Sentence comprehension

9.591; 24.945

September 13, 2004

Ted Gibson

Sentence processing

Research in sentence comprehension attempts to discover what representations are used in the understanding and production of sentences, and how these representations are used in the course of processing a sentence.

Puzzle: How do we process sentences?

Sometimes the context helps:

Please pass me the book on the table.

Puzzle: How do we process sentences?

Often the context does not help:

Giant lizard-like creatures are descending from spaceships and attacking Boston.

A man with below average intelligence will someday become the leader of the most powerful nation in the free world.

Today

1. How to address the question of how sentences are comprehended.

2. Information sources used in sentence comprehension.

3. Modularity in sentence comprehension? Syntax first?

4. The effects of plausibility, context and lexical frequency in on-line processing

How to uncover how the language processing mechanism works?

- Find input that the mechanism has difficulty with;
- Find input that the mechanism has little or no difficulty with.

How to uncover how the language processing mechanism works?

Useful evidence:

1. Ambiguous input that is easy / hard to process.
2. Unambiguous input that is easy / hard to process.

Easy to process temporary ambiguity

John knows Mary.
John knows Mary is intelligent.

The desert trains young people to be tough.
The desert trains are tough on young people.

Is the crowd in the room?
Is the crowd in the room happy?

Hard to process temporary ambiguity: Garden-path effects

The dog walked to the park chewed the bone.
(cf. The dog that was walked to the park chewed the bone.)

The horse raced past the barn fell.
(cf. The horse that was raced past the barn fell.

The cotton clothing is made of comes from Mississippi.
(cf. The cotton that clothing is made of comes from Mississippi.)

I put the candy on the table into my mouth.
(cf. I put the candy that was on the table into my mouth.)

Reading methods: Self-paced reading, eye-tracking

[illegible].

The --------- -------- -- --- ------
------ --- -- -- ----------.

--- defendant -------- -- --- ------
------ --- -- -- ----------.

--- ---------- examined -- --- ------
------ --- -- -- ----------.

--- ---------- --------- by --- ------
------ --- -- -- ----------.

--- --------- -------- -- the ------
------ --- -- -- ----------.

--- --------- -------- -- --- lawyer
------ --- -- -- ----------.

--- --------- -------- -- --- ------

turned --- -- -- ----------.

--- --------- -------- -- --- ------
------ out -- -- ----------.

--- --------- -------- -- --- ------
------ --- to -- ----------.

--- --------- -------- -- --- ------
------ --- -- be ----------.

--- --------- -------- -- --- ------
------ --- -- -- unreliable.

Did the defendant examine the lawyer?

The existence of garden-path effects provides evidence:

- That language is processed on-line, as it is heard or read
- That the human parser is not unlimited parallel. Rather, it must be ranked parallel or serial.

Hard to process unambiguous sentences
Nested (or *center-embedded*) structures

The reporter disliked the editor.

The reporter [who the senator attacked] disliked the editor.

\# The reporter [who the senator [who John met] attacked] disliked the editor.

Right-branching (non-nested) control:

John met the senator who attacked the reporter who disliked the editor.

Cross-linguistic generalization: Nested structures are hard; left- and right-branching structures are not.

Japanese:

Obasan-wa [bebiisitaa-ga [ani-ga imooto-o ijimeta] to itta] to omotteiru

aunt-top babysitter-nom older-brother-nom younger-sister-acc bullied that said that thinks

"My aunt thinks that the babysitter said that my older brother bullied my younger sister"

Less nested version: easier to understand

[bebiisitaa-ga [ani-ga imooto-o ijimeta to] itta to] obasan-wa omotteiru

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
  - Has an entity been introduced?
  - Information flow
  - Discourse coherence
- Intonational information

Information that is used in sentence comprehension

1. Syntax: Word order

The dog bit the boy.

vs.

The boy bit the dog.

Information that is used in sentence comprehension

2. Lexical (Word) information, e.g., frequency

Unambiguous sentences: more frequent, faster: “class” vs. “caste”

Ambiguity: more frequent usages are preferred
The old man the boats.

Syntactic argument structure frequencies
E.g., many verbs can take either an NP or a CP complement

Mary discovered / believed the answer was in the back of the book.

More difficulty in comprehending the disambiguating region “was in the ...” for the NP-biased verb “discover” than for the CP-biased verb “believe”.

Information that is used in sentence comprehension

2. Lexical (Word) information, e.g., frequency

Words with multiple senses of roughly equal frequency are comprehended slower (e.g., "pitcher") than unambiguous words or words which are highly frequency-biased towards one sense (e.g., "port").

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.”

Photos removed for copyright reasons.

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.

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!

Information that is used in sentence comprehension: Intonational or prosodic information

E.g., Intonational boundary information.

An intonational boundary: A perceptual break in an utterance. Typically (but not always) associated with lengthening and stress on the last word. A pause also typically occurs at the boundary between intonational phrases.

Information that is used in sentence comprehension: Intonational or prosodic information

The Anti-Attachment Hypothesis (Watson & Gibson, in press):

Listeners prefer not to attach an incoming word to a lexical head that is immediately followed by an intonational phrase boundary.

As a result, the presence of a boundary at a local attachment site increases processing difficulty; and the presence of a boundary after a word that has no subsequent attachments decreases processing difficulty.

Information that is used in sentence comprehension: Intonational or prosodic information

Prepositional phrase attachment ambiguity:
The bus driver angered the rider with a mean look.

How to interpret “with a mean look”?

Intonation can disambiguate:

Boundary before “the rider”: attachment of “with a mean look” to “angered” is more difficult. Therefore attachment to “the rider” is preferred.

Boundary before “with a mean look”: attachment of “with a mean look” to “the rider” is more difficult. Therefore attachment to “angered” is preferred.

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:

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

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:

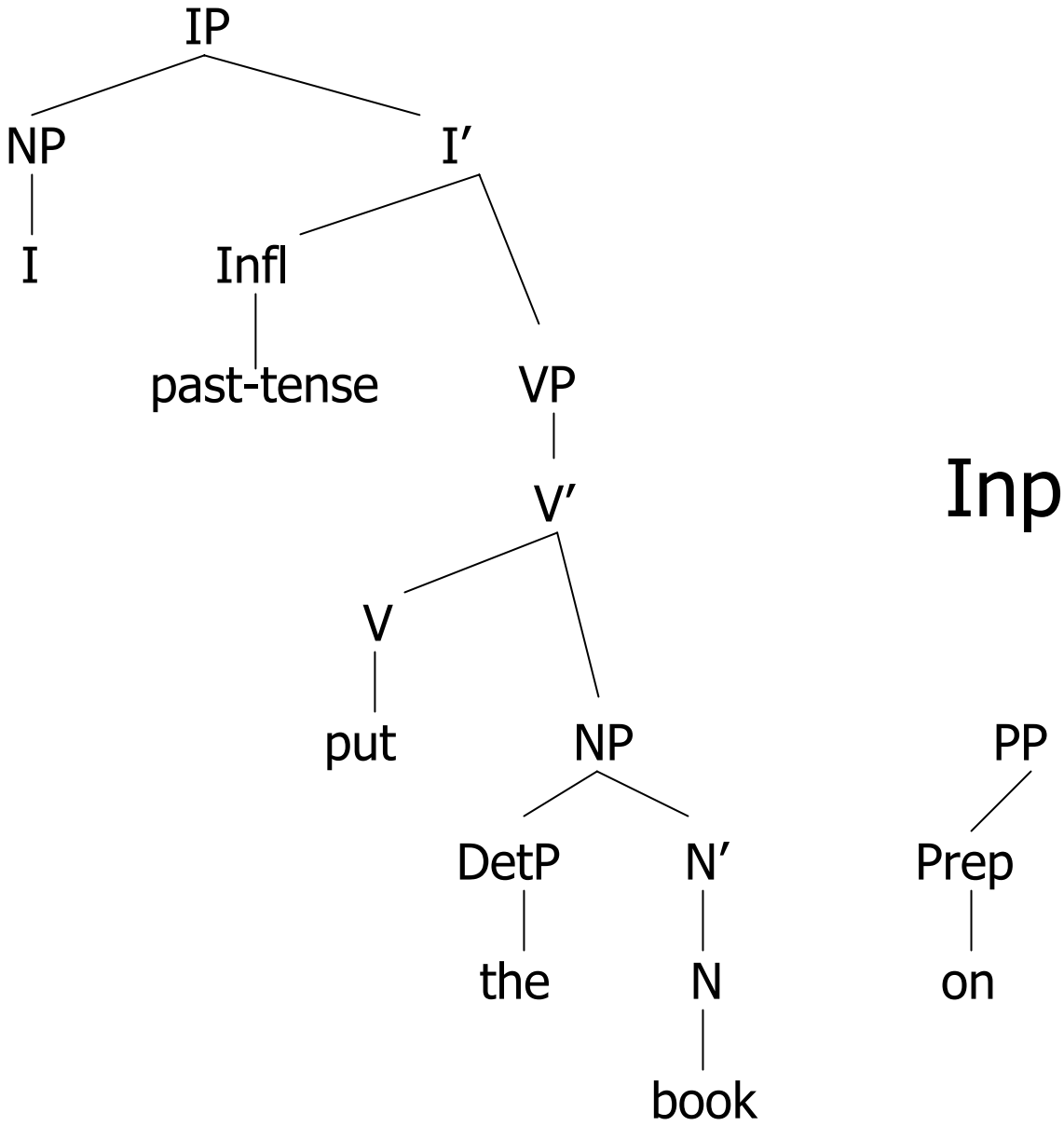
The psychologist convinced the patient **that he was having trouble with** to leave.

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

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

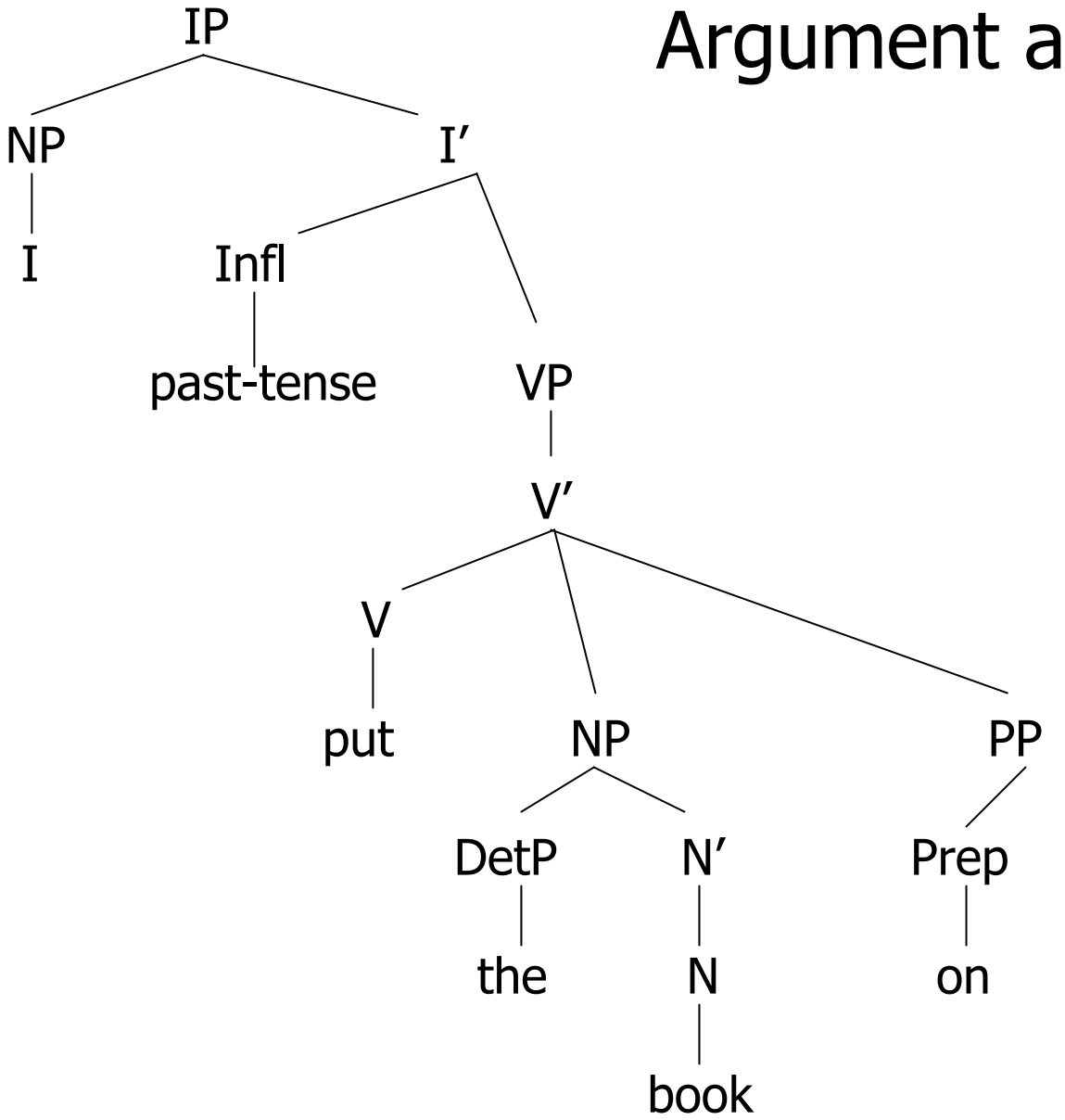
The horse **raced** past the barn fell.

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

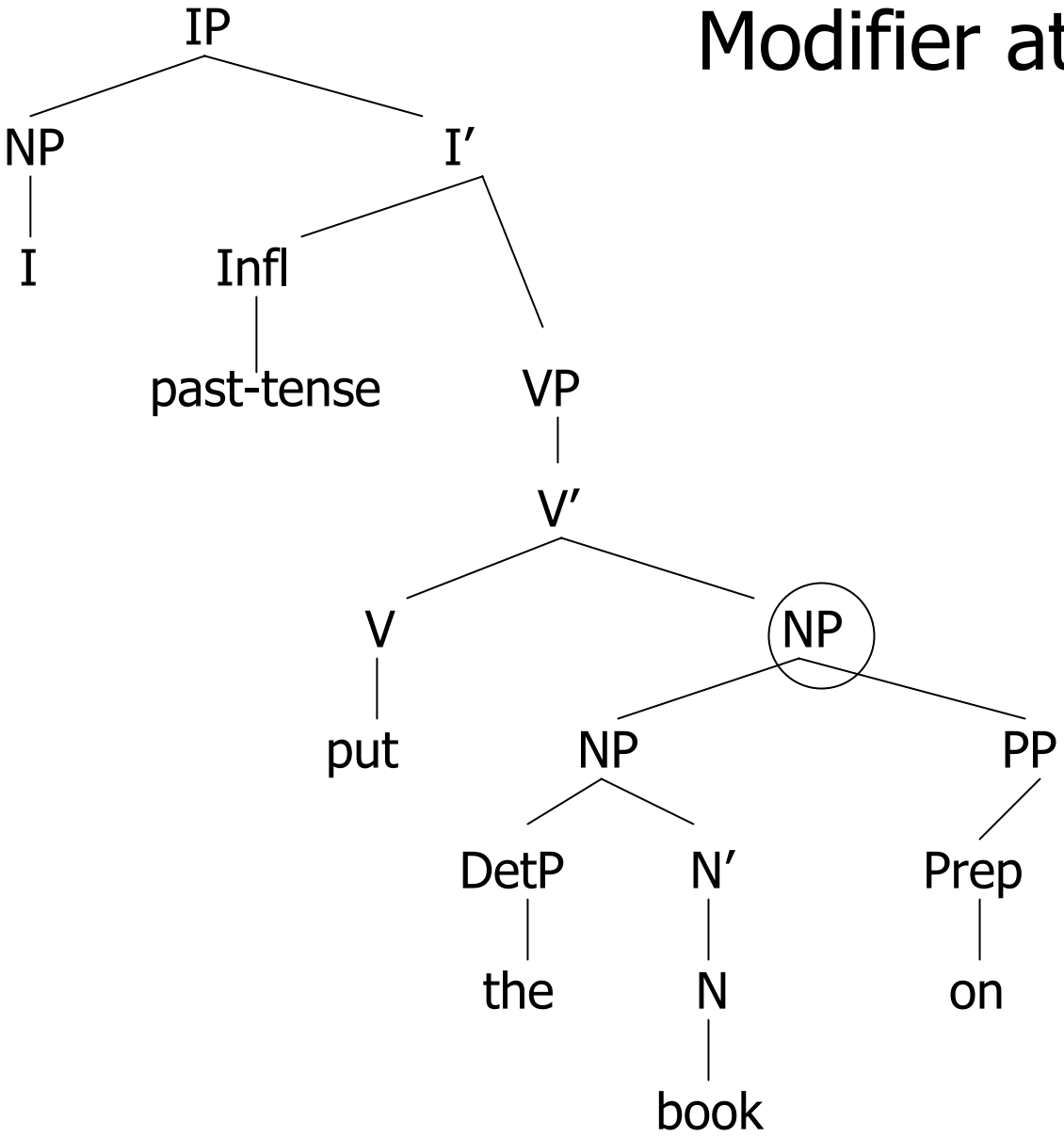


Input word: “on”

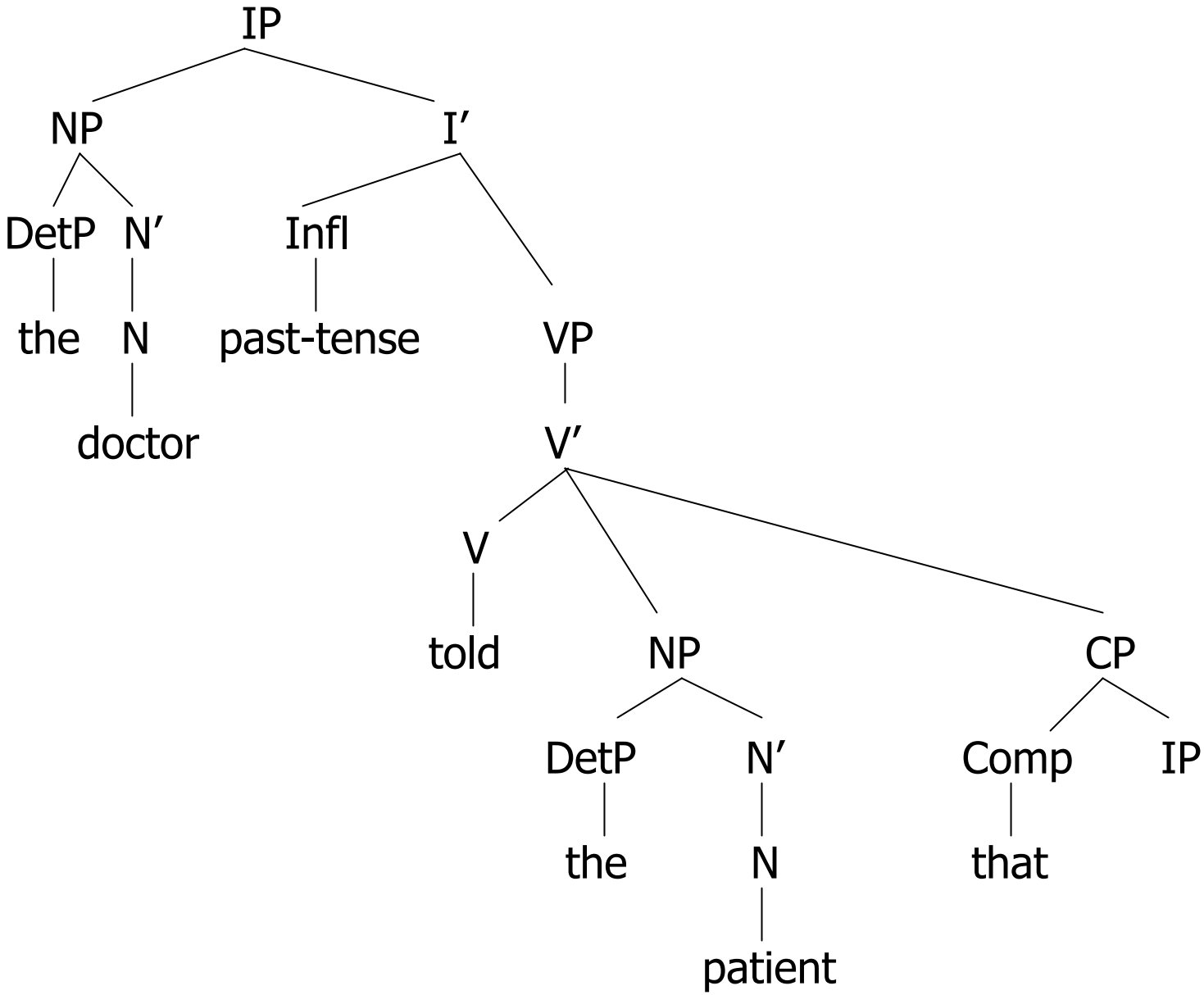
Argument attachment of “on”

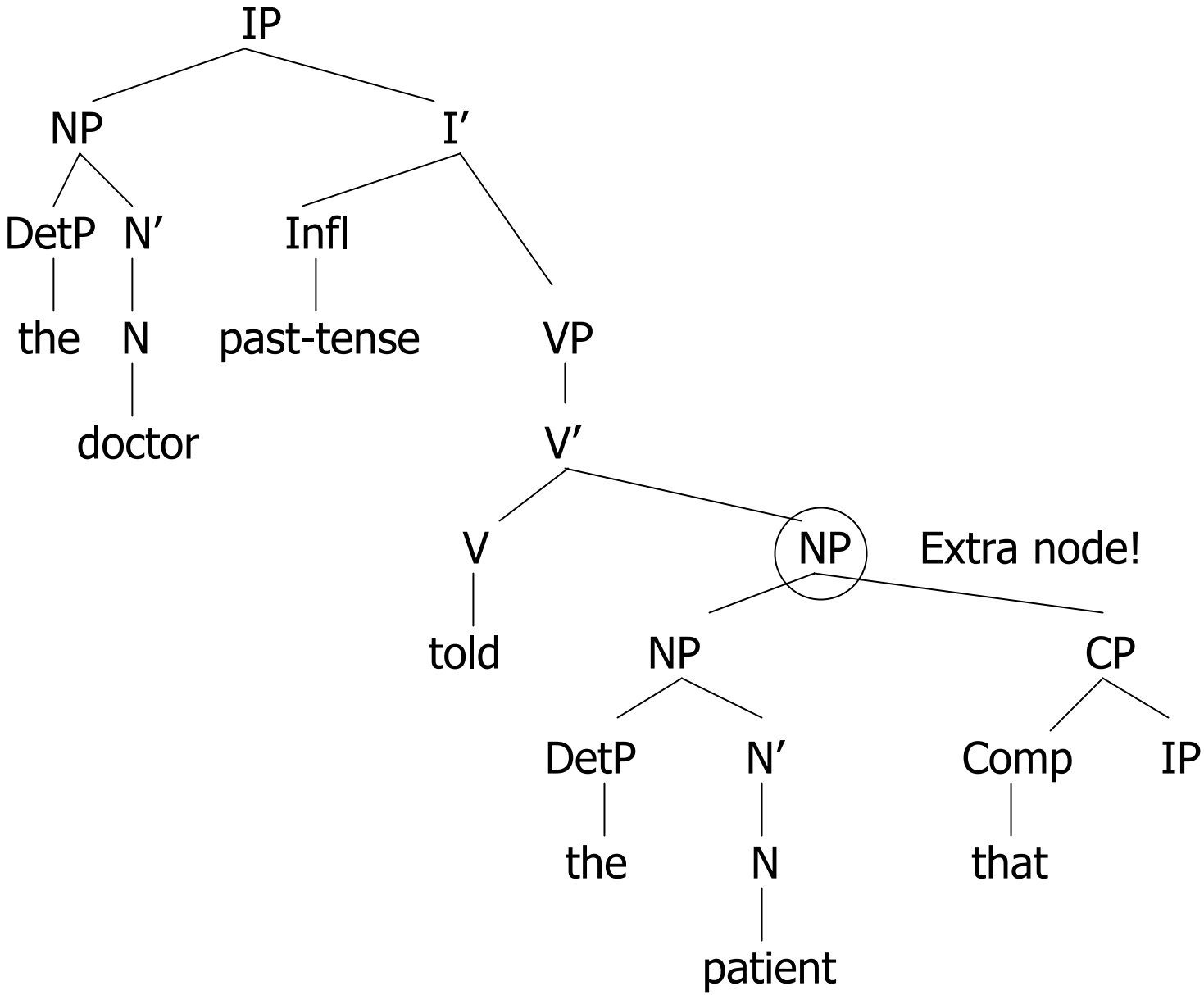


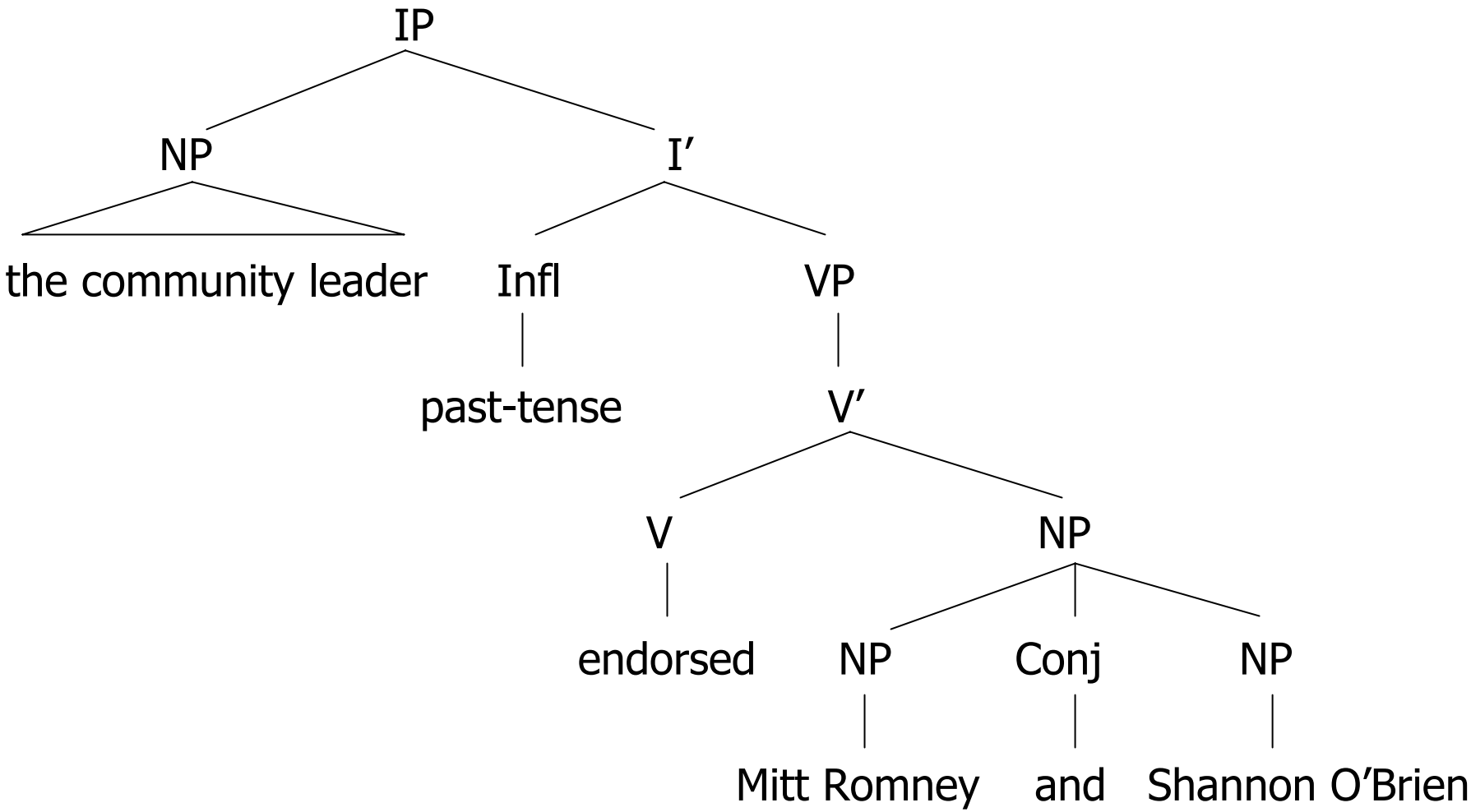
Modifier attachment of "on"

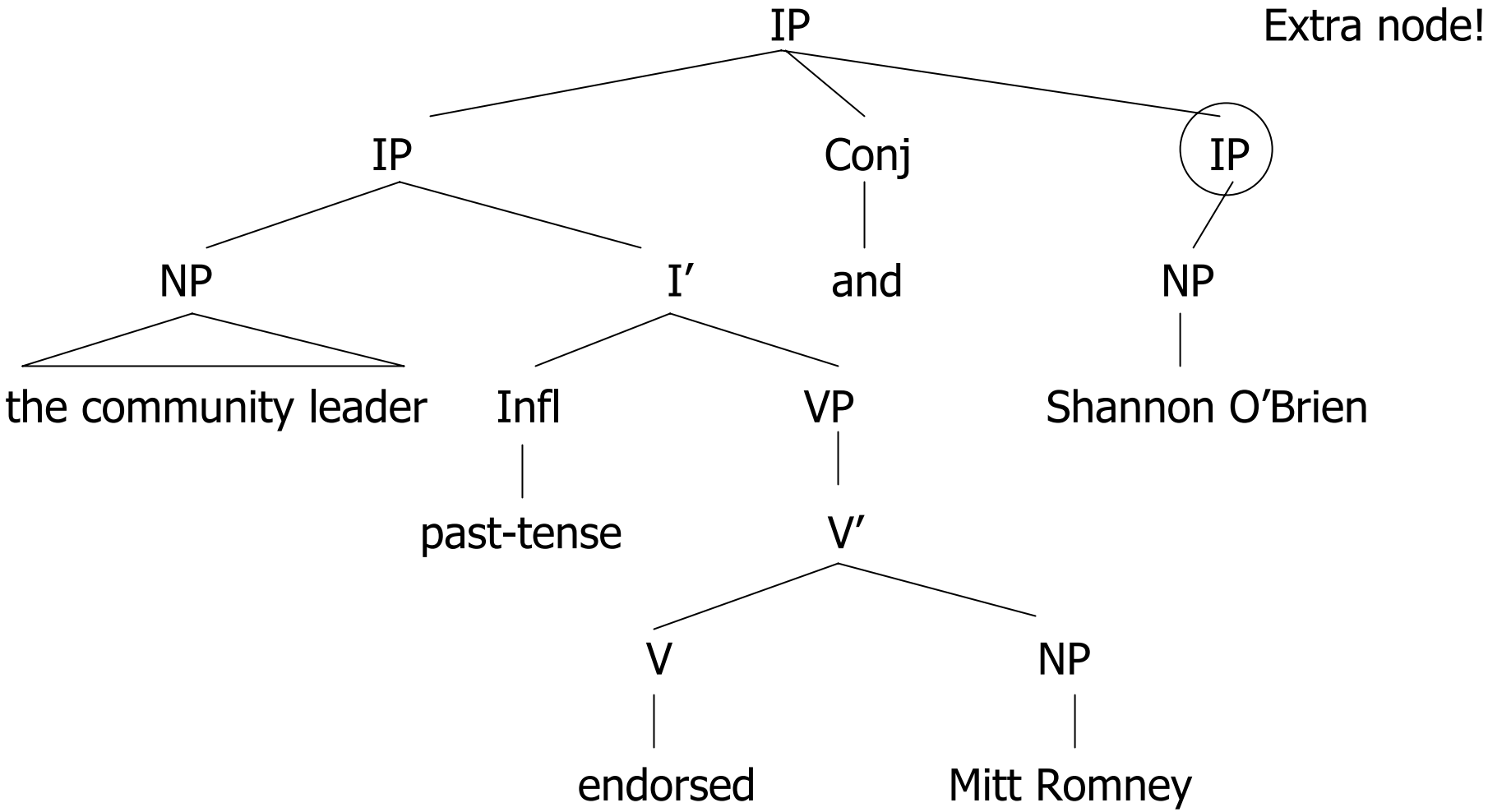


Extra node in the tree!

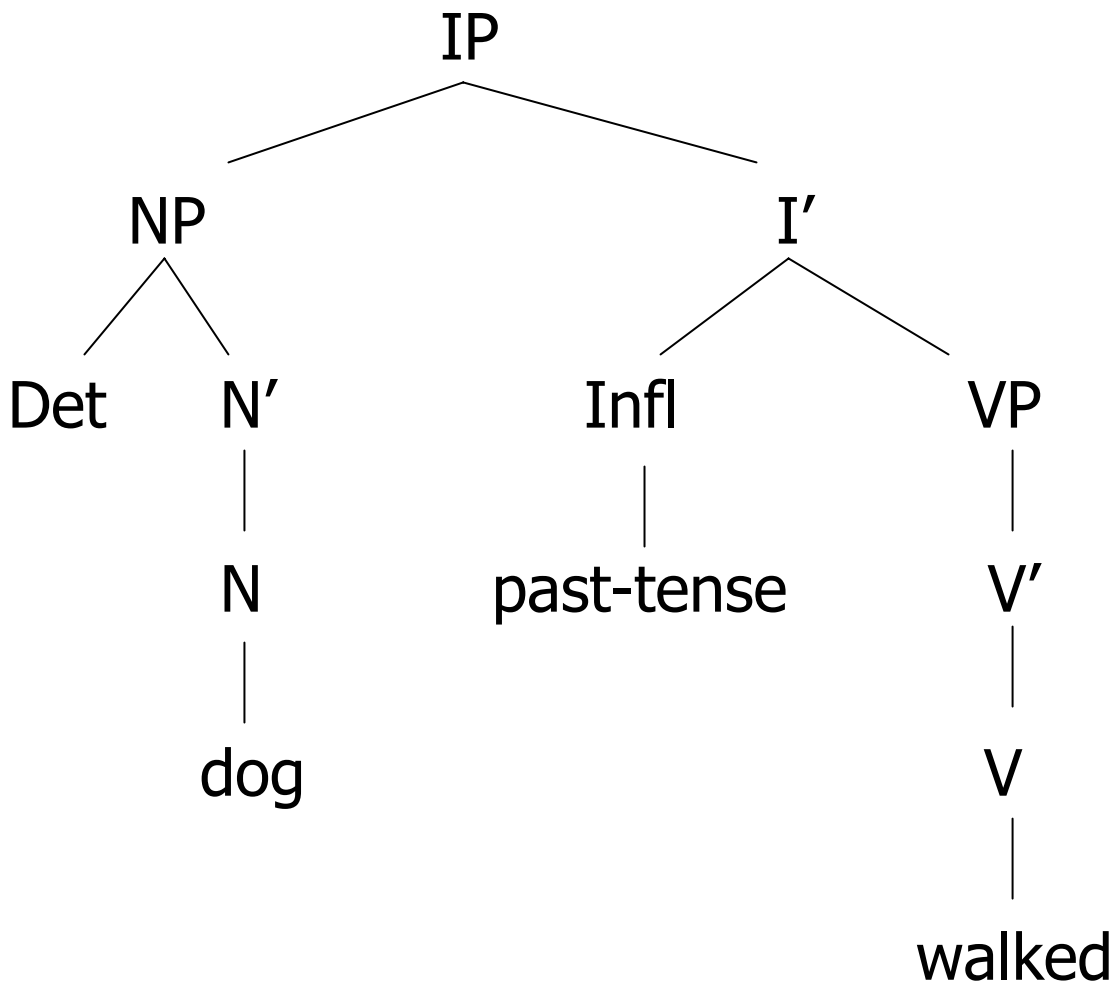

IP
NP
I’
DetP
N’
Infl
the
N
past-tense
VP
doctor
V’
V
told
NP
CP
DetP
N’
Comp
IP
the
N
that
patient


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

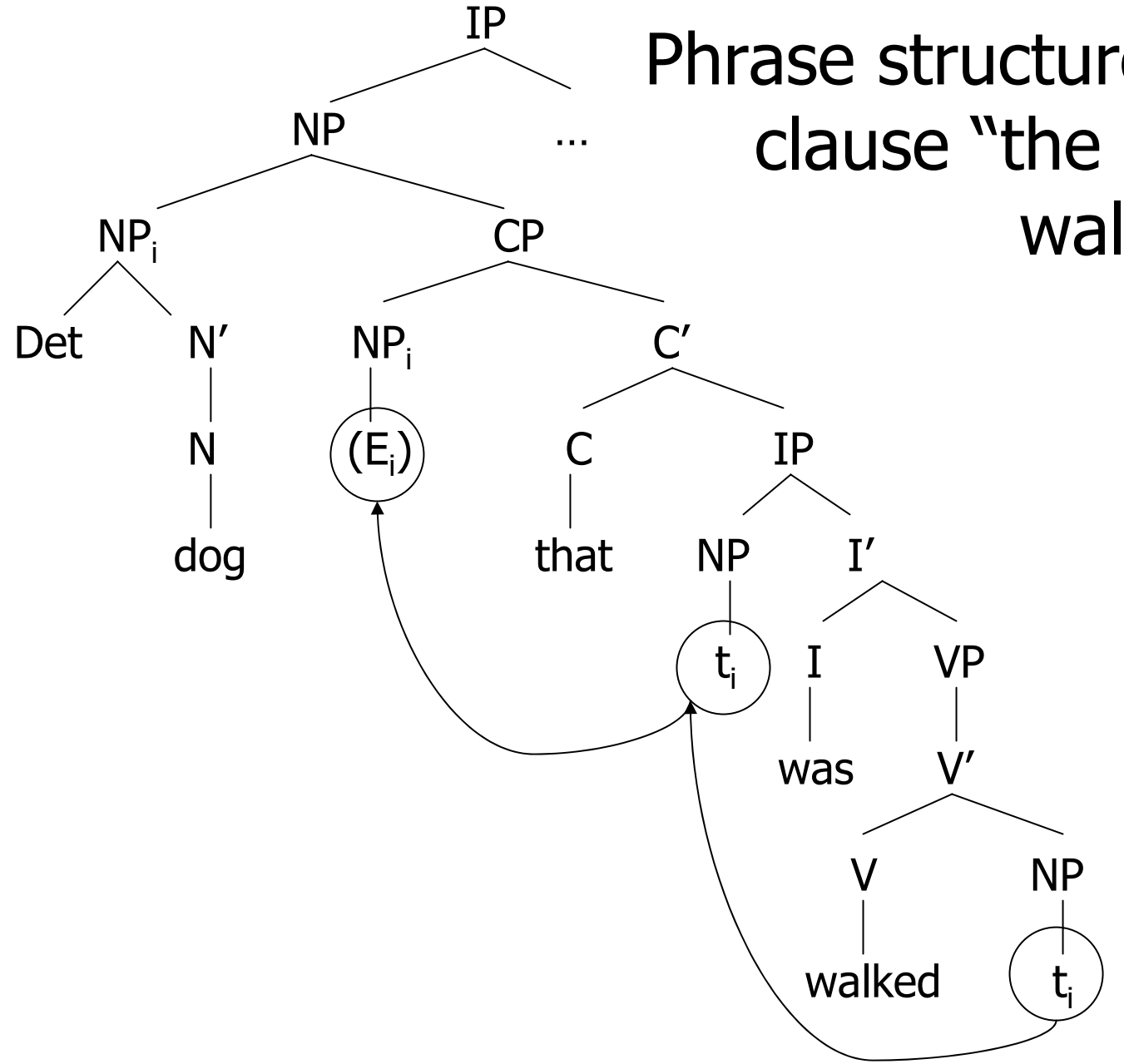

IP
NP
I’
the community leader
Infl
VP
past-tense
V’
V
NP
endorsed
NP
Conj
NP
Mitt Romney
and
Shannon O’Brien


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

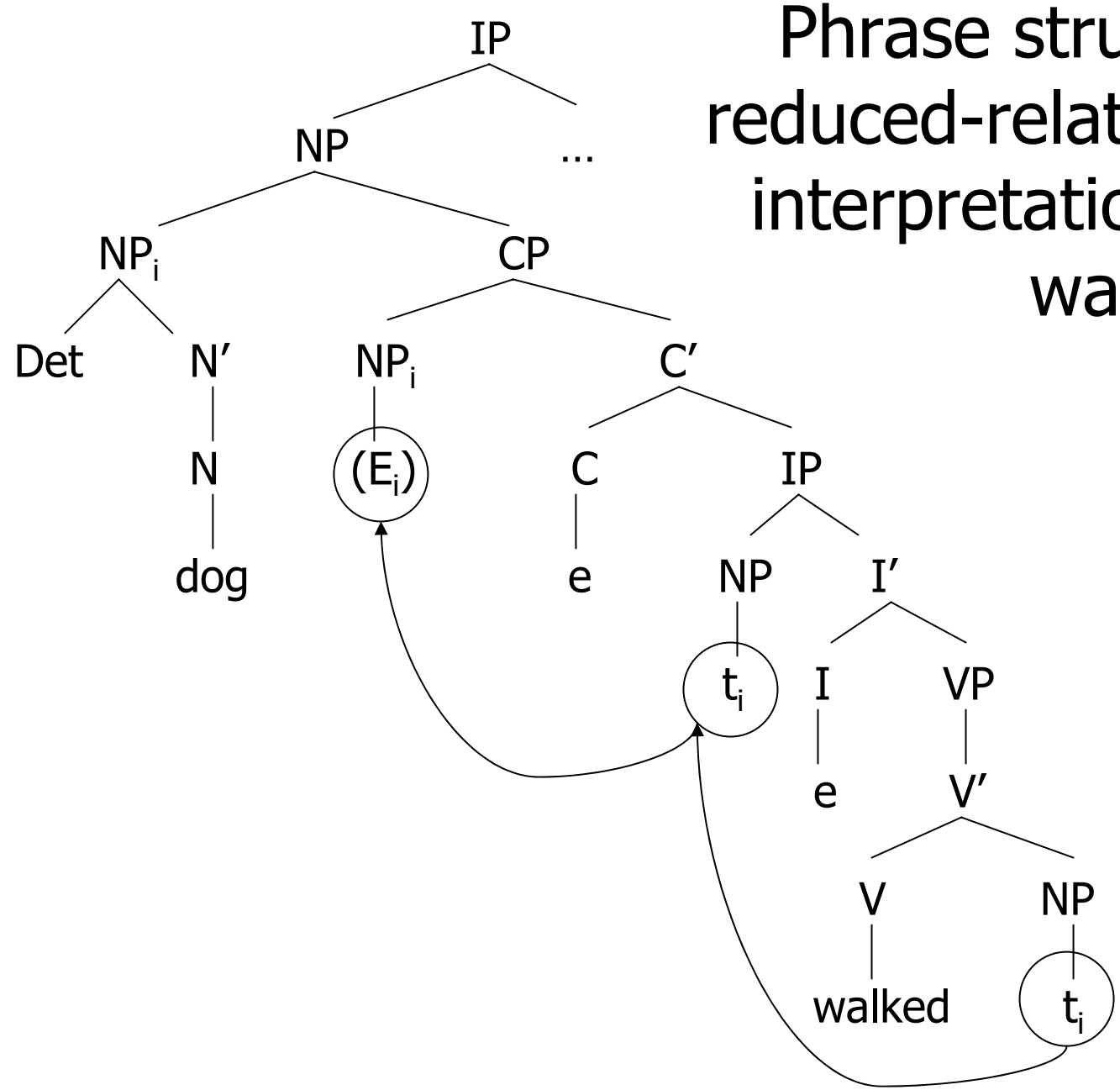
Phrase structure for the main-verb (MV) interpretation of "the dog walked"



Phrase structure for the relativ clause “the dog that was walked”



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



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.

Methodological issue:

All comparisons need appropriate controls.

Most published experimental work post-1990 includes good controls. Beware pre-1990.

Example of a problematic comparison (Frazier, 1979): NP/S ambiguity, reported in Frazier & Clifton (1995):

NP disambiguation: Sally was relieved when she found out the answer to the physics problem.

S disambiguation: Sally found out the answer to the physics problem was in the back of the book.

Methodological issue:

When investigating ambiguity, it is best to compare an ambiguous item with an unambiguous control.

NP disambiguation: Sally was relieved when she found out the answer to the physics problem.

S disambiguation: Sally found out the answer to the physics problem was in the back of the book.

Claim: S disambiguation harder than NP disambiguation. But there are no unambiguous controls. And the lexical material in the comparison pair is quite different.

Better comparison:

Sally found out **that** the answer to the physics problem was in the back of the book.

Methodological issue:

If there is no way to construct an unambiguous control, then control (a) word frequencies; (b) plausibility.

PP attachment:

The older campers questioned John's authority over the (group / summer) but they came to respect him after the big campfire in August.

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.

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

Graph removed for copyright reasons.

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

Graph removed for copyright reasons.

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.

How to resolve Ferreira & Clifton's (1986) observed data pattern?

Why are people slow at the ambiguous verb "examined / towed"?

If they are following the RR reading (for the "evidence examined" items, also used by TTG), then no difficulty is expected here.

If they are following the MV reading (for the "car towed" items), then no difficulty is expected either.

MacDonald, Pearlmutter & Seidenberg (1994)

Syntactic ambiguity resolution is just the same as lexical ambiguity resolution.

Duffy, Morris & Rayner (1988): Lexical ambiguity resolution.

Duffy, Morris & Rayner (1988): Lexical ambiguity resolution

Null contexts:

Equi-biased words (“pitcher”: baseball = jug) are read more slowly than biased words (“port”: harbor > wine).

Interpretation: parallel lexical access: causes some load: competition between readings

Duffy, Morris & Rayner (1988): Lexical ambiguity resolution

Biased contexts:

A biased context allows an equi-biased word to be read more quickly.

A biased context favoring the frequent reading of a biased word (e.g., the harbor sense of "port") does not speed RTs.

A biased context favoring the infrequent reading of a biased word (e.g., the wine sense of "port") slows RTs.

No context can get rid of the RTs coming from competition with the high-frequency alternative.

Conclusion: Frequency is a very important source of information.

Application of lexical ambiguity resolution to syntactic ambiguity resolution

People may be going slow on the verb in the "car towed" items in Ferreira & Clifton's study because there are two reasonable interpretations competing here:

Plausibility / Animacy information pushes somewhat toward the RR interpretation;

But the MV reading is much more frequent, and it is not completely implausible.

MacDonald, Pearlmutter & Seidenberg (1994)

- Meta-analysis of the items in 12 different studies which investigated the MV/RR ambiguity:

- 4 studies found no context effects (e.g., Ferreira & Clifton, 1986)

- 8 studies found context effects (e.g., Trueswell, Tanenhaus & Garnsey, 1994)

MacDonald, Pearlmutter & Seidenberg (1994)

- Meta-analysis of the items in 12 different studies which investigated the MV/RR ambiguity:
- The items in the studies that found no context effects had a mean past-participle frequency of 49.7%
- The items in the studies that found context effects had a mean past-participle frequency of 63.0%

This was a reliable difference ($p < .05$).

Trueswell (1996)

A direct test of the frequency hypothesis with respect to the MV/RR ambiguity

Experiment 1:

subjects: poor agents, good patients

manipulate the frequency of the past participle reading: high vs low past-part (high-pp / low-pp)

Prediction: no ambiguity effect for the high-pp items, but there should be an ambiguity effect for the low-pp items

Trueswell (1996)

Experiment 1 materials:

high past-participle frequency (high-pp):

The child (that was) adopted by the couple was happy to have a home.

low past-participle frequency (low-pp):

The audience (that was) entertained by the comedian left in high spirits.

Trueswell (1996)

Graph removed for copyright reasons.

Trueswell (1996)

Experiment 2:

subjects: good agents, good patients

manipulate the frequency of the past participle reading

Prediction: ambiguity effects for both the high-pp and low-pp items, maybe a bigger effect in the low-pp items

Trueswell (1996)

Experiment 2 materials:

high past-participle frequency (high-pp):

The person (that was) adopted by the couple was happy to have a home.

low past-participle frequency (low-pp):

The manager (that was) entertained by the comedian left in high spirits.

Trueswell (1996)

Graph removed for copyright reasons.

Constraint-satisfaction theories

MPS claim: The resolution of other ambiguities can be explained in terms of lexical frequency differences also.

The "constraint-satisfaction" theory is more of a framework than a theory: It provides a way to talk about ambiguity resolution. No one has yet provided much of a theory of (1) what the constraints are; and (2) how they interact.

Key differences between the garden-path theory and constraint satisfaction theories

A major difference is with respect to the time course of information: **Modularity**

Contrary to the garden-path theory, constraint-satisfaction theories predict immediate use of information sources other than syntax.

A response from people working in the garden-path theory: The evidence shows that other sources of information are used fast, but not necessarily first. Reanalysis is fast.

The constraint-satisfaction theories are more satisfactory here, because of parsimony.

Key differences between the garden-path theory and constraint satisfaction theories

Other differences:

GP theory is serial; CS theories are usually parallel.

GP theory includes the principles of Minimal Attachment and Late Closure as syntactic working memory principles. CS theories vary here.

Key differences between the garden-path theory and constraint satisfaction theories

A key prediction that separates the two kinds of theories (modularity):

CS theories predict that there should be circumstances under which the syntactically dominant reading gives rise to processing difficulty.

The garden-path theory predicts that such a situation is impossible. (Frazier & Clifton, 1995).