# Syntactic Analysis

Daniel Zeman

December 12–19, 2024

Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

unless otherwise stated

# Outline

1. Constituents vs. Dependencies

2. Universal Dependencies
   - A Tour through UD Syntax
   - Nonverbal Predicate and Copula
   - Core Arguments vs. Oblique Dependents
   - Ellipsis and Enhanced UD

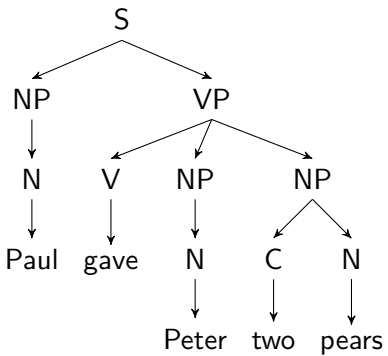
# Constituents vs. Dependencies

# Syntactic Structure

- Different shapes in different theories
- Typically a tree
  - Constituents (phrase tree structure)
  - Dependencies (dependency tree structure)

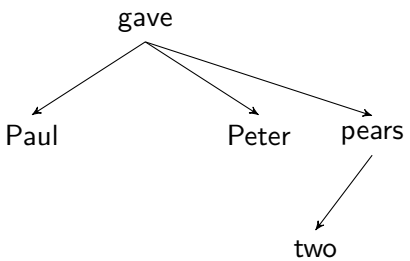
# Constituent Tree

(S (NP (N Paul)) (VP (V gave) (NP (N Peter)) (NP (C two) (N pears))))



# Dependency Tree

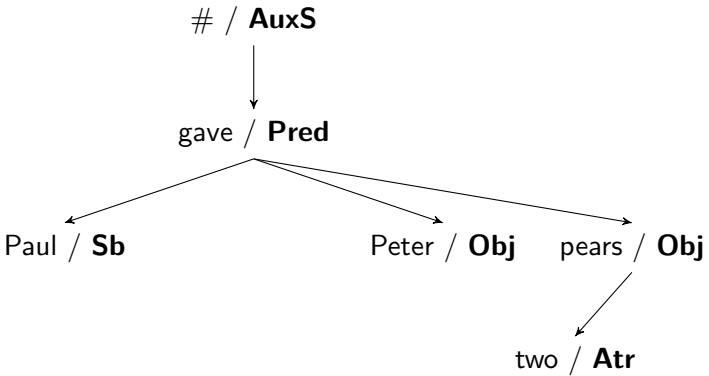
[gave,2] ( [Paul,1], [Peter,3], [pears,5] ( [two,4] ))



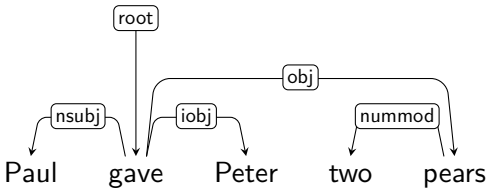

*Paul gave Peter two pears.*

# Dependency Tree with Labels

[#,0] ([gave,2] ([Paul,1], [Peter,3], [pears,5] ([two,4])), [.,6])



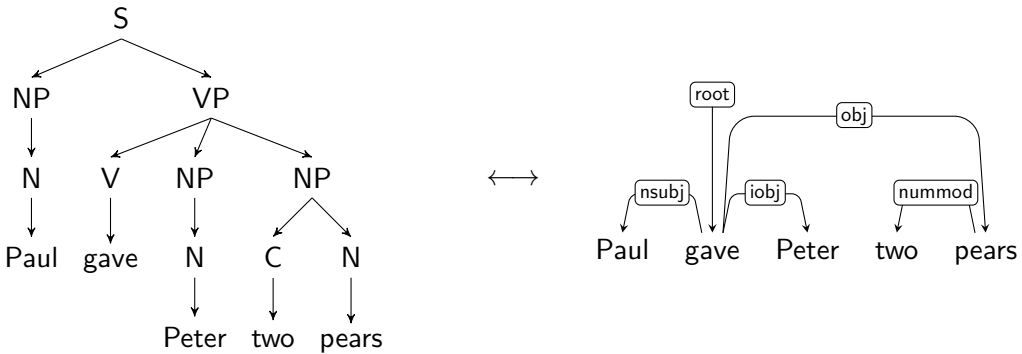
# Dependency Tree with Labels



# Constituents vs. Dependencies

- The two models are interconnected
- Sentence divided to phrases (constituents)
  - Recursive: phrases divided to smaller phrases
  - The smallest phrases are words
- There are dependencies (relations) between words (constituents)
  - Head of phrase = governing node, parent node
  - The other nodes are dependent nodes, children of the head

# Phrase vs. Dependency Trees



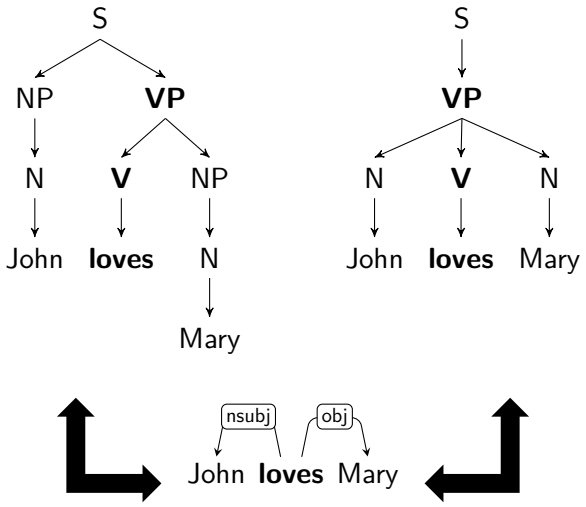
# Phrase vs. Dependency Trees

- Phrase trees
  - Usually do not mark the head
  - May not mark the function of the constituent in the superordinate constituent

# Phrase vs. Dependency Trees

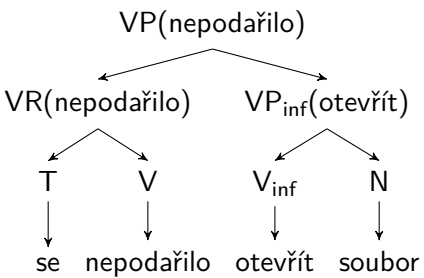
- Phrase trees
  - Usually do not mark the head
  - May not mark the function of the constituent in the superordinate constituent
- Dependency trees
  - Do not show nonterminals (phrase types)
    - Nor any other phrase-level features
  - Do not show "how the sentence is generated" (order, recursion, proximity of constituents)

# Example


S
NP
VP
N
V
NP
John
loves
N
Mary
S
VP
N
V
N
John
loves
Mary
nsubj
obj
John
loves
Mary

# Discontinuous Phrases

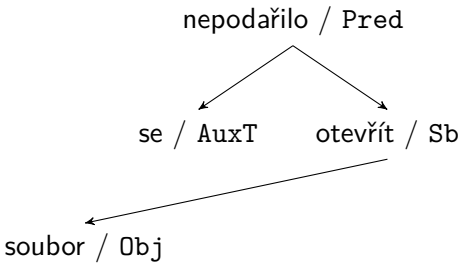
- Classical context-free grammar cannot describe them!
- They cannot be represented by bracketing.
- English example: *I found the best example ever.*
- Czech example: *(Soubor (se nepodařilo) otevřít).* "File couldn't be opened."



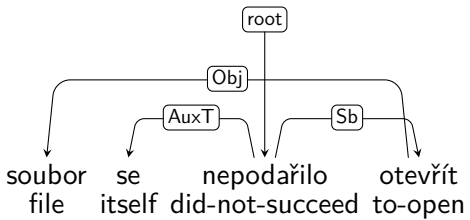
# Nonprojectivity

- Dependency tree including word order (x-coordinate of nodes).
- Projection to the base: the vertical from the node crosses a dependency (nonprojective edge).
- Formally:
  - Dependency $([g, x_g], [d, x_d])$ where $x_w$ is the order of the word $w$ in the sentence.
  - There exists a node $[n, x_n]$ that $x_g < x_n < x_d$ or $x_d < x_n < x_g$ and $[n, x_n]$ is not in subtree rooted by $[g, x_g]$.
- Informally: The string spanned by the subtree of the governing node is discontinuous, it contains gaps.

# Nonprojectivity Can Be Handled by a Dependency Tree!



# Nonprojectivity Can Be Handled by a Dependency Tree!

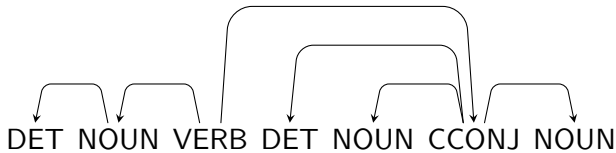


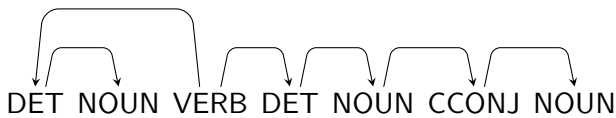
# Universal Dependencies

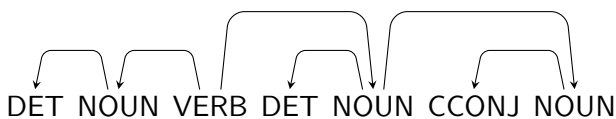
# Outline

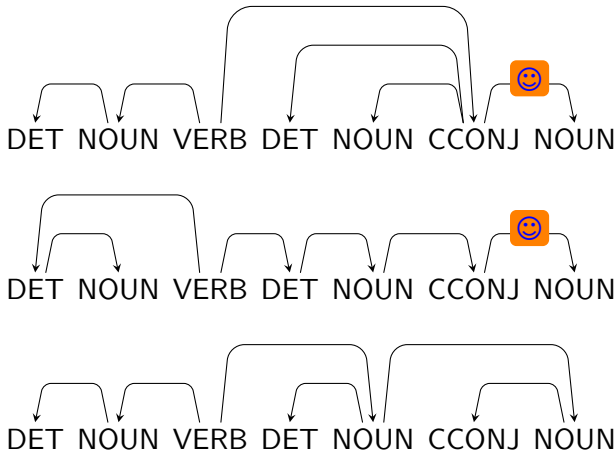
1 Constituents vs. Dependencies

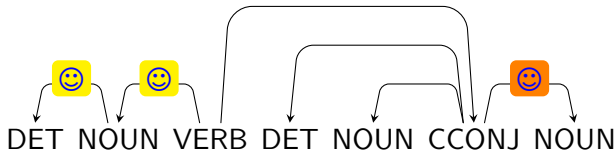
2 Universal Dependencies
- A Tour through UD Syntax
- Nonverbal Predicate and Copula
- Core Arguments vs. Oblique Dependents
- Ellipsis and Enhanced UD

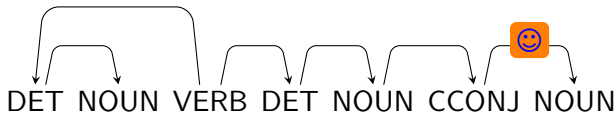

DET NOUN VERB DET NOUN CCONJ NOUN

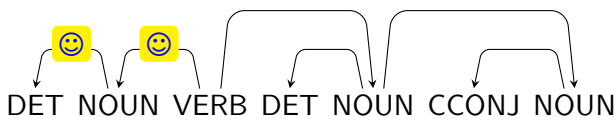

DET NOUN VERB DET NOUN CCONJ NOUN

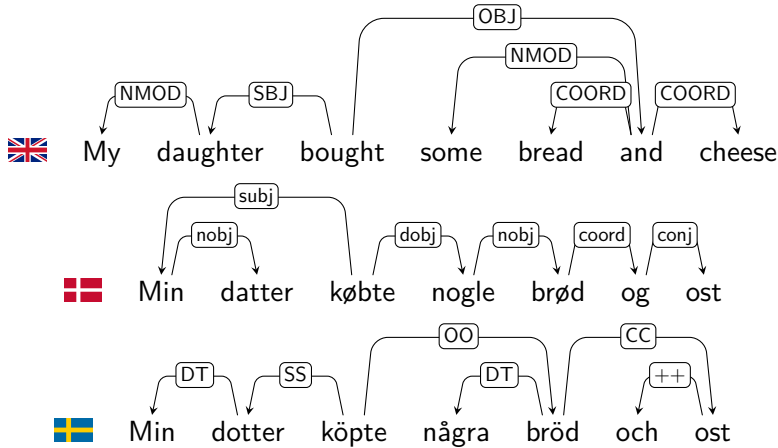

DET NOUN VERB DET NOUN CCONJ NOUN


DET NOUN VERB DET NOUN CCONJ NOUN
DET NOUN VERB DET NOUN CCONJ NOUN
DET NOUN VERB DET NOUN CCONJ NOUN


DET NOUN VERB DET NOUN CCONJ NOUN


DET NOUN VERB DET NOUN CCONJ NOUN


DET NOUN VERB DET NOUN CCONJ NOUN


OBJ
NMOD
NMOD
SBJ
COORD
COORD
My daughter bought some bread and cheese
subj
nobj
dobj
nobj
coord
conj
Min datter købte nogle brød og ost
OO
CC
DT
SS
DT
++
Min dotter köpte några bröd och ost

# Universal Dependencies

http://universaldependencies.org/

Milestones:

- 2008-05 Interset (morphological features)
- 2012-05 Google Universal POS tags
- 2012-05 HamleDT (harmonized Prague-style treebanks)
- 2013-08 Google Universal Dependency Treebank
- 2014-05 Universal Stanford Dependencies
- 2014-04 EACL Göteborg, kick-off meeting of UD
- 2014-10 UD guidelines version 1
- 2015-01 released first 10 treebanks
- every ~6 months new release
- 2016-12 UD guidelines version 2
- 2017-05 CoNLL Shared Task in parsing UD
- 2018-06 second Shared Task
- every ~6 months new release

# Universal Dependencies

- Same things annotated same way across languages...
- ... while highlighting different coding strategies

# Manning's Law

*The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:*

1. UD must be satisfactory on linguistic analysis grounds for individual languages.

*It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.*

# Manning's Law

*The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:*

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.

*It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.*

# Manning's Law

*The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:*

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.

*It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.*

# Manning's Law

*The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:*

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology.

*It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.*

# Manning's Law

*The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:*

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology.
5. UD must be suitable for computer parsing with high accuracy.

*It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.*

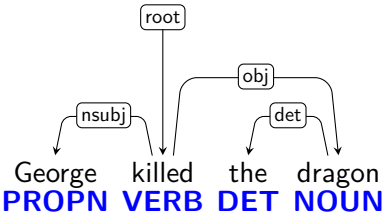
# Manning's Law

*The secret to understanding UD is to realize that the design is a very subtle compromise between approximately 6 things:*

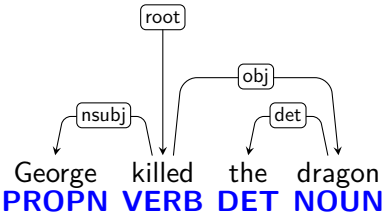
1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. ... it leads us to favor traditional grammar notions and terminology.
5. UD must be suitable for computer parsing with high accuracy.
6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...)

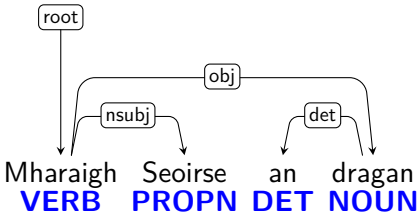
*It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.*

# Same Thing Same Way

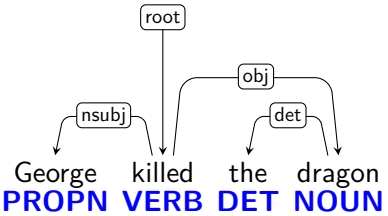


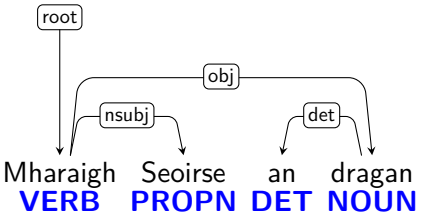
# Same Thing Same Way

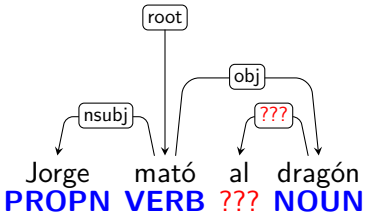


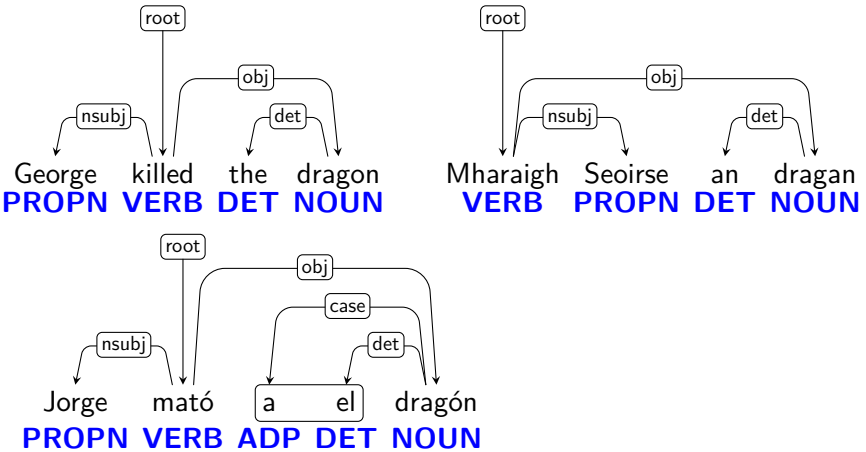
# Same Thing Same Way



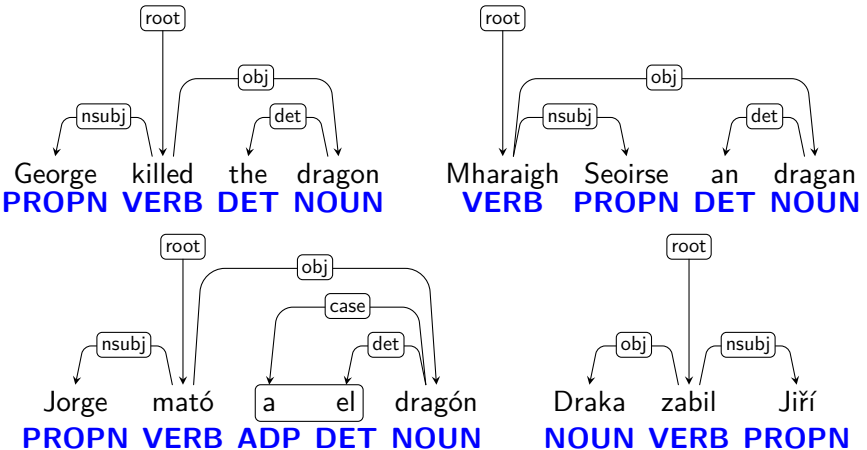



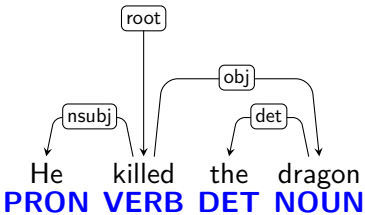
# Same Thing Same Way



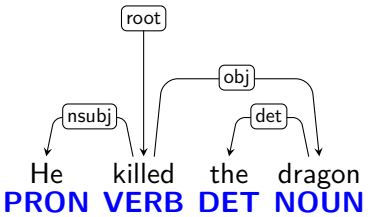
# Same Thing Same Way

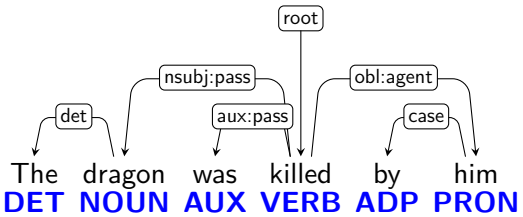


# Same Meaning $\neq$ Same Construction!

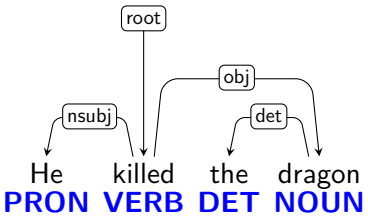


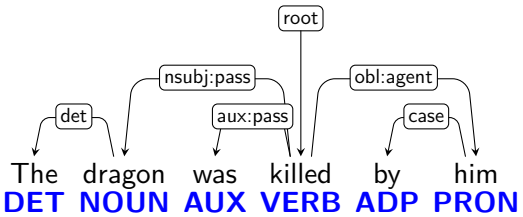
# Same Meaning $\neq$ Same Construction!

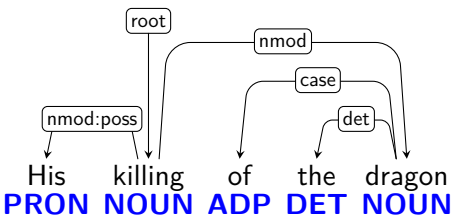


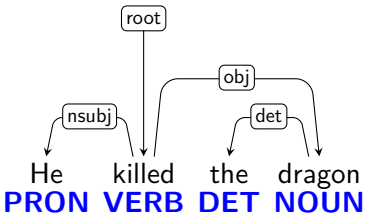
# Same Meaning ≠ Same Construction!

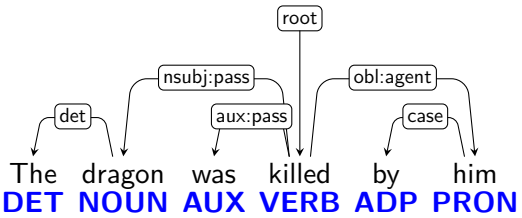


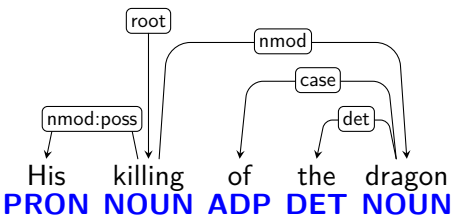



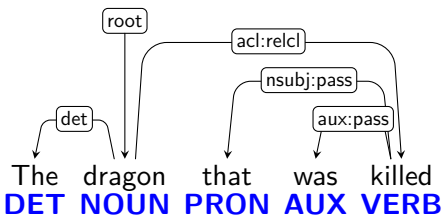



# Same Meaning $\neq$ Same Construction!

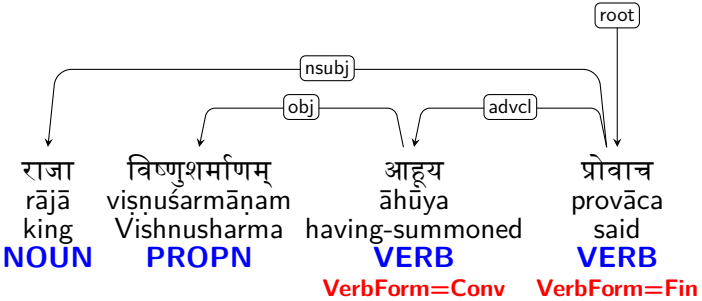


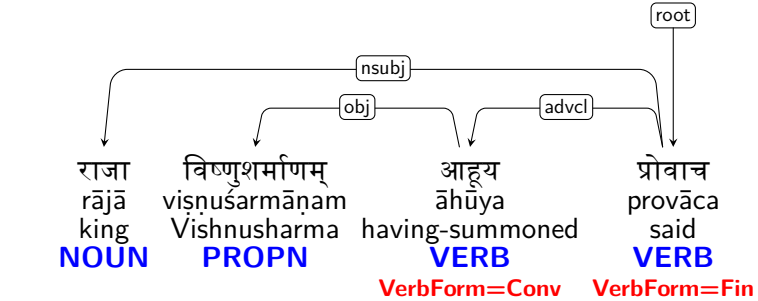



# Language-specific Preferences



# Language-specific Preferences



**Universal Dependencies**

# A Tour through UD Syntax

# Outline

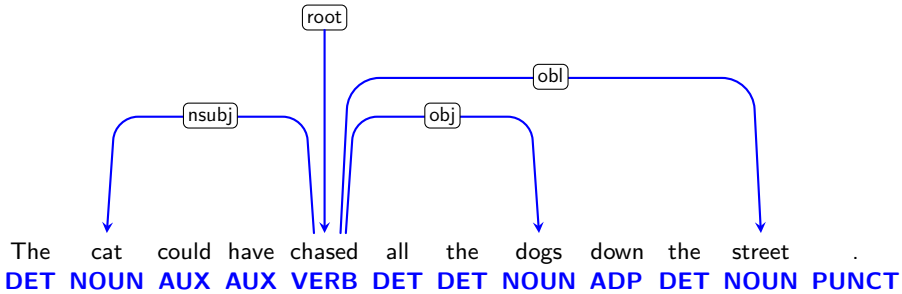
1 Constituents vs. Dependencies

2 Universal Dependencies
- A Tour through UD Syntax
- Nonverbal Predicate and Copula
- Core Arguments vs. Oblique Dependents
- Ellipsis and Enhanced UD

# Syntax

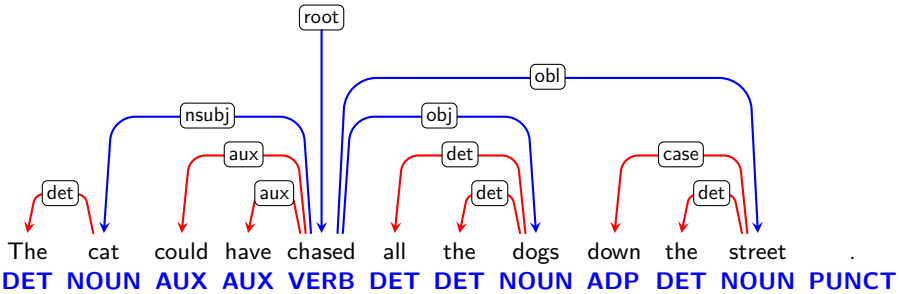
| The | cat | could | have | chased | all | the | dogs | down | the | street | . |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **DET** | **NOUN** | **AUX** | **AUX** | **VERB** | **DET** | **DET** | **NOUN** | **ADP** | **DET** | **NOUN** | **PUNCT** |

# Syntax



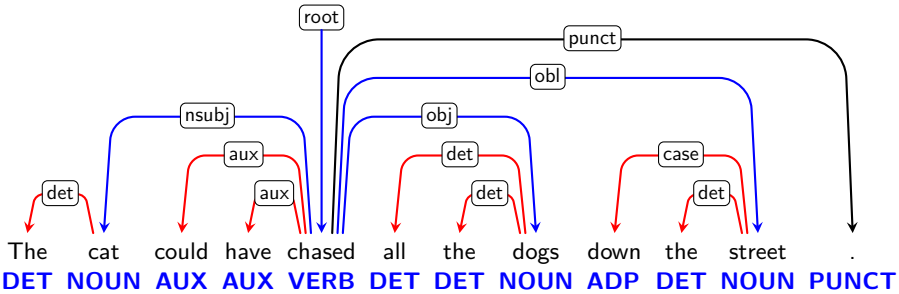

- Content words are related by dependency relations

# Syntax



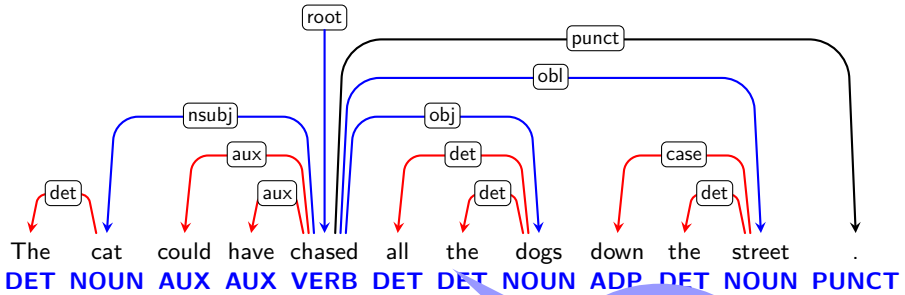

- Content words are related by dependency relations
- Function words attach to closest content words

# Syntax

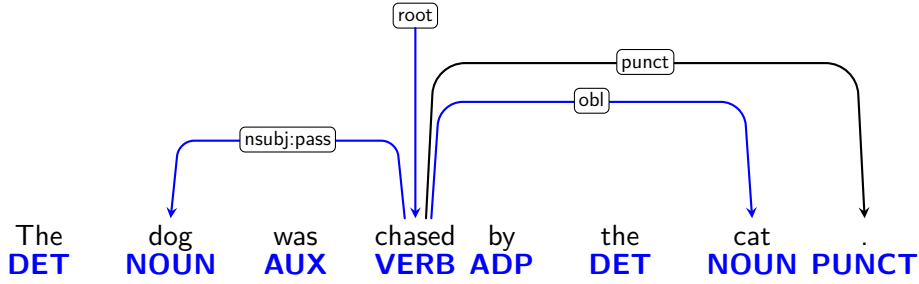


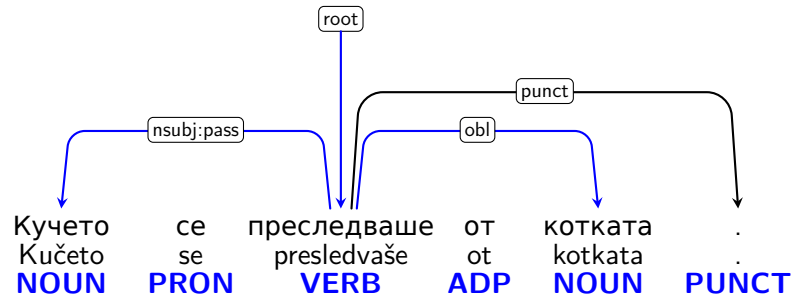

- Content words are related by dependency relations
- Function words attach to closest content words
- Punctuation attach to head of phrase or clause

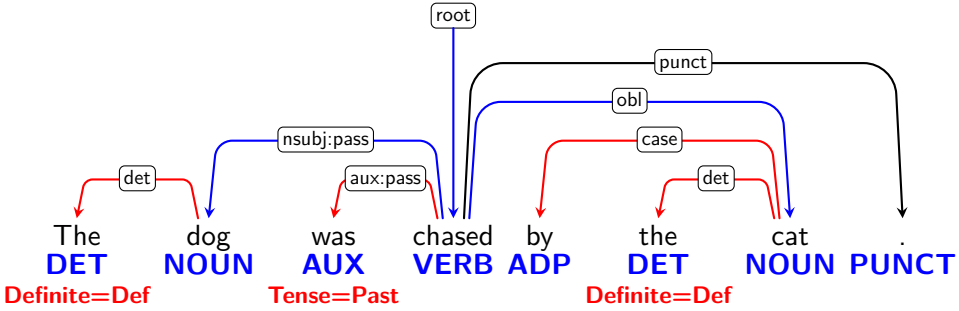
# Syntax

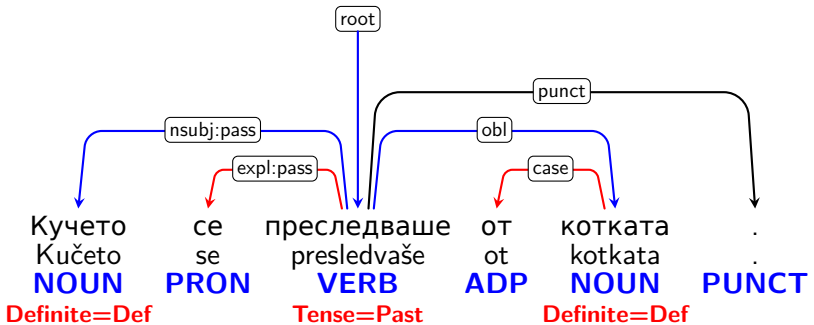


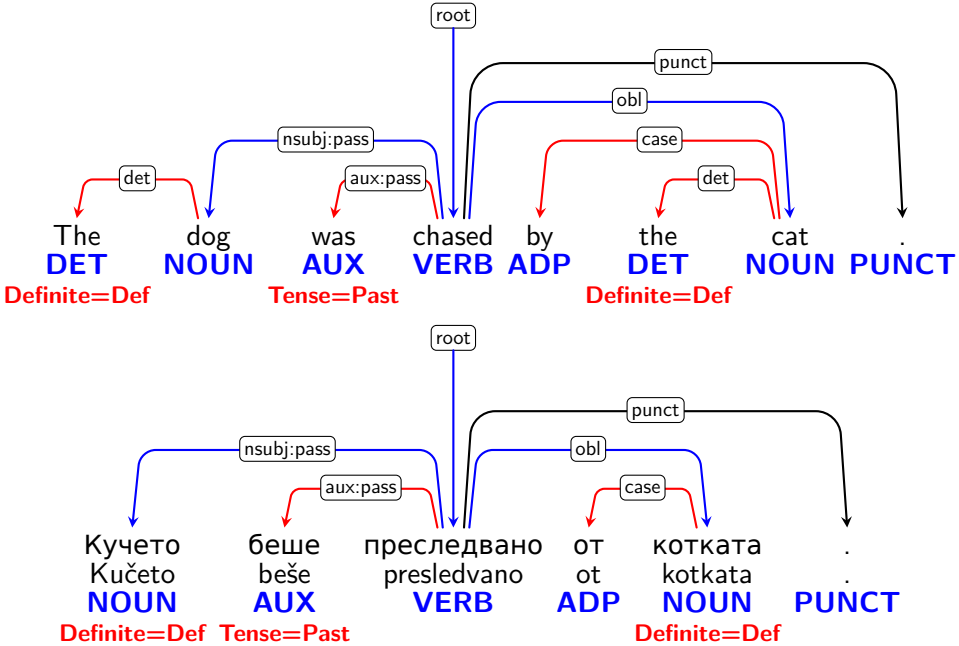

Not "dependency" in the strictly syntactic sense!

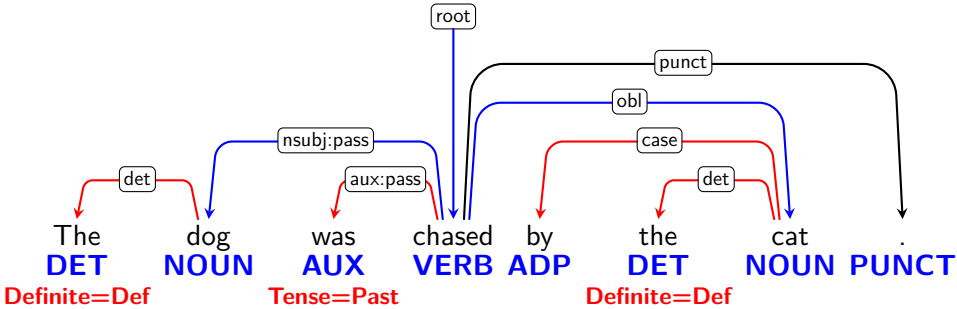

root
punct
obl
nsubj:pass
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT

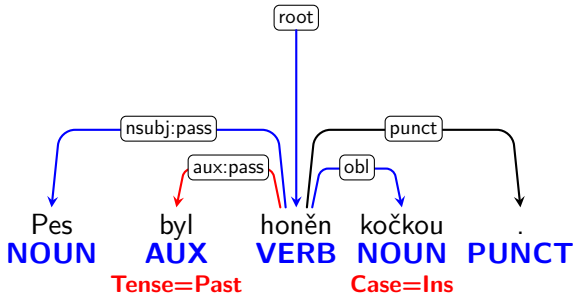

root
punct
nsubj:pass
obl
Кучето се преследваше от котката .
Kučeto se presledvaše ot kotkata .
NOUN PRON VERB ADP NOUN PUNCT


root
punct
obl
nsubj:pass
case
det
aux:pass
det
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def


root
punct
nsubj:pass
obl
expl:pass
case
Кучето се преследваше от котката .
Kučeto se presledvaše ot kotkata .
NOUN PRON VERB ADP NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def


root
punct
obl
nsubj:pass
case
det
aux:pass
det
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT
Definite=Def Tense=Past Definite=Def
root
punct
nsubj:pass
obl
aux:pass
case
Кучето беше преследвано от котката .
Kučeto beše presledvano ot kotkata .
NOUN AUX VERB ADP NOUN PUNCT
Definite=Def Tense=Past Definite=Def


root
punct
obl
nsubj:pass
case
det
aux:pass
det
The dog was chased by the cat .
DET NOUN AUX VERB ADP DET NOUN PUNCT
Definite=Def
Tense=Past
Definite=Def


root
nsubj:pass
punct
aux:pass
obl
Pes byl honěn kočkou .
NOUN AUX VERB NOUN PUNCT
Tense=Past
Case=Ins

## Dependents of Clauses (Verbal or Not)

| | Nominal | Clausal | Modifier | Function |
|---|---|---|---|---|
| **Core** | **nsubj** | csubj | | |
| **Non-Core** | **obl** | advcl | advmod | aux |
| | vocative | | discourse | cop |
| | dislocated | | | mark |
| | expl | | | |

## Dependents of Verbs, Adjectives and Adverbs

| | Nominal | Clausal | Modifier |
|---|---|---|---|
| **Core** | **obj** | ccomp | |
| | **iobj** | xcomp | |
| **Non-Core** | **obl** | advcl | advmod |
| | expl | | |

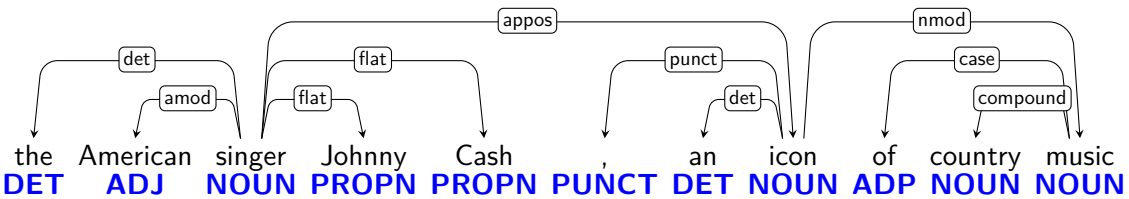
## Dependents of Nominals

| Nominal | Clausal | Modifier | Function |
|---|---|---|---|
| nmod | acl | amod | det |
| appos | | nummod | case |

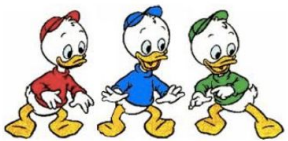
# Noun Phrase

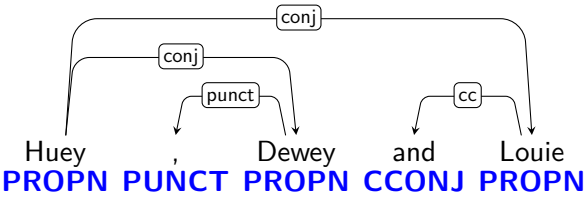
## Dependents of Nominals

| **Nominal** | **Clausal** | **Modifier** | **Function** |
|---|---|---|---|
| nmod | acl | amod | det |
| appos | | nummod | case |
| compound | | | clf |
| flat | | | |



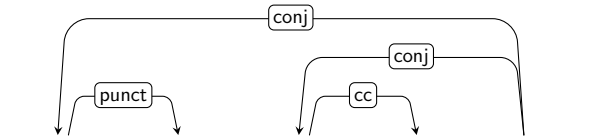
# Coordination






- Coordinate structures are headed by the first conjunct
  - Subsequent conjuncts depend on it via the conj relation
  - Conjunctions depend on the next conjunct via the cc relation
  - Punctuation marks depend on the next conjunct via the punct relation

# But Some Languages Might Prefer the Opposite



- Coordinate structures would be headed by the last conjunct
  - Preceding conjuncts would depend on it via the conj relation
  - Conjunctions would depend on the preceding conjunct
  - Punctuation marks would depend on the preceding conjunct

## Multiword Expressions

| ***Relation*** | ***Examples*** |
|---|---|
| *fixed* | *in spite of, as well as, ad hoc* |
| *flat* | *president Havel, New York, four thousand* |
| *compound* | *phone book, dress up* |
| *goeswith* | *notwith standing, with out* |

- UD annotation almost does not permit "words with spaces"
  - Multiword expressions are analyzed using special relations
  - The fixed, flat and goeswith relations are always head-initial
  - The compound relation reflects the internal structure
- Words with spaces may be allowed in v2:
  - Vietnamese (spaces delimit syllables, not words)
  - Numbers ("1 000 000")
  - Possibly other approved cases, e.g. multi-word abbreviations

## Other Relations

| ***Relation*** | ***Explanation*** |
|---|---|
| *parataxis* | *Loosely linked clauses of same rank* |
| *list* | *Lists without syntactic structure* |
| *orphan* | *Orphans in ellipsis linked together* |
| *reparandum* | *Disfluency linked to (speech) repair* |
| *foreign* | *Elements within opaque stretches of code switching* |
| *dep* | *Unspecified dependency* |
| *root* | *Syntactically independent element of clause/phrase* |

# Language-specific Relation Subtypes

- Language-specific relations are subtypes of universal relations added to capture important phenomena
- Subtyping permits us to "back off" to universal relations

## Language-specific Relation Subtypes

| Relation | Explanation |
|---|---|
| acl:relcl | Relative clause |
| compound:prt | Verb particle (dress up) |
| nmod:poss | Possessive nominal (Mary 's book) |
| obl:agent | Agent in passive (saved by the bell) |
| cc:preconj | Preconjunction (both ... and) |
| det:predet | Predeterminer (all those ...) |

**Universal Dependencies**

# Nonverbal Predicate and Copula

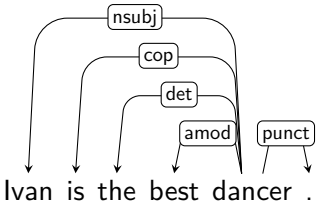
# Outline

1 Constituents vs. Dependencies

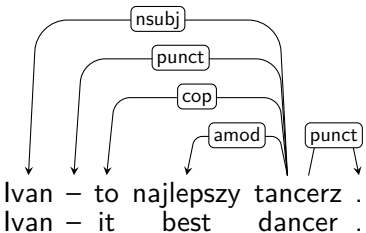
2 Universal Dependencies
- A Tour through UD Syntax
- Nonverbal Predicate and Copula
- Core Arguments vs. Oblique Dependents
- Ellipsis and Enhanced UD

# Nonverbal Predicate and Copula

- Some languages use a copula verb:

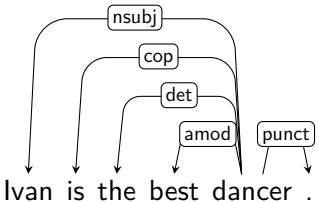



- Some languages use a copula pronoun:

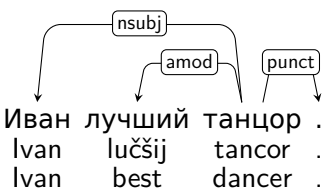


# Nonverbal Predicate and Copula

- Some languages use a copula verb:

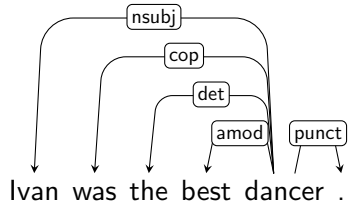



- Some languages omit the copula:

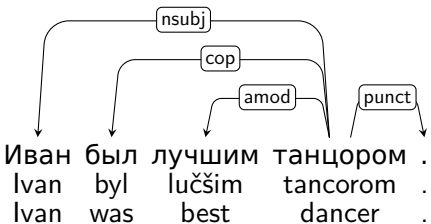


# Nonverbal Predicate and Copula

- Some languages use a copula verb:

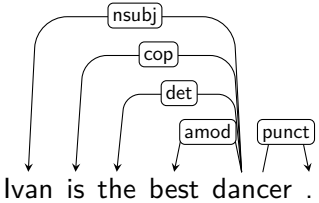



- Some languages use it only in some tenses:

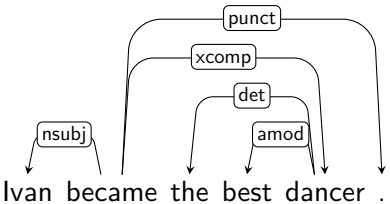


# Copula Verbs: We Are Restrictive!

- 🇬🇧 *To be* is copula:

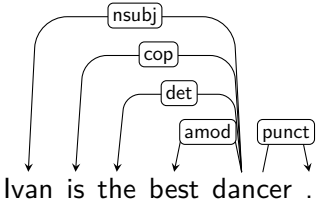



- 🇬🇧 *To become* is not copula:

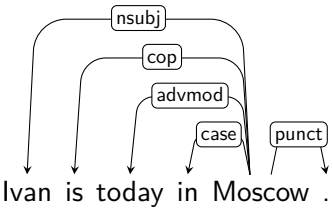


# Once Copula, Always Copula!

- 🇬🇧 This is parallel with Russian:

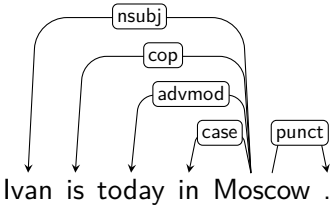



- 🇬🇧 This is also parallel with Russian:

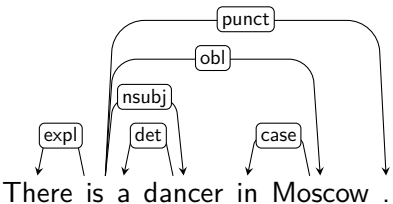


# Well, Almost...

- 🇬🇧 This is parallel with Russian:

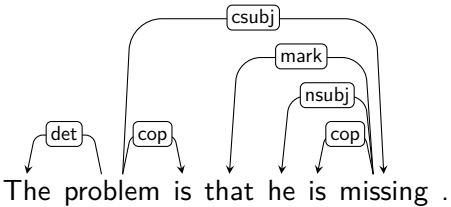



- 🇬🇧 But not with this in English:

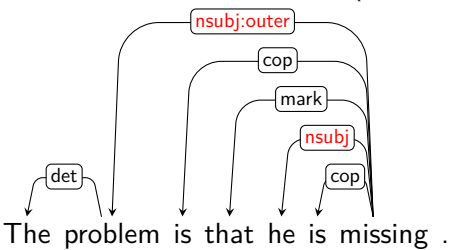


# Clauses and Copula

- A clause can be the subject:




- Or it can be annotated as the nonverbal predicate (note the two subjects):



**Universal Dependencies**

# Core Arguments vs. Oblique Dependents

# Outline

1 Constituents vs. Dependencies

2 Universal Dependencies
- A Tour through UD Syntax
- Nonverbal Predicate and Copula
- Core Arguments vs. Oblique Dependents
- Ellipsis and Enhanced UD

## Dependents of Clauses (Verbal or Not)

| | Nominal | Clausal | Modifier | Function |
|---|---|---|---|---|
| **Core** | **nsubj** | csubj | | |
| **Non-Core** | **obl** | advcl | advmod | aux |
| | vocative | | discourse | cop |
| | dislocated | | | mark |
| | expl | | | |

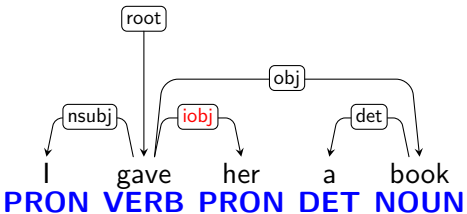
## Dependents of Verbs, Adjectives and Adverbs

| | Nominal | Clausal | Modifier |
|---|---|---|---|
| **Core** | **obj** | ccomp | |
| | **iobj** | xcomp | |
| **Non-Core** | **obl** | advcl | advmod |
| | expl | | |

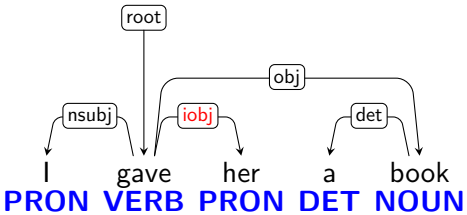
## Dependents of Nominals

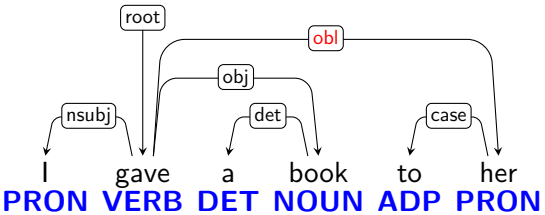
| Nominal | Clausal | Modifier | Function |
|---|---|---|---|
| nmod | acl | amod | det |
| appos | | nummod | case |

# Information Packaging

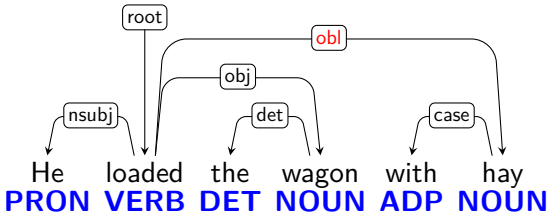


# Information Packaging

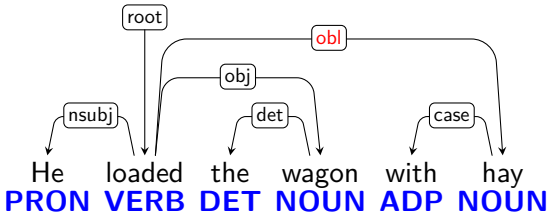


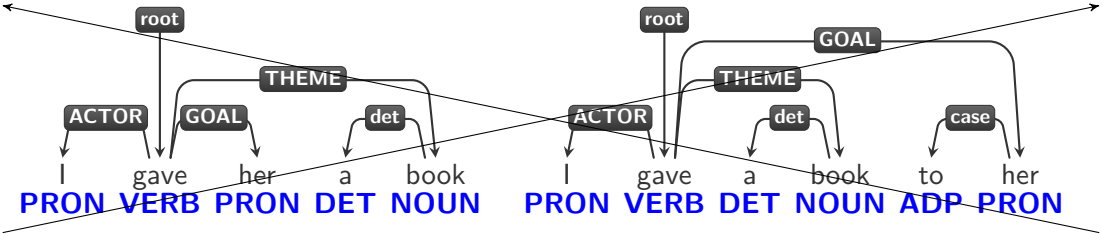
# Information Packaging



# Information Packaging



# UD is NOT about Semantic Roles!



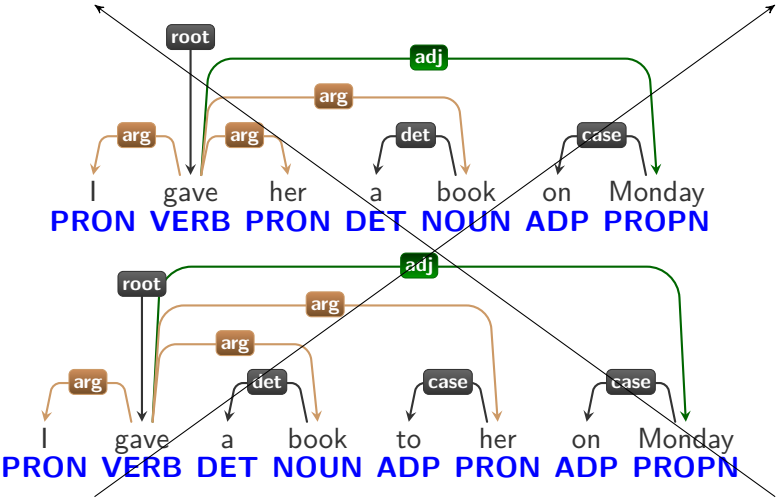
# Manning's Law – What If We Do Semantic Roles?

*The secret to understanding the design and current success of UD is to realize that the design is a very subtle compromise between approximately 6 things:*

1. UD must be satisfactory on linguistic analysis grounds for individual languages.
2. UD must be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. … it leads us to favor traditional grammar notions and terminology.
5. UD must be suitable for computer parsing with high accuracy.
6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, …)

*It's easy to come up with a proposal that improves UD on one of these dimensions. The interesting and difficult part is to improve UD while remaining sensitive to all these dimensions.*

# UD Avoids Argument-Adjunct Distinction!



# Avoiding an Argument-Adjunct Distinction

- From the guidelines:
  - Subtle, unclear, and frequently argued over
  - Questionable as a categorical distinction
  - Best practical solution is to eliminate it

# Avoiding an Argument-Adjunct Distinction

- From the guidelines:
  - Subtle, unclear, and frequently argued over
  - Questionable as a categorical distinction
  - Best practical solution is to eliminate it

- BUT:
  - Cannot be eliminated completely
  - Some people/data have it and want to keep it
    - It aligns well with traditional grammars
  - $\Rightarrow$ there is now a relation subtype `obl:arg`

# Avoiding an Argument-Adjunct Distinction

- From the guidelines:
  - Subtle, unclear, and frequently argued over
  - Questionable as a categorical distinction
  - Best practical solution is to eliminate it
- BUT:
  - Cannot be eliminated completely
  - Some people/data have it and want to keep it
    - It aligns well with traditional grammars
  - $\Rightarrow$ there is now a relation subtype `obl:arg`
- AND I will argue that
  - Core-oblique distinction is unclear and argued over too
  - (Though I will not propose to discard it.)

# So What Is Core and Why?

# Community Confusion

- UD v1 guidelines took core-oblique for granted
- English (simplified):
  - Bare noun phrase ⇒ core argument (`nsubj`, `obj`, `iobj`)
  - Prepositional phrase ⇒ oblique argument or adjunct (`obl`)

# Community Confusion

- UD v1 guidelines took core-oblique for granted
- English (simplified):
  - Bare noun phrase ⇒ core argument (`nsubj`, `obj`, `iobj`)
  - Prepositional phrase ⇒ oblique argument or adjunct (`obl`)
- Other languages: not necessarily! (Spanish, Japanese)
  - But some people simply took the English rule...
  - Manning's law: non-linguists!

# Community Confusion

- UD v1 guidelines took core-oblique for granted
- English (simplified):
  - Bare noun phrase ⇒ core argument (`nsubj`, `obj`, `iobj`)
  - Prepositional phrase ⇒ oblique argument or adjunct (`obl`)
- Other languages: not necessarily! (Spanish, Japanese)
  - But some people simply took the English rule...
  - Manning's law: non-linguists!
- Clash with traditional terminology
  - Grammars of German, Czech etc. define prepositional objects
  - But these are not necessarily core...
  - Yet some people took their national definition of object...

# Language-specific Coding Strategy

- Idea:
  - Oblique arguments are marked similarly to adjuncts (prepositions, certain morphological cases...)
  - Core arguments are marked differently
    - $\Rightarrow$ easy for annotators and non-linguists!

- Why are core arguments special?
  - They tend to be targeted by grammatical rules
    - Passivization
    - Control verbs
    - Reflexives
    - ...

# Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

# Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

- Andrews, 2007 (In Shopen: Language Typology)
  - Identify primary transitive predicates
    - We need semantic roles for this! (One-time only.)

# Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

- Andrews, 2007 (In Shopen: Language Typology)
  - Identify primary transitive predicates
    - We need semantic roles for this! (One-time only.)
  - Actor/agent = function A
  - Undergoer/patient = function P

# Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

- Andrews, 2007 (In Shopen: Language Typology)
  - Identify primary transitive predicates
    - We need semantic roles for this! (One-time only.)
  - Actor/agent = function A
  - Undergoer/patient = function P
  - Note the way they are coded
  - Note other grammatical rules that target them
  - Generalize to other predicates with same coding and rules

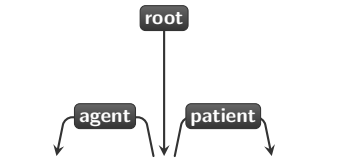
# Language-specific Coding Strategy

- Core vs. oblique is not defined in traditional grammar
- How shall we define it?

- Andrews, 2007 (In Shopen: Language Typology)
  - Identify primary transitive predicates
    - We need semantic roles for this! (One-time only.)
  - Actor/agent = function A
  - Undergoer/patient = function P
  - Note the way they are coded
  - Note other grammatical rules that target them
  - Generalize to other predicates with same coding and rules

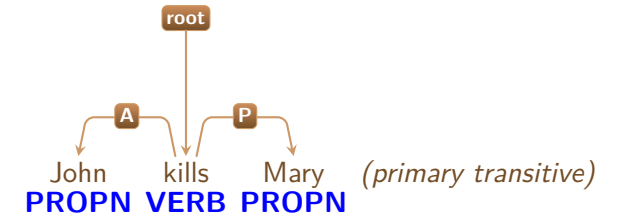
  - Then define:
    - function A ⇒ `nsubj`
    - function P ⇒ `obj`

# Transitive Predicates in English

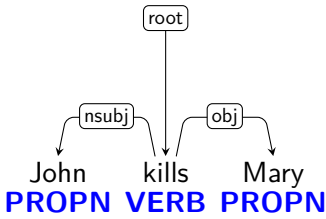



*(primary transitive)*

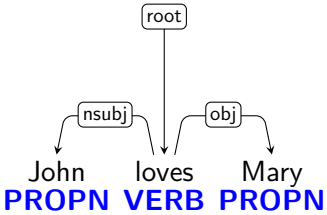
# Transitive Predicates in English



# Transitive Predicates in English

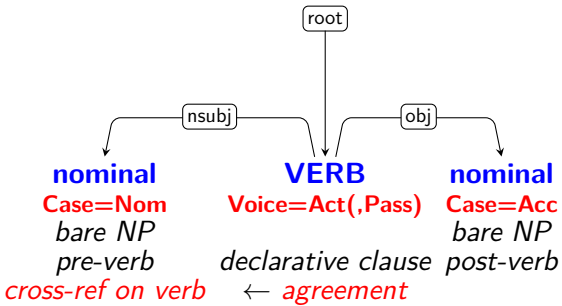



*(primary transitive)*

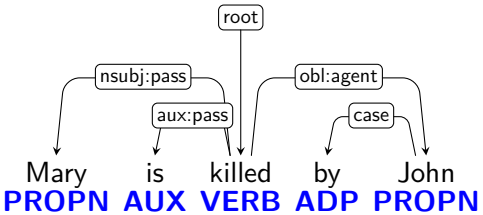


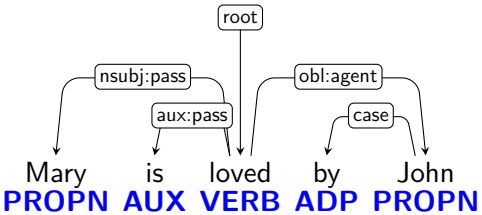

*(generalized transitive)*

# Transitive Predicates in English

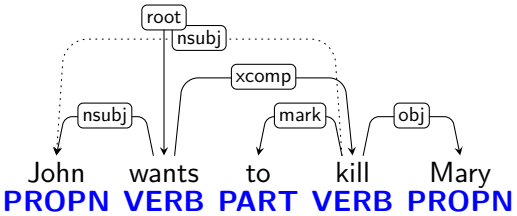


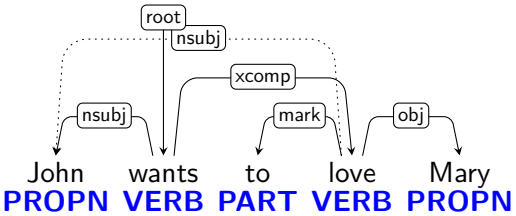
# Passivization in English



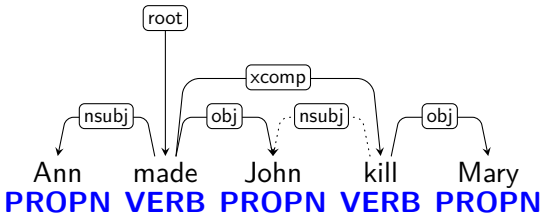



# Subject Control in English



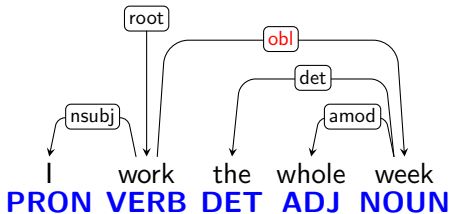



# Object Control in English



# Some Problems

- Some temporal adjuncts are bare noun phrases
  - *I work the whole week.*
  - *I work every Friday.*




- At least it cannot passivize:
  - **The whole week is worked by me.*
  - **Every Friday is worked by me.*
- But...

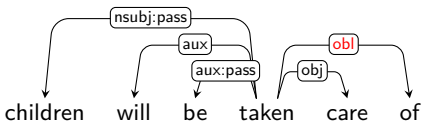
# Some Problems

- Some transitive verbs cannot passivize
  - *John has a new car.*
    - **A new car is had by John.*
  - *Friday does not suit me.*
    - **I am not suited by Friday.*

# Some Problems

- Some transitive verbs cannot passivize
  - *John has a new car.*
    - **A new car is had by John.*
  - *Friday does not suit me.*
    - **I am not suited by Friday.*

- Some prepositional verbs can passivize
  - *You can rely on Ben.*
    - *Ben can be relied on.*
  - *They will take care of your children.*
    - *Your children will be taken care of.*



# Tentative Summary?

- The borderline is inherently fuzzy
- No universally applicable and exact algorithm
- Better described in terms of probability

# Tentative Summary?

- The borderline is inherently fuzzy
- No universally applicable and exact algorithm
- Better described in terms of probability

- Core coding not favored by adjuncts
- Oblique coding similar to most adjuncts
- Passivization etc. may help...
- ... but does not work as strict criterion

# Tentative Summary?

- The borderline is inherently fuzzy
- No universally applicable and exact algorithm
- Better described in terms of probability

- Core coding not favored by adjuncts
- Oblique coding similar to most adjuncts
- Passivization etc. may help...
- ... but does not work as strict criterion

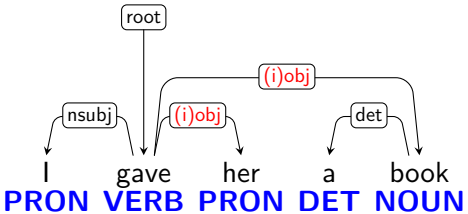
- Semantic roles needed when starting a new language
- Argument-adjunct needed to describe exceptions *(the whole week)*

# Intransitive Predicates

- Just one core argument
  - We already “know” how to find out if there are two
- ⇒ function S
  - Regardless of semantic role:
    - *John runs.*
    - *John sleeps.*
    - *John falls.*
- Then define:
  - function S ⇒ `nsubj`

# Ditransitive Predicates

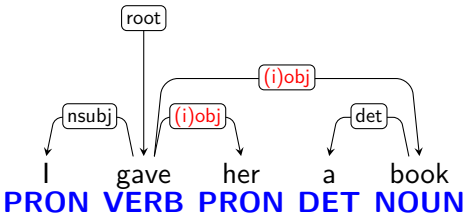
- Three core arguments
- Is one of them “least core”? ⇒ **`iobj`**
- (Alternatively, we could look at the semantic roles once again.)




- Passivization:
  - *She was given a book by me.*
  - *?A book was given her by me.*

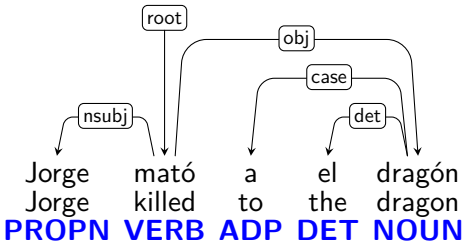
# Ditransitive Predicates

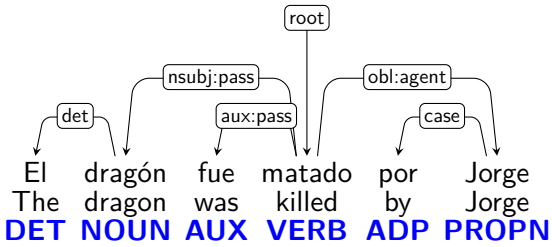
- Three core arguments
- Is one of them "least core"? ⇒ `iobj`
- (Alternatively, we could look at the semantic roles once again.)



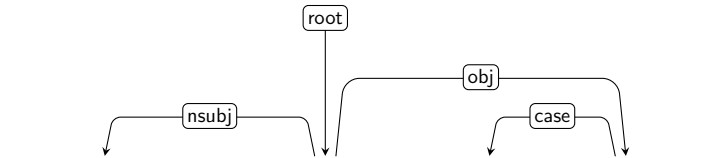

- Andrews (2007): *the status of the notion of 'indirect object' is problematic and difficult to sort out. The top priority is to work out what properties recipients and themes do and do not share with P arguments of primary transitive verbs.*

# Spanish

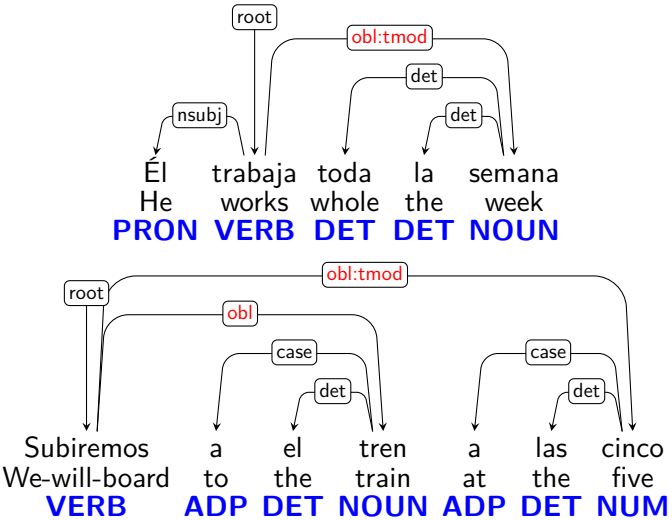

root
obj
case
nsubj
det
Jorge mató a el dragón
Jorge killed to the dragon
PROPN VERB ADP DET NOUN


root
nsubj:pass
obl:agent
det
aux:pass
case
El dragón fue matado por Jorge
The dragon was killed by Jorge
DET NOUN AUX VERB ADP PROPN

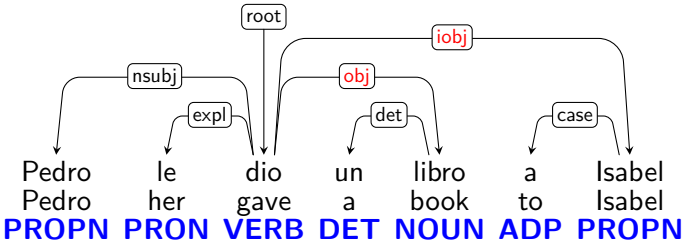
# Spanish Transitive Clauses

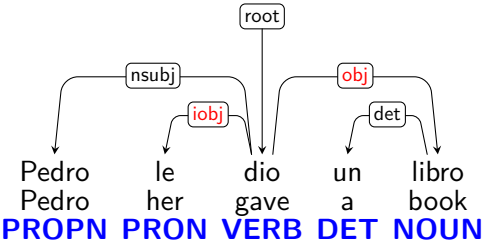


# Spanish Adjunct Exceptions

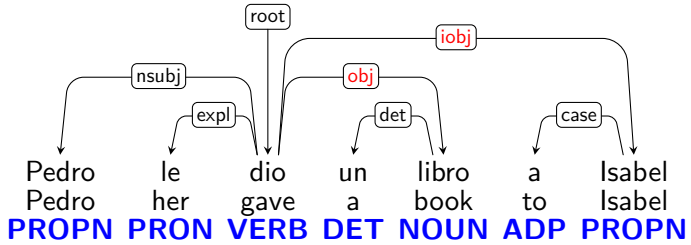


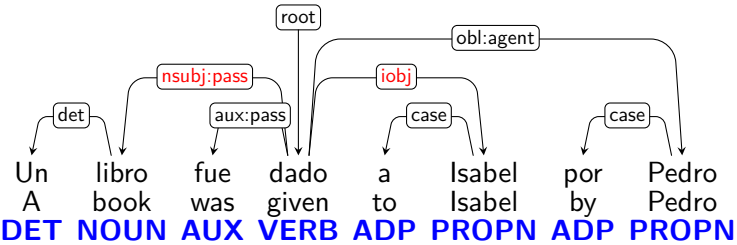
# Spanish Ditransitive Clauses



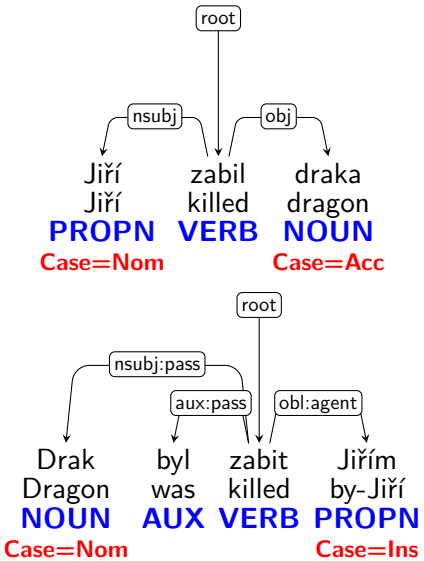



# Spanish Ditransitive Clauses

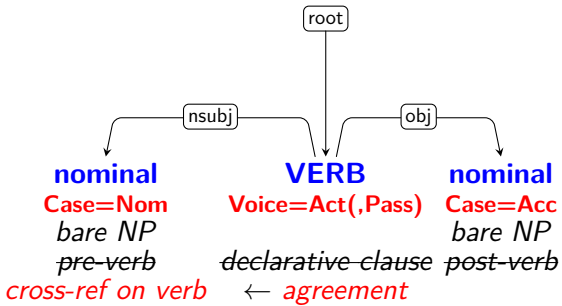


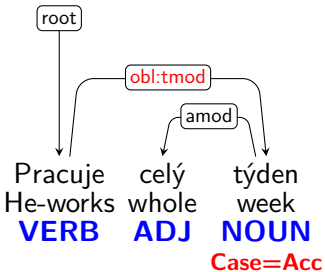
# Czech


root
nsubj
obj
Jiří
Jiří
PROPN
Case=Nom
zabil
killed
VERB
draka
dragon
NOUN
Case=Acc
root
nsubj:pass
aux:pass
obl:agent
Drak
Dragon
NOUN
Case=Nom
byl
was
AUX
zabit
killed
VERB
Jiřím
by-Jiří
PROPN
Case=Ins

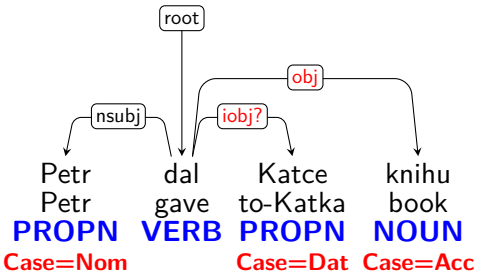
# Czech Transitive Clauses

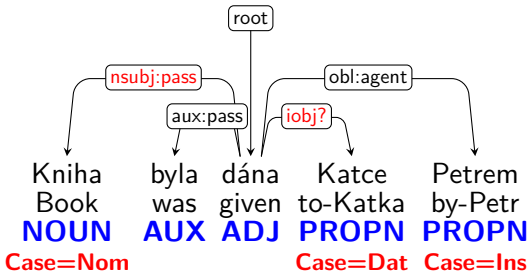


# Czech Adjunct Exceptions

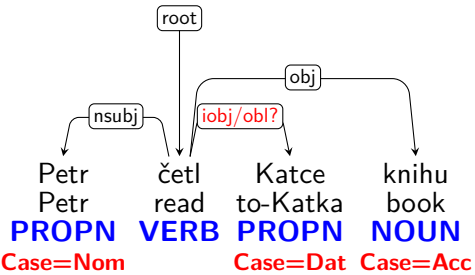


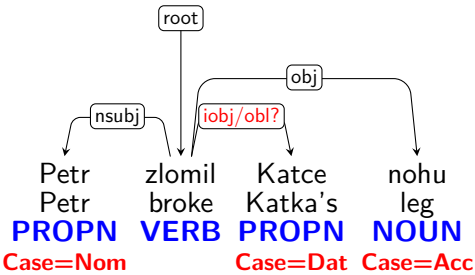
# Czech Ditransitive Clauses



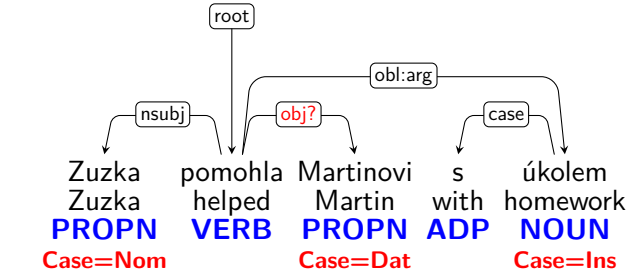



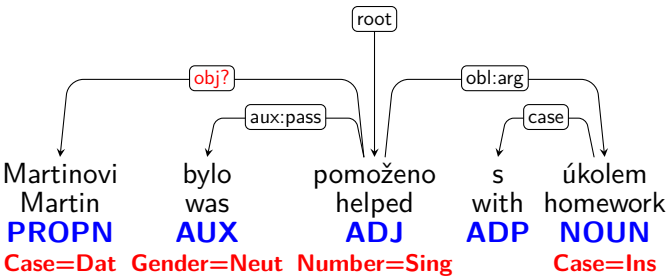
# Dative: Recipient vs. Beneficiary



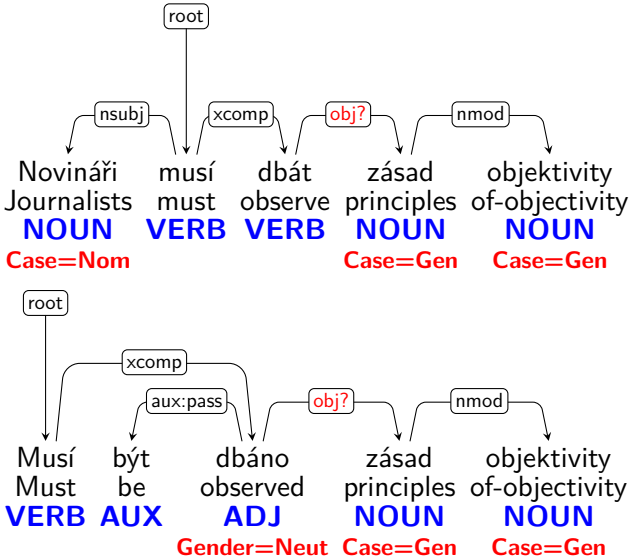



# Monotransitive with Dative?

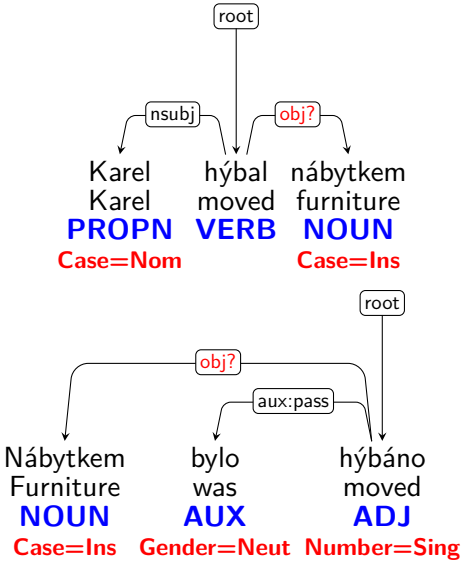


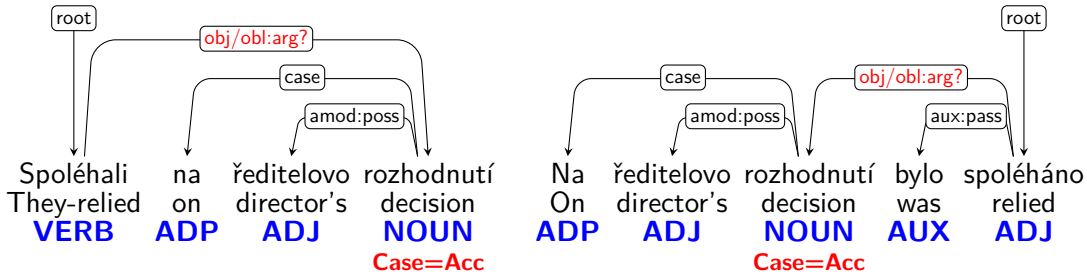
# Monotransitive with Genitive?



# Monotransitive with Instrumental?



# Monotransitive with Preposition?



# Tentative Summary 2

- There is a core-oblique scale:
- **Nom** > **Acc** > **Gen,Dat** > **Ins** > **preposition**

- Where is the borderline?

# Tentative Summary 2

- There is a core-oblique scale:
- **Nom** > **Acc** > **Gen,Dat** > **Ins** > **preposition**
- Where is the borderline?
- UD Czech 1.0: object = argument
  - Nom, Acc, Gen, Dat, Ins, ADP > "adverbial"

# Tentative Summary 2

- There is a core-oblique scale:
- **Nom** > **Acc** > **Gen,Dat** > **Ins** > **preposition**
- Where is the borderline?
- UD Czech 1.0: object = argument
  - Nom, Acc, Gen, Dat, Ins, ADP > "adverbial"
- UD Czech 2.1–2.5: bare NP > PP
  - Nom, Acc, Gen, Dat, Ins > ADP + adjuncts

# Tentative Summary 2

- There is a core-oblique scale:
- **Nom** > **Acc** > **Gen,Dat** > **Ins** > **preposition**

- Where is the borderline?

- UD Czech 1.0: object = argument
  - Nom, Acc, Gen, Dat, Ins, ADP > "adverbial"
- UD Czech 2.1–2.5: bare NP > PP
  - Nom, Acc, Gen, Dat, Ins > ADP + adjuncts
- UD Czech 2.6 (May 2020):
  - Nom, Acc > Gen, Dat, Ins, ADP + adjuncts

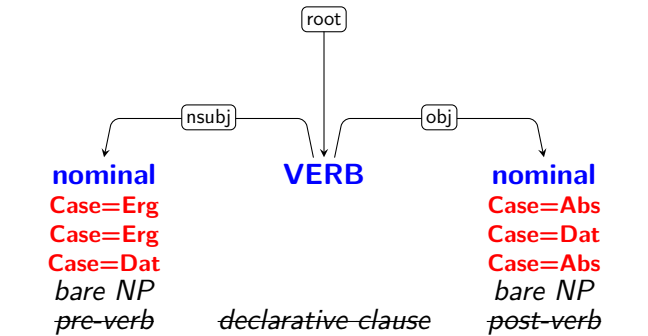
# Tentative Summary 2

- There is a core-oblique scale:
- **Nom** > **Acc** > **Gen,Dat** > **Ins** > **preposition**

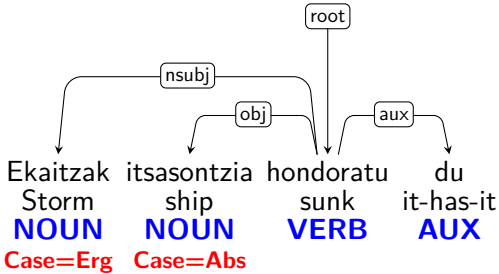
- Where is the borderline?

- UD Czech 1.0: object = argument
  - Nom, Acc, Gen, Dat, Ins, ADP > "adverbial"
- UD Czech 2.1–2.5: bare NP > PP
  - Nom, Acc, Gen, Dat, Ins > ADP + adjuncts
- UD Czech 2.6 (May 2020):
  - Nom, Acc > Gen, Dat, Ins, ADP + adjuncts
  - ⇒ No ditransitives in Czech!
  - (Exception: *učit* "to teach" takes two Acc.)

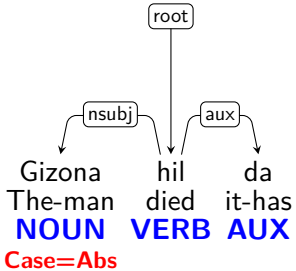
# Basque Transitive Clauses



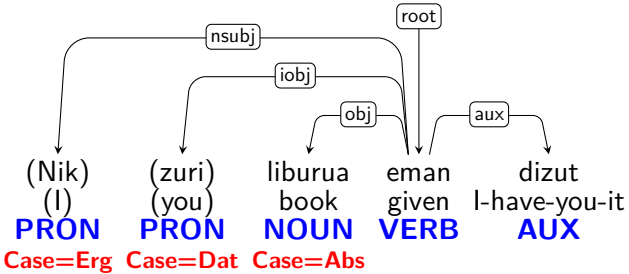
# Basque Transitive Clauses



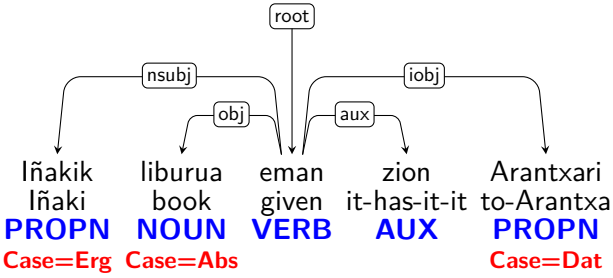
# Basque Intransitive Clauses



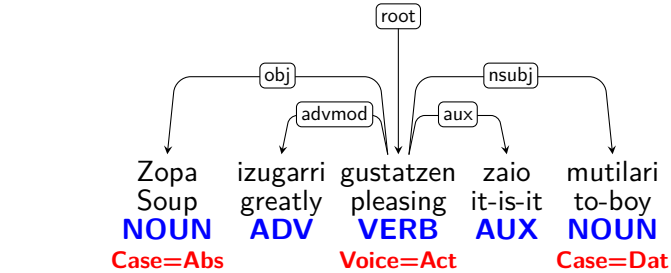
# Basque Ditransitive Clauses



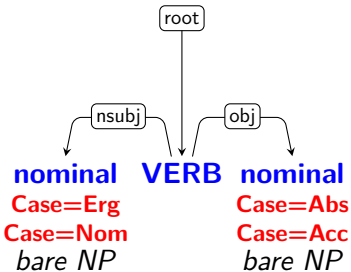
# Basque Ditransitive Clauses



# Basque Causative Applied to Dative Subject

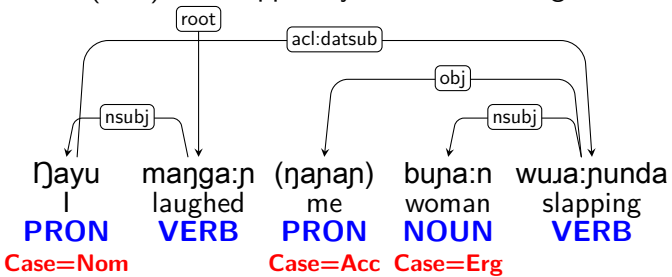


# Yidiɲ **Transitive Clauses**



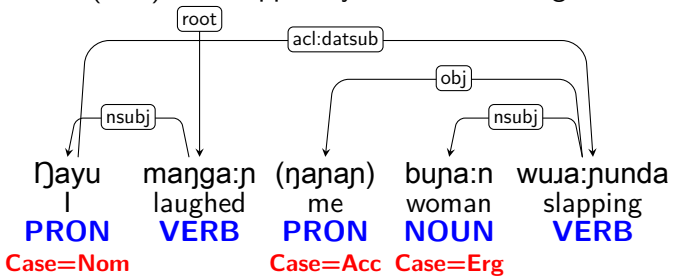
# Yidiɲ “Dative” Adnominal Clauses

“I, (who) was slapped by the woman, laughed”



# Yidiɲ “Dative” Adnominal Clauses

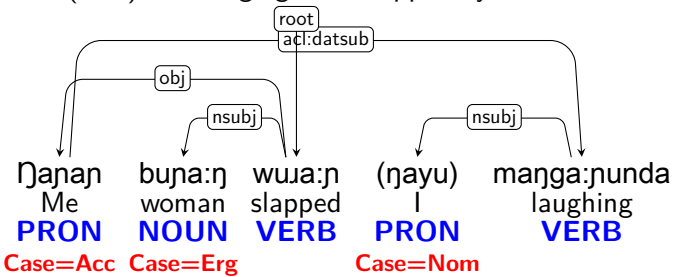
“I, (who) was slapped by the woman, laughed”




The coreferential (and elidable) NP must have S or P function.

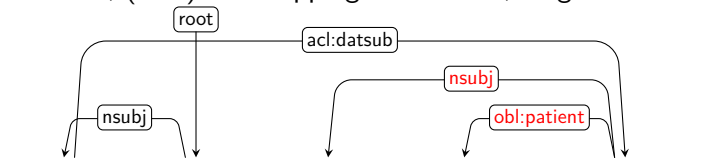
# Yidiɲ “Dative” Adnominal Clauses

“I, (who) was lauging, was slapped by the woman”



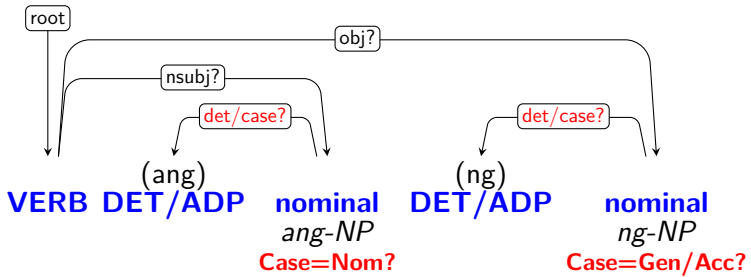

The coreferential (and elidable) NP must have S or P function.

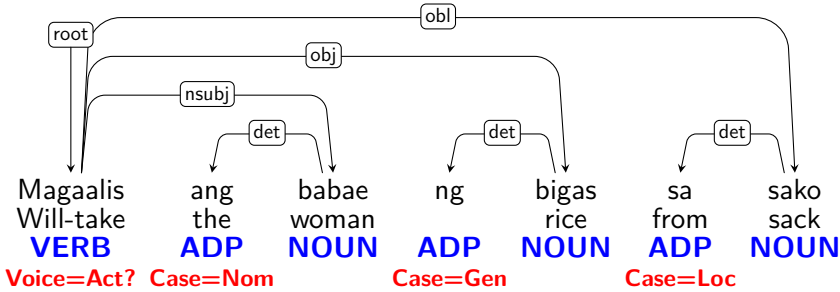
"I, (who) was slapping the woman, laughed"



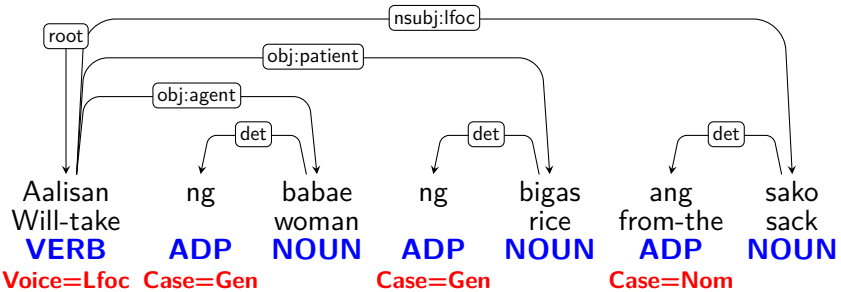

Original P is now oblique and original A is now S.

# Tagalog Transitive Clauses

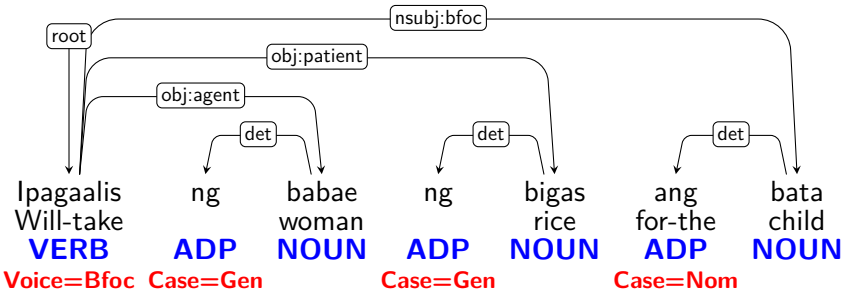



root
obl
obj
nsubj
det
det
det
Magaalis ang babae ng bigas sa sako
Will-take the woman rice from sack
VERB ADP NOUN ADP NOUN ADP NOUN
Voice=Act? Case=Nom Case=Gen Case=Loc
root
obl
nsubj:pass
obj:agent
det
det
det
Aalisin ng babae ang bigas sa sako
Will-take woman the rice from sack
VERB ADP NOUN ADP NOUN ADP NOUN
Voice=Pass? Case=Gen Case=Nom Case=Loc

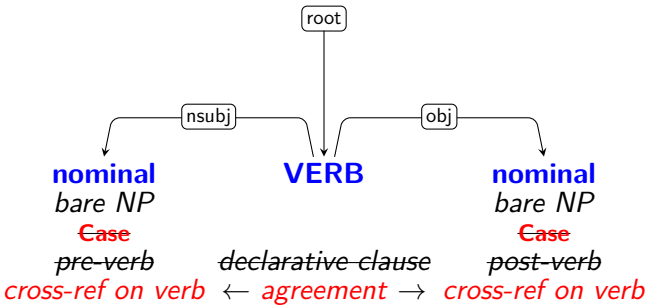
# Tagalog Locative Voice ⇒ Ditransitive!



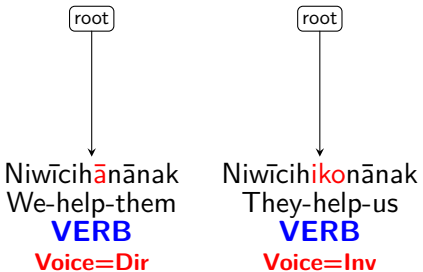
# Tagalog Benefactive Voice ⇒ Ditransitive!



# Plains Cree Transitive Clauses

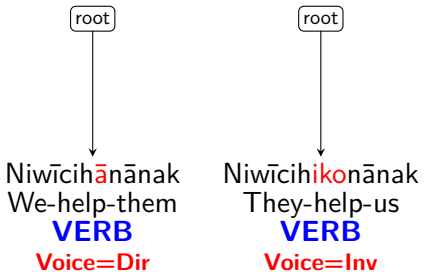


# Direct-Inverse Voice in Plains Cree




Animacy hierarchy: 1st person $>$ 3rd person

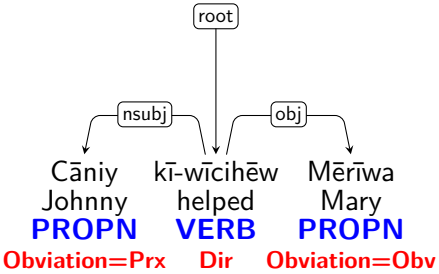
# Direct-Inverse Voice in Plains Cree

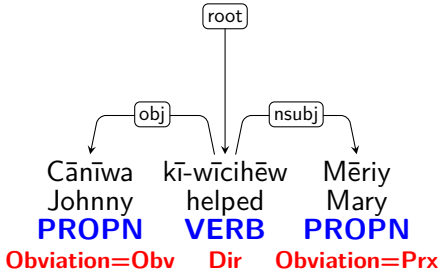


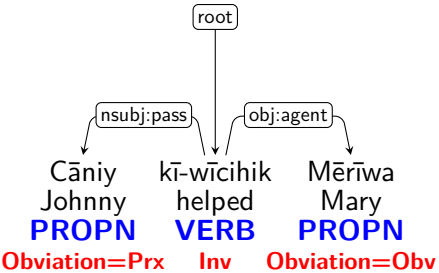

Animacy hierarchy: 1st person > 3rd person

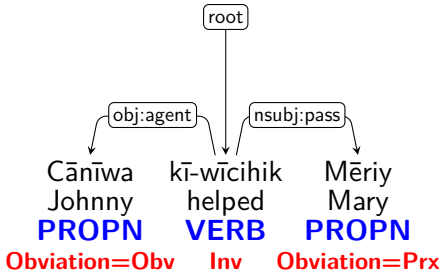
Should we set `nsubj` > `obj`?

# Direct-Inverse Voice in Plains Cree

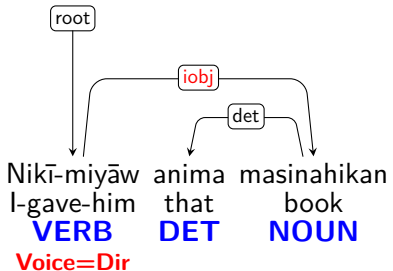


# Plains Cree Ditransitive Clauses




The theme (not the recipient) is indirect object because it is not cross-referenced on the verb (it is inanimate, while the verb references an animate object).

**Universal Dependencies**

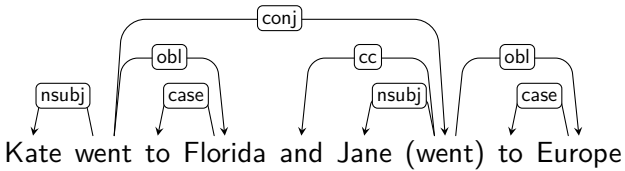
# Ellipsis and Enhanced UD

# Outline

1 Constituents vs. Dependencies

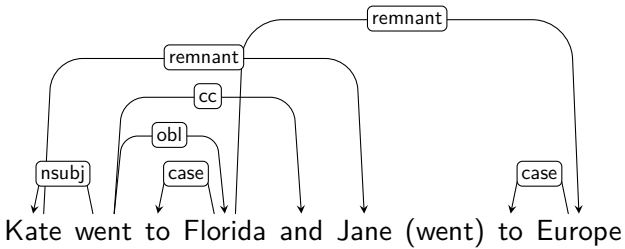
2 Universal Dependencies
- A Tour through UD Syntax
- Nonverbal Predicate and Copula
- Core Arguments vs. Oblique Dependents
- Ellipsis and Enhanced UD

# Deleted Predicates in Coordination

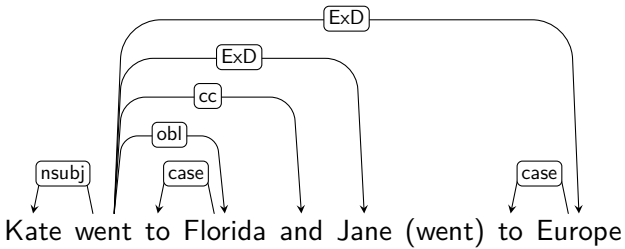



- Some treebanks would use an empty node to represent the second *went*.
- UD enhanced representation now allows empty nodes!
- But the basic representation sticks with the overt words.

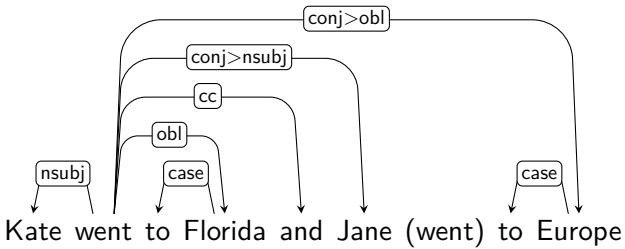
# UD V1: The `remnant` Relation



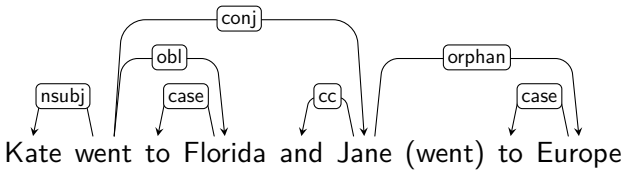
# PDT: The ExD Relation



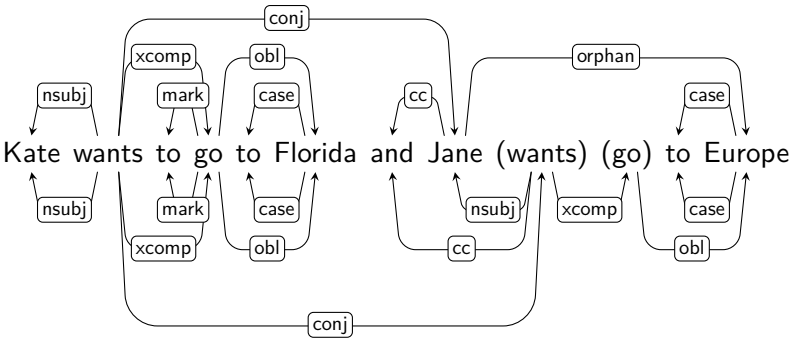
# Perseus Treebanks: Chained Relations



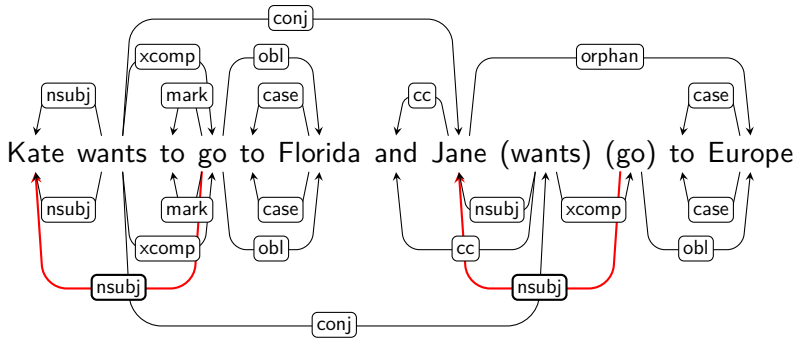
# UD V2: The `orphan` Relation



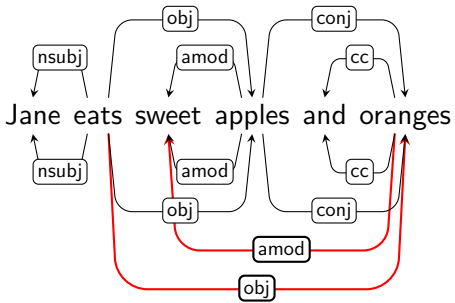
# Enhanced Dependencies: Gapping



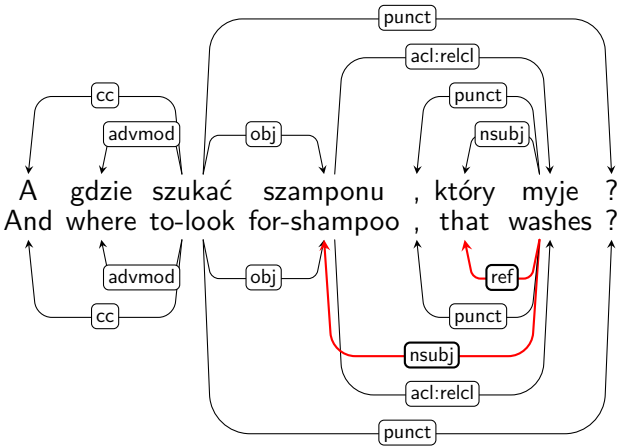
# Enhanced Dependencies: Gapping and Control



# Enhanced Dependencies: Coordination



# Enhanced Dependencies: Relative Clauses



# Basic Universal Dependencies: 138 (136) Languages and Growing

- **I.-E.:** Armenian (+West), Greek (+Ancient), Albanian, Gheg, Hittite, Breton, Irish, Manx, Scottish, Welsh, Afrikaans, Danish, Dutch, English, Faroese, Frisian, German, Gothic, Icelandic, Low Saxon, Norwegian, Swedish, Swiss German, Catalan, French, Galician, Italian, Latin, Ligurian, Neapolitan, Old French, Portuguese, Romanian, Spanish, Umbrian, Belarusian, Bulgarian, Church Slavonic, Croatian, Czech, Old Russian, Polish, Pomak, Russian, Serbian, Slovak, Slovenian, Ukrainian, Upper Sorbian, Latvian, Lithuanian, Kurmanji, Persian, Khunsari, Nayini, Soi, Urdu, Hindi, Kangri, Bhojpuri, Bengali, Marathi, Sinhala, Sanskrit
- **Dravidian:** Malayalam, Tamil, Telugu
- **Uralic:** Erzya, Estonian, Finnish, Hungarian, Karelian, Livvi, Komi Permyak+Zyrian, Moksha, Sámi North+Skolt
- **Turkic:** Kazakh, Old Turkish, Tatar, Turkish, Uyghur, Yakut
- Buryat
- Xibe
- Korean
- Japanese
- **Sino-T.:** Cantonese, Classical Chinese, Chinese
- **Tai-Kadai:** Thai
- **Aus.-As.:** Vietnamese
- **Austron.:** Indonesian, Javanese, Tagalog, Cebuano
- **Pama-Nyu.:** Warlpiri
- **Chu.-Kam.:** Chukchi
- **Esk.-Al.:** Yupik
- **U.-Az.:** Nahuatl
- **Mayan:** Kiche
- **Arawakan:** Apurinã
- **Arawan:** Madi
- **Tupian:** Akuntsu, Guajajara, Kaapor, Karo, Makurap, Mundurukú, Nheengatu, Tupinambá, Mbyá, Guaraní, Teko
- **M.-Je:** Xavante
- **Af.-As.:** Akkadian, Amharic, Arabic Standard+Levantine, Assyrian, Beja, Coptic, Hebrew (+Ancient), Maltese, Zaar
- **Niger-Congo:**

# Summary

- Constituent (phrase) trees … context-free grammar
- Dependency trees (or graphs)
  - Nonprojective dependencies
- Universal Dependencies
  - Unified annotation for all languages
    - Language-specific extensions
  - Content words higher than function words … better parallelism
  - Clauses – nominals – modifier words
  - Core arguments vs. oblique dependents

https://ufal.cz/courses/npfl094