Representation of Linguistic Knowledge

COGS 200

October 21, 2014

Learning objectives

- Two ways of representing linguistic knowledge

1. Linguistic knowledge in the mind
   – Generative approach
   – Underlying/superficial structure, transformations, three components

- Interlude: rule-governed translation

2. Linguistic knowledge from input
   – Exemplar approach

Linguistic knowledge representation Part 1: A Generative approach

- **Grammar** is the finite set of rules that can specify/**generate** the infinite set of sentences.
- *Sentence* = string of symbols in the language.
  - "Symbols" = **past, *the*, NP, ~ing, syllable,** f**, [ ]**, etc.
- *Language* is the set of strings of symbols generated by a grammar.
- A speaker's knowledge of language consists of that grammar.

Note on linguistic knowledge "representation"

- A representation can be thought of as a surrogate for world knowledge
  - Don't have to have a dog IN your head to know what a dog is
- Representation of linguistic knowledge is NOT a surrogate
  - You DO have language in your head
  - What is in your head IS language

Knowledge of language

- A speaker's knowledge of language consists of his/her grammar.
- Knowledge of language **cannot** be knowledge of set of sentences.
- Why not?

Knowledge of language

- So, what does the grammar look like?
- Kind of like Lego, but with rules

The building blocks

- Morphemes and words
  - **past, plural**
  - *un~, interest, ~ing*
  - *telescope*, *see*, *in*, *the*, *red*
- Category labels
  - N V P Art Adj

The building blocks

- Morphemes and words
  - **past, plural**
  - *un~, interest, ~ing*
  - *telescope, see, in, the, nice*
- Category labels
  - N V P Art Adj
- Lexicon
  - Mental "dictionary"
  - Organizes words and morphemes according to their categories, etc.

Rules: phrase structure

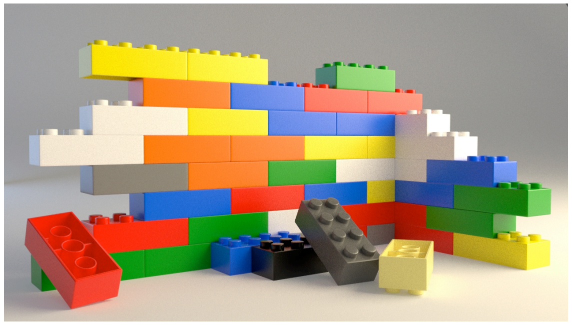
✓*I went to the store*, but **The to I store went.*

- S → NP VP
- NP → (Art) (Adj) N (PP)   [categories inside () are optional]
- VP → V (NP) (PP)
- PP → P NP

- Characterize acceptable patterns for sentences

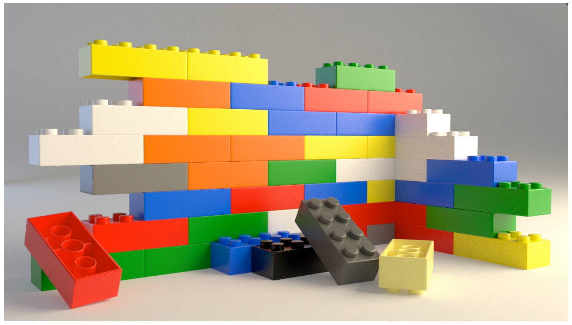
Phrase structure rules

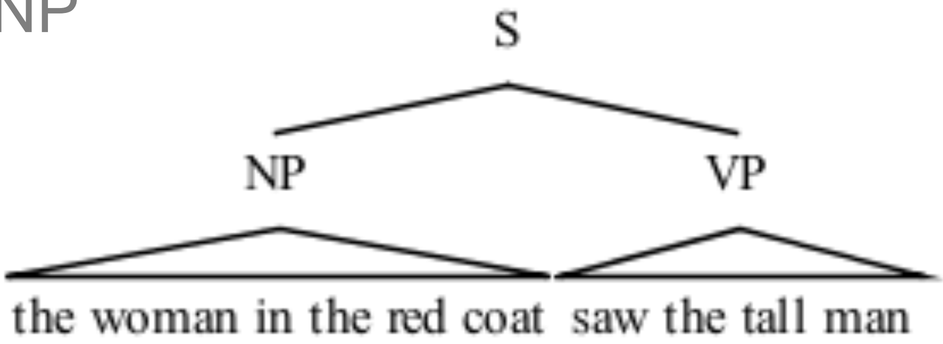
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP



Phrase structure rules

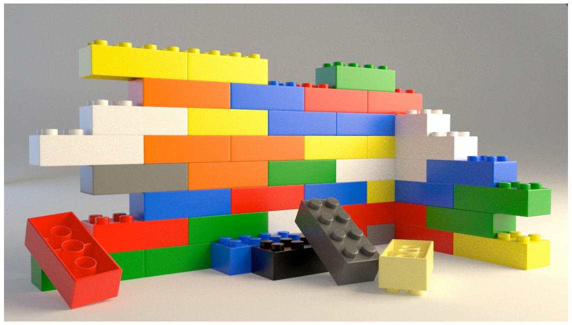
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

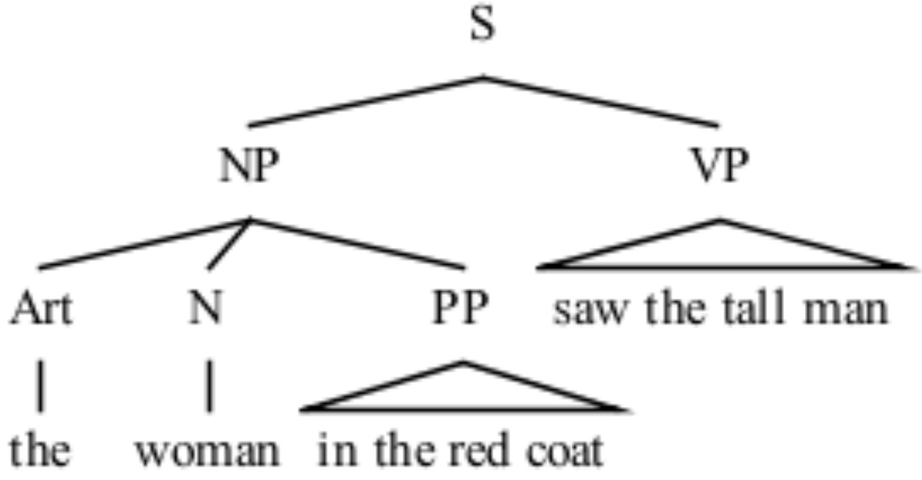




Phrase structure rules

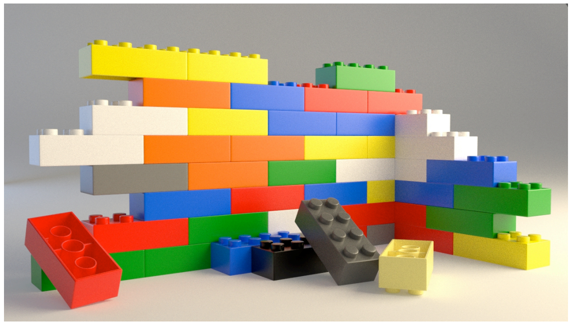
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

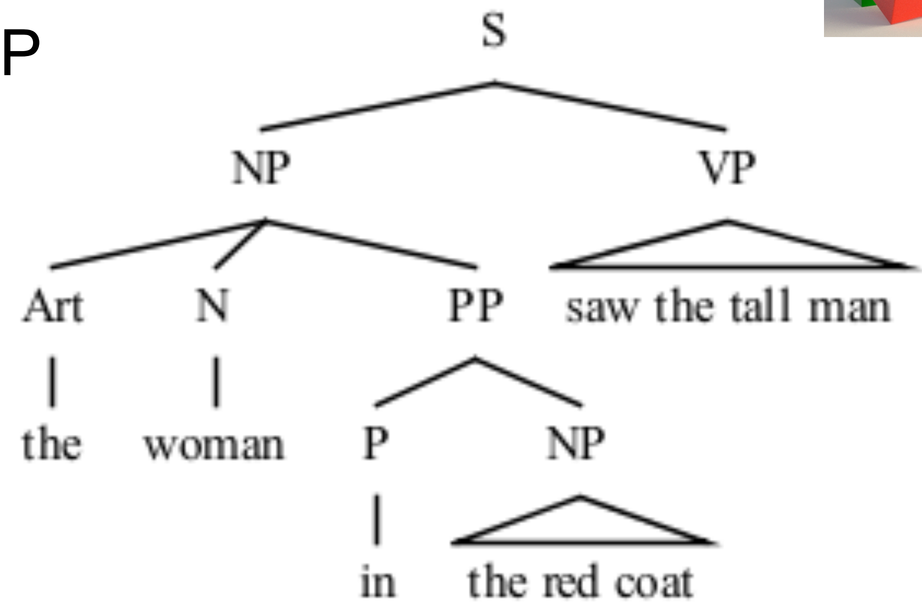




Phrase structure rules

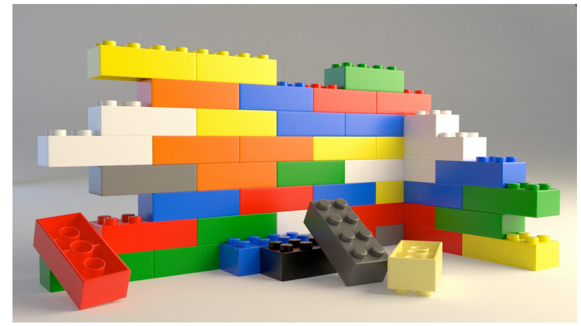
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

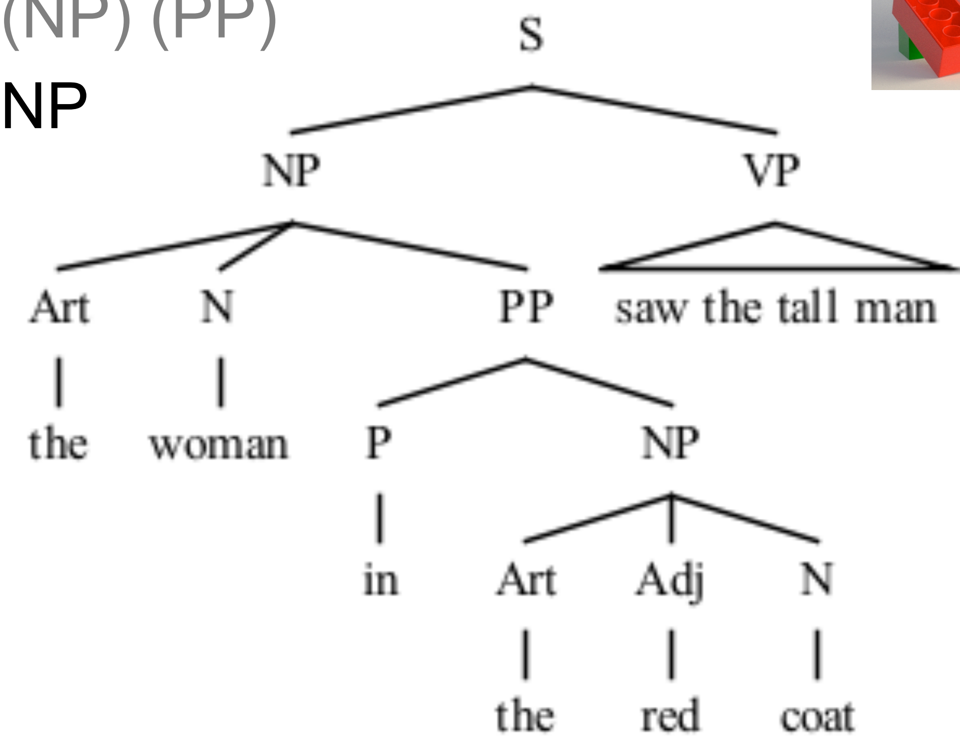




Phrase structure rules

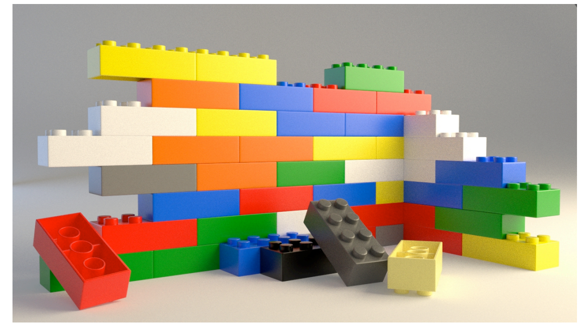
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

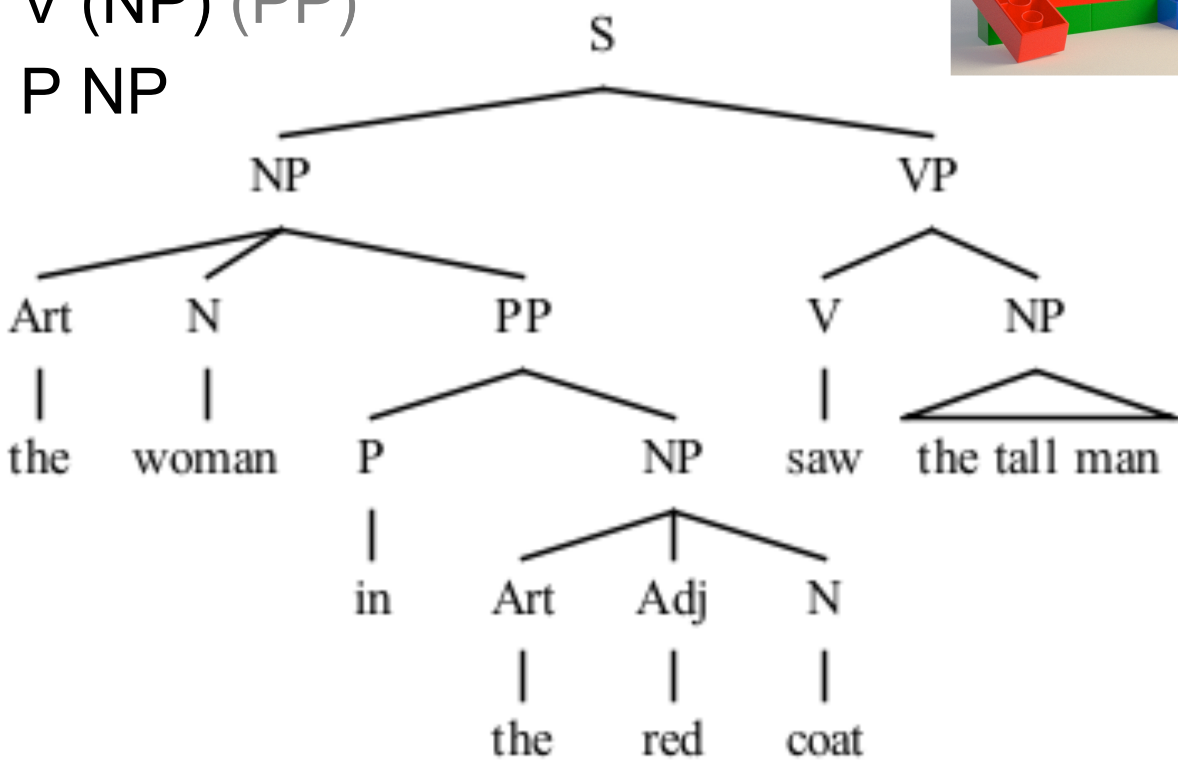




Phrase structure rules

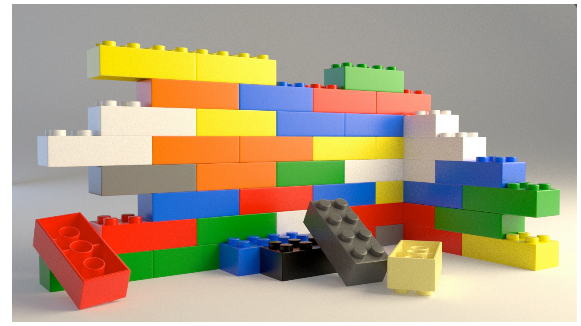
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

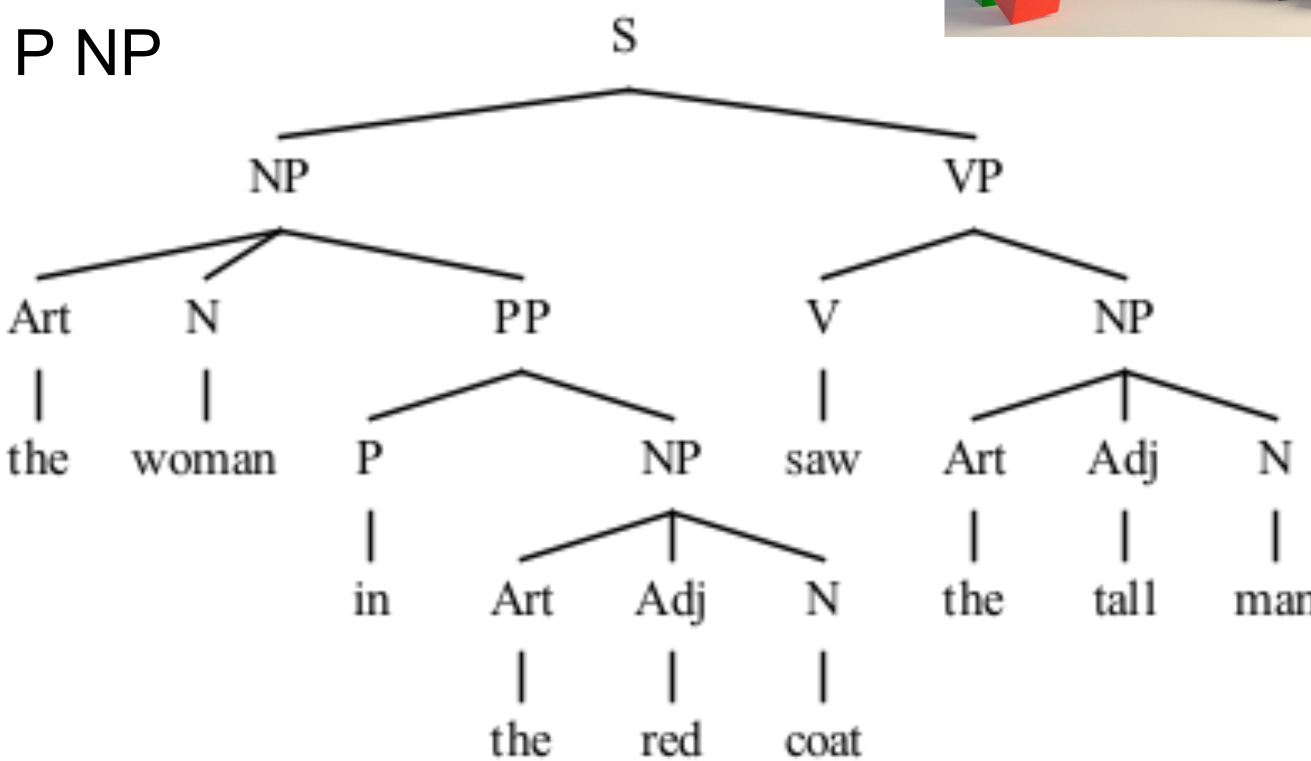




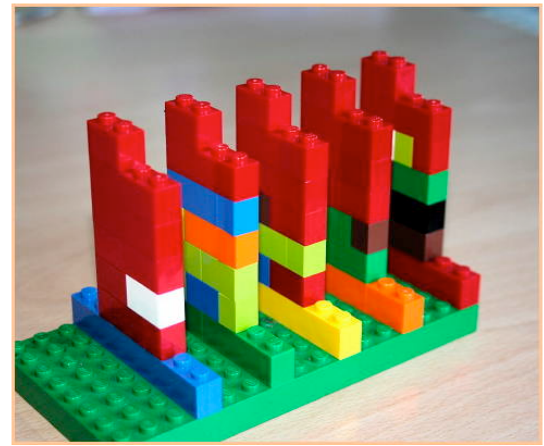
Phrase structure rules

- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

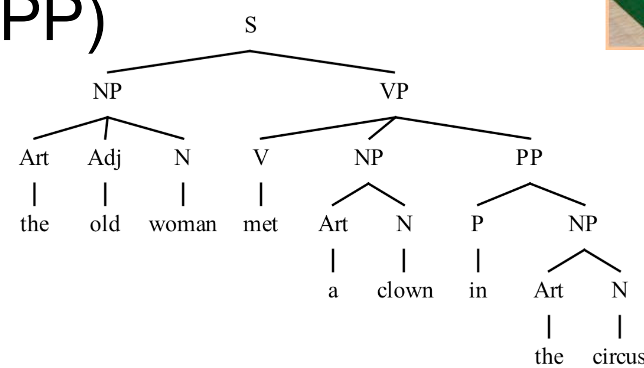


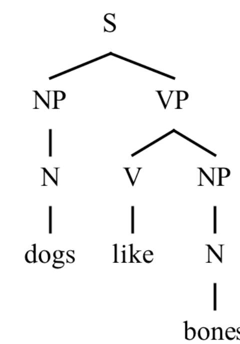


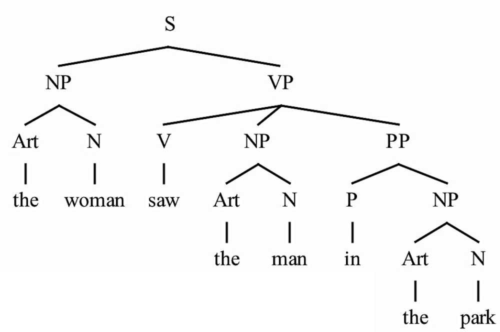
Phrase structure rules

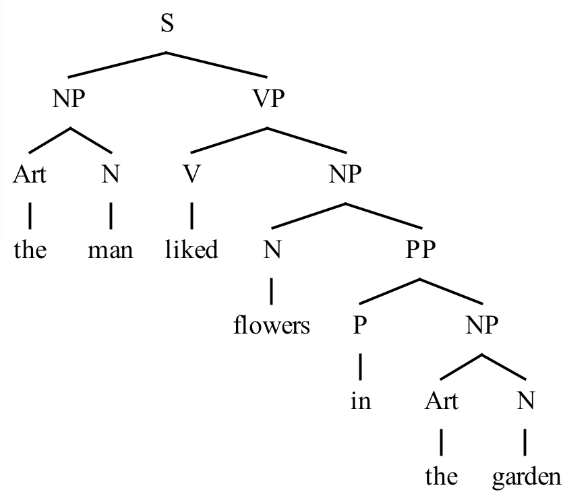


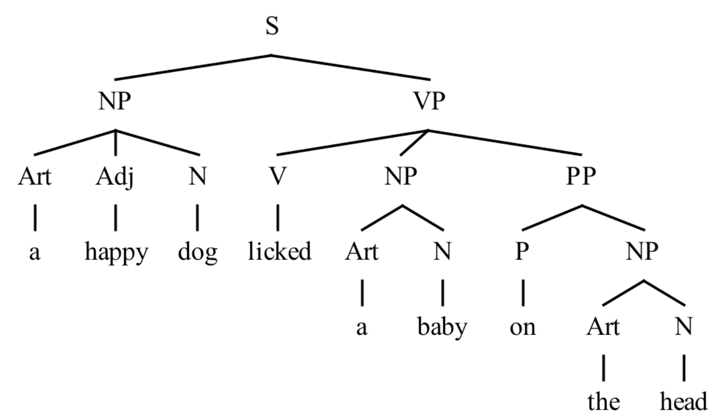
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

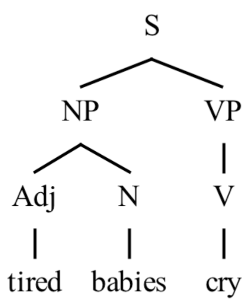




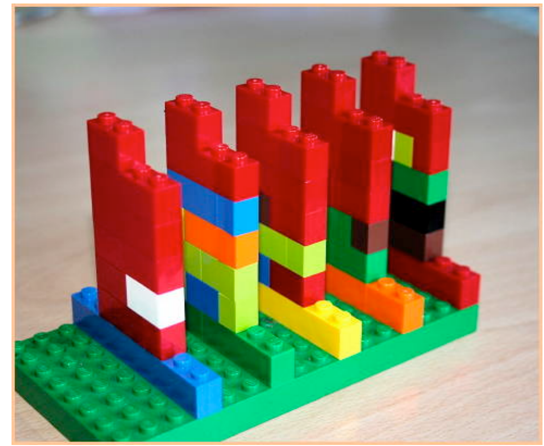






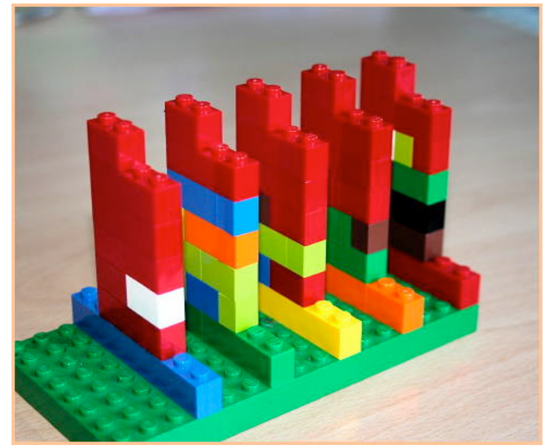


Phrase structure rules



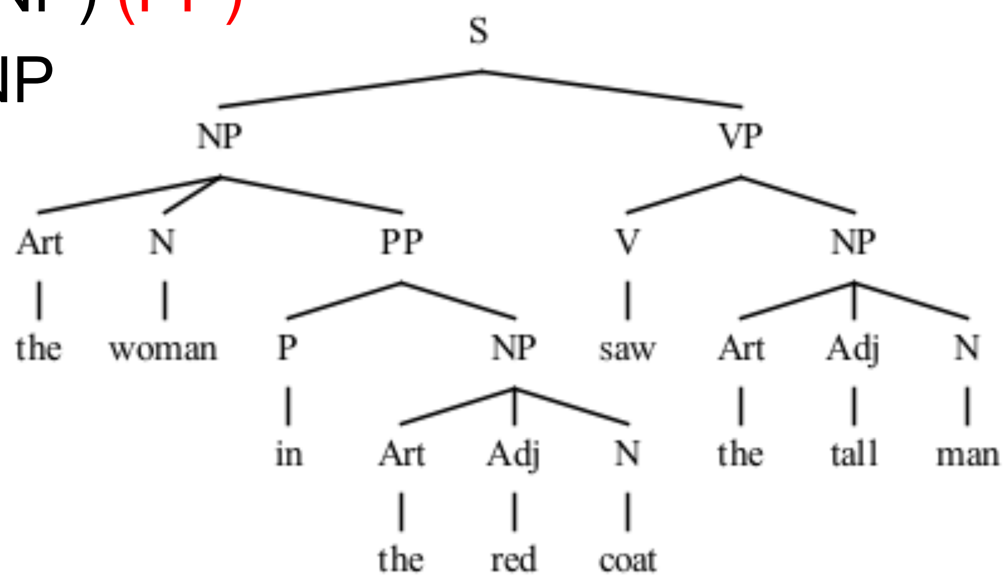
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

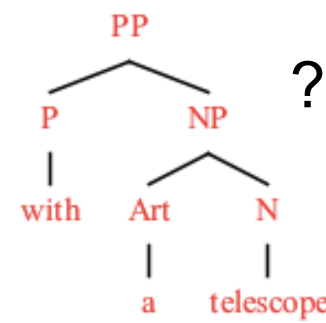
The woman in the red coat saw the tall man with a telescope.



Phrase structure rules

- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP





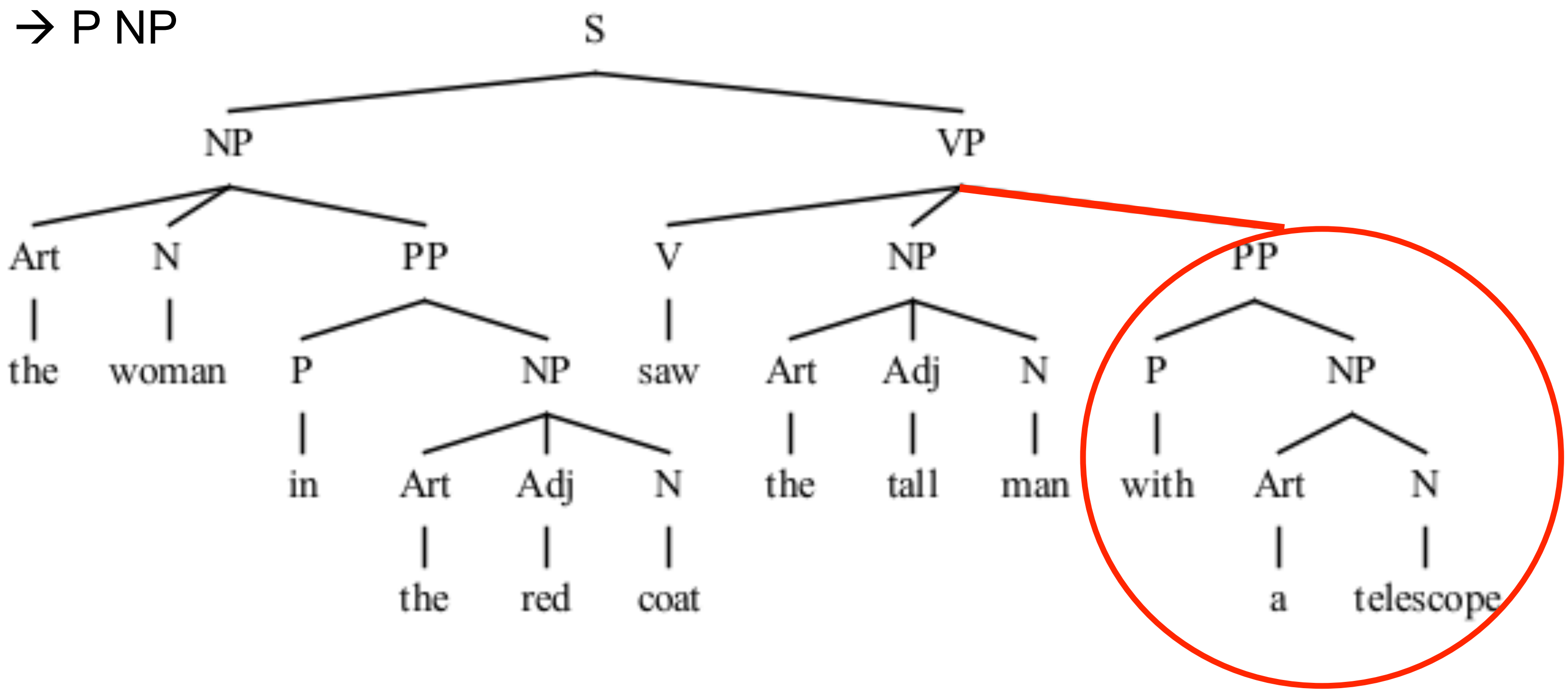
???

The woman in the red coat saw the tall man with a telescope.

Phrase structure and ambiguity

- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

"The woman in the red coat used a telescope to see the tall man."

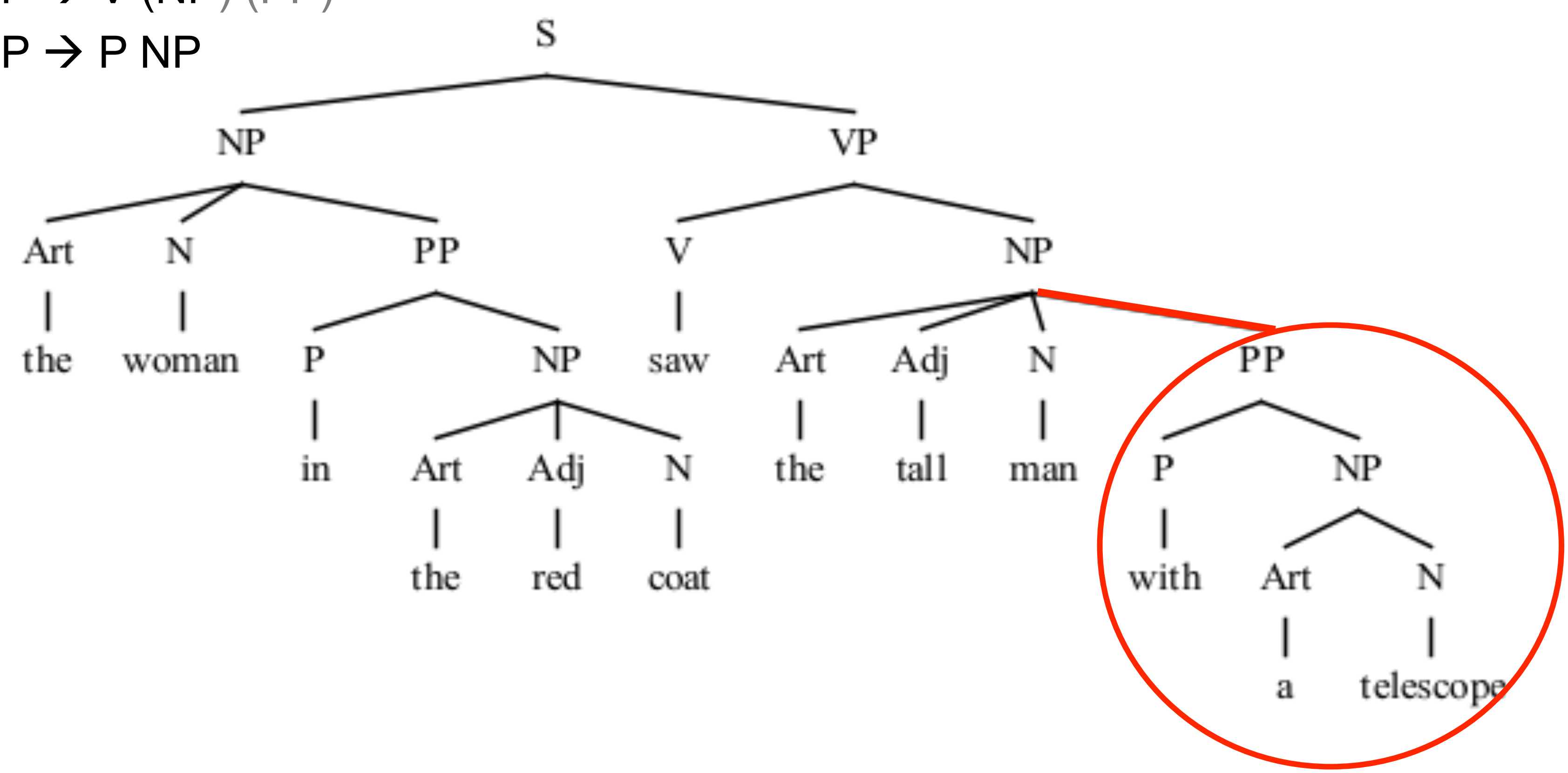


The woman in the red coat saw the tall man with a telescope.

Phrase structure and ambiguity

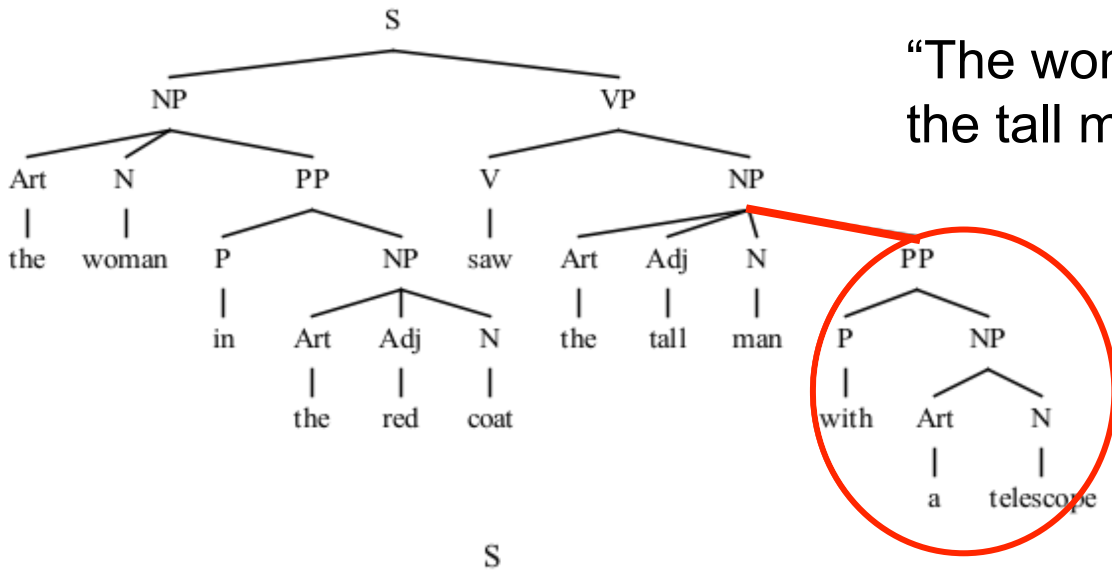
- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

"The woman in the red coat saw the tall man who had a telescope."

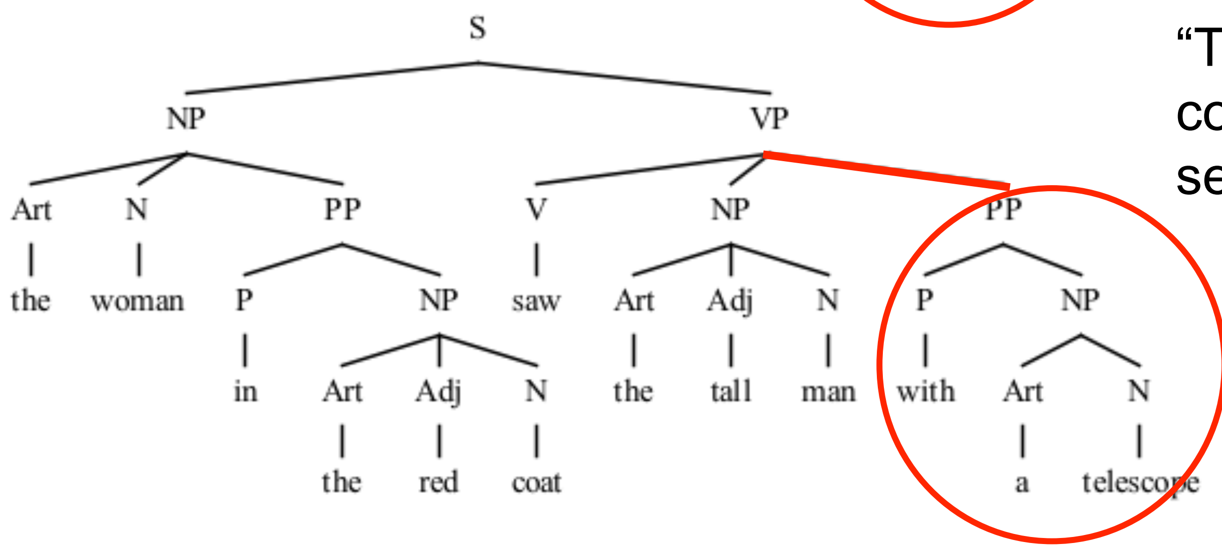


The woman in the red coat saw the tall man with a telescope.

Phrase structure and ambiguity



"The woman in the red coat saw the tall man who had a telescope."



"The woman in the red coat used a telescope to see the tall man."

Phrase structures reveal relationships among phrases and display meaning unambiguously.

Capturing relatedness

That tree is a pine.

The choir can sing beautifully.

Jane does really like sweets.

He will leave in the spring.

Capturing relatedness

That tree is a pine.

The choir can sing beautifully.

Jane does really like sweets.

He will leave in the spring.

That tree is a pine, isn't it?

The choir can sing beautifully, can't they?

Jane does really like sweets, doesn't she?

He will leave in the spring, won't he?

Capturing relatedness

That tree is a pine.	→	*That tree is a pine, isn't it?*
The choir can sing beautifully.	→	*The choir can sing beautifully, can't they?*
Jane does really like sweets.	→	*Jane does really like sweets, doesn't she?*
He will leave in the spring.	→	*He will leave in the spring, won't he?*

Capturing relatedness

- S → NP VP
- NP → (Art) (Adj) N (PP)
- VP → V (NP) (PP)
- PP → P NP

That tree is a pine, isn't it?

The choir can sing beautifully, can't they?

Jane does really like sweets, doesn't she?

He will leave in the spring, won't he?

Capturing relatedness

- S → NP VP
- NP → (Art) (Adj) N (PP)
- NP → Pro
- VP → V (NP) (PP)
- PP → P NP

- S → NP VP VP_{neg} NP_{pro}

That tree is a pine, isn't it?

The choir can sing beautifully, can't they?

Jane does really like sweets, doesn't she?

He will leave in the spring, won't he?

Capturing relatedness

That tree is a pine, isn't it?

The choir can sing beautifully, can't they?

Jane does really like sweets, doesn't she?

He will leave in the spring, won't he?

Capturing relatedness

That tree is a pine, isn't it?

The choir can sing beautifully, can't they?

Jane does really like sweets, doesn't she?

He will leave in the spring, won't he?

Capturing relatedness

That tree is a pine, isn't it?

The choir can sing beautifully, can't they?

Jane does really like sweets, doesn't she?

He will leave in the spring, won't he?

Capturing relatedness: transformations

That tree is a pine, isn't it?

The choir can sing beautifully, can't they?

Jane does really like sweets, doesn't she?

He will leave in the spring, won't he?

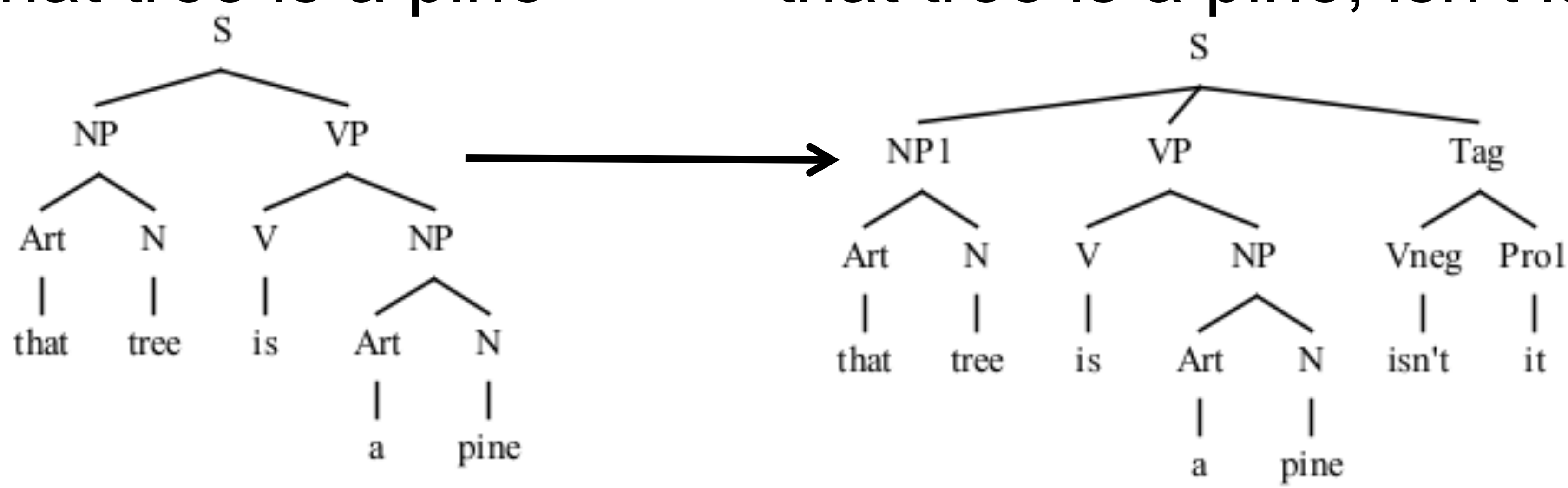
Capturing relatedness: transformations

Tag transformation:

Copy and negate verb; insert at end

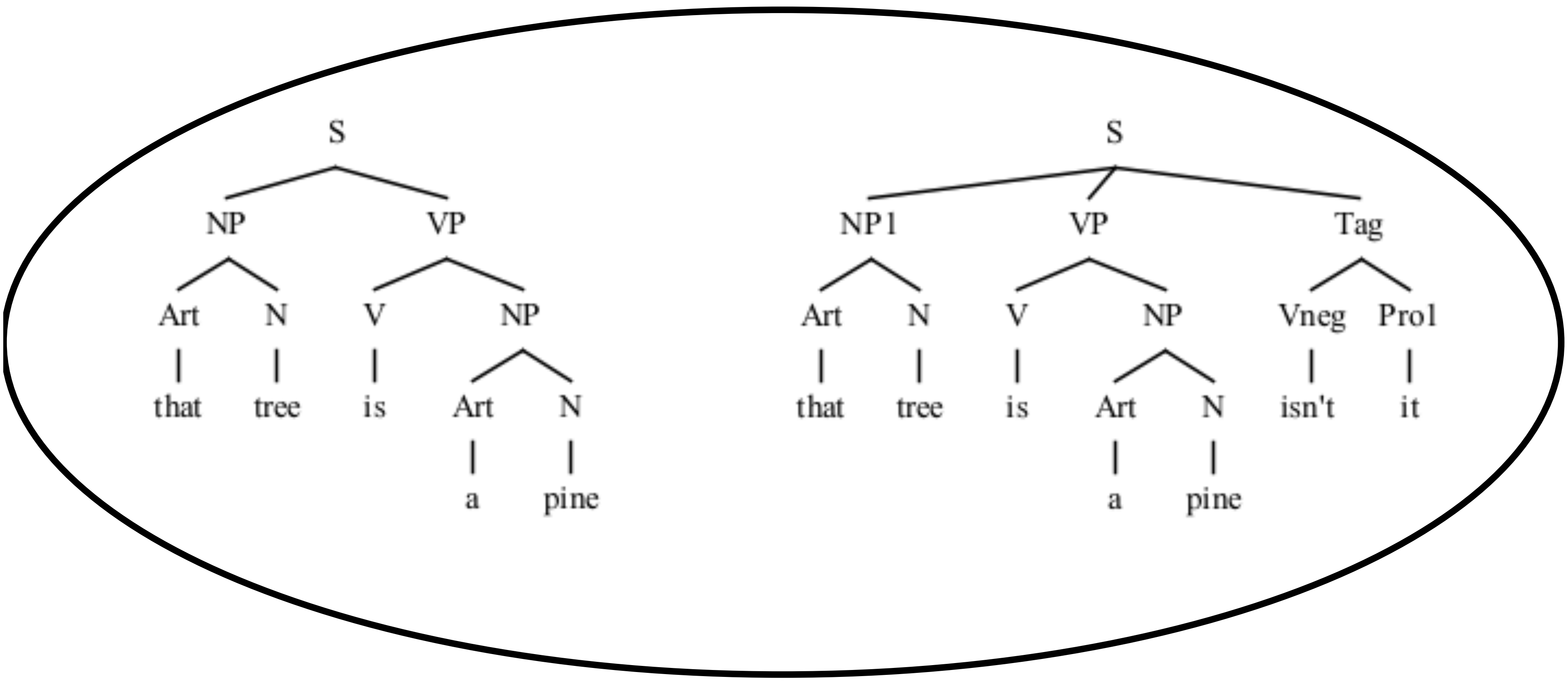
Copy NP1; make it into a pronoun; insert at end.

that tree is a pine ⟶ that tree is a pine, isn’t it



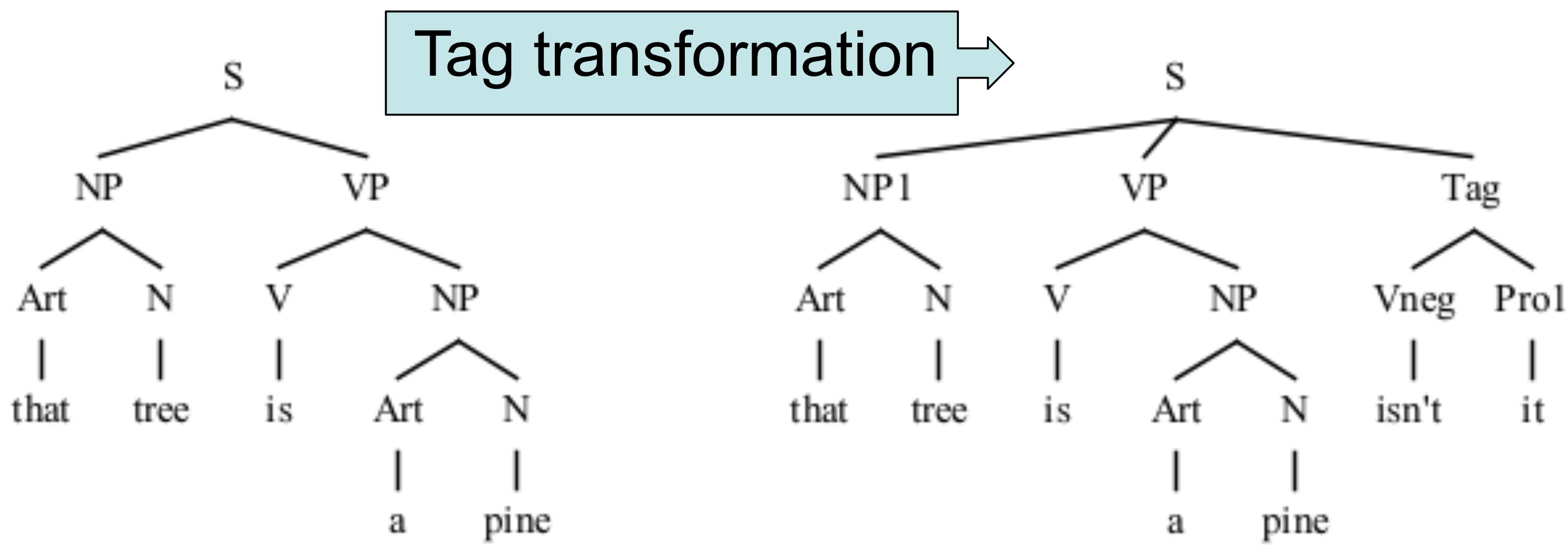
Capturing relatedness: transformations

That tree is a pine, isn't it is described by a **set** of phrase markers.



Capturing relatedness: transformations

More specifically, a **deep structure** derived from a **surface structure** by way of a **transformation**.



Deep structure

Surface structure

Capturing relatedness: one more example

- *Go to the store!*

- S → NP VP   ….. ????

- Who?
- When?

Capturing relatedness: one more example

- *Go to the store!*
- *Go to the store, won't you?*

Capturing relatedness: one more example

- *Go to the store!*
- *Go to the store, won’t you?*

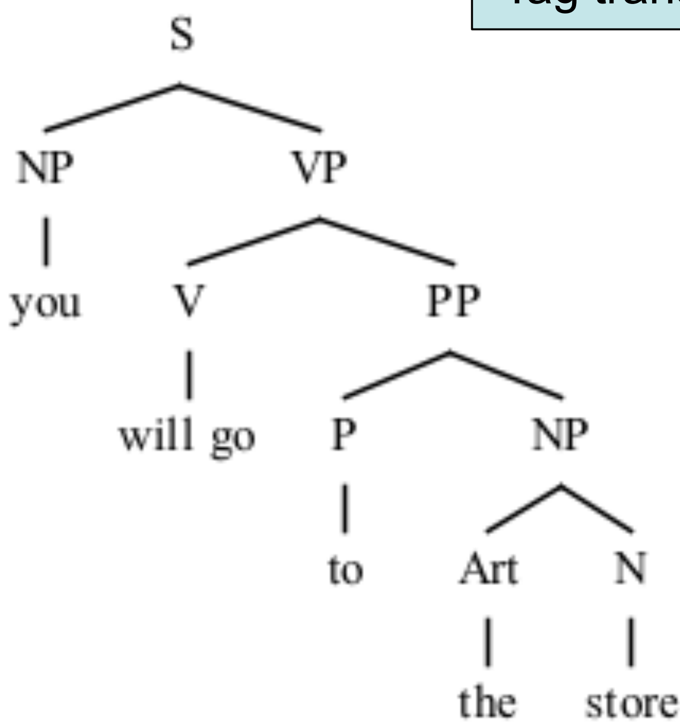
- *go to the store, won’t you?*

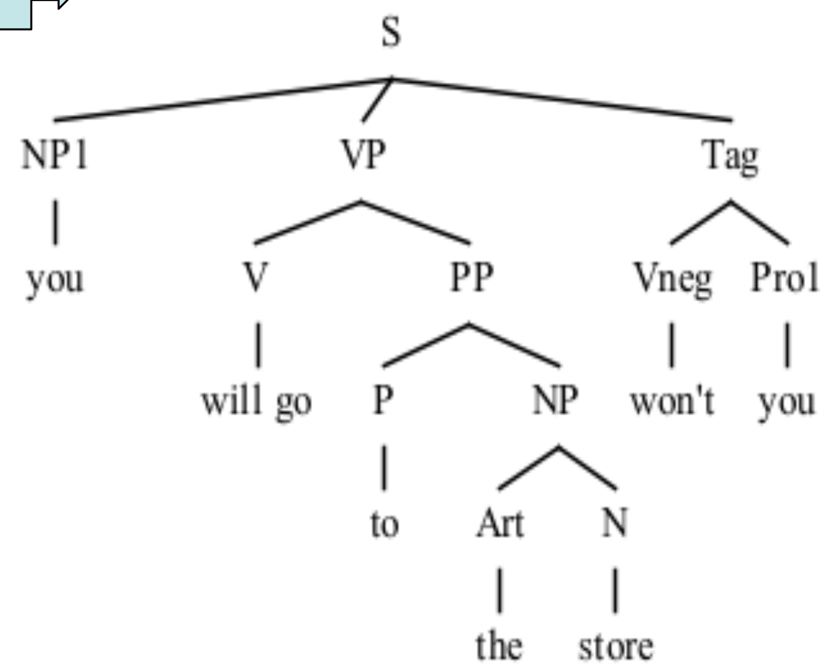
Capturing relatedness: one more example

- *Go to the store!*
- *Go to the store, won't you?*

- *you will go to the store, won't you?*

Tag transformation

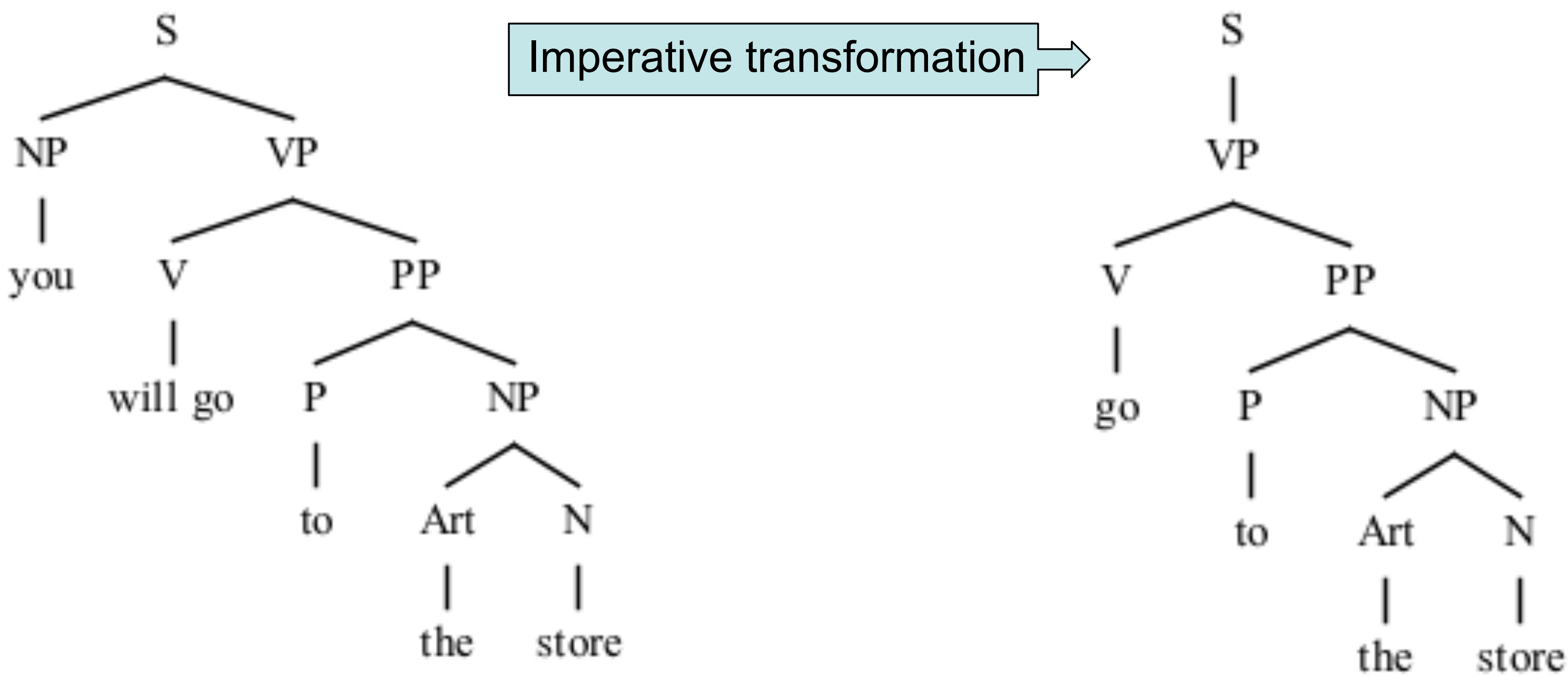




You will go to the store, won't you?

Deep structure

Surface structure

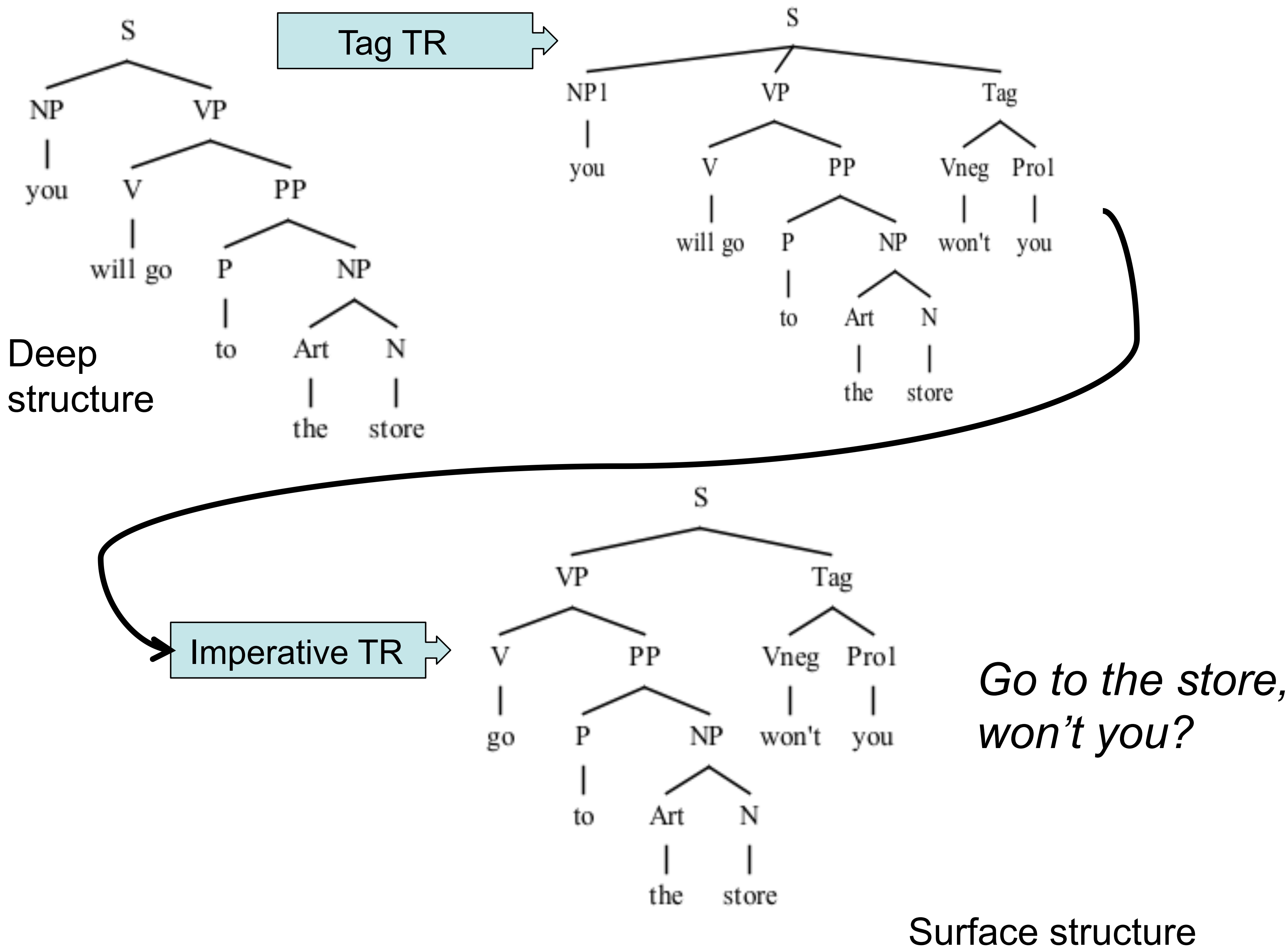


Go to the store.

Imperative transformation: Delete first NP and *will.*

Deep structure

Surface structure


Tag TR
Imperative TR
S
NP
VP
you
V
PP
will go
P
NP
to
Art
N
the
store
Deep structure
S
NP1
VP
Tag
you
V
PP
Vneg
Pro1
will go
P
NP
won't
you
to
Art
N
the
store
S
VP
Tag
V
PP
Vneg
Pro1
go
P
NP
won't
you
to
Art
N
the
store
Go to the store, won't you?
Surface structure

Capturing relatedness

you will go to the store

Imperative TR

Tag TR

Tag, Imperative TR

you will go to the store

go to the store

you will go to the store, won’t you

go to the store, won’t you

Deep structure

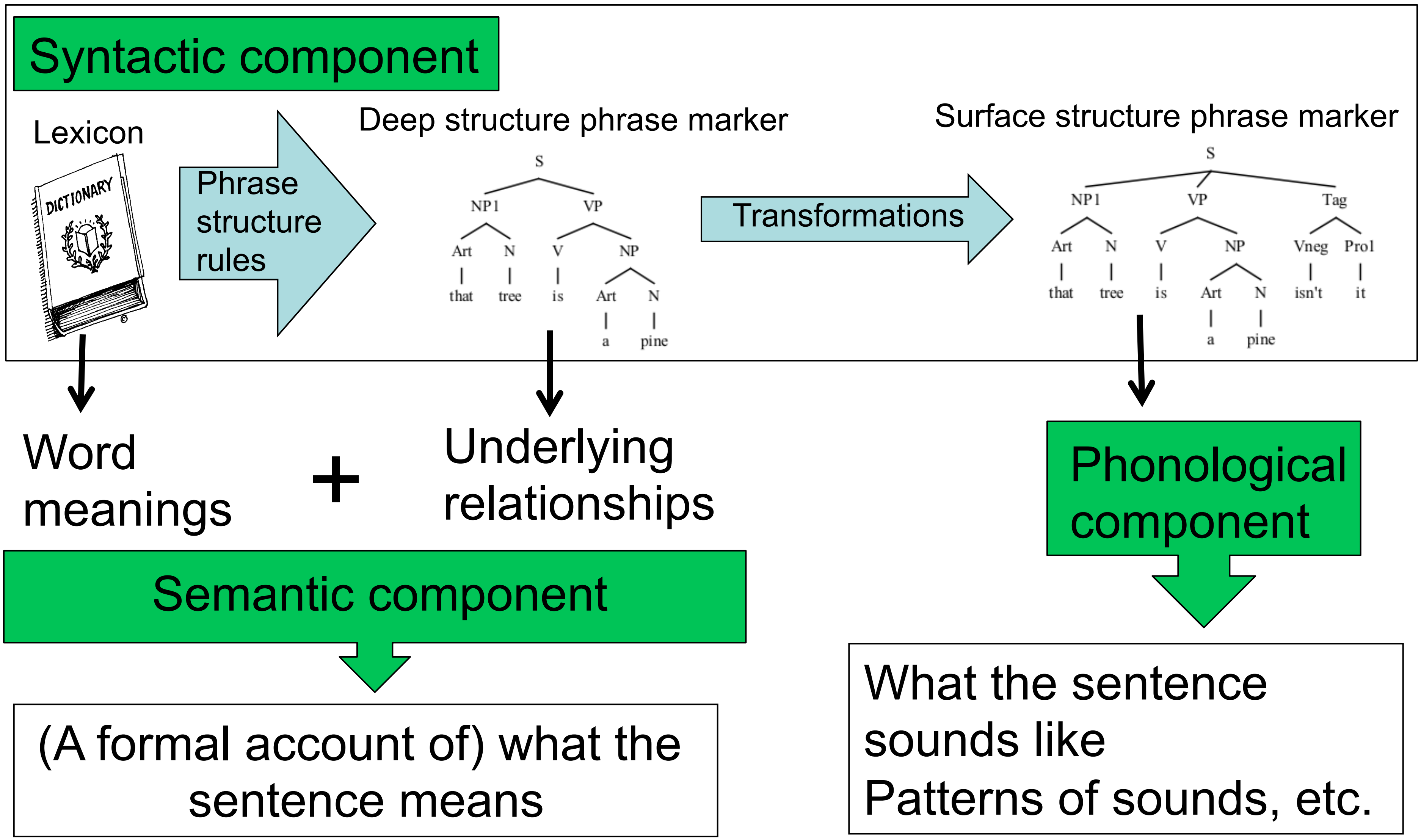
Surface structure

Capturing relatedness: What do transformations get us?

- S → NP VP
- S → NP VP VP_{neg} Np_{pro}
- S → VP
- S → VP VP_{neg} Np_{pro}
- No way to show relationships among
  - *You will go to the store*
  - *Go to the store*
  - *You will go to the store, won't you*
  - *Go to the store, won't you*

- S → NP VP
- Tag transformation
- Imperative transformation
- Transformations mark relationships between deep and surface structures

Linguistic knowledge representation



Sigh…

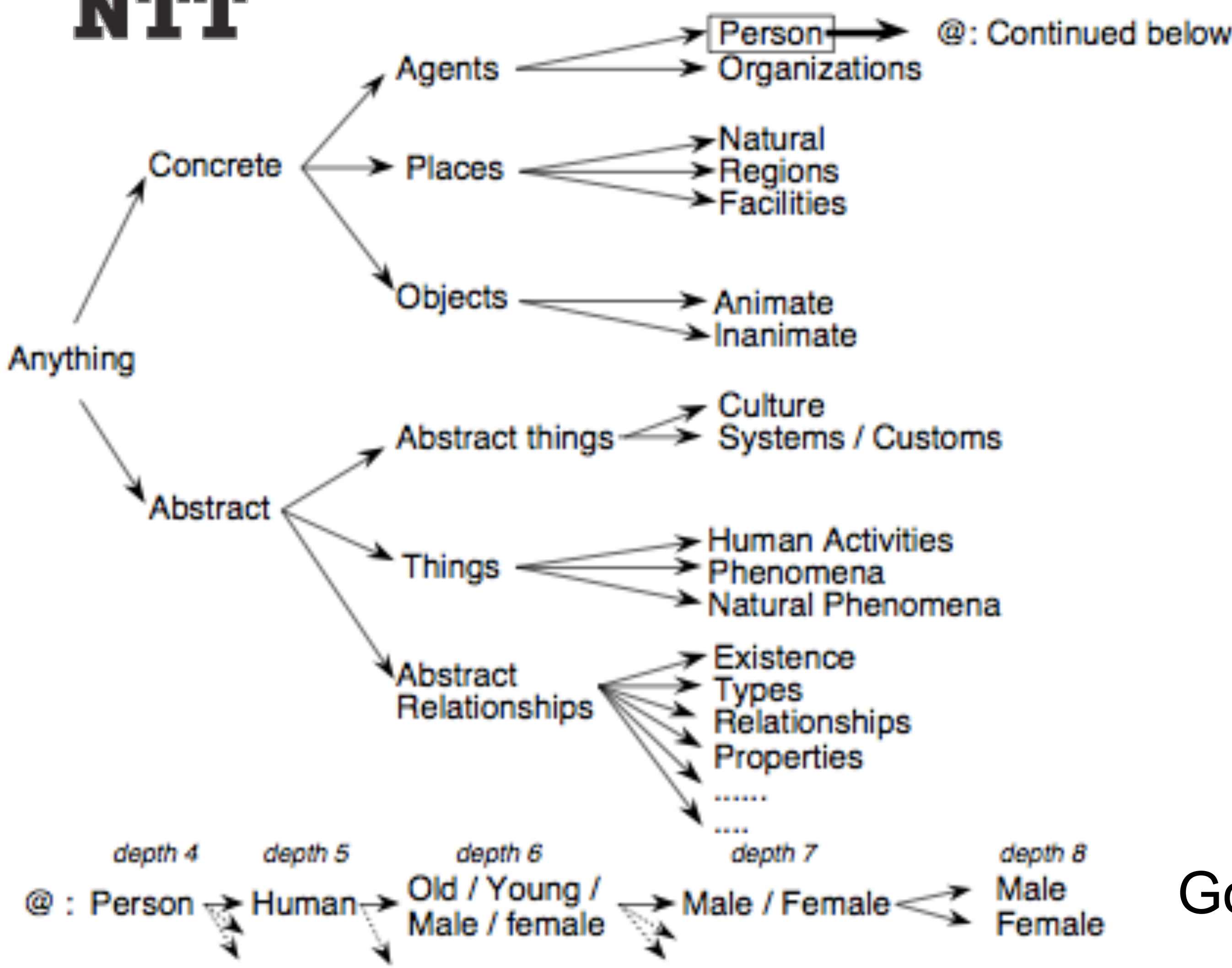
“A full account of the nature of all such rules has yet to be given [1964], although tremendous progress has been made in recent years, and the outlines of correct solutions appear to be relatively clear.” (Postal)

“A full account of the nature of all such rules” ***still*** “has yet to be given” [2014]. (me)

Interlude
Rule-based machine translation: ALT-JE

- Research arm of the Nippon Telegraph and Telephone Corporation
- 1980s ~ ??
- Proprietary, product-oriented

Semantic hierarchy



Goes to 12 levels

Translation rules, using valency patterns

IF						THEN					
	J-Verb	=	"yaku"				Subj	=	N_1		
	N_1 (Subj)	≡	"Person"				E-Verb	=	"bake"		
	N_2 (Obj)	≡	"Bread" or "Cake"				Obj	=	N_2		
IF						THEN					
	J-Verb	=	"yaku"				Subj	=	N_1		
	N_1 (Subj)	≡	"Person"				E-Verb	=	"roast"		
	N_2 (Obj)	≡	"Meat"				Obj	=	N_2		
IF						THEN					
	J-Verb	=	"yaku"				Subj	=	N_1		
	N_1 (Subj)	≡	"Person"				E-Verb	=	"broil"		
	N_2 (Obj)	≡	"Fish" or "Seafood"				Obj	=	N_2		
IF						THEN					
	J-Verb	=	"yaku"				Subj	=	N_1		
	N_1 (Subj)	≡	"Agents"				E-Verb	=	"cremate"		
	N_2 (Obj)	≡	"Person" or "Animals"				Obj	=	N_2		
IF						THEN					
	J-Verb	=	"yaku"				Subj	=	N_1		
	N_1 (Subj)	≡	"Agents" or "Machines"				E-Verb	=	"burn"		
	N_2 (Obj)	≡	"Places" or "Objects" or "Locations"				Obj	=	N_2		

Figure 2: Translation rules for the Japanese verb "YAKU". These rules are composed manually by human experts. " ≡ " indicates "an instance of".

NTT's ALT J-E

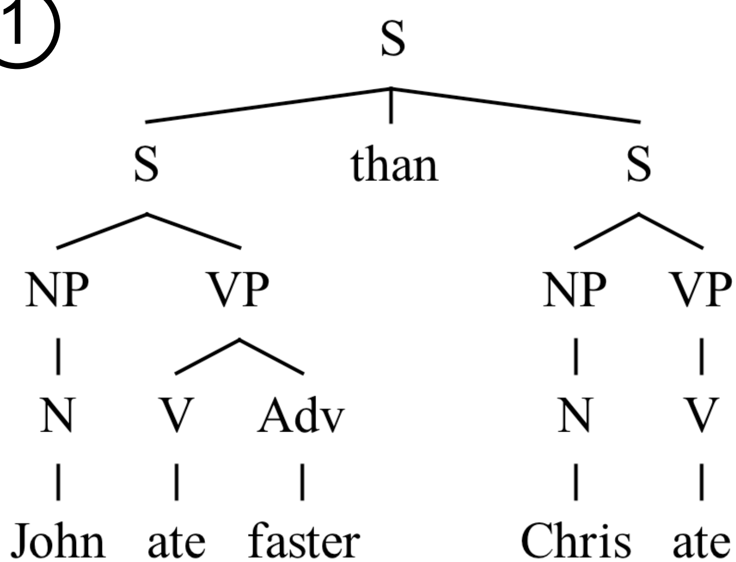
- 3000 semantic categories; 12 levels
- 11,214 valency patterns for >5000 verbs
  - Doesn't include idioms
  - Expect need @ 27,000 patterns to cover 80% of Japanese verbs (or any language, for that matter)
- Automate category construction over dictionaries (not corpora)
- Uses frequency to decide senses not possible to decide using semantic categories
  - *hato* = "pigeon" or "dove"; "pigeon" more frequent

Psychological reality in Linguistics

- What is the mental status of phrase structure trees and transformations?
  - Actual participants in mental processing
  - Model with no implications for actual processing
- I.e., we behave AS IF we follow rules…
  - …and we do.
  - …but we don't know whether we do or not.
- Which is it?

Transformations as psychologically real

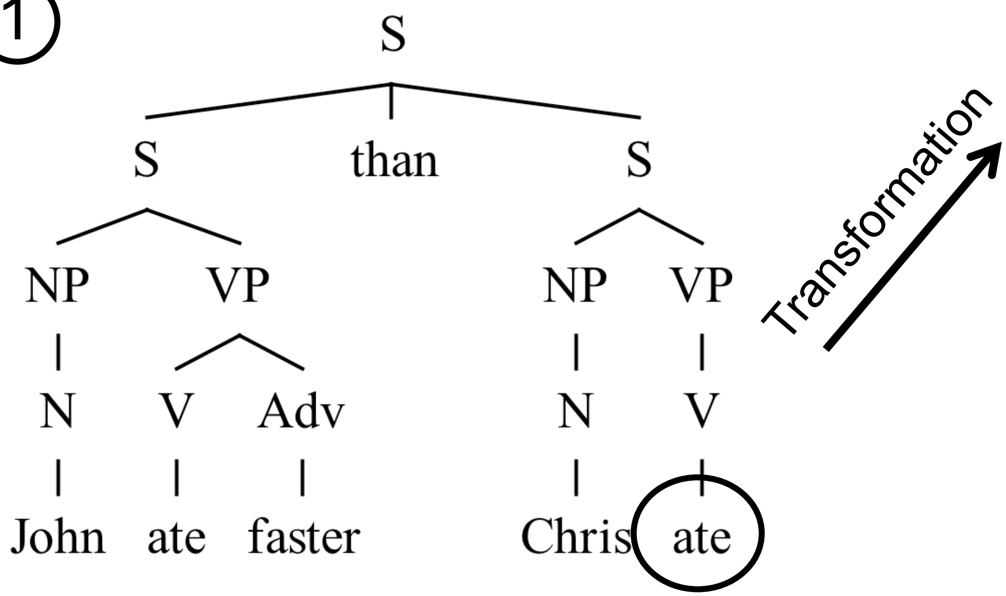
1



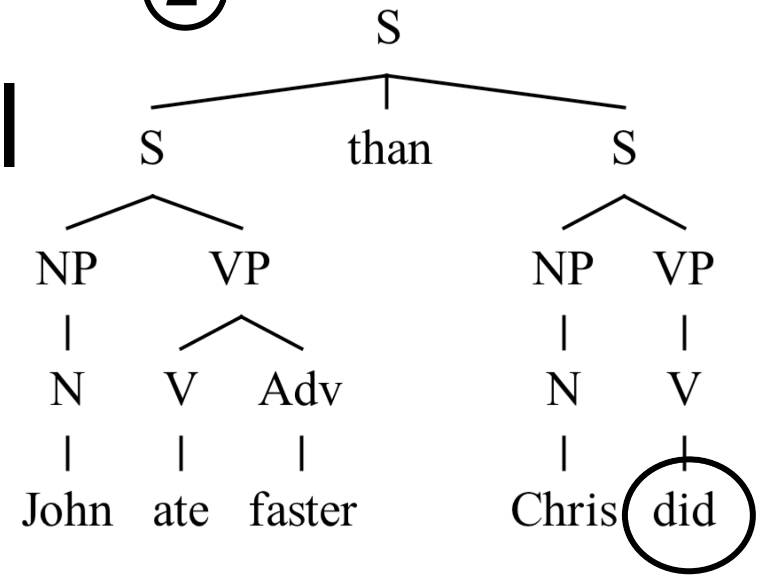
Deep structure

Transformations as psychologically real

1



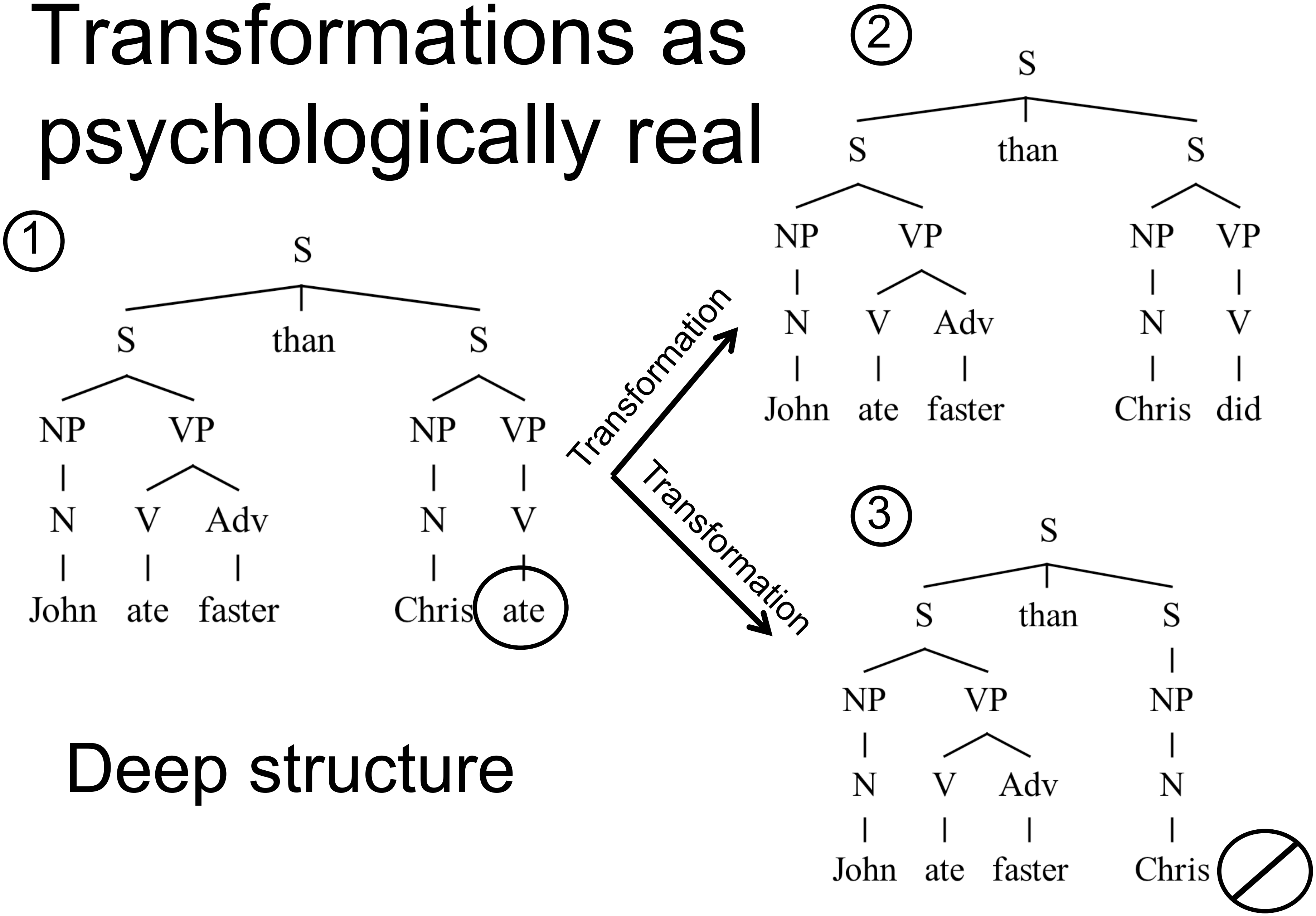
2



Surface structure

Deep structure

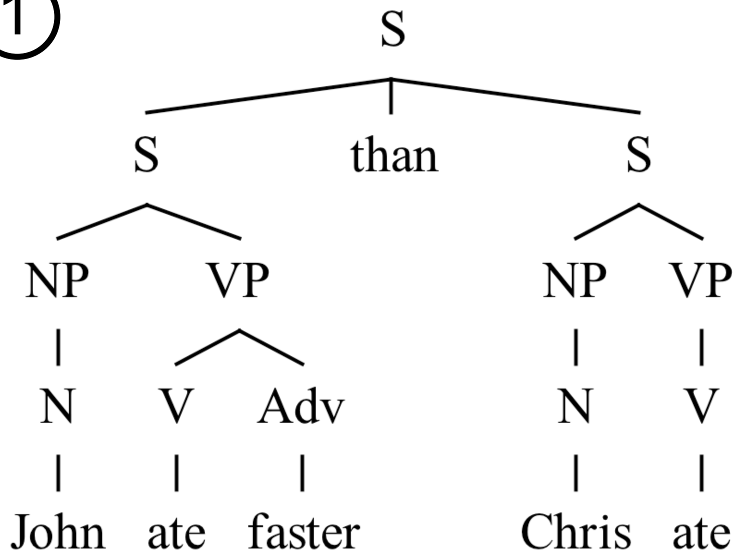
Transformations as psychologically real



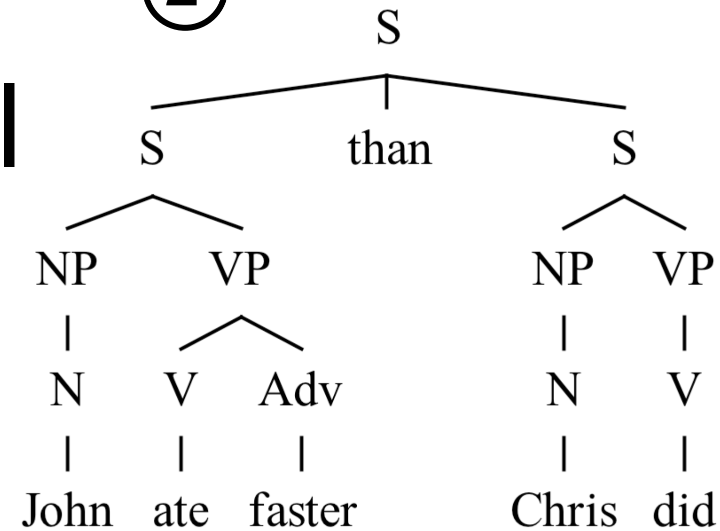
Deep structure

Transformations as psychologically real

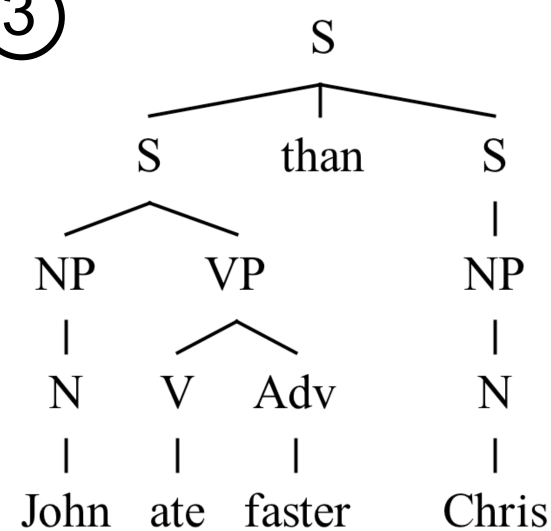
①



②



③



If transformations are psychologically real, it should be faster to process ① than ② or ③. In fact subjects respond to ① more slowly than either ② or ③.

Dependency

- Generative approach assumes compositionality
  - Build from smallest possible, independent pieces
- However, need to account for dependencies among words (bigger chunks)
  - **Idioms**: *baker's dozen; a toss-up, crack someone up, drive someone up a wall, see the forest for the trees*
  - **Collocations**: *drenched to the ?, sharp as a ?, a glimmer of ?, **build** a house* vs. ***create** a novel*
  - *Akimbo, scrambled, inclement, precinct*
  - **Deviant syntax**: *All of a sudden, How dare you?, Beware of the dog*
- Doesn't work well to build these phrases from their component parts

Linguistic knowledge representation Part 2: Exemplar theory

- Speakers record all linguistic experiences
  - “Warehouse” analogy, no more Legos
- Of varying sizes
  - Not just a collection of smallest independent bits
- Keep track of the frequencies of expressions

Memorizing (= recording)

- **Words** have to be memorized anyway
- Why not also…
  - *Hi, how are you?; see you later; ‘sup?*
  - Lyrics to *Jingle Bells, O Canada, Rigoletto*
- Where do you draw the line?
- Generative approach draws it at smallest independent part
- Exemplar theory does not
  - Individually determined

Generating novel sentences?

- Store all *previous* stuff
  - If you have heard *hi, how are you?* you can produce *hi, how are you?*
- How do speakers generate **new** stuff?
  - How did I produce *soundless noisy democracy barks gently*?
  - Generative grammar generates from lexical items, phrase structure rules and transformations
- (Actually, how do speakers generate **new** stuff that is **grammatical***?)*

Data-oriented parsing

- Keep a store of representations
  of all previous language experiences
- Derive sub-structures from those representations of previous language
- Re-combine into new structures
- Keep track of
  - Frequency
  - Recency
  - Modelled by probability
  - Can set probability at zero if something is way in the past (thus, forgetting)

Data-oriented parsing requires:

- Representation format for well-formed utterances
- Decomposition operations that create all possible fragments
- Composition operations that re-combine them
- Probability model that assigns likelihood

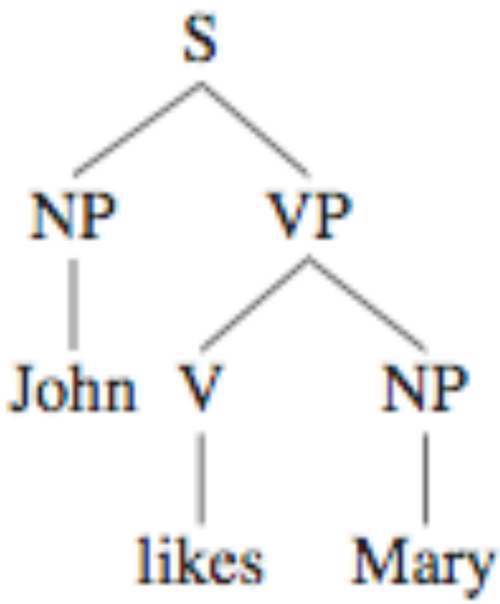
Data: Tagged corpora

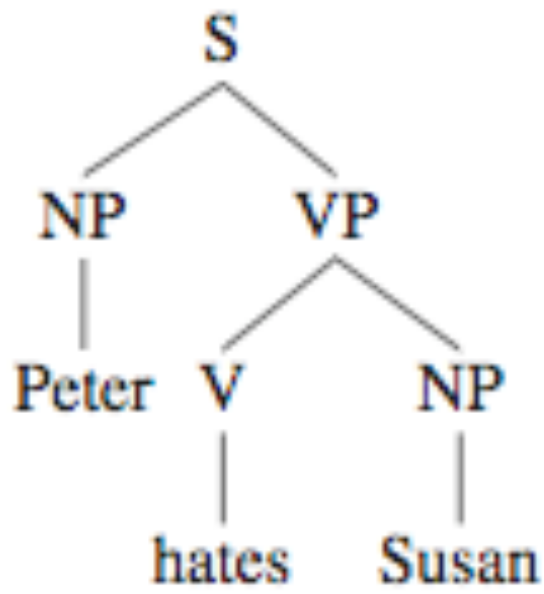
(Penn Treebank, Switchboard corpus)

```
( (CODE SpeakerA2 .))
( (INTJ Okay
          .
          E_S))
( (CODE SpeakerB3 .))
( (SBARQ (INTJ Well)
           (WHNP-1 what)
           (SQ do
              (NP-SBJ you)
              (VP think
                    (NP *T*-1)
                    (PP about
                       (NP (NP the idea)
                              (PP of
                                 ,
                                 (INTJ uh)
                                 ,
                                 (S-NOM (NP-SBJ-2 kids)
                                           (VP having
                                                 (S (NP-SBJ *-2)
                                                    (VP to
                                                       (VP do
                                                             (NP public service work))))
                                                 (PP-TMP for
                                                          (NP a year)))))))))
           ?
```

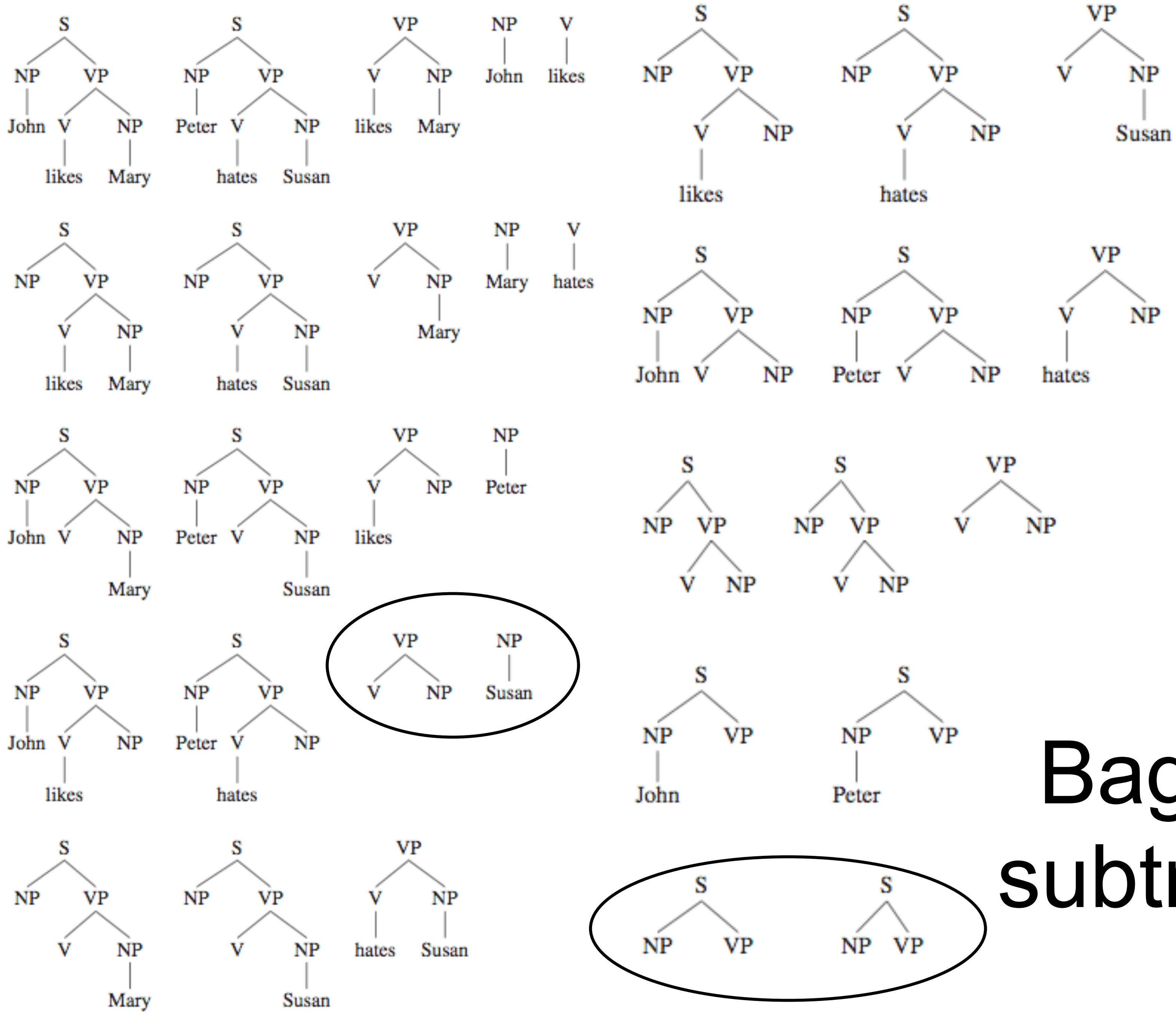
“Okay.”
“Well, what do you think about the idea of uh kids having to do public service work for a year?”

Example

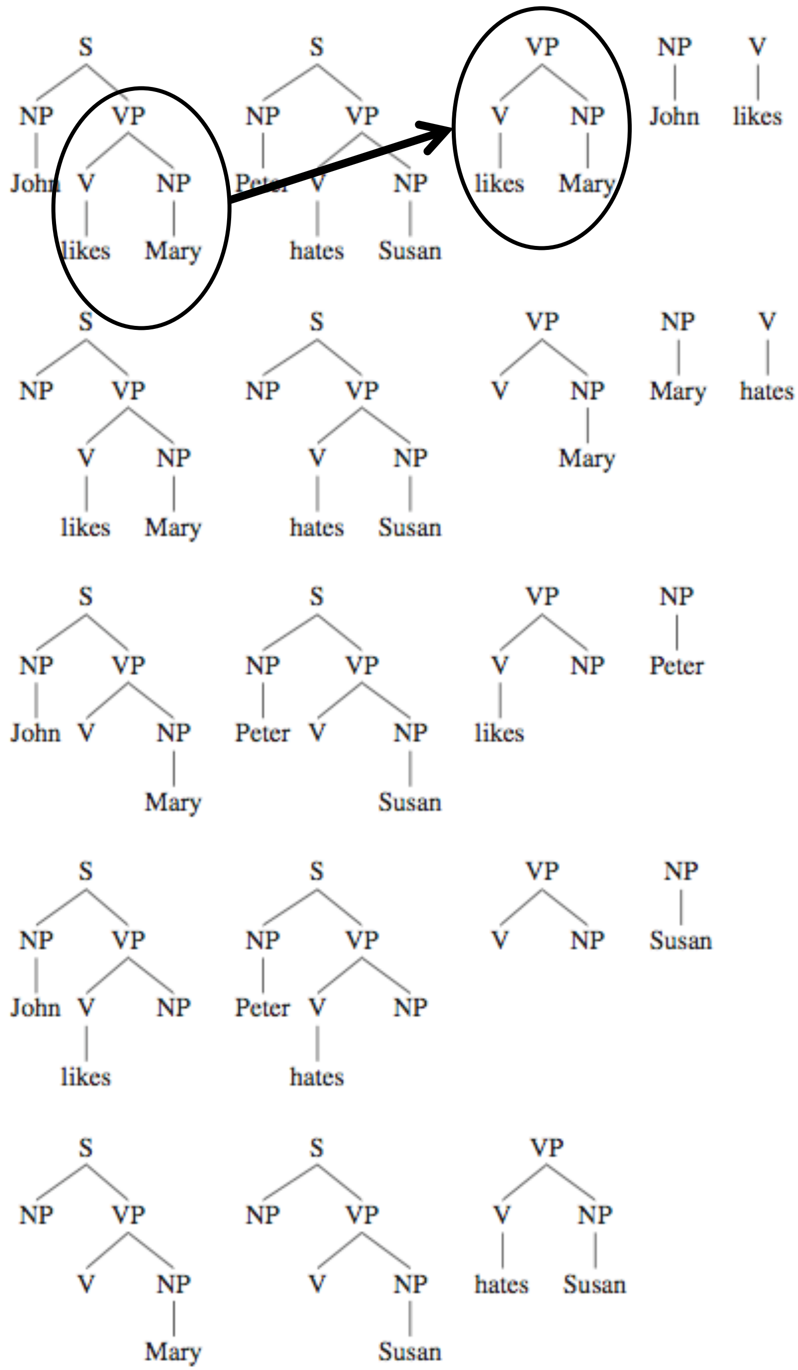


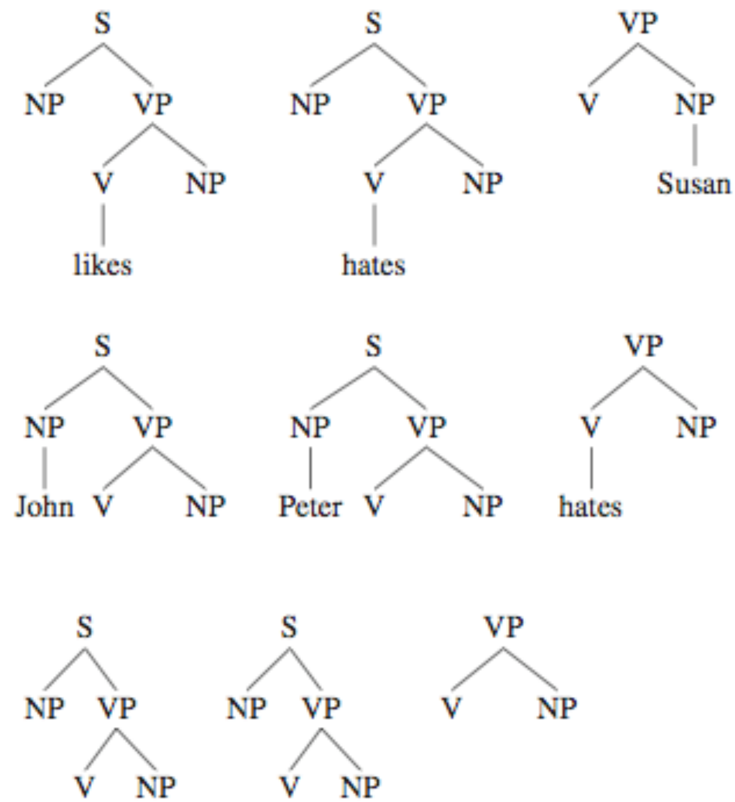


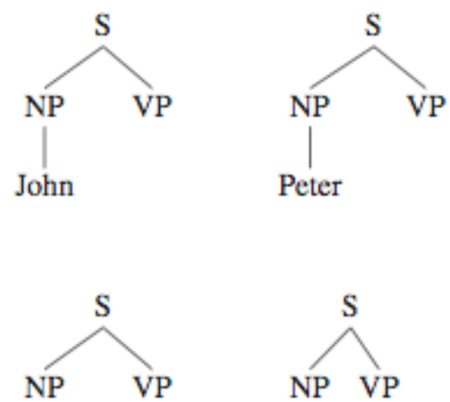
- Note: the structures may look the same as the generative tree structures, but the implications are different.
  - Generative approach: structures are crucial; words are incidental
  - Exemplar approach: words AND structures are recorded



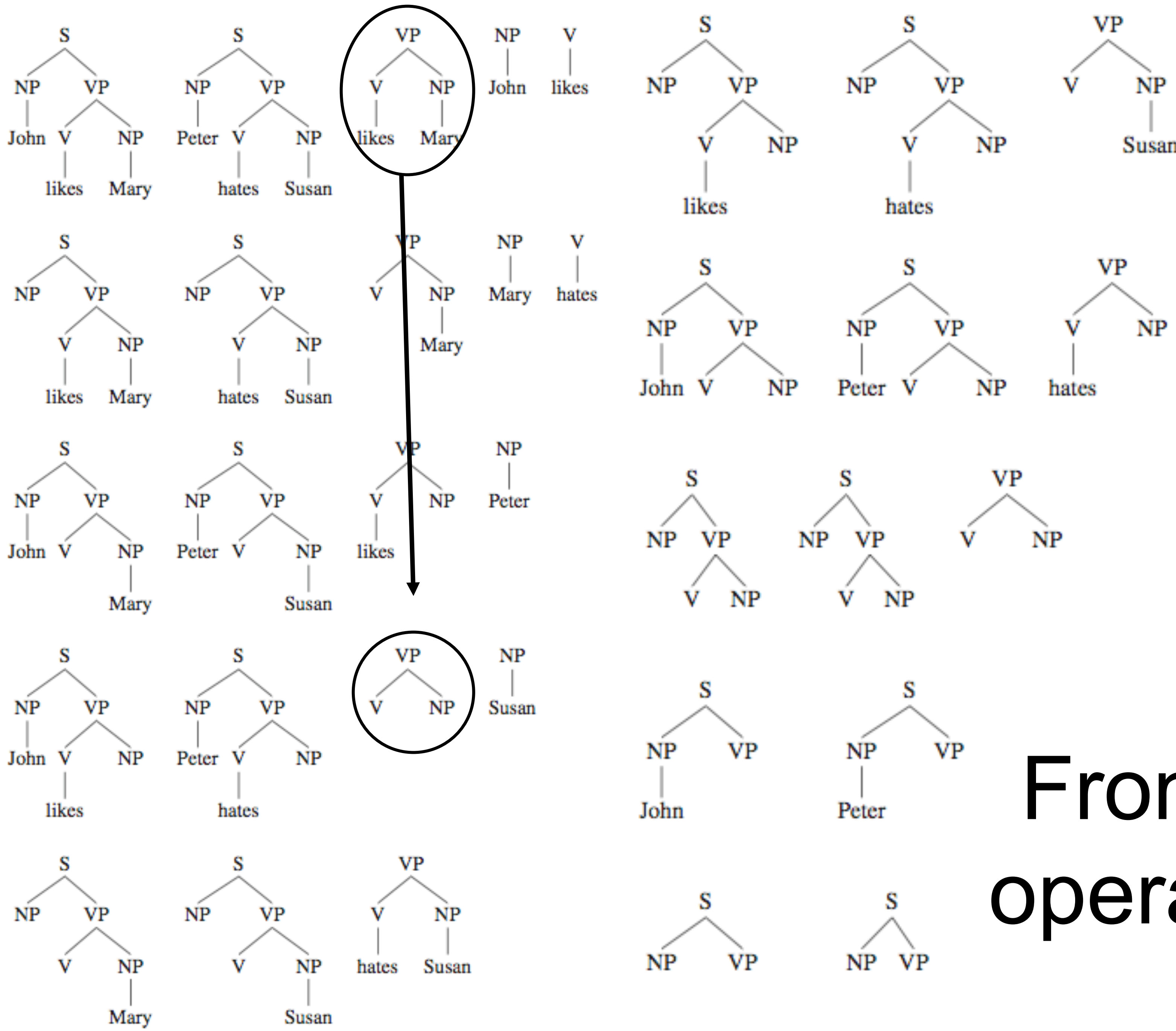
Bag of subtrees



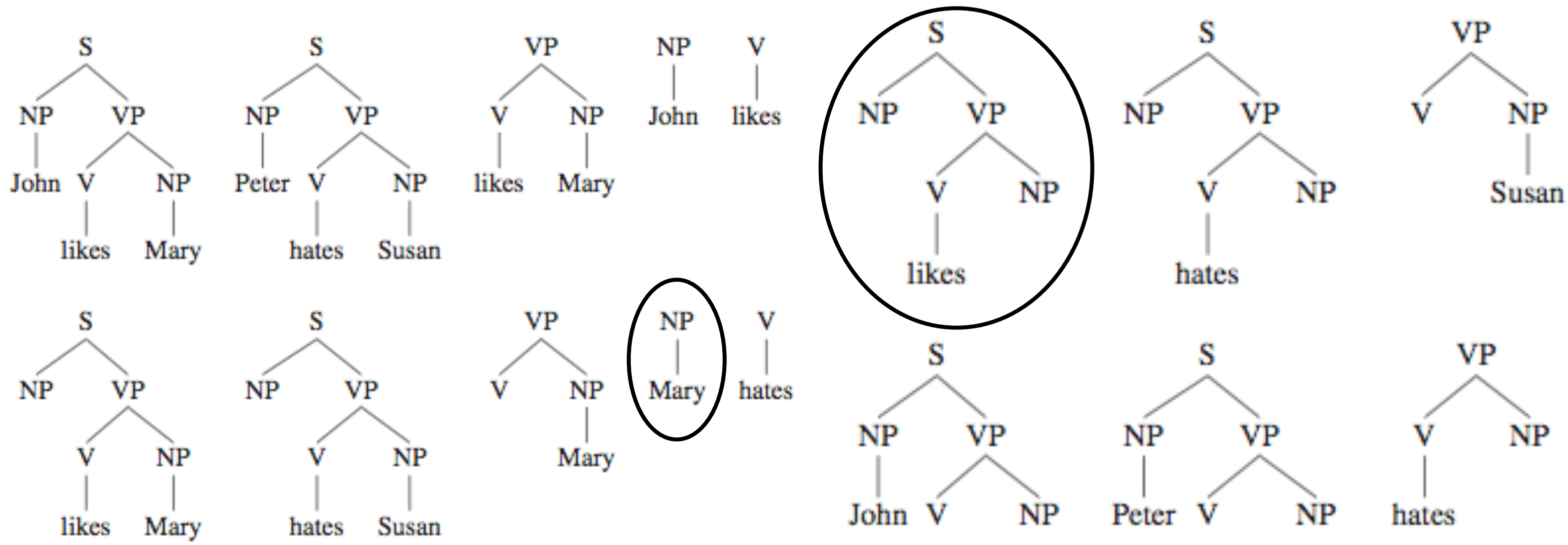


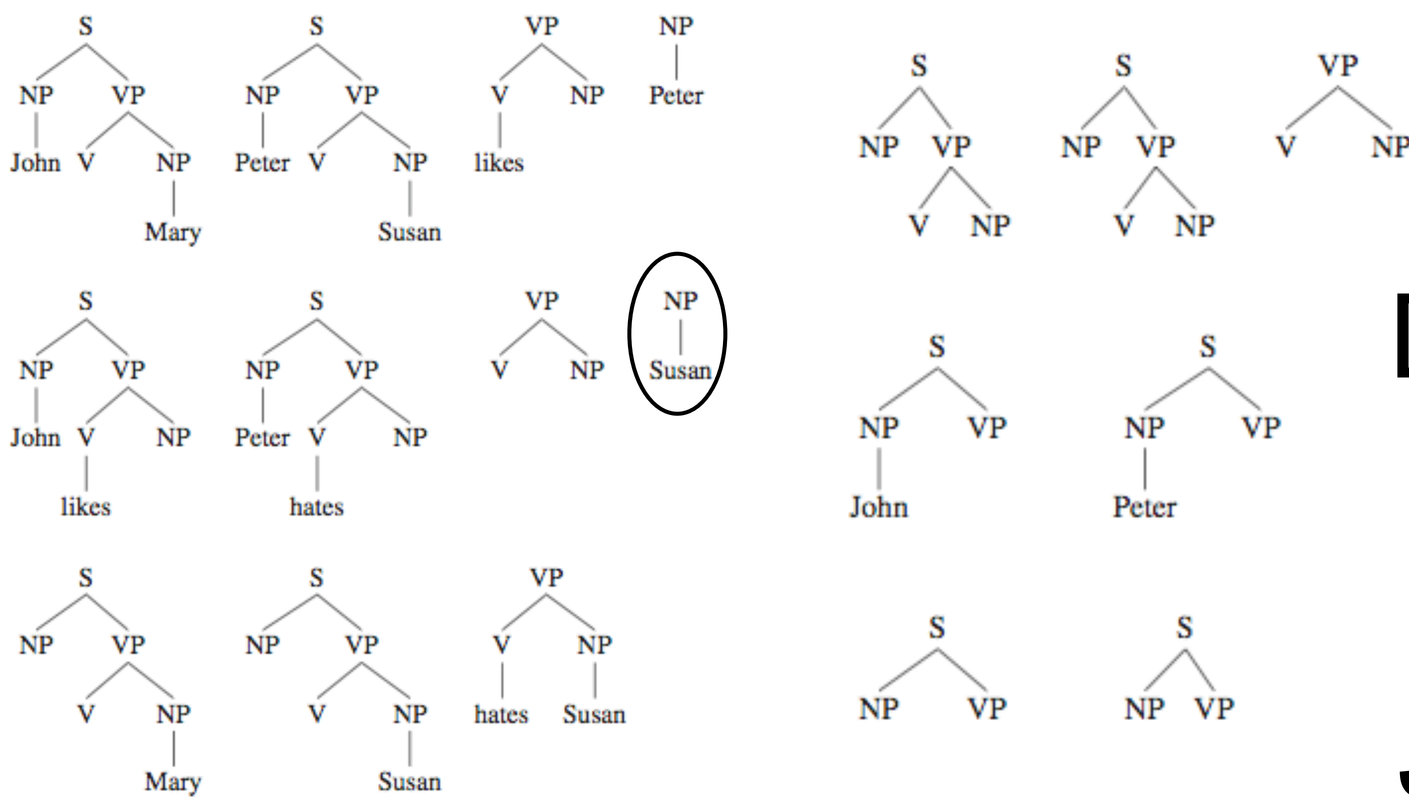


Root operation



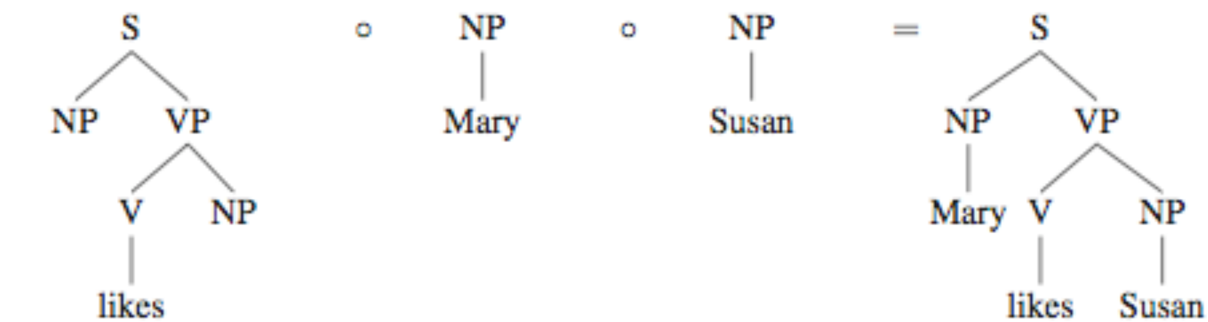
Frontier operation

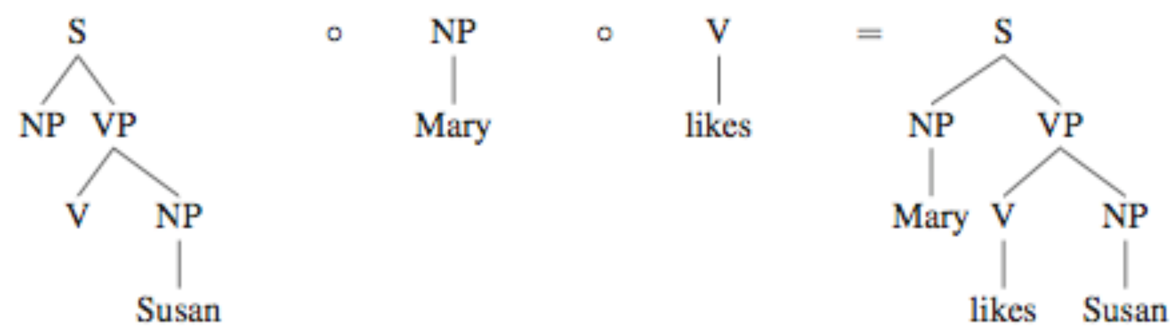


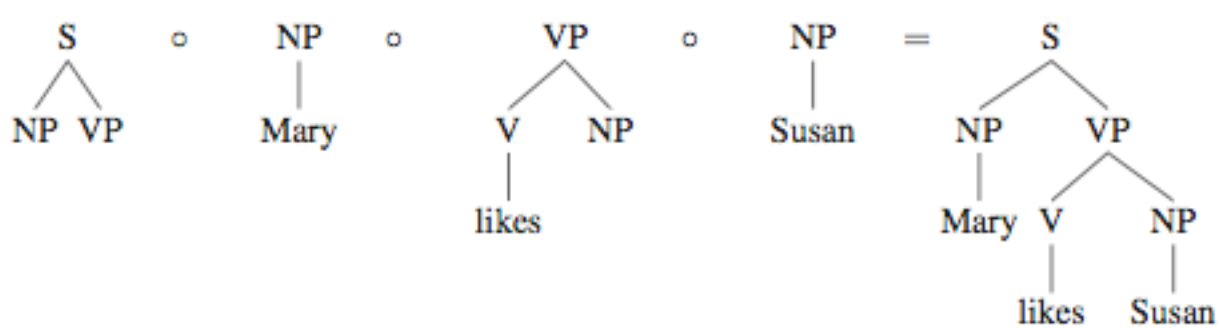


Derive *Mary likes Susan*

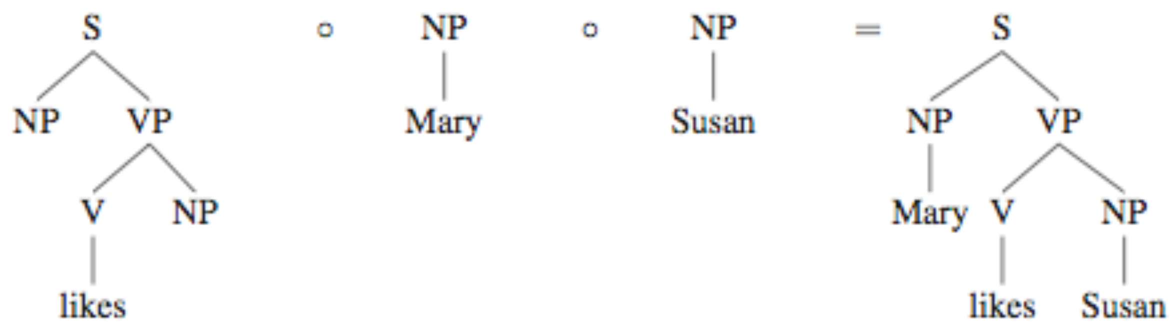
Derive *Mary likes Susan*







Probability of *Mary likes Susan*



20 subtrees with S at top in bag of subtrees = 1/20	X	4 subtrees with NP at top in bag of subtrees = 1/4	X	4 subtrees with NP at top in bag of subtrees = 1/4	=	1/320

- Will also need to figure in probabilities of other derivations

Notion of probability

- Similarity between a given sentence and the exemplars in the corpus
  - What do we mean by corpus here?
- Correlates with
  - Number of trees that share fragments with the sentence
  - How big those fragments are
- Speakers maximize probability in production and perception

Claimed advantages

- Allows for gradient of judgments (stochasticity)
  - Gradual degradation of “grammaticality”
    - *How old are you?* vs. *What age are you?* vs. *What age do you have?* vs. *How many years do you have?*
- Fragments can be any size
  - Can capture idioms/collocations, etc., of any size
  - “Pre-fabs” = 55% to 70% of English
- (Handles structural ambiguity just like generative approach)

Psychological reality of exemplars

- Experimental evidence
  - Present three-word sentences
    - From a collected corpus
  - Matched for plausibility, word complexity, structural complexity, etc.
  - Task: grammatical or not?
  - Subjects faster with more frequently heard sentences (Bod, 2000)
- High frequency expressions change faster: *I dunno* vs. *I don't notice*
- High frequency forms resist regularization: *keep/kept* vs. *leap/leapt* (Bybee, 2006)

Exemplars and machine translation

- For you to think about…
- The “fit” between an exemplar approach and a statistical approach to machine translation using “big data”

Bayesian Inference

Bayesian inference is an iterative process in which new evidence repeatedly updates an initial probability distribution

Before any evidence is taken into account, one starts with some belief about H, expressed as an initial prior probability

At each iteration step, the prior probability is taken to be the posterior probability from the previous iteration

See http://en.wikipedia.org/wiki/Bayesian_inference

- What constitutes “big data” for an exemplar approach to human language processing?

Knowledge representation

- EXEMPLAR THEORY
- Well-formed linguistic representation
- Subtrees, any size
- De-composing, composing operations
- Record/re-use all linguistic experience
- Corpus: linguistic input
- "Goodness" = probability

- GENERATIVE GRAMMAR
- Phrase structure, transformations
- Smallest, independent
- Phrase structure rules, transformations
- Generate linguistic experience
- Corpus: intuitive judgments
- "Goodness" = intuition

Learning objectives

- Two ways of representing linguistic knowledge

1. Linguistic knowledge in the mind
   - Generative approach
   - Creates utterances from rules operating over minimal building blocks

- Interlude: rule-governed translation
  - Really a failure or just not enough time, money, computer speed and memory, and energy?

2. Linguistic knowledge from input
   - Exemplar approach
   - Creates utterances by re-using pieces (or any size) previously recorded