

Extending Grammars with Features

CSC485

Lecture 18

Announcements

- A2 Extension!
 - **3-day extension** and is now due on **Sunday, November 10, at 5 pm.**
- Essay 4 has been posted:
https://www.cs.toronto.edu/~niu/teaching/csc485/essay_4.pdf

Agreement and Inflection

- Problem: Agreement phenomena.

Nadia {washes/***wash**} the dog.

The boys {***washes**/wash} the dog.

You {***washes**/wash} the dog.

- **Morphological inflection** of verb must match subject noun in person and number.

Subject-Verb Agreement

Present tense

	Singular	Plural
1	<i>I wash</i>	<i>we wash</i>
2	<i>you wash</i>	<i>you wash</i>
3	<i>he/she/it washes</i>	<i>they wash</i>
1	<i>I am</i>	<i>we are</i>
2	<i>you are</i>	<i>you are</i>
3	<i>he, she, it is</i>	<i>they are</i>

Subject-Verb Agreement

Past tense

	Singular	Plural
1	<i>I washed</i>	<i>we washed</i>
2	<i>you washed</i>	<i>you washed</i>
3	<i>he, she, it washed</i>	<i>they washed</i>
1	<i>I was</i>	<i>we were</i>
2	<i>you were</i>	<i>you were</i>
3	<i>he, she, it was</i>	<i>they were</i>

Determiner-Noun Agreement

singular

this cat

that cat

a cat

every cat

plural

these cats

those cats

many cats

five cats

Number in Inuktitut

- Some languages have more than two grammatical numbers.
- E.g. Inuktitut has three: singular, dual, and plural.



matu	<i>door</i>
matu uk	<i>doors (two)</i>
matu it	<i>doors (three or more)</i>
nuvuja	<i>cloud</i>
nuvuja ak	<i>clouds (two)</i>
nuvuja it	<i>clouds (three or more)</i>
inuk	<i>person</i>
Inu uk	<i>people (two)</i>
inui t	<i>people (three or more)</i>
qajaq	<i>kayak</i>
qaja ak	<i>kayaks (two)</i>
qaja t	<i>kayaks (three or more)</i>
lqaluk	<i>fish</i>
lqalu uk	<i>fish (two)</i>
lqalu it	<i>fish (three or more)</i>

Agreement Features

- English agreement rules are fairly simple.
 - Subject : verb w.r.t. person and number.
 - No agreement required between verb and object.
- Many languages have other agreements.
 - E.g., German: Article and adjective ending depends on noun gender and case:

Agreement features

Nominative Case (Subject Case)

Masculine	Feminine	Neuter	Plural
der	die	das	die

der neu**e** Wagen

die schön**e** Stadt

das alt**e** Auto

die neu**en** Bücher

the new car

the beautiful city

the old car

the new books

Masculine	Feminine	Neuter	Plural
ein	eine	ein	keine

ein neu**er** Wagen

eine schön**e** Stadt

ein alt**es** Auto

keine neu**en** Bücher

a new car

a beautiful city

an old car

no new books

Agreement Features

- E.g., Chinese: Numeral classifiers, often based on shape, aggregation, ...:

两条鱼	liang tiao yu	‘two classif-long-ropelike fish’
两条河	liang tiao he	‘two classif-long-ropelike rivers’
两条腿	liang tiao tui	‘two classif-long-ropelike legs’
两条裤子	liang tiao kuzi	‘two classif-long-ropelike pants’
两只胳膊	liang zhi gebo	‘two classif-general arms’
两件上衣	liang jian shangyi	‘two classif-clothes-above-waist tops’
两套西装	liang tao xizhuang	‘two classif-set suits’

两条人命	liang tiao renming	‘two classif-long-ropelike human lives’
两条狗	liang tiao gou	‘two classif-long-ropelike dogs’



Agreement Features

- English agreement rules are fairly simple.
- Many languages have other agreements.
- Some languages have multiple grammatical genders.
 - E.g. Chichewa has genders for men, women, bridges, houses, diminutives, men inside houses, etc. Between 12-18 in total.
- Some languages overtly realize many of these distinctions.
 - E.g. some Hungarian verbs have as many as 4096 inflected forms.

	<i>Present</i>		<i>Past</i>		<i>Future</i>	
	<i>indefinite</i>	<i>definite</i>	<i>indefinite</i>	<i>definite</i>	<i>indefinite</i>	<i>definite</i>
1s	várok	várom	vártam	vártam	fogok várni	fogom várni
2s	vársz	várod	vártál	vártad	fogsz várni	fogod várni
3s	vár	várja	várt	várta	fog várni	fogja várni
1p	várunk	várjuk	vártunk	vártuk	fogunk várni	fogjuk várni
2p	vártok	várjátok	vártatok	vártátok	fogtok várni	fogjátok várni
3p	várnak	várják	vártak	várták	fognak várni	fogják várni

black: stem; blue: personal endings; red: past tense marker.

kutyába, kutyára, kutyához, kutyából, kutyáról, kutyától,
kutyáig, kutyaként, kutyául, kutyaképpen, kutyástul,
kutyásan, kutyailag, kutyaé, kutyáébb, kutyáét,
kutyáénak, kutyáéval, kutyáéért, kutyáéban, kutyáén,
kutyáénál, kutyáéba, kutyáéra, kutyáéhoz, kutyáéból,
kutyáéről, kutyáétól, kutyáéig, kutyáéként, kutyáéképpen,
kutyáéstul, kutyáésan, kutyáésilag, kutyás, kutyásabb, ...

Why???



Our
Father
in heaven.



utyásnak, kutyással, kutyásért, kutyásé, kutyásban, kutyáson, kutyásnál, kutyásba, kutyásra,
utyáshoz, kutyásból, kutyásról, kutyástól, kutyásig, kutyásként, kutyásul, kutyásképpen,
utyásilag, kutyásé, kutyásébb, kutyásénak, kutyáséval, kutyáséban, kutyásén, kutyásénál,
utyáséba, kutyáséra, kutyáséhez, kutyáséból, kutyáséről, kutyásétől, kutyáséig, kutyáséként,
utyáséképpen, kutyáséilag, kutyátlan, kutyátlanabb, kutyátlannak, kutyátlannal, kutyátlanért,
utyátlanban, kutyátlanon, kutyátlannál, kutyátlanba, kutyátlanra, kutyátlanhoz, kutyátlanból,
utyátlanról, kutyátlantól, kutyátlanig, kutyátlanként, kutyátlanul, kutyátlanképpen, kutyátlanilag,
utyátlané, kutyátlanébb, kutyátlanénak, kutyátlanéval, kutyátlanéért, kutyátlanéban, kutyátlanén,
utyátlanénál, kutyátlanéba, kutyátlanéra, kutyátlanéhez, kutyátlanéból, kutyátlanéről, kutyátlanétől,
utyátlanéig, kutyátlanéként, kutyátlanéképpen, kutyátlanésilag, kutyásság, kutyásságibb,
utyásságnak, kutyássággal, kutyásságért, kutyásságban, kutyásságon, kutyásságnál, kutyásságba,
utyásságra, kutyássághoz, kutyásságból, kutyásságról, kutyásságtól, kutyásságig, kutyásságként,
utyásságul, kutyásságképpen, kutyásságilag, kutyássága, kutyásságább, kutyásságának,
utyásságával, kutyásságáért, kutyásságában, kutyásságán, kutyásságánál, kutyásságába,
utyásságára, kutyásságához, kutyásságából, kutyásságáról, kutyásságától, kutyásságáig,
utyásságaként, kutyásságaul, kutyásságaképpen, kutyásságáilag, kutyásságéa, kutyásságébb,
utyásságáénak, kutyásságáéval, kutyásságáéért, kutyásságáéban, kutyásságáén, kutyásságáénál,
utyásságáéba, kutyásságáéra, kutyásságáéhez, kutyásságáéból, kutyásságáéről, kutyásságáétől,
utyásságáéig, kutyásságáéként, kutyásságáéképpen, kutyásságáéilag, kutyásságé, kutyásságébb,
utyásságénak, kutyásságéval, kutyásságéért, kutyásságéban, kutyásságén, kutyásságénál,
utyásságéba, kutyásságéra, kutyásságéhez, kutyásságéból, kutyásságéről, kutyásságétől,
utyásságéig, kutyásságéként, kutyásságéképpen, kutyásságéilag, kutyátlanság, kutyátlanságibb,
utyátlanságnak, kutyátlansággal, kutyátlanságért, kutyátlanságban, kutyátlanságon,
utyátlanságnál, kutyátlanságba, kutyátlanságra, kutyátlansághoz, kutyátlanságból,
utyátlanságról, kutyátlanságtól, kutyátlanságig, kutyátlanságként, kutyátlanságul,
utyátlanságképpen, kutyátlanságilag, kutyátlansága, kutyátlanságább, kutyátlanságának,
utyátlanságával, kutyátlanságáért, kutyátlanságában, kutyátlanságán, kutyátlanságánál,
utyátlanságába, kutyátlanságára, kutyátlanságához, kutyátlanságából, kutyátlanságáról,



utyátlanságától, kutyátlanságáig, kutyátlanságaként, kutyátlanságául, kutyátlanságaképpen,
utyátlanságáilag, kutyátlanságáé, kutyátlanságáébb, kutyátlanságáénak, kutyátlanságáéval,
utyátlanságáében, kutyátlanságáén, kutyátlanságáénál, kutyátlanságáéba, kutyátlanságáéra,
utyátlanságáéhoz, kutyátlanságáéból, kutyátlanságáéről, kutyátlanságáétől, kutyátlanságáéig,
utyátlanságaéként, kutyátlanságaéképpen, kutyátlanságáéilag, kutyátlanságé, kutyátlanságébb,
utyátlanságénak, kutyátlanságéval, kutyátlanságéért, kutyátlanságében, kutyátlanságén,
utyátlanságénál, kutyátlanságéba, kutyátlanságéra, kutyátlanságéhoz, kutyátlanságéből,
utyátlanságéről, kutyátlanságétől, kutyátlanságéig, kutyátlanságéként, kutyátlanságéilag,
kutyátlanodott, elkutyátlanodottabb, elkutyátlanodottnak, elkutyátlanodottal, elkutyátlanodottért,
kutyátlanodottban, elkutyátlanodotton, elkutyátlanodottnál, elkutyátlanodottba, elkutyátlanodottra,
kutyátlanodotthoz, elkutyátlanodottból, elkutyátlanodottról, elkutyátlanodotttól, elkutyátlanodottig,
kutyátlanodottként, elkutyátlanodottul, elkutyátlanodottképpen, elkutyátlanodottilag,
kutyátlanodotté, elkutyátlanodottébb, elkutyátlanodotténak, elkutyátlanodottéval,
kutyátlanodottéért, elkutyátlanodottében, elkutyátlanodottén, elkutyátlanodotténál,
kutyátlanodottéba, elkutyátlanodottéra, elkutyátlanodottéhoz, elkutyátlanodottéből,
kutyátlanodottéről, elkutyátlanodottétől, elkutyátlanodottéig, elkutyátlanodottéul,
kutyátlanodottéképpen, elkutyátlanodottéilag, elkutyátlanodottság, elkutyátlanodottságabb,
kutyátlanodottságnak, elkutyátlanodottsággal, elkutyátlanodottságért, elkutyátlanodottságban,
kutyátlanodottságon, elkutyátlanodottságnál, elkutyátlanodottságba, elkutyátlanodottságra,
kutyátlanodottsághoz, elkutyátlanodottságból, elkutyátlanodottságról, elkutyátlanodottságtól,
kutyátlanodottságig, elkutyátlanodottságként, elkutyátlanodottságul, elkutyátlanodottságképpen,
kutyátlanodottságilag, elkutyátlanodottsága, elkutyátlanodottságabb, elkutyátlanodottságának,
kutyátlanodottságával, elkutyátlanodottságáért, elkutyátlanodottságában, elkutyátlanodottságán,
kutyátlanodottságánál, elkutyátlanodottságába, elkutyátlanodottságára, elkutyátlanodottságához,



kutyátlanodottságából, elkutyátlanodottságáról, elkutyátlanodottságától, elkutyátlanodottságáig,
kutyátlanodottságaként, elkutyátlanodottságaul, elkutyátlanodottságaképpen, elkutyátlanodottságáilag
kutyátlanodottságáé, elkutyátlanodottságáébb, elkutyátlanodottságáénak, elkutyátlanodottságáéval,
kutyátlanodottságáéért, elkutyátlanodottságáében, elkutyátlanodottságáén, elkutyátlanodottságáénál,
kutyátlanodottságáéba, elkutyátlanodottságáéra, elkutyátlanodottságáéhoz, elkutyátlanodottságáéból,
kutyátlanodottságáéről, elkutyátlanodottságáétól, elkutyátlanodottságáéig, elkutyátlanodottságaéként,
kutyátlanodottságáéilag, megkutyásodott, megkutyásodottabb, megkutyásodottnak, megkutyásodottal
megkutyásodottért, megkutyásodottban, megkutyásodotton, megkutyásodottnál, megkutyásodottba,
megkutyásodottra, megkutyásodotthoz, megkutyásodottból, megkutyásodottról, megkutyásodottól,
megkutyásodottig, megkutyásodottként, megkutyásodottul, megkutyásodottképpen, megkutyásodottilag
megkutyásodotté, megkutyásodottébb, megkutyásodotténak, megkutyásodottéval, megkutyásodottéért,
megkutyásodottságabb, megkutyásodottságnak, megkutyásodottsággal, megkutyásodottságért,
megkutyásodottságban, megkutyásodottságon, megkutyásodottságnál, megkutyásodottságba,
megkutyásodottságra, megkutyásodottságához, megkutyásodottságból, megkutyásodottságról,
megkutyásodottságtól, megkutyásodottságig, megkutyásodottságként, megkutyásodottságul,
megkutyásodottságképpen, megkutyásodottságilag, megkutyásodottsága, megkutyásodottságabb,
megkutyásodottságának, megkutyásodottságával, megkutyásodottságáért, megkutyásodottságában,
megkutyásodottságán, megkutyásodottságánál, megkutyásodottságába, megkutyásodottságára,
megkutyásodottságához, megkutyásodottságából, megkutyásodottságáról, megkutyásodottságától,
megkutyásodottságáig, megkutyásodottságaként, megkutyásodottságaul, megkutyásodottságaképpen,
megkutyásodottságáilag, megkutyásodottságáé, megkutyásodottságáébb, megkutyásodottságáénak,
megkutyásodottságáéval, megkutyásodottságáéért, megkutyásodottságáében, megkutyásodottságáén,
megkutyásodottságáénál, megkutyásodottságáéba, megkutyásodottságáéra,
megkutyásodottságáéhoz, megkutyásodottságáéból, megkutyásodottságáéről,
megkutyásodottságáétól, megkutyásodottságáéig, megkutyásodottságaéként, ...





Inflectional morphology

- Word may be inflected ...
 - ... to indicate paradigmatic properties, e.g. singular / plural, past / present, ...
 - ... to indicate some (other) semantic properties
 - ... to agree with inflection of other words.
- Each (open-class) word-type has a **base form** / **stem** / **lemma**.
- Each occurrence of a word includes inflection by a (possibly null) morphological change.

CFG Grammar

S → NP VP

NP → Det N

NP → Det Adj N

NP → NP PP

VP → V

VP → V NP

PP → P NP

Det → *the* | *a* | *an*

Adj → *old* | *red* | *happy* | ...

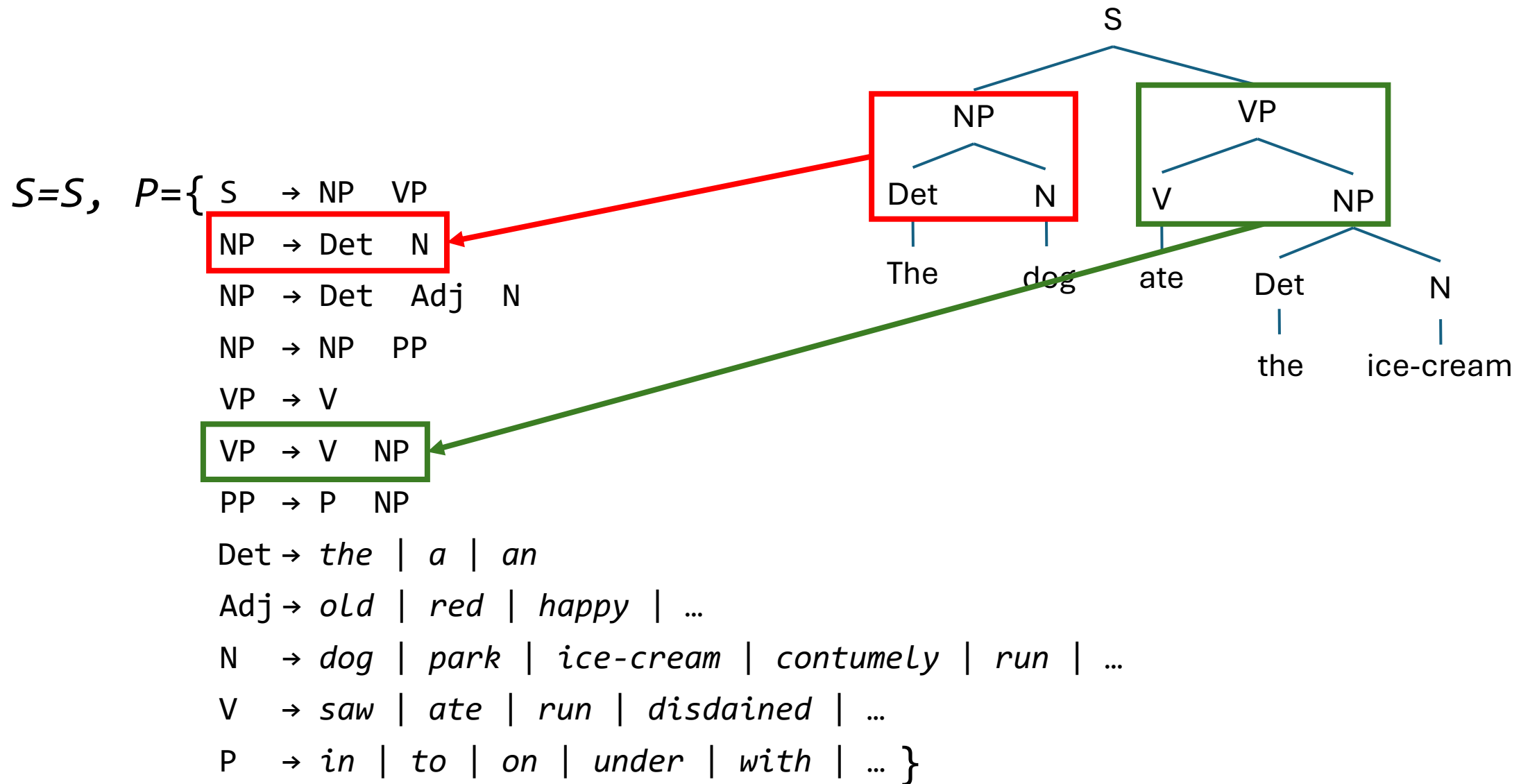
N → *dog* | *park* | *ice-cream* | *contumely* | *run* | ...

V → *saw* | *ate* | *run* | *disdained* | ...

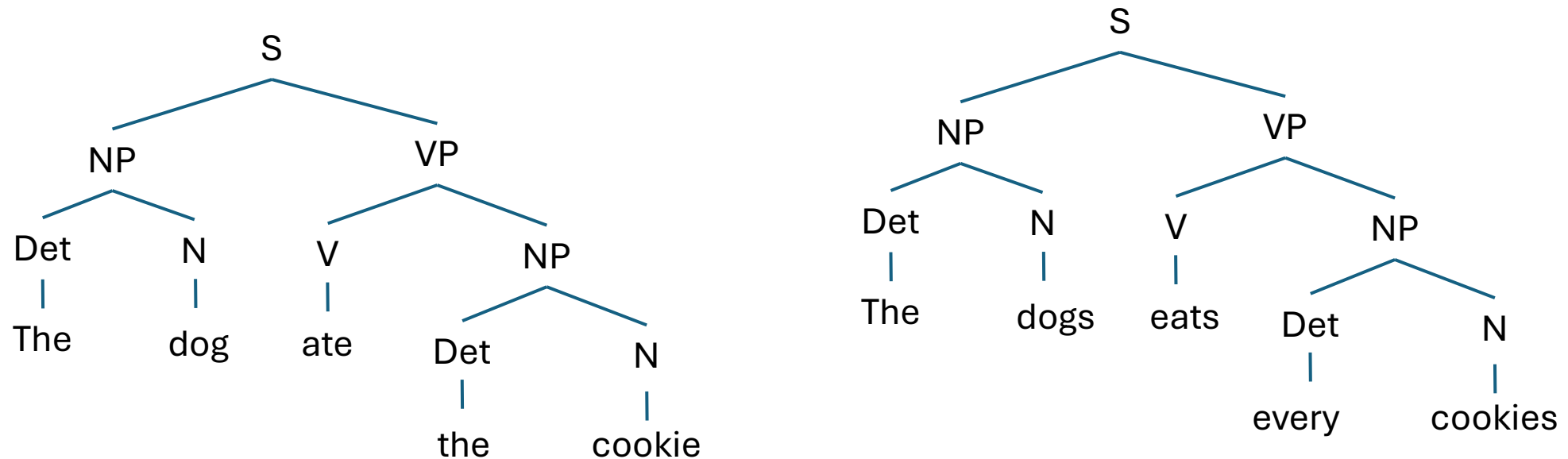
P → *in* | *to* | *on* | *under* | *with* | ... }

Lexical categories:

NT's that rewrite
as a single T.



How do we enforce agreement?



Rule proliferation

- **Problem:** How to account for this in grammar.
- **Possible solution:** Replace all NPs, Vs, and VPs throughout the grammar.

$S \rightarrow NP \quad VP$

$VP \rightarrow V$

$VP \rightarrow V \quad NP$

$NP \rightarrow \text{you} \mid \text{me} \mid \textit{dog} \mid \text{dogs} \mid \text{bear} \mid \text{bears} \dots$

$V \rightarrow \textit{washes} \mid \textit{wash} \mid \textit{washed} \mid \textit{is} \mid \text{was} \dots$

S → NP3s VP3s

S → NP3p VP3p

S → NP2 VP2

S → NP1s VP1s

S → NP1p VP1p

NP3s → dog | bear ...

NP3p → dogs | bears ...

NP2 → you

NP1s → me

...

V3s → *is* | *was* | *washes* | *washed* ...

V3p → are | were | wash | washed ...

V1s → am | was | wash | washed ...

...

Rule proliferation

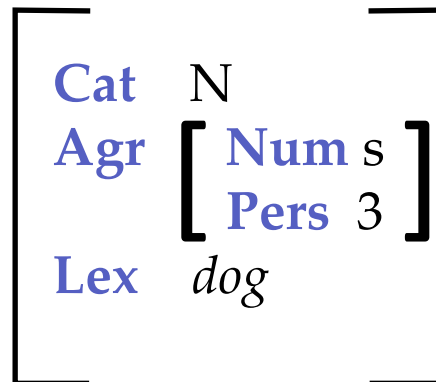
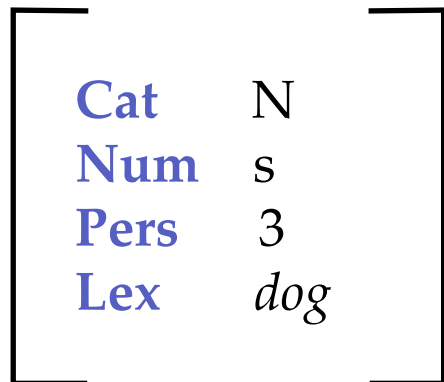
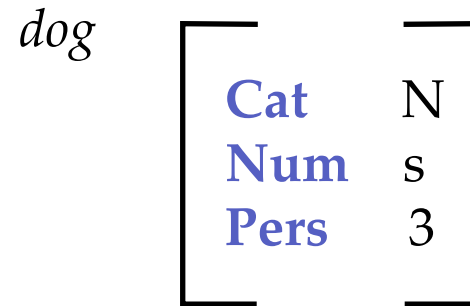
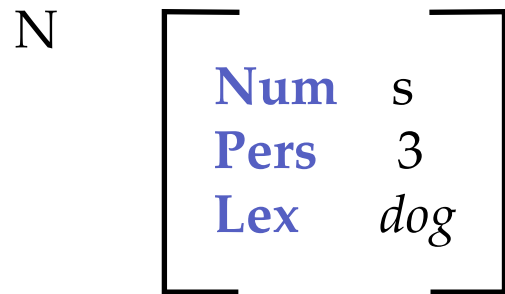
- **Drawback 1:** the result is big ... really big.
- **Drawback 2:** Losing the generalization:
 - All these Ss, NPs, VPs have the same structure.
 - Doesn't depend on particular verb, noun, and number.
- CF rules collapse together structural and featural information.
- All information must be completely and directly specified.
 - E.g., can't just say that values must be equal for some feature without saying exactly what values.

Feature structures

- Solution: Separate feature information from syntactic, structural, and lexical information.
- A **feature structure** is a list of pairs:
 [feature-name feature-value]
- Feature-values may be atoms or feature structures.
- Can consider syntactic category or word to be bundle of features too.
- Can represent syntactic structure.

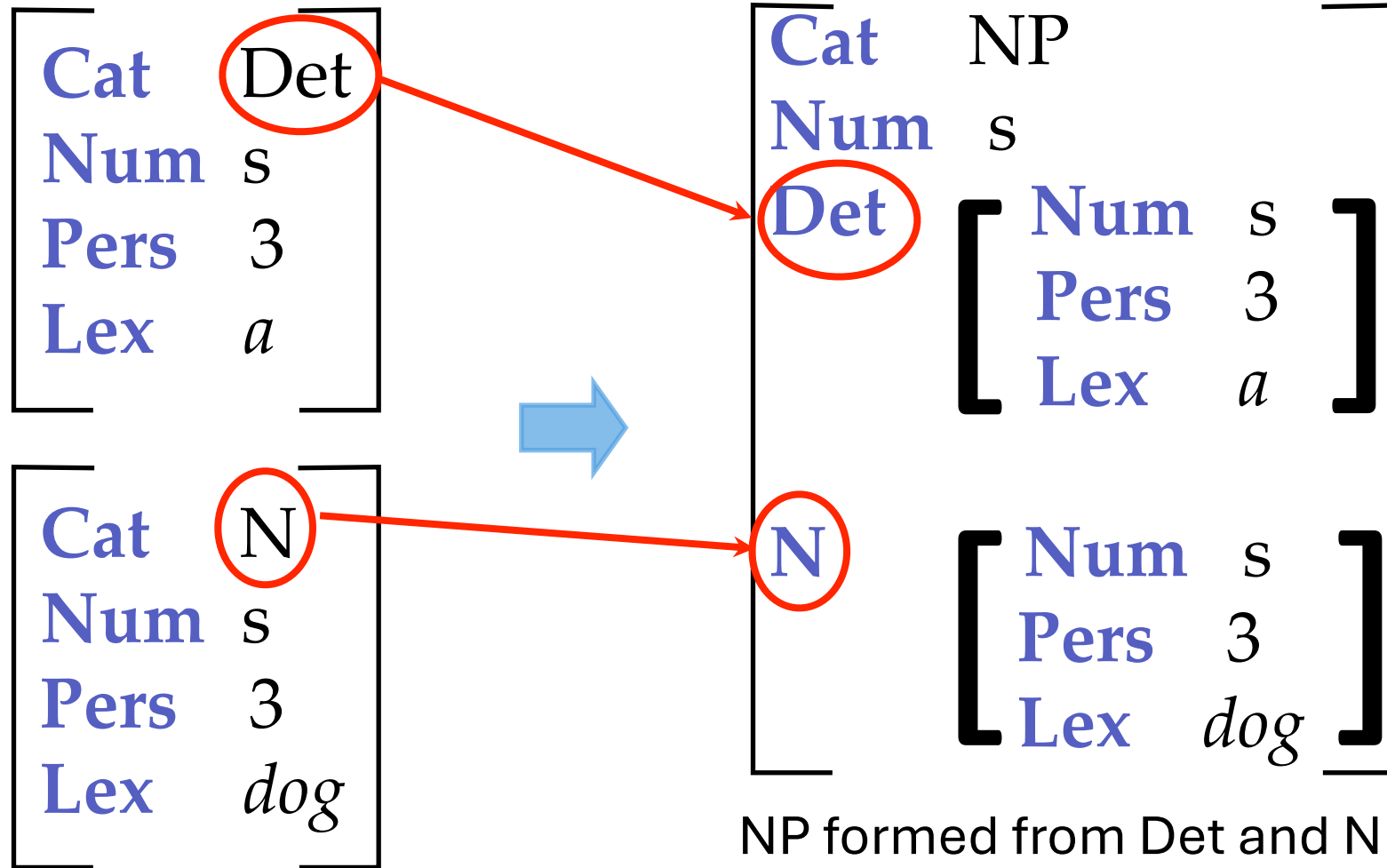
Feature structures

- *Drawback*: many equivalent notations.



Feature paths:
features of
features; e.g.,
(Agr Pers 3)

Feature structures



NP formed from Det and N.
Feature values in components become
feature names in new constituent.

Components of feature use

- **1. Lexical specification:**

Description of **properties** of a word:
morphological, syntactic, semantic, ...

dog: $\begin{bmatrix} \text{Cat} & \text{N} \\ \text{Agr} & 3\text{s} \end{bmatrix}$

sleeps: $\begin{bmatrix} \text{Cat} & \text{V} \\ \text{Agr} & 3\text{s} \end{bmatrix}$

dogs: $\begin{bmatrix} \text{Cat} & \text{N} \\ \text{Agr} & 3\text{p} \end{bmatrix}$

sleep: $\begin{bmatrix} \text{Cat} & \text{V} \\ \text{Agr} & \{1\text{s}, 2\text{s}, 1\text{p}, 2\text{p}, 3\text{p}\} \end{bmatrix}$

Or: $\text{N} \rightarrow \text{dog}$
 $(\text{N } \text{Agr}) = 3\text{s}$

$\text{V} \rightarrow \text{sleeps}$
 $(\text{V } \text{Agr}) = 3\text{s}$

$\text{N} \rightarrow \text{dogs}$
 $(\text{N } \text{Agr}) = 3\text{p}$

$\text{V} \rightarrow \text{sleep}$
 $(\text{V } \text{Agr}) = \{1\text{s}, 2\text{s}, 1\text{p}, 2\text{p}, 3\text{p}\}$

Components of feature use

- **2. Agreement:**

- Constraints on co-occurrence in a rule — within or across phrases.
- Typically are equational constraints.

$$\begin{aligned} \text{NP} &\rightarrow \text{Det } \text{N} \\ (\text{Det } \text{Num}) &= (\text{N } \text{Num}) \end{aligned}$$

$$\begin{aligned} \text{S} &\rightarrow \text{NP } \text{VP} \\ (\text{NP } \text{Agr}) &= (\text{VP } \text{Agr}) \end{aligned}$$

Components of feature use

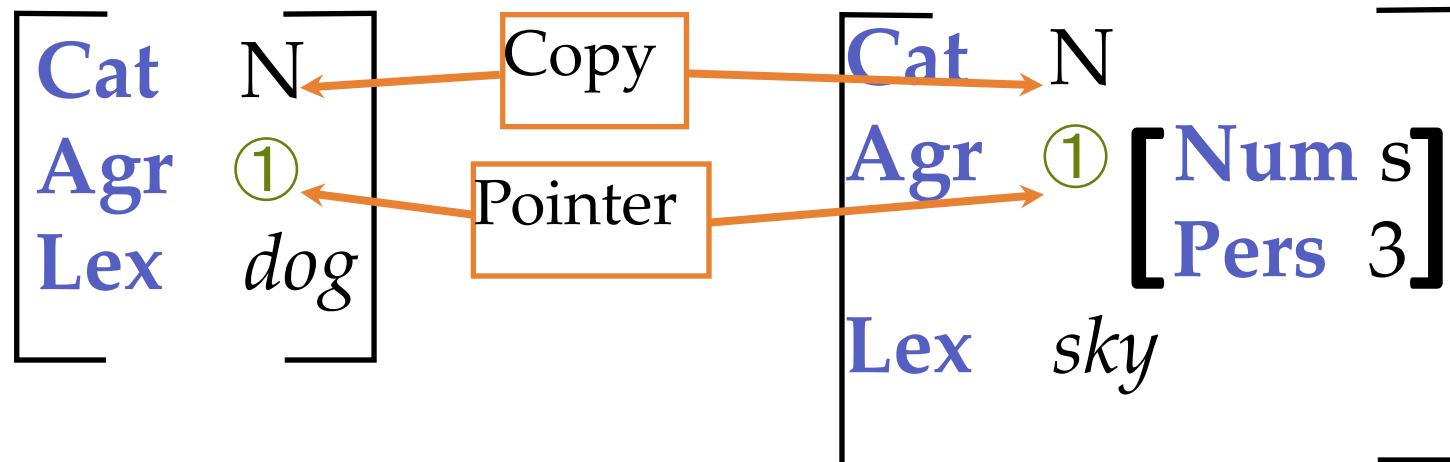
- 3. Projection:
 - **Sharing of features** between the head of a phrase and the phrase itself.

$$\begin{aligned} \text{VP} &\rightarrow \text{V} \dots \\ (\text{VP } \text{Agr}) &= (\text{V } \text{Agr}) \end{aligned}$$

- Head features:
 - **Agr** is typical, but so is the head-word itself as a feature.
(Common enough that there's usually a mechanism for "declaring" head features and omitting them from rules.)

Constraints on feature values

- What does it mean for two features to be “equal”?
 - A copy of the value or feature structure, or a pointer to the same value or feature structure (re-entrancy, shared feature paths).



Constraints on feature values

- But: It may be sufficient that two features are not equal, just compatible — that they can be unified.

• E.g.,

Cat	N
Pers	3
Num	s

 and

Cat	N
Pers	3
Gndr	F

TRALE: Parsing with Feature

- We have such an engine for parsing with feature: TRALE
- TRALE Tutorial: <https://www.cs.toronto.edu/~niu/csc485/trale/>
- More information in A3 tutorials.

Quiz

Why does rule proliferation pose a challenge in context-free grammar parsing?

- A. It makes grammar too restrictive
- B. It loses structural generalizations
- C. It requires additional lexical categories
- D. It simplifies sentence parsing

Subsumption of feature structures

- Feature structure X **subsumes** feature structure Y if Y is consistent with, and at least as specific as X .
 - Also say that Y extends X .
 Y can add (non-contradictory) features to those in X .
- Definition: X subsumes Y ($X \sqsubseteq Y$) iff there is a simulation of X inside Y , i.e., a function s.t.:
 - $\text{sim}(X) = Y$
 - If X is atomic, so is Y and $X = Y$
 - Otherwise, for all feature values $X.f$: $Y.f$ is defined, and sim simulates $X.f$ inside $Y.f$.

Subsumption of feature structures

- Examples:

$$\begin{bmatrix} \text{Cat} & \text{N} \\ \text{Pers} & 3 \end{bmatrix} \sqsubseteq \begin{bmatrix} \text{Cat} & \text{N} \\ \text{Pers} & 3 \\ \text{Gndr} & \text{F} \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} \text{Cat} & \text{N} \\ \text{Pers} & 3 \\ \text{Num} & \text{s} \end{bmatrix} \not\sqsubseteq \begin{bmatrix} \text{Cat} & \text{N} \\ \text{Pers} & 3 \\ \text{Gndr} & \text{F} \end{bmatrix}$$

$$\begin{bmatrix} \text{Cat} & \text{VP} \\ \text{Agr} & \textcircled{1} \\ \text{Subj} & [\text{Agr} \textcircled{1}] \end{bmatrix} \sqsubseteq \begin{bmatrix} \text{Cat} & \text{VP} \\ \text{Agr} & \textcircled{1} \\ \text{Subj} & \text{Agr} \textcircled{1} \left[\begin{bmatrix} \text{Pers} & 3 \\ \text{Num} & \text{s} \end{bmatrix} \right] \end{bmatrix}$$

Unification

- The *unification* of X and Y ($X \sqcup Y$) is the most general feature structure Z that is subsumed by both X and Y .
 - Z is the smallest feature structure that extends both X and Y .
- Unification is a constructive operation.
 - If any feature values in X and Y are incompatible, it fails.
 - Else it produces a feature structure that includes all the features in X and all the features in Y .

Unification

$$\begin{bmatrix} \text{Cat} & \text{N} \\ \text{Pers} & 3 \\ \text{Num} & \text{s} \end{bmatrix} \sqcup \begin{bmatrix} \text{Cat} & \text{N} \\ \text{Pers} & 3 \\ \text{Gndr} & \text{F} \end{bmatrix} = \begin{bmatrix} \text{Cat} & \text{N} \\ \text{Pers} & 3 \\ \text{Num} & \text{s} \\ \text{Gndr} & \text{F} \end{bmatrix}$$

Sample grammar fragment

$S \rightarrow NP VP$

$(NP \text{ Agr}) = (VP \text{ Agr})$

$NP \rightarrow Det N$

$(NP \text{ Agr}) = (N \text{ Agr})$

$(Det \text{ Agr}) = (N \text{ Agr})$

$VP \rightarrow V$

$(VP \text{ Agr}) = (V \text{ Agr})$

$Det \rightarrow a$

$[\text{Agr } 3s]$

$N \rightarrow dog$

$[\text{Agr } 3s]$

$V \rightarrow sleep$

$[\text{Agr } ^3s]$

$Det \rightarrow all$

$[\text{Agr } 3p]$

$N \rightarrow dogs$

$[\text{Agr } 3p]$

$V \rightarrow sleeps$

$[\text{Agr } 3s]$

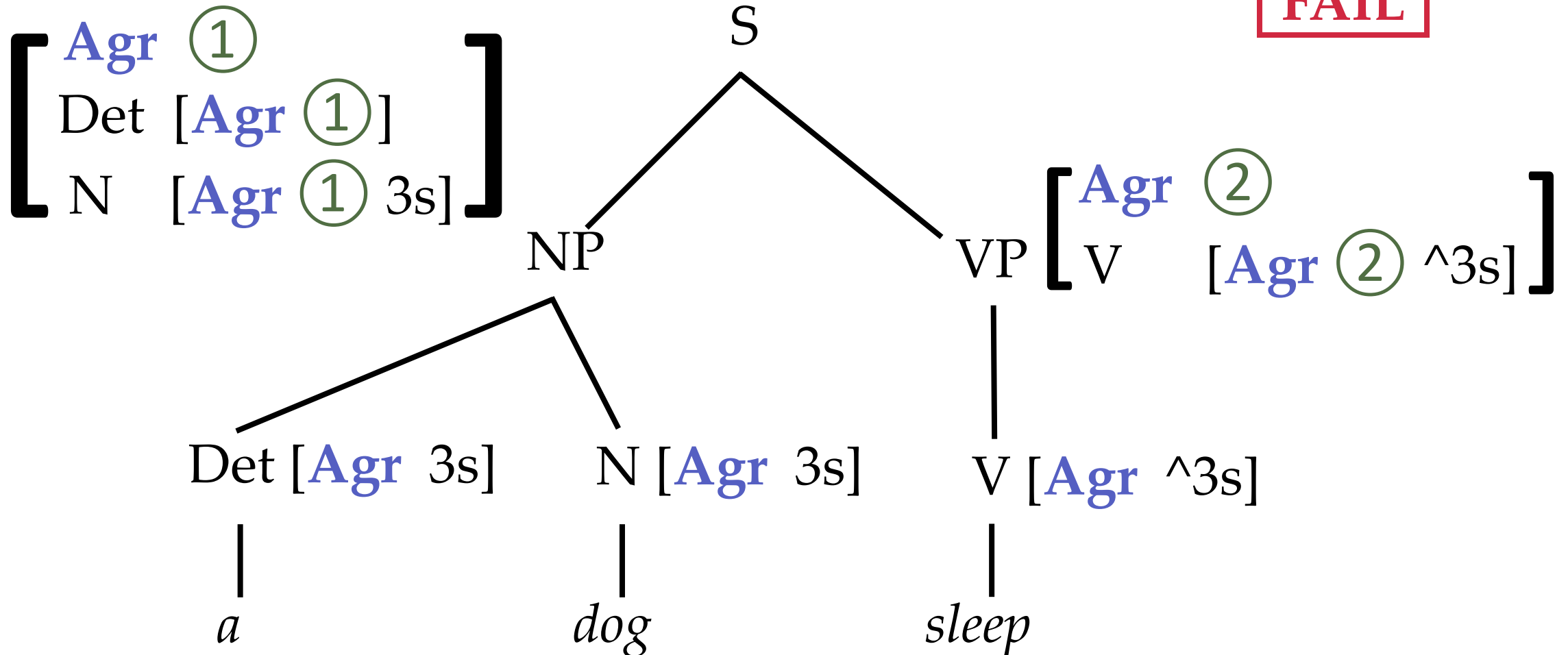
$Det \rightarrow the$

$[\text{Agr } \{3s, 3p\}]$

Mismatched features fail

$[Agr \textcircled{1}] \sqcup [Agr \textcircled{2}]$

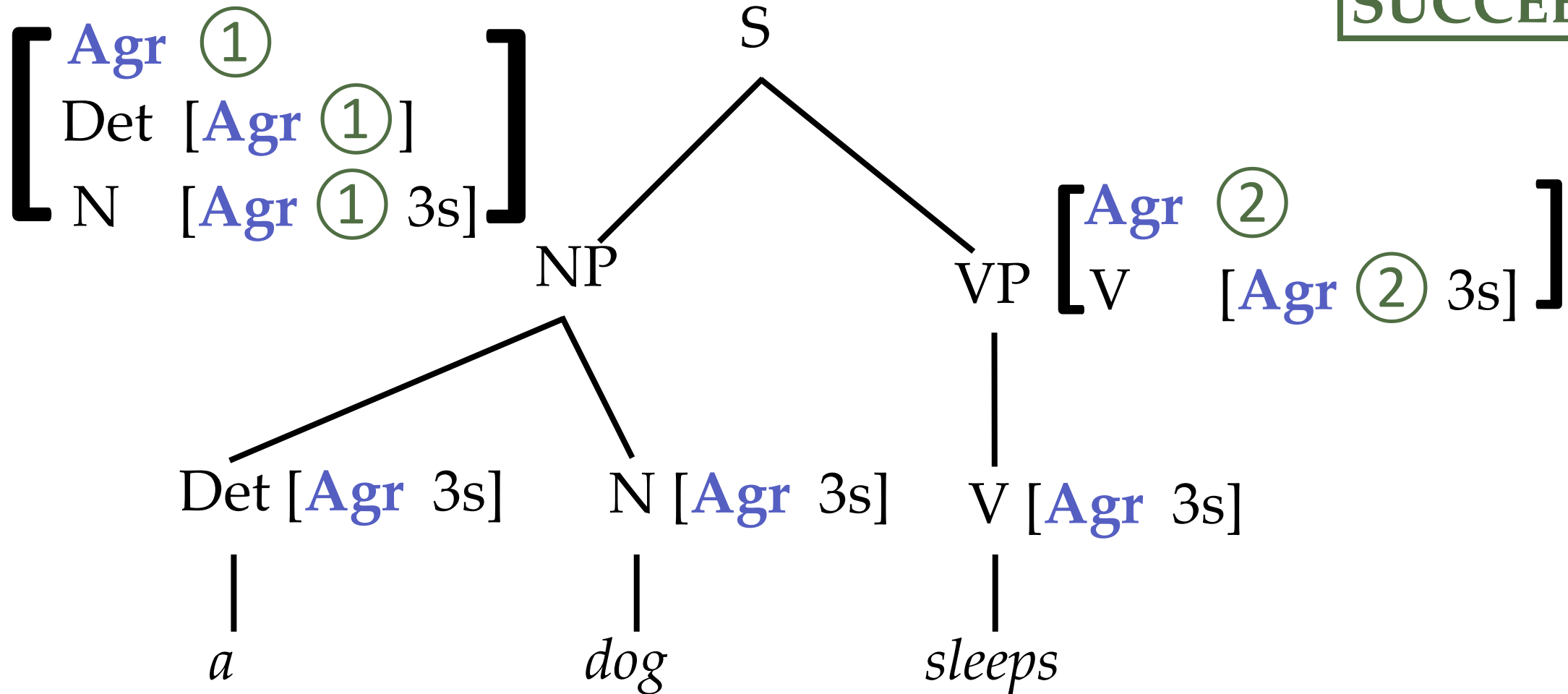
FAIL



Mismatched features fail

$[Agr \textcircled{1}] \sqcup [Agr \textcircled{2}]$

SUCCEED



Advantages of this approach

- Distinguishes structure from “functional” info.
- Allows for economy of specification:
 - Equations in rules:
 $S \rightarrow NP \ VP$
 $(NP \text{ Agr}) = (VP \text{ Agr})$
 - Sets of values in lexicon:
 $N \rightarrow \text{fish}$
 $(N \text{ Agr} \{3s, 3p\})$
- Allows for indirect specification and transfer of information, e.g., head features.

Must unify with



Features and the lexicon

- Lexicon may contain each inflected form.
 - Feature values and base form listed.
- Lexicon may contain only base forms.
 - Process of morphological analysis maps inflected form to base form plus feature values.
 - Time–space trade-off, varies by language.
- Lexicon may contain ***semantics*** for each form.

Morphological analysis

- Morphological analysis is simple in English.
 - Reverse the rules for inflections, including spelling changes.

dogs → *dog* [**Agr** 3p]

dog → *dog* [**Agr** 3s]

berries → *berry* [**Agr** 3p]

buses → *bus* [**Agr** 3p]

eats → *eat* [**Agr** 3s, **Tns** pres]

ripped → *rip* [**Tns** past]

tarried → *tarry* [**Tns** past]

running → *run* [**Tns** pp]

- Irregular forms will always have to be explicitly listed in lexicon.

children → *child* [**Agr** 3p] *sang* → *sing* [**Tns** past]

Morphology in other languages

- Rules may be more complex in other (even European) languages.
- Languages with compounding (e.g., German) or agglutination (e.g., Finnish) require more-sophisticated methods.
 - E.g., *Verdauungsspaziergang*, a stroll that one takes after a meal to assist in digestion.

English auxiliary system

This kale

must

have

been

being

eaten

modal

perfect

progressive

passive

main verb



“I would never do such a thing unless
You were already having been going to do that.”

Semantics as a lexical feature

- Add a Sem feature:

Cat	N
Num	s
Pers	3
Lex	<i>dog</i>
Sem	dog

Typewriter font
for semantic objects



- The meaning of dog is dog.
- The meaning of *chien* and *Hund* are both dog.
- The meaning of dog is G52790.
- Not uncommon, but pretty unimpressive ...

Verb subcategorization

- **Problem:** Constraints on verbs and their complements.

*Nadia **told / instructed / *said / *informed** Ross to sit down.*

*Nadia ***told / *instructed / said / *informed** to sit down.*

*Nadia **told / *instructed / *said / informed** Ross of the
requirement to sit down.*

*Nadia **gave / donated** her painting to the museum.*

*Nadia **gave / *donated** the museum her painting.*

*Nadia **put / ate** the cake in the kitchen.*

*Nadia ***put / ate** the cake.*

Verb tense and aspect

- **Tense** and **aspect** markings on verb:
 - Locate the event in time (relative to another time).
 - Mark the event as complete/finished or in progress.

Nadia rides the horse.

— In progress now.

Nadia rode the horse.

— Completed before now.

Nadia had ridden the horse.

— Completed before now.

Nadia was riding the horse.

— In progress before now.

⋮

Verb tense and aspect

- **Tense:** past or present
- **Aspect:** simple, progressive, or perfect

Nadia ...

	Simple	Progressive	Perfect
Present	<i>rides</i>	<i>is riding</i>	<i>has ridden</i>
Past	<i>rode</i>	<i>was riding</i>	<i>had ridden</i>

... the horse

Verb tense and aspect

- **Tense:** past or present
- **Aspect:** simple, progressive, or perfect

Nadia ...

Nadia ...

		Auxiliary verb	
	Simple	Progressive	Perfect
Present	<i>rides</i>	<i>is riding</i>	<i>has ridden</i>
Past	<i>rode</i>	<i>was riding</i>	<i>had ridden</i>

... the horse

Verb tense and aspect

- **Tense:** past or present
- **Aspect:** simple, progressive, or perfect

Nadia ...

	Simple	Progressive	Perfect
Present	<i>rides</i>	<i>is riding</i>	<i>has ridden</i>
Past	<i>rode</i>	<i>was riding</i>	<i>had ridden</i>
		<i>... the horse</i>	

In progress

Complete

Present participle

Past participle

Verb tense and aspect

- **Tense:** past or present
- **Aspect:** simple, progressive, or perfect
Nadia ...

Auxiliary verbs

	Simple	Perfect progressive (continuous)
Present	<i>rides</i>	<i>has been riding</i>
Past	<i>rode</i>	<i>had been riding</i>

... the horse

Modal verbs

- Modal verbs: Auxiliary verbs that express degrees of certainty, obligation, possibility, prediction, etc.

Nadia

{could, should, must, ought to, might, will, ...}

{ride, be riding, have ridden, have been riding}

the horse.

English auxiliary system

- Structure (so far):

[MODAL] [HAVE] [BE] MAIN-VERB

- General pattern:

VP \rightarrow AUX VP

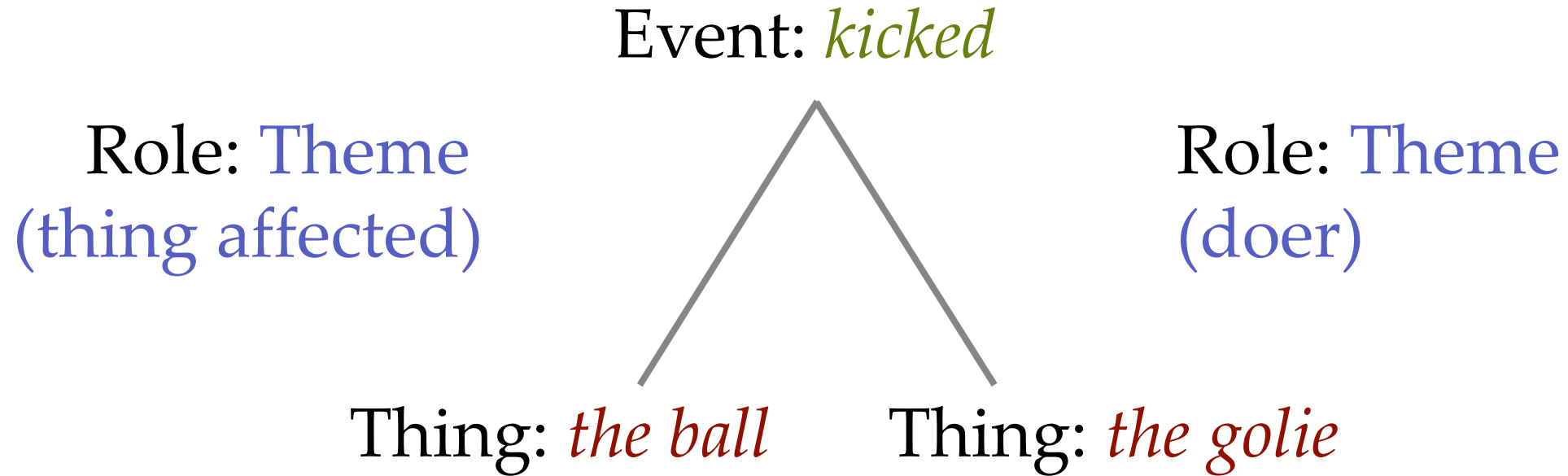
AUX \rightarrow MODAL | HAVE | BE

- Use features to capture necessary agreements.

Diathesis

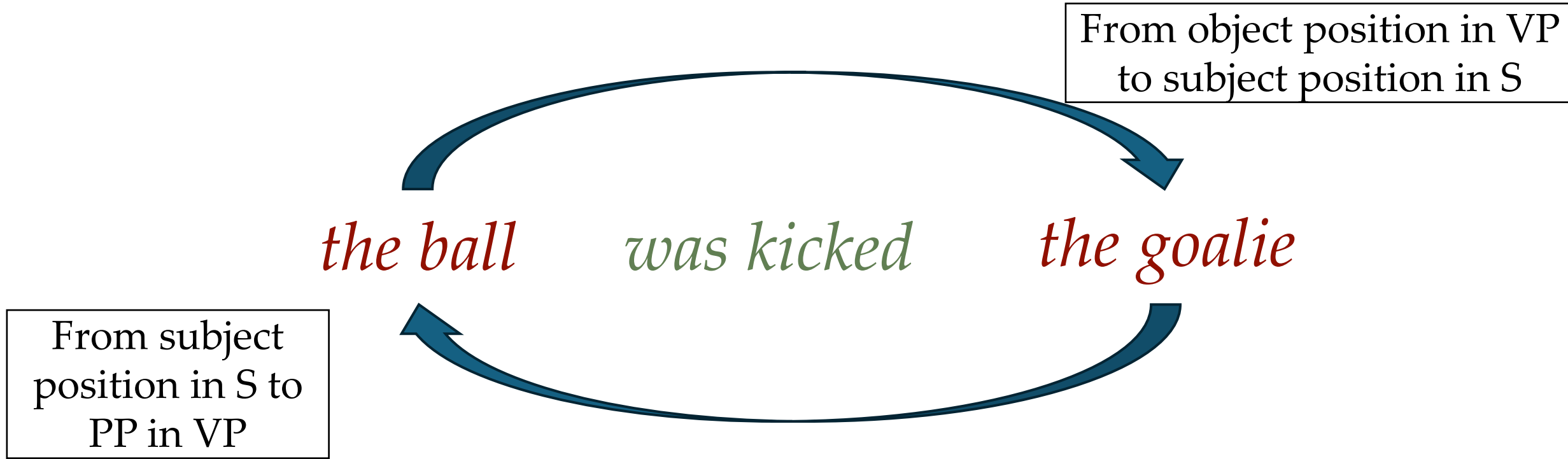
aka voice: the relationship between the verb and its arguments.

The ball was kicked by the goalie.



kick (agent=goalie, theme=ball)

Diathetic alternation



But the semantic representation doesn't change

Voice

- Voice: System of assigning thematic roles to syntactic positions.
 - English has active and passive voices.
- Passive expressed with *be*+past participle.
Other auxiliaries may also apply, including progressive *be*.
- *Nadia was kissed.* *Nadia was being kissed.*
Nadia had been kissed. *Nadia had been being kissed.*
Nadia could be kissed. *Nadia could have been being kissed.*
- Structure:
[MODAL] [HAVE] [BE1] [BE2] MAIN-VERB

Some useful features

- **VForm**: The tense/aspect form of a verb: passive, pastprt, ...
- **CompForm**: The tense/aspect form of the complement of an auxiliary.

Augmenting rules for passive voice

- For all rules of the form:

$VP \rightarrow V \ NP \ X$
(V **Subcat**) = $_y$



$VP \rightarrow V \ X$
(V **Subcat**) = $_y$
(V **VForm**) = passive
(VP **VForm**) = passive

Metarule to ease grammar coding

- Augment Aux+VP rules:

$VP \rightarrow AUX \ VP$
(AUX **Root**) = Be2
(AUX **CompForm**) = (VP₂ **VForm**)
(VP₂ **VForm**) = passive

The GAP feature for passive voice

S → NP VP

(NP **Agr**) = (VP **Agr**)

(VP **VForm**) = passive

(VP **Gap Cat**) = NP

(VP **Gap Agr**) = (NP **Agr**)

(VP **Gap Sem**) = (NP **Sem**)

VP → AUX VP

(VP₁ **Agr**) = (AUX **Agr**)

(VP₁ **VForm**) = (VP₂ **VForm**)

(VP₁ **Gap**) = (VP₂ **Gap**)

(AUX **Lex**) = be2

(VP₂ **VForm**) = passive

V → *kicked*

(V **VForm**) = {pastprt, passive}

(V **Subcat**) = _np

(V **Lex**) = *kick*

(V **Sem**) = *kick*

VP → V NP

(VP **VForm**) = (V **VForm**)

(VP **Gap**) = (NP **Gap**)

(V **Subcat**) = _np

Empty string



NP → ε

(NP **Gap Cat**) = NP

(NP **Gap Agr**) = (NP **Agr**)

(NP **Gap Sem**) = (NP **Sem**)

NP → *cans*

(NP **Agr**) = 3p

(NP **Lex**) = *can*

(NP **Sem**) = *cans*

AUX → *were*

(AUX **Agr**) = 3p

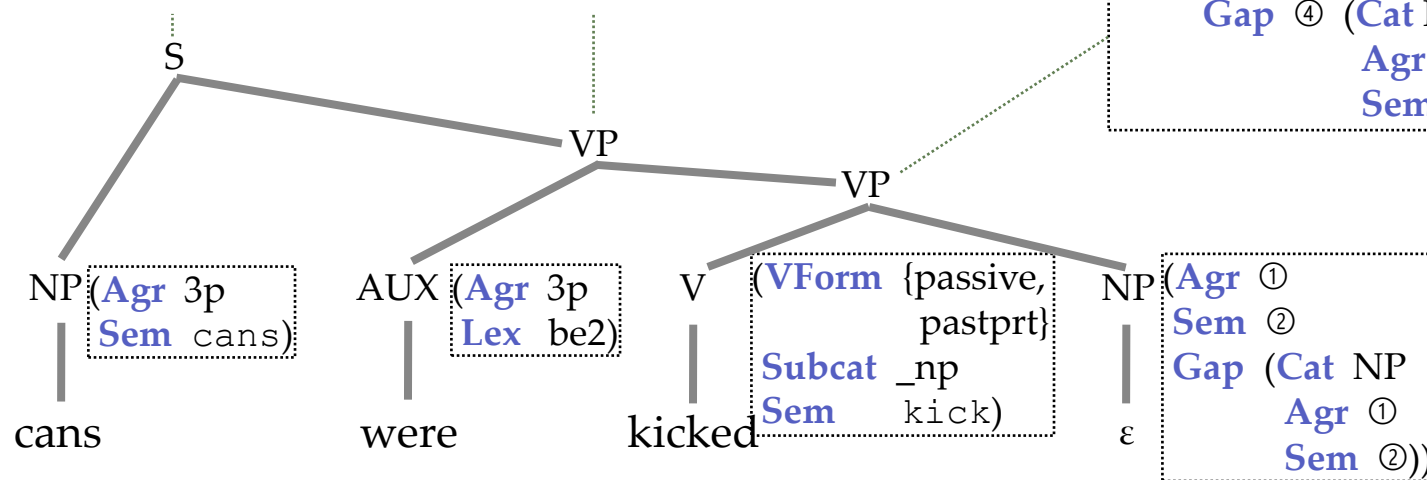
(AUX **Lex**) = be2

(NP (Agr ① 3p
Sem ② cans)
VP (Agr ①
VForm ③
Gap ④
AUX (Agr ① 3p
Lex be2)
VP (VForm ③
Gap ④
V (VForm ③ passive
Subcat _np
Sem kick)
NP (Agr ①
Sem ②
Gap ④ (Cat NP
Agr ①
Sem ②))))

(Agr ⑤
VForm ③
Gap ④
AUX (Agr ⑤ 3p
Lex be2)
VP (VForm ③
Gap ④
V (VForm ③ passive
Subcat _np
Sem kick)
NP (Agr ①
Sem ②
Gap ④ (Cat NP
Agr ①
Sem ②))))

Note: The green ①'s of the S were ⑤'s until the 4th constraint of the rule $S \rightarrow NP VP$. The 5th constraint fills in the Sem of the Gap ②.

(VForm ③
Gap ④
V (VForm ③ {passive,
pastprt}
Subcat _np
Sem kick)
NP (Agr ①
Sem ②
Gap ④ (Cat NP
Agr ①
Sem ②))))



Other cases of gap percolation

- Other constructions involve NPs in syntactic configurations where they would not get the right thematic roles using linear order alone.

Nadia seems to like Ross.

Nadia seems to be liked.

Nadia is easy to like.

Who did Nadia like?

I fed the dog that Nadia likes to walk.

- Can use grammar rules with gap features to ensure correct structure/interpretation of these as well.

Summary

- Features help capture syntactic constructions in a general and elegant grammar.
- Features can encode the compositional semantics of a sentence as you parse it.
- Features can accomplish mapping functions between syntax and semantics that simplify the interpretation process.