Sentence Processing & Production

BCS 152

October 31 2018

Homework 3 Reminder!!!

- Due **today at 11:59pm**

- Read instructions carefully & submit all of the required documents

Frequency and Predictability in Sentence Processing

I’m going to show you some strings of letters. Tell me as fast as you can if it’s an English word or not.

apple

cucumber

lesick

orange

durian

Frequency tends to ease processing

- People are faster to identify ‘apple’ and ‘orange’ as a word than ‘cucumber’ and ‘durian’

Now I’m going to show you pairs of words. Tell me as fast as you can if it’s a valid English phrase or not.

Examples:

alcoholic beverages (valid phrase)

book teach (not a valid phrase)

stainless steel

psychic nephew

grass green

armchair linguist

- Frequency effect extends to *phrases* as well
- Faster to identify ‘stainless steel’ than ‘psychic nephew’
- Can frequency give us insight into how people process sentences?

From last time...

The horse raced past the barn fell.

- *race* occurs as a main verb more frequently than as a past participle
- *horse* is most often the agent *racing* actions, not the theme
- The constraints are biased against reduced relative
- **What happens if the biases (subject plus verb) go the other way?**

Lexical biases

The salmon released in the stream spawned

The landmine buried in the sand exploded

- GP effect completely goes away for these sentences!
- Evidence **against** two-stage model – more than just syntactic information is considered

Trueswell & Tanenhaus (1994)

- Comprehenders use the *frequency* of a structure during processing

- People simply don't say things like '*The horse raced past the barn fell'* that often!

What is the nature of the frequency effect in sentence processing?

Expectation-Based Syntactic Comprehension (Levy, 2008)

Roger Levy

- The processing difficulty of a word in a sentence is proportional to how probable it is given the prior context
- Surprisal = -log(p(word | context))
  - as probability gets lower, surprisal gets higher (don't worry about the math!)

How do we know the probabilities?

- Train a simple computational model on a large corpus of text to learn:
- Given some *context*, what is the *probability* of encountering a particular word?
  - P(fell | The horse raced past the barn)

What is context?

- In principle: anything that could be relevant
  - e.g., factors that we saw affect parsing last week
  - visual context, semantics, etc

Put the apple on the towel...

P(‘into’ | two apples) > P(‘into’| one apple)

modifier interpretation

What is context?

- In practice: it's hard to quantify these rich contexts
  - We don't have good models of semantics
  - We don't know what the visual context was with just a text corpus
- Instead, use simpler context like syntactic structure or words
- 2 types of common models:
  - Probabilistic Context-Free Grammar (PCFG)
  - *n-gram* model

PCFGs

- Like phrase structure rules, but with probabilities attached to them (learn the probabilities from corpus data)

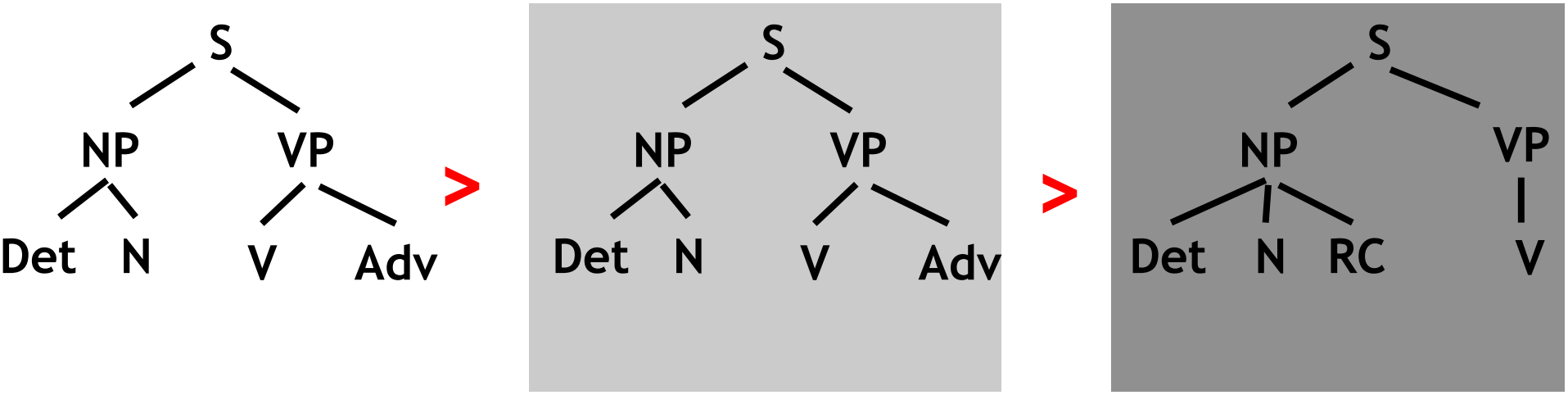
NP → Det N	P = 0.3
NP → Det N PP	P = 0.1
NP → Det Adj N	P = 0.1
…	
N → ‘dog’	P = 0.001
N → ‘cucumber’	P = 0.00001

N-gram Models

- n-gram = n words
- bigram (2): P(fell | barn)
- trigram (3): P(fell | the barn)
- 4-gram: P(fell | past the barn)
- 5-gram: P(fell | raced past the barn)
- ...

Expectation-Based Comprehension Theory

The horse raced past the barn.....



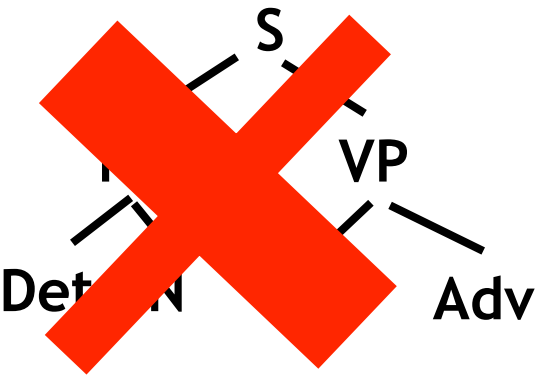
'The horse raced past the barn yesterday'

'The horse raced past the barn quickly'

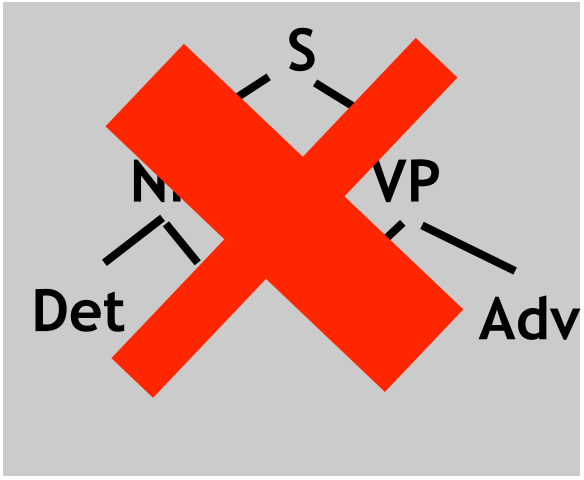
'The horse raced past the barn fell'

Decreasing Probability

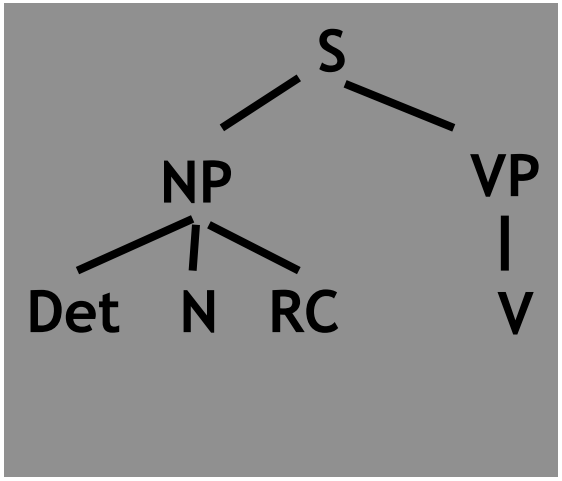
The horse raced past the barn fell



‘The horse raced past the barn yesterday’



‘The horse raced past the barn quickly’



‘The horse raced past the barn fell’

Decreasing Probability

- The less probable the structure, the more costly it is to re-rank your parse hypotheses

The baby read the book slept.

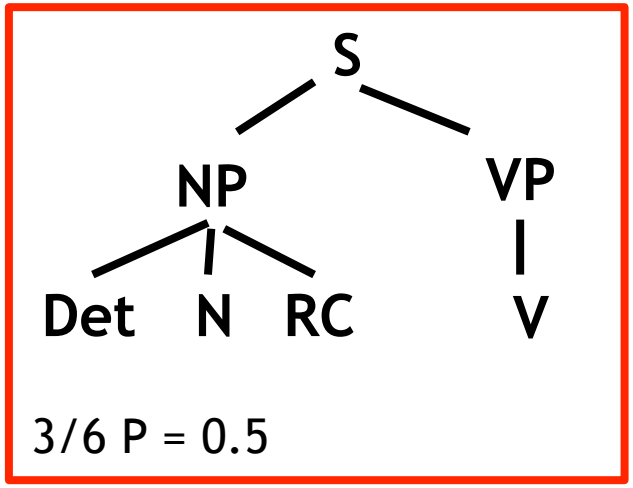
The man read the pamphlet.

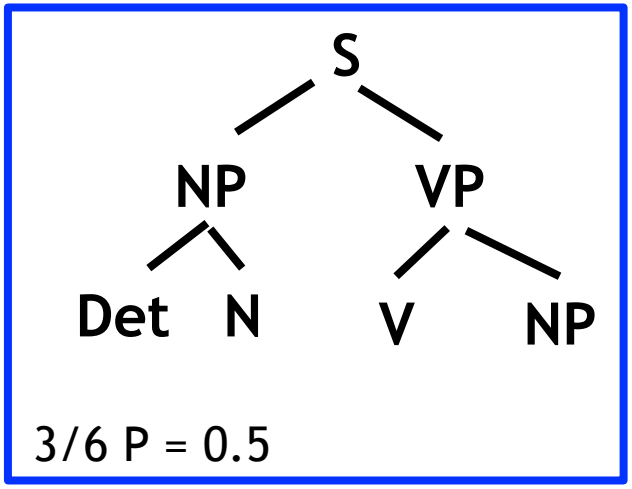
The baby read the book cried.

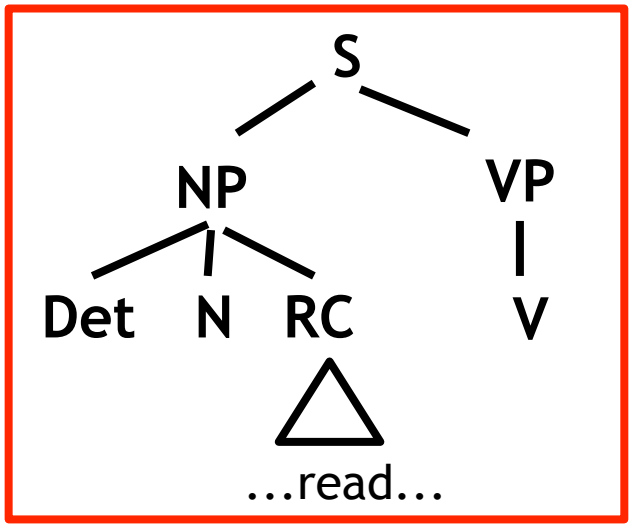
The jockey raced the horse.

The girl raced the marathon.

The horse raced past the barn fell.

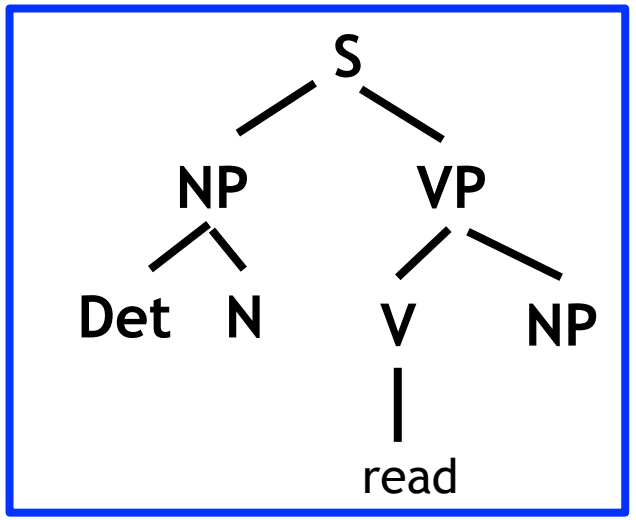






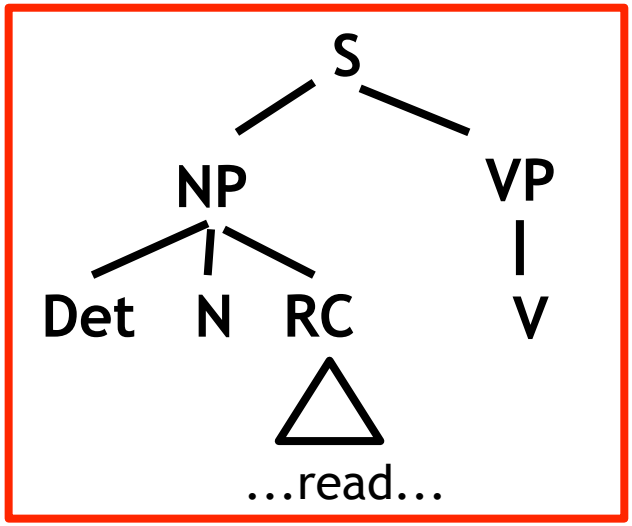
2/3 sentences

P = 0.6667

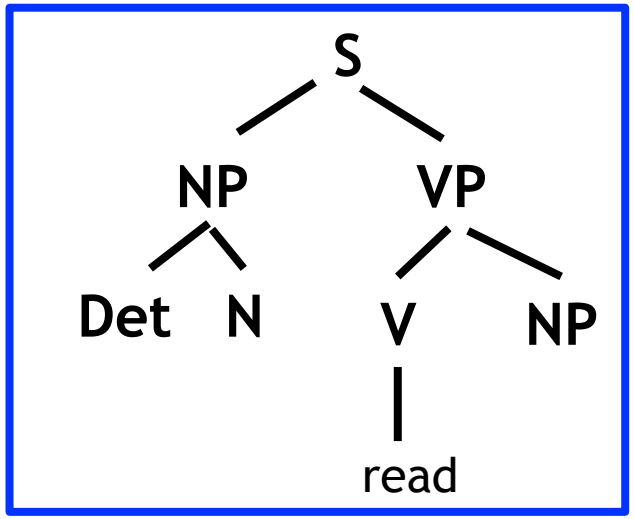


1/3 sentences

P = 0.3333

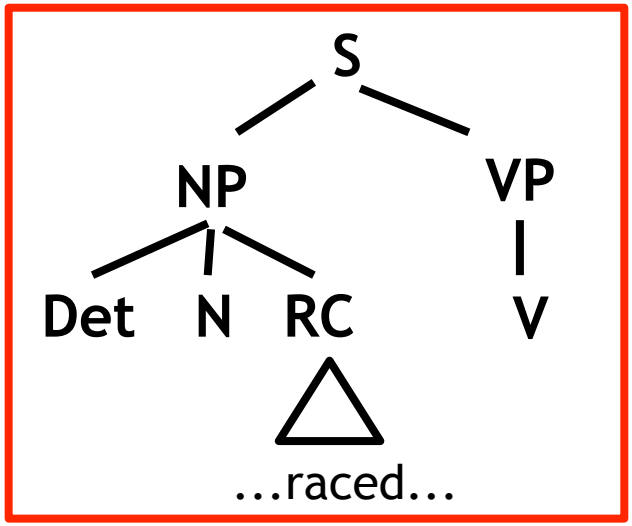


2/3 sentences
P = 0.6667

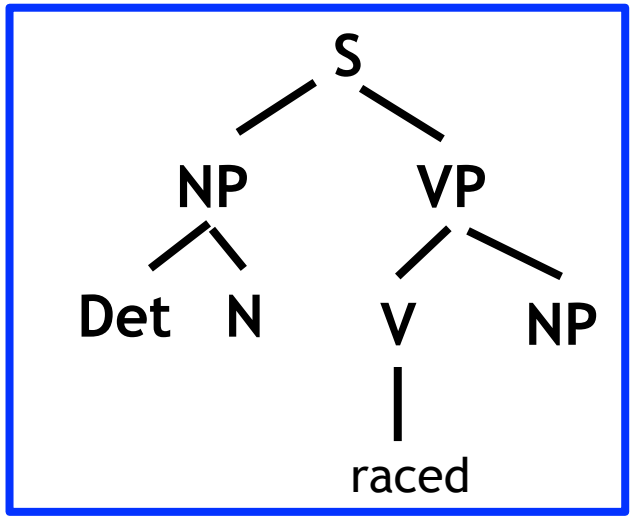


1/3 sentences
P = 0.3333

P(‘slept’ | ‘…read…’) = 0.6667



1/3 sentences
P = 0.3333



2/3 sentences
P = 0.6667

P(‘fell’ | ‘…raced…’) = 0.333

Expectation-Based Theory explains lack of GP effect

The salmon released in the stream spawned

The landmine buried in the sand exploded

Trueswell & Tanenhaus (1994)

Another case: distance from subject to verb

The player that the coach met bought the house.

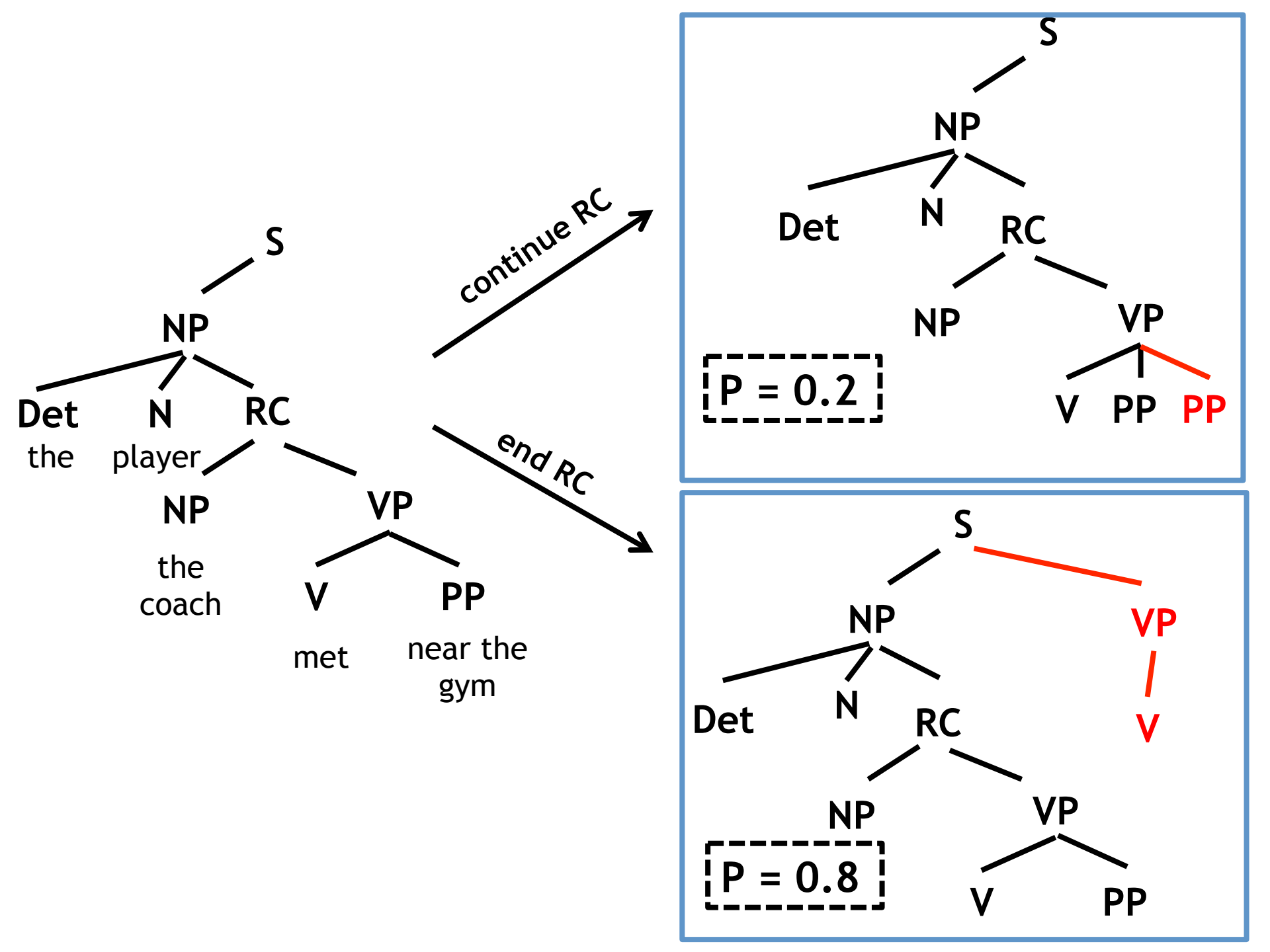
The player that the coach met by the gym near the river at 8 o'clock bought the house.

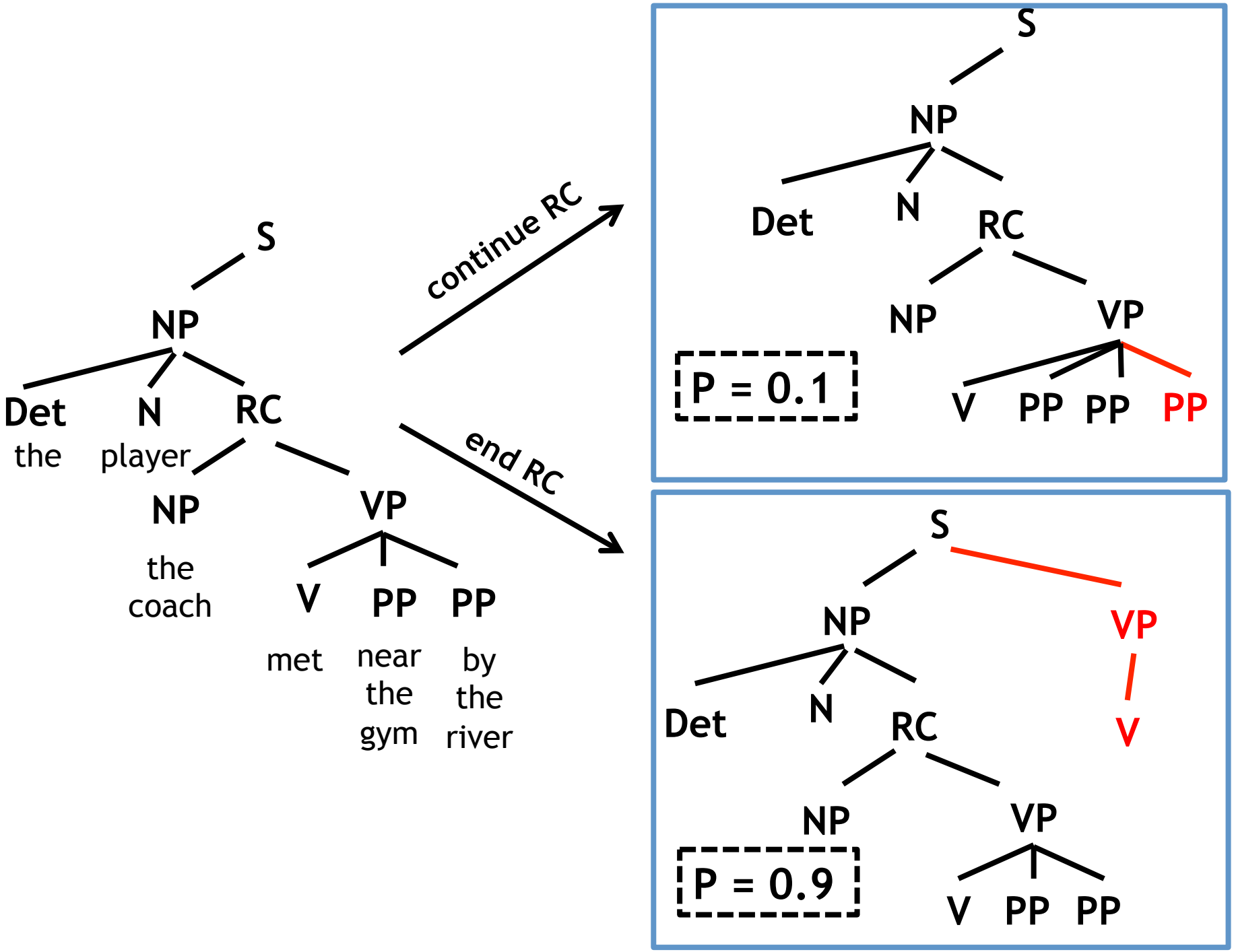
Intuition: as RC gets longer, processing 'bought' might be harder because it's farther away from its subject

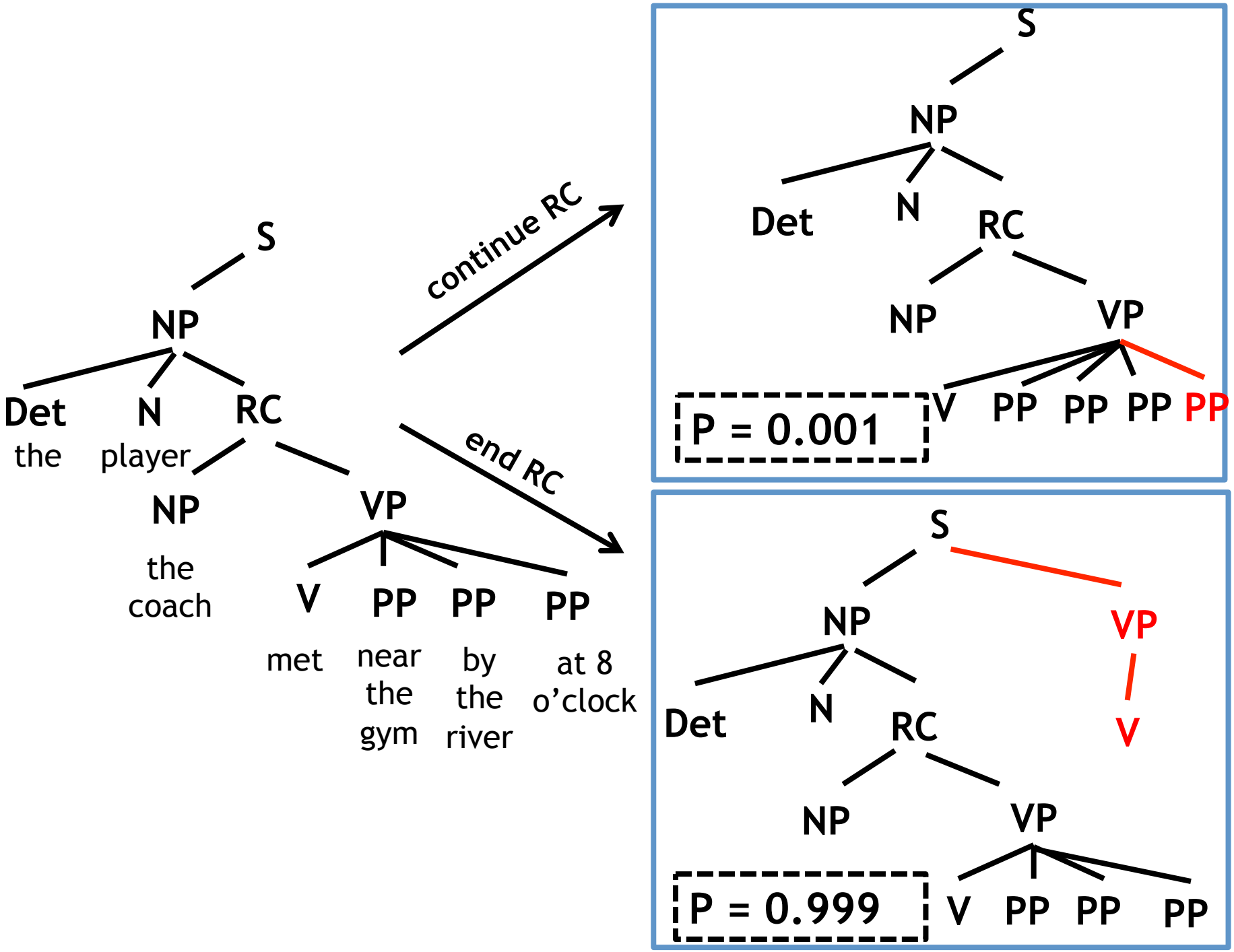
The player [RC that the coach met [PP at 8 o’clock]] bought the house.

The player [RC that the coach met [PP by the river] [PP at 8 o’clock]] bought the house.

The player [RC that the coach met [PP near the gym] [PP by the river] [PP at 8 o’clock]] bought the house.


S
NP
Det
the
N
player
RC
NP
the
coach
VP
V
met
PP
near the
gym
continue RC
end RC
S
NP
Det
N
RC
NP
VP
V
PP
PP
P = 0.2
S
NP
VP
Det
N
RC
V
NP
VP
P = 0.8
V
PP


S
NP
Det
the
N
player
RC
NP
the
coach
VP
V
met
PP
near
the
gym
PP
by
the
river
continue RC
end RC
S
NP
Det
N
RC
NP
VP
P = 0.1
V
PP
PP
PP
S
NP
VP
Det
N
RC
V
NP
VP
P = 0.9
V
PP
PP


S
NP
Det
the
N
player
RC
NP
the
coach
VP
V
met
PP
near
the
gym
PP
by
the
river
PP
at 8
o'clock
continue RC
end RC
S
NP
Det
N
RC
NP
VP
P = 0.001
V
PP
PP
PP
PP
S
NP
Det
N
RC
NP
VP
P = 0.999
V
PP
PP
PP
VP
V

Surprisal predicts processing the verb should actually be ***easier*** the longer the RC gets!

As more PPs are encountered, the probability that the RC will end gets higher & higher

Surprisal prediction wins out

	Number of PPs intervening between embedded and matrix verb		
	1 PP	2 PPs	3 PPs
DLT prediction	Easier	Harder	Hardest
Surprisal	13.87	13.54	13.40
Mean reading time (ms)	510 ± 34	410 ± 21	394 ± 16

Reading time at 'bought' decreases with more PPs added to relative clause

Another example

- Verb-specific differences in structural preferences
  - DO-biased vs. PO-biased verbs
  - ‘I gave my friend the book’ > ‘I gave the book to my friend’
  - ‘I kicked the ball to my teammate’ > ‘I kicked my teammate the ball’
- Surprisal accounts for difference in processing times

Why is expectation-based theory such a compelling account?

- Low-probability events are by definition less likely to happen
- By processing high-probability things very fast, we optimize our processing times overall

Do we need the assumption of syntax for the theory to make these predictions?

- No! Simple n-gram models that don't know anything about syntactic structure produce similar results

- Another appeal of theory: don't need to assume any particular syntactic structure exists in people's minds

What happens when our prediction is wrong?

‘The horse raced past the barn fell.’

Higher probability for linguists than non-linguists!

Do we always process it so slow?

Fine et al. (2013)

- Give subjects a bunch of exposure to GP sentences
- Does GP effect reduce over time?

Fine et al. (2013)

	Block 1	Block 2
Filler-First Group	**The girl kicked the ball to her teammate. (x16)**	**The horse raced past the barn fell. (x10)**
RC-First Group	**The horse raced past the barn fell. (x16)**	**The horse raced past the barn fell. (x10)**

In Block 2, does RC-First Group have smaller GP effect than Filler-First Group?

GP Effect

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

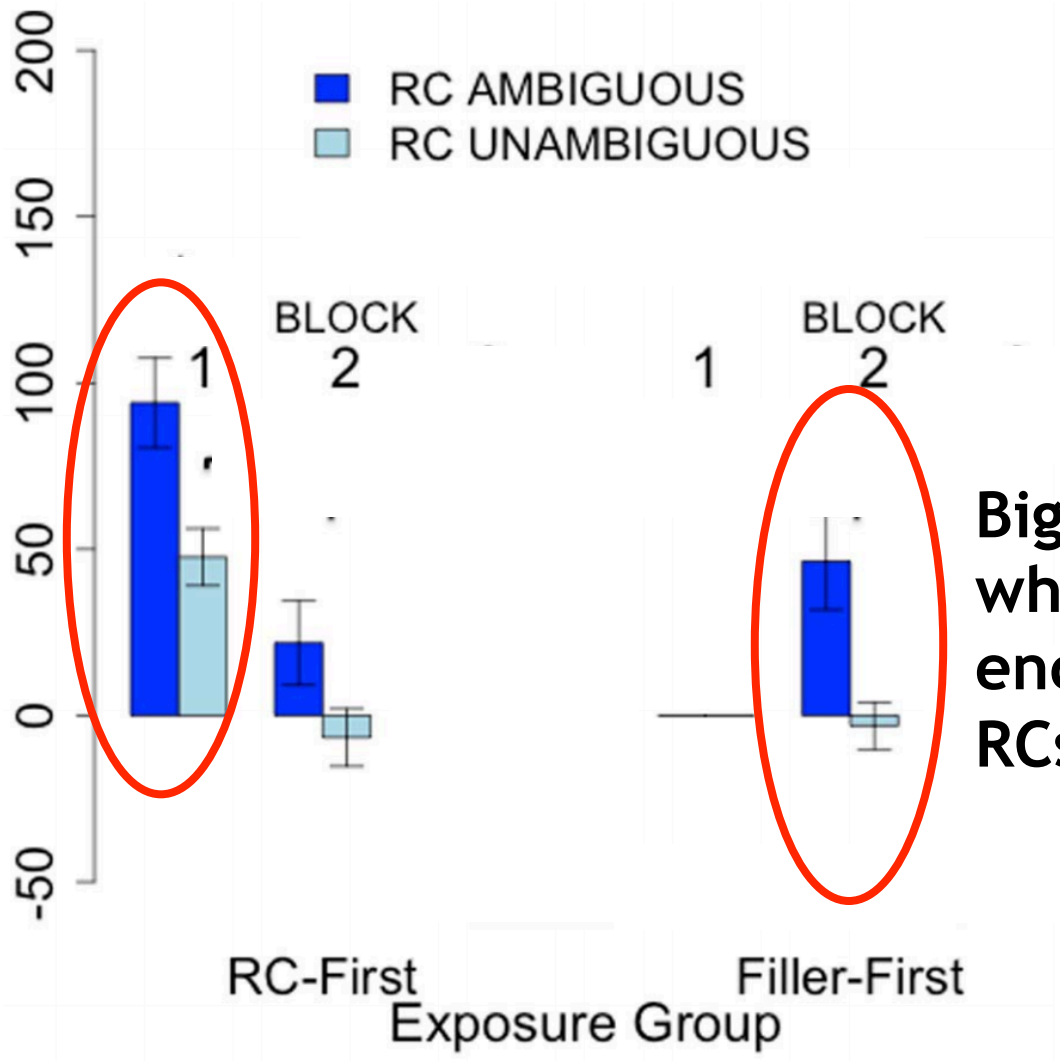
(ambiguous RC)

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

(unambiguous RC)

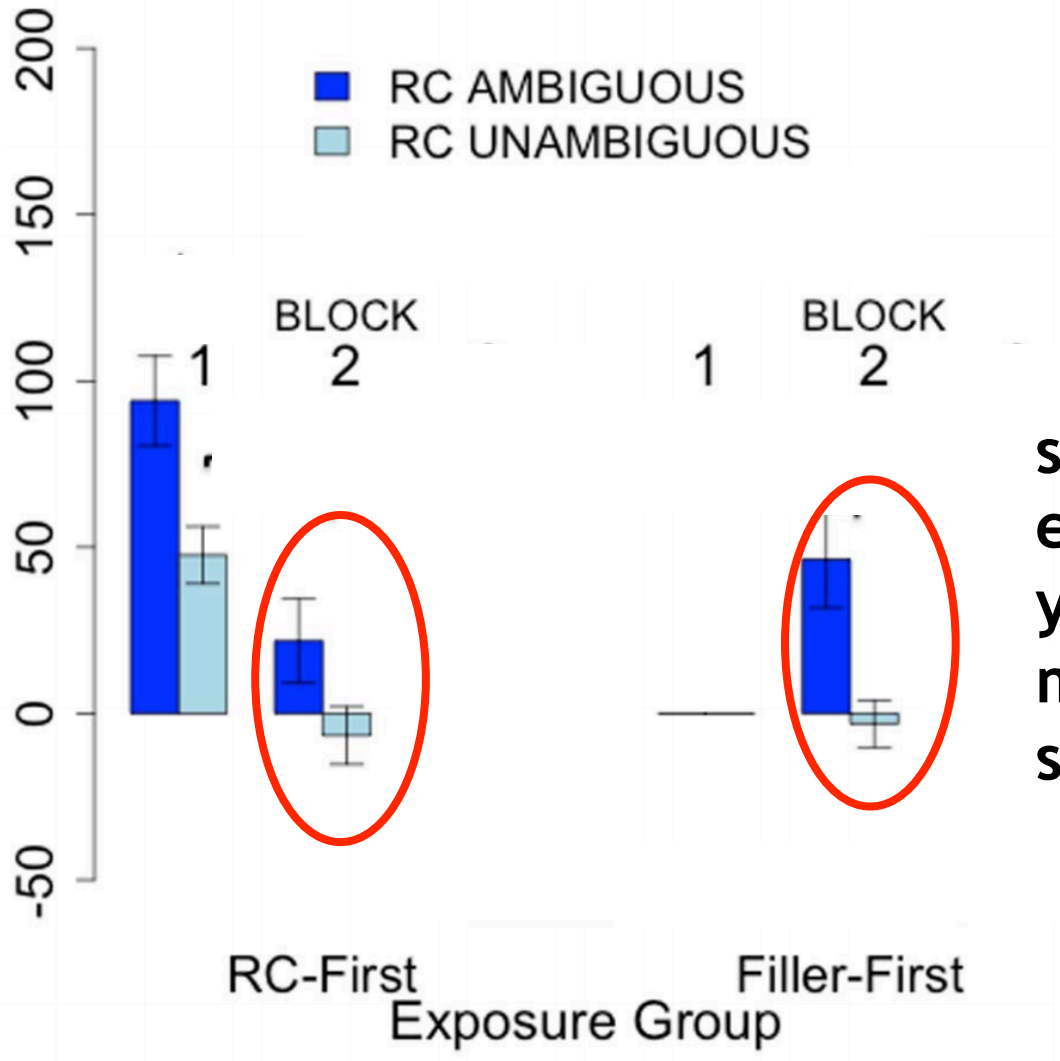
Unambiguous RT – Ambiguous RT = Garden-Path Effect

Fine et al. (2013)



Big GP effect when first encountering RCs

Fine et al. (2013)



smaller GP effect once you've seen many GP sentences

- Comprehenders continue to optimize their processing by *updating* probability of structural occurrence

- Even the hardest GP sentences can become easier if you hear them enough!

Some food for thought

Where is the expectation-based theory going?

- Studies like Levy (2008) can't account for many effects on processing we've seen

- Semantic context
- Visual context
- Discourse context

- If we can find a good way to model these types of information, does the expectation-based theory still stand?

Are RTs a good measure of sentence comprehension?

- Sentences that should be harder to understand can be ‘easier’ to process

The player that the coach met by the gym near the river at 8 o’clock bought the house.

- Can we trust RTs as a measure of comprehension difficulty? What exactly do they index?

Sentence Production

(1) Language production is *incremental* – we plan as we go along!

(2) What we say next is influenced by how cognitively available/accessible it is

(3) Do we produce things that are easy for our listeners to understand, or do we simply produce what's easy?

Incrementality

- Phonological encoding of a word takes place shortly before articulation (Dell, 1986; Garrett, 1975; Levelt, 1989)
  - **no buffering of whole sentences**
- So, when are words phonologically encoded (while talking)?
- Do speakers know how much time they have to phonologically encode the next word?

Griffin (2003)

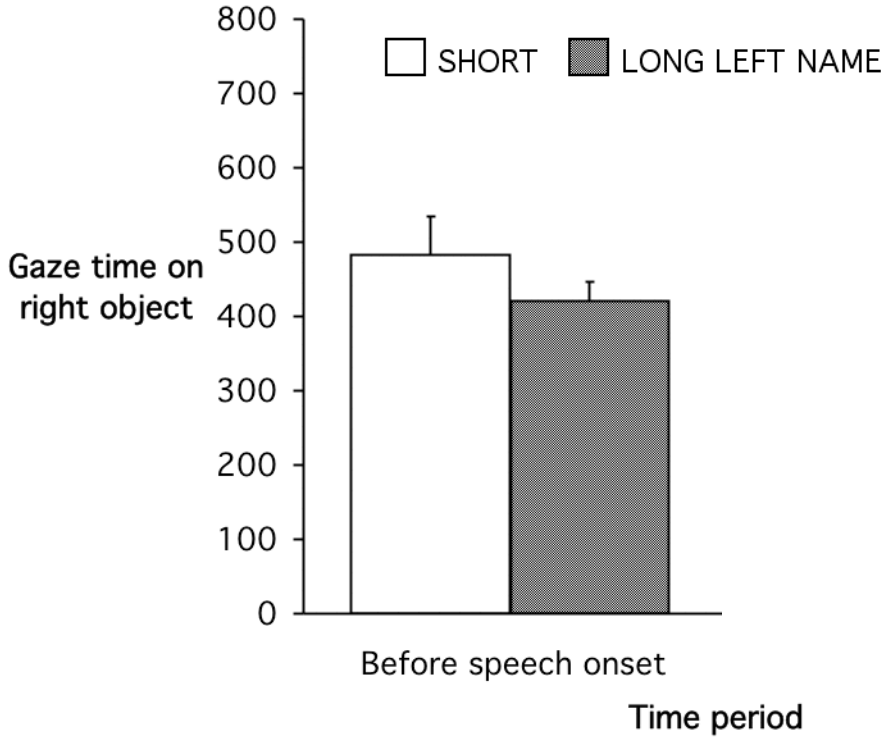
- How do speakers incrementally plan groups of words?

chef → tank

chandelier → tank

Griffin (2003)

- Speakers “know” how much time they will have to produce a word while they are articulating the next word



Avoiding suspension of speech

- Speakers dislike stopping and employ a variety of strategies to avoid it:
  - Reordering constituent

    She gave {him the key/ the key to him}

  - Inserting optional words where grammar permits

    I read a book (that) she wrote.

  - Slowing down speech rate
  - Inserting filled pauses and parentheticals (*uh*, *um*, *I mean*, *you know*)

- The longer the phrase you're trying to access, the longer the filler word you insert (Clark & Fox Tree, 2002)
  - 'The uh chandelier'
  - 'The um gold chandelier'
  - 'The uuumm big gold chandelier in the dining room'

Principle of Immediate Mention

(Ferreira and Dell, 2000)

"Production proceeds more efficiently if syntactic structures are used that permit quickly selected lemmas to be mentioned as soon as possible."

What's a 'quickly selected lemma'?

- Words that are more cognitively *available*
  - More frequent words
  - Shorter words
  - Previously mentioned ('given') words
  - More concrete/'imageable' words

Frequency

I handed the girl the icepick >> I handed the icepick to the girl

Length

salt and pepper >> pepper and salt

Previous mention (givenness)

[...] Roger [....] . I gave Roger the book. >>

[...] Roger [....] . I gave the book to Roger.

Concreteness/Imageability

There's more money in the economy this year >> The economy has more money this year.

- Evidence from naturalistic speech (sports commentating):

  That will bring [to the plate] [Barry Bonds]

… gives speaker extra time to check his scorecard and see who the next batter is!

Wasow (1997)

When we speak, do we keep in mind what will be easy for our listener?

- **Egocentric production:** speakers *only* say what is easiest for production, regardless of how easy it is for comprehension

- **Audience design:** speakers take their listener into account during production and try to reduce ambiguity

Avoiding Garden-Paths

Jane showed the letter to Mary to her mother. (Prepositional Object, PO)

Jane showed her mother the letter to Mary. (Double Object, DO)

Arnold et al. (2004)

Avoiding Garden-Paths

Participant reads sentence silently (neither PO or DO structure):

"Andy's note to Terry was read to the entire class by the teacher."

Prompted to answer: "What did the teacher do?"

Arnold et al. (2004)

- Measure how often they produce DO vs. PO
  - The teacher read Andy's note to Terry to the entire class (PO, garden-path)
  - The teacher read the entire class Andy's note to Terry (DO, not garden-path)
- Compare to condition that wouldn't produce garden-path in a PO construction
  - "Andy's note was read to the entire class by the teacher."
  - PO doesn't produce any garden path ("The teacher read Andy's note to the entire class.")

Arnold et al. (2004)

	No Garden Path
Prepositional object	27%
Double object	73%

- Speakers *don't* avoid producing a garden-path sentences!
- Supports egocentric theory of production

Arnold et al. (2004)

Another ambiguity

The coach knew you...
The coach knew you missed practice.

This is easily avoidable:
The coach knew *that* you missed practice.

Ferreira & Dell (2000): How do comprehension ease and availability trade off?

- Recall: when a word is *available*, want to produce it as soon as possible
- When word is less available, we can put off having to produce it by using filler words

What happens when a word is available, but a filler word reduces ambiguity for their listener?

I knew (that) I missed practice.

You knew (that) you missed practice.

- Repetition makes word available
- Structure is always unambiguous (*I knew I.)

You knew (that) I missed practice.

- Lack of repetition makes word less available
- Structure is unambiguous (*You knew I.)

I knew (that) you missed practice.

- Lack of repetition makes word less available
- Structure is ambiguous

I knew (that) I missed practice.

You knew (that) you missed practice.

Egocentric & Audience Design: very few *that* productions because structure is unambiguous and word is available

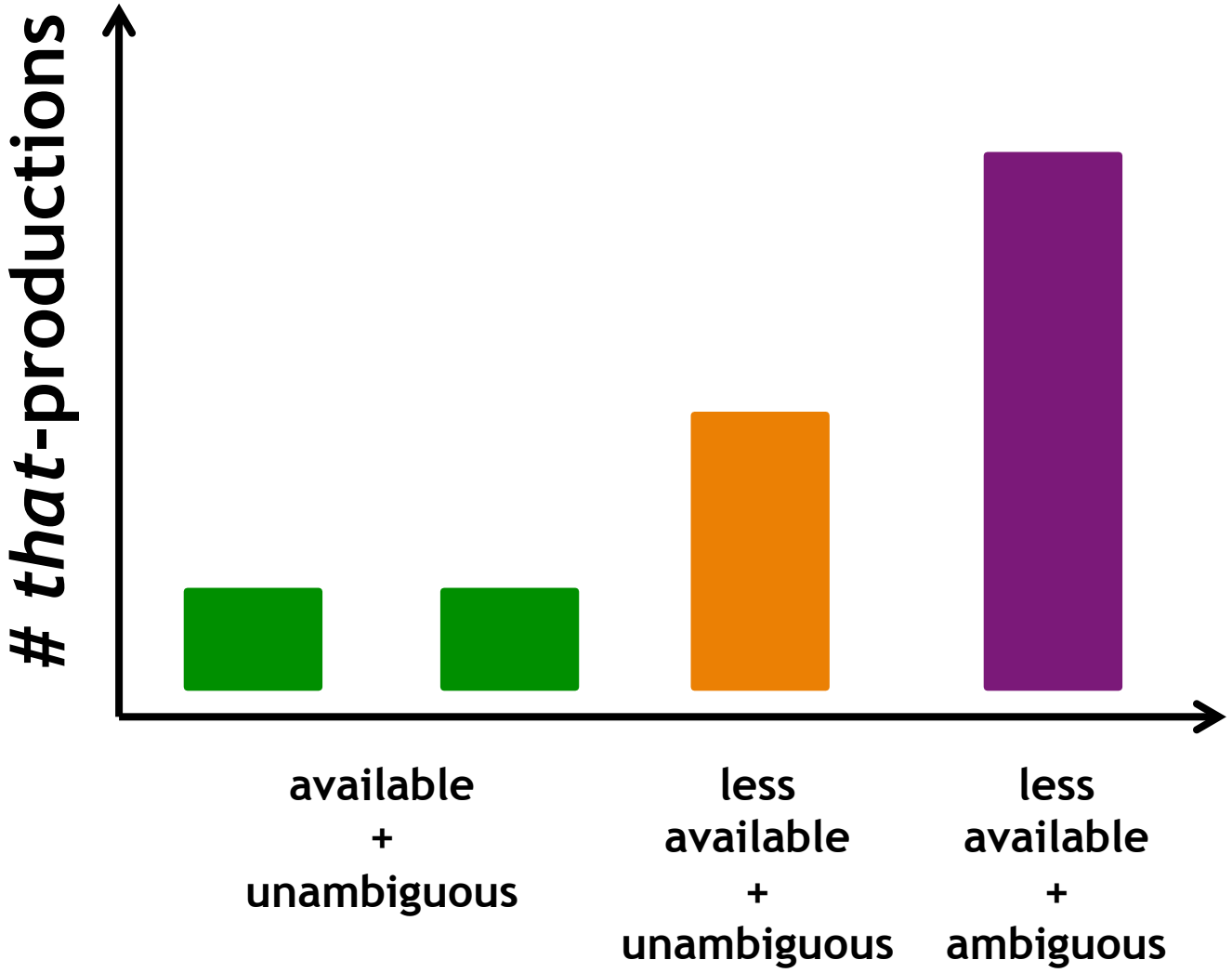
You knew (that) I missed practice.

Egocentric & Audience Design: more *that* productions because structure is unambiguous but word is less available

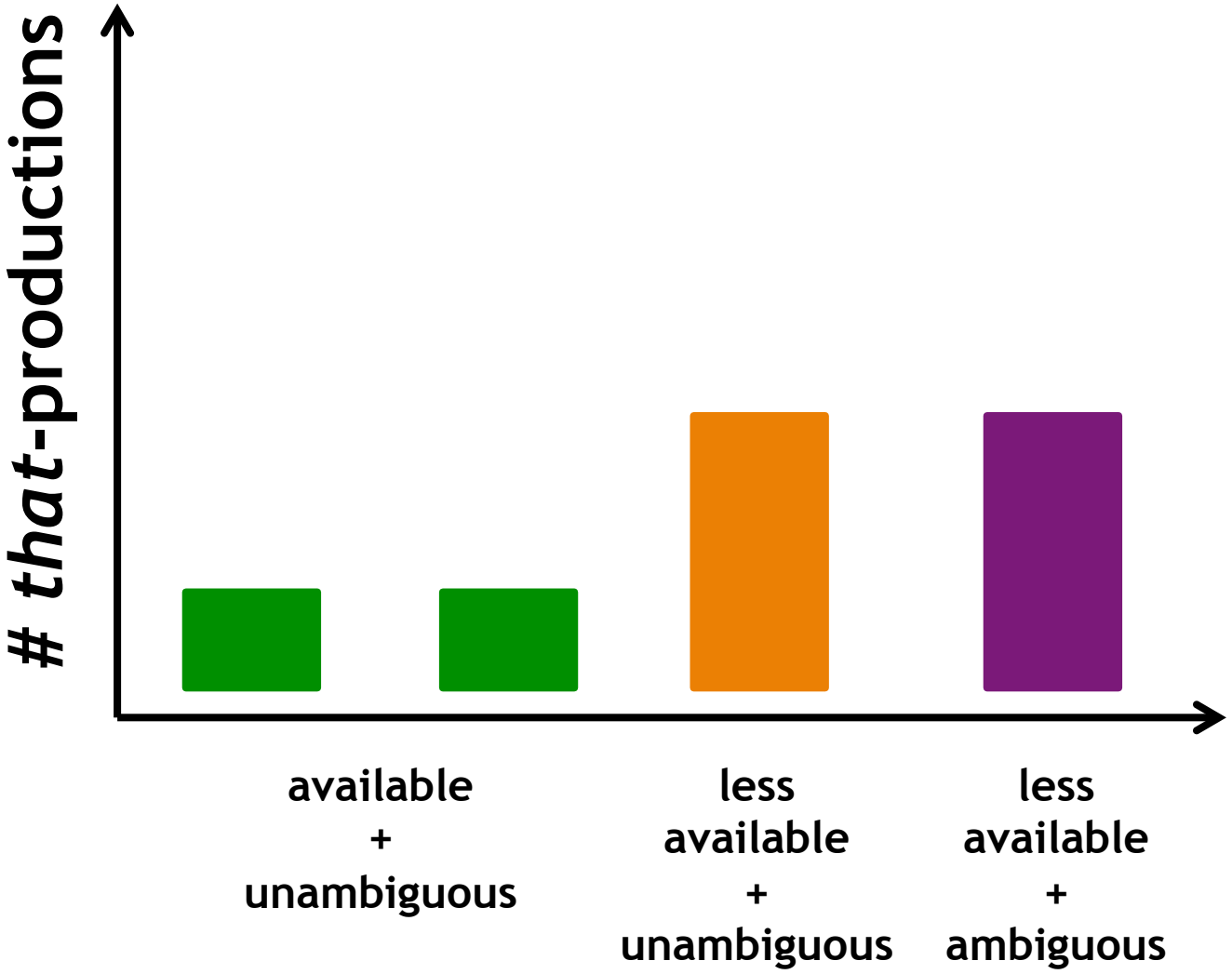
I knew (that) you missed practice.

Egocentric: as many *that* productions as above
Audience Design: even more *that* productions because structure is ambiguous

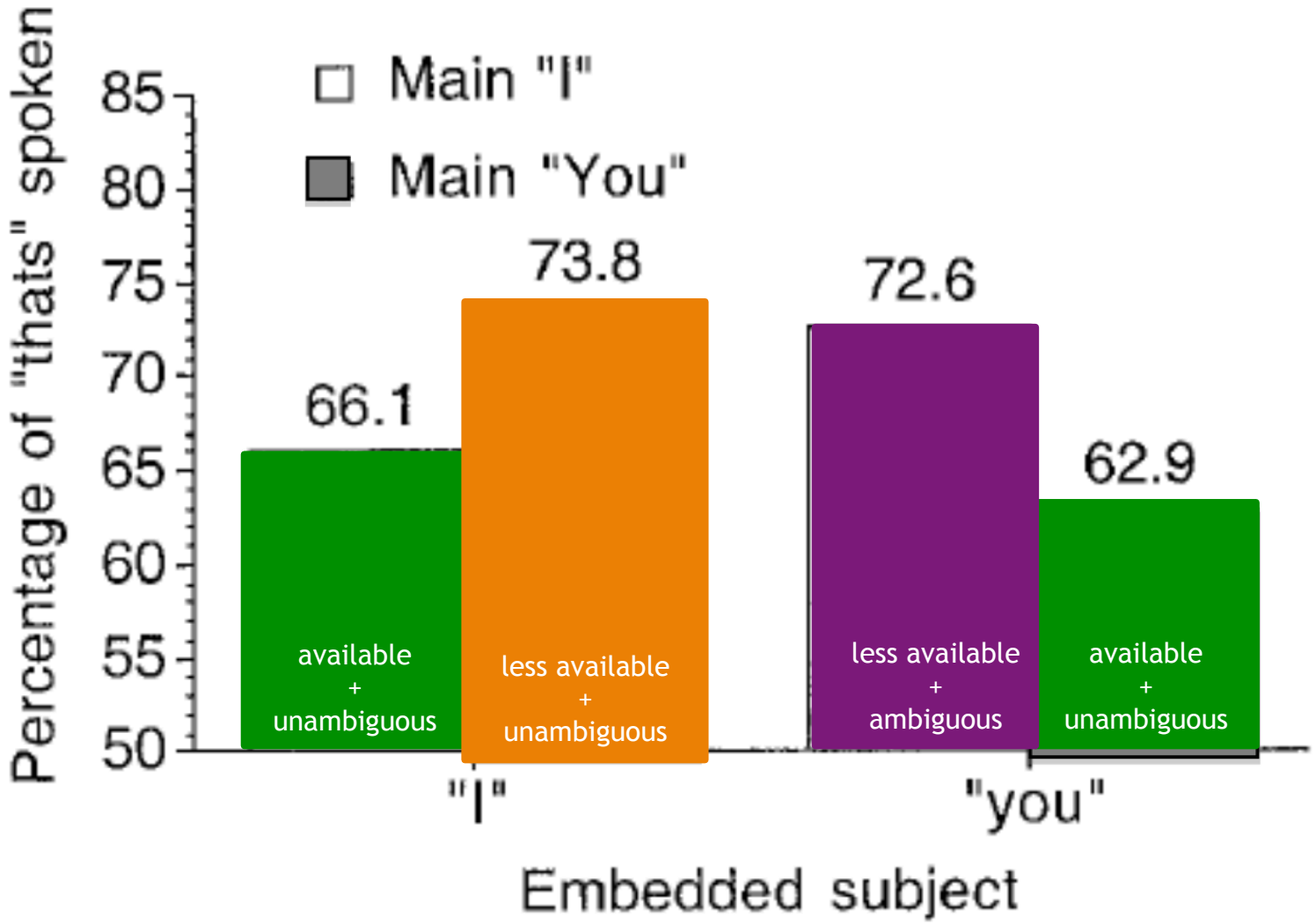
Audience Design Prediction



Egocentric Prediction



Ferreira & Dell (2000) Results



- Evidence for egocentric theory

- Speakers seem to be concerned with producing what's easy, not what is most useful for listeners