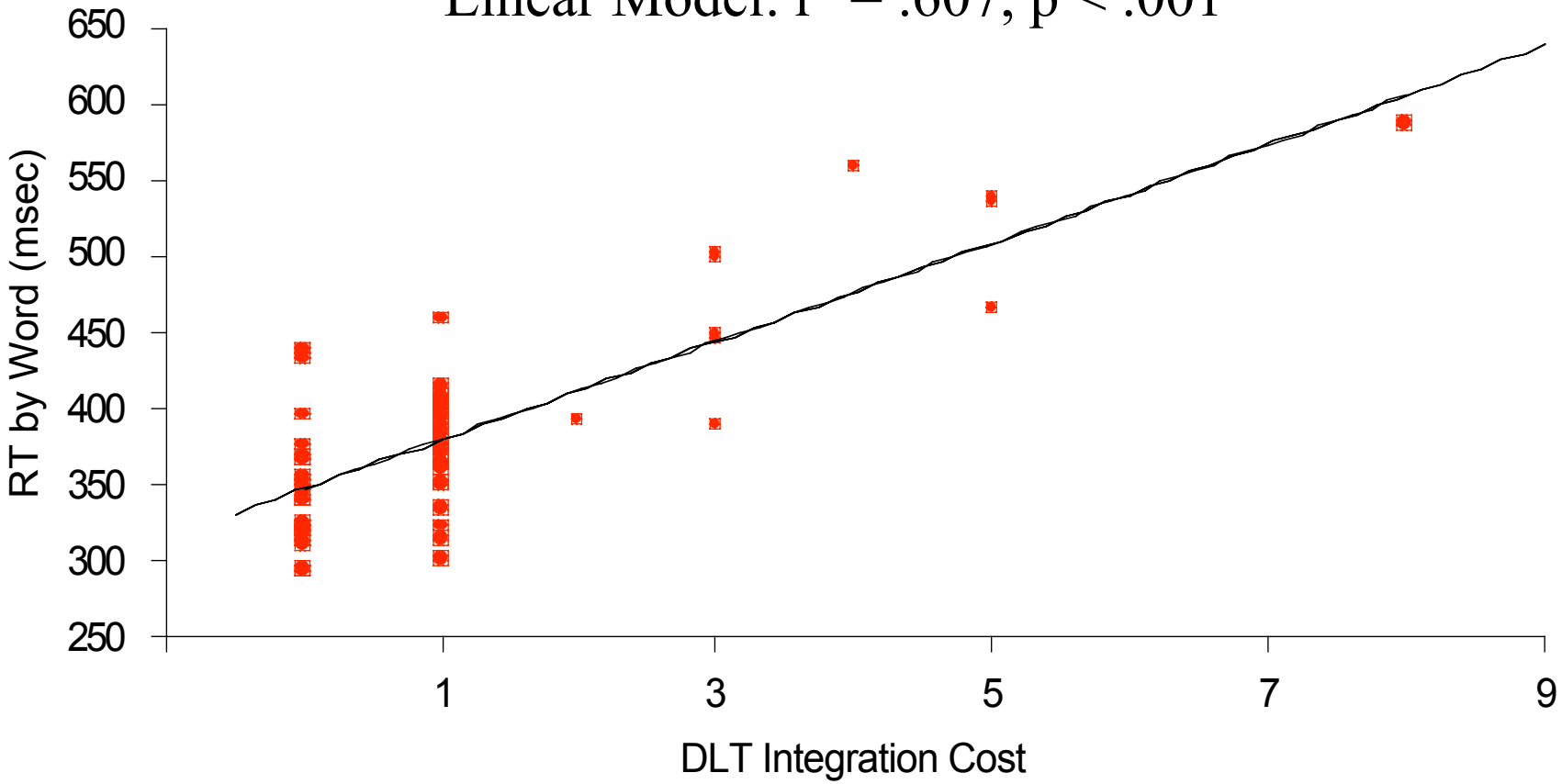
Experiment 2: DLT vs. RTs by Regions

Linear Model: $r^2 = .721$, $p < .001$



Experiment 2: DLT vs. RTs by Words

Linear Model: $r^2 = .607$, $p < .001$



Other effects of dependency distances

- Many construction comparisons cross-linguistically:
  - Heaviness effects: Put the heavy constituent later (Bever, 1970; Hawkins, 1994)
  - Nested vs. non-nested structures
  - Nested vs. cross-serial dependencies (Bach et al. 1986)

Length / heaviness effects

See e.g., Bever (1970), Hawkins (1994):

"Save the hardest for last."

(1) ? I gave [the beautiful green pendant that's been in the jewelry store window for weeks] [to my mother].

(2) I gave [my mother] [the beautiful green pendant that's been in the jewelry store window for weeks].

Nesting complexity effects

(1) The reporter disliked the editor.

(2) The reporter [who the senator attacked] disliked the editor.

(3) The reporter [who the senator [who John met] attacked] disliked the editor.

(4) John met the senator [who attacked the reporter [who disliked the editor]].

Locality account of nesting complexity

Nested structures have longer distance dependencies than non-nested structures.

An old puzzle: relative clauses (RCs) and complement clauses (CCs)

(1) RC within CC: difficult, but processable

The fact [that the employee [who the manager hired] stole office supplies] worried the executive.

(2) CC within RC: unprocessable

The executive [who the fact [that the employee stole office supplies] worried] hired the manager.

Assumption: the cost of an integration peaks after a some intervening material.

The worst integration steps are therefore those that involve **multiple** integrations at the same point.

Maximal integration step in (1): 1 long integration, at "worried"; 2 short integrations at "hired".

Maximal integration step in (2): **2 long integrations, at "worried".** (There is also 1 long integration at "hired".)

Nested vs. Cross-serial dependencies

Cross-serial dependencies are easier to process than nested dependencies (Bach, Brown & Marslen-Wilson, 1986)

Nested vs. Cross-serial dependencies

Nested (German)

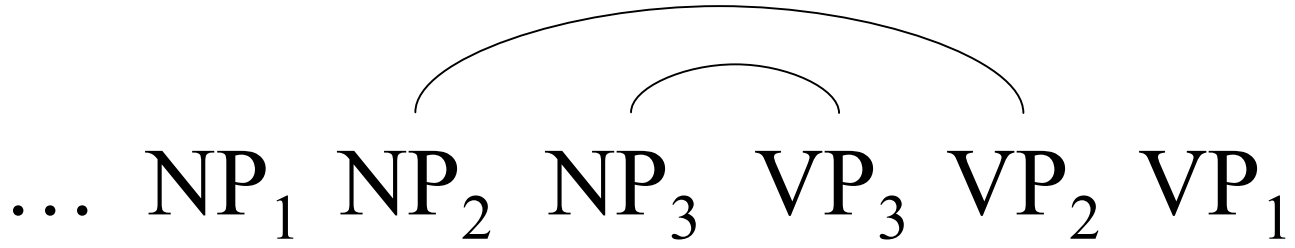
$$\ldots \; NP_1 \; NP_2 \; NP_3 \; VP_3 \; VP_2 \; VP_1$$

Nested vs. Cross-serial dependencies

Nested (German)

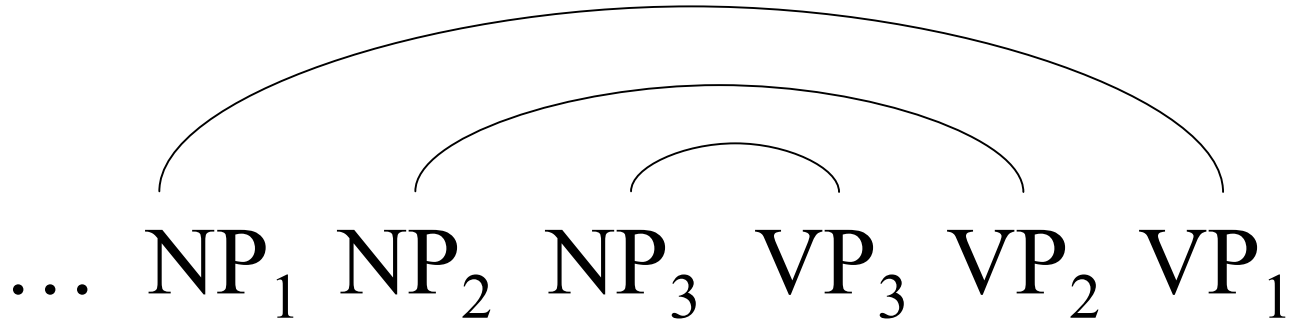
Nested vs. Cross-serial dependencies

Nested (German)



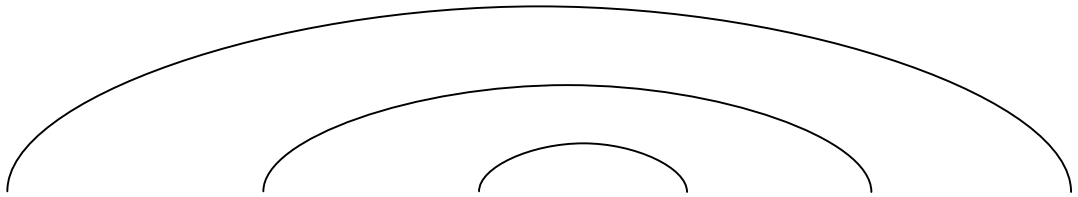
Nested vs. Cross-serial dependencies

Nested (German)



Nested vs. Cross-serial dependencies

Nested (German)



Johanna hat den Männern Hans die Pferde füttern lehren helfen.

Joanna has the men Hans the horses feed teach helped

“Joanna helped the men teach Hans to feed the horses.”

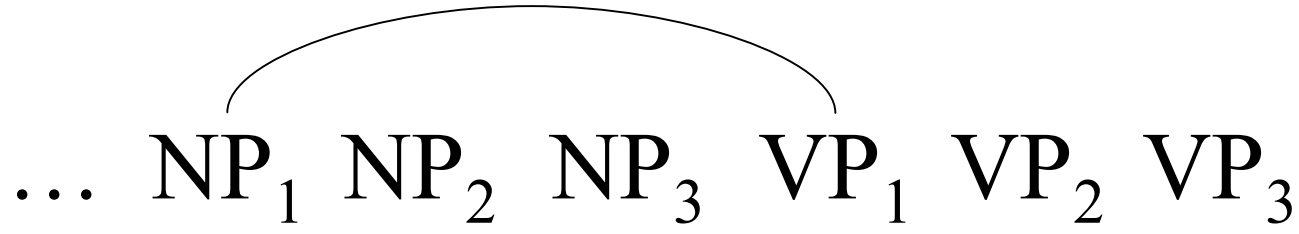
Nested vs. Cross-serial dependencies

Cross-serial (Dutch)

$\ldots\ NP_1\ NP_2\ NP_3\ VP_1\ VP_2\ VP_3$

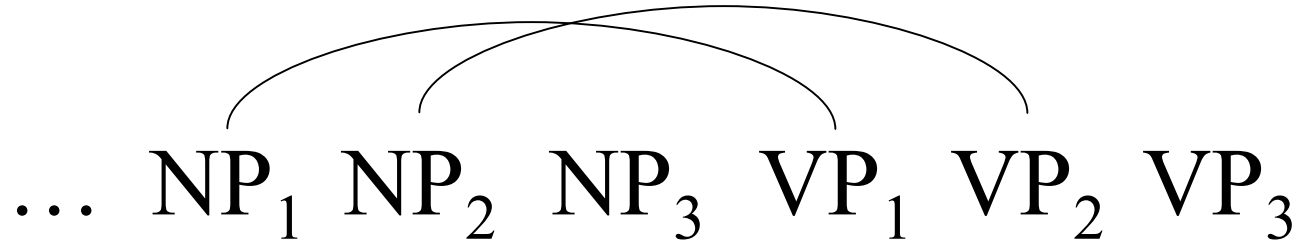
Nested vs. Cross-serial dependencies

Cross-serial (Dutch)



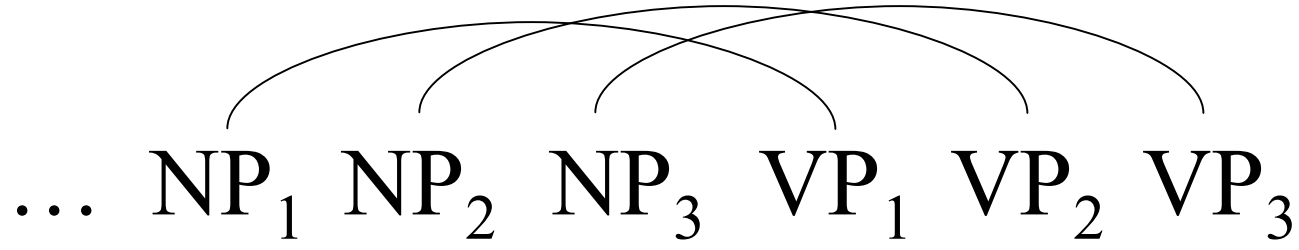
Nested vs. Cross-serial dependencies

Cross-serial (Dutch)



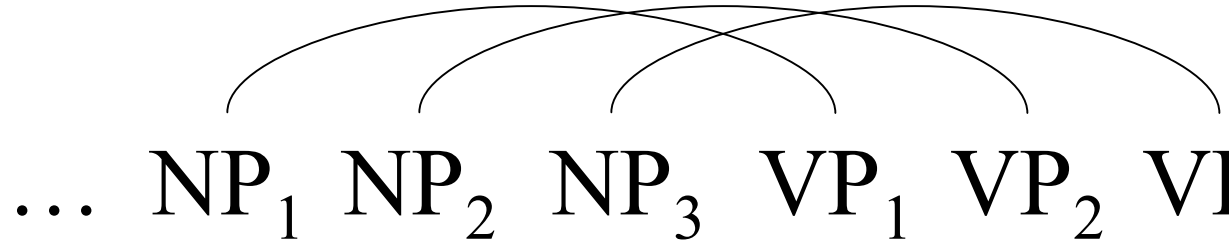
Nested vs. Cross-serial dependencies

Cross-serial (Dutch)



Nested vs. Cross-serial dependencies

Cross-serial (Dutch)



Jeanine heeft de mannen Hans die paarden helpen leren voeren.

Joanna has the men Hans the horses helped teach feed

“Joanna helped the men teach Hans to feed the horses.”

Open question: How to quantify distance?

- Words / syllables? (Gibson, 1998; Hawkins, 1994)
- Discourse structure? (Gibson, 1998; 2000; Warren & Gibson 2002)
  - New discourse referents (new nouns, verbs)
- Intervening interfering NPs? (Gordon et al. 2001, 2002; cf. Lewis)
  - Similar NPs (syntactic, semantic) cause more interference

Nested pronoun generalization
(Bever, 1974; Kac, 1981)

The low complexity of doubly-nested examples like:

(1) The reporter who everyone that I met trusts said the president won't resign yet. (Bever, 1974)

(2) A book that some Italian who I've never heard of wrote will be published soon by MIT Press. (Frank, 1992)

Cf. (3) # The reporter who the senator who John met trusts said that the president won't resign yet.

The relative ease of (1) and (2) can be explained by a discourse-based distance metric or an interference-based distance metric.

Null contexts:

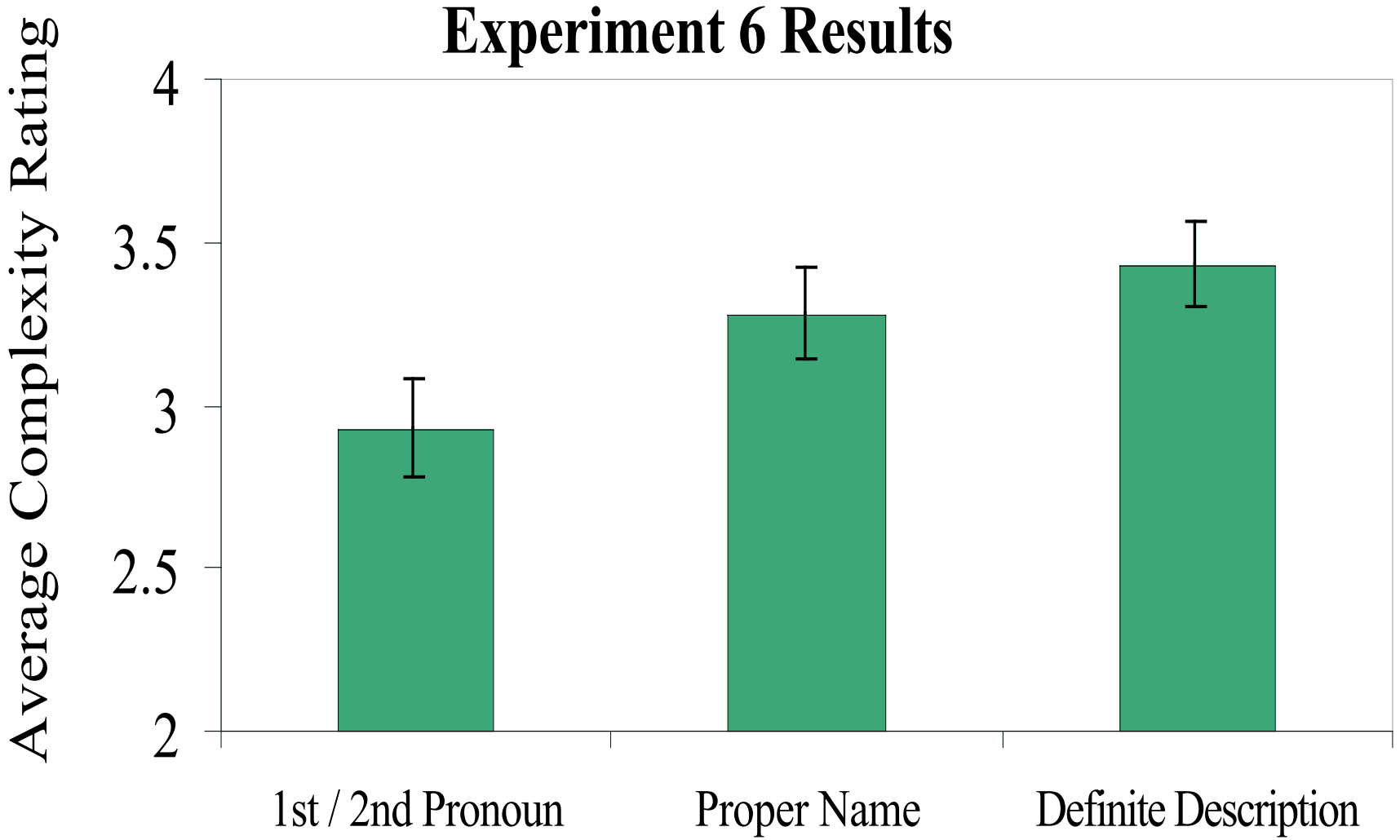
The low complexity of nested pronouns

Experiment 6 Materials (Warren & Gibson, 2002)

First / second person pronoun
The reporter who the senator who you met attacked disliked the editor.

Proper name
The reporter who the senator who John met attacked disliked the editor.

Definite description
The reporter who the senator who the professor met attacked disliked the editor.


Experiment 6 Results
Average Complexity Rating
4
3.5
3
2.5
2
1st / 2nd Pronoun
Proper Name
Definite Description

A discourse-based distance metric

(1) The reporter who the senator who **you** met trusts said that the president won't resign yet.

(2) # The reporter who the senator who **John** met trusts said that the president won't resign yet.

The relative ease of (1) vs. (2) can be explained in terms of a discourse-based distance metric: count the number of new NPs and verbs between two elements that need to be integrated.

Maximal integration cost in both sentences is at "trusts": 2 integrations: a) between "the senator" and "trusts"; and b) between "who" and "trusts".

A discourse-based distance metric

(1) The reporter who the senator who **you** met trusts said that the president won't resign yet.

(2) # The reporter who the senator who **John** met trusts said that the president won't resign yet.

In (1):

Integrating "trusts" to "senator": 2 new NPs/verbs ("trusts", "met") since "the senator".

Integrating "trusts" to "who": 3 new NPs/verbs ("trusts", "met", "senator") since "who".

Total cost: 2+3 = 5 units.

In (2):

Integrating "trusts" to "senator": 3 new NPs/verbs ("trusts", "met", **"John"**) since "the senator".

Integrating "trusts" to "who": 4 new NPs/verbs ("trusts", "met", **"John"**, "senator") since "who".

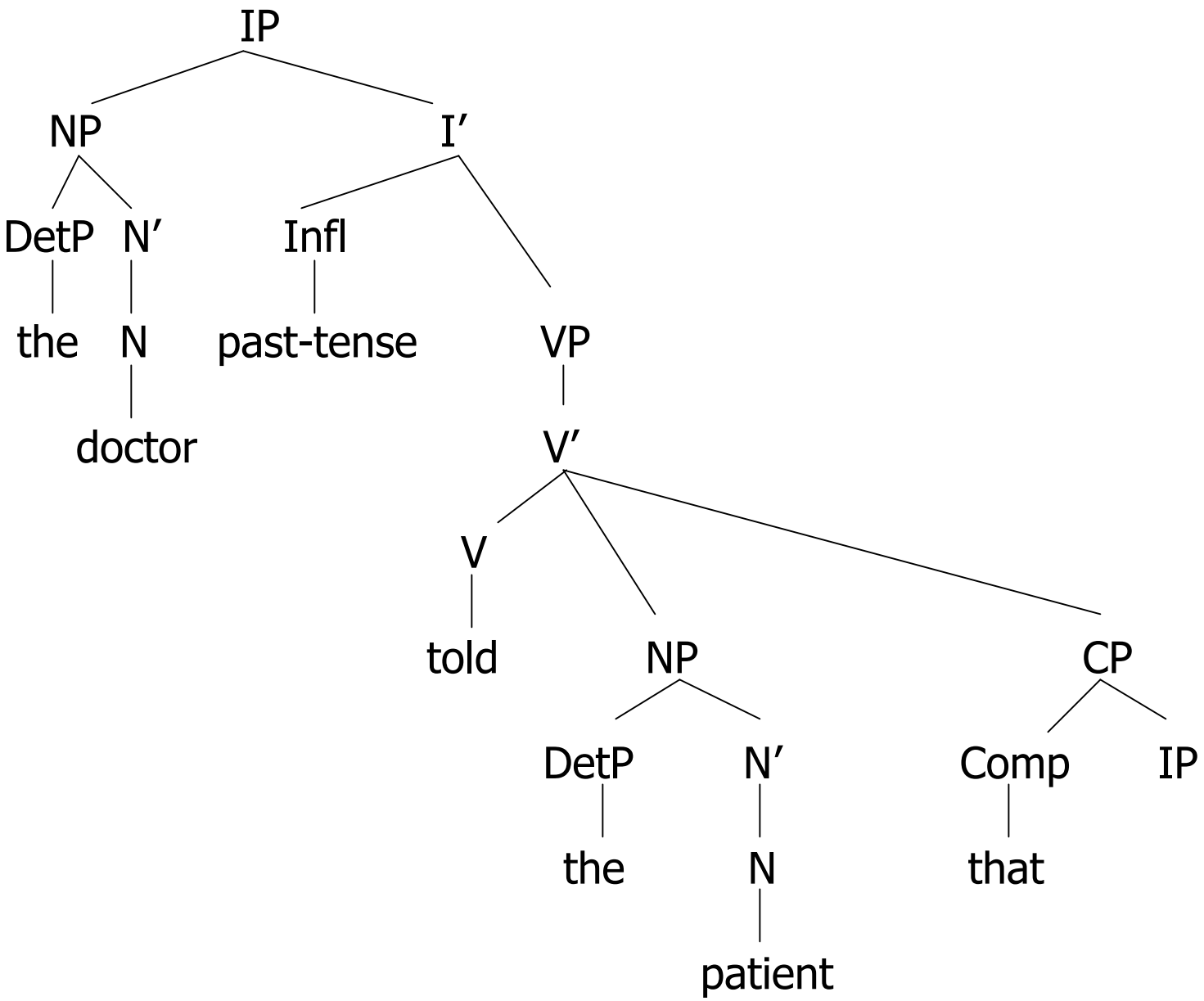
Total cost: 3+4 = 7 units.

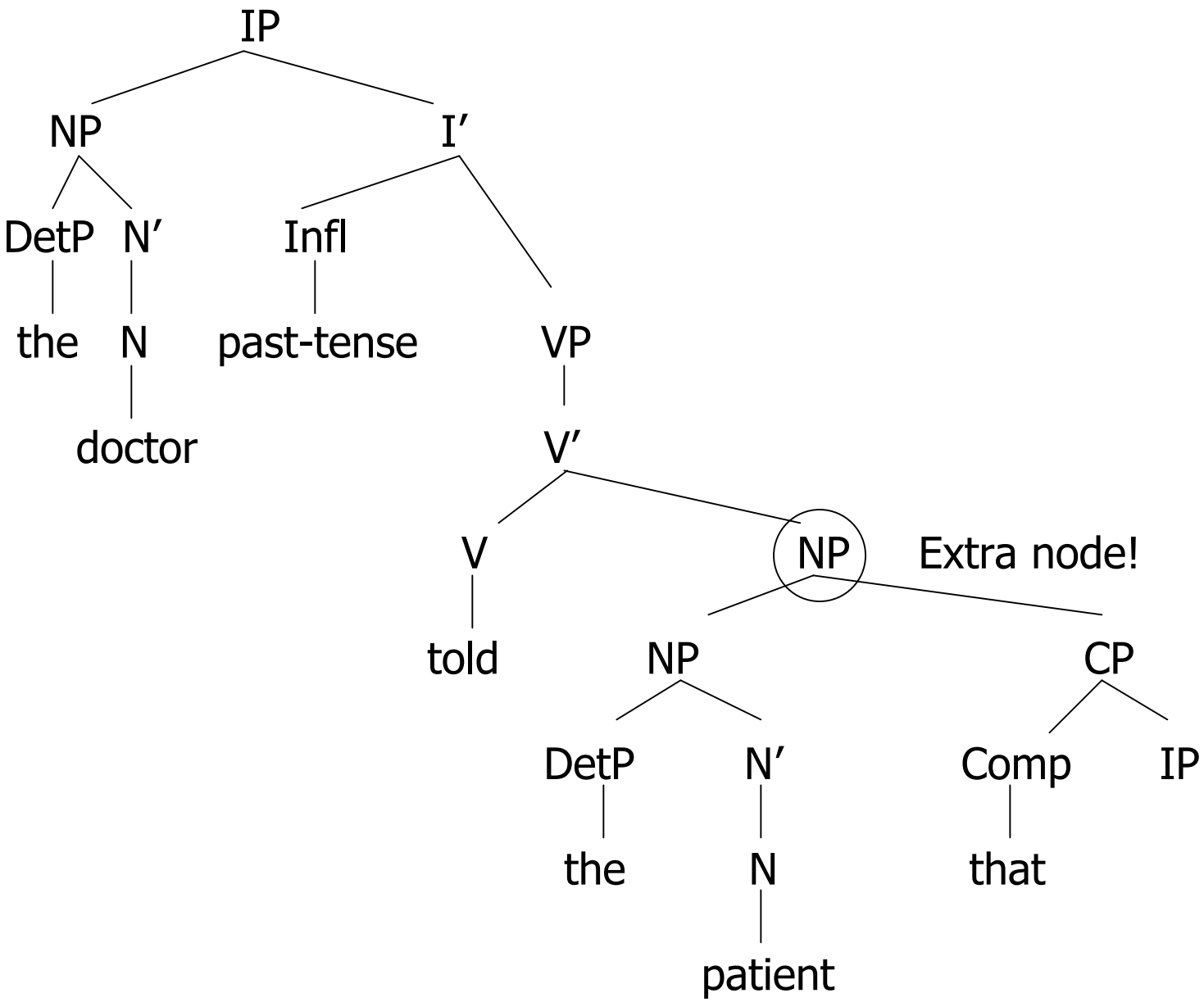
Application to technical writing

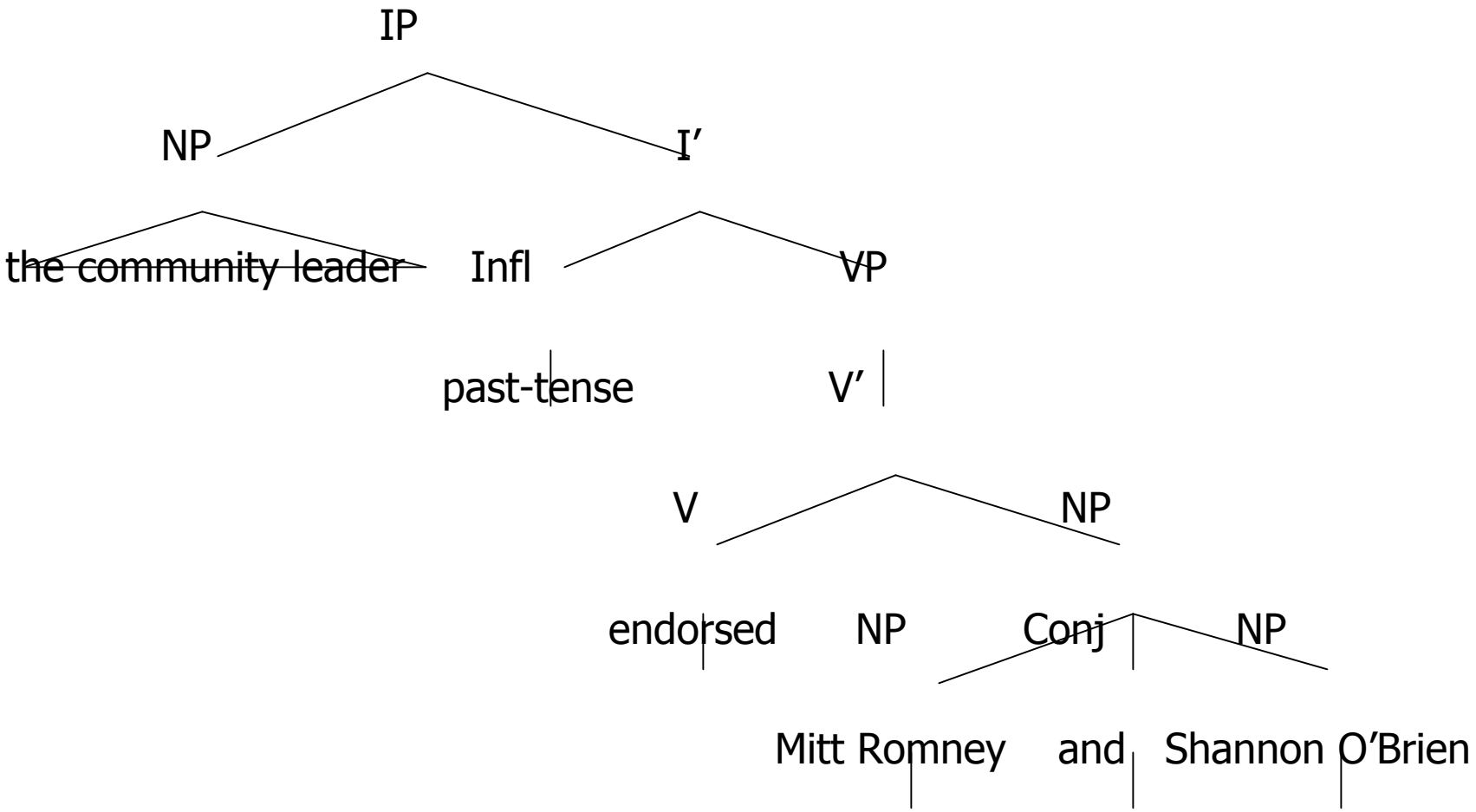
- To help the reader, keep syntactic dependencies close together.

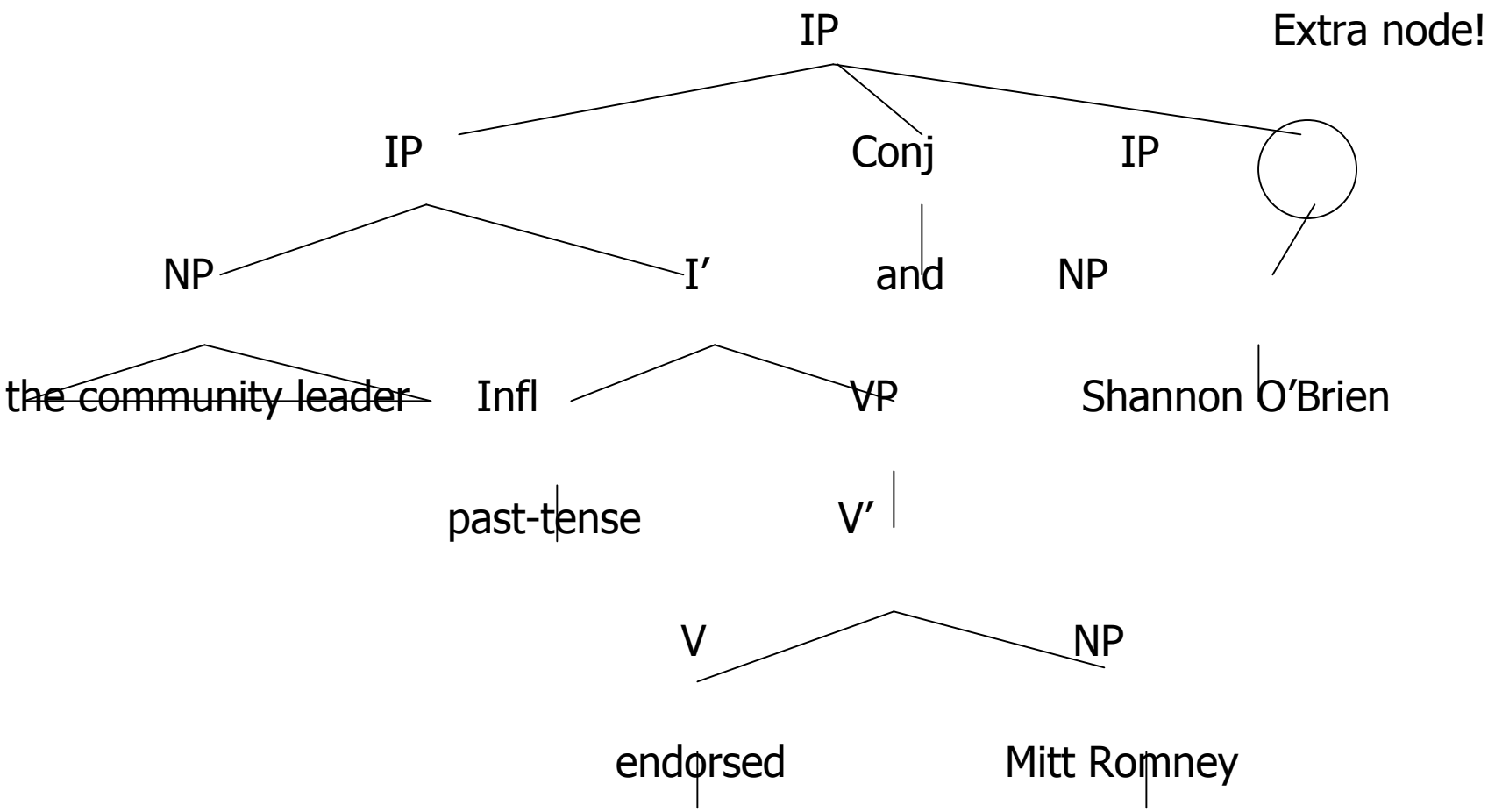
Next time: Syntactic storage

Appendix: Minimal Attachment examples


IP
NP
I’
DetP
N’
Infl
the
N
past-tense
VP
doctor
V’
V
told
NP
CP
DetP
N’
Comp
IP
the
N
that
patient


IP
NP
I’
DetP
N’
Infl
the
N
past-tense
VP
doctor
V’
V
NP
Extra node!
told
NP
CP
DetP
N’
Comp
IP
the
N
that
patient


IP
NP
I’
the community leader
Infl
VP
past-tense
V’
V
NP
endorsed
NP
Conj
NP
Mitt Romney
and
Shannon O’Brien


IP
Extra node!
IP
Conj
IP
NP
I’
and
NP
the community leader
Infl
VP
Shannon O’Brien
past-tense
V’
V
NP
endorsed
Mitt Romney

Examples of Minimal Attachment preferences

NP / S ambiguity:

?# Sally discovered **the answer to the physics problem** was in the back of the book.

Noun-noun (NN) / Relative clause (RC)

The **cotton clothing is made of** grows in Mississippi.

Others:

The teacher told the children **the ghost story had frightened** that it wasn't true.

Principle 2: **Late Closure** (= Locality): When possible, attach incoming lexical items into the clause or phrase currently being processed.

Adverbial attachment:

The bartender told the detective that the suspect left the country yesterday.

NP / Zero ambiguity

While Mary was mending the sock fell off her lap.

Minimal Attachment and Late Closure are principles of ambiguity resolution only. They do not extend to processing unambiguous structures.

The dependency locality theory (DLT) principles apply to both ambiguous and unambiguous structures.

More on the DLT later.