



Tutorial on Universal Dependencies

Cross-linguistically consistent syntactic annotation

Joakim Nivre¹ Daniel Zeman² Filip Ginter³ **Francis M. Tyers^{4,5}**

¹Department of Linguistics and Philology, Uppsala University, Sweden

²Institute of Formal and Applied Linguistics, Charles University, Prague, Czechia

³Department of Information Technology, University of Turku, Finland

⁴Giela ja kultuurra instituhtta, UiT Norgga árktalaš universitehta, Tromsø, Norway

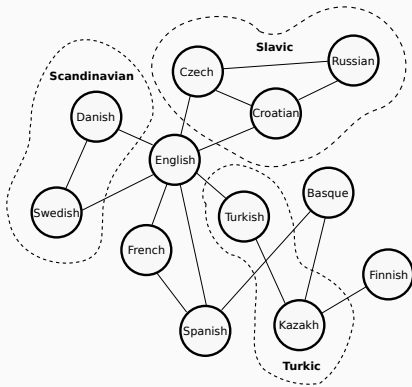
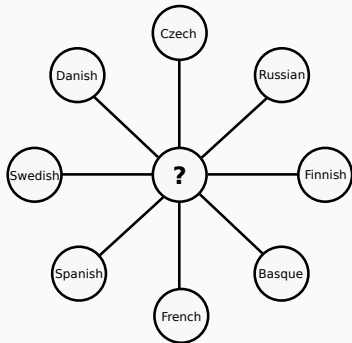
⁵Arvutiteaduse instituut, Tartu Ülikool, Estonia

How can we do cross-lingual syntax ?

- Find the structures we expect to find in all languages
- Describe how they are dealt with
 - Using a representation that facilitates cross-linguistic parallelism
- Allow language-specific extensions



Approaches



UD tries to standardise between languages, and particularly:

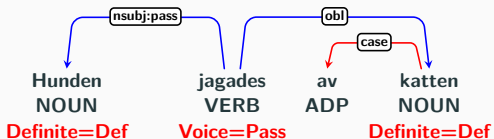
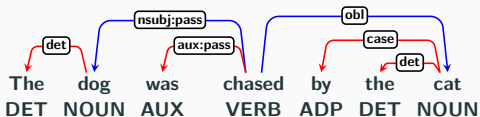
- Within language groups
- Between typologically similar constructions



- Basic principles:
 - The primacy of content words
 - Clauses, nominals and modifier words
 - Core arguments vs. other dependents
- Universal and language-specific relations
- Basic and enhanced dependencies



The Primacy of Content Words



Three Types of Structures

- Clauses headed by a (possibly non-verbal) predicate

the cat chased the dog

the dog is cute

- Nominal phrases

the cat on the mat

my best friend

- Modifier words

very skillfully

not always



Core Arguments

- Arguments of basic intransitive and transitive verbs
- Distinguished by one or more of the following properties:
 - Verbs usually only agree with core arguments
 - Core arguments normally appear without adpositions
 - Certain cases, traditionally called nominative, accusative, and absolutive are typically reserved core arguments
 - Core arguments often occupy special positions in the clause
 - Syntactic phenomena like control, relativisation and passivisation can be restricted to core arguments
- UD distinguishes core arguments from oblique dependents
- UD does **not** distinguish arguments from adjuncts



Syntactic Relations

	Nominal	Clause	Modifier Word	Function Word
Core Predicate Dep	nsubj obj iobj	csubj ccomp xcomp		
Non-Core Predicate Dep	obl vocative expl dislocated	advcl	advmod* discourse	aux cop mark
Nominal Dep	nmod appos nummod	acl	amod	det clf case
Coordination	MWE	Loose	Special	Other
conj cc	fixed flat compound	parataxis list	orphan goeswith reparandum	punct root dep

* Generalized modifier of predicates and (non-nominal) modifiers



A Two-Level Architecture

- Universal relations
 - Broad categories to allow cross-linguistic comparison
- Language-specific relations
 - Subtypes to capture language-specific phenomena

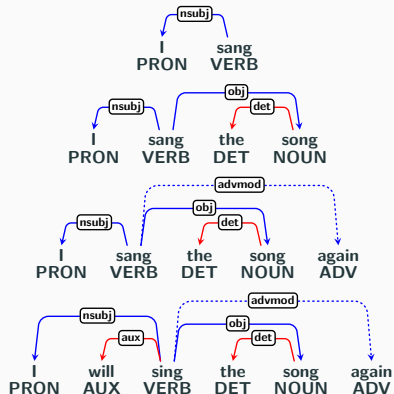
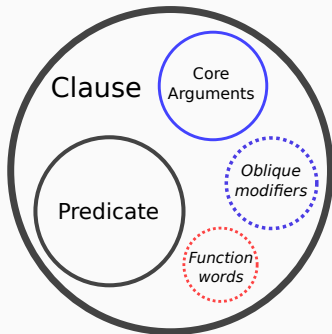
Universal	Language-Specific
acl	acl:relcl
compound	compound:prt
nmod	nmod:poss



Basic clauses



Clauses

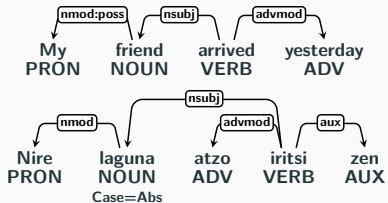


- Basic universal structure
- Distinguish *core* from *oblique*
- Function words are leaf nodes



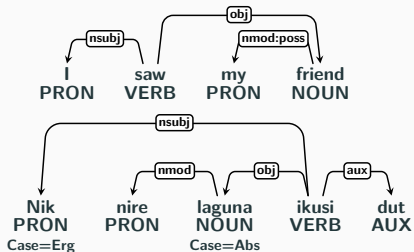
Intransitive

- Single argument, nsubj

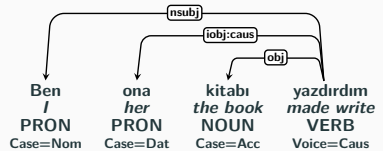
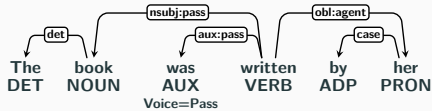
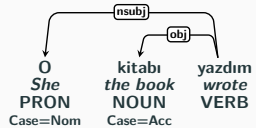
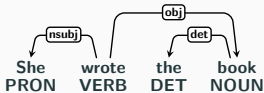


Transitive

- Two arguments,
 - nsubj: The proto-Agent
 - obj: The proto-Patient
- Language-internal criteria
 - Case marking
 - Word order
 - Agreement
 - Valency changes
- Not exclusively case/semantic role



Valency Changing

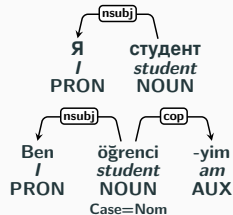
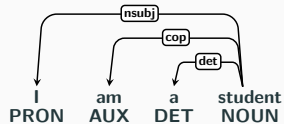


- Annotate syntax, not semantic role
- Optional use of subtypes for:
 - Passive
 - Causative



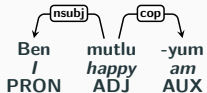
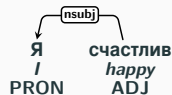
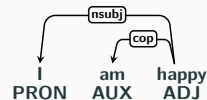
Non-Verbal Predication

- At most one copula item
 - Semantically empty
- Predicate is head
- Types:
 - **Equational**
 - Attributinal
 - Locational



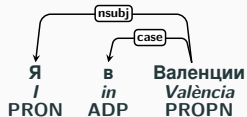
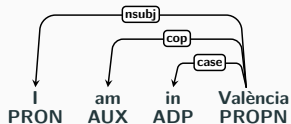
Non-Verbal Predication

- At most one copula item
 - Semantically empty
- Predicate is head
- Types:
 - Equational
 - **Attributional**
 - Locational



Non-Verbal Predication

- At most one copula item
 - Semantically empty
- Predicate is head
- Types:
 - Equational
 - Attributinal
 - **Locational**

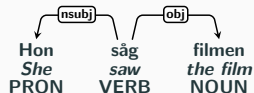


Nominal phrases



Nominal Phrases

- Basic cross-lingual structure
- Headed by a noun, proper noun or pronoun
- Does not take core arguments



Adnominal Modifiers

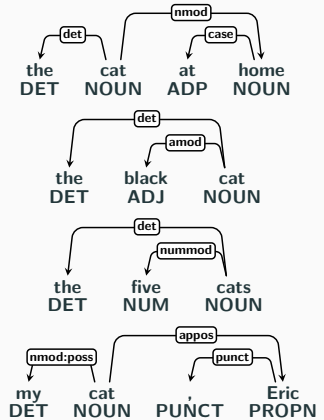
- The type of dependent determines the label for the adnominal modifier:

Nominal nmod

Adjectival amod

Numeral nummod

- Additional label, appos, for apposition



Adnominal Modifiers

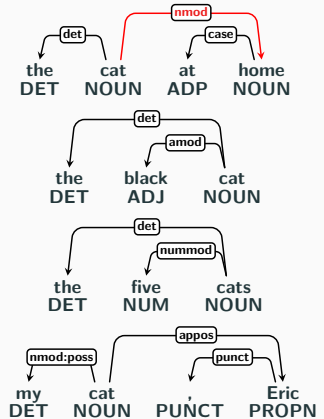
- The type of dependent determines the label for the adnominal modifier:

Nominal nmod

Adjectival amod

Numeral nummod

- Additional label, appos, for apposition



Adnominal Modifiers

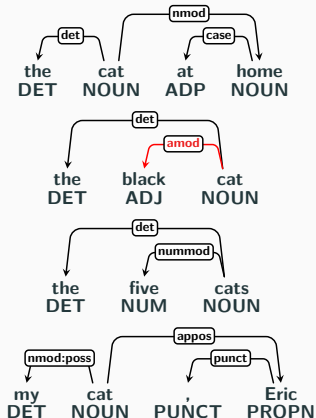
- The type of dependent determines the label for the adnominal modifier:

Nominal nmod

Adjectival amod

Numeral nummod

- Additional label, appos, for apposition

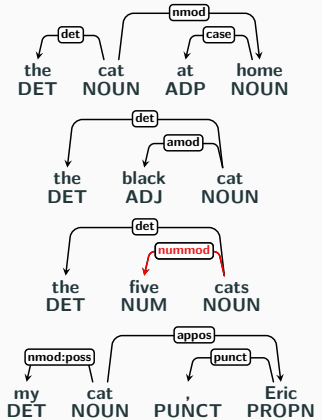


Adnominal Modifiers

- The type of dependent determines the label for the adnominal modifier:

Nominal	nmod
Adjectival	amod
Numeral	nummod

- Additional label, appos, for apposition



Adnominal Modifiers

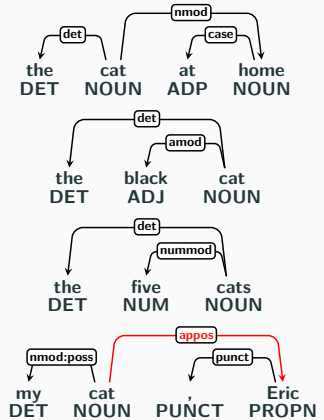
- The type of dependent determines the label for the adnominal modifier:

Nominal nmod

Adjectival amod

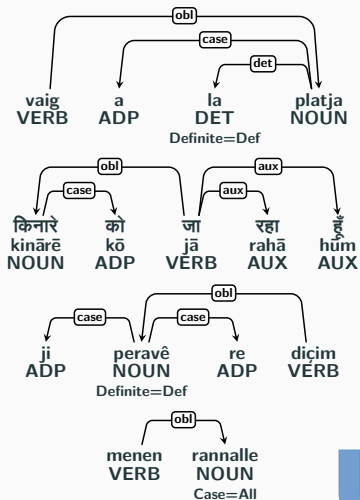
Numeral nummod

- Additional label, appos, for **apposition**



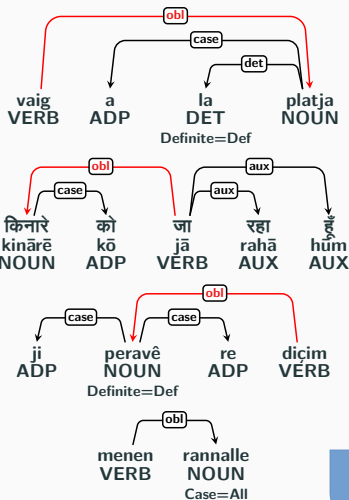
Adpositional Phrases

- Single part-of-speech tag, ADP
- Adpositions attach to the head of the nominal phrase
 - relation: case
- Adpositional phrases are attached to
 - clausal heads: obl
 - nominal heads: nmod
- Parallel with oblique NPs



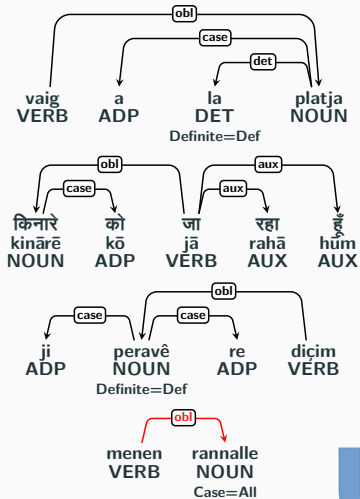
Adpositional Phrases

- Single part-of-speech tag, ADP
- Adpositions attach to the head of the nominal phrase
 - relation: case
- Adpositional phrases are attached to
 - clausal heads: **obl**
 - nominal heads: **nmod**
- Parallel with oblique NPs



Adpositional Phrases

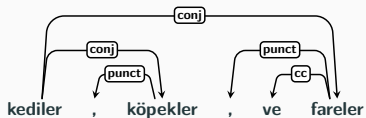
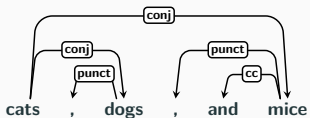
- Single part-of-speech tag, ADP
- Adpositions attach to the head of the nominal phrase
 - relation: case
- Adpositional phrases are attached to
 - clausal heads: obl
 - nominal heads: nmod
- Parallel with oblique NPs



Complex clauses



Coordination



- Co-ordination is not a dependency relation
- The “head” (parent node) is the first conjunct



Subordination

Complex clauses involving subordination arise because a core or non-core dependent is realised as a clausal structure

Four basic types:

Core	Non-core
Clausal subjects	Adverbial clause modifiers
Clausal complements	Adnominal clause modifiers

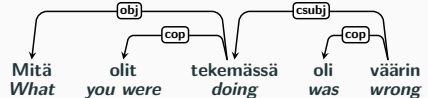
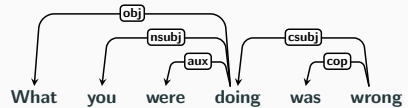
These may be either finite or non-finite



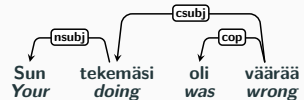
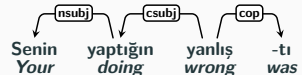
Clausal Subjects

- When the role of subject is filled by a clause
- Relation is csubj
- Sometimes blurry:
 - “His writing surprised me”

Finite

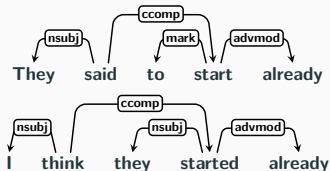


Non-finite

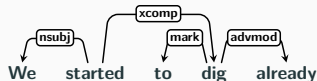


Clausal Complements

Open



Closed

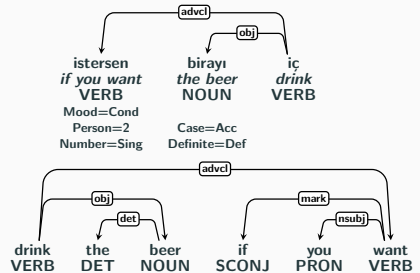


- **Open:** The subject of the subordinate clause is not co-referential with an argument of the matrix clause
- **Closed:** The subject of the subordinate clause is controlled by the matrix clause

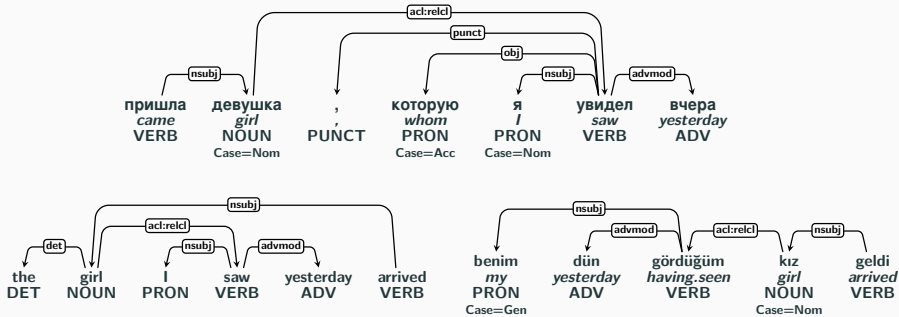


Adverbial Clauses

- Clause modifier of verb or other predicate
- Non-core complement, e.g.
 - Temporal clause
 - Conditional clause
 - Purpose clause
 - ...



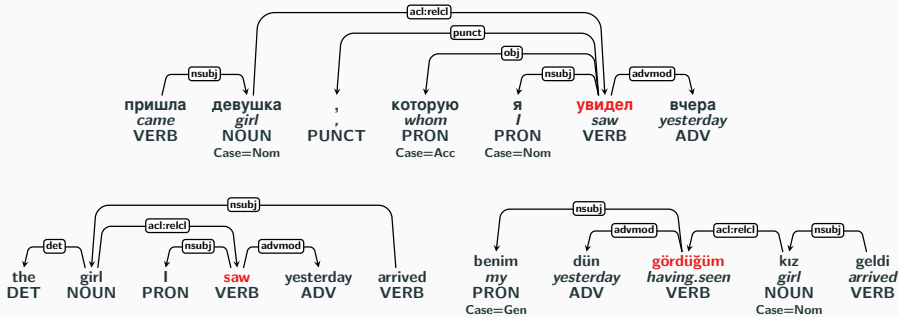
Adnominal Clauses



- The predicate is head of the relative clause
- The relative clause depends on the nominal it modifies
- Arguments (incl. relative pronouns) and modifiers annotated as in main clauses



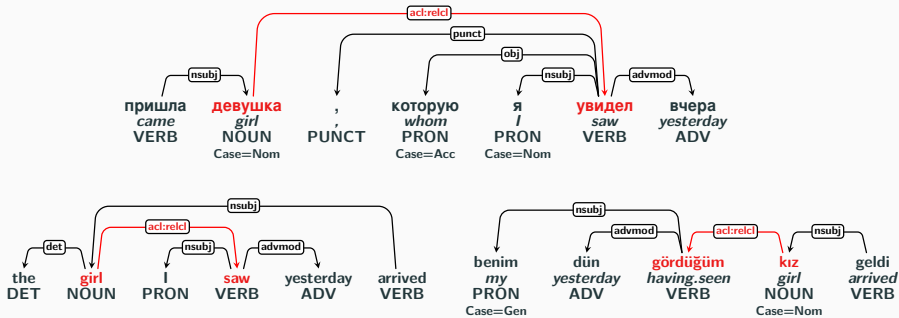
Adnominal Clauses



- The predicate is head of the relative clause
- The relative clause depends on the nominal it modifies
- Arguments (incl. relative pronouns) and modifiers annotated as in main clauses



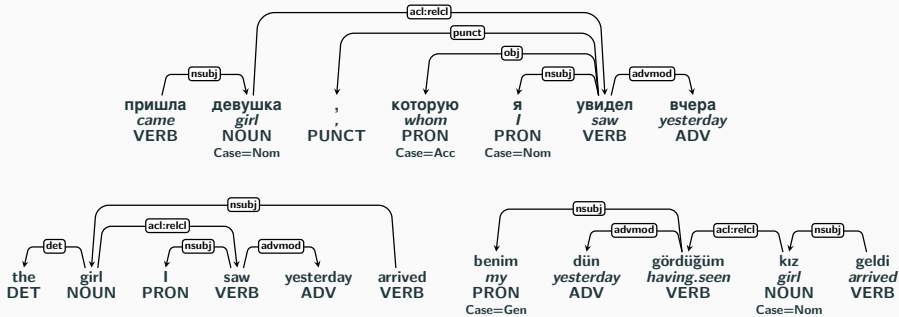
Adnominal Clauses



- The predicate is head of the relative clause
- The relative clause depends on the nominal it modifies
- Arguments (incl. relative pronouns) and modifiers annotated as in main clauses



Adnominal Clauses



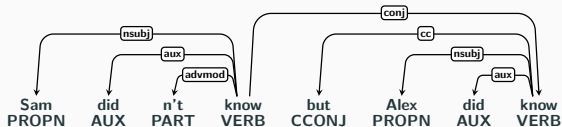
- The predicate is head of the relative clause
- The relative clause depends on the nominal it modifies
- Arguments (incl. relative pronouns) and modifiers annotated as in main clauses



Ellipsis



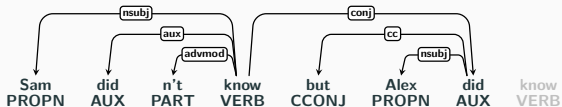
Promotion



- No empty nodes in the basic representation
- Function words are promoted when their heads are elided
- Real dependencies recovered in the enhanced representation



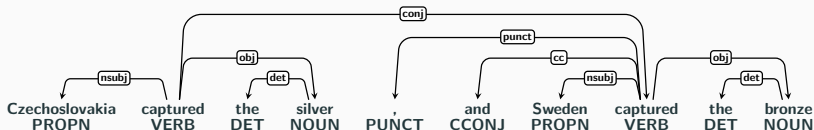
Promotion



- No empty nodes in the basic representation
- Function words are promoted when their heads are elided
- Real dependencies recovered in the enhanced representation



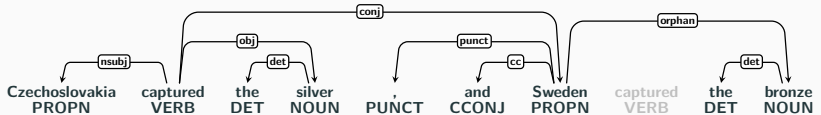
Orphan



- When promotion would result in a misleading dependency relation
 - e.g. an object depending on a subject
- Typical case: core argument promotion in predicate ellipsis
- Maintains clausal integrity



Orphan



- When promotion would result in a misleading dependency relation
 - e.g. an object depending on a subject
- Typical case: core argument promotion in predicate ellipsis
- Maintains clausal integrity

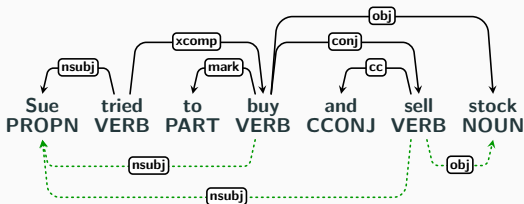


Enhanced dependencies



Enhanced Dependencies

- An extended dependency graph containing:
 - Null nodes for elided predicates
 - Additional subject relations for control and raising constructions
 - Propagation of dependents over coordination
 - Coreference in relative clause constructions
 - Labels augmented with function word information



Questions?



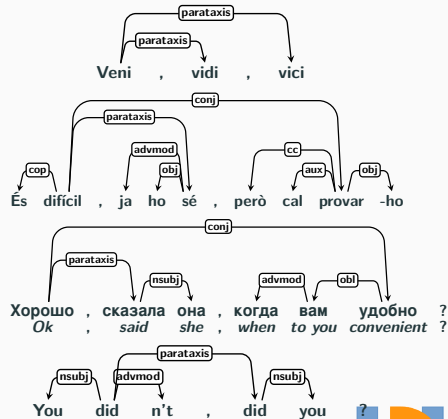
Loose Joining Relations

Parataxis:

- Side-by-side sentences
- Parentheticals
- Some kinds of reported speech
- Tag questions

List:

- Lists of related items



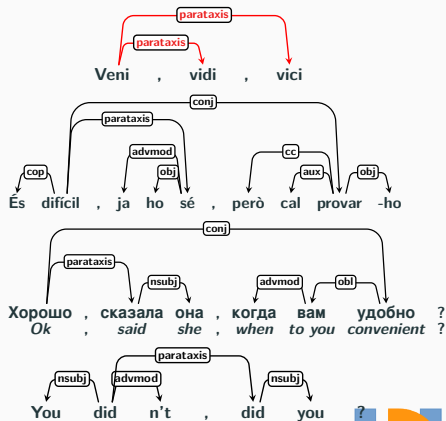
Loose Joining Relations

Parataxis:

- Side-by-side sentences
- Parentheticals
- Some kinds of reported speech
- Tag questions

List:

- Lists of related items



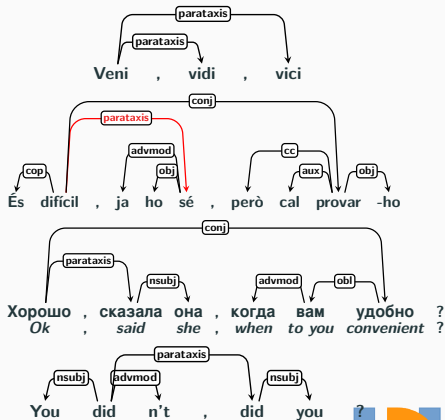
Loose Joining Relations

Parataxis:

- Side-by-side sentences
- **Paraneticals**
- Some kinds of reported speech
- Tag questions

List:

- Lists of related items



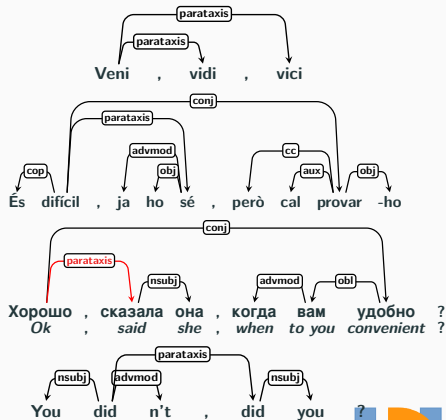
Loose Joining Relations

Parataxis:

- Side-by-side sentences
- Parentheticals
- Some kinds of reported speech
- Tag questions

List:

- Lists of related items



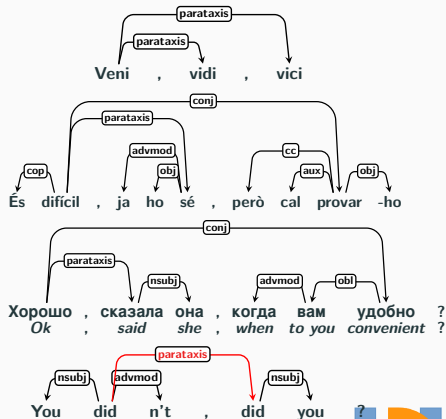
Loose Joining Relations

Parataxis:

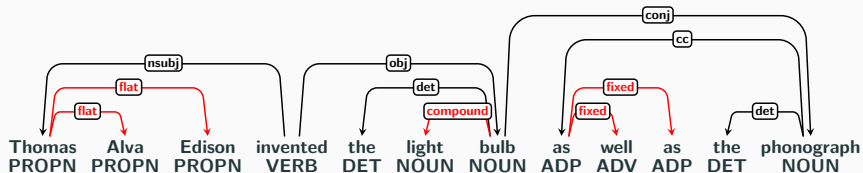
- Side-by-side sentences
- Parentheticals
- Some kinds of reported speech
- **Tag questions**

List:

- Lists of related items



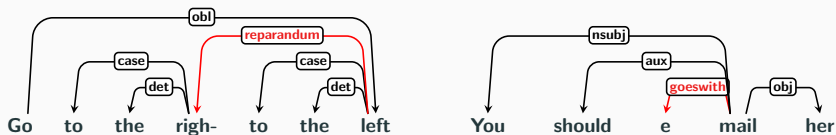
Multiword Expressions



- Most MWEs are given a regular syntactic analysis
- Three special relations:
 - `fixed` = fixed grammaticised expressions
 - `flat` = semi-fixed expressions with no clear head
 - `compound` = lexical compounds (normally headed)



Disfluencies



The reparandum relation:

- Disfluencies that are overridden in speech repair.

The goeswith relation:

- Parts of words due to orthographical or processing errors.

