A probabilistic approach to Lexical-Functional Grammar

Ronald M. Kaplan

Xerox Palo Alto Research Center

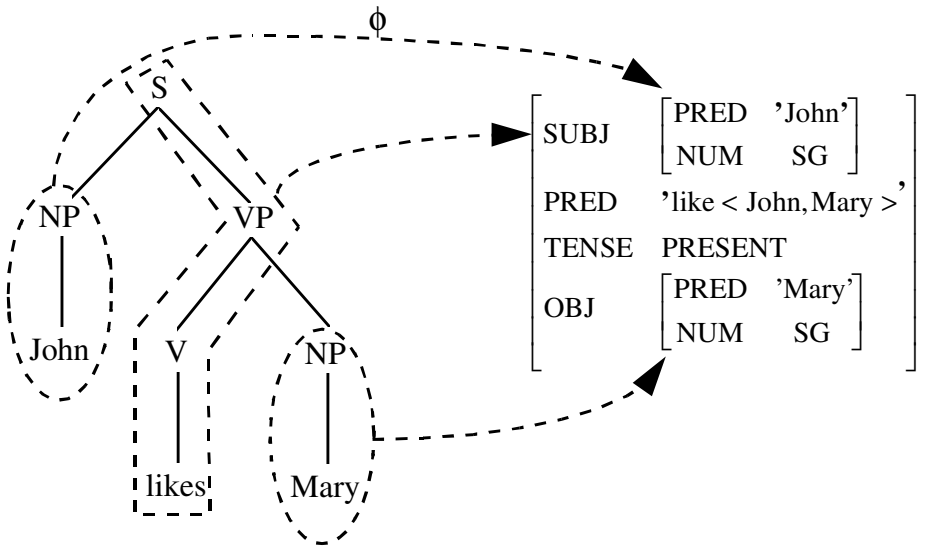
Based on joint work with

Rens Bod, Khalil Sima'an, Remko Scha

1. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Linguistic Theories provide

Representations



Formal encoding of grammatical relations

Rules

$$S \rightarrow \underset{(\uparrow SUBJ) = \downarrow}{NP} \quad \underset{\uparrow = \downarrow}{VP}$$

$$VP \rightarrow \underset{\uparrow = \downarrow}{V} \quad \underset{(\uparrow OBJ) = \downarrow}{NP}$$

Determine representations for all possible utterances

Usual goal: *minimal, nonredundant* set of *independent* generalizations with *free interactions*

Carry explanatory burden

2. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Competence Hypothesis

- Language user *applies* internalized rules to produce internal representations

- Language user *acquires* rules by abstraction of grammatical experience guided by universal principles and constraints

3. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Alternative view: Representations only, no rules

- Language user acquires *examples of representations* from syntactic experience
- Language user applies *operations on representations* to produce representations for new utterances
- Linguistic theory specifies representations and operations
- Rules perhaps appear in scientific discourse, but are not part of native speaker's "competence"

4. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Productivity from examples

(following Scha, Bod: *Data Oriented Parsing*)

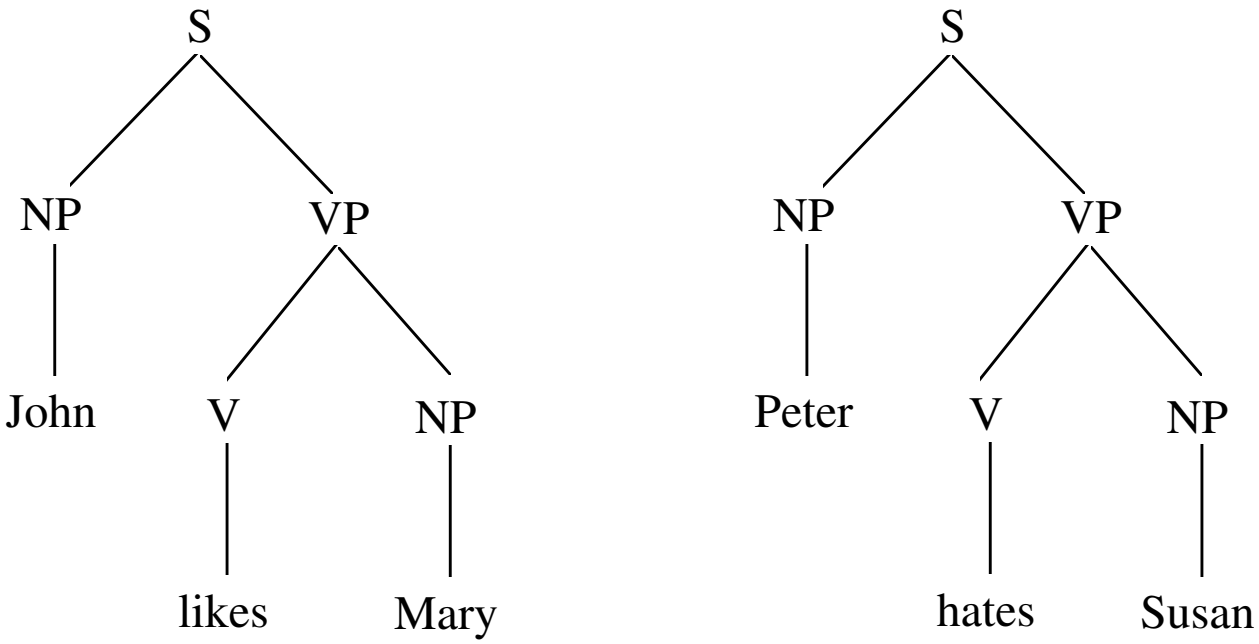
Given: corpus annotated with representations
(e.g. phrase structures)

1. Break structures into fragments--remember them

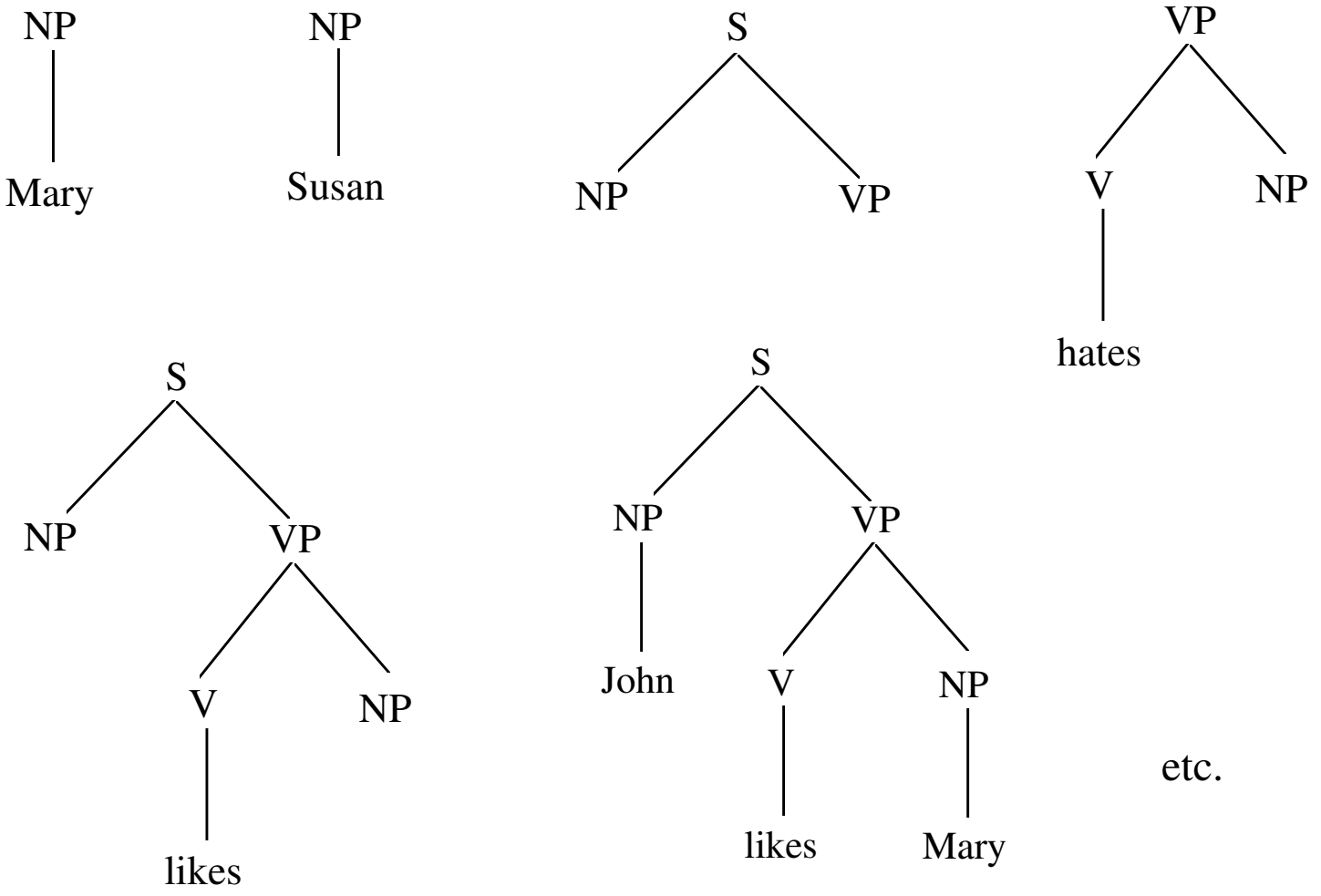
2. Combine fragments to get structures for new sentences

DOP illustration

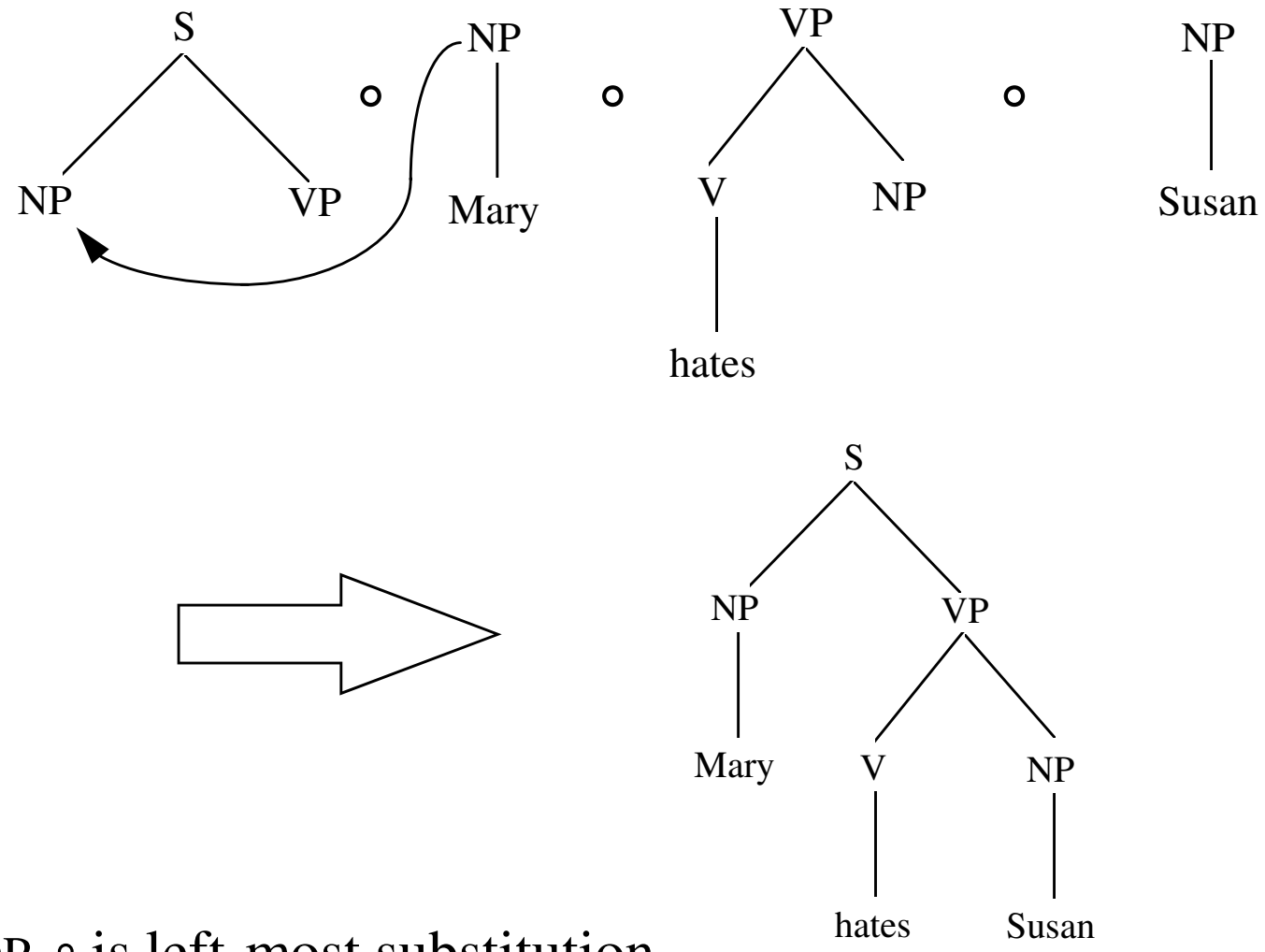
Given: corpus annotated with representations:



1. Break structures into fragments



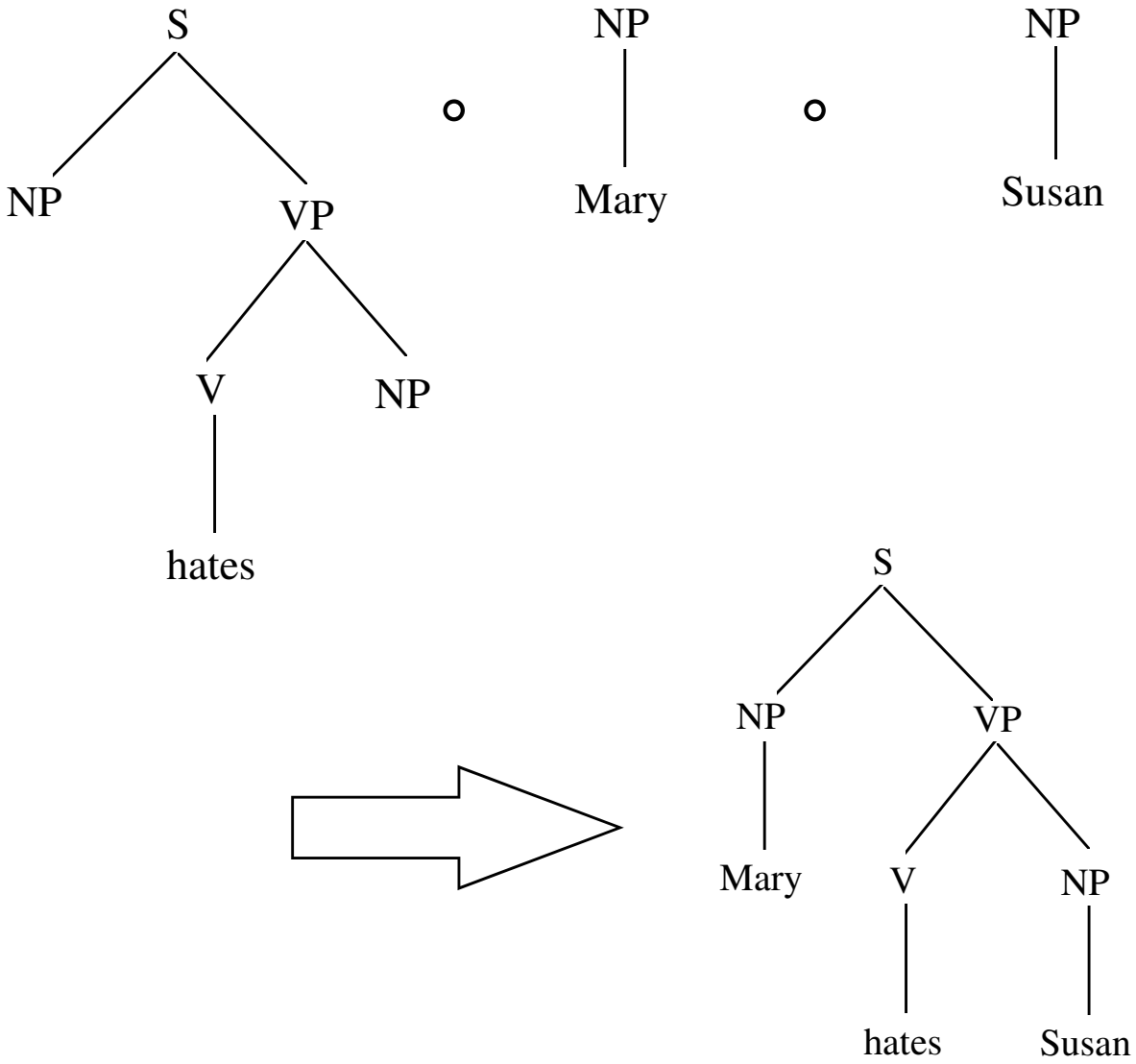
2. Combine fragments to get structures for new utterances



In DOP, ∘ is left-most substitution

8. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Another derivation of the same structure:

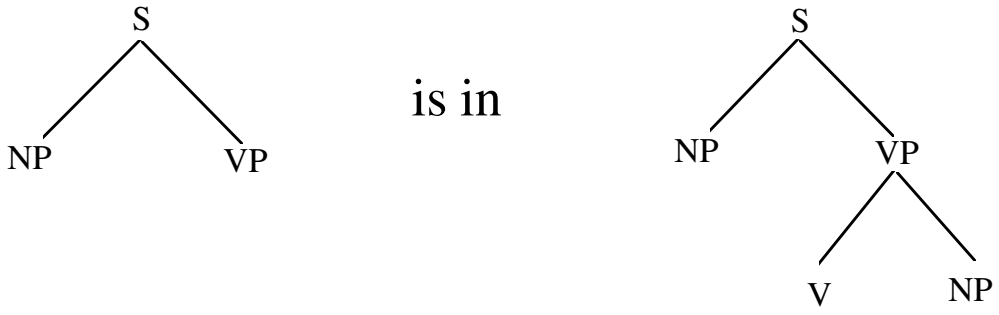


9. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Observations

- Fragments are not minimal
  - Range from context-free rule equivalent (S → NP VP) to whole-utterance structure.
  - Some large fragments may represent idiosyncratic constructions, others may not. We don't care.
  - We don't even care how many fragments there are (in principle).

    TAG?
- Fragments are redundant, with overlapping information.



- Multiple results, not derivations, correspond to ambiguity

10. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Probabilities

Resolve ambiguities, implicitly identify most useful fragments

- Frequency affects language-user interpretations: governs choice among several grammatical alternatives

  Mehler & Carey (68)....Tanenhaus and Trueswell (95)

- Typically, probabilities are defined on rules (stochastic grammars)

- DOP: Probabilities are defined on representations, not rules

  Scha (90) .... Bod (95)

11. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

A corpus-oriented, representation-based approach requires

1. A theory of well-formed utterance *representations*.
2. A definition of productive representation *fragments*.
3. A definition of a fragment-combination *operation* ∘.
4. A *probability model* for utterance representations.

A linguistic theory provides 1, 2, 3
but no other descriptive devices

For DOP

1. Representations: phrase structure trees.

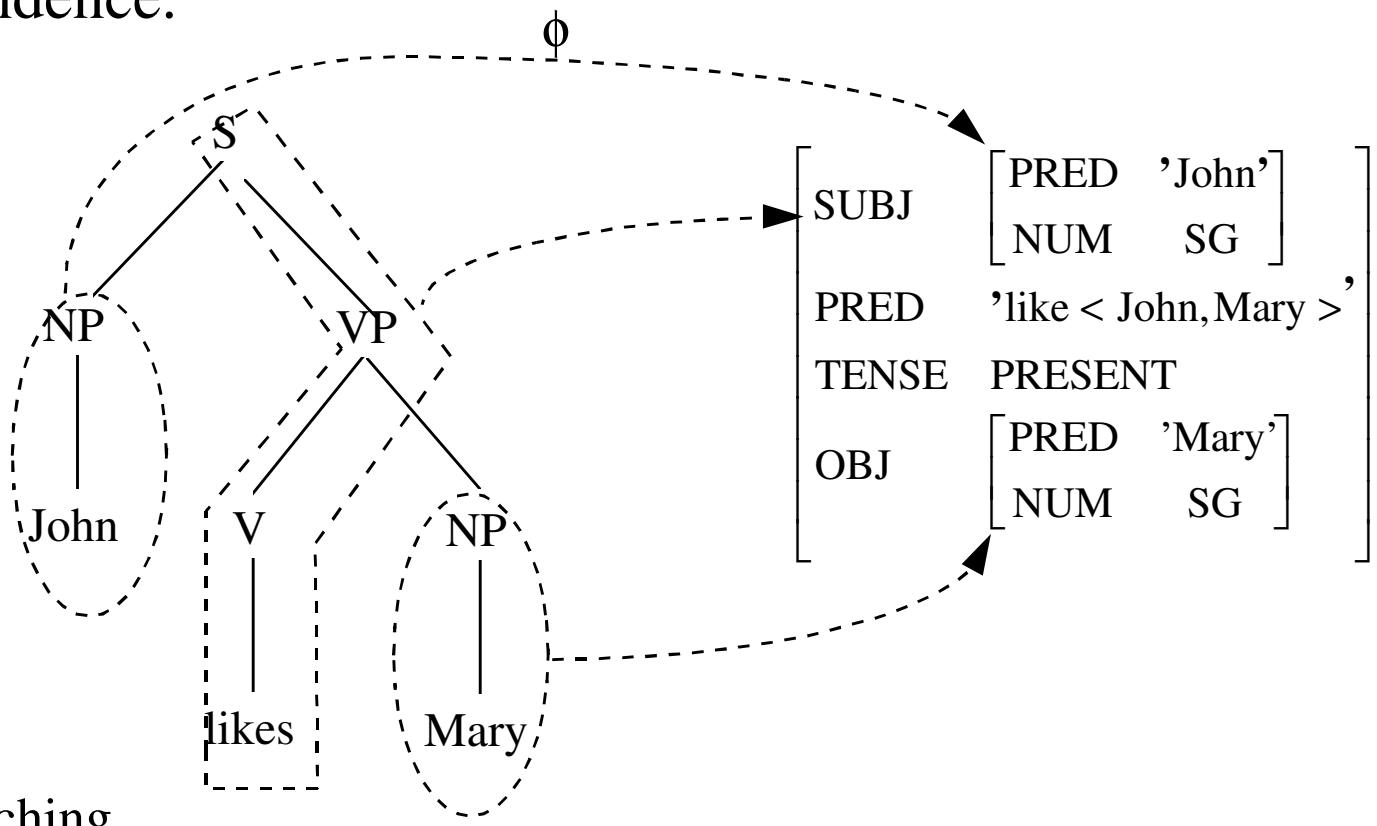
2. Fragments: connected subtrees.

3. Operation $\circ$: substitution of leftmost matching category.

4. Probability model: ...later.

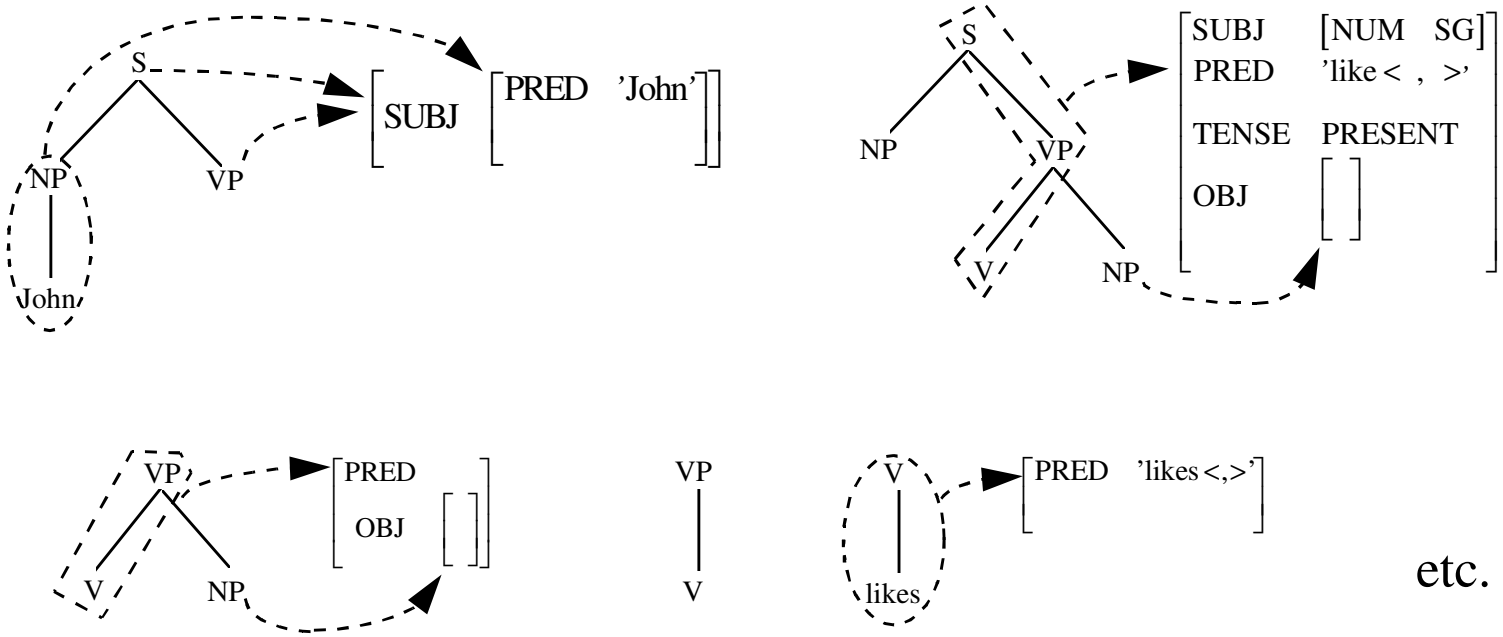
For LFG:

1. Representations: valid* c-structures and f-structures in correspondence.



*No nonbranching dominance chains

2. Fragments: loosely, connected subtrees in correspondence with connected sub-f-structures



Intuition says: some possible fragments are implausible

Examples of theory-based restrictions

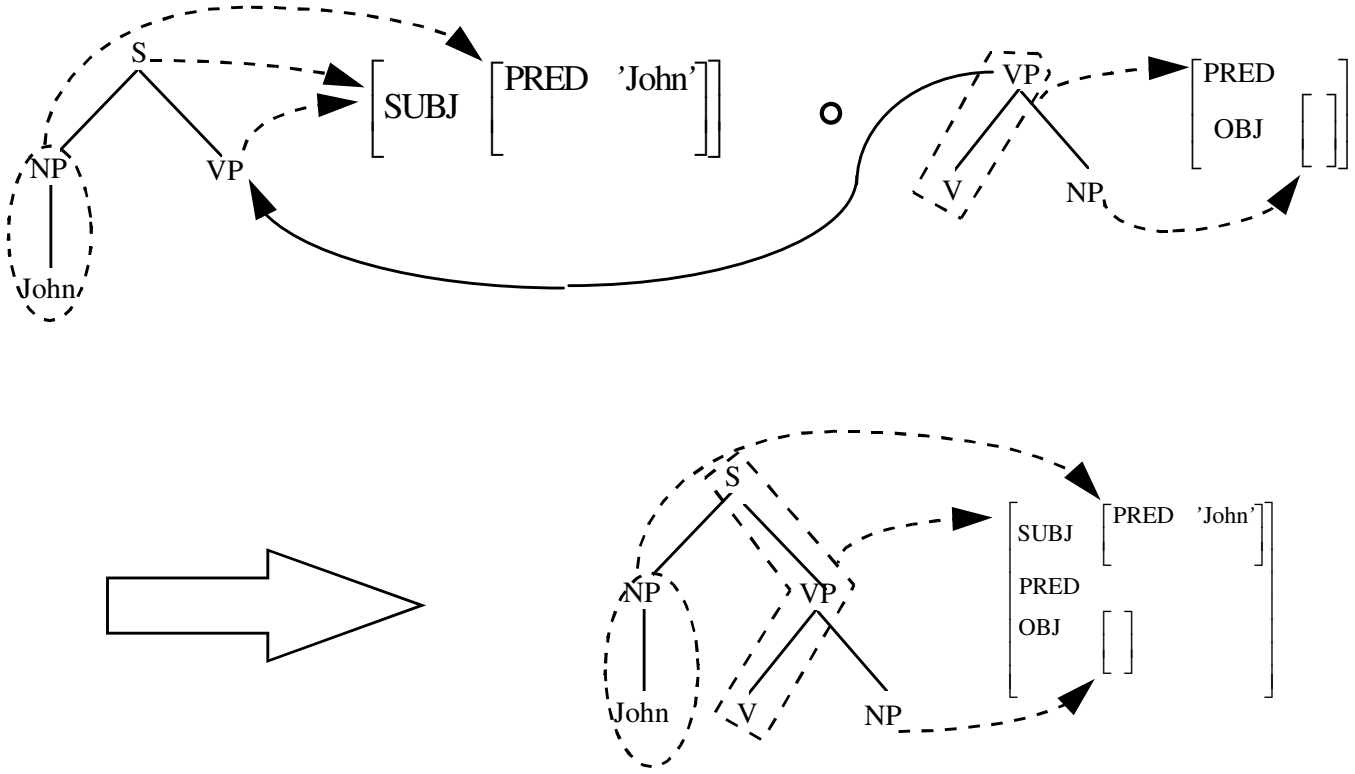
Lexical predicates: If a fragment includes an f-structure lexical predicate, the fragment must also include a corresponding lexical node.

Head chains: If a f ragment includes node *n* corresponding to f-structure *f*, then all other nodes under *n* corresponding to *f* must be included.

Control: If a fragment contains one path of a control identity, it must contain the other.

Sisters: If a fragment includes a node *n*, it must include all of *n*'s sisters (from DOP).

3. Operation: Left-most substitution of matching categories followed by unification of corresponding fragment f-structures



Derivation

A *derivation* for an utterance *u* is a sequence of fragments $\langle f_1, f_2 ... f_n \rangle$ such that the composition operator $\circ$ applied from left to right results in a valid representation *R* whose yield is *u*:

$$R = (...((f_1 \circ f_2) \circ ...) \circ f_n)$$

$$= \langle \text{c-structure}, \phi, \text{f-structure} \rangle$$

Theory of representation defines "valid":

e.g. no nonbranching dominance chains,
complete and coherent f-structure.

Theory of representation defines "yield":

e.g the terminal string of the c-structure.

4. Probability Model

Let *C* be a corpus of structures and *Bag(C)* be the bag containing all fragments derived from *C*. #(*f*) is the number of times that fragment *f* appears in the bag.

The probability of each fragment is estimated by its corpus frequency:

$$P(f) = \frac{\#(f)}{\sum_{g \in Bag(C)} \#(g)}$$

Probability of a derivation

A *derivation* for an utterance *u* results in a representation *R* whose yield is *u*.

- We assume a fragment sequence $s = \langle f_1, f_2 ... f_n \rangle$ is constructed from the bag by random sampling with replacement. Then its sequence probability is

$$P(s) = \prod_i P(f_i)$$

- There may be infinitely many sequences that result in no representation or which result in a representation whose yield is not *u*. We are not interested in those. For a given derivation *d* of *u* we obtain

$$P(d | d \text{ yields } u) = \frac{P(d)}{\sum_{s \text{ yields } u} P(s)}$$

The linguistic theory must guarantee for every *u* a maximum derivation length. (E.g. no nonbranching chains)

20. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.

Probability of an utterance representation

In general there are many derivations of a particular representation *R* for an utterance *u*. Assuming these derivations are independent, we have

$$P(R) = \sum_{d \text{ results in } R} P(d | d \text{ yields } u)$$

We assign the most probable *R* as the best analysis of *u*.

The most probable *R*: the one most likely to have been derived.

Other approaches

- Stochastic grammars: Assign probabilities to rules

  The most probable *R*: the one with the most probable derivation

- Johnson (1996): Assign probabilities to f-structure relations

  The most probable *R*: the one with the most probable f-structure independent of any derivation

  "Model theory vs. proof theory"

Summary

- A productive system based on representations, not rules
- Clear, but different, role for linguistic theory
- Different claims about what a native-speaker “knows”, what needs to be explained
- Theory of acquisition combined with theory of processing

(Although it may be impractical...)

23. R. M. Kaplan, *A probabilistic approach to LFG*, LFG Colloquium and Workshops, Rank Xerox Research Centre, Grenoble, August 1996.