Context-Free Translation Models

Adam Lopez
University of Edinburgh → Johns Hopkins University

Finite-State Models

Finite-State Models

- Very simple models get us pretty far!

Finite-State Models

- Very simple models get us pretty far!
- There’s no data like more data.

Finite-State Models

- Very simple models get us pretty far!
- There’s no data like more data.
- Word-based models follow intuitions, but not all.

Finite-State Models

- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.

Finite-State Models

- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.
- *All* of these models are weighted regular languages.

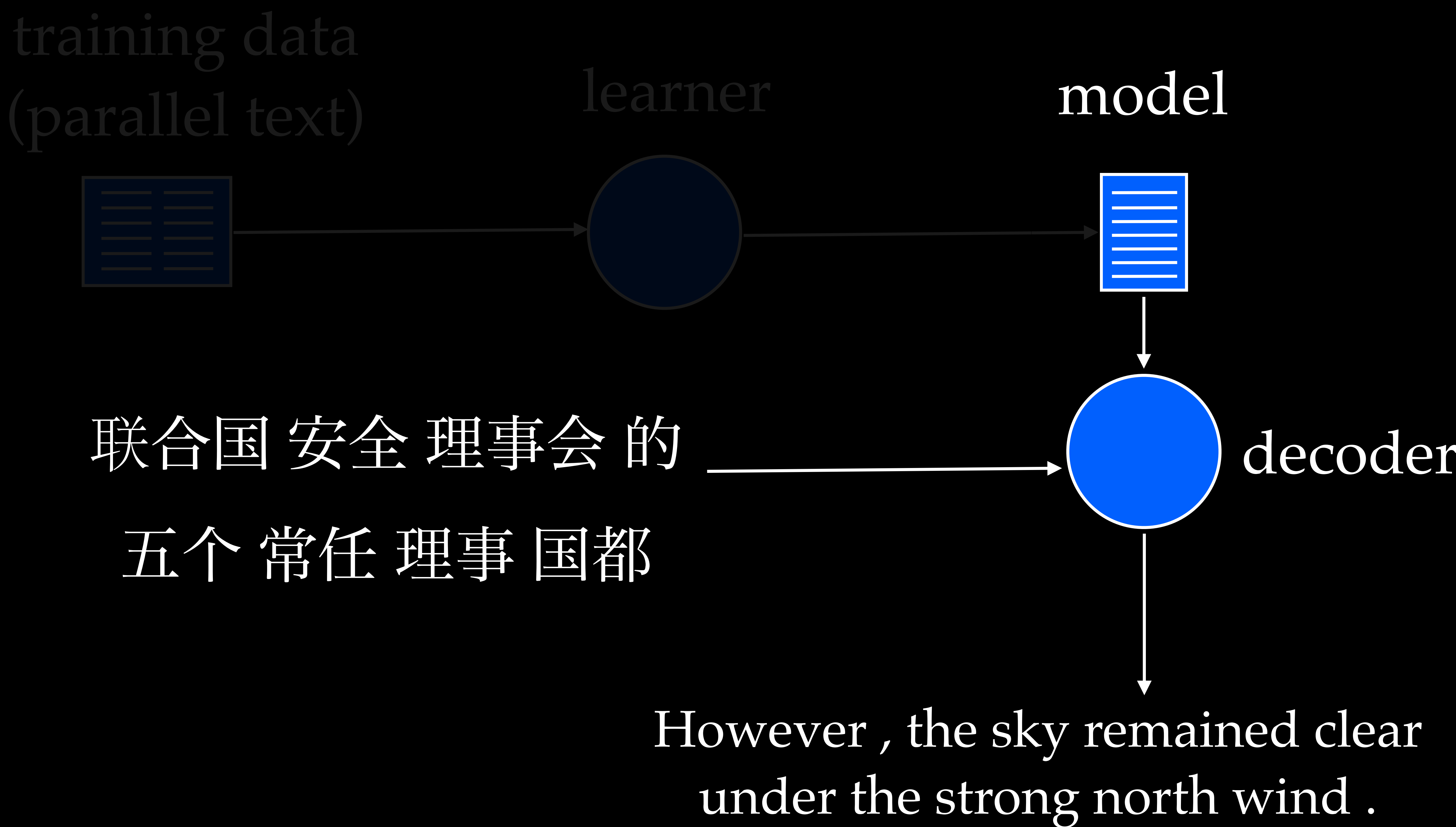
Finite-State Models

- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.
- *All* of these models are weighted regular languages.
- Need dynamic programming with approximations.

Finite-State Models

- Very simple models get us pretty far!
- There's no data like more data.
- Word-based models follow intuitions, but not all.
- Phrase-based models are similar, but more effective.
- *All* of these models are weighted regular languages.
- Need dynamic programming with approximations.
- Is this the best we can do?

Overview


training data
(parallel text)
learner
model
decoder
联合国 安全 理事会 的
五个 常任 理事 国都
However , the sky remained clear
under the strong north wind .

Two Problems

- Exact decoding requires exponential time.
  - This is a consequence of arbitrary permutation.
  - But in translation reordering is not arbitrary!
- Parameterization of reordering is weak.
  - No generalization!

la empresa tiene enemigos fuertes en Europa .

the company has strong enemies in Europe .

Garcia and associates .

Garcia y asociados .

Carlos Garcia has three associates .

Carlos Garcia tiene tres asociados .

his associates are not strong .

sus asociados no son fuertes .

Garcia has a company also .

Garcia tambien tiene una empresa .

its clients are angry .

sus clientes estan enfadados .

the associates are also angry .

los asociados tambien estan enfadados .

the clients and the associates are enemies .

los clientes y los asociados son enemigos .

the company has three groups .

la empresa tiene tres grupos .

its groups are in Europe .

sus grupos estan en Europa .

the modern groups sell strong pharmaceuticals .

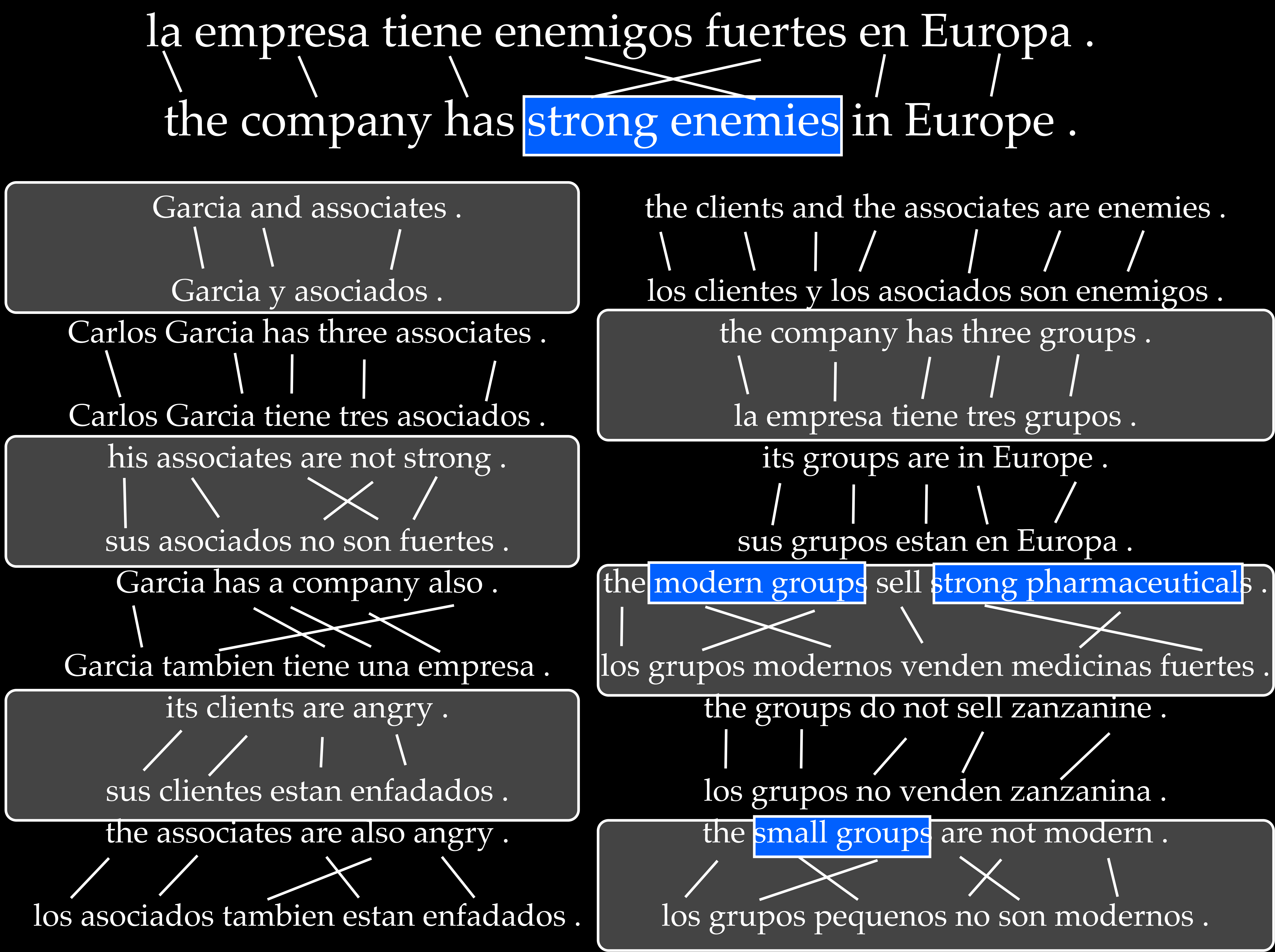
los grupos modernos venden medicinas fuertes .

the groups do not sell zanzanine .

los grupos no venden zanzanina .

the small groups are not modern .

los grupos pequenos no son modernos .



la empresa tiene enemigos fuertes en Europa .
the company has strong enemies in Europe .
Garcia and associates .
Garcia y asociados .
Carlos Garcia has three associates .
Carlos Garcia tiene tres asociados .
his associates are not strong .
sus asociados no son fuertes .
Garcia has a company also .
Garcia tambien tiene una empresa .
its clients are angry .
sus clientes estan enfadados .
the associates are also angry .
los asociados tambien estan enfadados .
the clients and the associates are enemies .
los clientes y los asociados son enemigos .
the company has three groups .
la empresa tiene tres grupos .
its groups are in Europe .
sus grupos estan en Europa .
the modern groups sell strong pharmaceuticals .
los grupos modernos venden medicinas fuertes .
the groups do not sell zanzanine .
los grupos no venden zanzanina .
the small groups are not modern .
los grupos pequenos no son modernos .

la empresa tiene enemigos fuertes en Europa .

the company has strong enemies in Europe .

Same pattern:

NN JJ → JJ NN

sus asociados no son fuertes .

Garcia has a company also .

Garcia tambien tiene una empresa .

its clients are angry .

sus clientes estan enfadados .

the associates are also angry .

los asociados tambien estan enfadados .

sus grupos estan en Europa .

the modern groups sell strong pharmaceuticals .

los grupos modernos venden medicinas fuertes .

the groups do not sell zanzanine .

los grupos no venden zanzanina .

the small groups are not modern .

los grupos pequenos no son modernos .

la empresa tiene enemigos fuertes en Europa .

the company has **strong enemies** in Europe .

Same pattern:

NN JJ → JJ NN

Finite-state models do not capture this generalization.

sus asociados no son fuertes .

Garcia has a company also .

Garcia tambien tiene una empresa .

its clients are angry .

sus clientes estan enfadados .

the associates are also angry .

los asociados tambien estan enfadados .

sus grupos estan en Europa .

the **modern groups** sell **strong pharmaceuticals** .

los grupos modernos venden medicinas fuertes .

the groups do not sell zanzanine .

los grupos no venden zanzanina .

the **small groups** are not modern .

los grupos pequenos no son modernos .

Context-Free Grammar

Context-Free Grammar

S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu

Context-Free Grammar

S

S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu

Context-Free Grammar

S

S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu

Context-Free Grammar

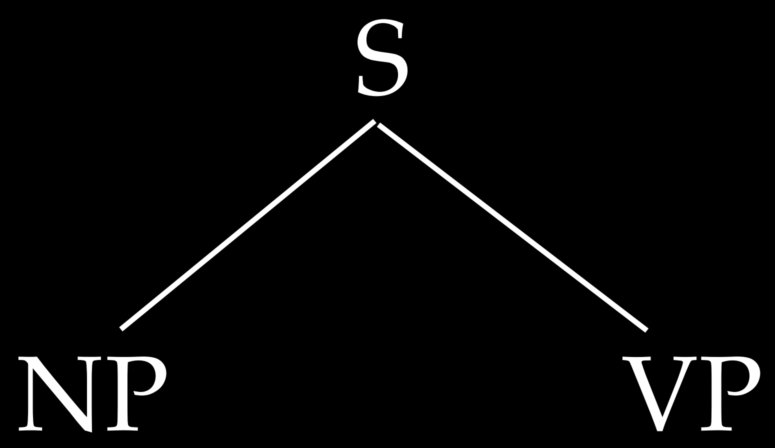
S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu



Context-Free Grammar

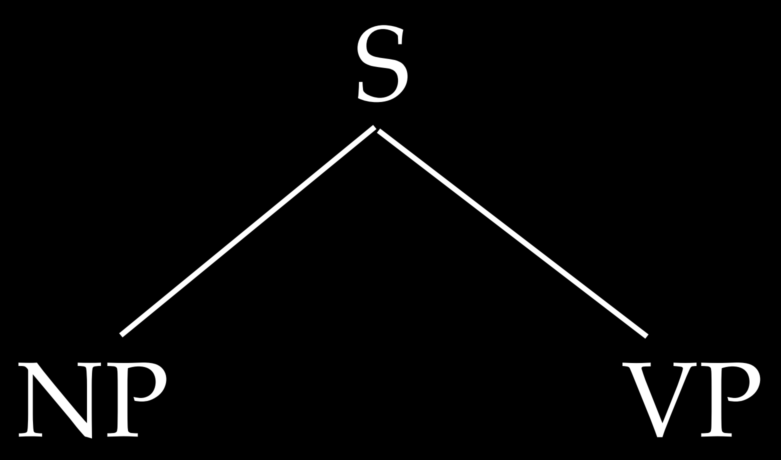
S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu



Context-Free Grammar

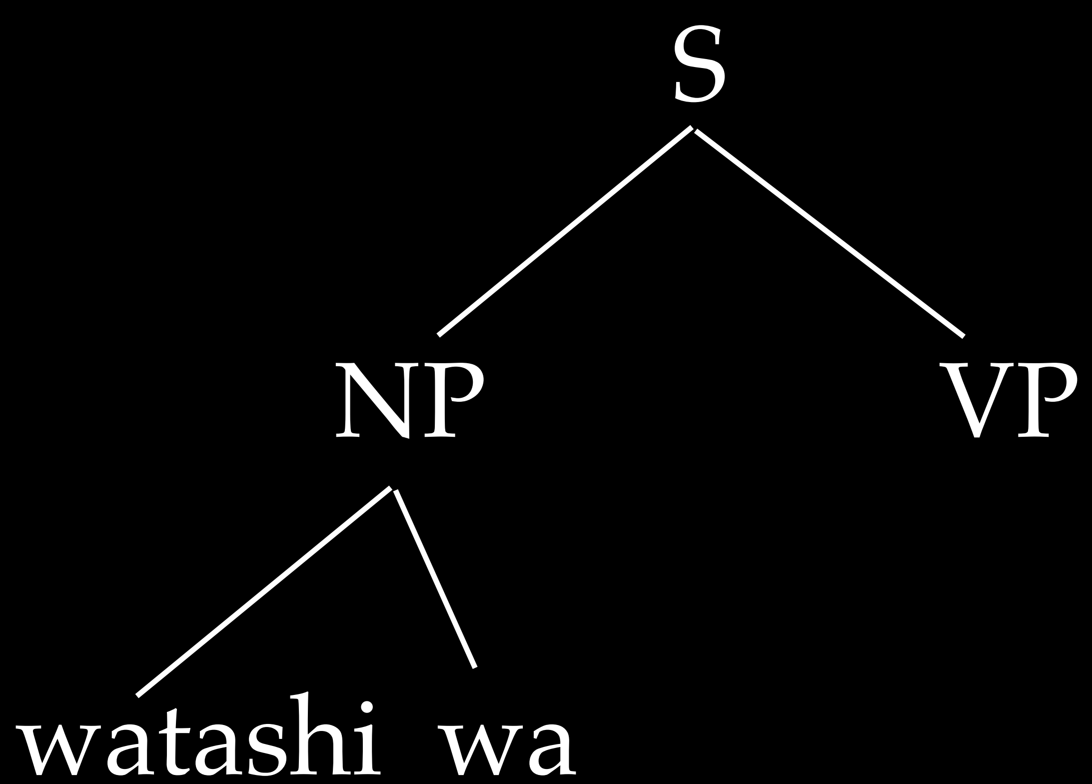
S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu



Context-Free Grammar

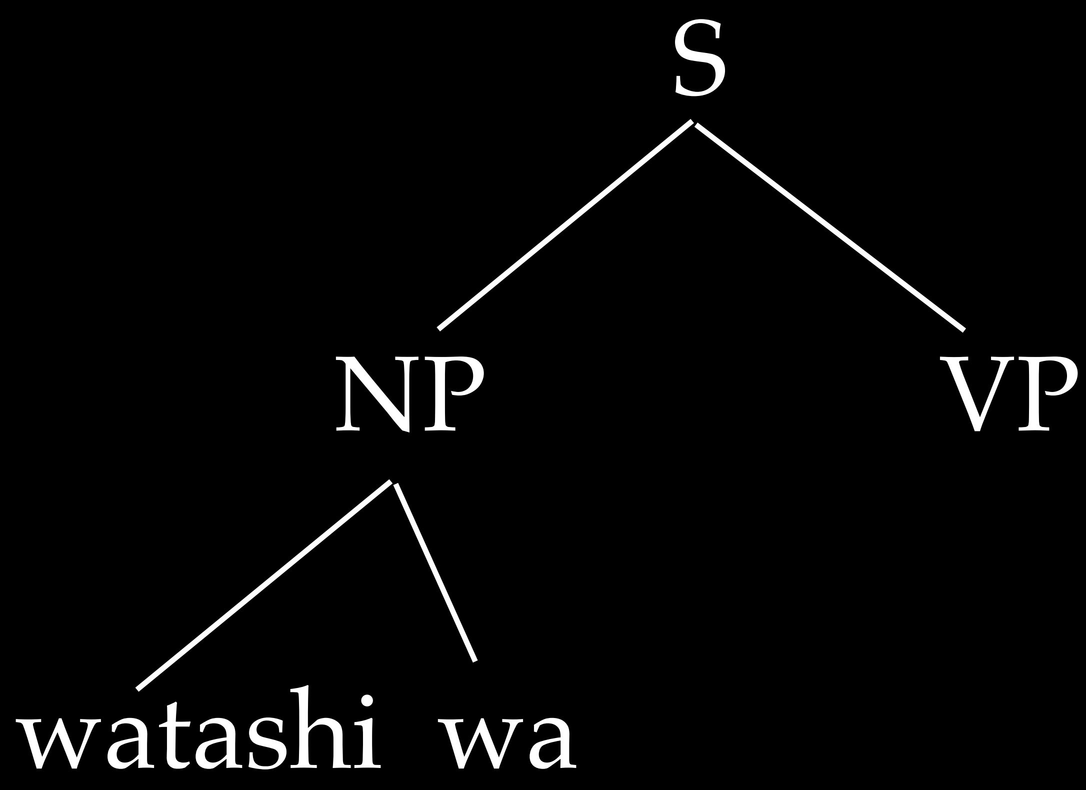
S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu



Context-Free Grammar

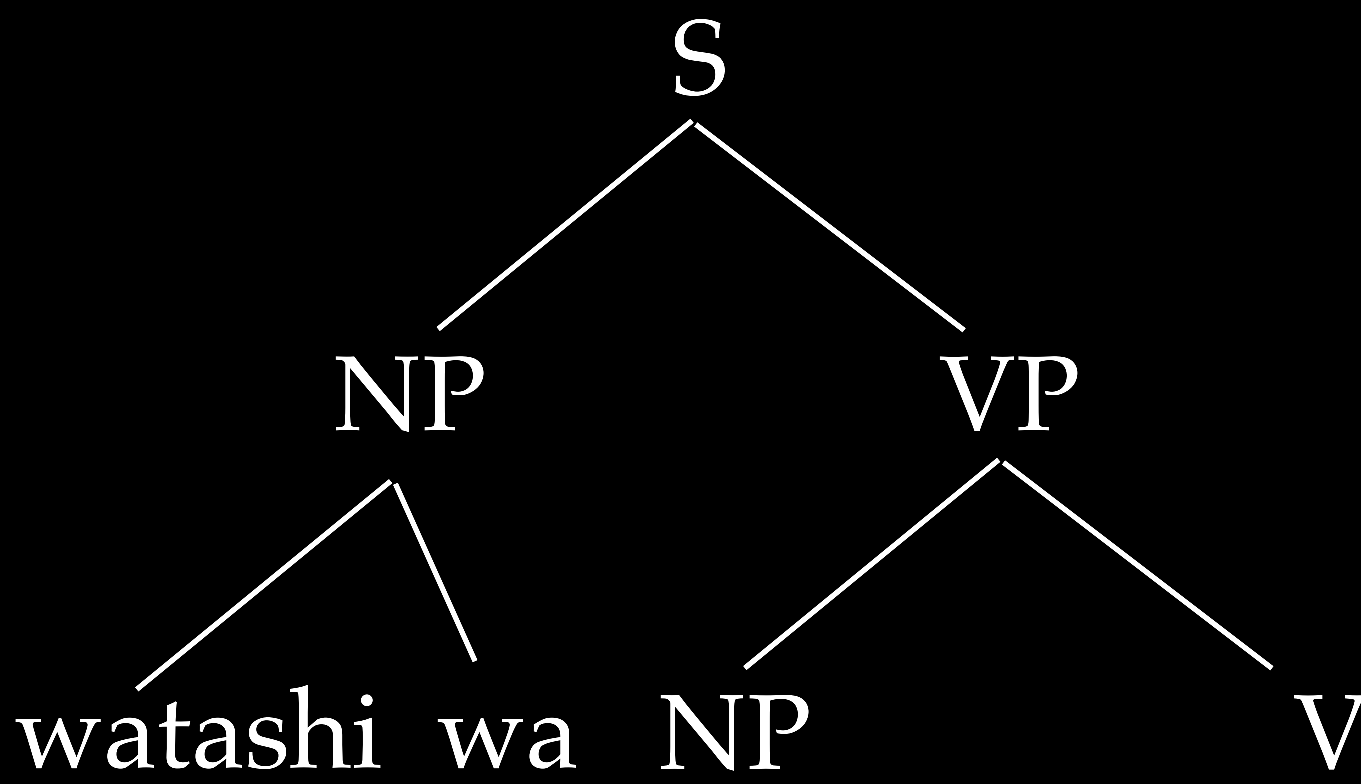
S → NP VP

NP → watashi wa

NP → hako wo

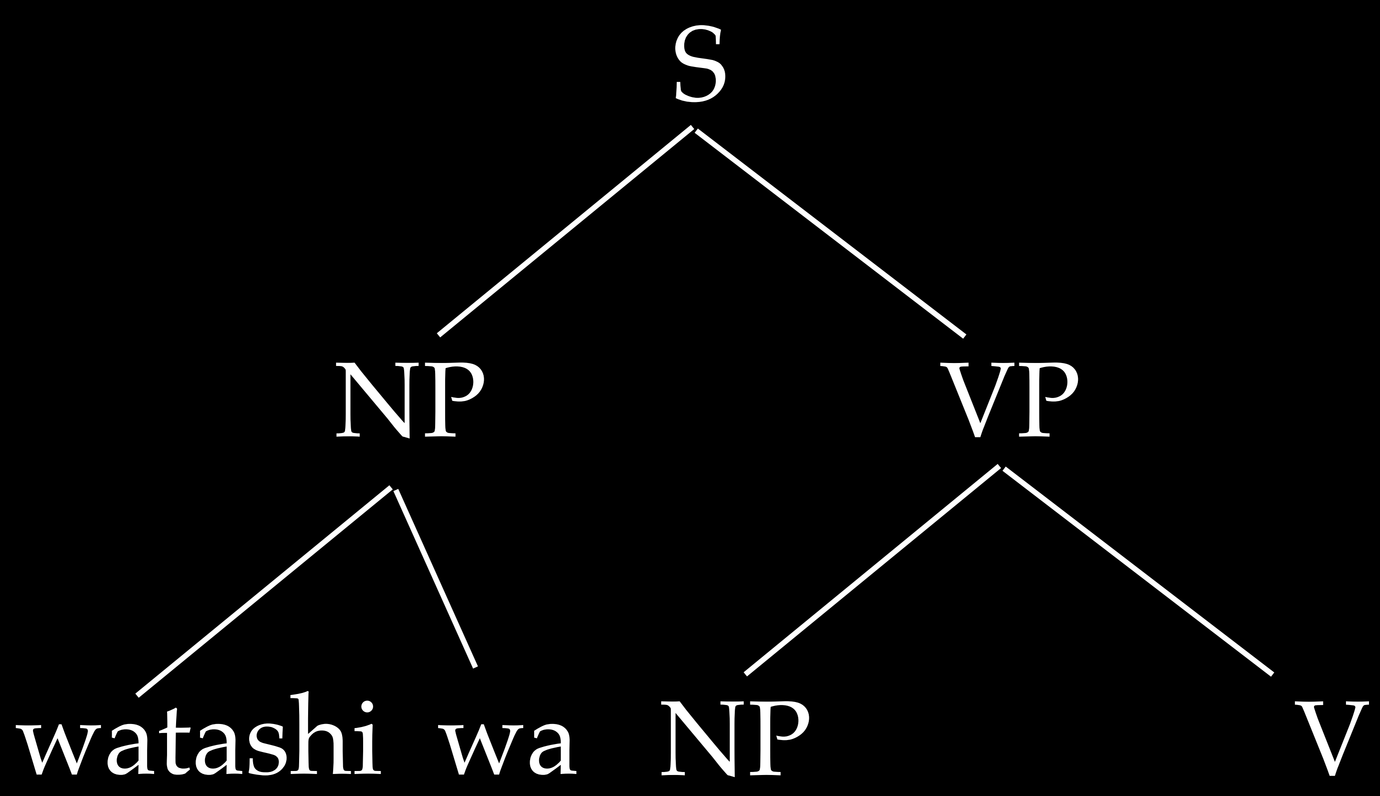
VP → NP V

V → akemasu



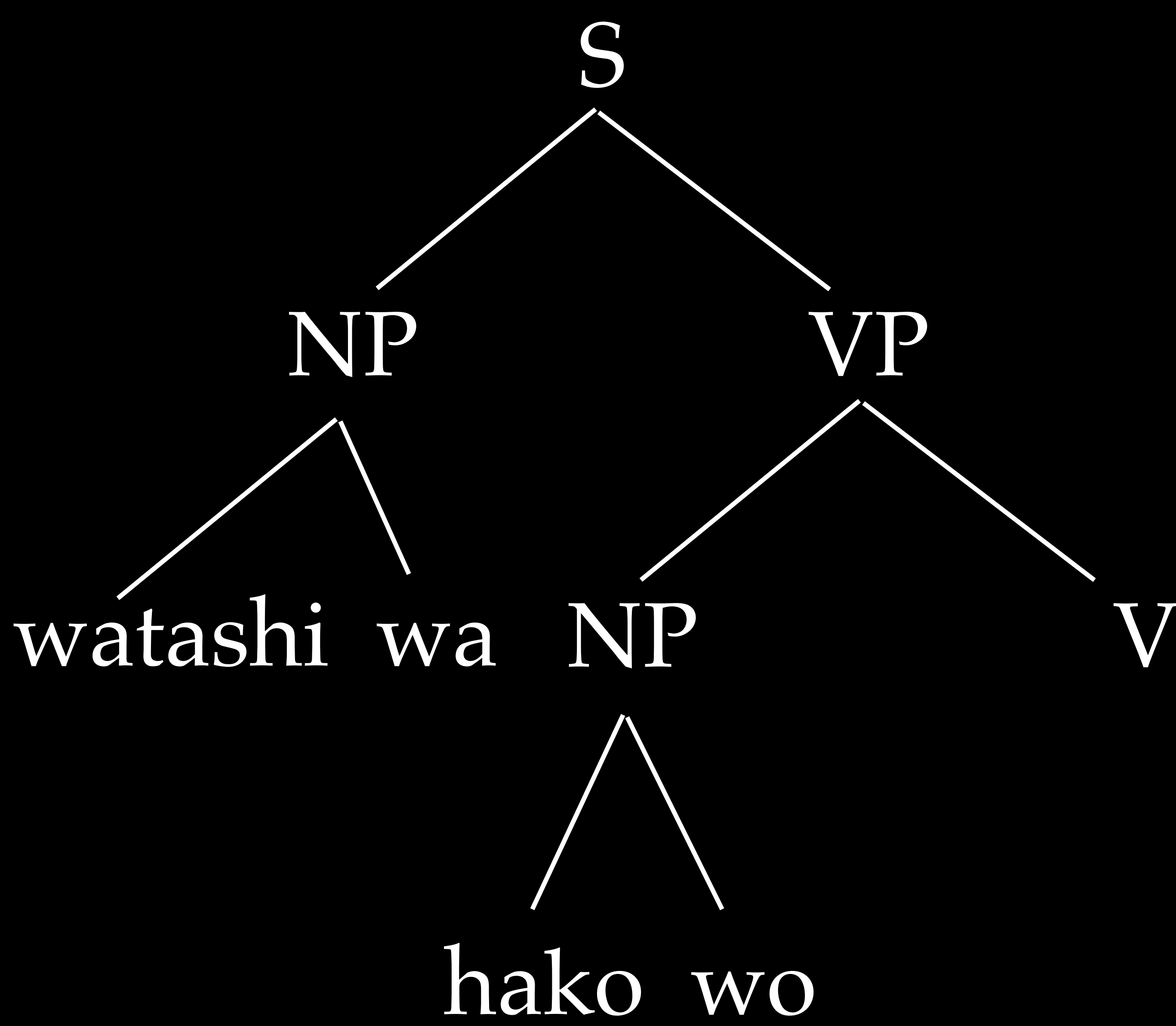
Context-Free Grammar

S → NP VP
NP → watashi wa
NP → hako wo
VP → NP V
V → akemasu



Context-Free Grammar

S → NP VP
NP → watashi wa
NP → hako wo
VP → NP V
V → akemasu



Context-Free Grammar

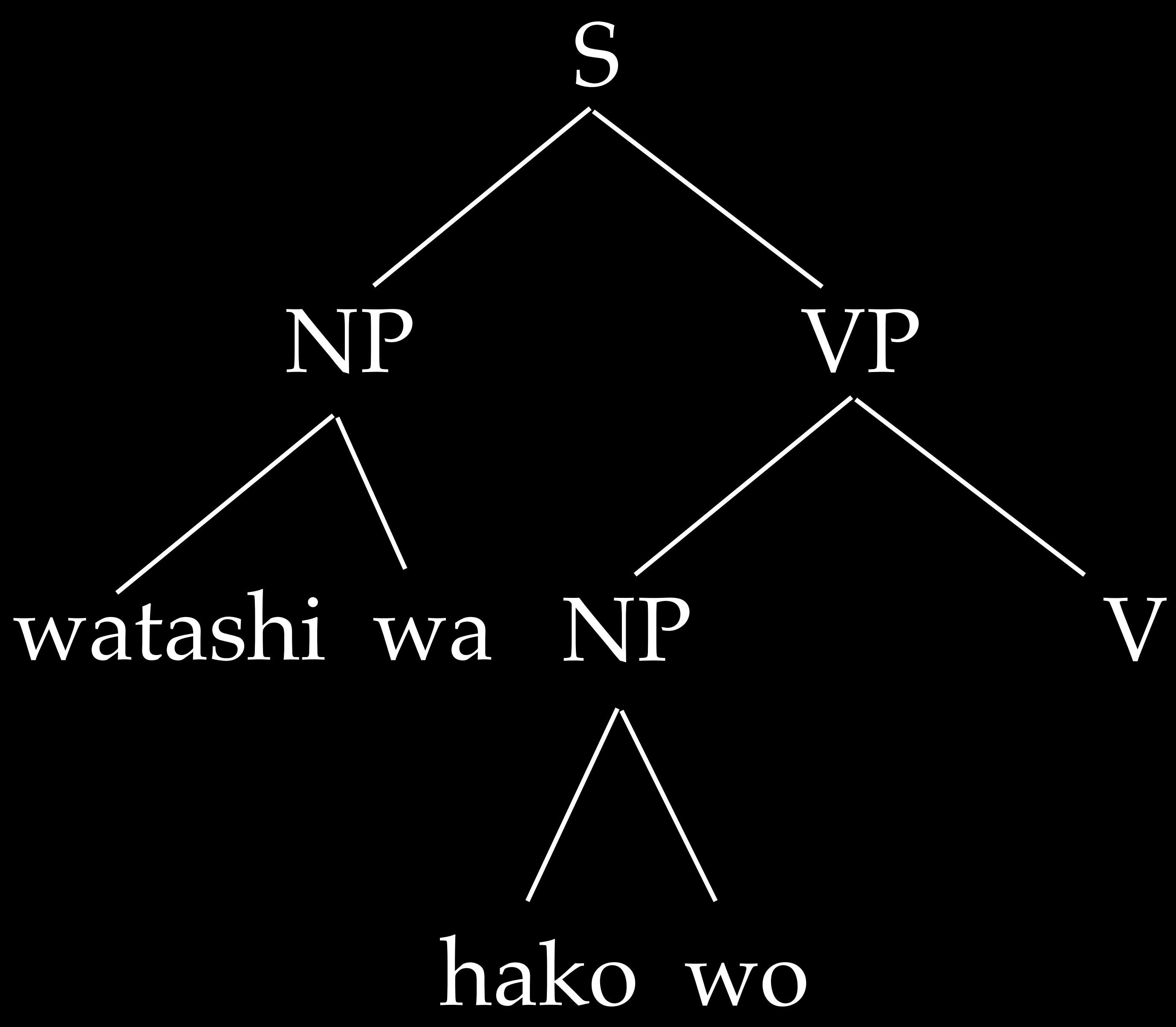
S → NP VP

NP → watashi wa

NP → hako wo

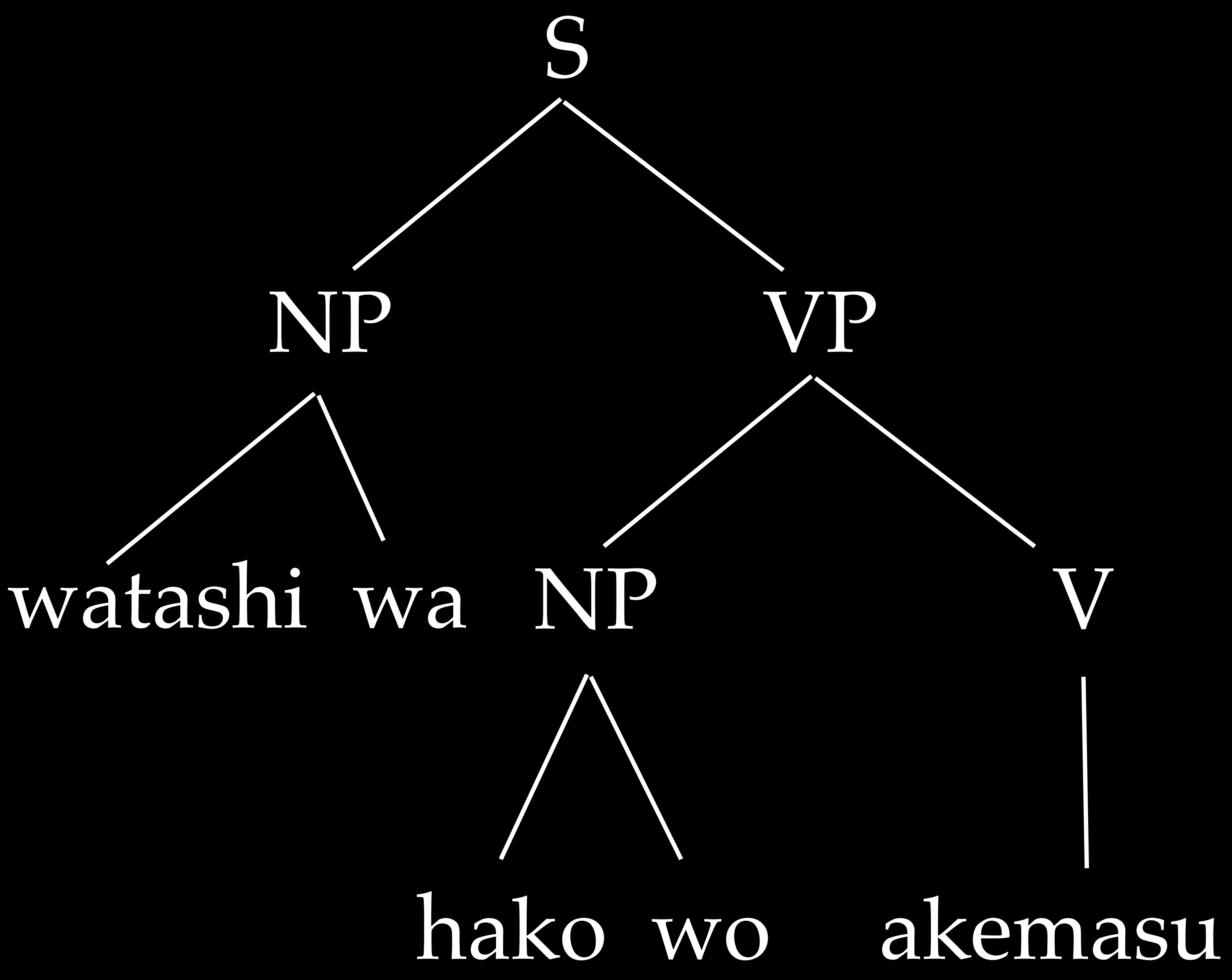
VP → NP V

V → akemasu



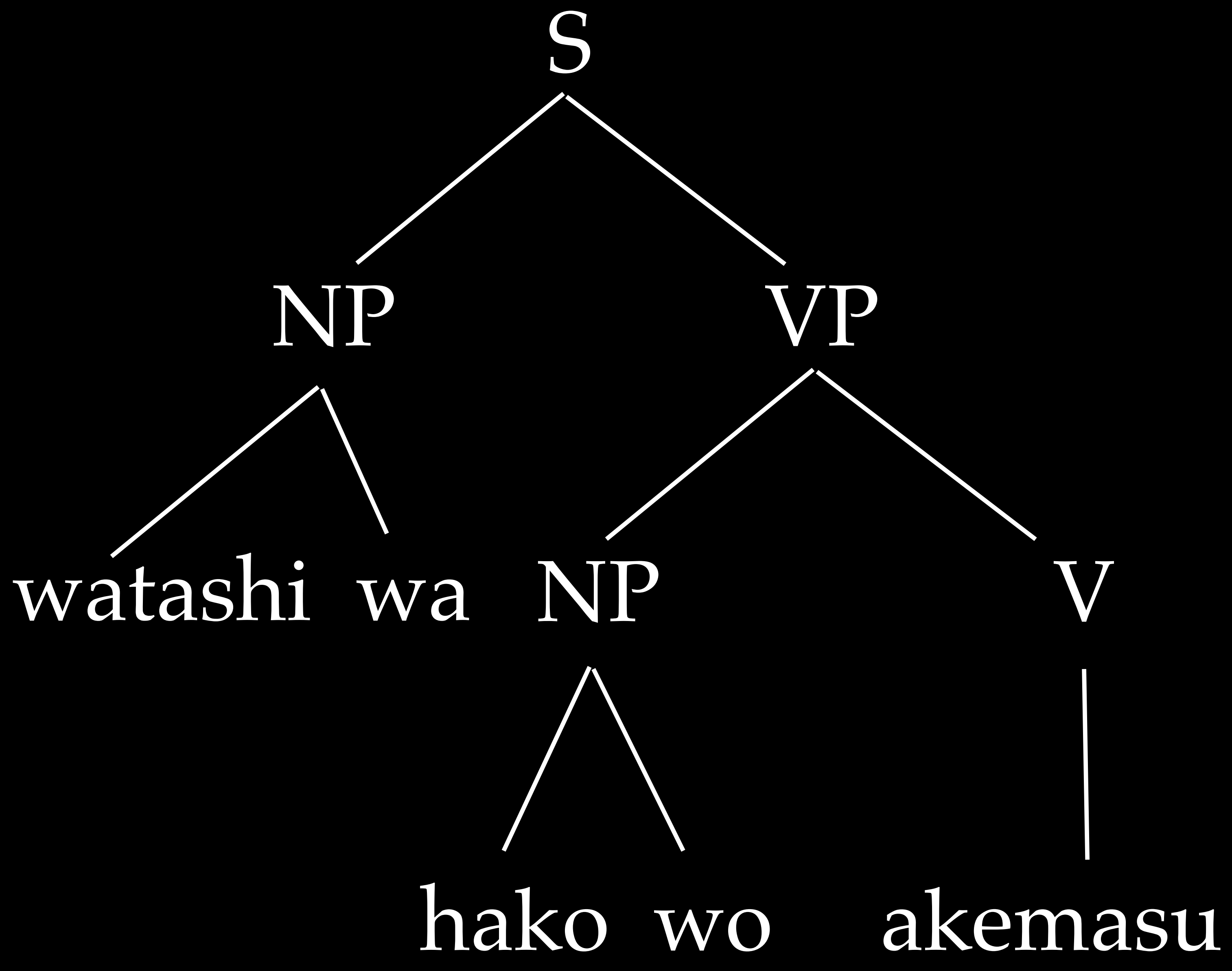
Context-Free Grammar

S → NP VP
NP → watashi wa
NP → hako wo
VP → NP V
V → akemasu



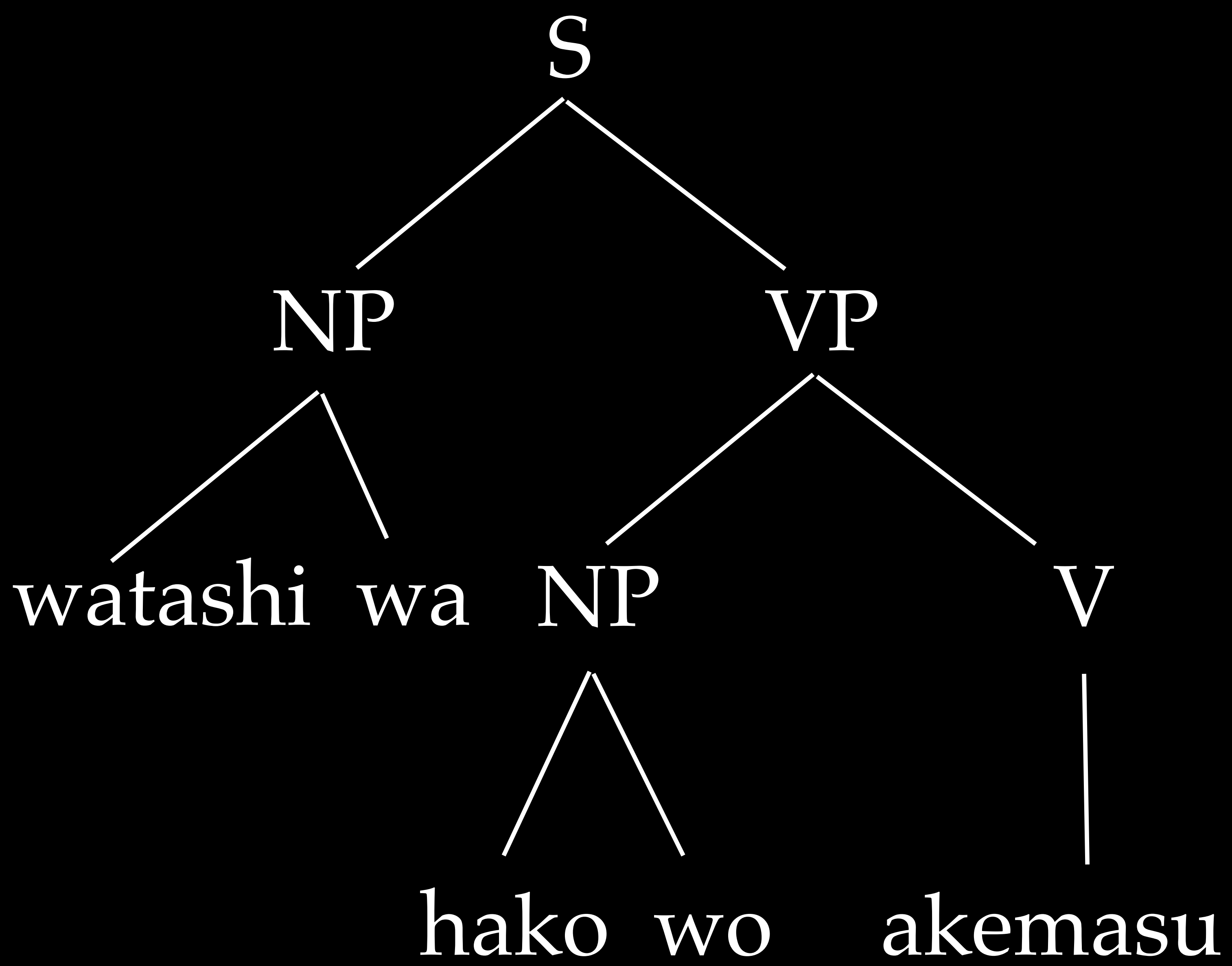
Context-Free Grammar

S → NP VP
NP → watashi wa
NP → hako wo
VP → NP V
V → akemasu



Context-Free Grammar

S → NP VP
NP → watashi wa
NP → hako wo
VP → NP V
V → akemasu



watashi wa hako wo akemasu

Synchronous Context-Free Grammar

S → NP VP

NP → watashi wa

NP → hako wo

VP → NP V

V → akemasu

Synchronous Context-Free Grammar

S → NP VP	S → NP VP
NP → watashi wa	NP → I
NP → hako wo	NP → the box
VP → NP V	VP → V NP
V → akemasu	V → open

Synchronous Context-Free Grammar

$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar

$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar

S S

$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar

S .... S

$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar

S .... S

$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar

$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

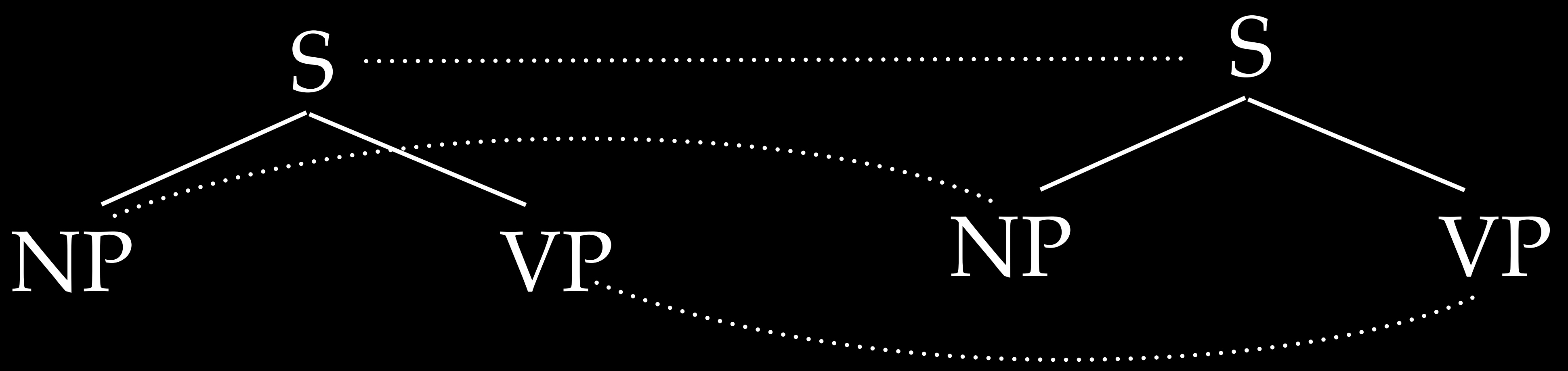
NP → watashi wa / I

NP → hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

V → akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

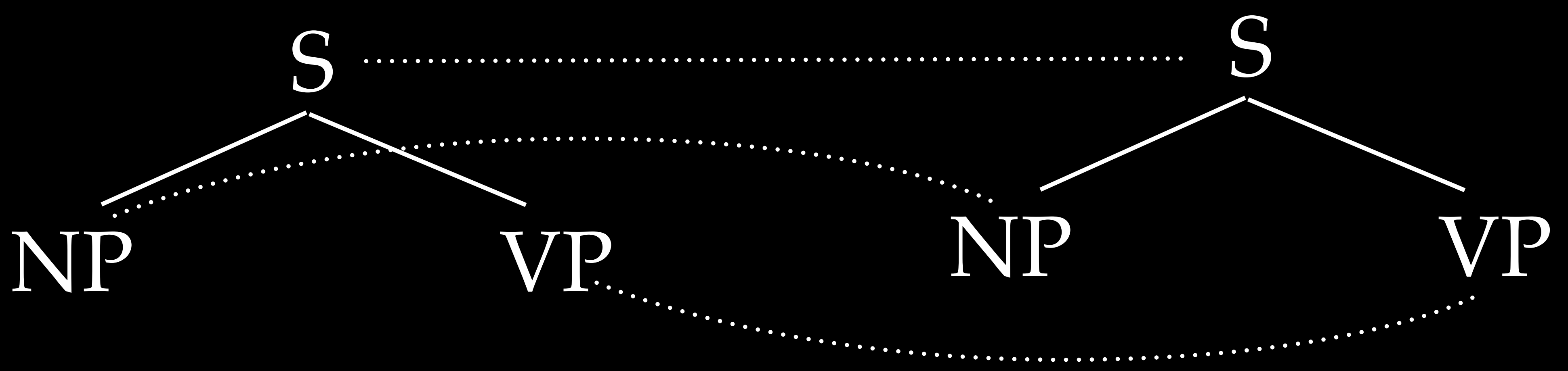
NP → watashi wa / I

NP → hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

V → akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

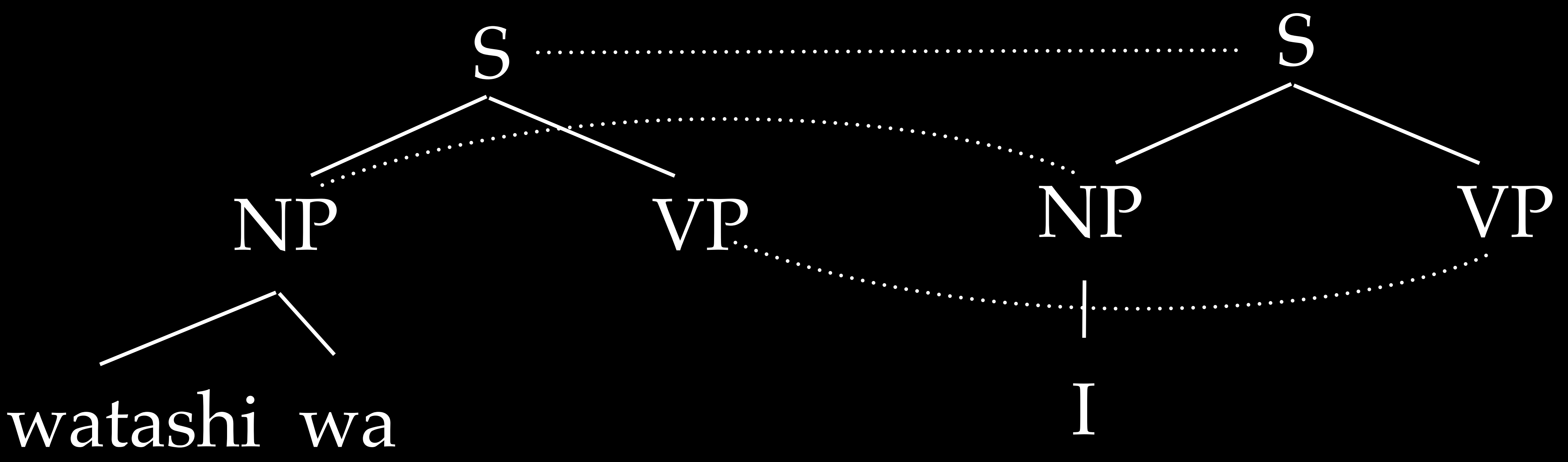
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

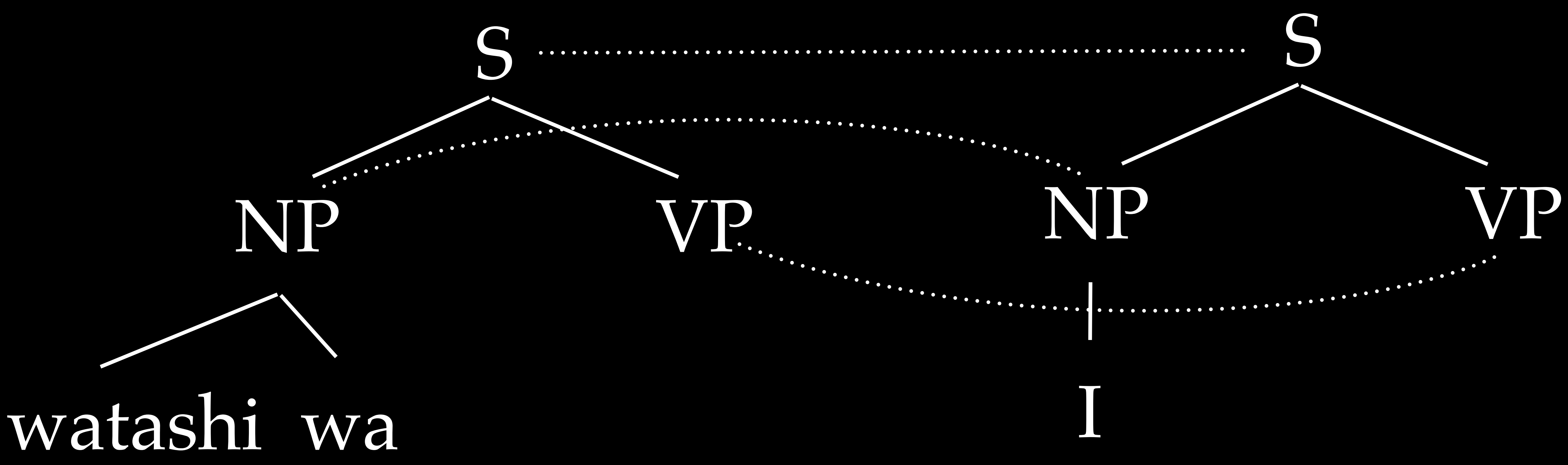
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

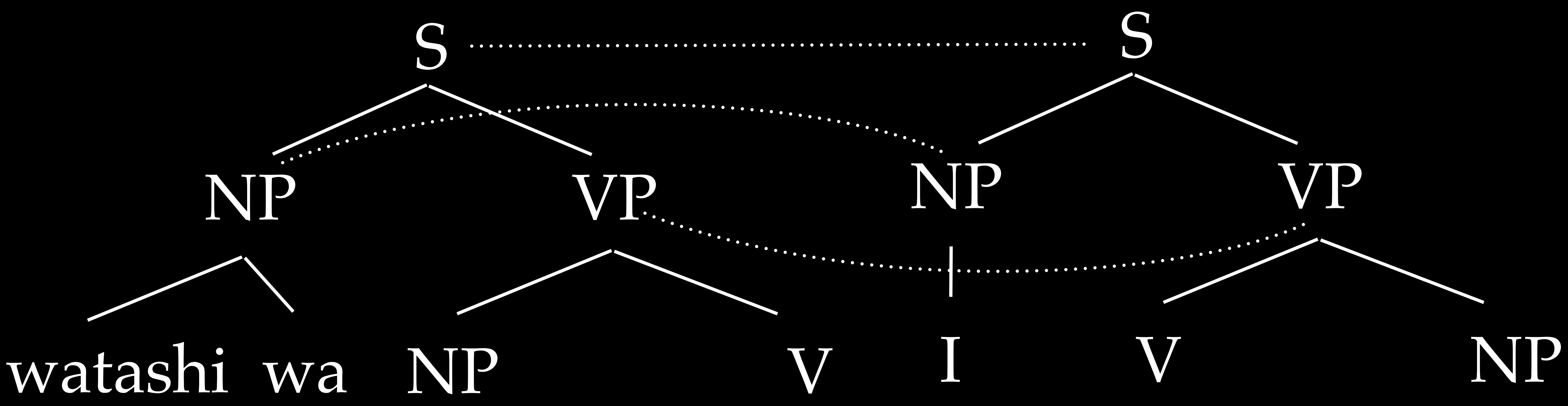
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

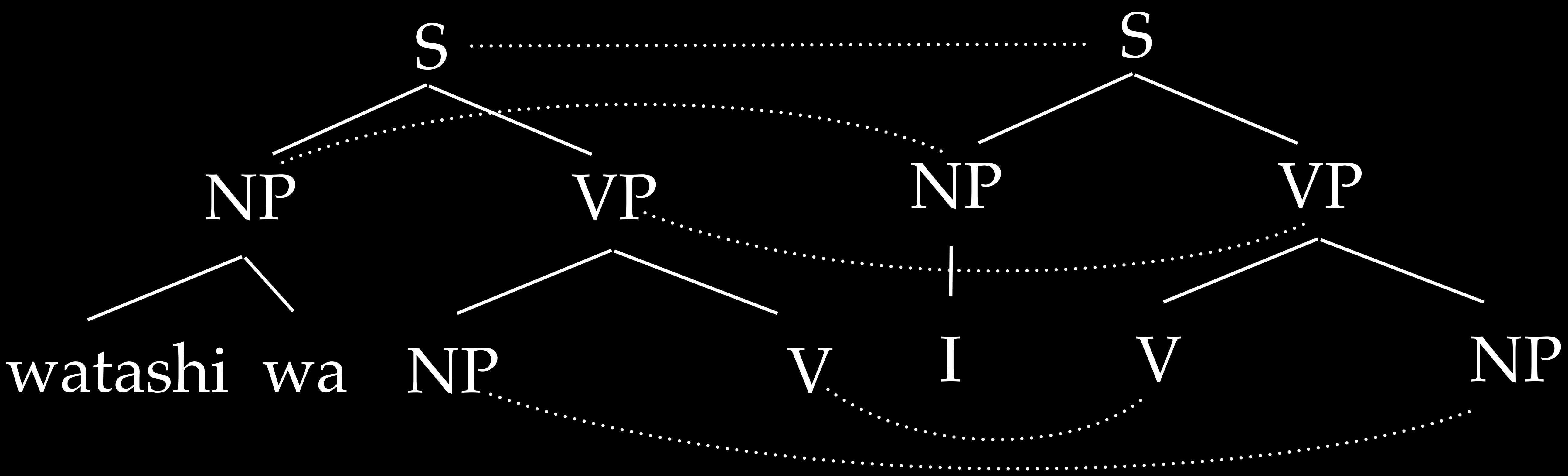
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

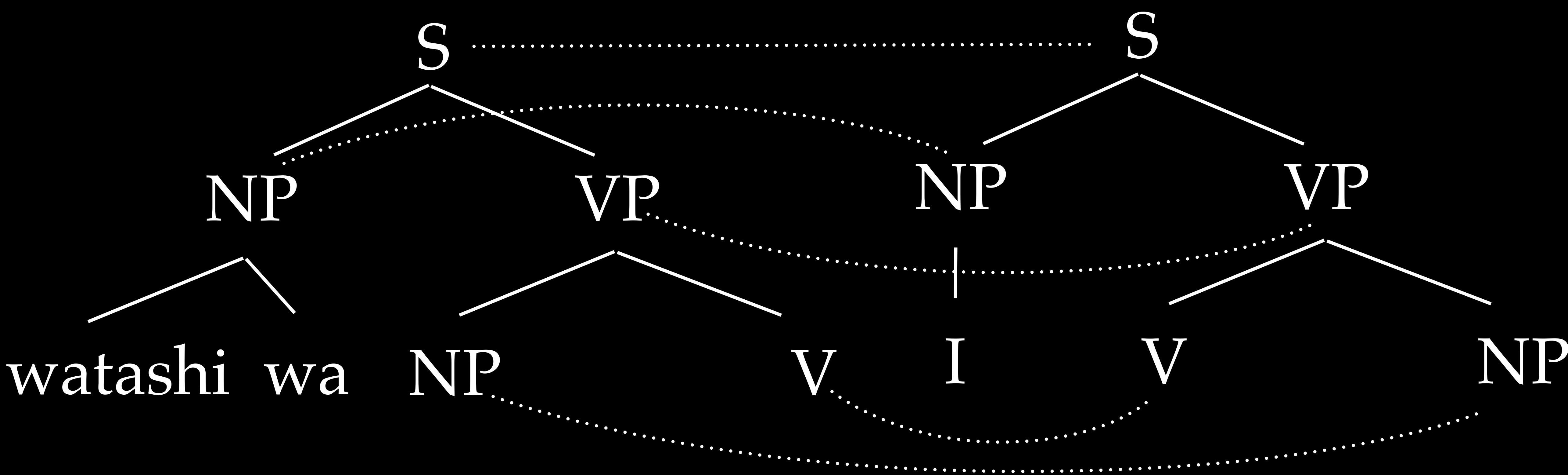
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

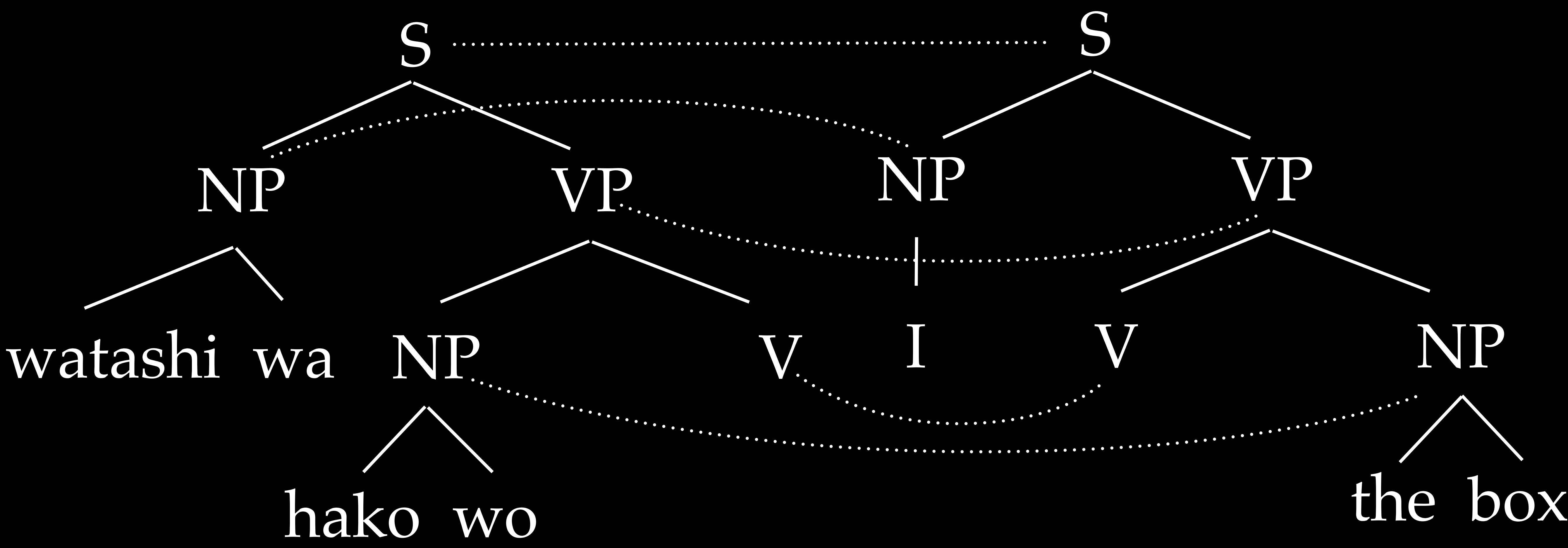
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

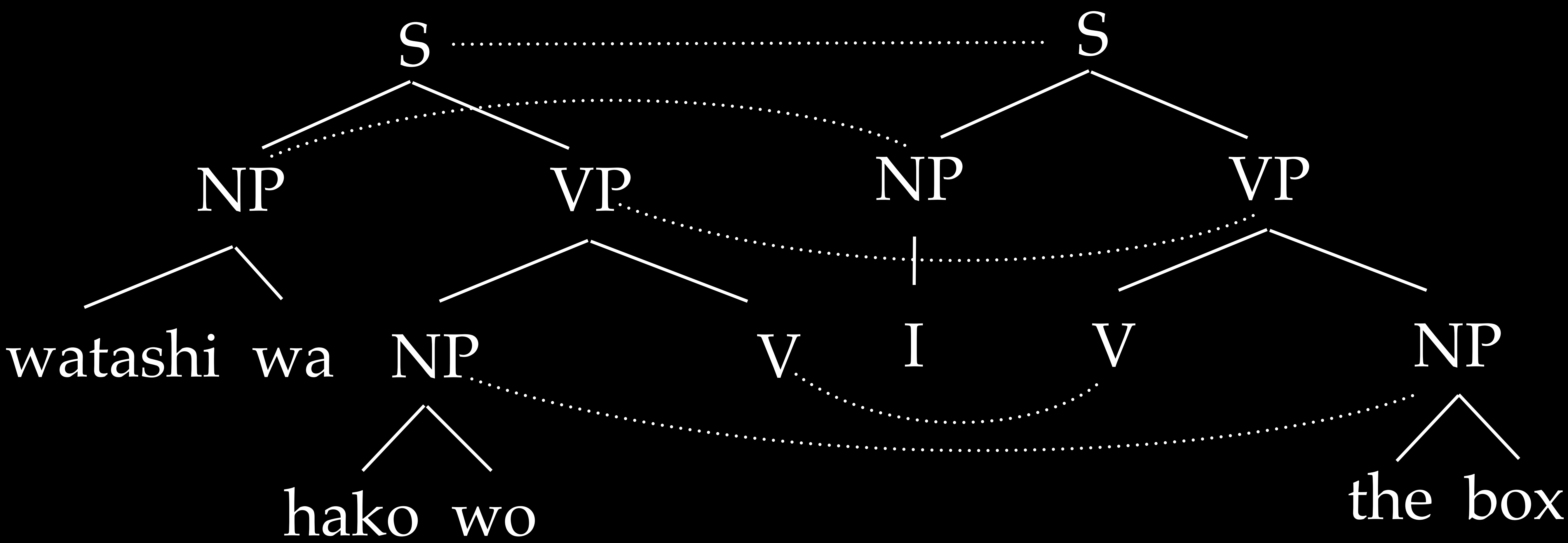
$NP \rightarrow$ watashi wa / I

NP → hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

V → akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

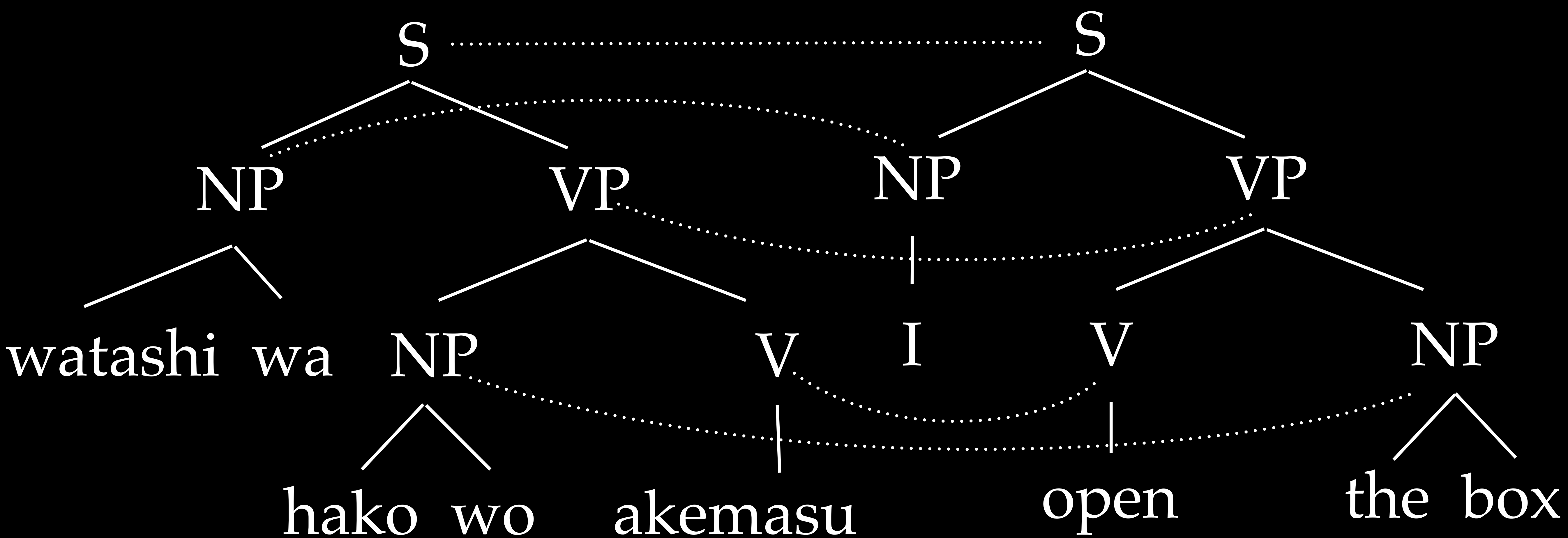
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar



$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

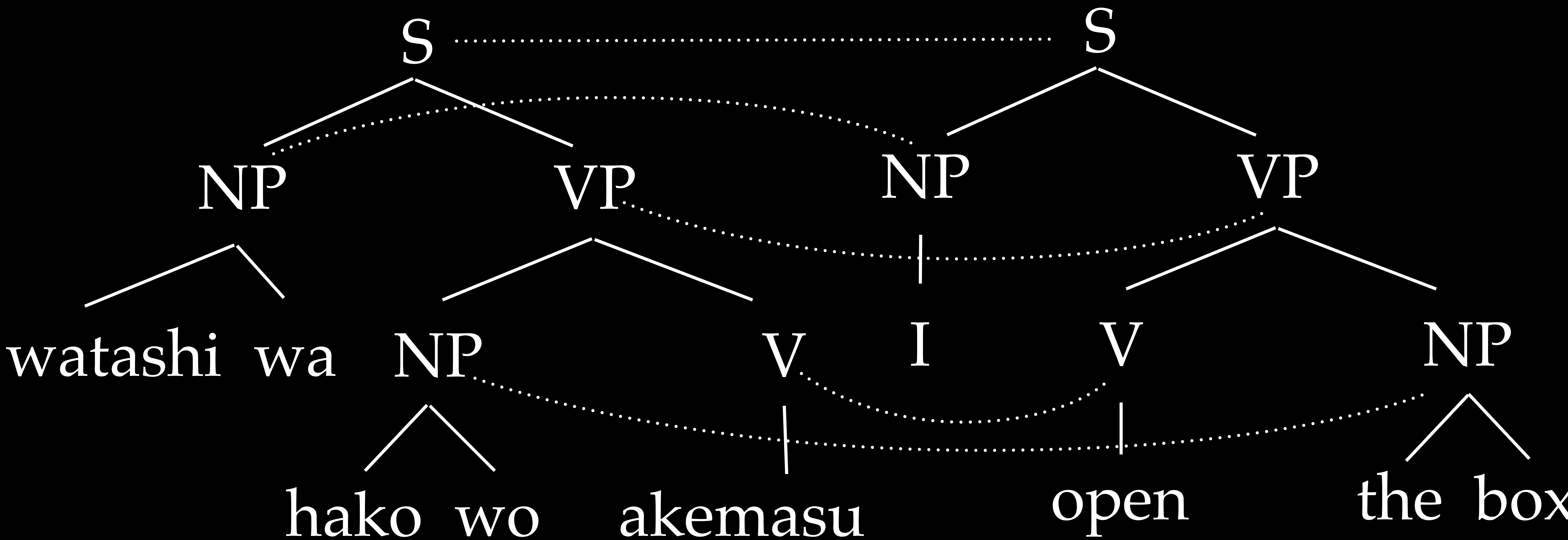
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

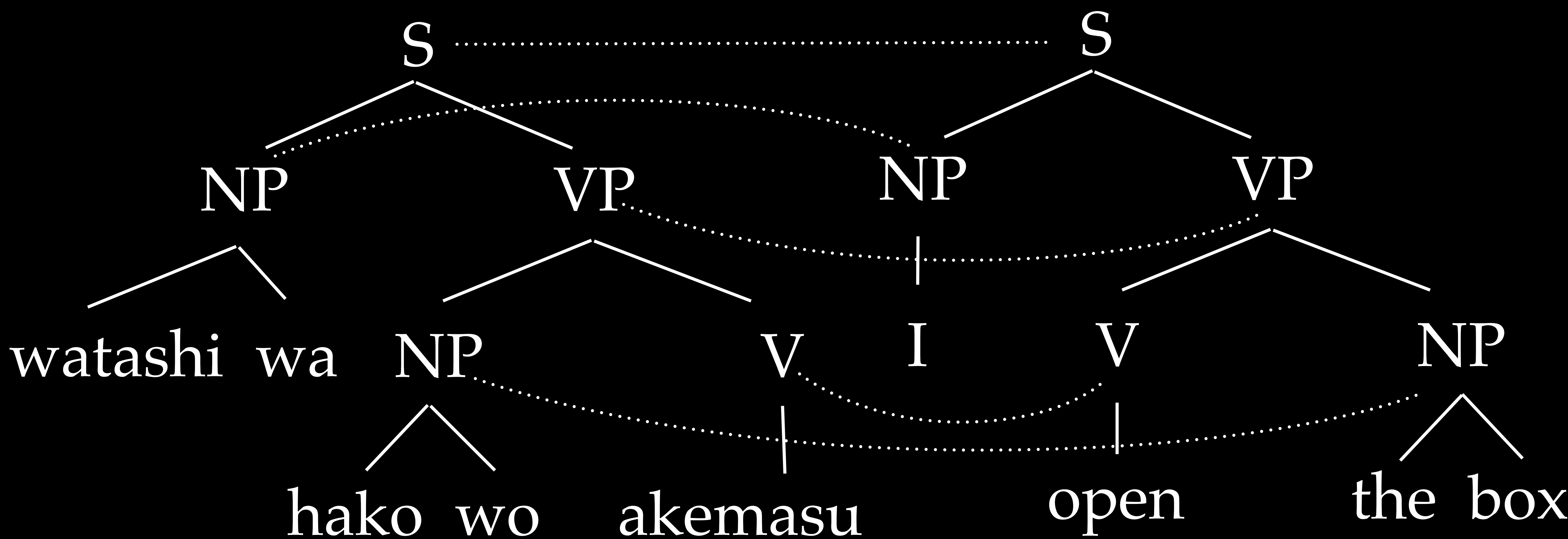
$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

$V \rightarrow$ akemasu / open

Synchronous Context-Free Grammar

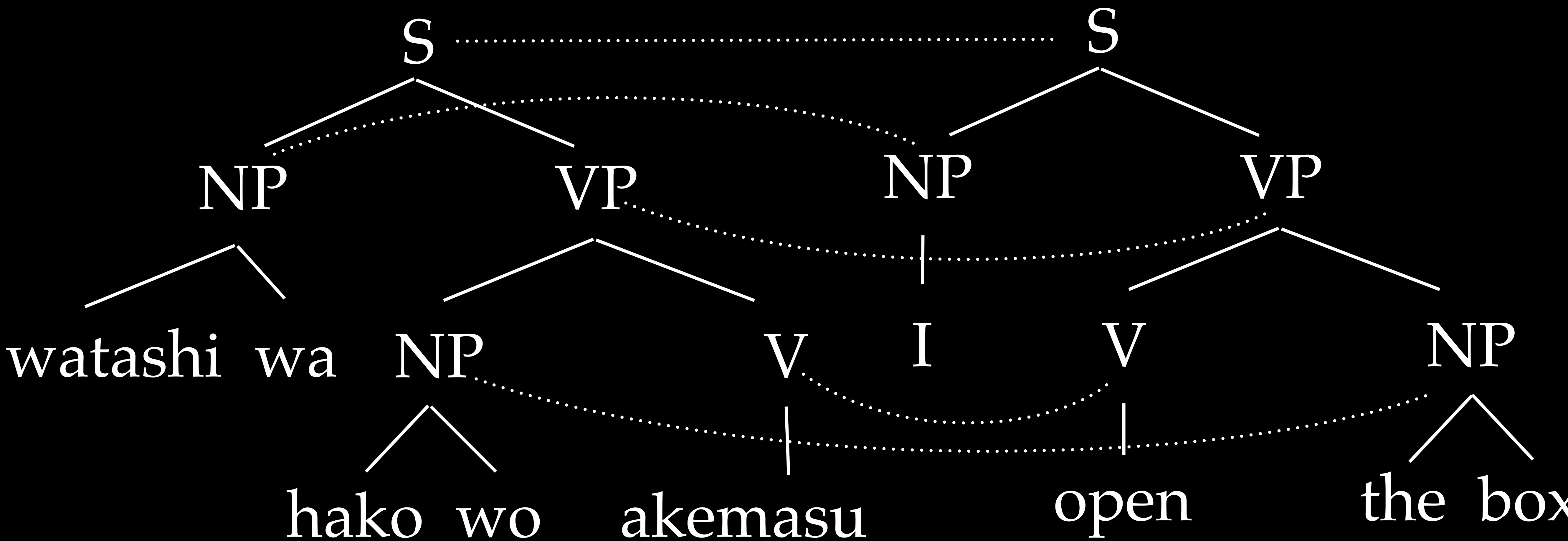


Synchronous Context-Free Grammar



watashi wa hako wo akemasu

Synchronous Context-Free Grammar

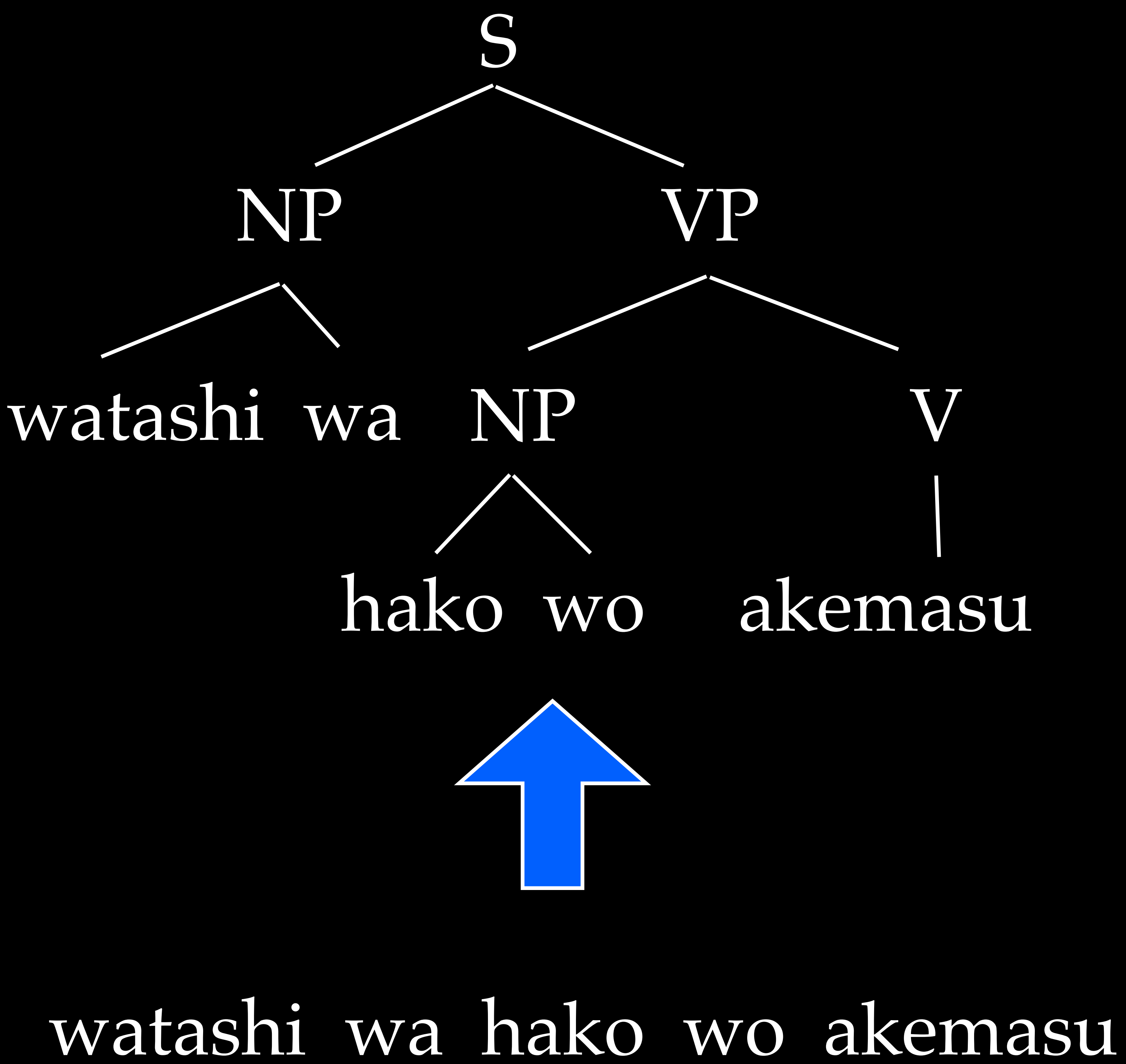


watashi wa hako wo akemasu I open the box

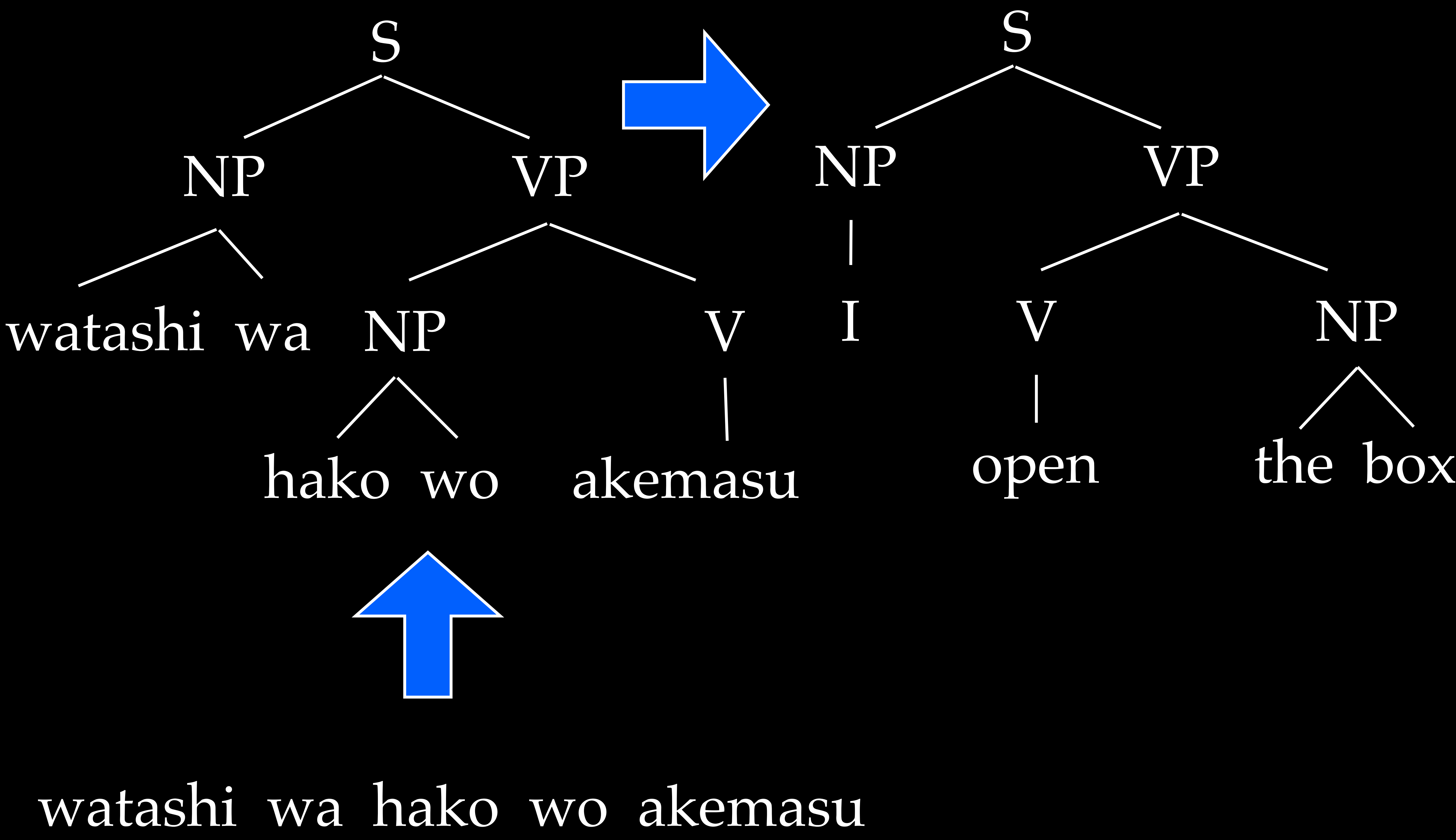
Translation as Parsing

watashi wa hako wo akemasu

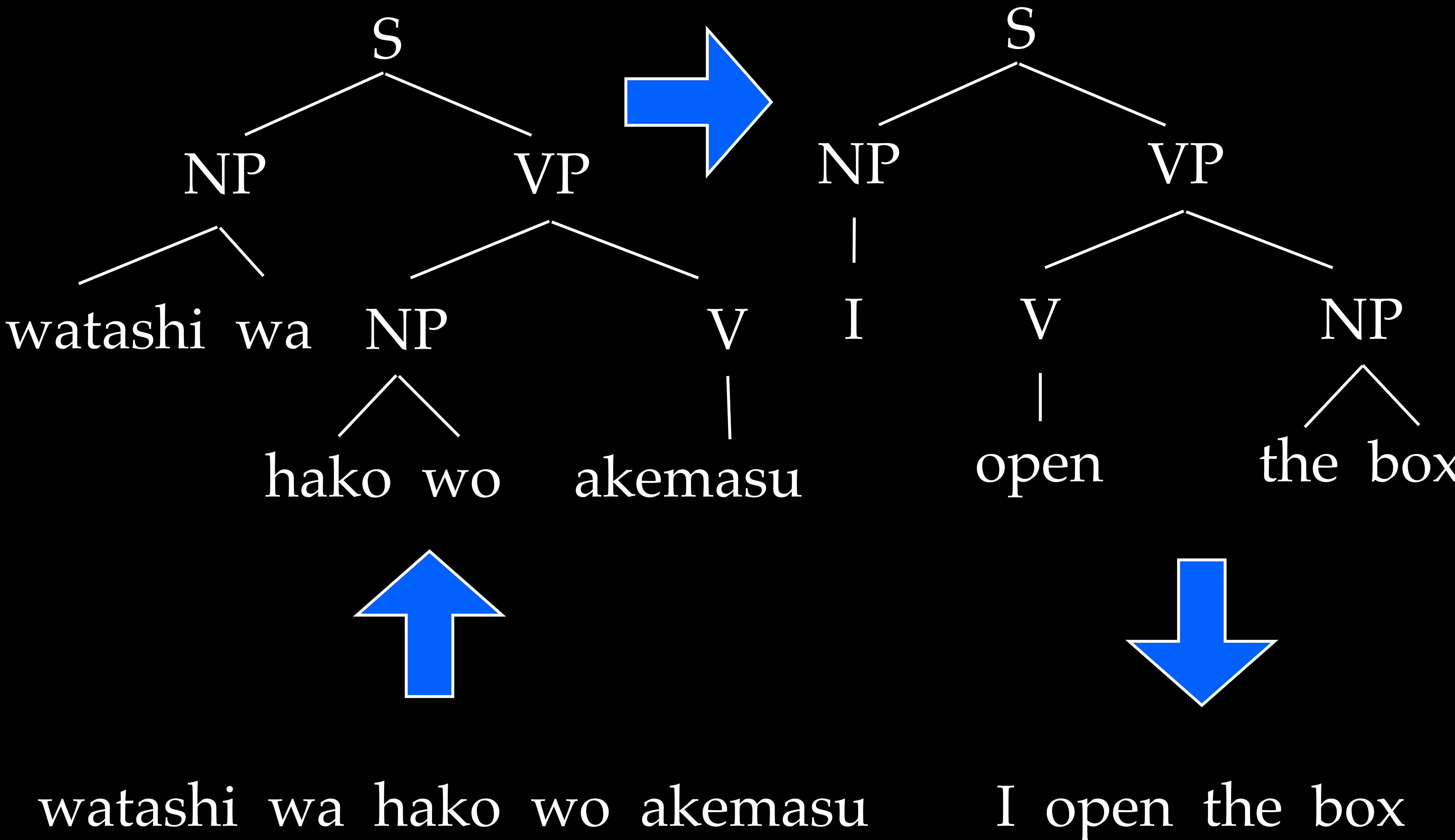
Translation as Parsing



Translation as Parsing



Translation as Parsing



Decoding

Decoding

- In general, there are an exponential number of possible parse trees for a sentence.

Decoding

- In general, there are an exponential number of possible parse trees for a sentence.
- Dynamic programming to the rescue!

Parsing

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

I_1 saw_2 her_3 $duck_4$

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

I_1 saw_2 her_3 duck_4

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

I_1 saw_2 her_3 $duck_4$

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

I_1 saw_2 her_3 $duck_4$

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$PRP_{0,1} \leftarrow (w_1 = \text{I}) \wedge (PRP \rightarrow \text{I})$$

I_1 saw_2 her_3 duck_4

Parsing

NN → duck
NP → PRP$ NN
PRP → her
PRP → I
PRP$ → her
S → PRP VP
SBAR → PRP VB
VB → duck
VP → VBD NP
VP → VBD SBAR
VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$PRP_{0,1} \leftarrow (w_1 = \mathrm{I}) \wedge (PRP \rightarrow \mathrm{I})$$

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$PRP_{0,1}$

↑

I_1 saw_2 her_3 duck_4

Parsing

- NN → duck
- NP → PRP$ NN
- PRP → her
- PRP → I
- PRP$ → her
- S → PRP VP
- SBAR → PRP VB
- VB → duck
- VP → VBD NP
- VP → VBD SBAR
- VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$VBD_{1,2}$

$PRP_{0,1}$

I_1 saw_2 her_3 duck_4

Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

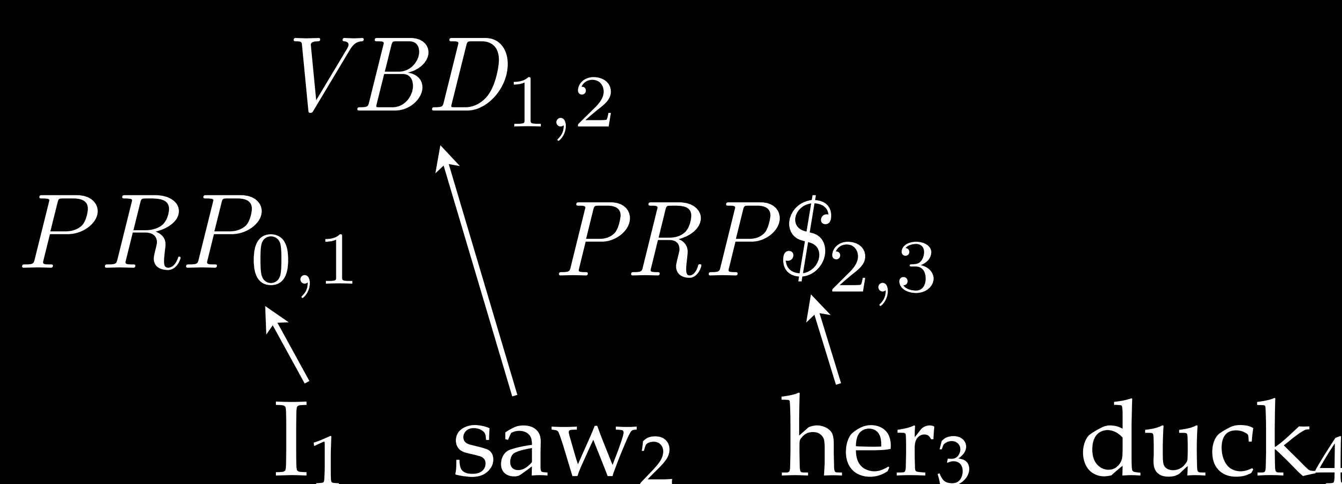
VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

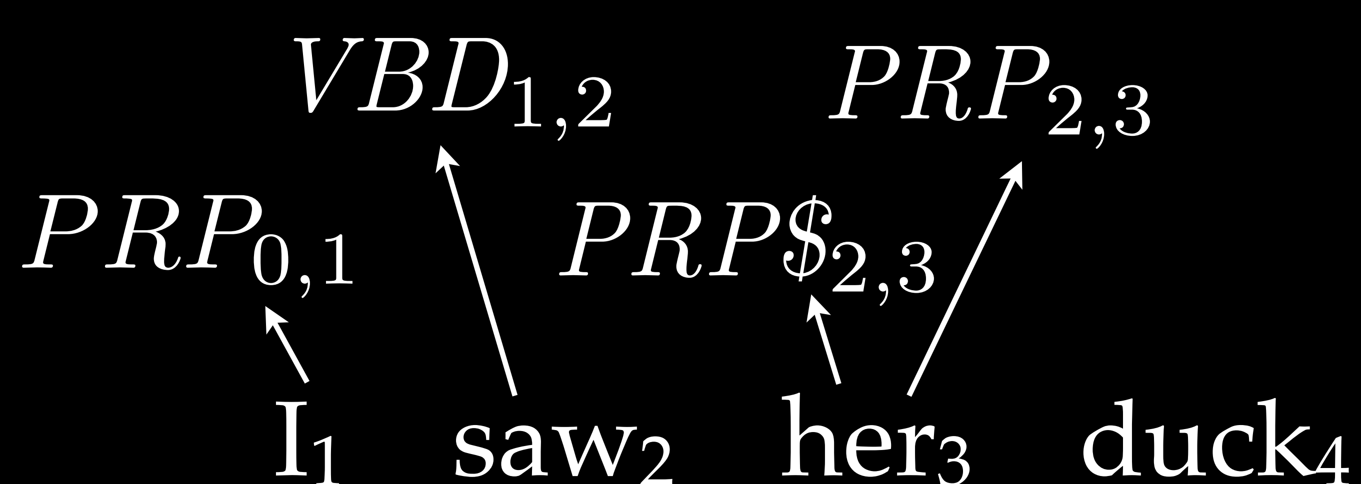
VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

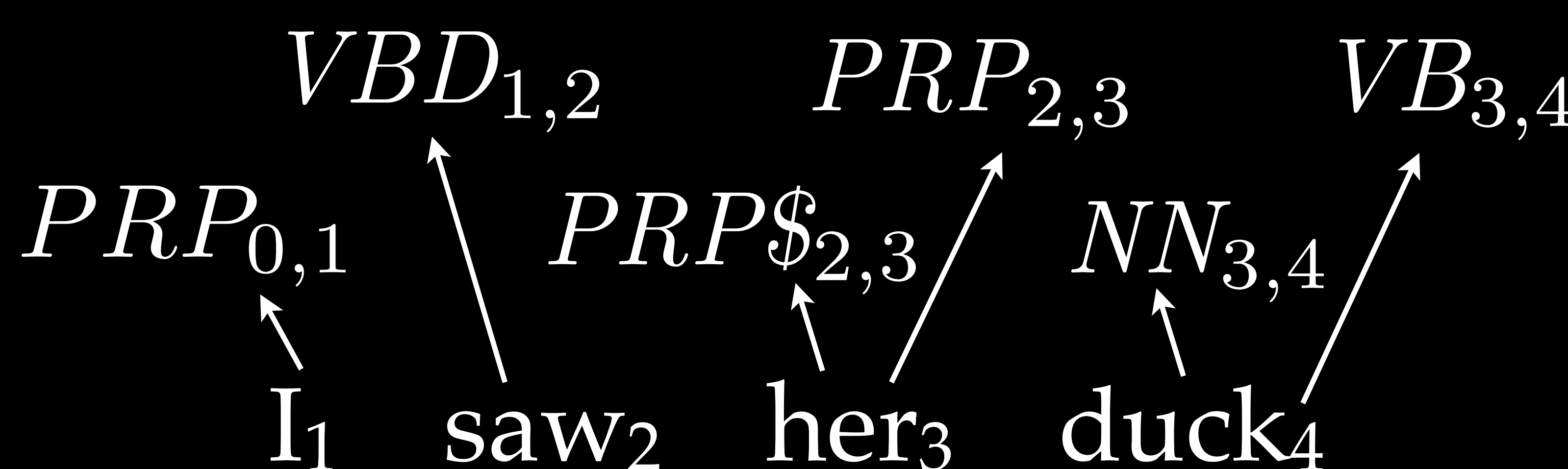
VB → duck

VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

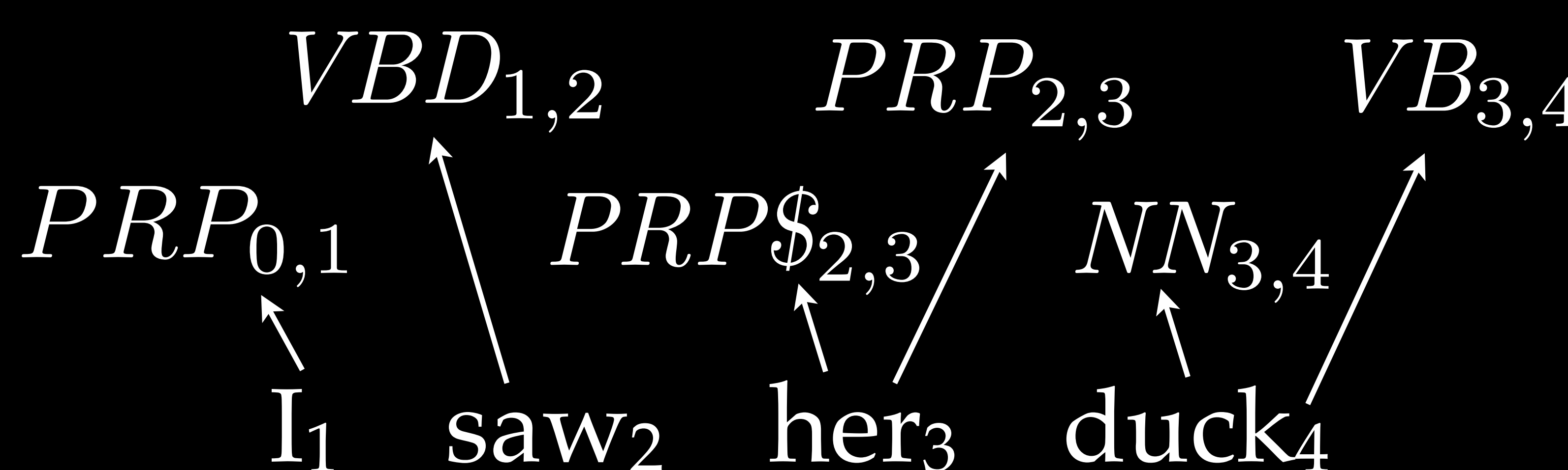
VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow YZ)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

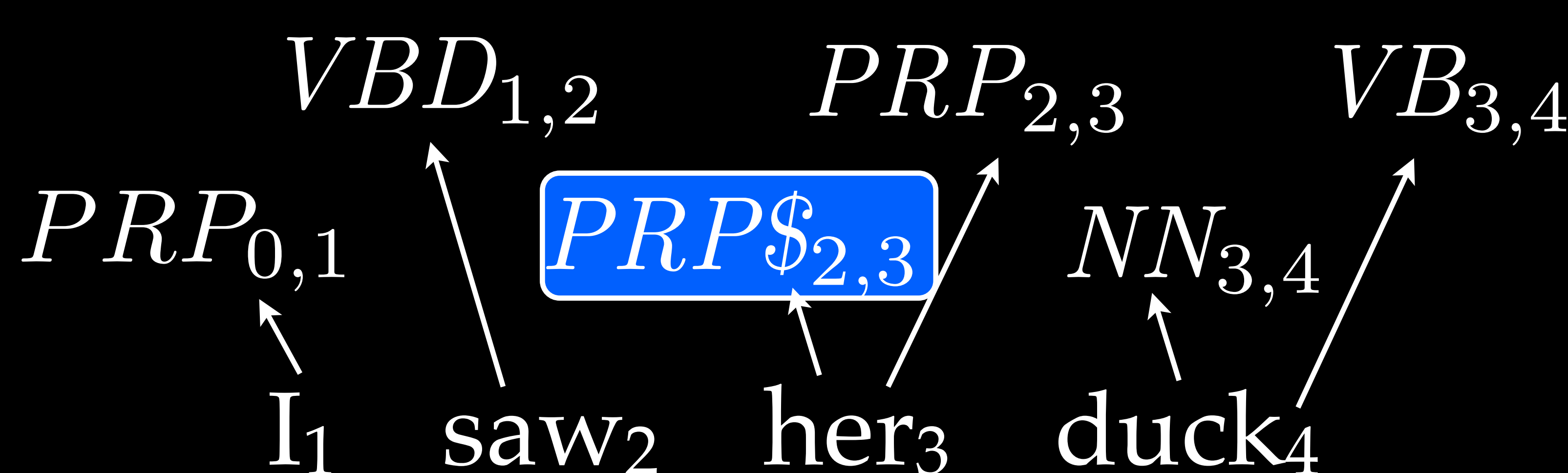
VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow YZ)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

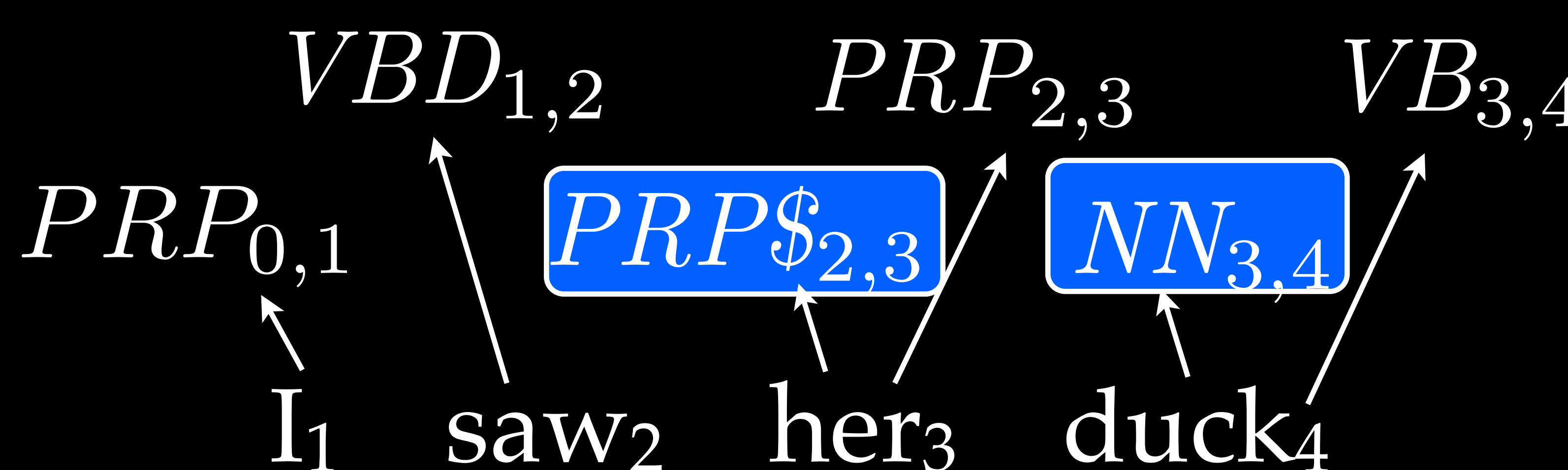
VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow Y Z)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

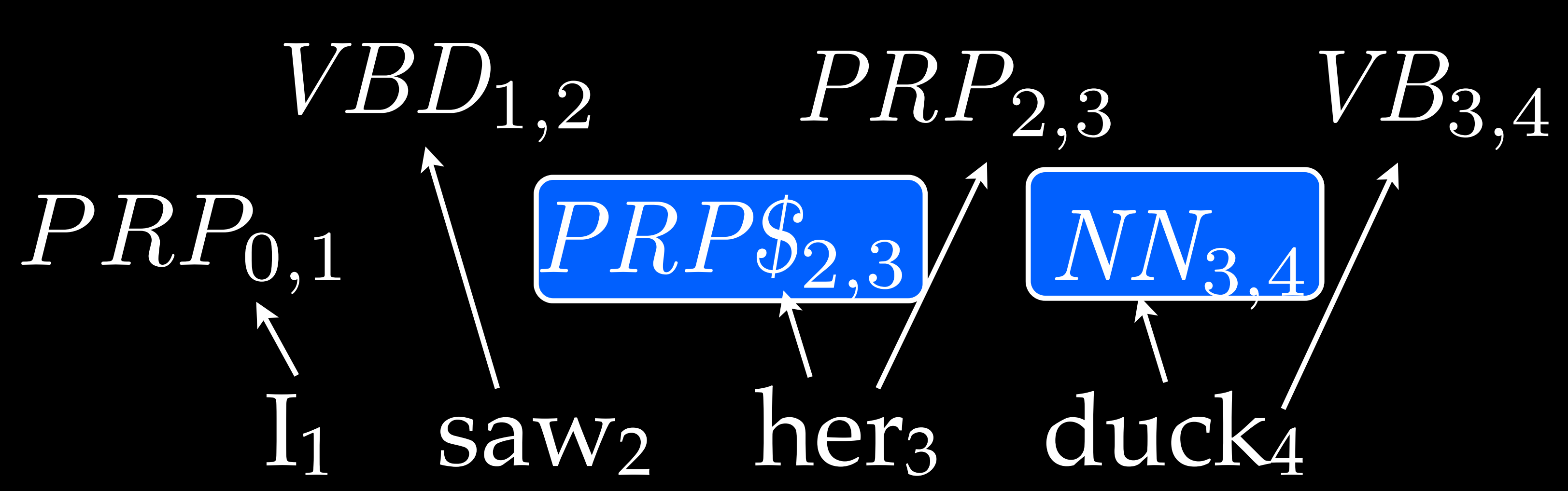
VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow Y Z)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

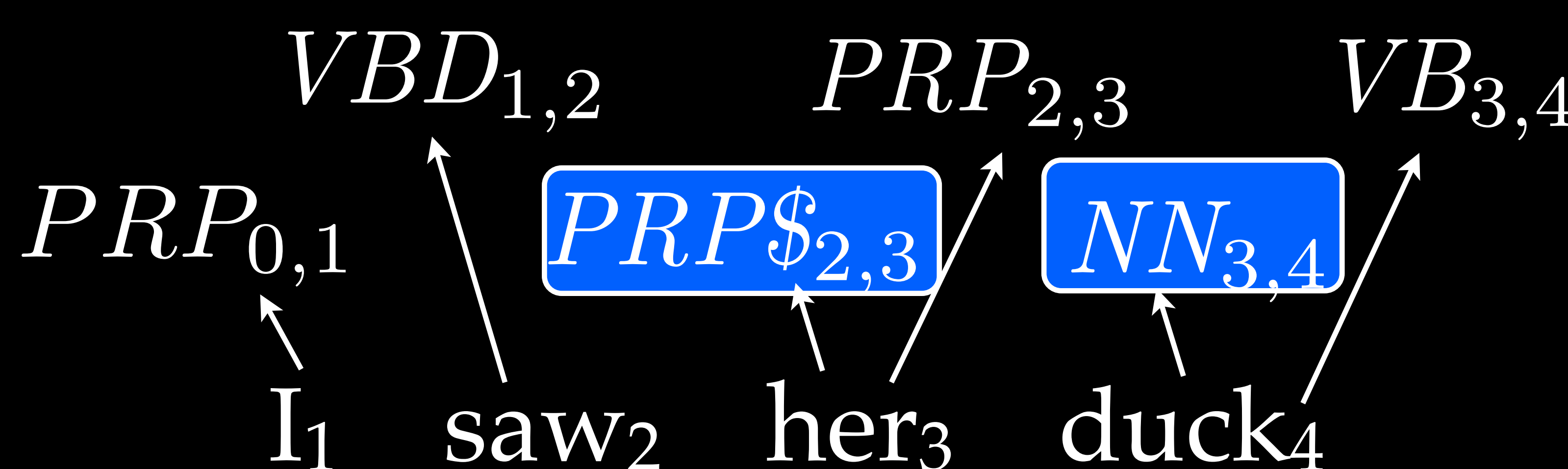
VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow Y Z)$$

$$NP_{2,4} \leftarrow PRP\$_{2,3} \wedge NN_{3,4} \wedge (NP \rightarrow PRP\$\ NN)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

VP → VBD NP

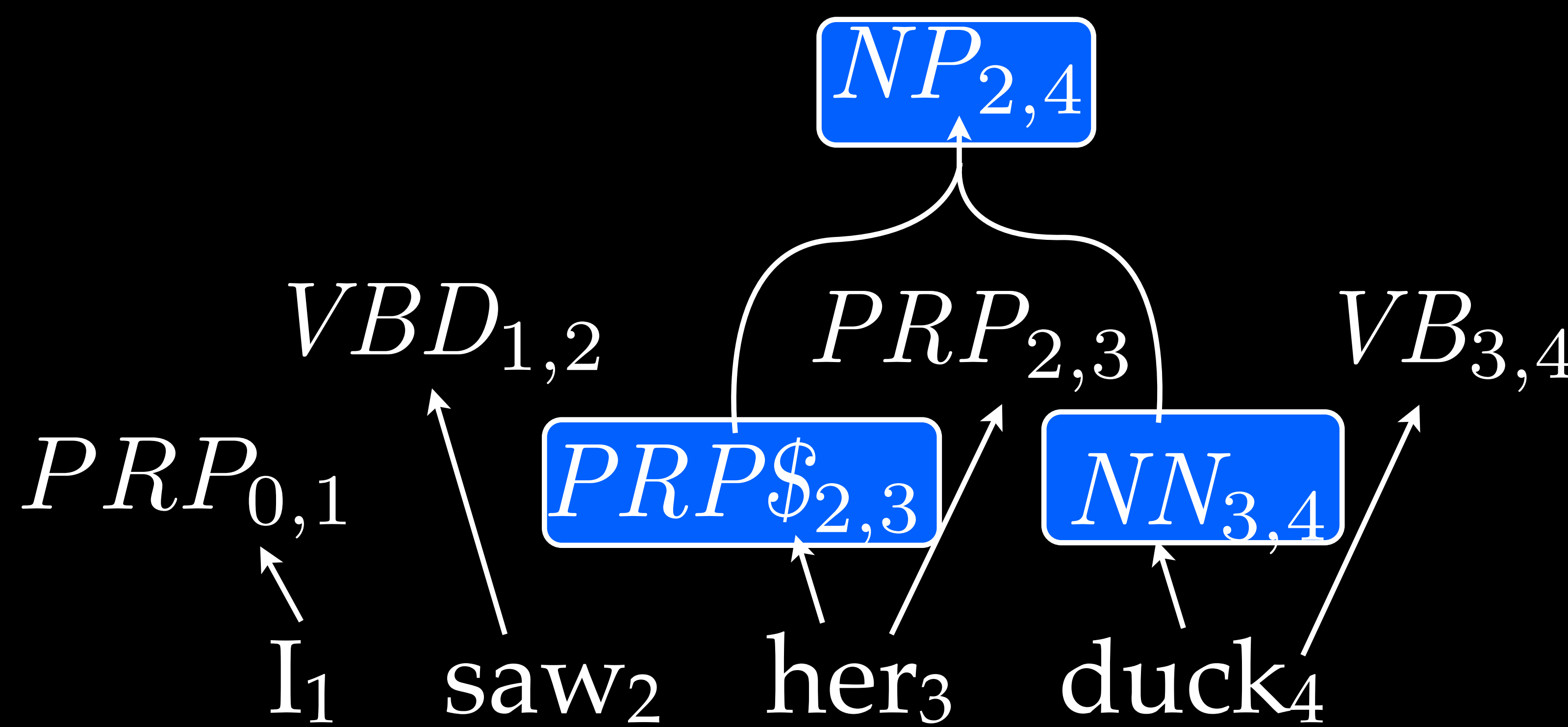
VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow Y Z)$$

$$NP_{2,4} \leftarrow PRP\$_{2,3} \wedge NN_{3,4} \wedge (NP \rightarrow PRP\$\ NN)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

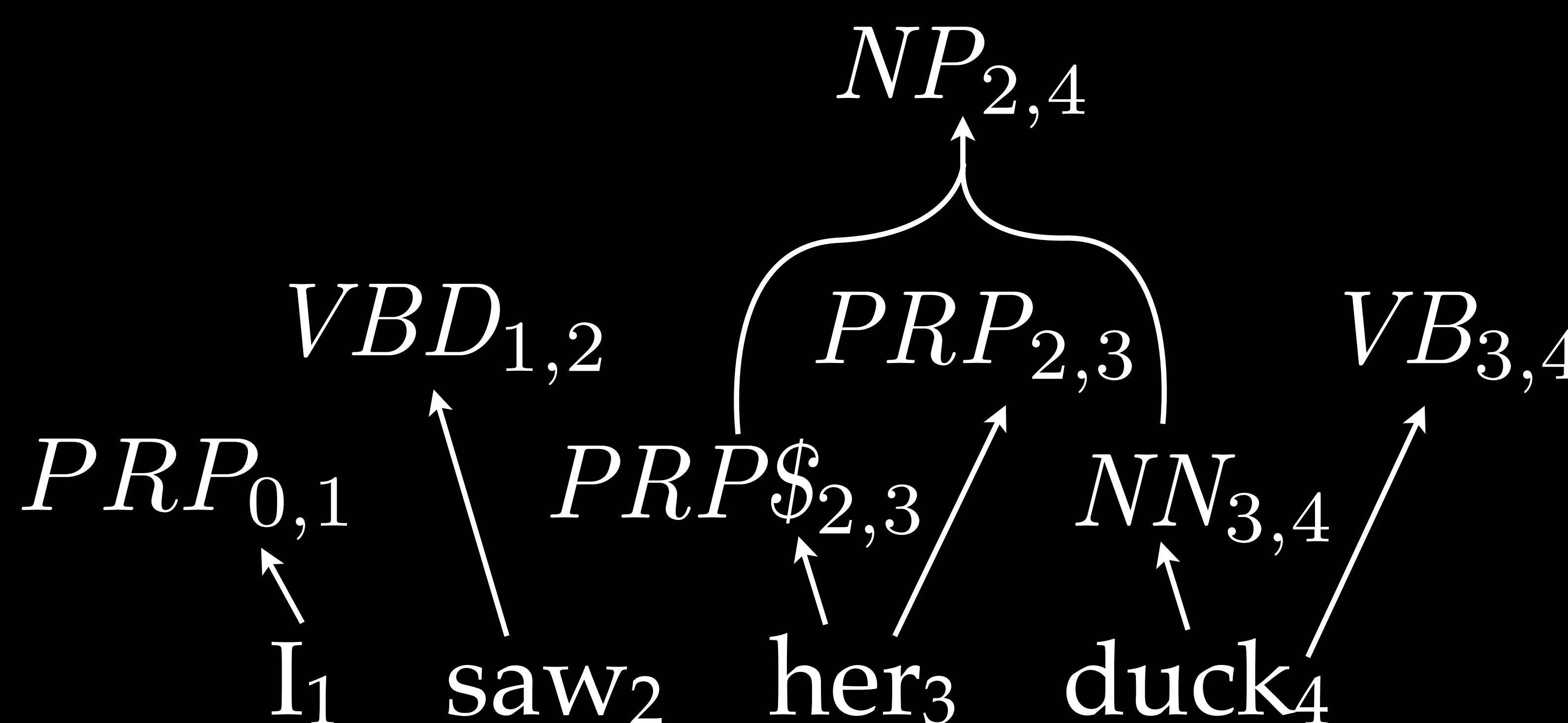
VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

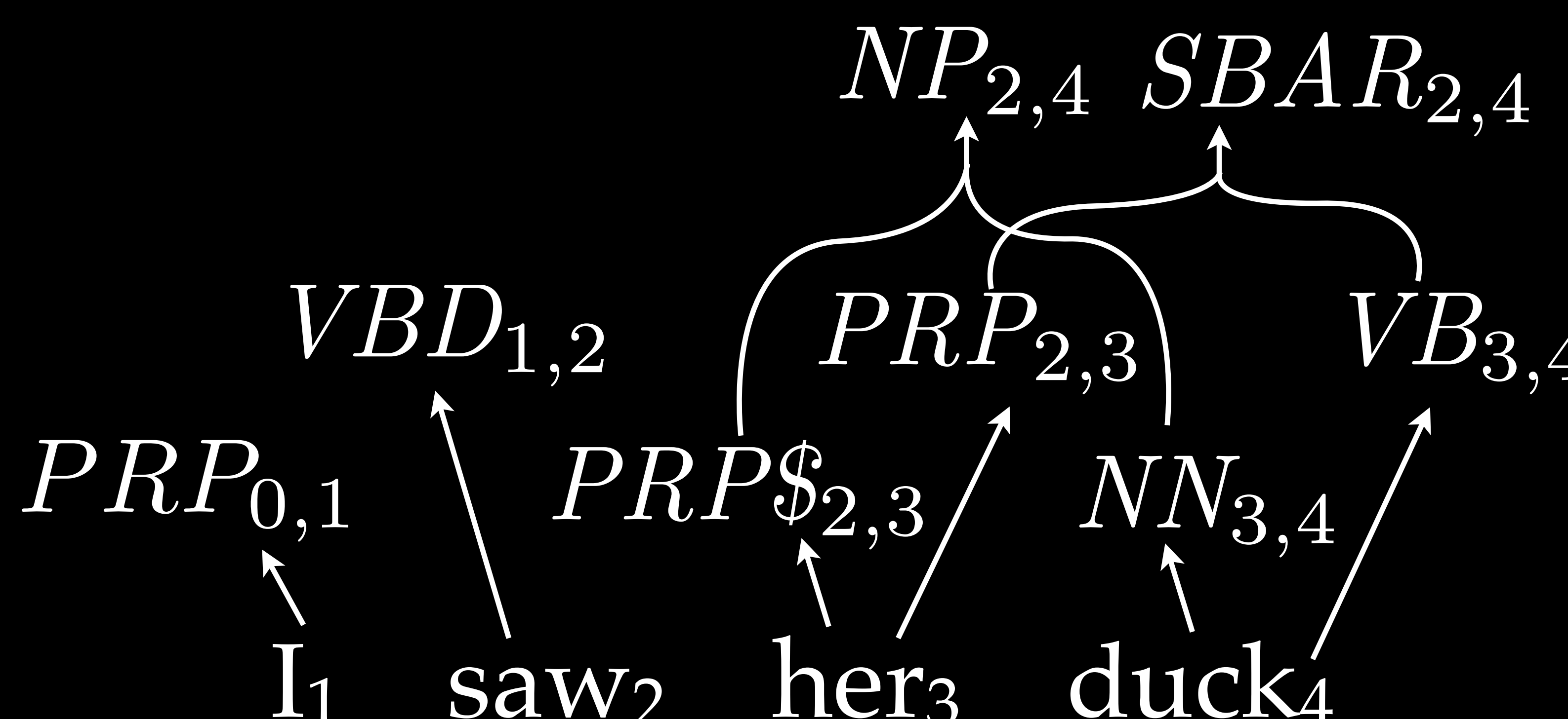
$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow YZ)$$



Parsing

NN → duck
NP → PRP$ NN
PRP → her
PRP → I
PRP$ → her
S → PRP VP
SBAR → PRP VB
VB → duck
VP → VBD NP
VP → VBD SBAR
VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$
$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow YZ)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

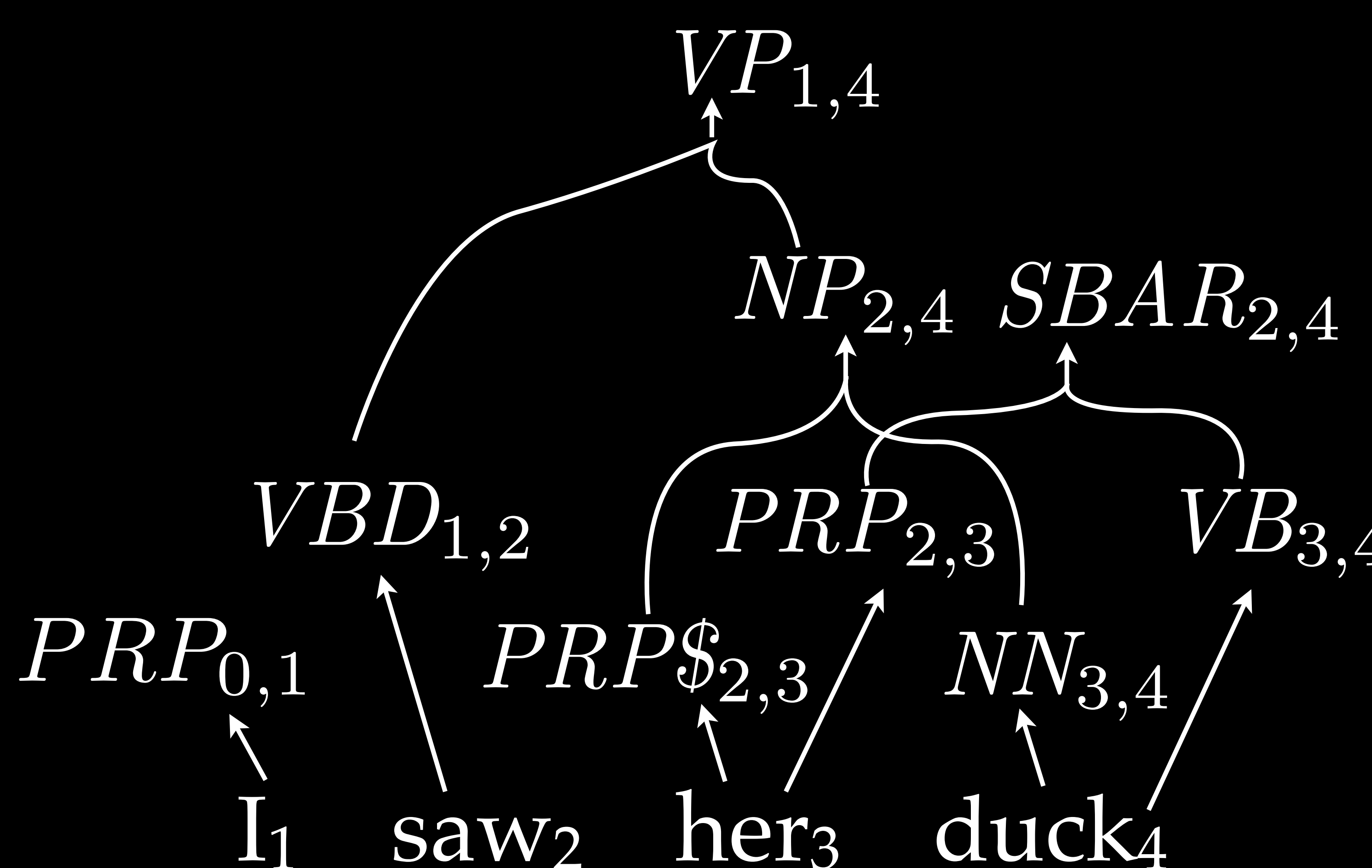
VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow YZ)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

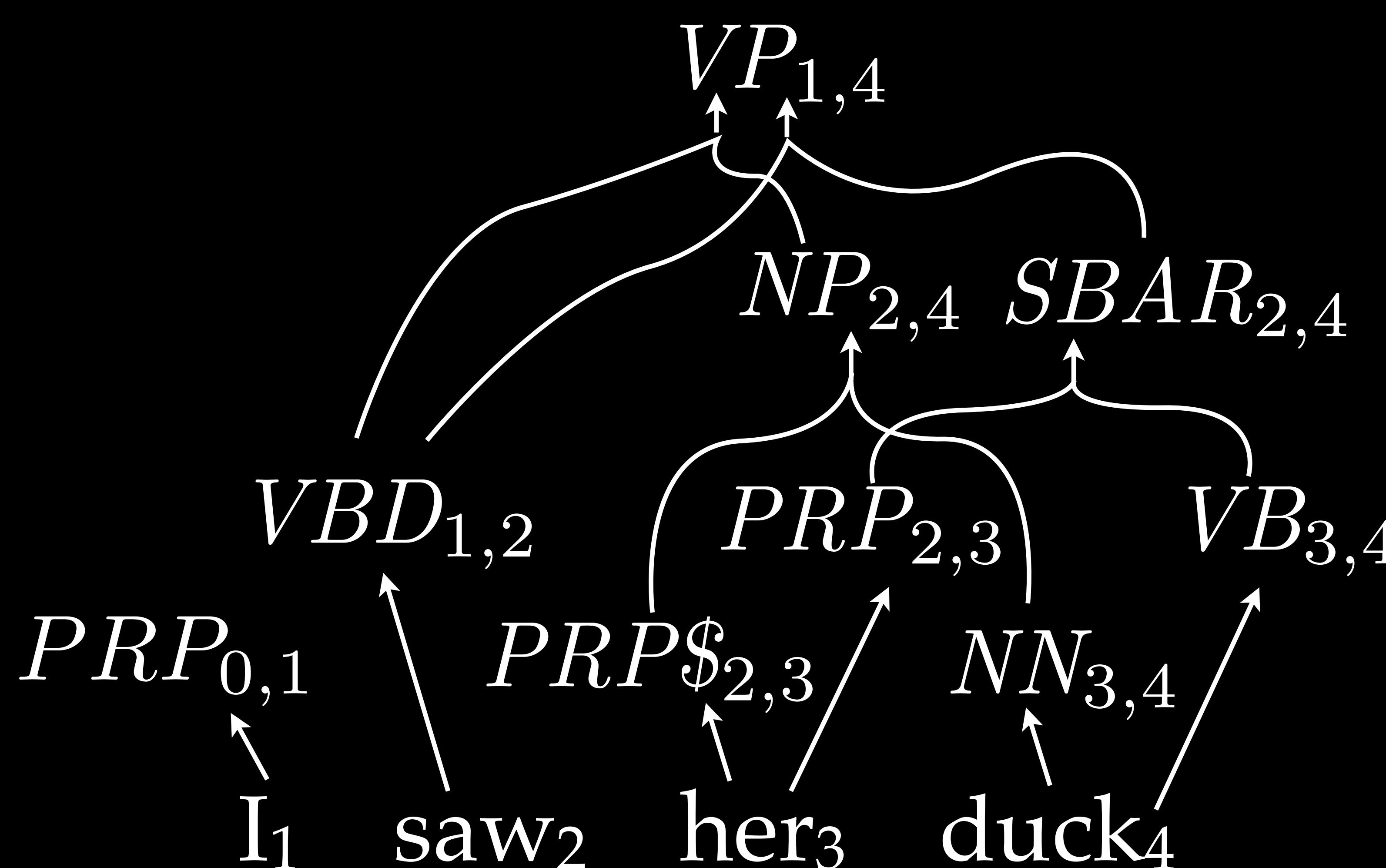
VP → VBD NP

VP → VBD SBAR

VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow YZ)$$



Parsing

NN → duck

NP → PRP$ NN

PRP → her

PRP → I

PRP$ → her

S → PRP VP

SBAR → PRP VB

VB → duck

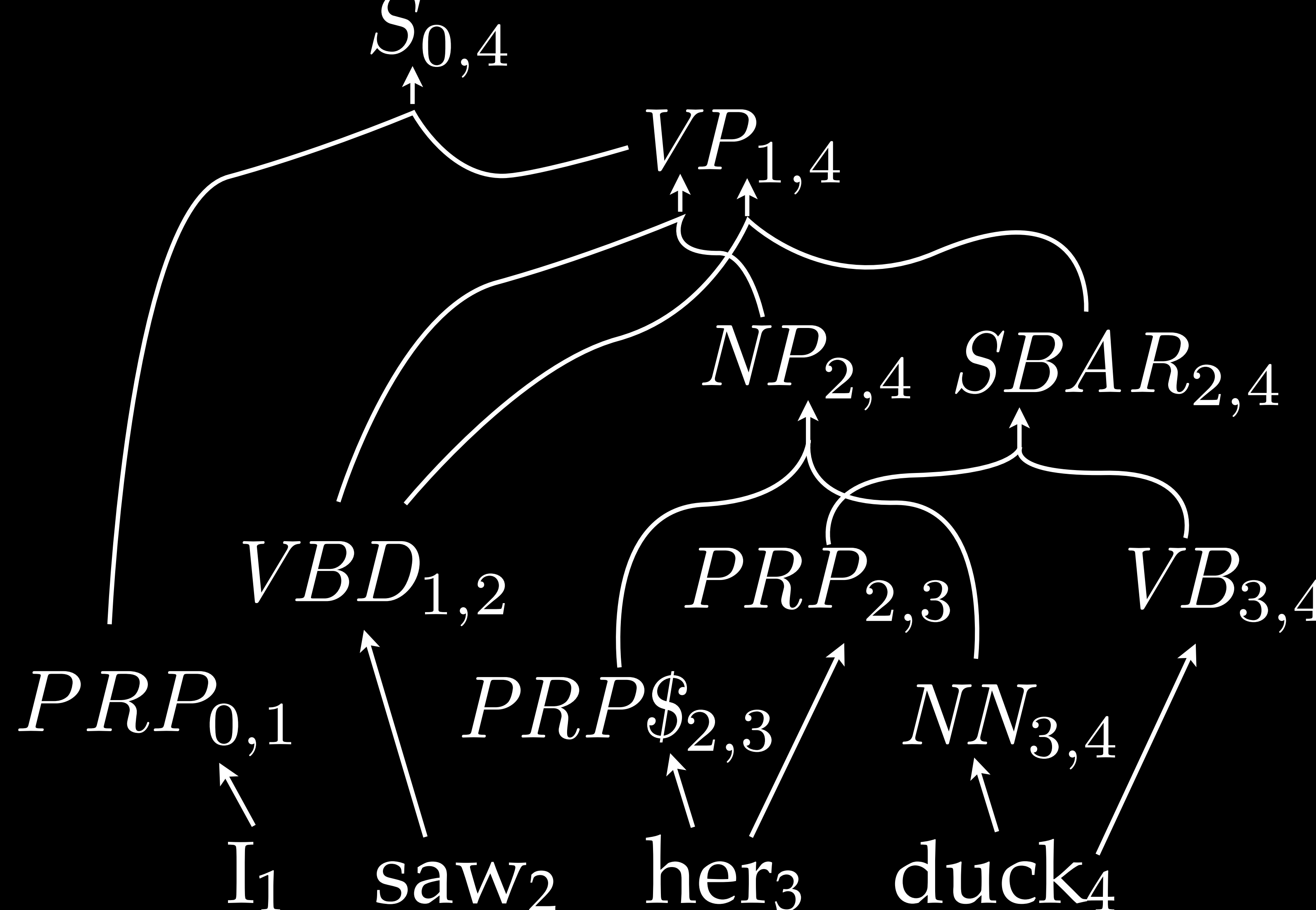
VP → VBD NP

VP → VBD SBAR

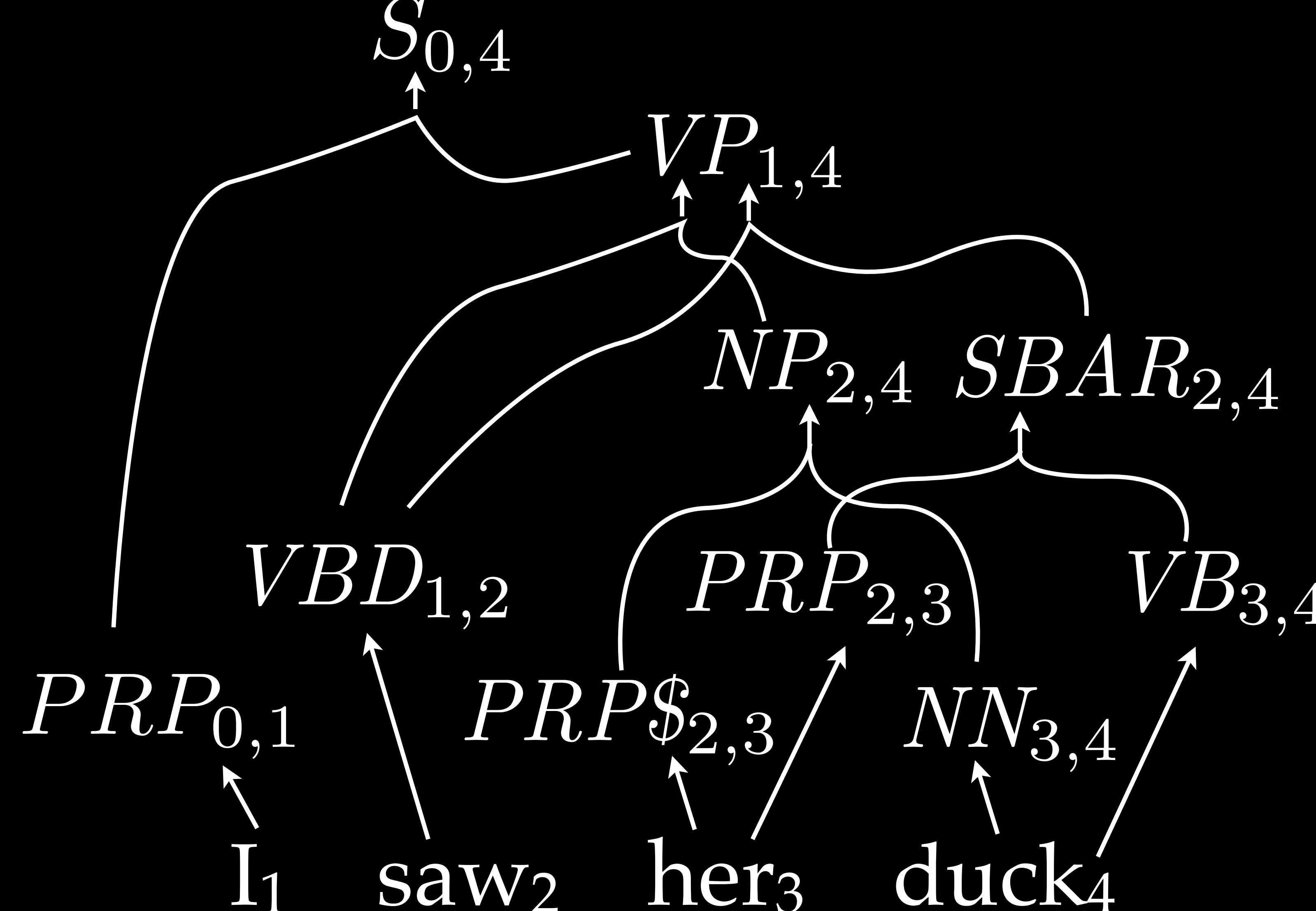
VBD → saw

$$X_{i,i+1} \leftarrow (w_{i+1} = w) \wedge (X \rightarrow w)$$

$$X_{i,j} \leftarrow Y_{i,k} \wedge Z_{k,j} \wedge (X \rightarrow YZ)$$

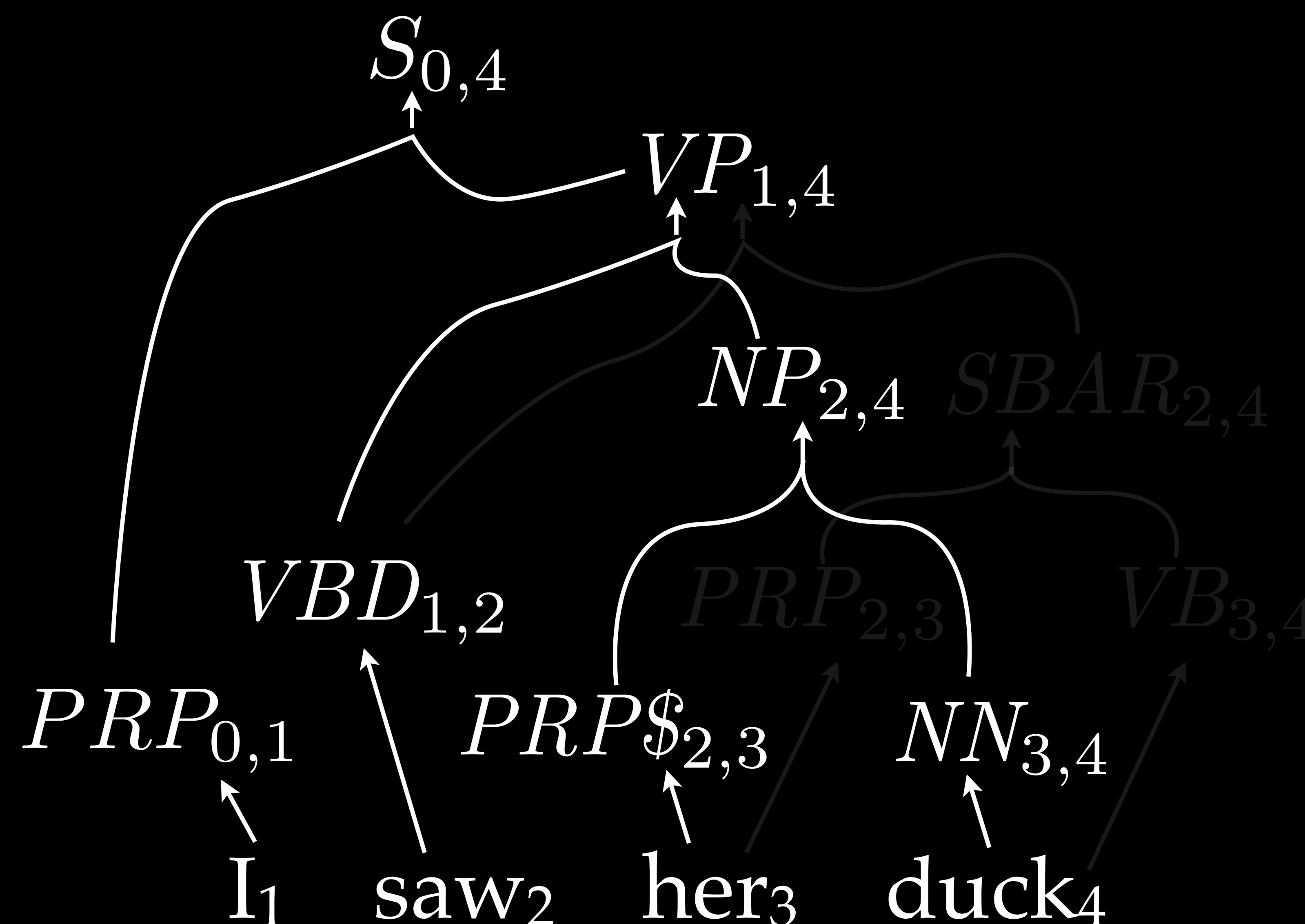


Parsing

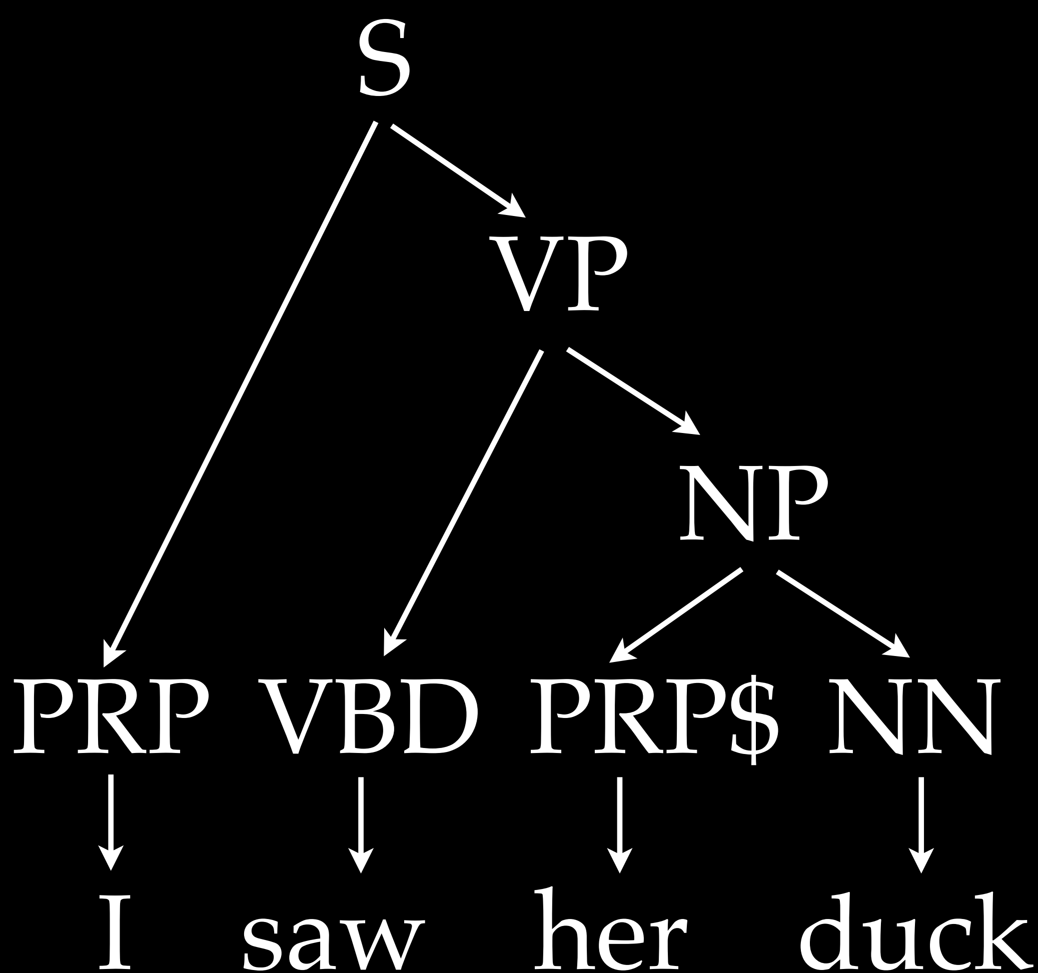


$S_{0,4}$
$VP_{1,4}$
$NP_{2,4}$ $SBAR_{2,4}$
$VBD_{1,2}$ $PRP_{2,3}$ $VB_{3,4}$
$PRP_{0,1}$ $PRP\$_{2,3}$ $NN_{3,4}$
I1 saw2 her3 duck4

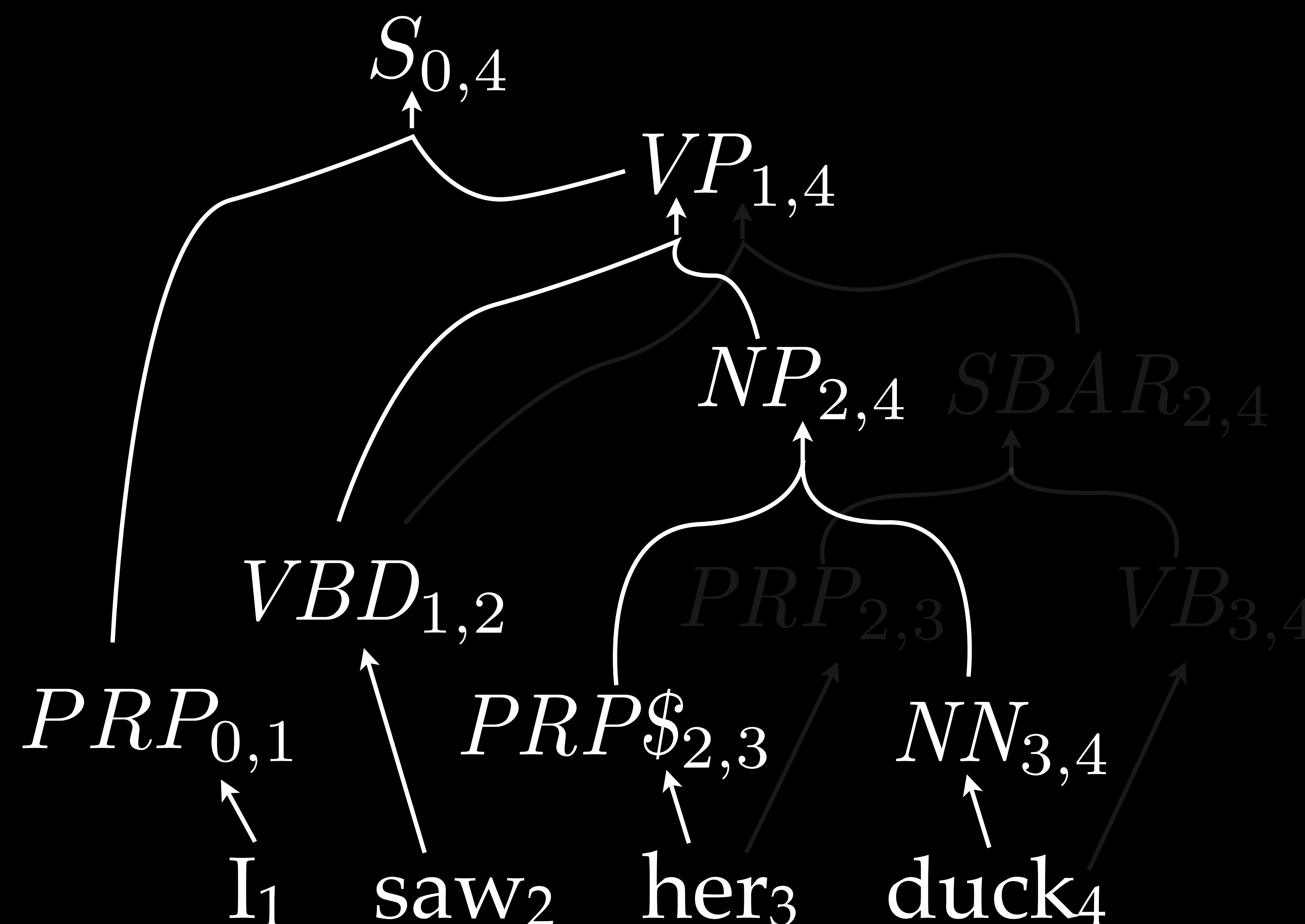
Parsing


$S_{0,4}$
$VP_{1,4}$
$NP_{2,4}$
$SBAR_{2,4}$
$VBD_{1,2}$
$PRP_{2,3}$
$VB_{3,4}$
$PRP_{0,1}$
$PRP\$_{2,3}$
$NN_{3,4}$
I1 saw2 her3 duck4

Parsing

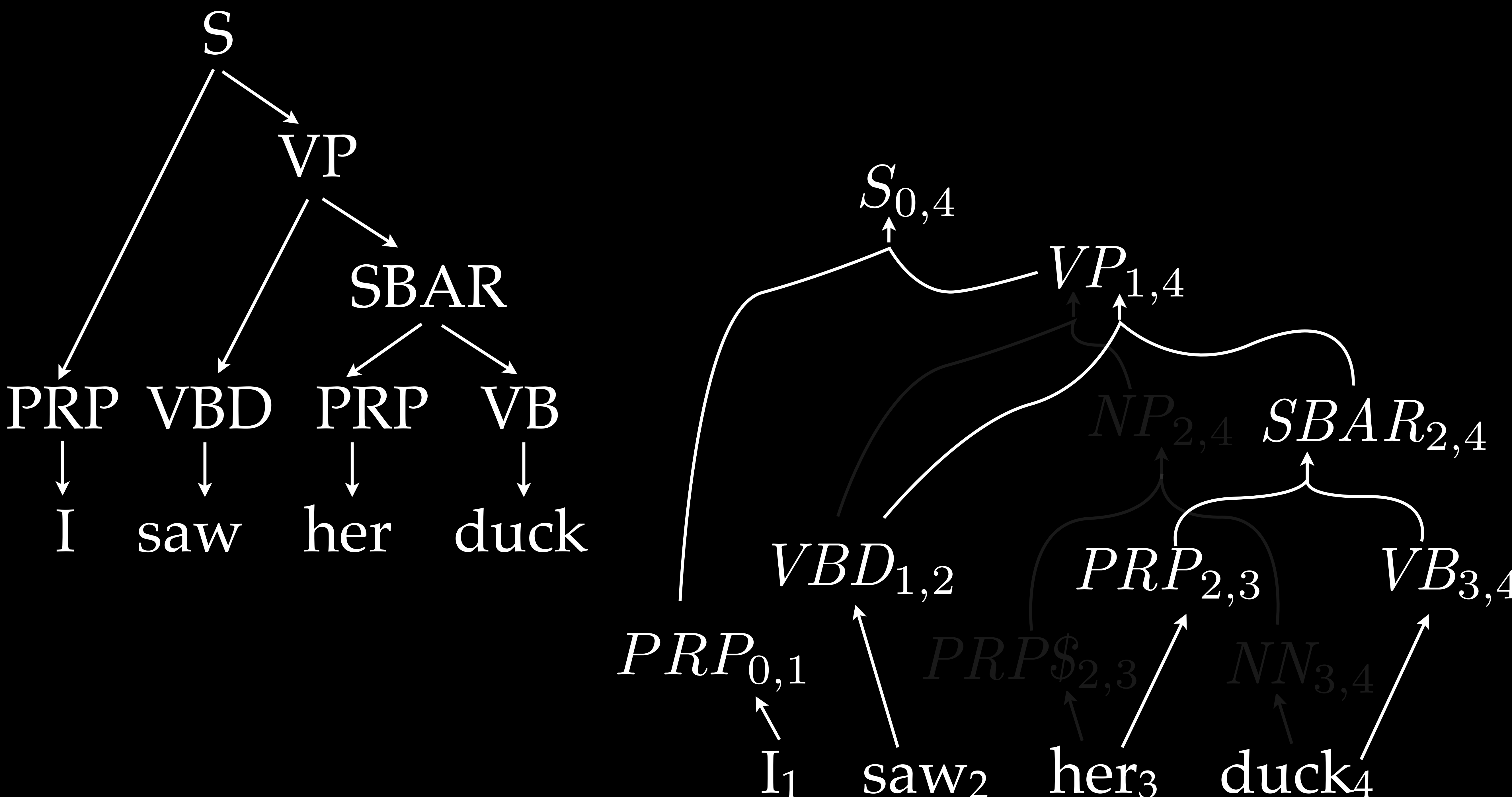


S
VP
NP
PRP VBD PRP$ NN
I saw her duck



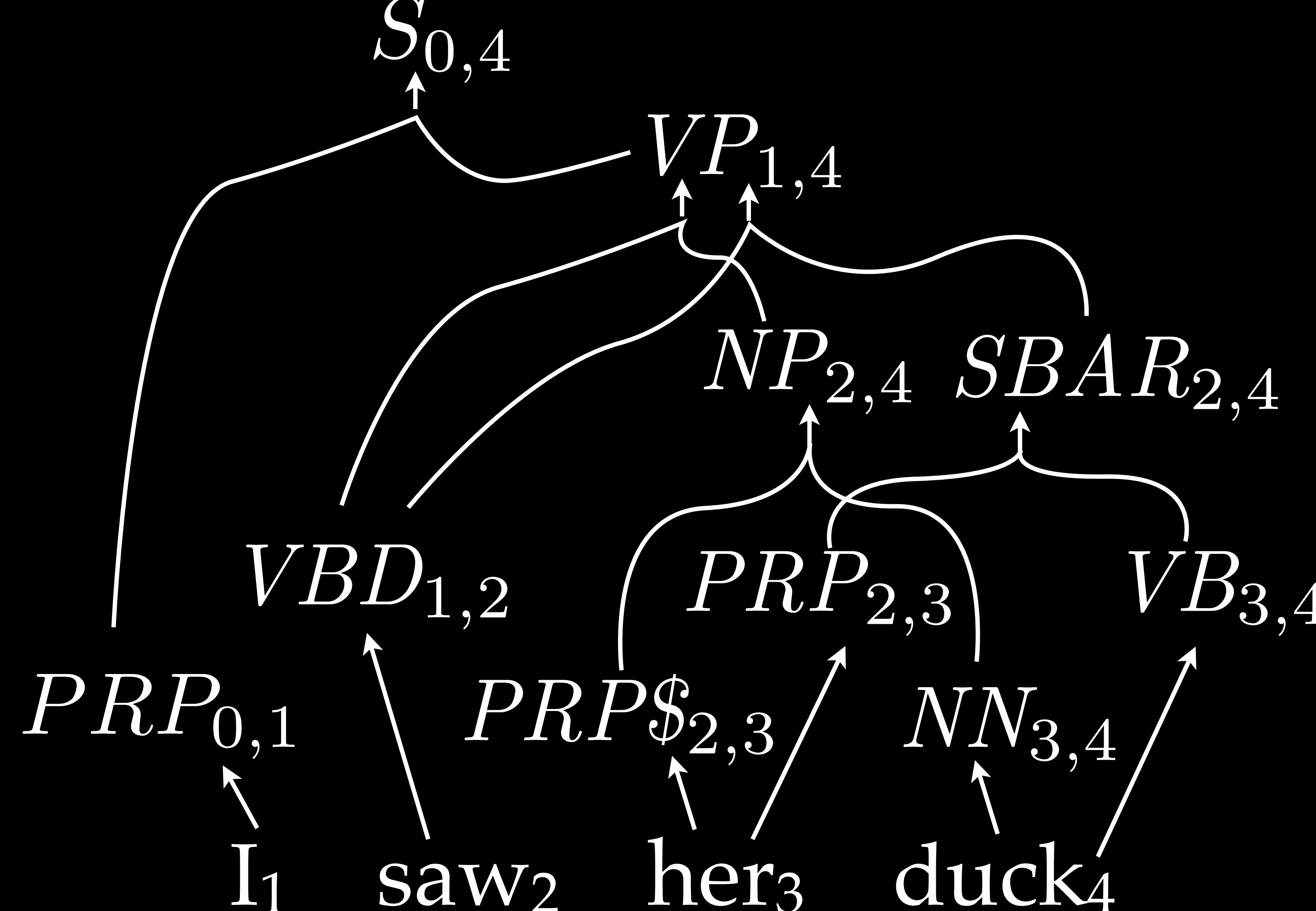
$S_{0,4}$
$VP_{1,4}$
$NP_{2,4}$ $SBAR_{2,4}$
$VBD_{1,2}$ $PRP_{2,3}$ $VB_{3,4}$
$PRP_{0,1}$ $PRP\$_{2,3}$ $NN_{3,4}$
I_1 saw_2 her_3 $duck_4$

Parsing


S
VP
SBAR
PRP VBD PRP VB
I saw her duck
$S_{0,4}$
$VP_{1,4}$
$NP_{2,4}$
$SBAR_{2,4}$
$VBD_{1,2}$
$PRP_{2,3}$
$VB_{3,4}$
$PRP_{0,1}$
$PRP\$_{2,3}$
$NN_{3,4}$
I_1 saw_2 her_3 $duck_4$

Parsing

Analysis

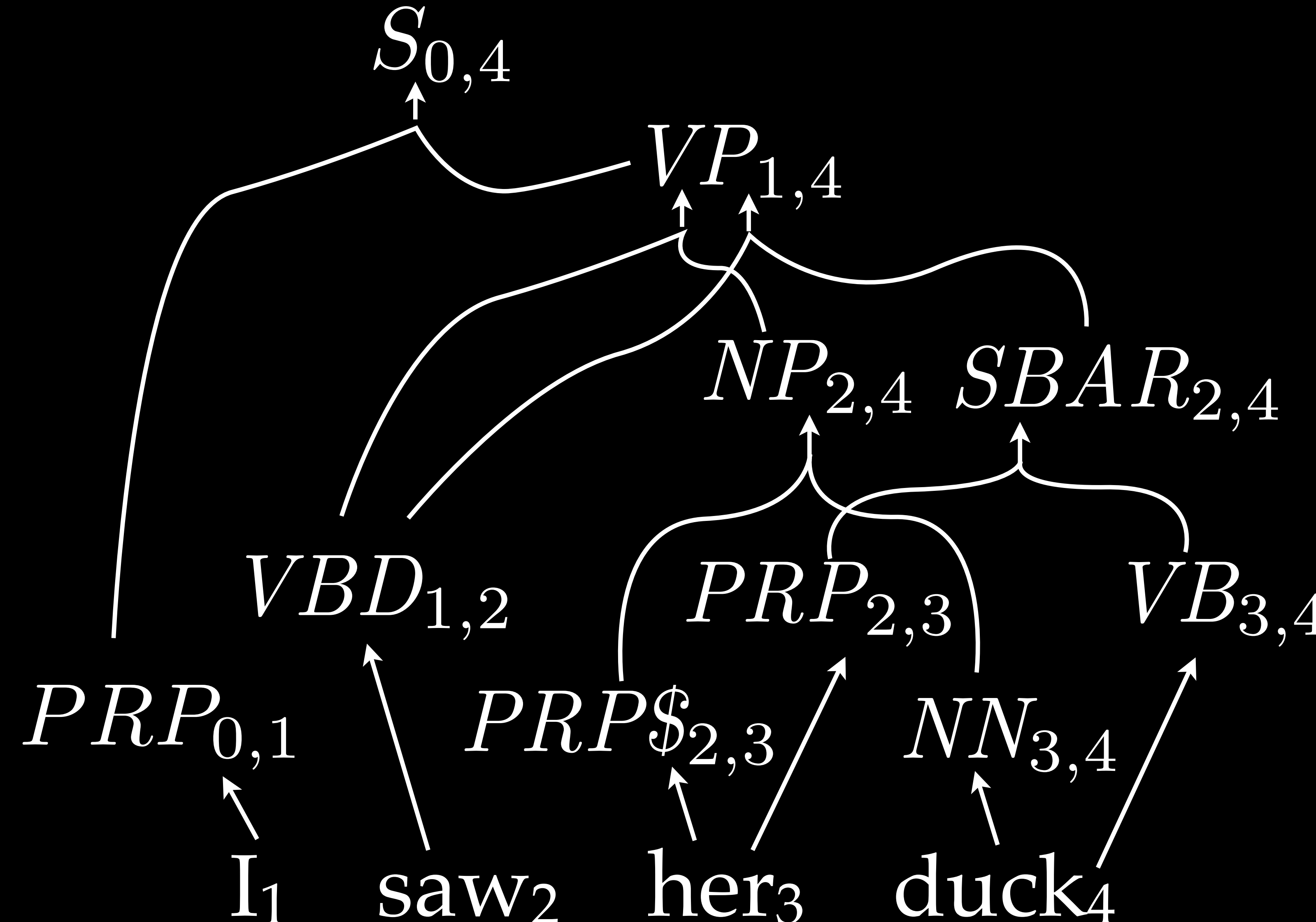


Parsing

Analysis

$O(Nn^2)$ nodes

$O(Gn^3)$ edges



Language Models Again

- Language models are finite-state (i.e. regular).
- Our translation model is context-free.
- We can again compute full model via ***intersection***.
- Result is also context-free.
- Bad news for context-free language models and context-free translation models...
  - Context-free languages not closed under intersection.
  - Computation is in PSPACE!

Language Models Again

- Basic DP strategy: nodes include category, span, and left and right language model context.
- While polynomial, this still tends to be too slow to do exactly.
- Various forms of pruning are generally used.
- Finding efficient algorithms is currently an area of very active research.

The Big Question

The Big Question

Where do the categories come from?

The Big Question

Where do the categories come from?

Answer #1: there are no categories!

The Big Question

Where do the categories come from?

Answer #1: there are no categories!

$X \rightarrow X_1\ X_2\ /\ X_1\ X_2$

$X \rightarrow X_1\ X_2\ /\ X_2\ X_1$

X → watashi wa / I

X → hako wo / the box

X → akemasu / open

The Big Question

Where do the categories come from?

Answer #1: there are no categories!

$X \rightarrow X_1\ X_2\ /\ X_1\ X_2$ ← Keep order

$X \rightarrow X_1\ X_2\ /\ X_2\ X_1$

$X \rightarrow$ watashi wa / I

$X \rightarrow$ hako wo / the box

$X \rightarrow$ akemasu / open

The Big Question

Where do the categories come from?

Answer #1: there are no categories!

$X \rightarrow X_1\ X_2\ /\ X_1\ X_2$ ← Keep order

$X \rightarrow X_1\ X_2\ /\ X_2\ X_1$ Swap order

X → watashi wa / I

X → hako wo / the box

X → akemasu / open

The Big Question

Where do the categories come from?

Answer #1: there are no categories!

$X \rightarrow X_1\ X_2\ /\ X_1\ X_2$ ← Keep order

$X \rightarrow X_1\ X_2\ /\ X_2\ X_1$ ← Swap order

$X \rightarrow$ watashi wa / I

$X \rightarrow$ hako wo / the box

$X \rightarrow$ akemasu / open

The Big Question

Where do the categories come from?

Answer #1: there are no categories!

$X \rightarrow X_1\ X_2\ /\ X_1\ X_2$ ← Keep order

$X \rightarrow X_1\ X_2\ /\ X_2\ X_1$ ← Swap order

$X \rightarrow$ watashi wa / I

$X \rightarrow$ hako wo / the box

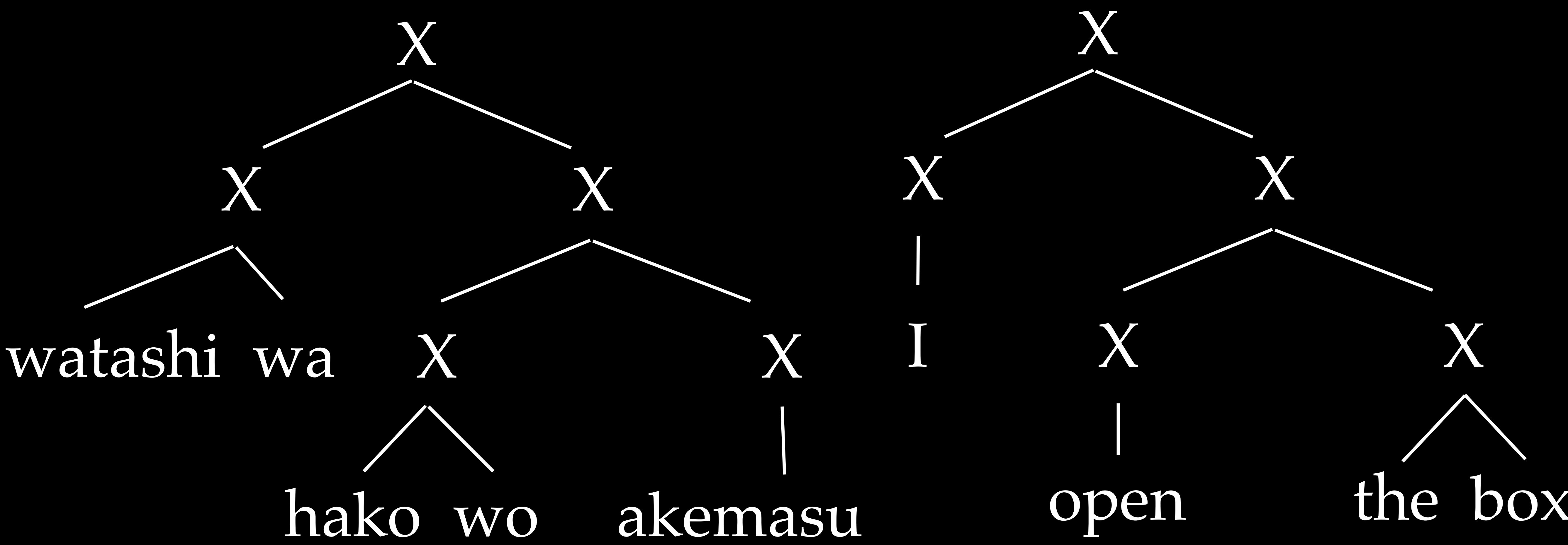
$X \rightarrow$ akemasu / open

← Translate words or phrases

The Big Question

Where do the categories come from?

Answer #1: there are no categories!



Inversion Transduction Grammar

Inversion Transduction Grammar

Parsing is polynomial. We must be giving up *something* in order to acheive polynomial complexity.

Inversion Transduction Grammar

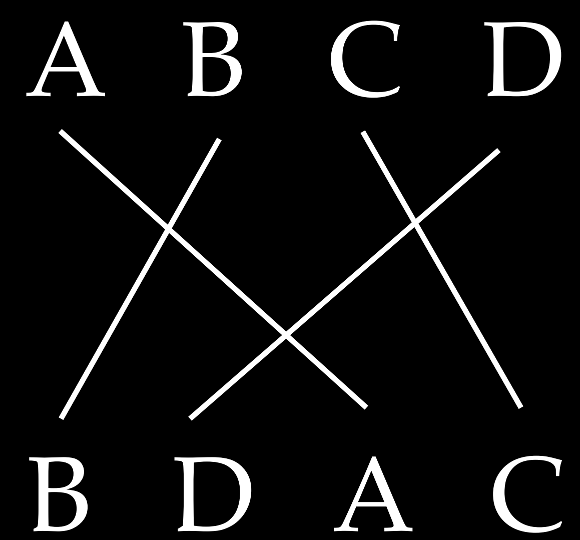
Parsing is polynomial. We must be giving up *something* in order to acheive polynomial complexity.

A B C D

B D A C

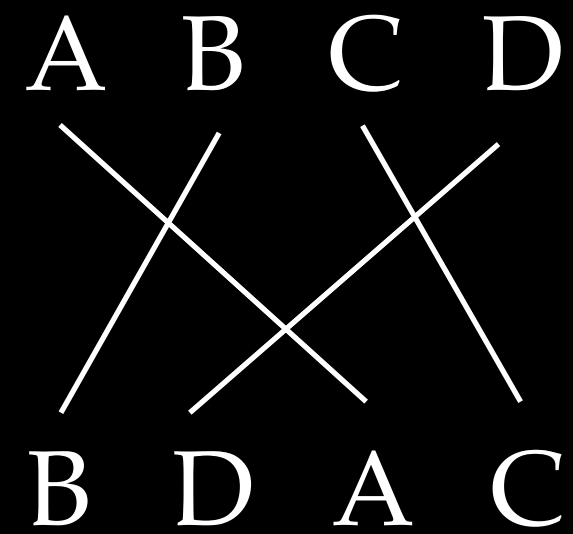
Inversion Transduction Grammar

Parsing is polynomial. We must be giving up *something* in order to acheive polynomial complexity.



Inversion Transduction Grammar

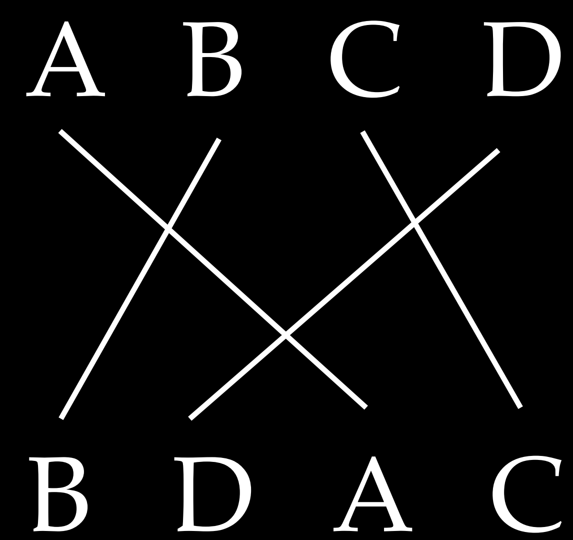
Parsing is polynomial. We must be giving up *something* in order to acheive polynomial complexity.



ITG cannot produce this kind of reordering.

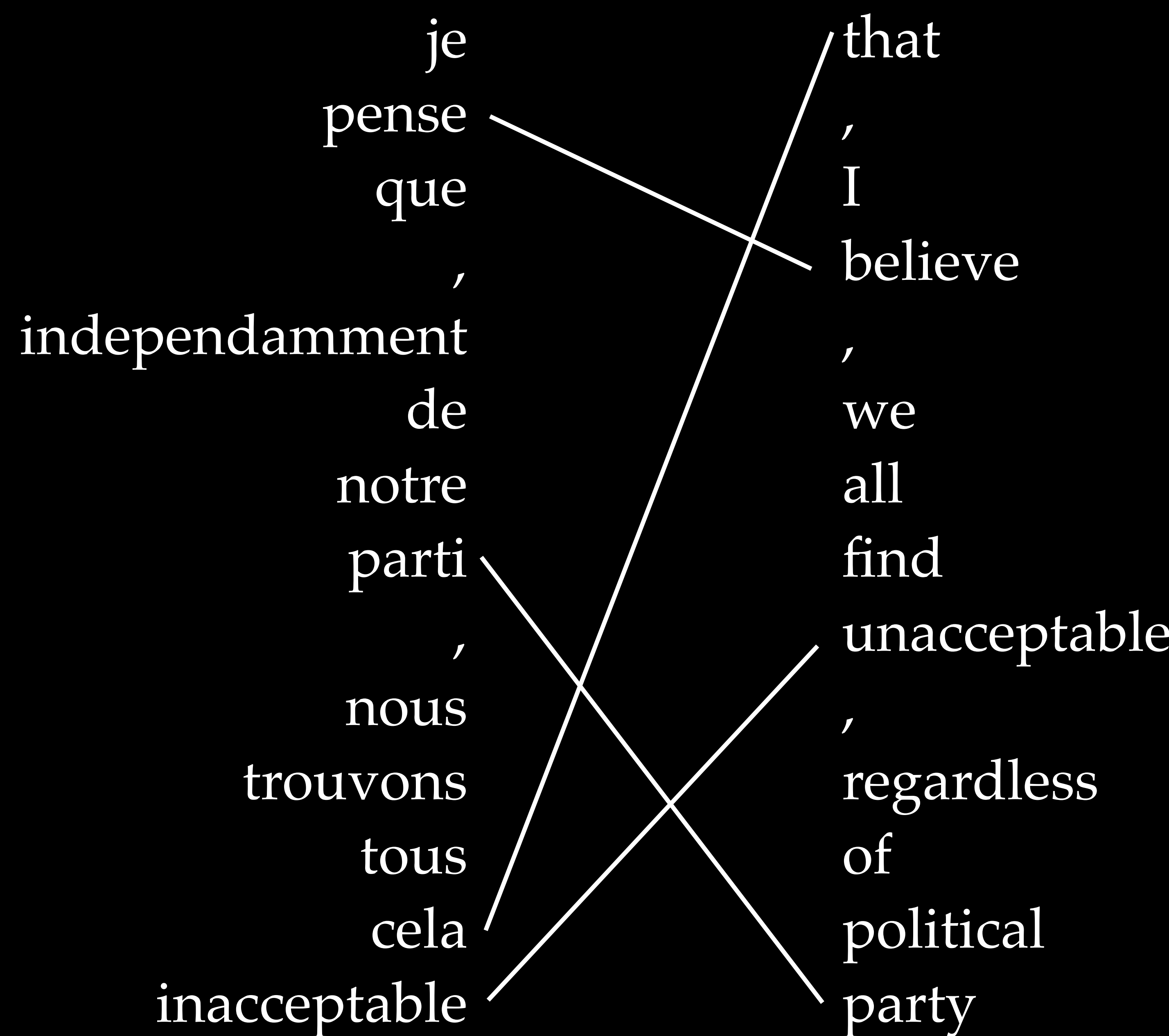
Inversion Transduction Grammar

Parsing is polynomial. We must be giving up *something* in order to acheive polynomial complexity.



ITG cannot produce this kind of reordering.
Does this matter? Do such reorderings occur in real data?

Inversion Transduction Grammar

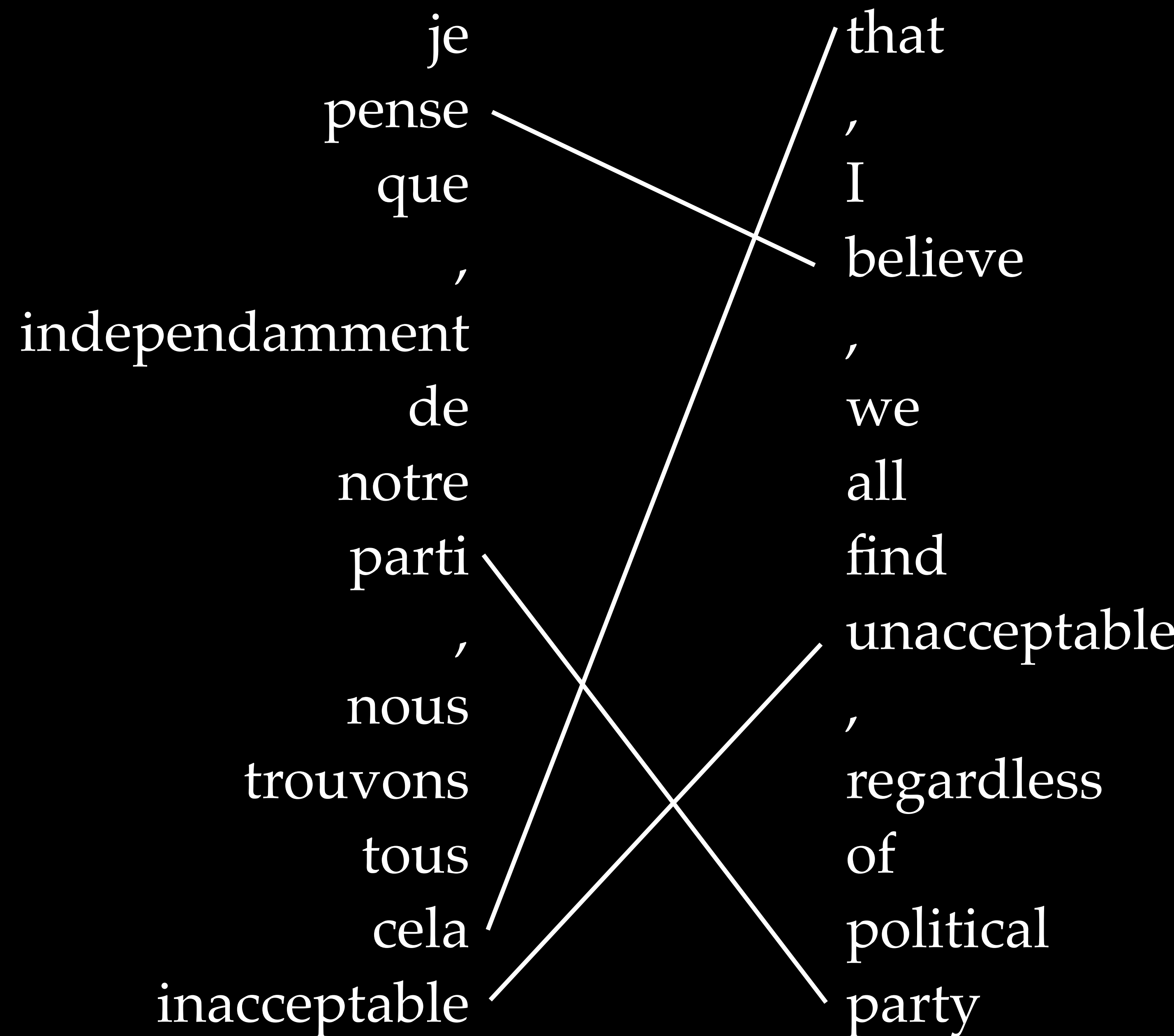


ITG cannot produce this kind of reordering.

Does this matter? Do such reorderings occur in real data?

YES!

Inversion Transduction Grammar



ITG cannot produce this kind of reordering.

Does this matter? Do such reorderings occur in real data?

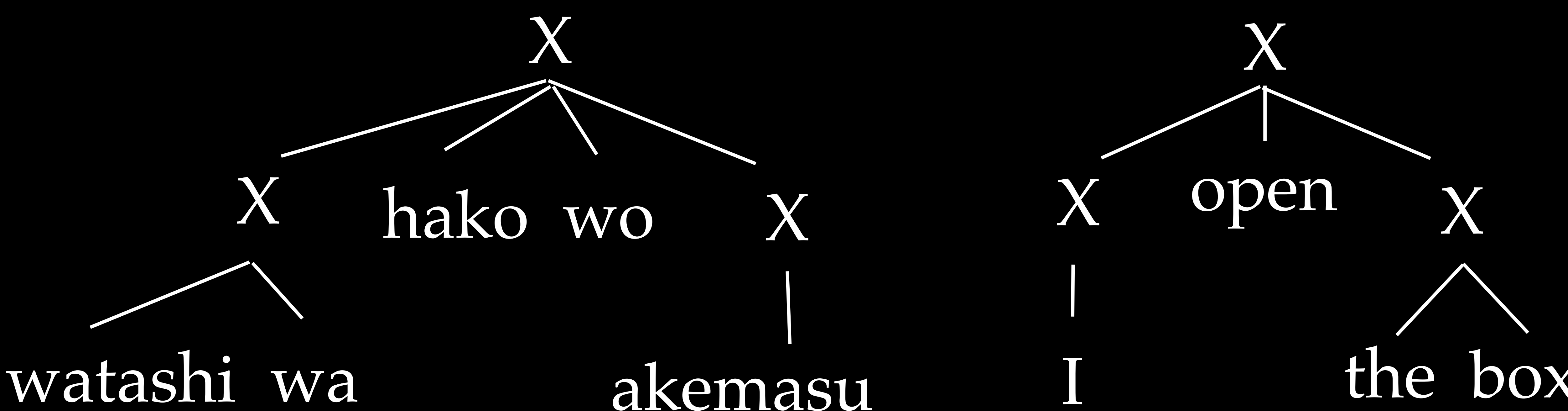
YES! (but they're very rare)

Hierarchical Phrase-Based Translation

$X \rightarrow X_1$ hako wo X_2 / X_1 open X_2

$X \rightarrow$ hako wo / the box

$X \rightarrow$ akemasu / open



The Big Question

Where do the categories come from?

The Big Question

Where do the categories come from?

Answer #2: from a parser.

The Big Question

Where do the categories come from?

Answer #2: from a parser.

$S \rightarrow NP_1\ VP_2\ /\ NP_1\ VP_2$

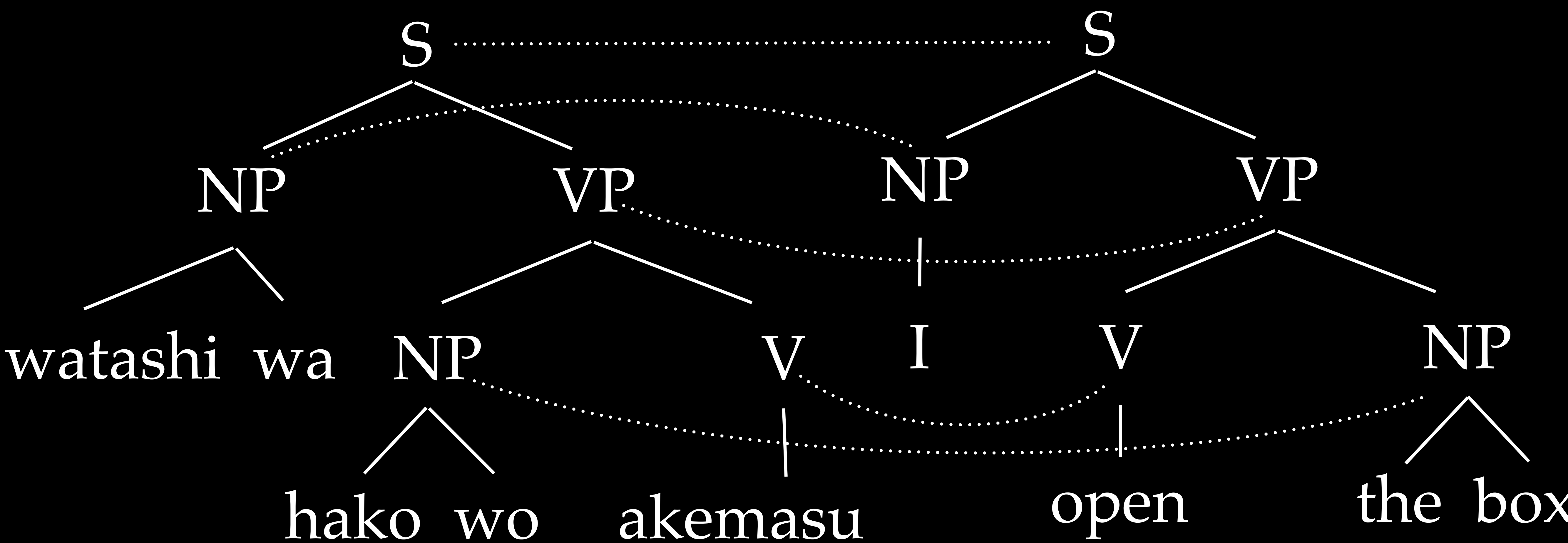
$NP \rightarrow$ watashi wa / I

$NP \rightarrow$ hako wo / the box

$VP \rightarrow NP_1\ V_2\ /\ V_1\ NP_2$

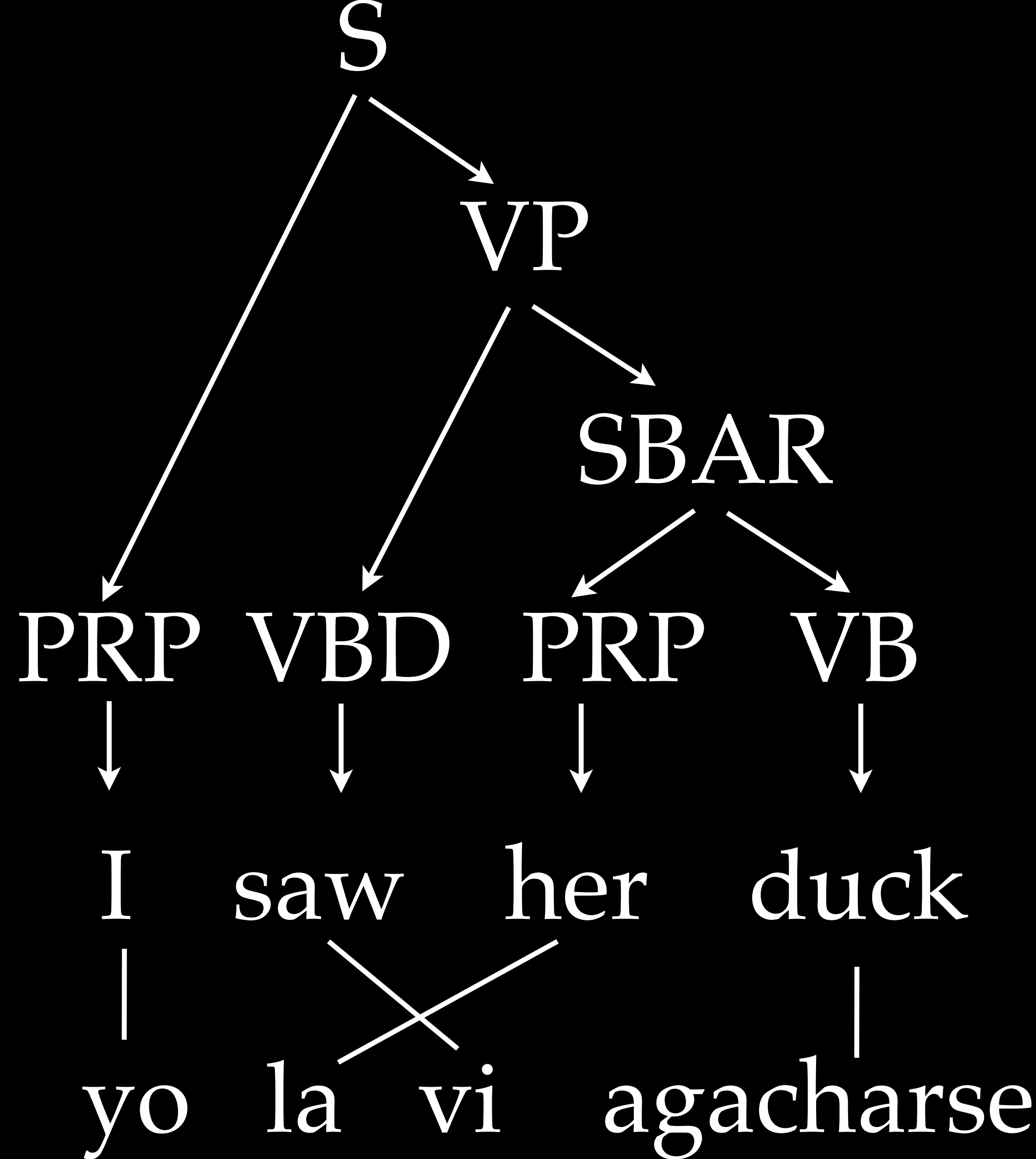
$V \rightarrow$ akemasu / open

Syntax-based Translation



Are reorderings in real data consistent with isomorphisms on linguistic parse trees?

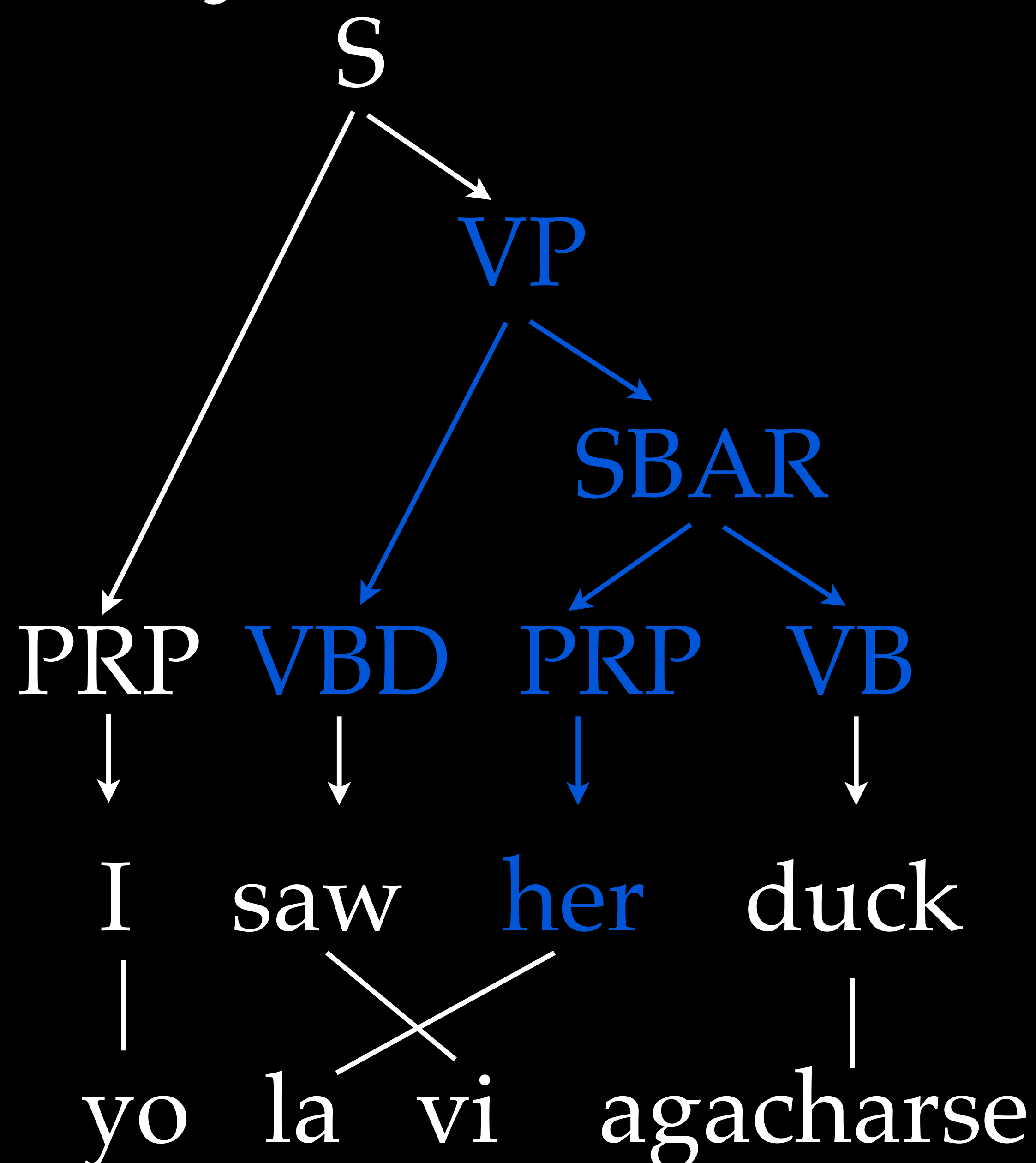
Syntax-based Translation



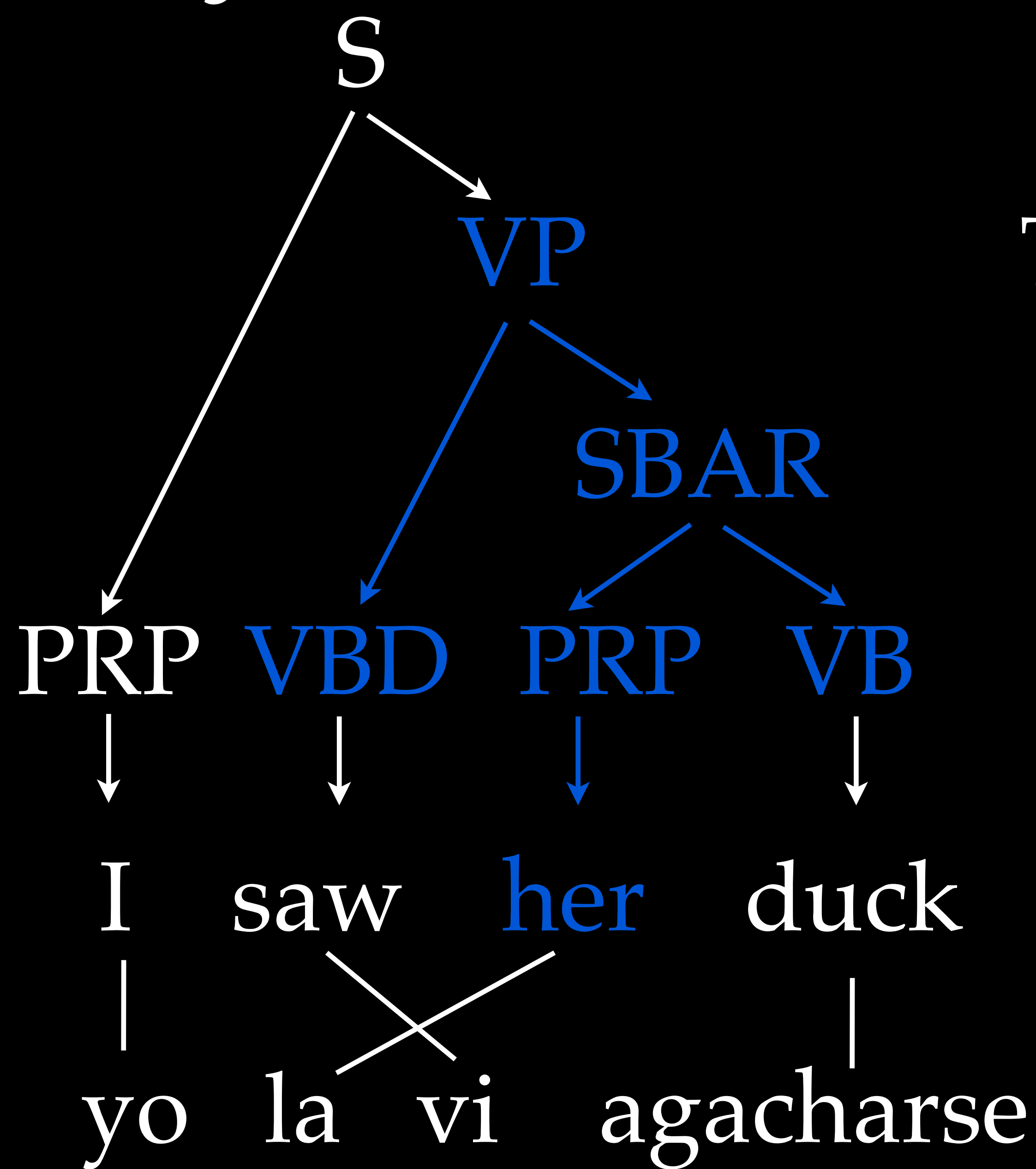
Are reorderings in real data consistent with isomorphisms on linguistic parse trees?

Of course not.

Syntax-based Translation

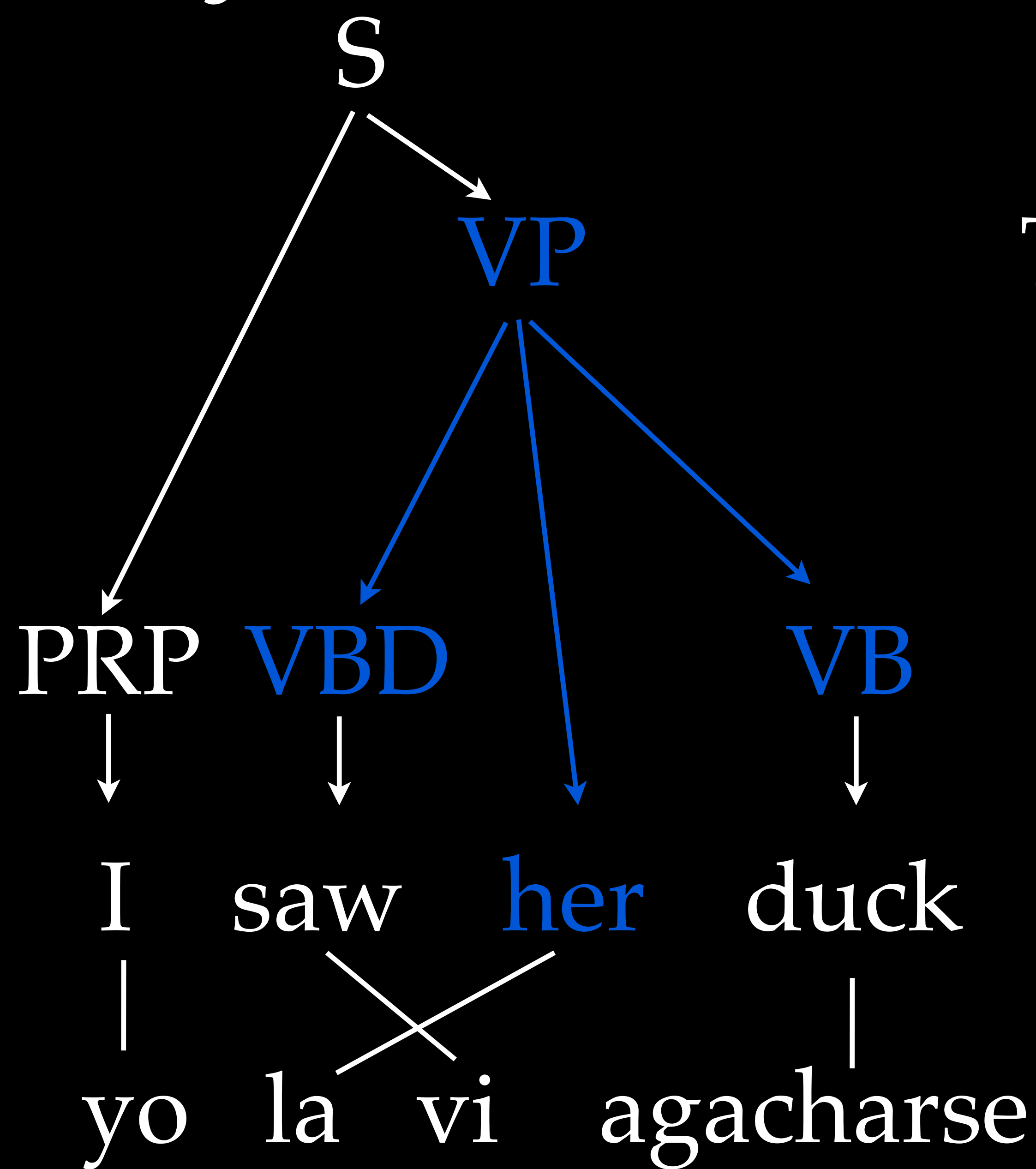


Syntax-based Translation



Tree substitution grammar

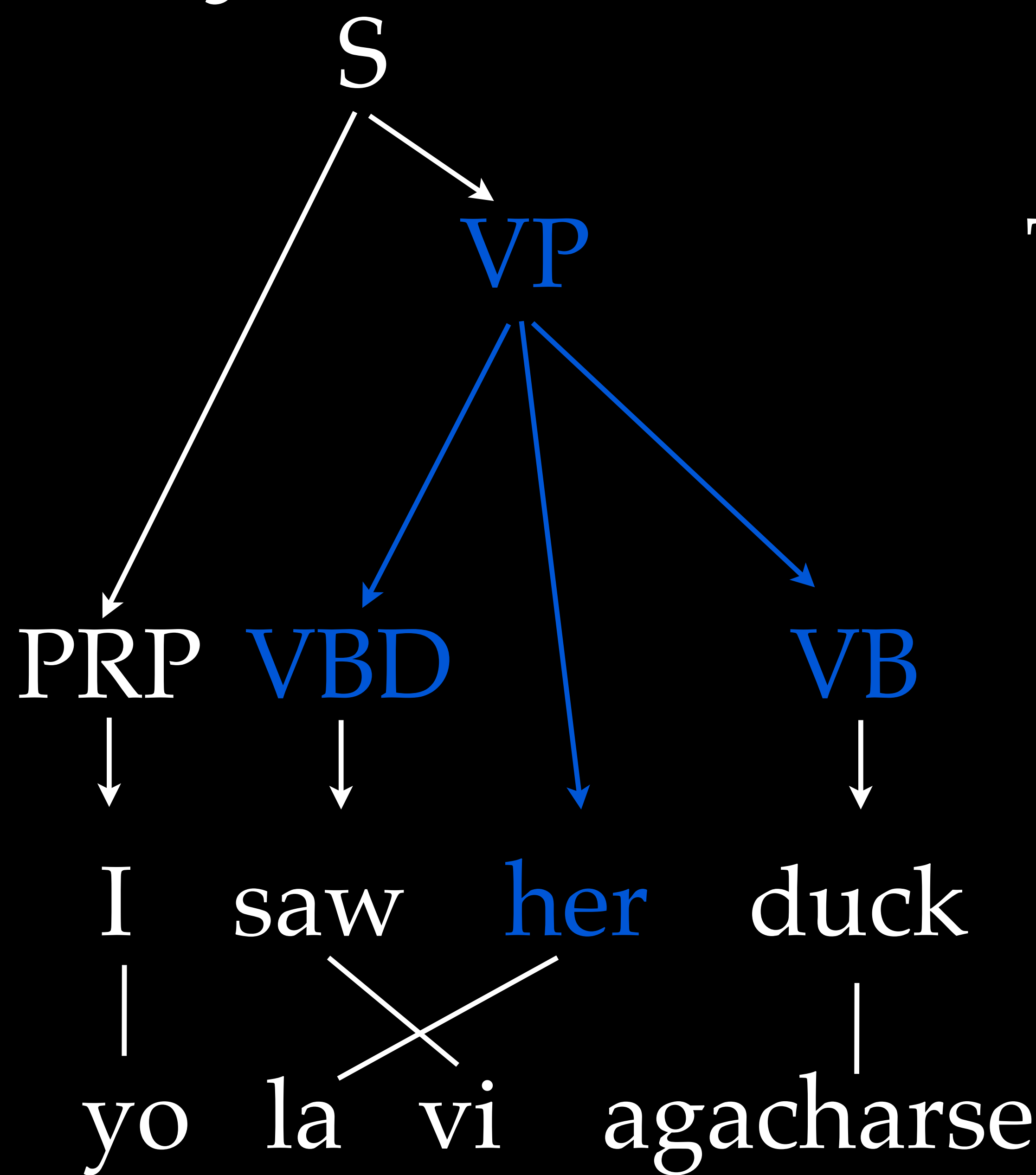
Syntax-based Translation



Tree substitution grammar

weakly equivalent SCFG

Syntax-based Translation



Tree substitution grammar

weakly equivalent SCFG

$VBD \rightarrow$ saw / vi

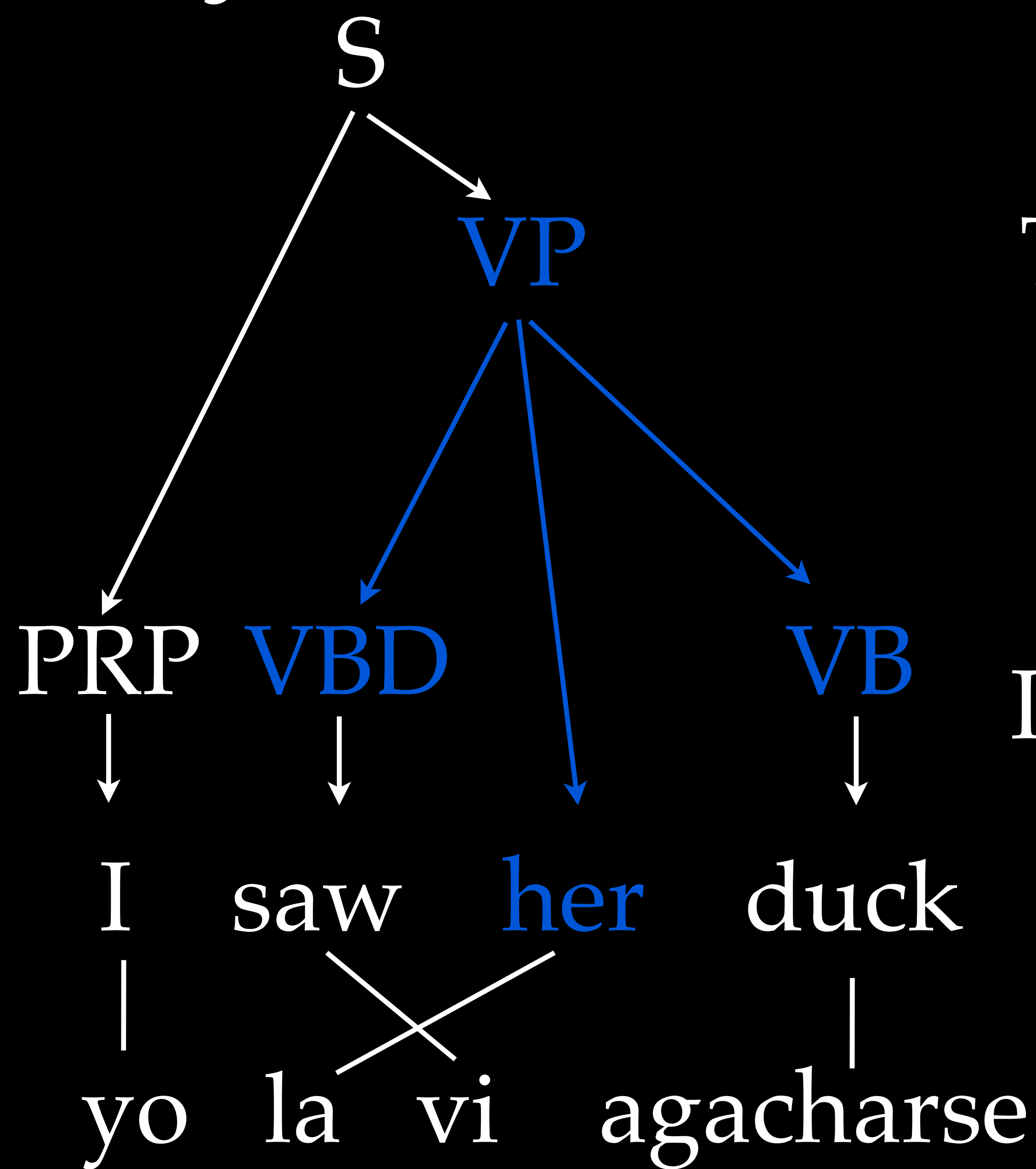
$VB \rightarrow$ duck / agacharse

$S \rightarrow PRP_1\ VP_2\ /\ PRP_1\ VP_2$

$PRP \rightarrow$ I / yo

$VP \rightarrow VBD_1$ her VB_2 / la $VBD_1\ VB_2$

Syntax-based Translation



Tree substitution grammar

weakly equivalent SCFG

Problem: we need a parser!

$$VBD \rightarrow \text{saw} \ / \ \text{vi}$$

$$VB \rightarrow \text{duck} \ / \ \text{agacharse}$$

$$S \rightarrow PRP_1 \ VP_2 \ / \ PRP_1 \ VP_2$$

$$PRP \rightarrow \text{I} \ / \ \text{yo}$$

$$VP \rightarrow VBD_1 \ \text{her} \ VB_2 \ / \ \text{la} \ VBD_1 \ VB_2$$

The Big Question

Where do the categories come from?

The Big Question

Where do the categories come from?

Answer #3: they are automatically induced!

The Big Question

Where do the categories come from?

Answer #3: they are automatically induced!

This is an area of active research.

www.clsp.jhu.edu/workshops/ws10/groups/msgismt/

Another Big Question...

Where do the grammars come from?

Recap: Expectation Maximization

- Arbitrarily select a set of parameters (say, uniform).
- Calculate ***expected counts*** of the unseen events.
- Choose new parameters to maximize likelihood, using expected counts as proxy for observed counts.
- Iterate.
- Guaranteed that likelihood is monotonically nondecreasing.

Can we apply it to other models?

- Sure, why not?
- The derivation structure of each model is simply a latent variable.
- We simply apply EM to each model structure.

Recap: Expectation Maximization

- Arbitrarily select a set of parameters (say, uniform).
- Calculate ***expected counts*** of the unseen events.
- Choose new parameters to maximize likelihood, using expected counts as proxy for observed counts.
- Iterate.
- Guaranteed that likelihood is monotonically nondecreasing.

Recap: Expectation Maximization

- Arbitrarily select a set of parameters (say, uniform).
- Calculate ***expected counts*** of the unseen events.
- Choose new parameters to maximize likelihood, using expected counts as proxy for observed counts.
- Iterate.
- Guaranteed that likelihood is monotonically nondecreasing.

Recap: Expectation Maximization

- Arbitrarily select a set of parameters (say, uniform).
- Calculate *expected counts* of the unseen events.
- Choo hood, using d counts.
- Iterate.
- Guaranteed that likelihood is monotonically nondecreasing.

BAD:
Objective function is highly non-convex

Recap: Expectation Maximization

- Arbitrarily select a set of parameters (say, uniform).
- Calculate ***expected counts*** of the unseen events.
- Choose new parameters to maximize likelihood, using expected counts as proxy for observed counts.
- Iterate.
- Guaranteed that likelihood is monotonically nondecreasing.

Recap: Expectation Maximization

- Arbitrarily select a set of parameters (say, uniform).
- Calculate ***expected counts*** of the unseen events.
- Choose new parameters to maximize likelihood, us… …nts.
- It…
- G… n…

WORSE:
Computing expectations from a phrase-based model, given a sentence pair, is NP-Complete (by reduction to SAT; DeNero & Klein, 2008)

Recap: Expectation Maximization

- Arbitrarily select a set of parameters (say, uniform).
- Calculate ***expected counts*** of the unseen events.
- Choose new parameters to maximize likelihood, us … nts.
- It
- G
  n

Computing expectations from an SCFG model, given a sentence pair, is at least *$O(n^6)$*

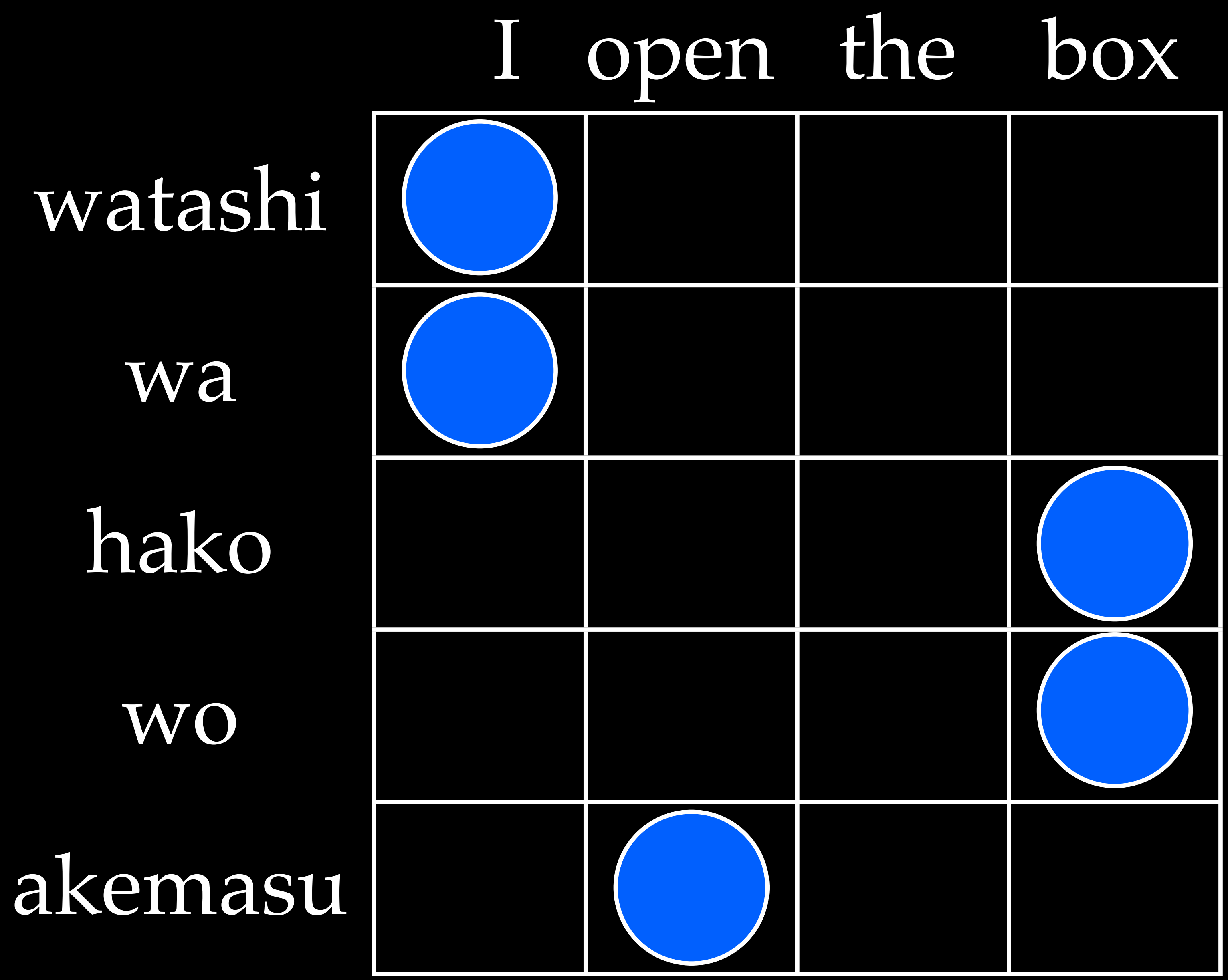
Now What?

- Option #1: approximate expectations
  - Restrict computation to some tractable subset of the alignment space (arbitrarily biased).
  - Markov chain Monte Carlo (very slow).

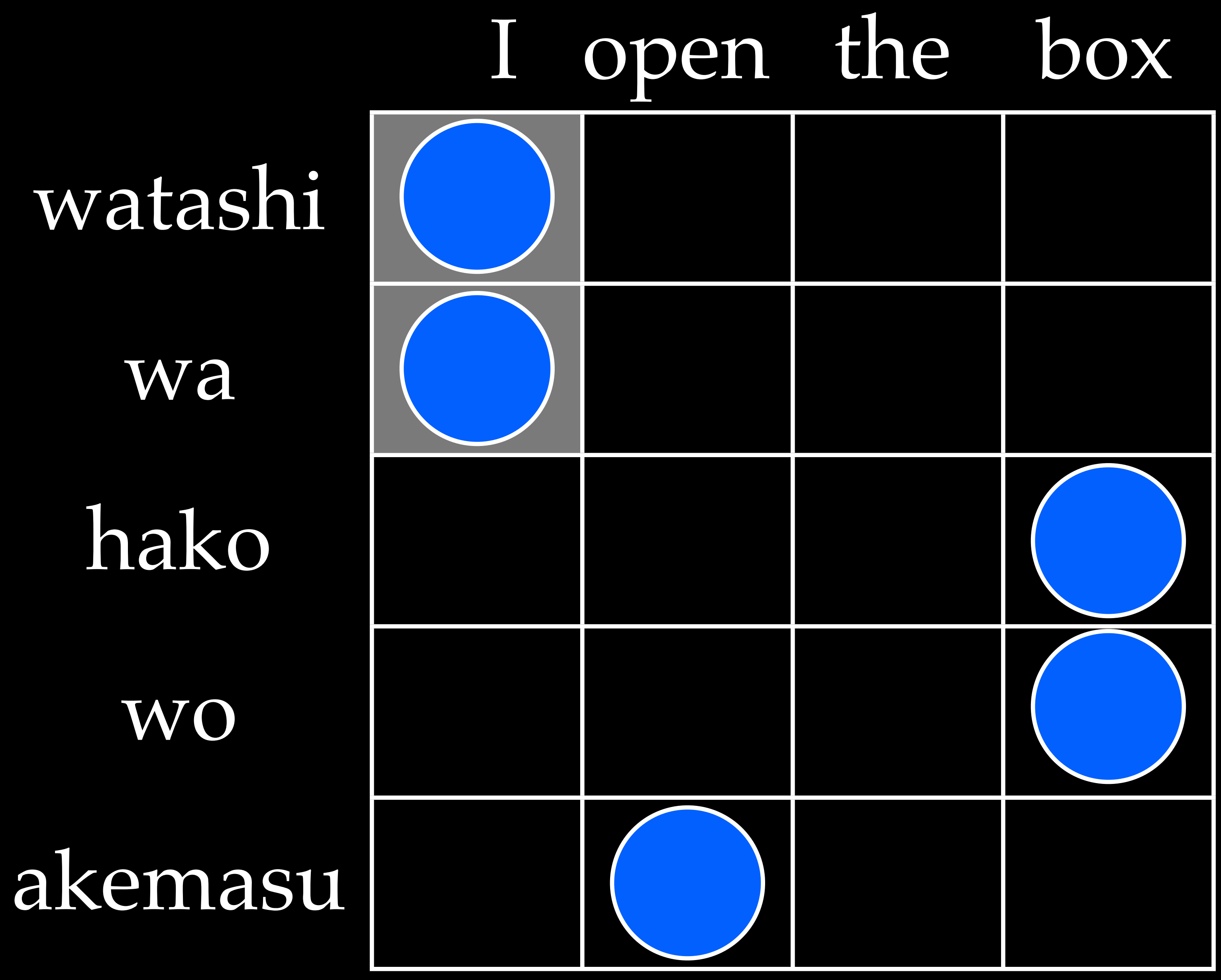
Now What?

- Option #2: change the problem definition
  - We already know how to learn word-to-word translation models efficiently.
  - Idea: learn word-to-word alignments, extract most probable alignment, then treat it as observed.
  - Learn phrase translations consistent with word alignments.
  - Decouples alignment from model learning -- is this a good thing?

Phrase Extraction

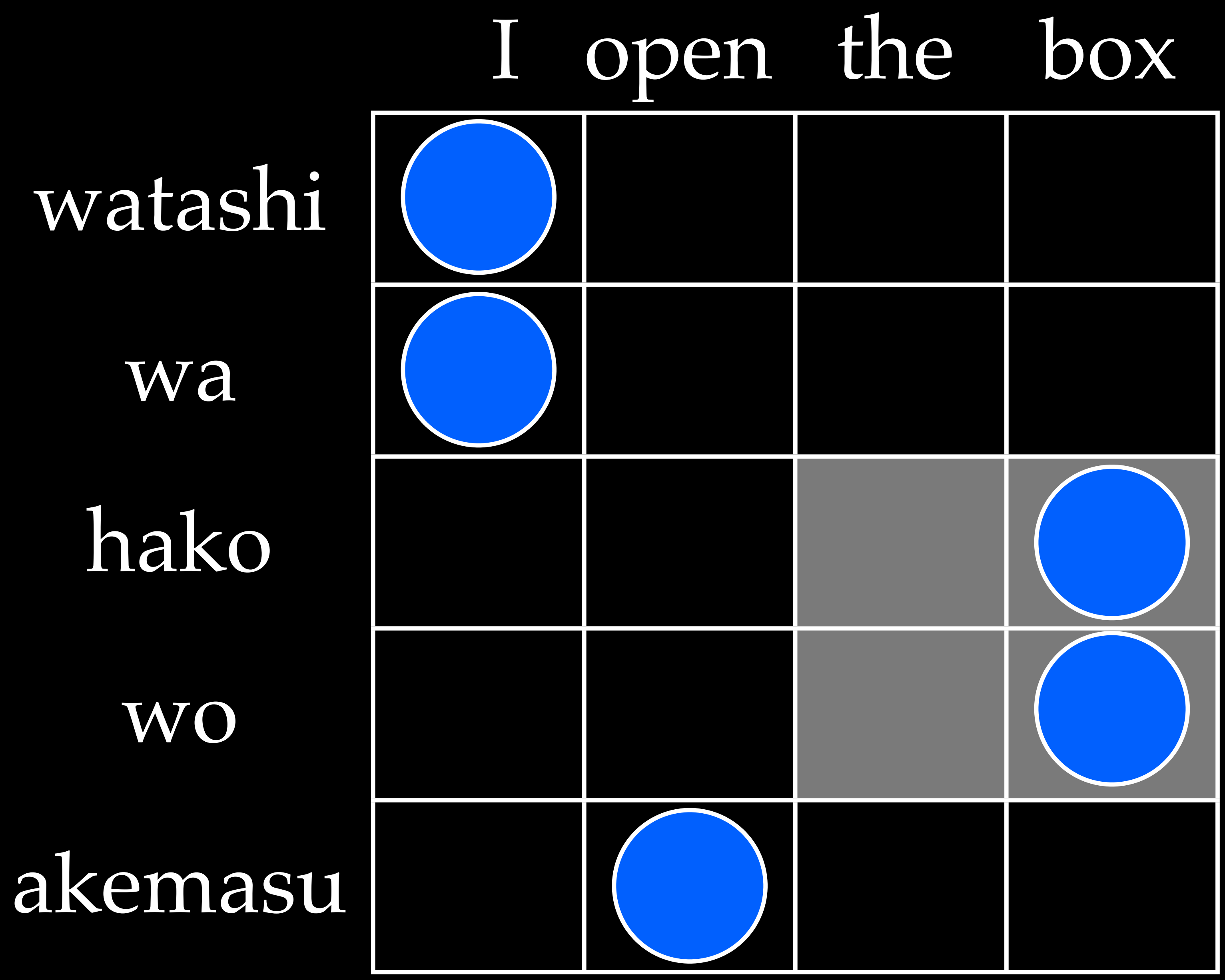


Phrase Extraction



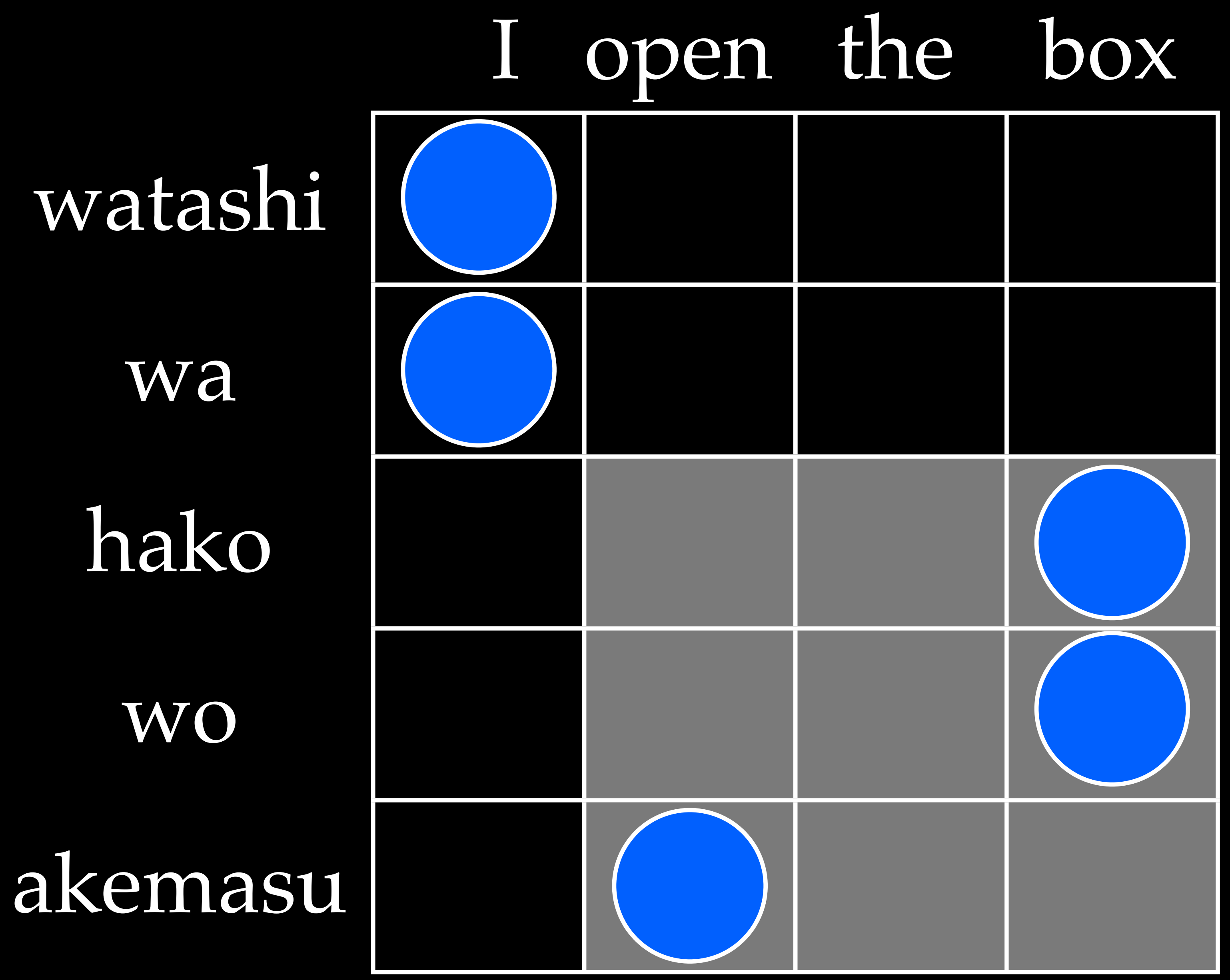
watashi wa / I

Phrase Extraction



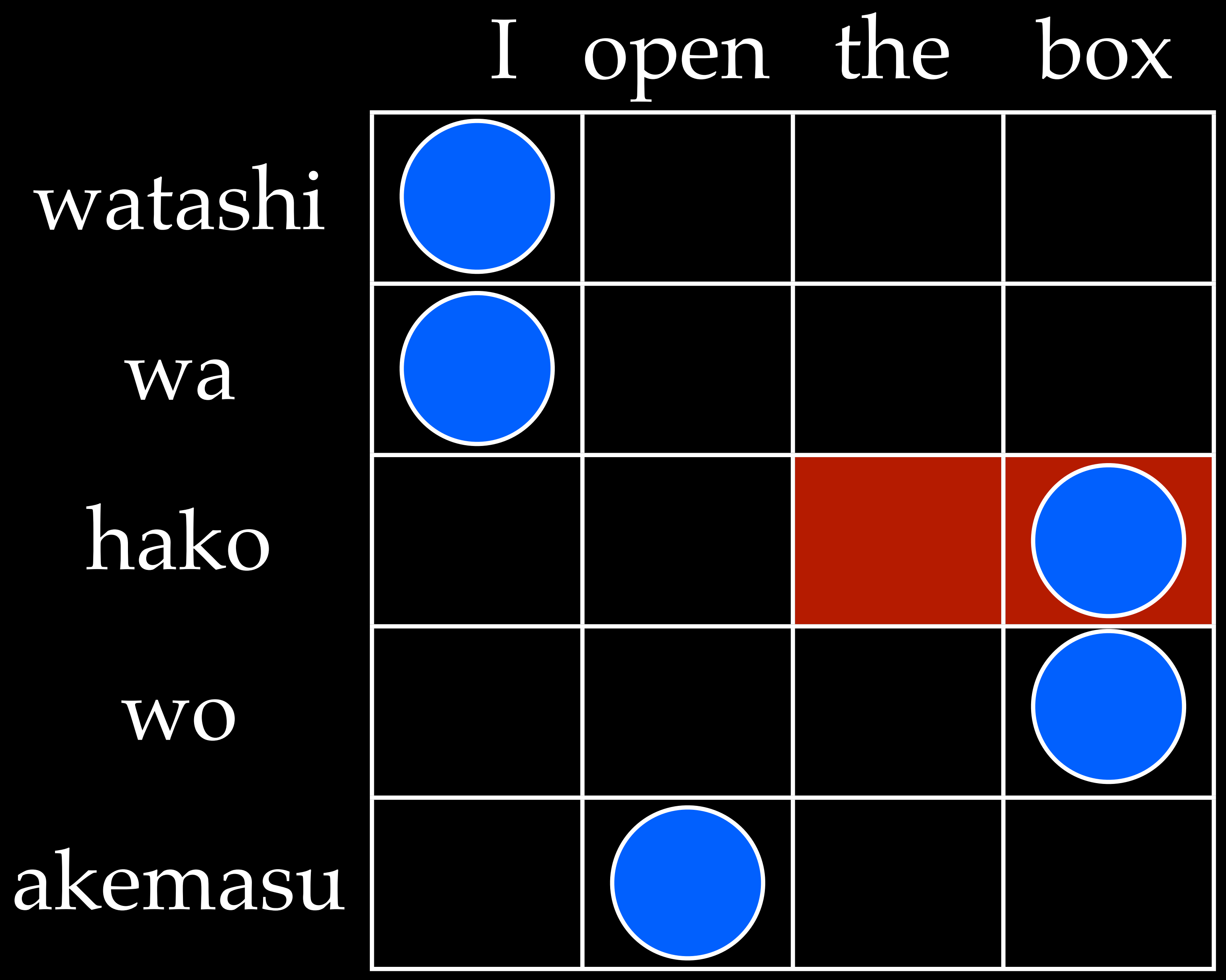
hako wo / the box

Phrase Extraction

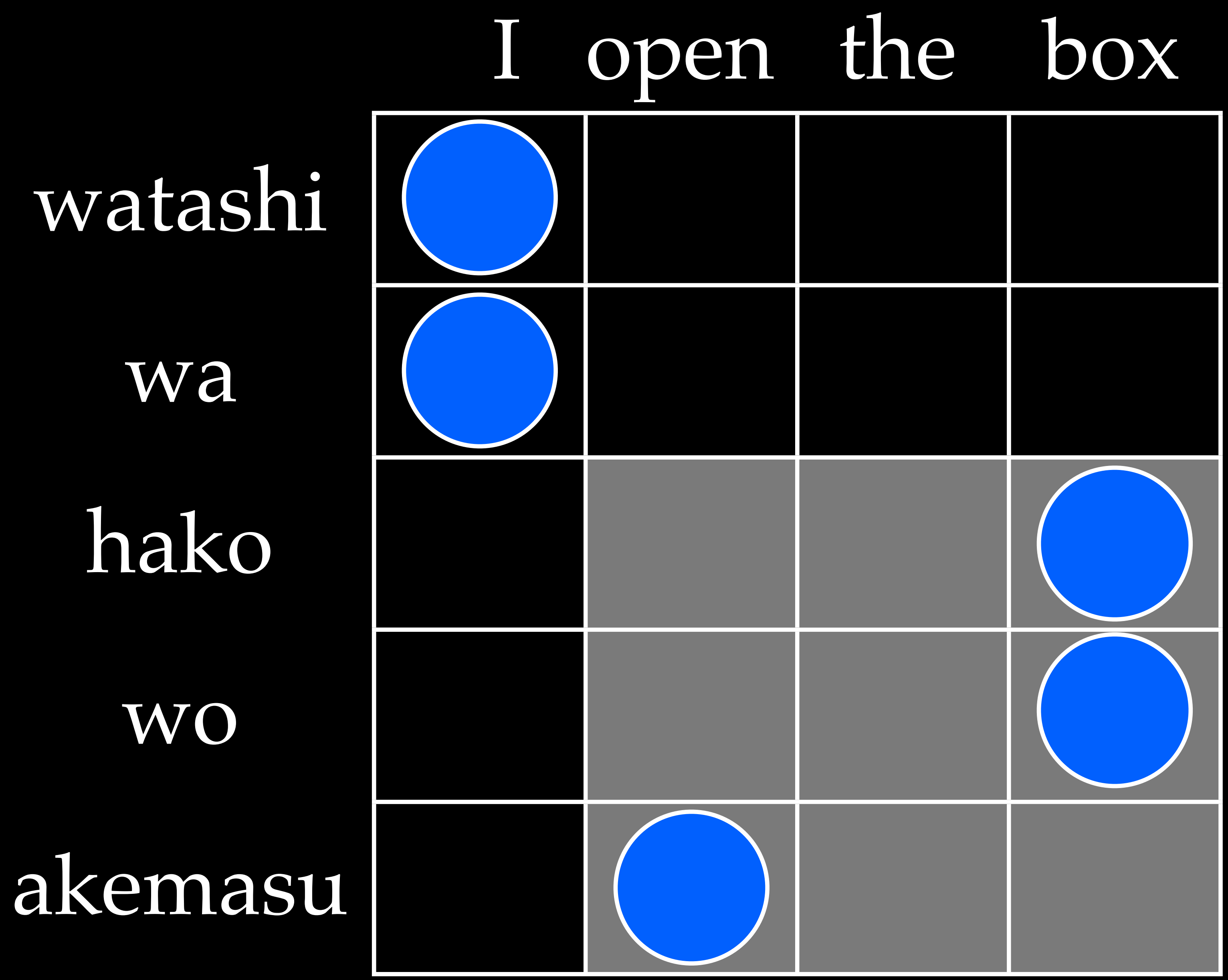


hako wo akemasu / open the box

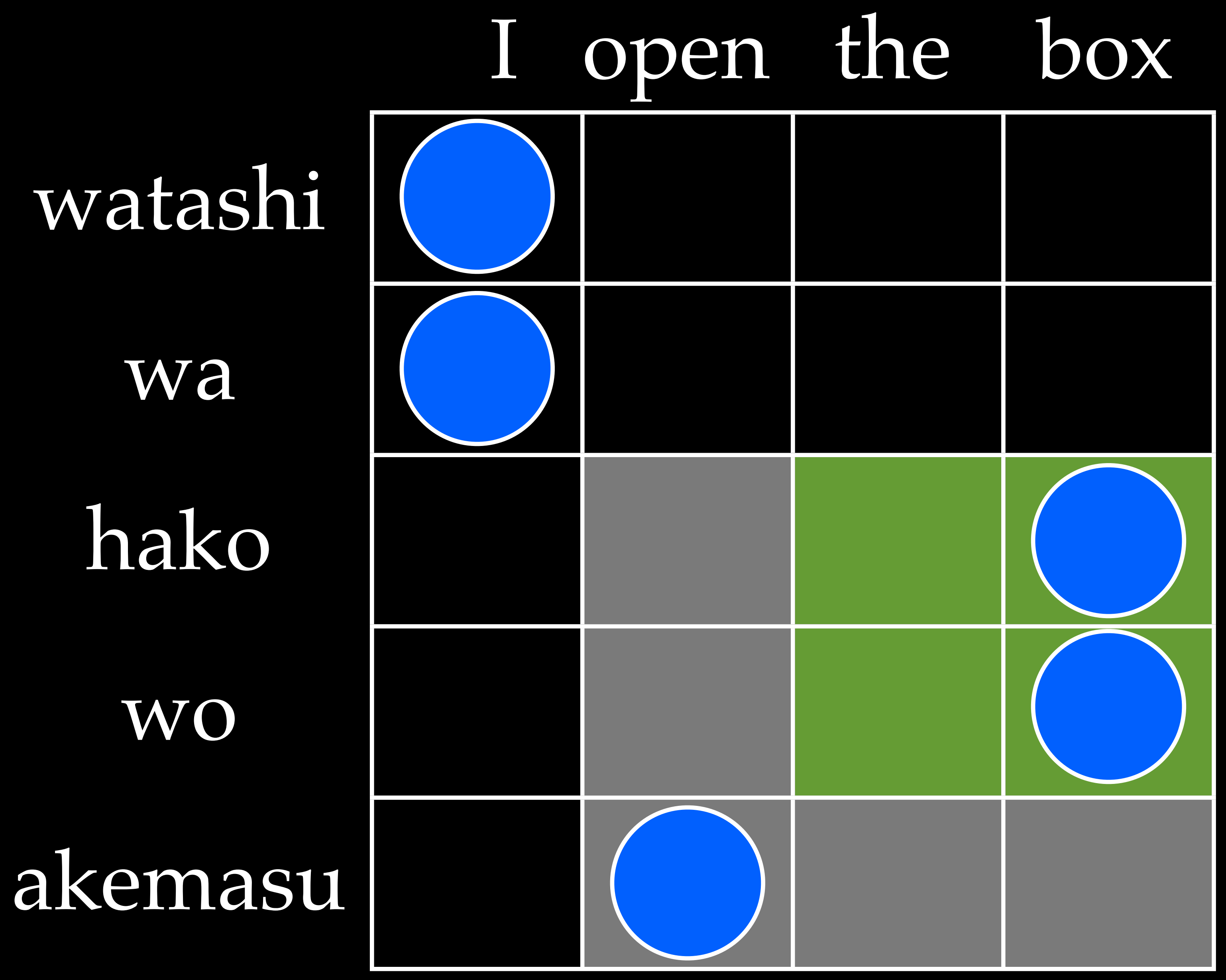
Phrase Extraction



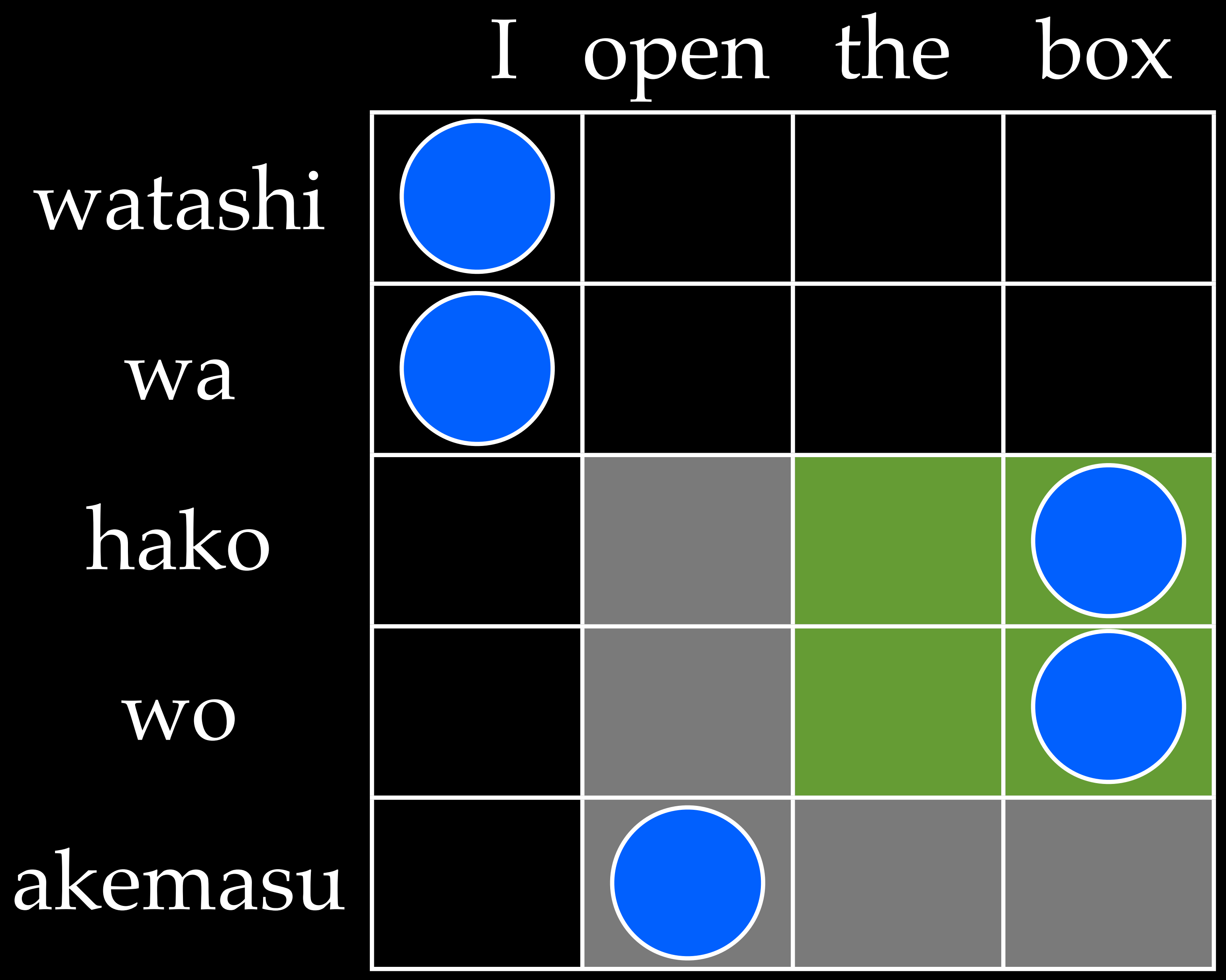
Hierarchical Phrase Extraction



Hierarchical Phrase Extraction

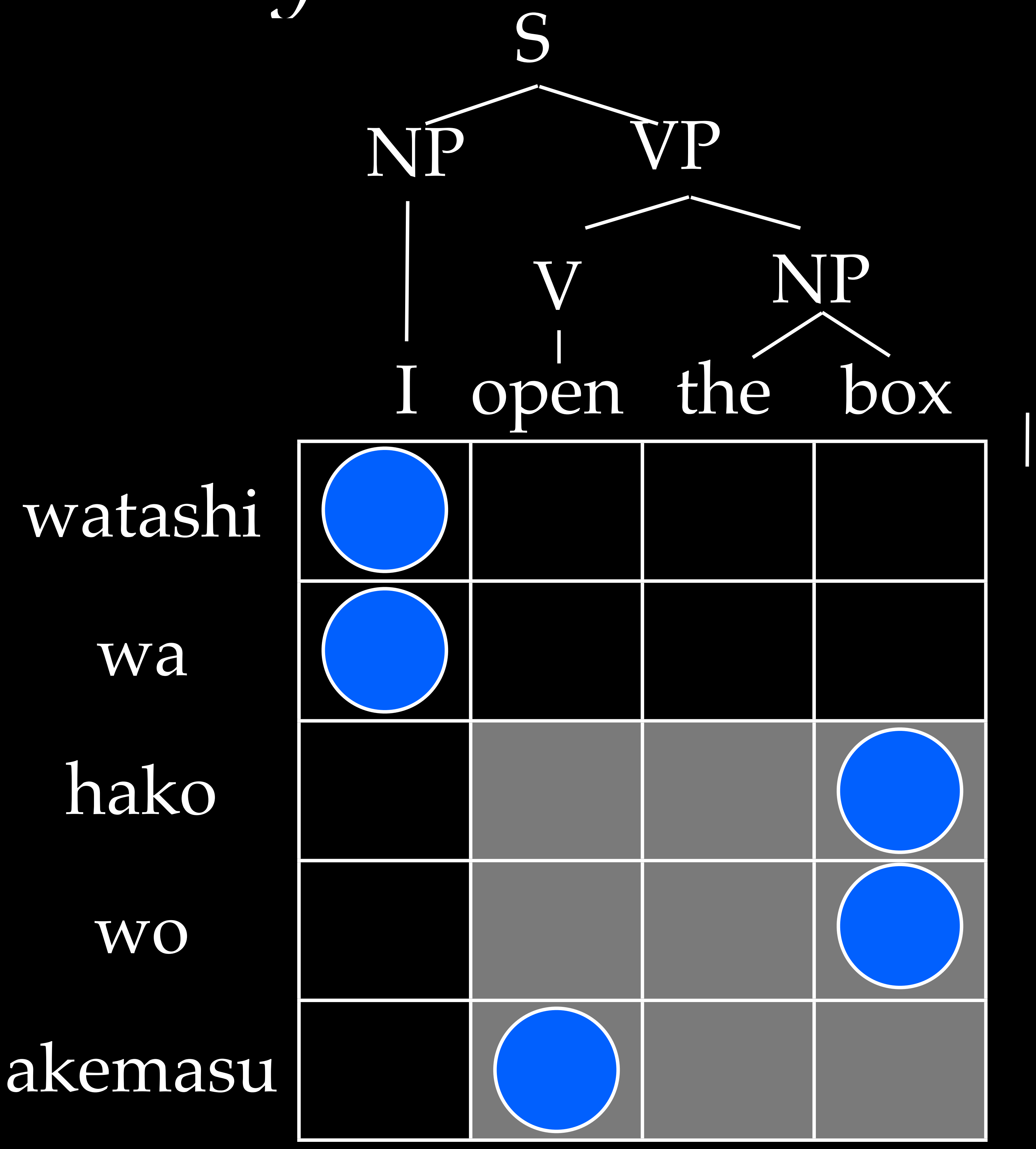


Hierarchical Phrase Extraction

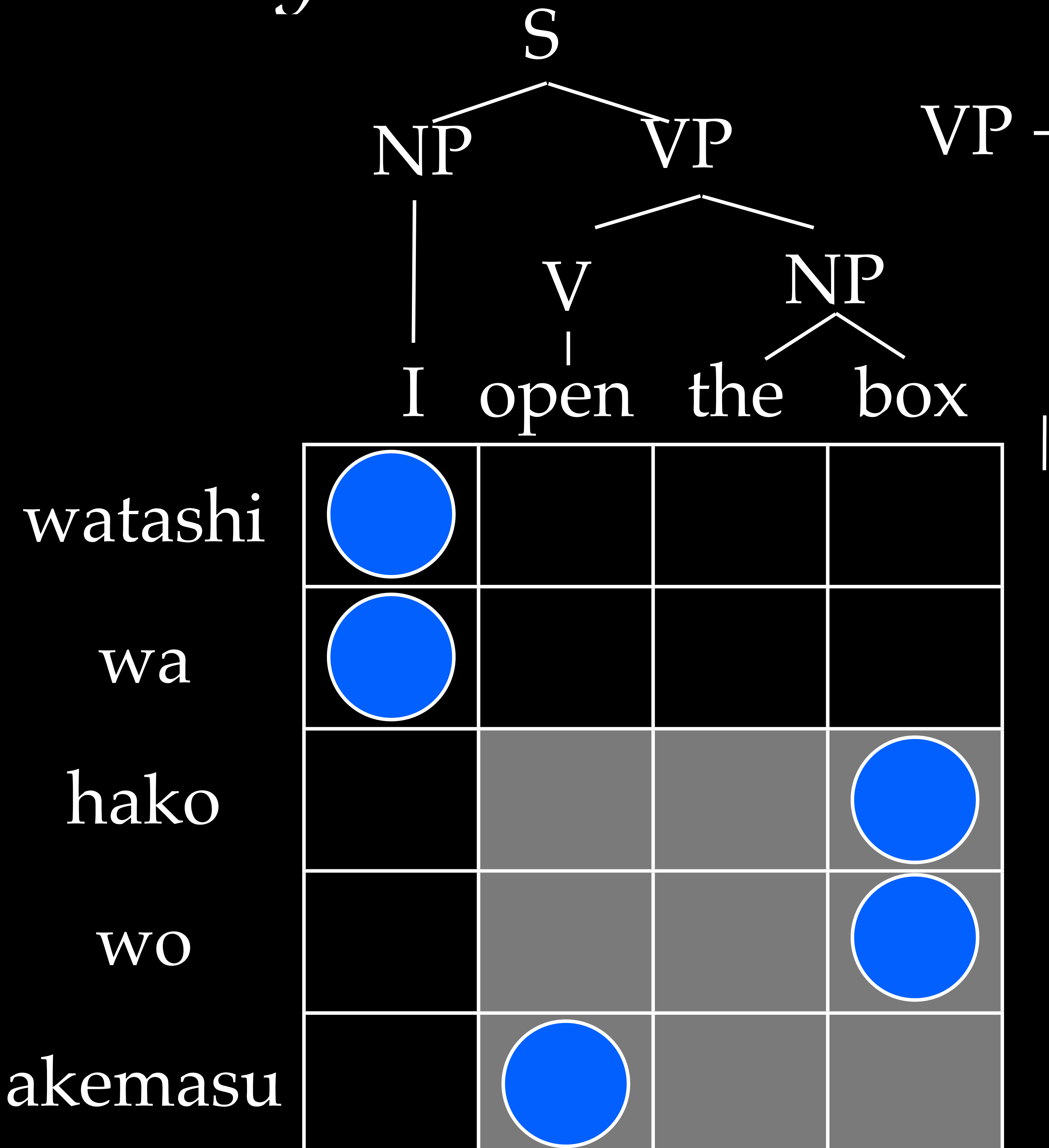


X_1 akemasu / open X_1

Syntactic Phrase Extraction

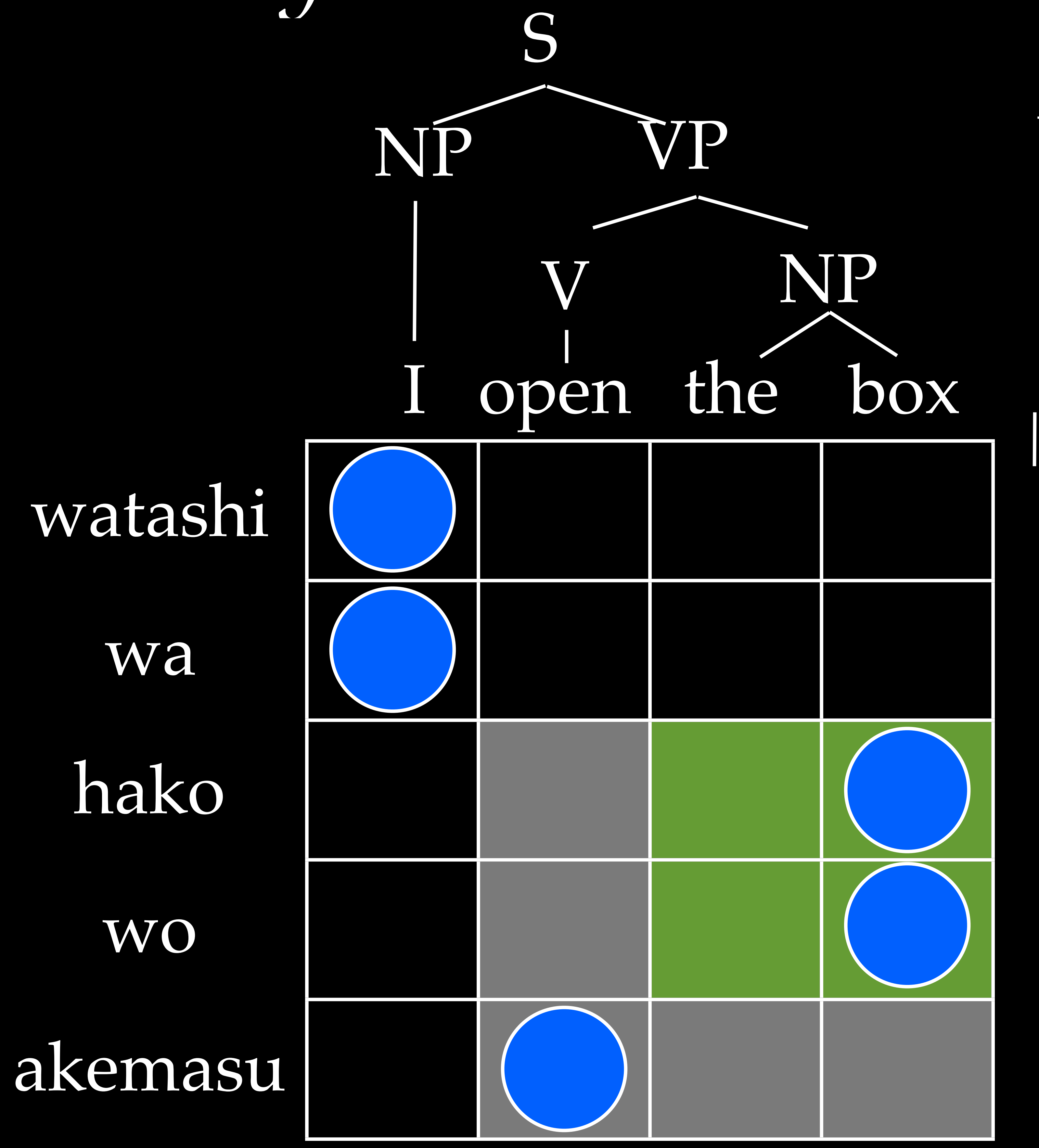


Syntactic Phrase Extraction



VP → hako wo akemasu / open the box

Syntactic Phrase Extraction



$VP \rightarrow NP_1$ akemasu / open NP_1

Summary

- Unsupervised learning over intractable models turns out to be a hard problem.
- Heuristic methods are widely used, but they offer no useful guarantees and are highly biased.
- Finding more elegant approximations is a topic of ongoing research.

Implementations

- Synchronous context-free translation models
  - Moses -- www.statmt.org/moses
  - cdec -- www.cdec-decoder.org
  - Joshua -- www.cs.jhu.edu/~ccb/joshua

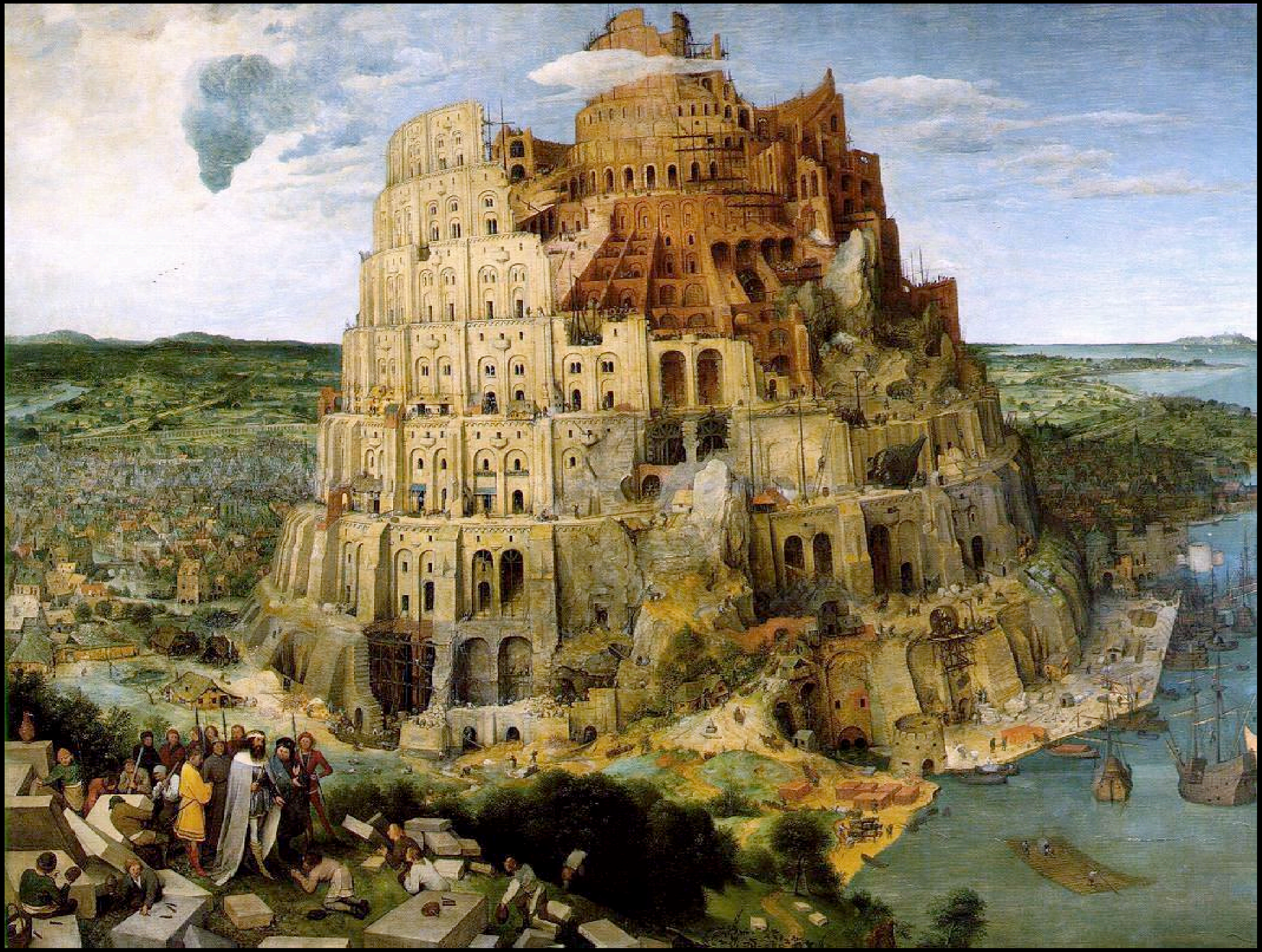
Datasets

- Proceedings of the European Parliament
  - www.statmt.org/europarl
- Linguistic Data Consortium
  - www.ldc.upenn.edu

Summary

- Many probabilistic translation models can be thought in terms of weighted (formal) languages.
- Dynamic programming is a common (though not universal!) decoding strategy.
- With these concepts in mind, you might be able to define models that capture other translation phenomena (e.g. morphosyntactics, semantics).

Recap



The Tower of Babel

Pieter Brueghel the Elder (1563)

Google

Language Tools

Translated search

Type a search phrase in your language. Google will find results in other languages and translate them for you to read.

Search for: [ ] Translate and Search

Search pages written in:
- (•) **Automatically selected languages**
- ( ) Specific languages

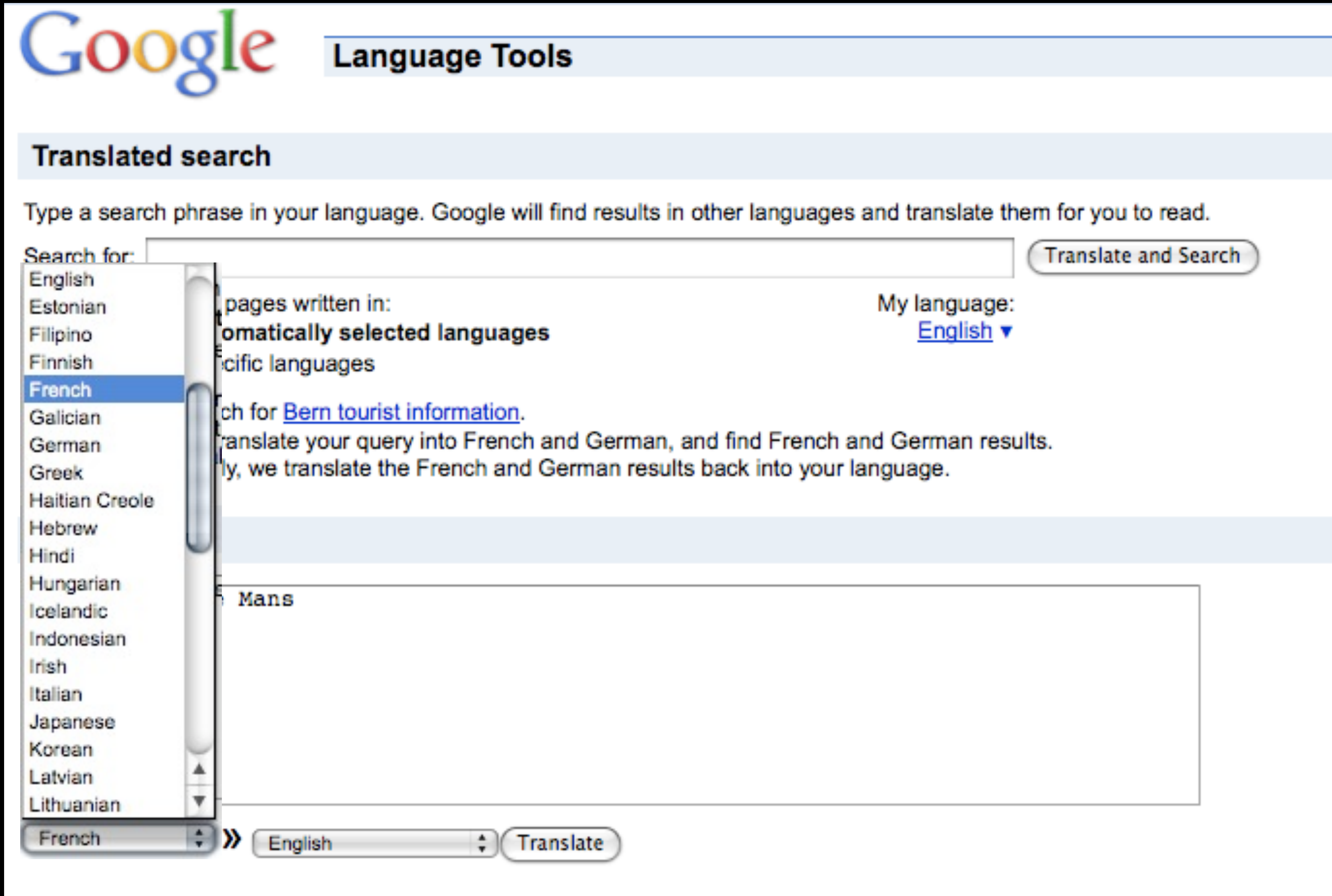
My language:
English ▾

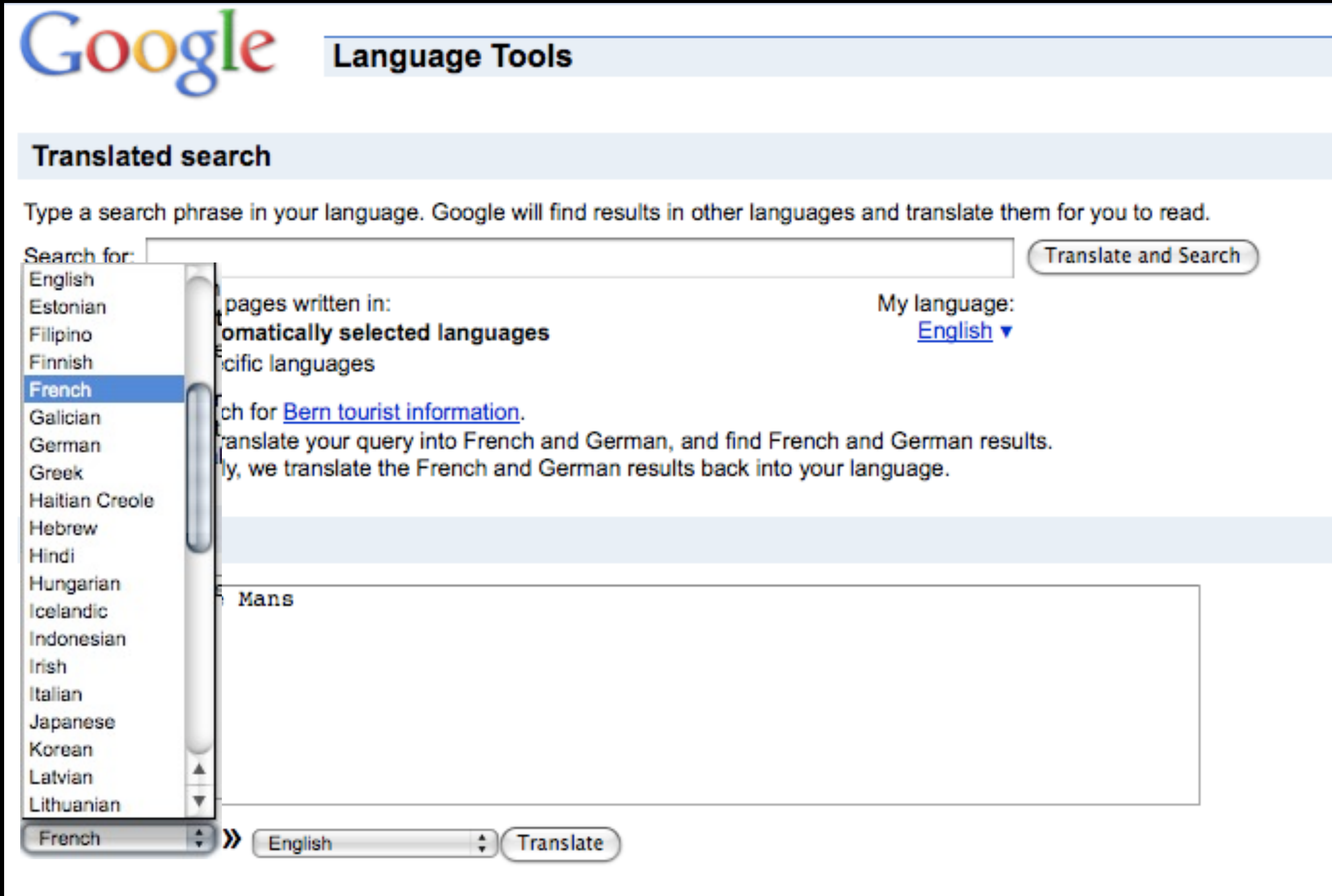
Example: 1. Search for Bern tourist information.
2. We translate your query into French and German, and find French and German results.
3. Finally, we translate the French and German results back into your language.

Translate text

Bienvenue à Le Mans

French » English Translate


Google
Language Tools
Translated search
Type a search phrase in your language. Google will find results in other languages and translate them for you to read.
Search for:
Translate and Search
pages written in:
omatically selected languages
cific languages
My language:
English ▼
ch for Bern tourist information.
ranslate your query into French and German, and find French and German results.
ly, we translate the French and German results back into your language.
Mans
English
Estonian
Filipino
Finnish
French
Galician
German
Greek
Haitian Creole
Hebrew
Hindi
Hungarian
Icelandic
Indonesian
Irish
Italian
Japanese
Korean
Latvian
Lithuanian
French
»
English
Translate

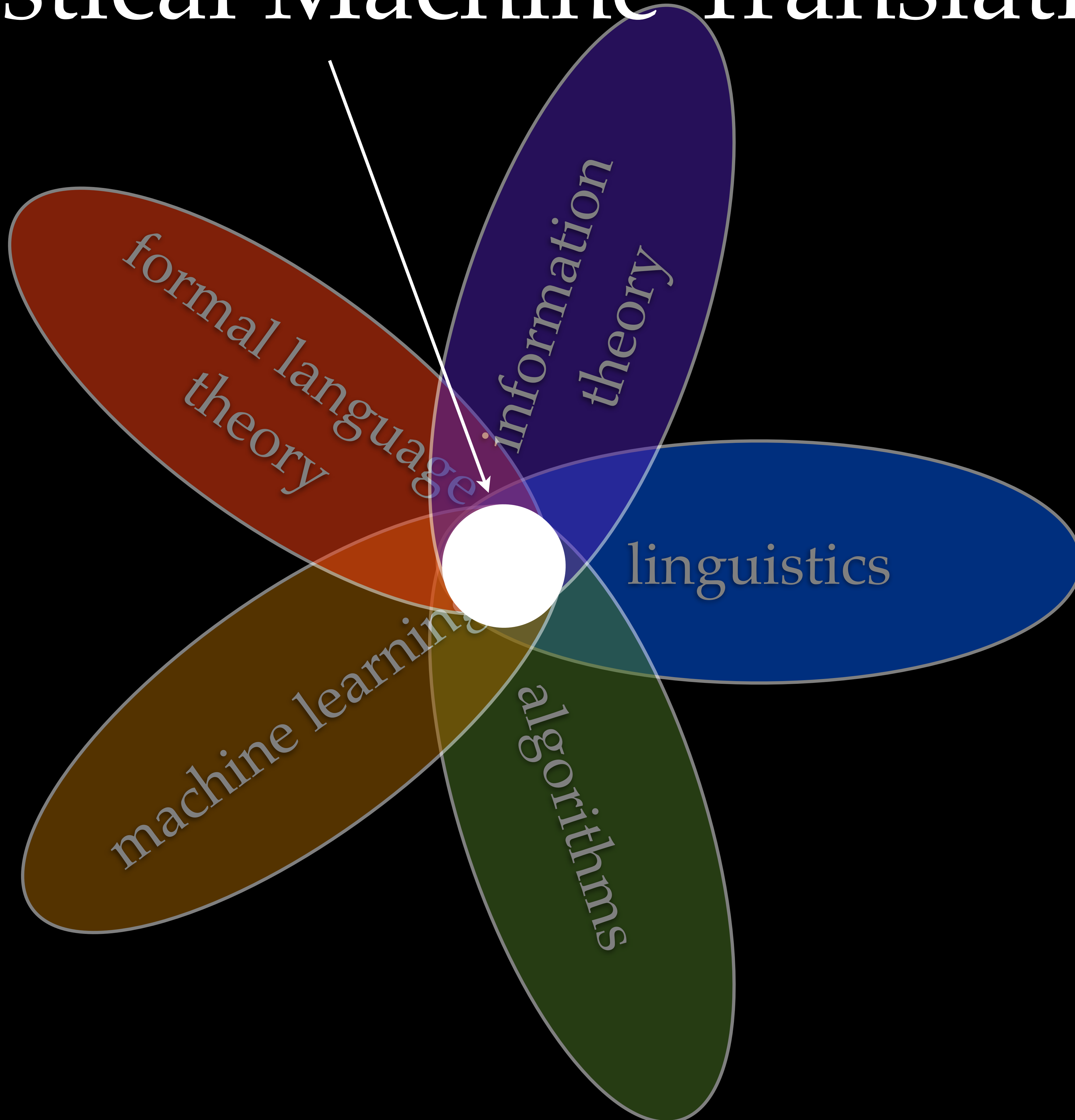


2756 language pairs!

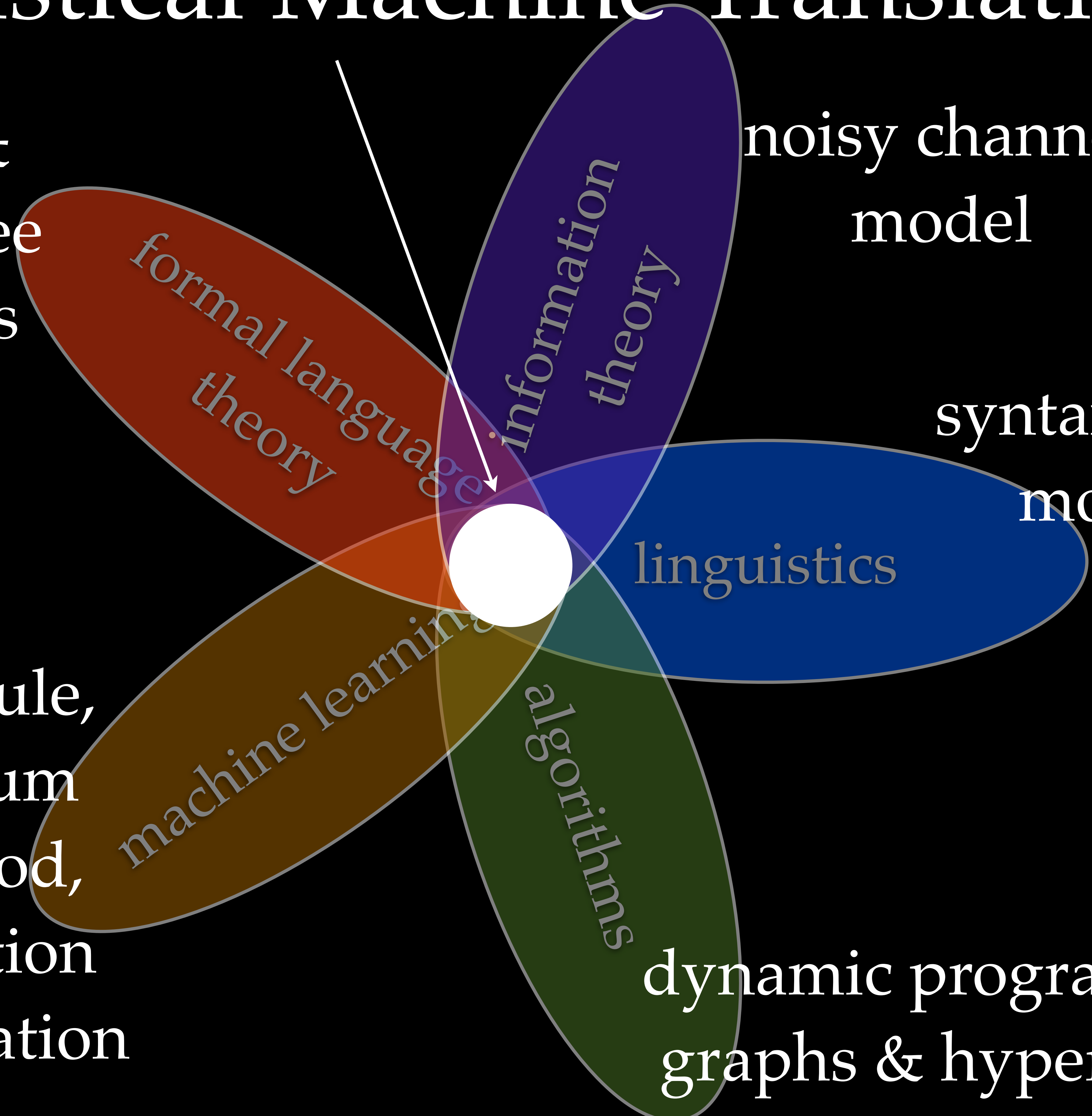
Statistical Machine Translation

Develop a statistical ***model*** of translation that can be ***learned*** from ***data*** and used to ***predict*** the correct English translation of new Chinese sentences.

Statistical Machine Translation



Statistical Machine Translation



The Data Deluge

- We are overwhelmed with data, but we can harness it to solve real problems.
- Formal tools help us model the data.
- Probabilistic tools help us learn models and make predictions.
- Algorithmic optimization methods make it all run.
- Tradeoffs: model expressivity vs. tractability.

We aren't there yet!

▲ ahgwijjim and gangdoenjang hobakipssam (from left). / Visual media reporters yigyeongmin kmin@chosun.com

In the evening, a cup of soju haemuljjim enjoy together, it is ahgwijjim. Crunchy bean sprouts and parsley, Styela clava toktok popping, flesh-year-old angler dotomhan tossed two sisters, grandma's homemade progress to the tremendous flavor. Agencies also direct fermentation soak for dessert. Sweet and rich, cool. The province is not meant to taste and a big shame assumptions are made to a home.

We aren't there yet!

- We still need:
  - Better models of translation
  - Based on linguistic insights
  - Better approximations
  - Better algorithms

Research in both ASR and MT continues. The statistical approach is clearly dominant. The knowledge of linguists is added wherever it fits. And although we have made significant progress, we are very far from solving the problems.

Fred Jelinek

18 November 1932 — 14 September 2010