

Syntax

Introduction to Linguistics

Dustin A. Chacón

- **Observation:** Speakers have intuitions that phrases are well-formed / ill-formed (*)

(1) Colorless green ideas sleep furiously

(2) *Ideas colorless green sleep furiously

(3) Big red balloon

(4) *Red big balloon

(5) Bobby said that Shelly owed herself another chance

(6) *Bobby said that Shelly owed himself another chance

(7) The man that drinks coffee is named Dale

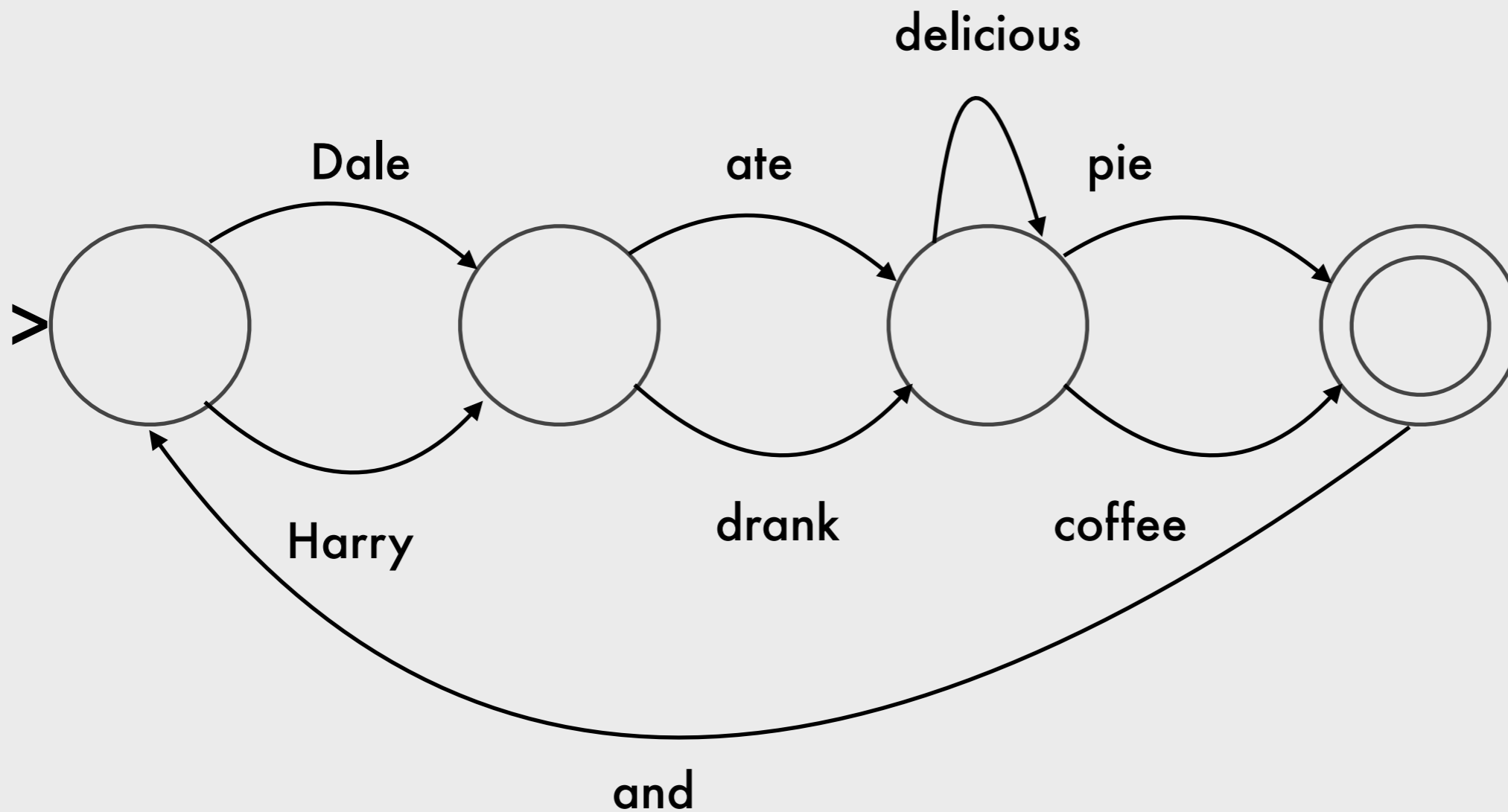
(8). The coffee that Dale drinks is black as a moonless night at midnight

(9). *The man drinks coffee is named Dale

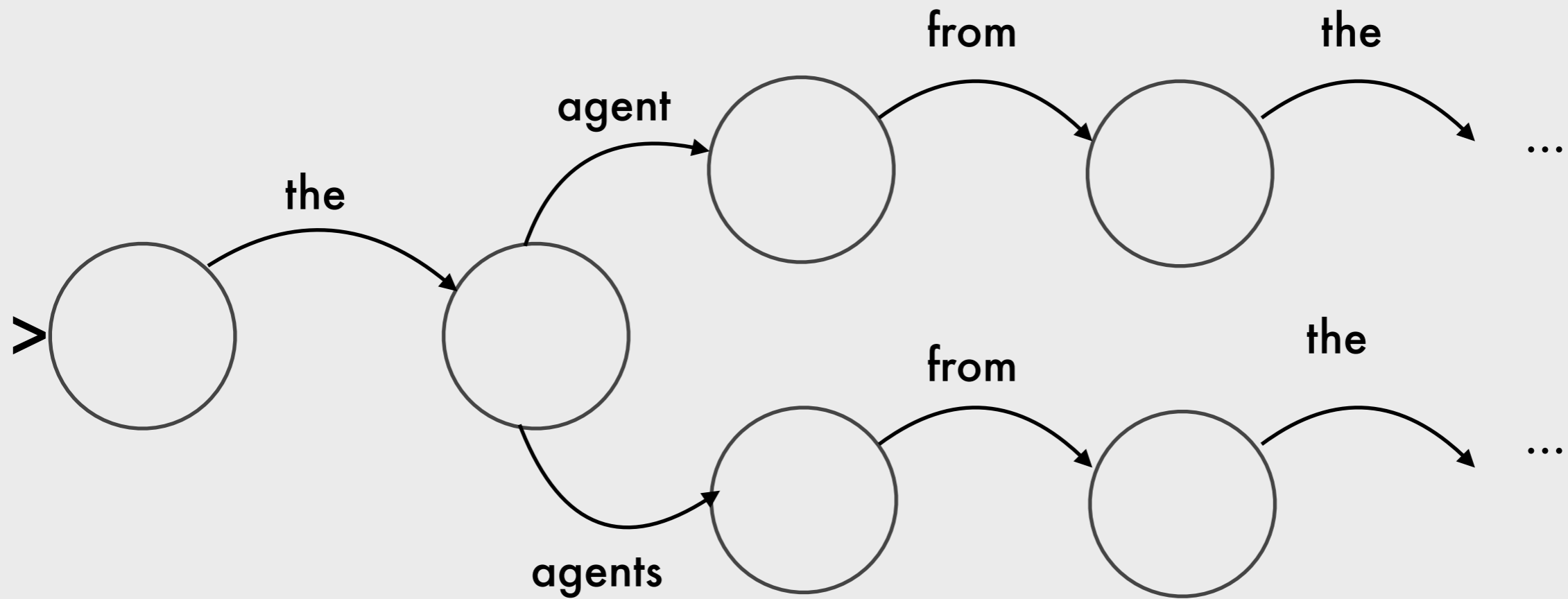
(10). The coffee Dale drinks is black as a moonless night at midnight

- The **syntax** of a language are a set of rules that determine possible sound-meaning pairs
- Certain sound sequences are not **generated** by the syntax of a language; i.e., they're **ungrammatical**
- Speakers of natural languages can utter/understand an indefinite number of sentences...
 - (1) The purple hippo that the velociraptor liked ate a banana
 - (2) Colorless green ideas sleep furiously
- ... and sentences may be indefinitely long
 - (1) Lucy ate a chocolate bunny
 - (2) Harry heard that [Lucy ate a chocolate bunny]
 - (3) Dale said that [Harry heard that [Lucy ate a chocolate bunny]]

- One kind of device that can generate the set of possible strings in a language is a **finite-state automata**:



- Finite-State Automata are not powerful enough to capture **long-distance dependencies**
 - (1) The **agent** from the FBI **likes** coffee
 - (2) The **agents** from the FBI **like** coffee
 - (3) *The **agent** from the FBI **like** coffee
 - (4) *The **agents** from the FBI **likes** coffee
- The above can be captured with a FSA, with paths diverging after the states that accept *agent / agents* –



- There can be an indefinite number of modifiers, each doubling the size of the FSA w/ massive redundancy
 - (1) The agents [that Harry likes] like...
 - (2) The agents [that the sheriff [that lives in Twin Peaks] likes] like...
 - (3) The agents [that the sheriff [that lives in [the city [that's located in Washington]] likes] like ...
- FSAs miss the generalization that words are collected into **phrases** and phrases are in **grammatical relations** with each other

- Most theories of syntax make sure of **context-free grammars (CFGs)**:
 1. A "start" symbol, S
 2. Rewrite rules
 3. A set of non-terminal symbols
 4. A set of terminal symbols

$S \rightarrow NP VP$

$NP \rightarrow Det N$

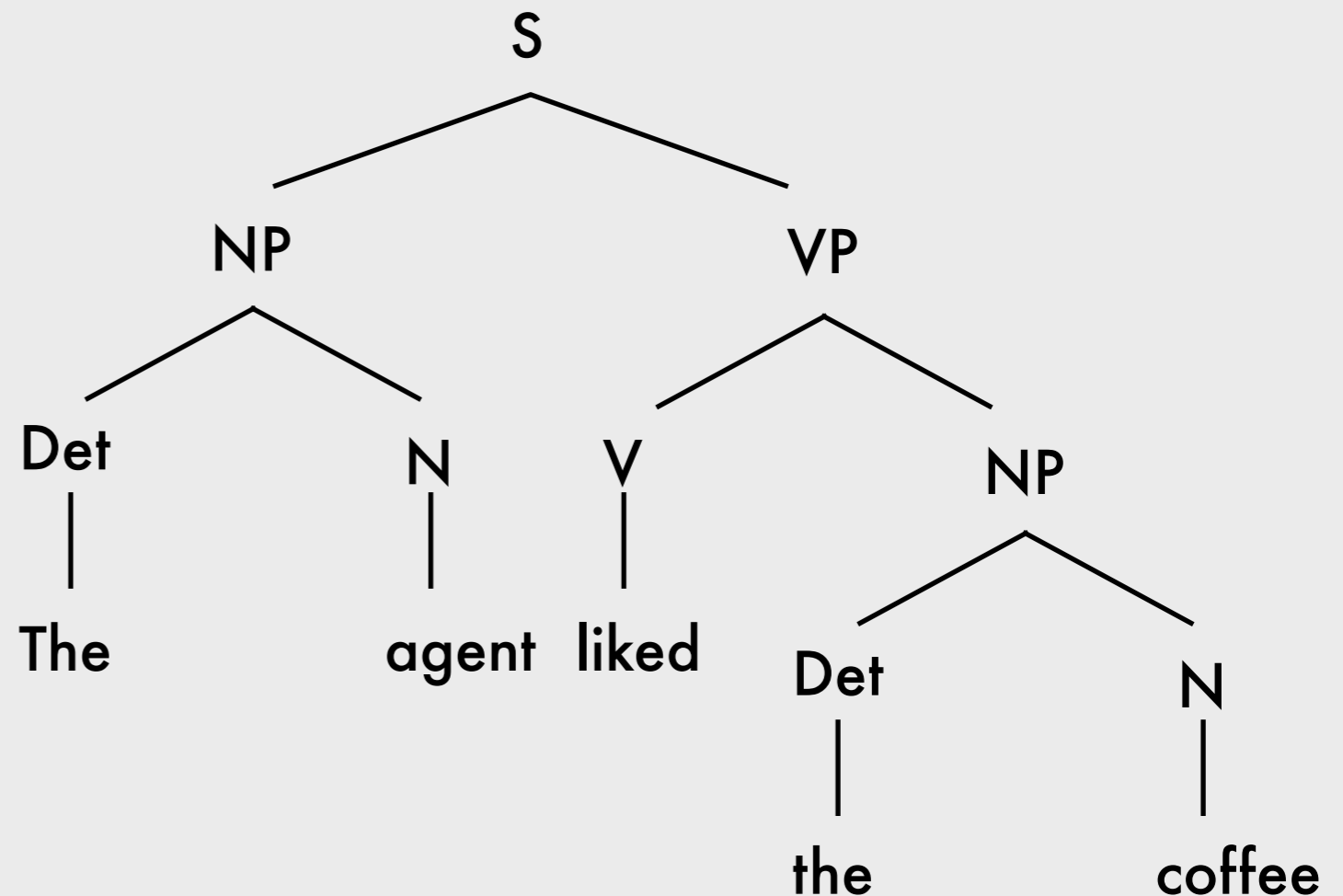
$VP \rightarrow V$

$VP \rightarrow V NP$

$Det \rightarrow the, a, \dots$

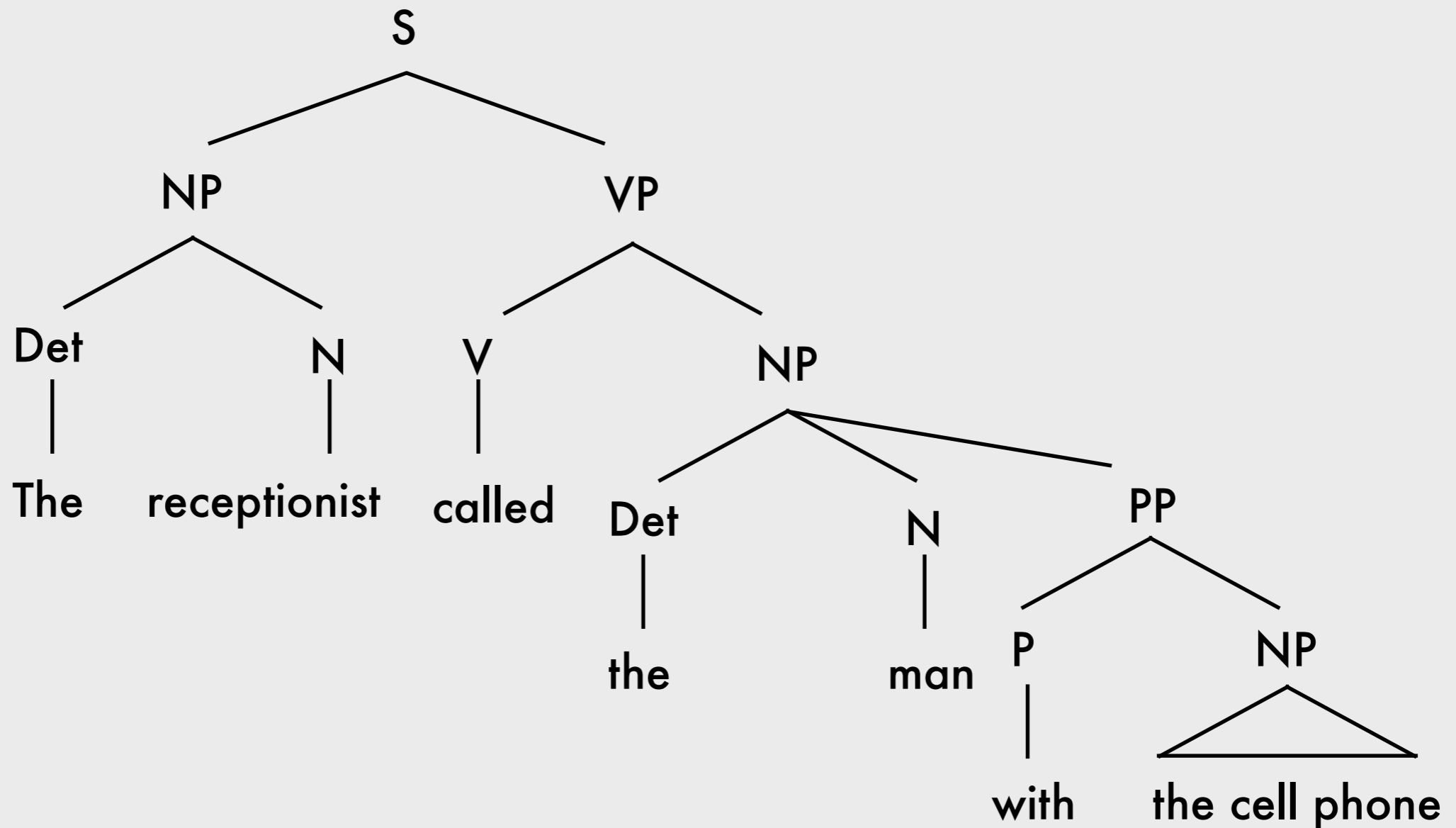
$N \rightarrow agent, man, coffee, pie \dots$

$V \rightarrow liked, ate, drank, \dots$

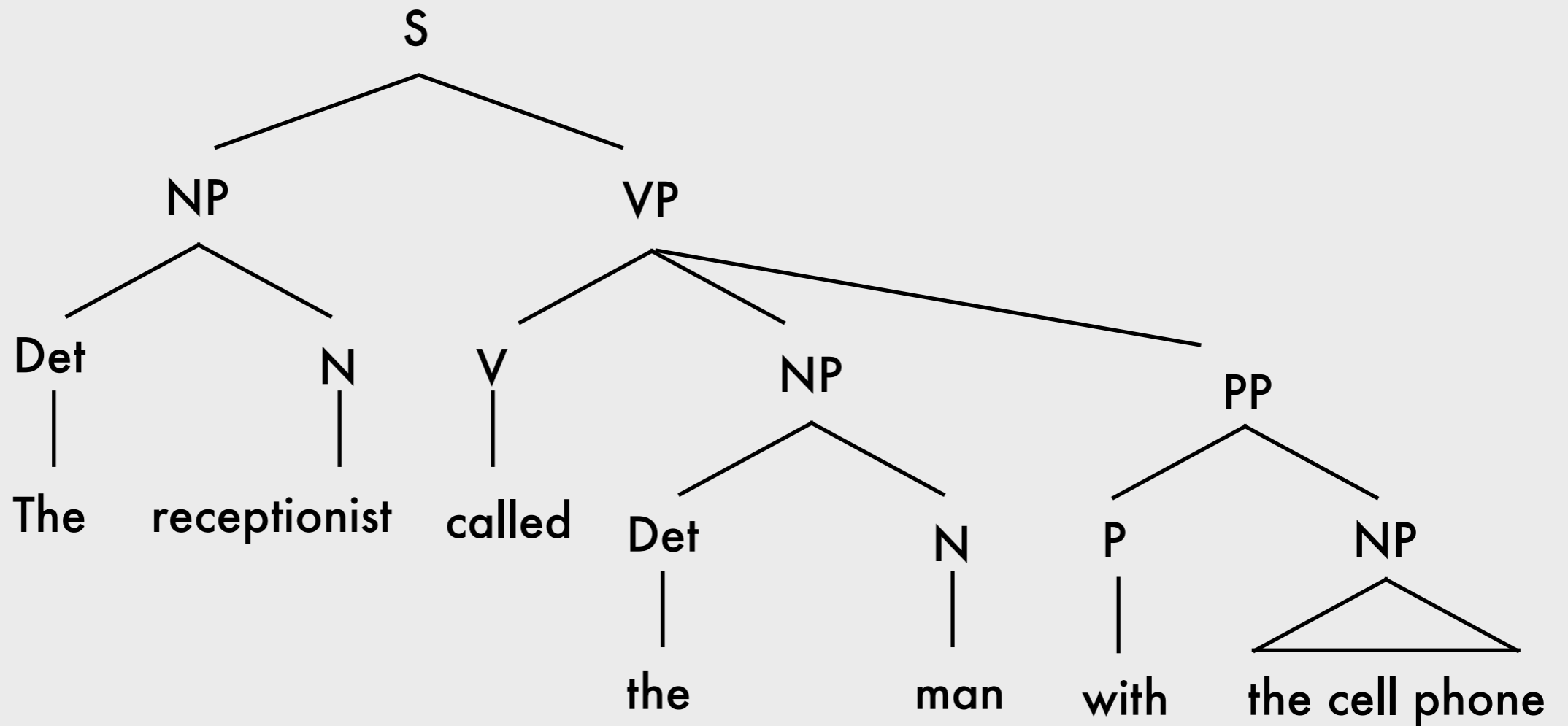


- **Constituency tests** show that grammatical generalizations are sensitive to these intermediate categories (NP, VP, ...)
- **Clefting** targets NPs, PPs, AdvPs:
 - (1) It was [_{NP} Dale] that ate the delicious pie very quickly at the RR Diner
 - (2) It was [_{PP} at the diner] that Dale ate the delicious pie very quickly
 - (3) It was [_{AdvP} very quickly] that ate the delicious pie at the RR Diner
 - (4) *It was [_{AdjP} delicious] that Dale ate the pie very quickly at the RR Diner
- "Do so anaphora" refer to VPs only; pronouns refer to NPs:
 - (5) [[_{NP} Dale] [_{VP} ate the pie at the RR diner], and Shelly *did* so too
 - (6) ... and he really enjoyed it.

- The grammatical structure is closely linked to the semantics:



- The grammatical structure is closely linked to the semantics:



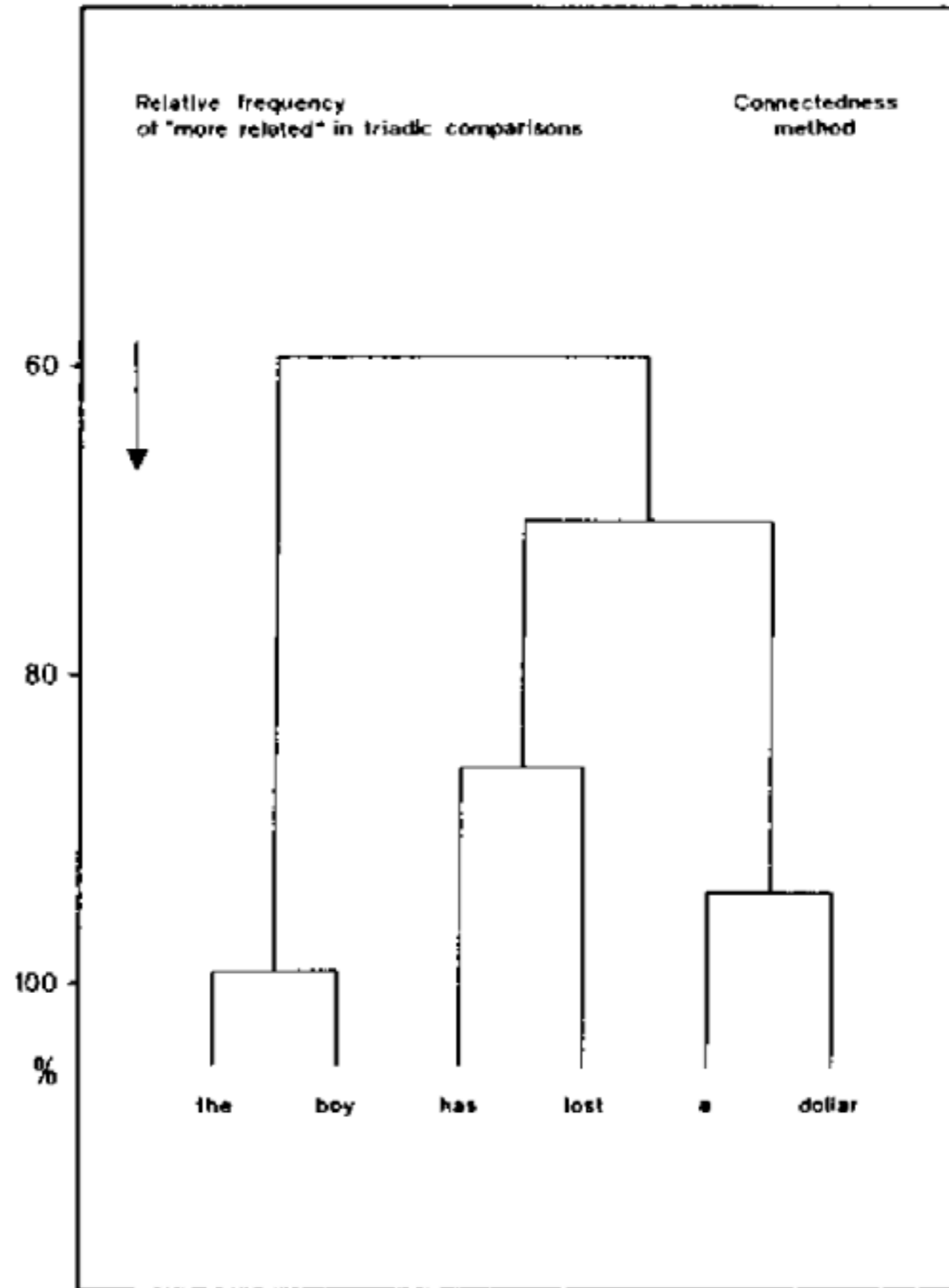
(1) The receptionist called the man with the phone

(2) It was [_{VP} call the man with the phone] that the receptionist did

(3) It was call [_{NP} the man with the phone] that the receptionist did

(4) [_{NP} The man] was called with the phone by the receptionist

- Levelt (1970)



- Within a phrase, there are three pieces – the **head**, **arguments**, and **adjuncts**

- The **head** of the phrase is the word that determines the properties of the whole phrase; i.e., a noun phrase is plural if its head noun is plural:

(1) [_{NP} The agents_[+PL] from the FBI] like_[+PL] coffee

(2) [_{NP} The agent_[-PL] from the FBI] likes_[-PL] coffee

- The head of a phrase may require one other phrase within it; the **argument** or **complement**

(3) Dale [_{VP} devoured [_{NP} the pie]]

(4) *Dale [_{VP} devoured]

- A word may **select** the form of its argument –

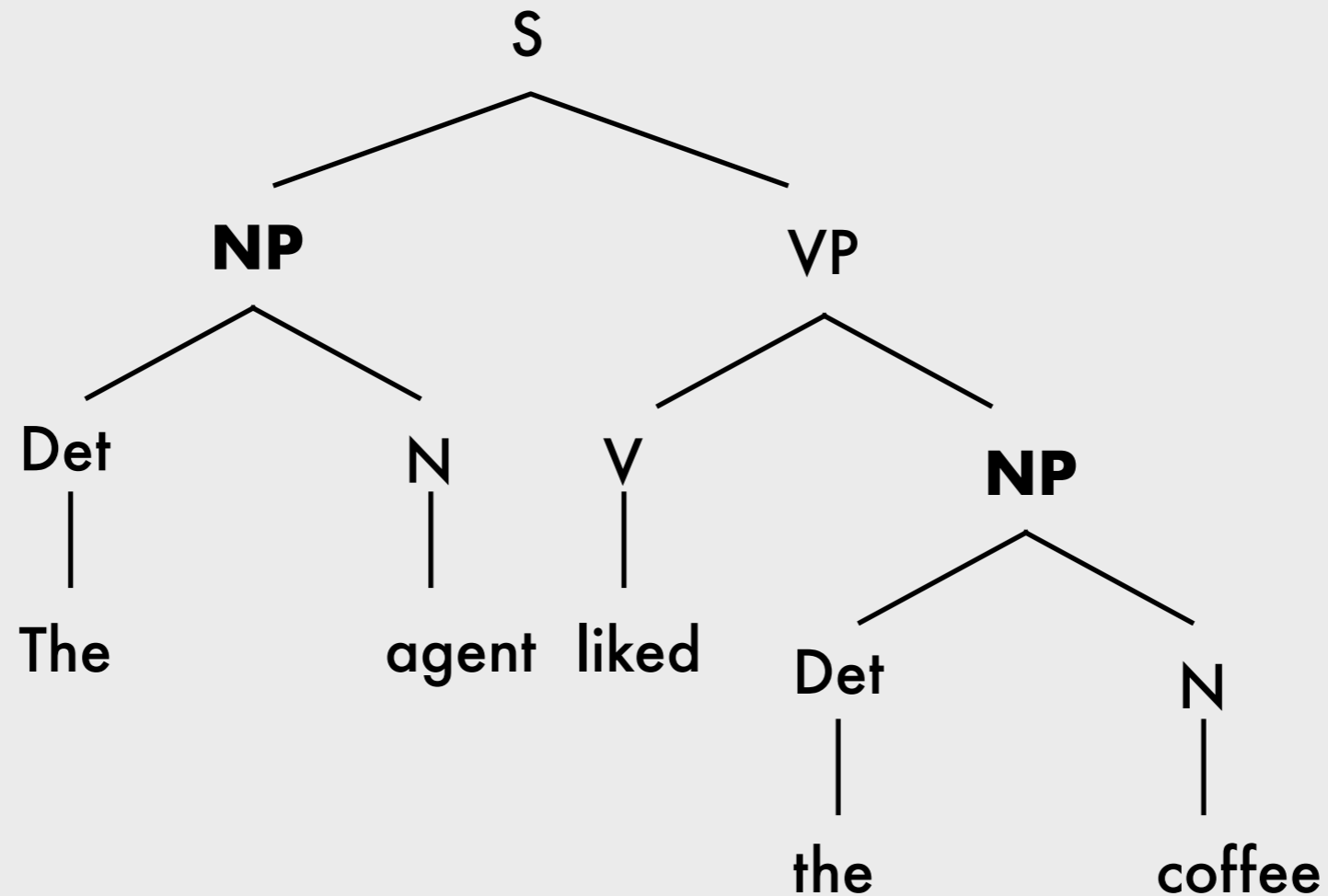
(5) Andy [_{VP} depended [_{PP} on/*in Dale]]

- Other phrases are **adjuncts** – these are unselected, indefinite, and (sometimes) reorderable

- (1) Dale [_{VP} devoured the pie [_{AdvP} very swiftly] [_{PP} at the RR Diner] [_{PP} with Harry]]
- (2) Dale [_{VP} devoured the pie [_{PP} with Harry] [_{AdvP} very swiftly] [_{PP} at the RR Diner]]
- (3) Dale [_{VP} devoured the pie [_{PP} with Harry] [_{PP} at the RR Diner] [_{AdvP} very swiftly]]

- In HW#1, you will find more differences between arguments and adjuncts

- Phrases form **dependencies** which are not directly encoded in the CFG; i.e., “subject” is the NP sister to the VP; “object” is the NP next to the verb



- *Devour* requires an NP complement, which is understood as the object being eaten:

(1) Dale [_{VP} devoured [_{NP} the pie]]

(2) *Dale [_{VP} devoured]

- If the object is a question word, the fronted question word can satisfy the NP complement requirement “long distance”:

(3) **What** did Dale [_{VP} devour] ?

(4) ***What** did Dale [_{VP} devour [_{NP} the pie]] ?

- Most theories of syntax posit multiple representations to capture these facts; e.g., **movement rules**:

(5) Dale [_{VP} devoured **what**] → **what** did Dale [_{VP} devour what]