

LIN7209 – Syntax

Theory and Architecture

Adèle Hénot-Mortier (based on David's original materials)

23/09/2025

Queen Mary University of London

Logistics at a glance

- Adèle (she/her, a.mortier@qmul.ac.uk, ArtsOne 113).
- Tuesday 9-11, ArtsTwo 3.17.
- Course updated regularly on [QMPlus](#).
- Office hours Monday and Friday 13-14 starting 29/09.

Readings

- The **background reading** is David's **Core Syntax**. I encourage you to read the relevant Chapter(s) (give in **syllabus**) prior to class.
- Each week has a **[main reading]** (see **syllabus**). We'll go through it in class and will serve as the backbone for the discussion. Try to read it both before and after class.
- **(Extra reading)** can be done whenever, but ideally should be done by the end of the class.
- We'll work on data assignments based on what we have covered in some of the class sessions and this will be the basis of the assessment.

Assessment

- You'll have to write up a portion of the **data analysis** exercises we do in class in the form of a short linguistics paper (~4000 words).
- Criteria for assessment:
 - **professional presentation** (examples, glosses, citation, bibliography etc.);
 - **clear argumentation** connecting data, analysis and theory.
- It is QMUL policy to penalize **late submissions** 5 marks per day. In addition, submissions received more than 7 days after the due date will receive a mark of 0.
- Any request for an **extension** must be made no later than 24h after the due date with valid supporting evidence.

Introduction

What's Syntax?

- (1) The quick brown fox jumps over the lazy dog.
 - (2) Colorless green ideas sleep furiously.
- How do we assign a meaning to these sequences of words?



- A “bruteforce” solution: syntax encodes pairs of form-meaning mappings.
 - Immediate issue: that would require a lot of memory space, in fact an **infinite amount of space**, given that sentences can be arbitrarily long, so there’s an infinite number of them.
- (3) Jo told Ed that Al believes that ... that Lu is sick.

A big picture question

- What should a theory of syntax look like then?
- Question *a priori* independent from a specific syntactic framework.
- At the most basic level, Syntax must involve a number of **primitives**, but also a number of **productive rules/principles**, allowing us to generate an infinite number of outputs with finite means (**generative** syntax).
- Core ingredients we will review today:
 - Constituent structure;
 - Argument structure;
 - Dependency.
- We'll also quickly review how generative syntax built on these ingredients and evolved into the so-called Minimalist framework.
- Before diving this, let's further define **what really falls into Syntax**.

Syntax and processing

Agreement attraction

(4) [?]* The cat near the windows **have** bitten the dog.
 subject attractor

- Sometimes, non-subjects linearly intervene between the subject and the verb, and make us produce the “wrong” agreement. This is called **agreement attraction**.
- Should we consider (4) partially grammatical then?
- Not really: after careful consideration, most if not all speakers would agree that (4) deserves a *.
- Additionally, kids are exposed to such sentences but do not overproduce them.
- The *processing* of two linearly adjacent nominals can be confusing when determining agreement–this appears independent from Syntax.

Garden-pathing (I)

- (5) The old man the boat, while the young laze about on the shore.
- Sentences may be hard to make sense of, because we want to interpret them in a certain way, but something towards the end of the sentence is **not consistent with this initial interpretation**.
 - In (5), we want to understand *The old man* as a noun phrase (\simeq *the man that is old*), rather than a noun phrase plus a verb (\simeq *the old people operate...*).
 - This is because *old* is **more likely** to be used as an adjective as opposed to a noun, and *man*, as a noun as opposed to a verb. Consequently, the structure we first infer is different from the one that eventually makes sense.
 - The continuation *while the young...* helps disambiguate in favor of the unlikely yet consistent interpretation.

Garden-pathing (II)

- (6) The woman brought the sandwich from the kitchen tripped.
- Here, the problem comes from the fact *brought* is understood as a past tense **main verb**, whereas it's intended to be a **past participle** within a reduced relative! Consequently, the main verb *tripped* does not make sense at first.
 - Should we consider (5) and (6) partially ungrammatical then?
 - Not really: syntactically innocuous substitutions drastically improve these sentences!

(5') The **elderly** man the boat.

(6') The woman **given** the sandwich from the kitchen tripped.

Ungrammaticality vs. processing difficulties

- Agreement attraction and garden-path effects outline the fact that sentences can “feel” off due to processing difficulties, rooted in our idea of what’s likely (**statistics**).
- Usually, these difficulties are overcome after more careful consideration, or by changing some superficial aspects of the sentence to reduce ambiguities.
- A sentence like (7) cannot be “rescued” that way.

(7) * Herself likes Jo.

- So processing difficulties are not issues with the underlying structures, but with the likelihood assigned to particular instantiations of the structure.

Syntax and the Parser

- We need to distinguish the **syntactic module** from the processing module (the **parser**).
- The parser uses the rules provided by Syntax, but also extra heuristics (e.g. use the most likely rule at any given point) to make processing faster.
- Sentences can be ungrammatical but wrongly accepted/produced by the parser (**attraction**), or grammatical but rejected by the parser (**garden-pathing**).
- This relates to the standard difference between syntactic **competence** (our abstract, internal capacity to produce any sentence), and **performance** (what we actually produce, given our cognitive limitations).
- Let's now review a few core components of Syntax, starting with constituent structure.

Constituency

Evidence 1: Basic word order

- In English, many sentences follow the SVO pattern.

(8) The cat bit the dog.
 AGENT PATIENT

- Null hypothesis: the form-meaning relation is purely linear: the AGENT comes first, then the predicate, then the PATIENT.
- Easily refuted: passives swap agent and patient, yet mean roughly the same thing as their active counterparts!

(8') The dog was bitten by the cat.
 PATIENT AGENT

- Syntax must encode something about how the different elements of a sentence **interact** beyond linear order.

Extension: verb-second

- Many Germanic languages but also Ingush (Northeast Caucasian) and O'odham (Uto-Aztecán) are verb-second (V2). From Yiddish (V2 in main and embedded clauses):

- (9) Ikh **hob** gezen mitvokh, az ikh **vel** nit kenen kumen donershtik.
I have seen Wednesday that I will not can come Thursday
'I saw on Wednesday that I wouldn't be able to come on Thursday.'
- (10) Mitvokh **hob** ikh gezen, az donershtik **vel** ikh nit kenen kumen.
Wednesday have I seen that Thursday will I not can come.
'On Wednesday I saw that on Thursday I wouldn't be able to come.'

- Whatever comes first (before the verb), the sentence roughly has the same meaning (modulo emphasis). Again evidence for "non-linear" structure.
- English has vestigial V2. Can you think of some examples?



Evidence 2: question formation

- In English polar questions, subject and auxiliary get swapped.

- (11) a. The cat **has** bitten the dog.
 b. **Has** the cat bitten the dog?

- Null hypothesis: swap the linearly first noun phrase with the linearly first auxiliary in the sentence.
- Fairly easily refuted: when the first auxiliary is part of a relative clause modifying the subject, it is not the target of inversion!

- (12) a. The cat [that **is** lying on the mat] **has** bitten the dog.
 b. * **Is** the cat [that lying on the mat] **has** bitten the dog?
 c. **Has** the cat [that is lying on the mat] bitten the dog?

- Syntax must encode what it means to be the “main” auxiliary—some **hierarchy**.

Evidence 3: agreement

- In English, third person main verbs overtly agree in number with their subject.

(13) The cat (near the door) has bitten the dog.

- Null hypothesis: agree with the linearly closest noun phrase.
- Fairly easily refuted: an attraction configuration whereby the closest noun differs in number with the subject!

(4) a. *The cat near the windows **have** bitten the dog.

b. The cat near the windows **has** bitten the dog.

- Extra evidence that syntax must encode some hierarchy.

Encoding hierarchy via constituent structure

- A straightforward implementation of hierarchical relations between words is **constituent structure**.
- Word or groups of words are **bracketed** to form constituents. In our case, a well-formed bracketing is s.t.:
 - [w] with w a word is a well-formed constituent;
 - [C] is well formed iff C is a concatenation of one or more well-formed constituents.
- This is an inductive definition: it applies to arbitrarily long sentences.

(14) [[[The][[cat][[near][[the][windows]]]]][[has][[bitten][[the][dog]]]]]

- This gets heavy real quick so obvious/irrelevant brackets may be dropped.
- Brackets may be **labeled** (e.g. $[_{DP} \text{the windows}]$), but some theories choose not to encode this as part of constituent structure, instead encoding this kind of information if the words.

Standard constituency tests

(15) [The cat near the window] has bitten the dog.



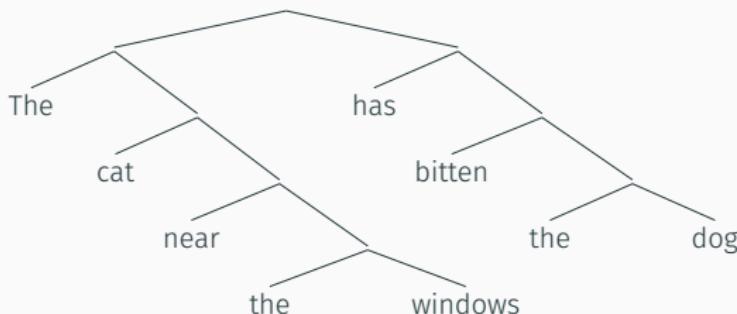
- Substitution: *Sachou has bitten the dog.*
- Clefting: *It's the cat near the window that has bitten the dog.*
- Question formation: *Who bit the dog? The cat near the window.*
- Coordination: *Sachou and the cat near the window have bitten the dog.*

(16) The cat near the window [has bitten the dog].

- Ellipsis: *The cat near the window has bitten the dog, and Sachou too* (has bitten the dog)

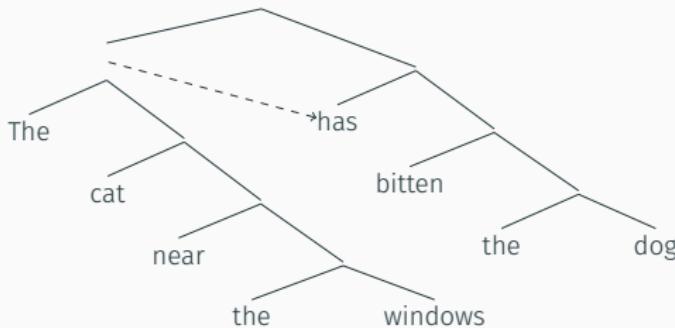
A convenient and equivalent view: tree structure

- Constituents form trees. Trees can be defined inductively:
 - a single node is a tree (“terminal”/“leaf”)
 - a node branching into one or more trees (“non-terminal”), is a tree.
- The mapping between trees and constituents can then be established as follows:
 - $[w]$ is a single node labeled w .
 - $[C]$ with $C = C_0 \dots C_k$ is a tree whose branches are the trees obtained from C_0 , $C_1 \dots$ and C_k respectively.



Hierarchies in trees

- Having trees makes available hierarchical relations:
 - N dominates N'** if N is the mother of N' or N is the mother of a node dominating N' ;
 - N C-Commands N'** if N and N' are not in any kind of dominance relation, and whatever dominates N dominates N' .

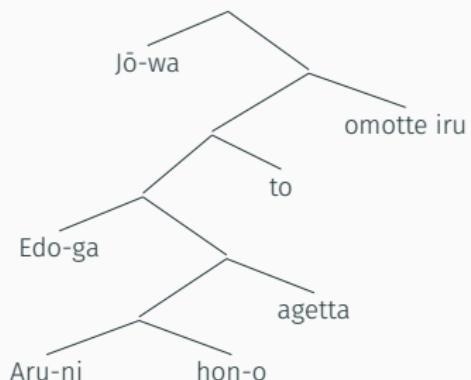
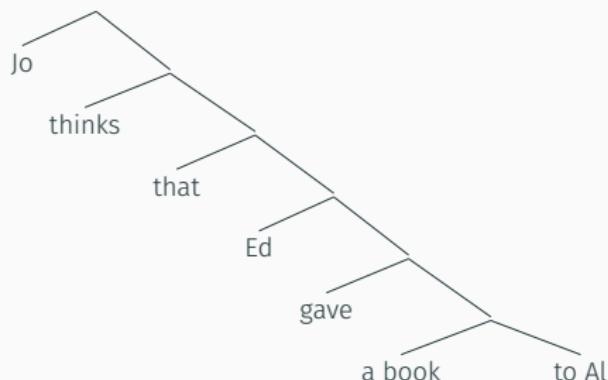


- Trees are expressive enough to explain why *has* is singular in (14) and the above tree: it must agree with the entire C-Commanding constituent *The cat near the windows*, rather than with the deeply embedded constituent *the windows*.

What about ordering then? English vs. Japanese

- Ideally, syntactic representations should minimally differ across languages.

- (17) a. Jo thinks that Ed gave a book to Al.
b. Jō-wa Edo-ga Aru-ni hon-o ageta to omotte iru.



Advantages of an unordered tree structure

- Sentences must be produced sequentially (**linearization**), but their trees don't come with an inherent notion of ordering between sister nodes. Being the left/right node on paper does not mean being left/right in the sentence.



- Separating tree structure from linearization allows different languages to express similar sentences with the **same core structure**: they just happen to linearize these structures differently.

How structure still constraints order

- Tree structure constrains the space of possible linearizations, in ways that are empirically validated.
 - The distribution of attested $\{S, V, O\}$ word orders supports a $[S[VO]]$ core structure.
 - The distribution of attested $\{Dem, Num, Adj, Noun\}$ orders (along with experimental evidence) supports a $[Dem[Num[Adj Noun]]]$ core structure.

One order, two structures

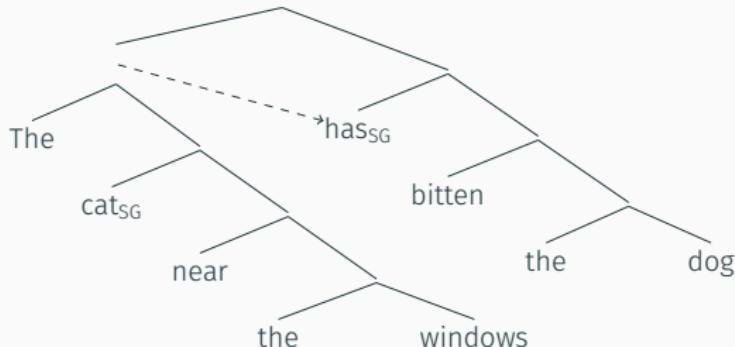
(18) Sleepy cats and dogs



- A tree can have different linearizations, and the same linearization, can be compatible with different trees!
- Such **structurally ambiguous** sentences, are also **semantically ambiguous**.

Argument structure

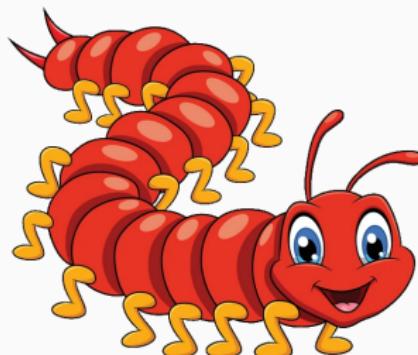
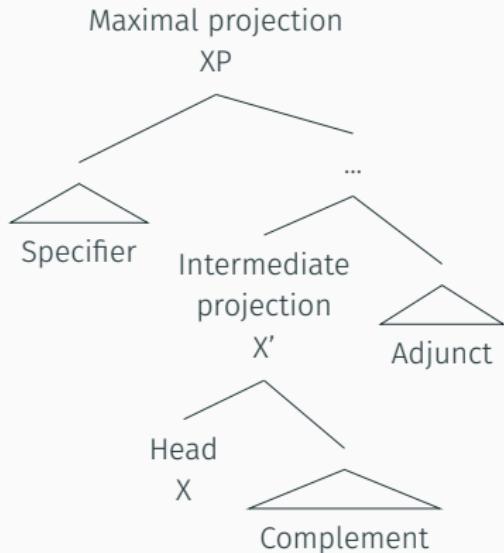
Back to agreement



- To explain why *has* must be singular, we need to explain why the constituent *the cat near the window* should be seen as singular.
- Intuition: this constituent inherits the properties of one of its lexical elements; here the (singular) noun *cat*.

Heads and structural relations

- Constituents tend to inherit the distribution and properties of one of their elements—the **head**.
 - The other elements of the constituent can then be defined by their structural relation to the head.



The head of a sentence

- Constituents inherit the behavior of their head—what about entire sentences then? Are they like subjects? VPs?

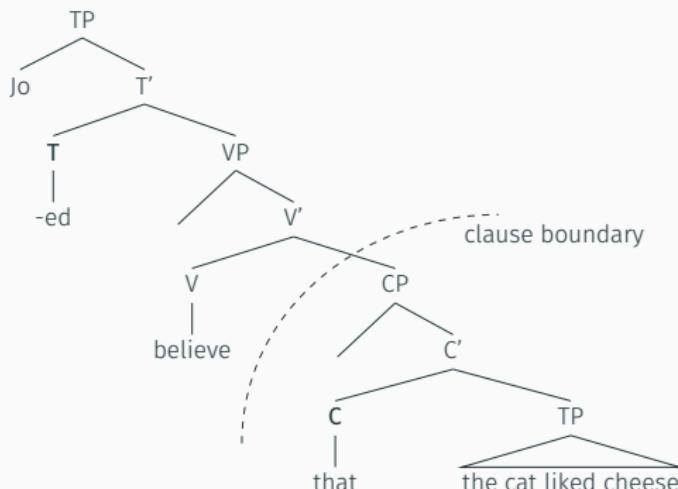


- (11) a. The cat **has** bitten the dog.
b. **Has** the cat bitten the dog?
- (19) a. The cat likes cheese.
b. **Does** the cat like cheese?
- (20) The cat **liked** cheese.

- Questions (and some declarative sentences) contain an auxiliary that's neither the subject nor the VP proper, and encodes **tense**. Tense can also surface as an affix.
- Sentences behave like **tensed phrases**; they are headed by a tense (T) head. The Specifier of T (Spec-TP) eventually hosts the subject.

Functional heads

- Functional heads like T correspond to “functional”, **closed-class**, morpho-phonologically weak material: tense, aspect, complementizer, voice...
- Usually located **high** in the structure, though certain frameworks assume that functional heads are virtually everywhere (think nouns, verbs, adjectives etc...)



Arguments

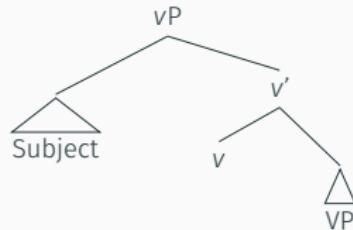
- Arguments are the elements that a head (typically verbal) requires for the sentence to be well-formed.
- **Internal** argument: forms a constituent with the head; complement.
- **External** argument: not part of the head's maximal projection; typically subject.

The case of subjects

- If a given constituent is a true argument of a head, then the head may be sensitive to its semantic characteristics.
- The interpretation of V is often sensitive to the semantic characteristic of its object, but never to those of its subject!

- (21) a. throw a baseball.
b. throw support behind a candidate.
c. throw a party.

- This implies subjects are not true arguments of the verb, and motivates a structure whereby subjects are not specifiers of V, but instead specifiers of a higher functional head v.



Structural vs. thematic relations

- In English, predicates always have an overt subject, but may or may not have an object. Those that do are **transitive**; those that don't are **intransitive**.
- Should we equate subject (a structural relation) with **AGENT** (a thematic/semantic relation)?

- (22) The cat was bitten by the dog. Passive
 PATIENT AGENT
- (23) a. Le chat_{PATIENT} est tombé.
 The cat is fallen.
 'The cat fell.' Intransitive unaccusative
- b. Le chat_{AGENT} a grogné.
 The cat has growled.
 'The cat growled.' Intransitive unergative

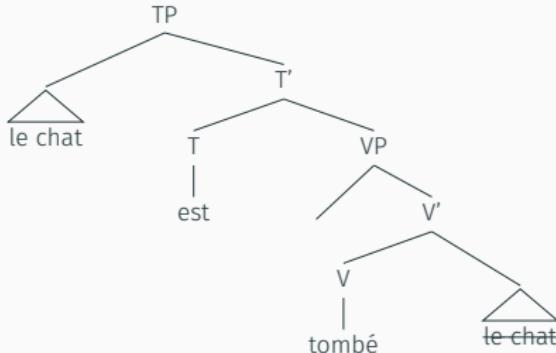
Dependencies

The unaccusative/unergative distinction

- (23) a. Le chat_{PATIENT} est tombé.
The cat is fallen.
'The cat fell.'

Intransitive unaccusative

- How come subjects of intransitives are sometimes sometimes PATIENT?
- Can be explained by a **movement dependency**: a PATIENT starts its life as complement of the verb, then moves up to Spec-vP/Spec-TP because English *needs* an overt subject!



Subcategorization

- (24) a. * Jo hit
b. Jo hit { the ball / *the concept to time }.
- (25) a. Jo believes { that / *whether } Sachou likes cheese.
b. Jo wonders { *that / whether } Sachou likes cheese.

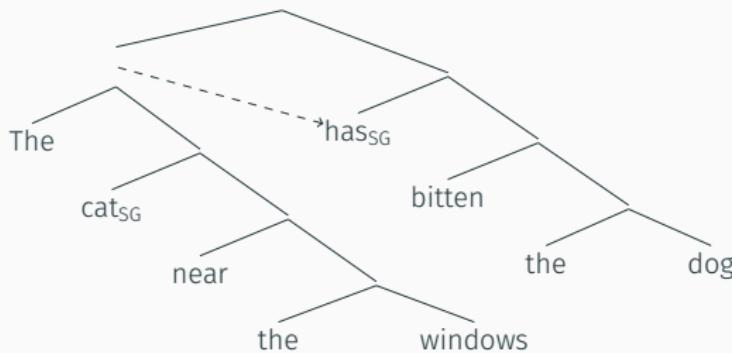
- Some predicates are transitive and can have an object, some are intransitive and cannot.
- Transitive predicates may impose further conditions on their object: animacy, a syntactic category...
- This is known as subcategorization or selection. This constitutes a **local, positional dependency** between a head and its complement.

Bound readings / coreference

- (26) Every kitten wanted its food. $\forall x : \text{kitten}(x). \text{want}(x, \text{food}(x))$
- (27) Jo told Lu to love herself.
~~ Jo told Lu to love Lu.
~~ *Jo told Lu to love Jo.

- Possessive/reflexive pronouns **corefer** with specific elements:
 - variable introduced by a quantified expression (e.g. *every kitten*);
 - referent introduced by a referring expression (e.g. *Lu*).
- The possibility of a dependency is structurally determined by C-Command, and in the case of English reflexives, restricted to a local domain (here, excluding *Jo*).
- (Variable) binding is thus another kind of **positional dependency** that Syntax must encode.

Agreement again!



- We have established that two elements in an agreement relation must be in C-Command relation (structural).
- On top of encoding the necessary configuration for agreement, Syntax must encode what agreement is: a **formal dependency** between constituents.

What does formal mean?

- In English, plural is often signaled with a /s/ morpheme. Is the presence of this morpheme on the subject the right condition for plural agreement on the main verb?

(28) The sheep are bleating.

- Some subjects devoid of overt plural morphology still trigger agreement.
- Syntax needs to encode plurality more abstractly, as a plural “**feature**”.
- Formal dependencies lie agreement operate on these abstract features.
- Morphology maps abstract features to overt realizations.

Beyond subject-verb agreement

- Some languages like Tsez (Nakh-Daghestanian) will primarily agree with their object.
- Some languages will agree with any argument, as soon as one of them bears the critical feature (omnivorous agreement).

- (29) a. nits-ikákomimm-ok-innaan-**a**
1-love_{TA}-INV-1Exc-**Sg**
‘She loves us (exclusive).’
- b. nits-ikákomimm-a-nnaan-**a**
1-love_{TA}-3OBJ-1EXC-**Sg**
‘We (exclusive) love **her**.’

Case

- Another kind of formal dependency is **case**, e.g. nominative vs. accusative.

(30) a. She saw him.
 Nom Acc

b. * Him saw she/her.
 Acc Nom Acc

- Is case thematic? Structural?

(31) a. She fell.
 Nom, PATIENT

b. He was bitten.
 Nom, PATIENT

- Case seems to depend on structure, not thematic relations.
- Basque, Georgian, Mayan, Tibetan, Sumerian, some Indo-European and many Indo-Aryan languages display **absolutive-ergative case systems**, in which the subject of an intransitive and the object of a transitive are marked with the absolute, while the subject of transitives are ergative.

Formalizing the core concepts

Three big historical landmarks

- Phrase structure rules
- Transformational grammar
- Minimalism

Phrase structure rules

- Production rules of the form “mother node → children nodes” specifying how to build syntax trees.

S → NP VP

NP → Det N

$$\text{VP} \rightarrow \text{V} \text{ NP}$$

Det → the

• • •

- Issue: powerful system, but oblivious of certain inherent differences in argument structure, e.g. (in)transitivity.

- (32) a. The shooting of the hunters. ambiguous
b. The growling of lions. unambiguous

Transformations (I)

- (32) a. The shooting of the hunters. ambiguous
 b. The growling of lions. unambiguous

- Transformational grammar: the above nominalizations are derived from sentences *via* specific transformation rules.
- Because *shoot* is transitive, (32a) can be derived from either *The hunters shoot ...* or from *... shoot the hunters*, capturing the ambiguity.
- Because *growl* is intransitive, (32b) can be only derived from *The lions growl*, hence the absence of an ambiguity.

Transformations (II)

- Transformations affect structure but not meaning. Meaning is determined at the level of “**deep structure**”.
- Applications: active-passive alternation, *tough*-constructions, raising constructions...dependencies in general.

(8) The cat bit the dog.
 AGENT PATIENT

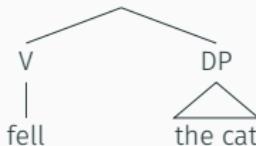
(8') The dog was bitten by the cat.
 PATIENT AGENT

- (33) a. It's tough to please Jo.
 b. Jo is tough to please.
- (34) a. It seems that Jo is happy.
 b. Jo seems to be happy.

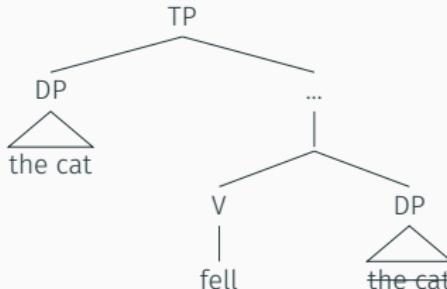
- But a bulky system which poses the question of the division of labor between deep and surface structure (post-transformations).

Minimalism: Merge

- Core components of Syntax enabled by one operation: **MERGE**.
 - Structure building (before: phrase structure rules) is **EXTERNAL MERGE**—you merge two constituents that have not been merged before.

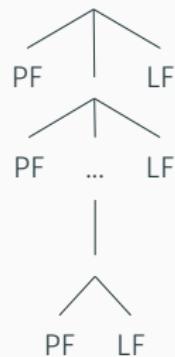


- Movement dependencies (before: transformations) are **INTERNAL MERGE**—you re-MERGE a constituent that was already part of the tree.



Minimalism: a “narrow” Syntax

- Things that are not enabled by MERGE are pushed out of syntax. MERGE produces a structure that is sent to the **interfaces** (“spell-out”):
 - The **phonological** component, which determines how the structure gets pronounced (**Phonological Form**);
 - The **semantic** component, which determines how the structure gets interpreted (**Logical Form**).
- Depending on the size of the structure, there may be multiple spell-out points (“**phases**”).



References i

-  Bock, K. (1995).
Producing agreement.
Current Directions in Psychological Science, 4(2):56–61.
-  Culbertson, J. and Adger, D. (2014).
Language learners privilege structured meaning over surface frequency.
Proceedings of the National Academy of Sciences, 111(16):5842–5847.
-  Dillon, B., Staub, A., Levy, J., and Clifton, C. (2017).
Which noun phrases is the verb supposed to agree with?: Object agreement in american english.
Language, 93(1):65–96.
-  Dryer, M. S. and Haspelmath, M., editors (2013).
WALS Online (v2020.4).
Zenodo.
-  Grishin, P. (2023).
Omnivorous third person agreement in algonquian.
Glossa: a journal of general linguistics, 8(1).

References ii

-  Kratzer, A. (1996).
Severing the External Argument from its Verb, page 109–137.
Springer Netherlands.
-  Nevins, A. (2011).
Multiple agree with clitics: person complementarity vs. omnivorous number.
Natural Language & Linguistic Theory, 29(4):939–971.