LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Universal Dependencies

Renhao Pei

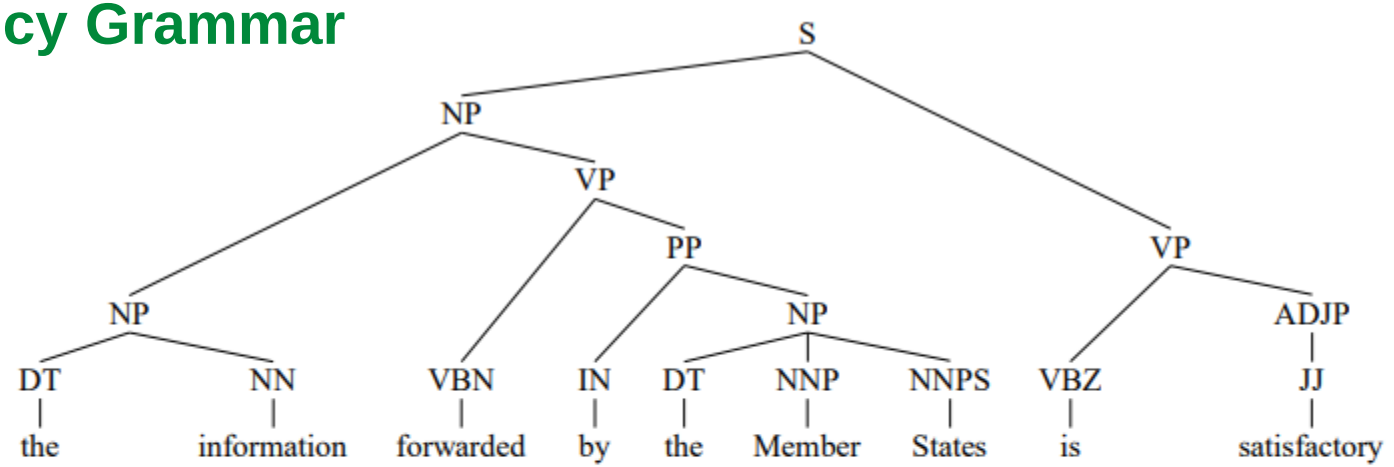
19.12.2023

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

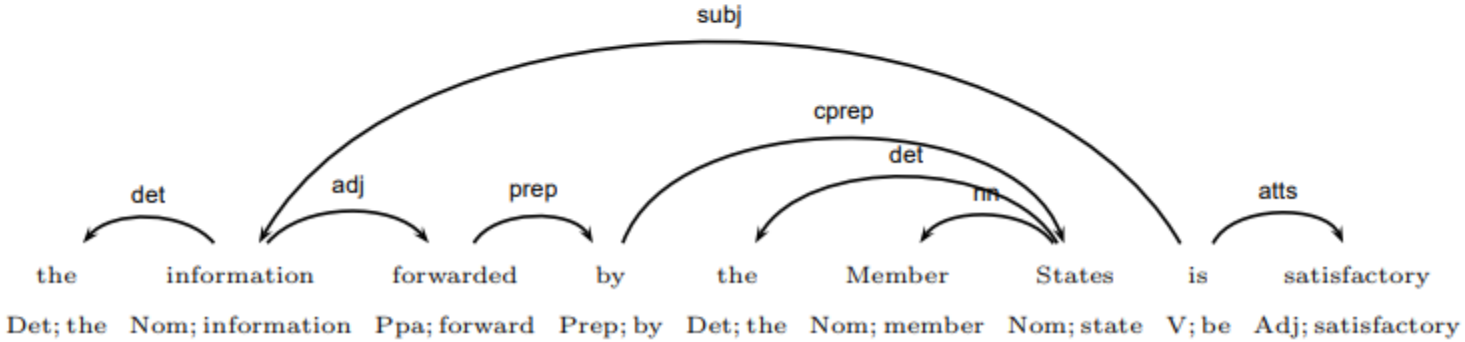
Table of Content

Overview of Dependency Grammar

- Dependency vs. Constituency



- Dependency:
- a binary asymmetrical relation
  (from Head to Dependent)



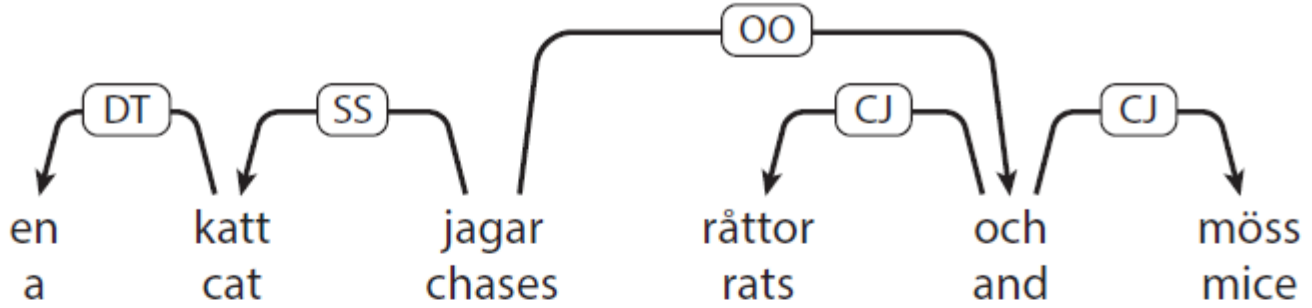
- One idea that unites the dependency grammar tradition: syntactic structure can be reduced to binary dependency relations
  - But how? there are many different frameworks… (including UD and MTM)

Overview of Dependency Grammar

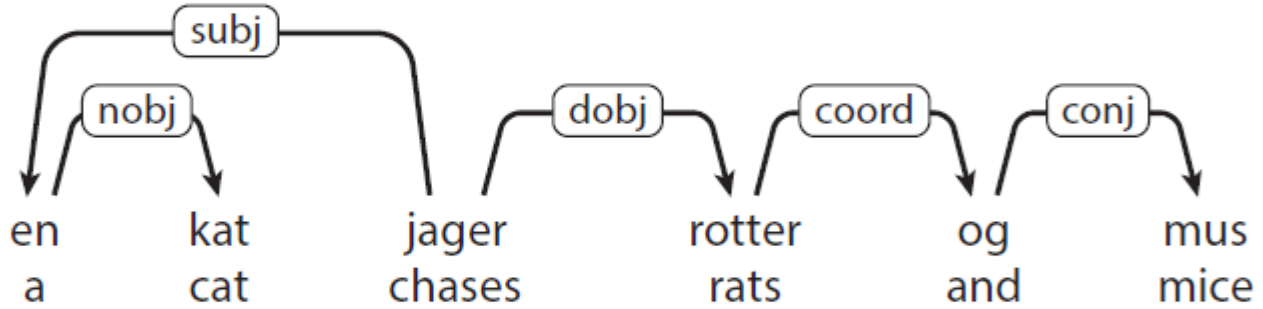
Three parallel sentences annotated according to:

- the Swedish Treebank
- the Danish Dependency Treebank
- the Stanford Dependencies

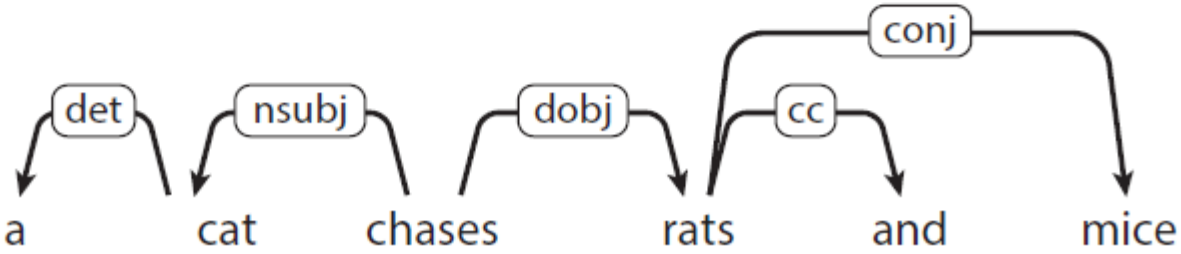
Swedish



Danish



English



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Motivation to UD

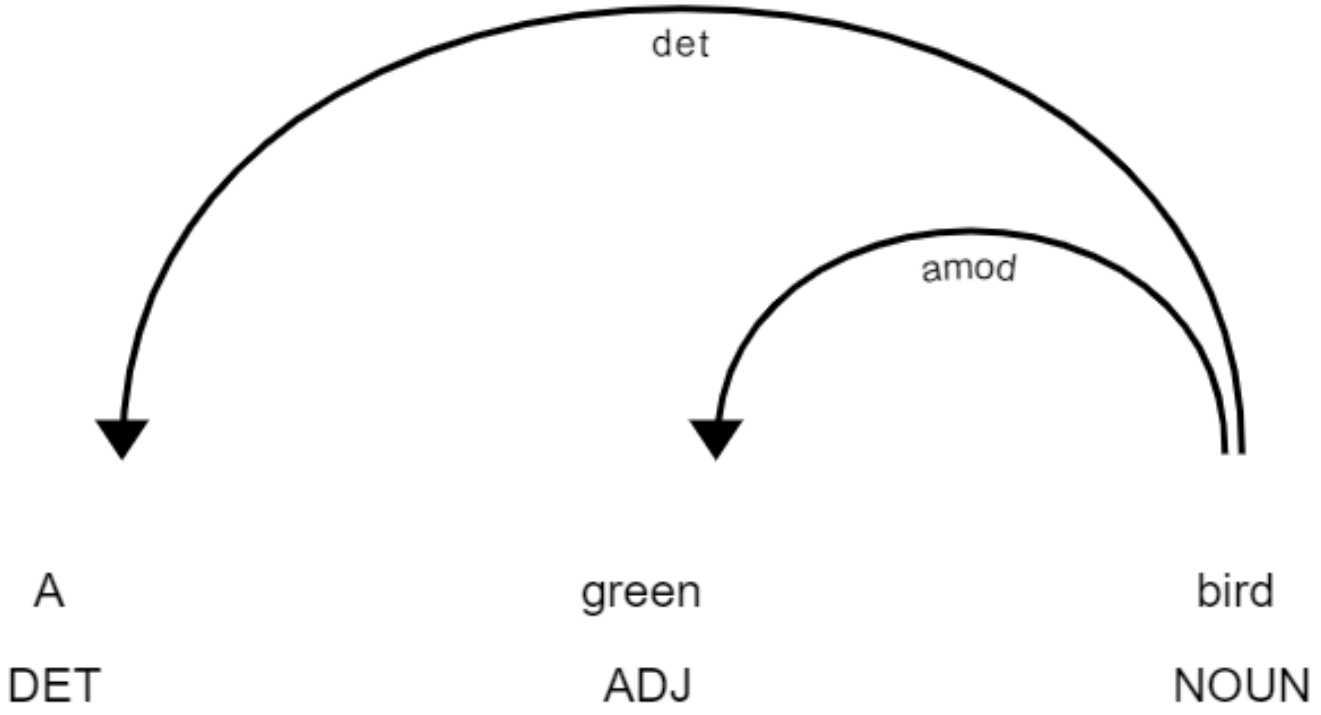
- We need a **universally consistent** framework to annotate of similar constructions across languages

- UD: not only a theoretical **framework**, but also a **corpus**
  - Currently **148** languages (as in v2.13 released 15 Nov 2023), still growing!
  - Language-specific extensions are allowed

- UD emerged from joint effort of the NLP community, based on:
  - Stanford dependencies
  - Google universal part-of-speech tags
  - the *Interset* interlingua for morphosyntactic tag sets

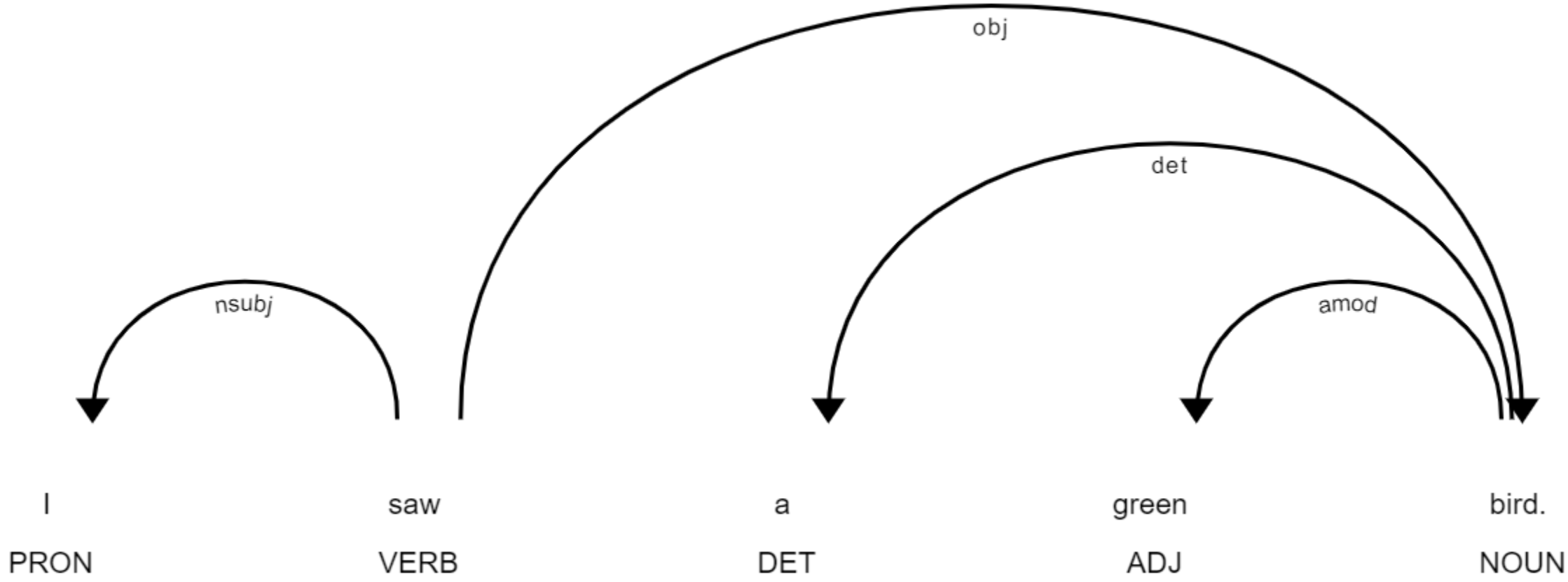
UD basics

- Nominals: the primary means for referring to **entities**
  - often built around nouns



UD basics

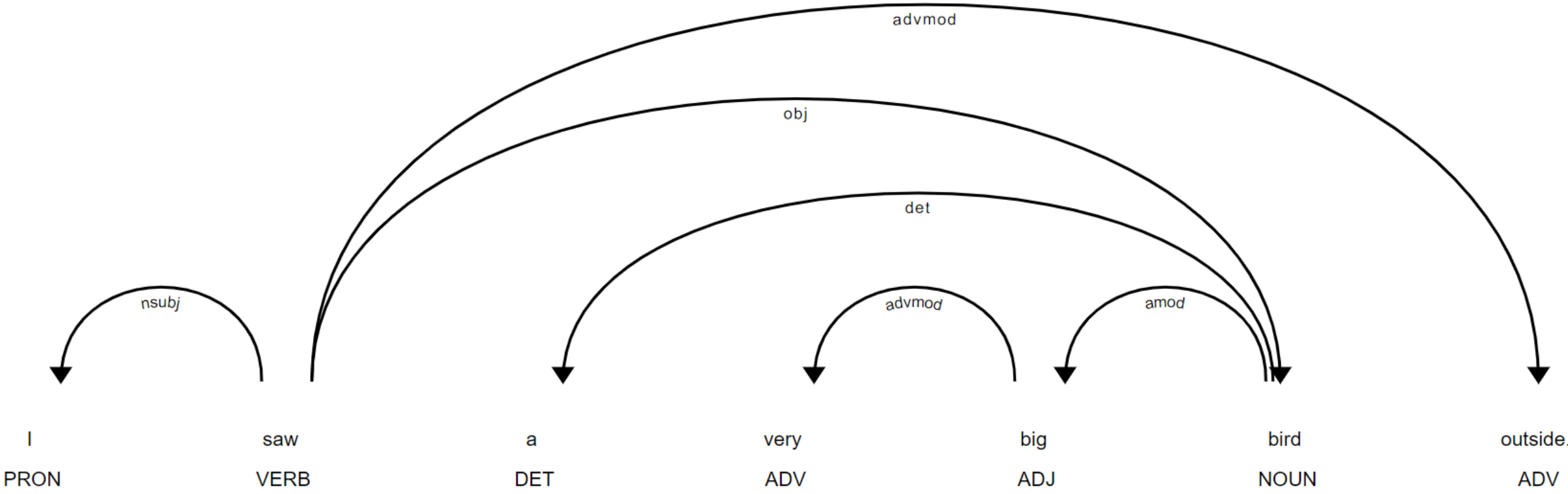
- Clauses: the primary means for referring to **events**
  - Often build around verbs



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

UD basics

- Modifiers: attributive modifiers of nominals, clauses, and other modifiers



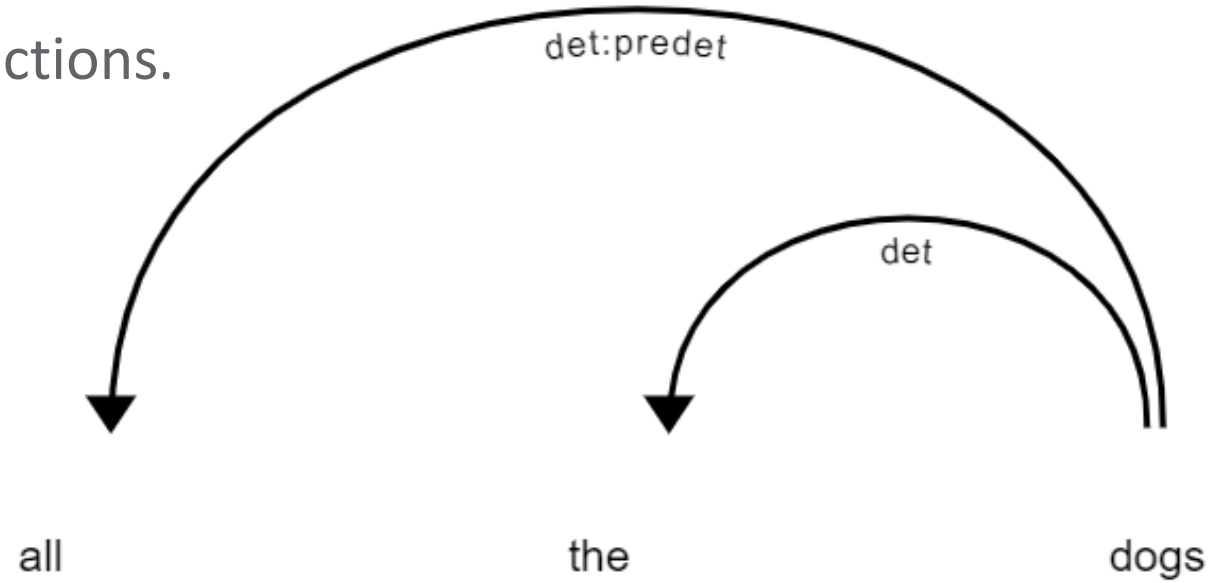
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

UD basics

Language-specific extensions

- UD allows relation subtypes (marked by a colon)

to further capture language-specific constructions.



- E.g. *predeterminers* as a subtype of determiners, can be annotated as *det:predet*

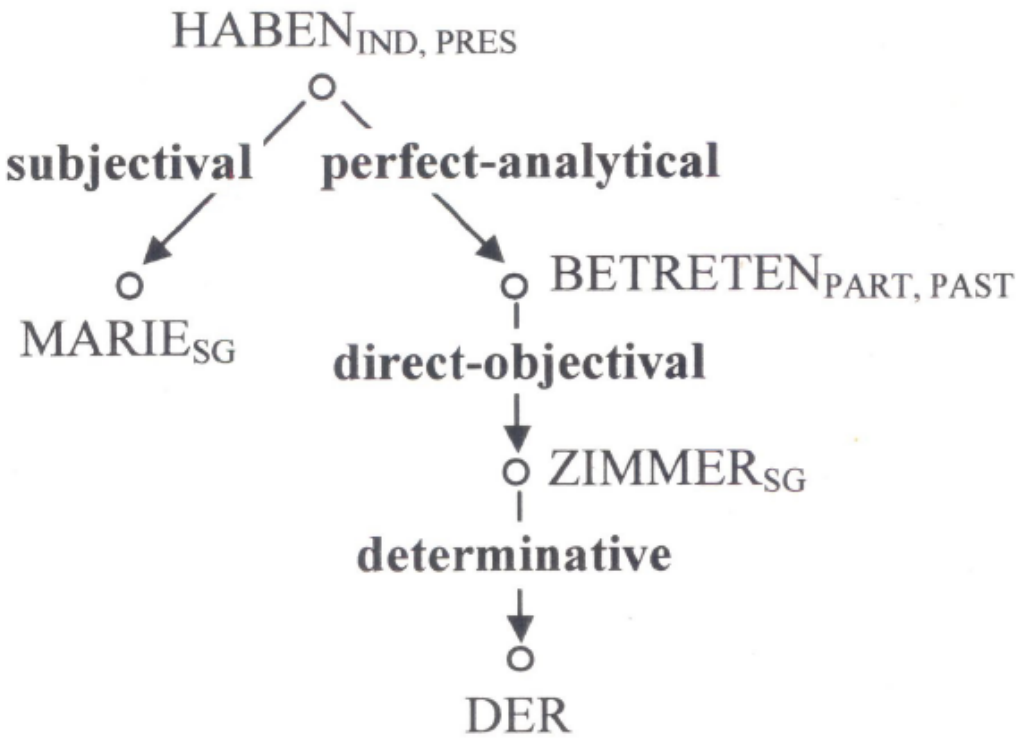
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

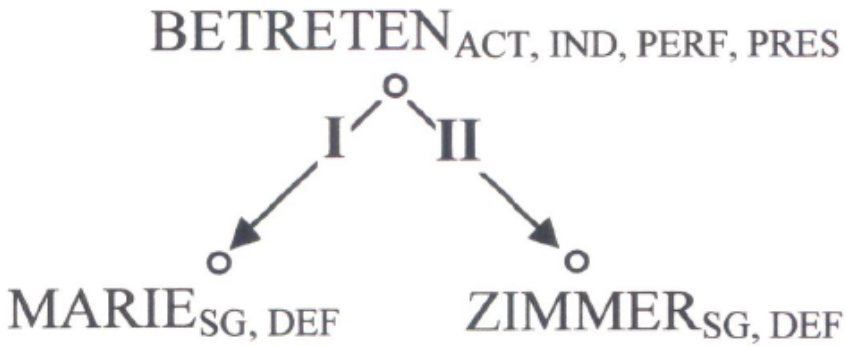
Comparing UD and MTM

Levels of Representation

- MTM has 2 levels for syntactic representation:
  - surface syntactic representation (SSynt)
  - deep syntactic representation(DSynt), which is more semantically oriented

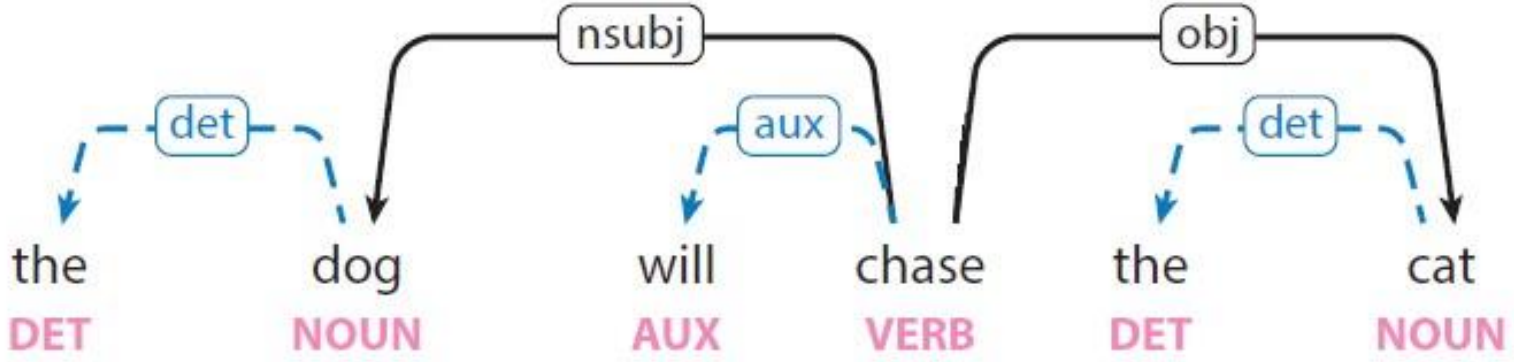
- UD has (mostly) only one level of syntactic representation known as ‘basic’ dependencies





Comparing UD and MTM

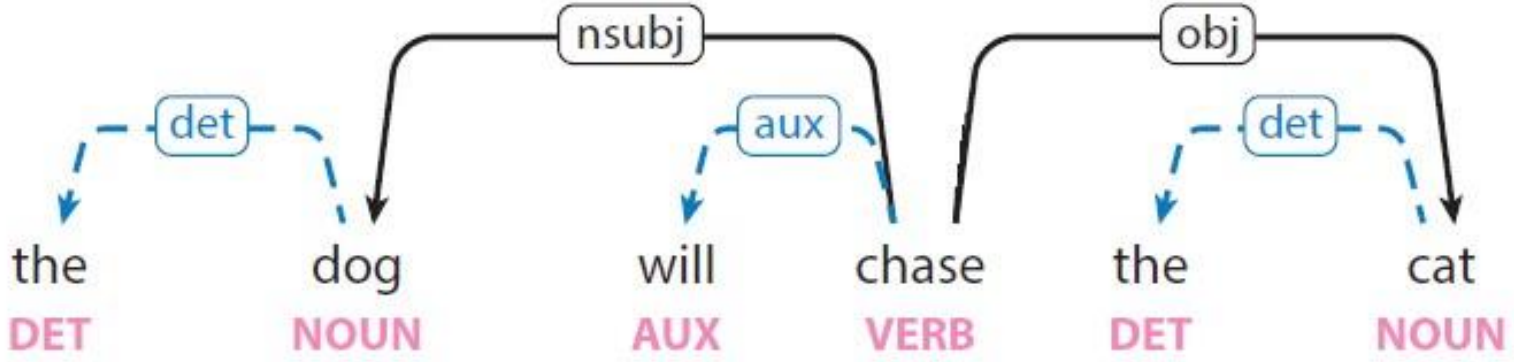
- UD merges the two levels (in MTM) into one level, while this one level has two sets of relations:
  - 1. syntactic dependency relations: argument and modifier relations like *nsubj*, *obj*…
  - 2. functional relations: relations between a lexical head and its grammatical markers, like *det*, *aux*…



Comparing UD and MTM

- Compared to MTM:
  - DSynt: the blue relations would be omitted
  - SSynt: ‘will’ would be the head, and ‘chase’ and ‘dog’ will be its dependent.

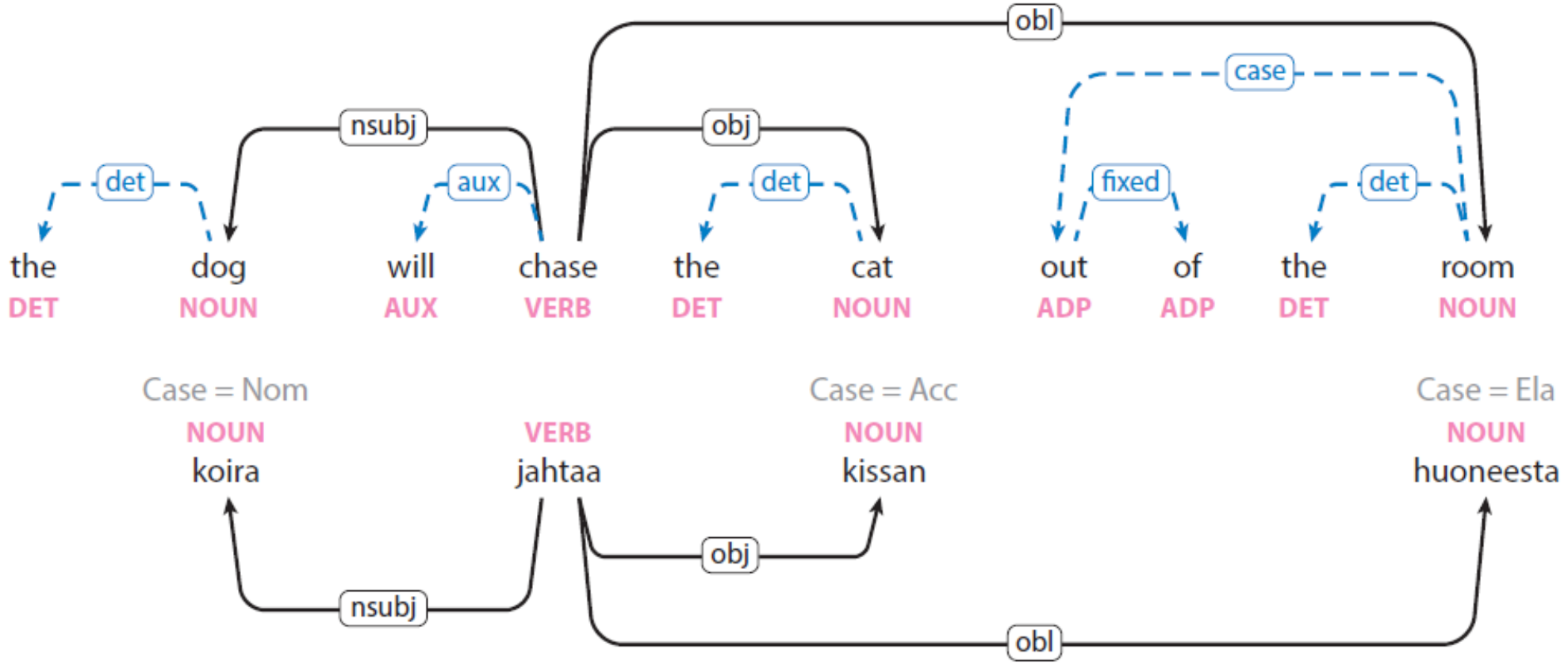
- UD gives priority to **content words**, and function words are attached to the content word as dependent.
  - ‘chase’ is the head of ‘will’



Comparing UD and MTM

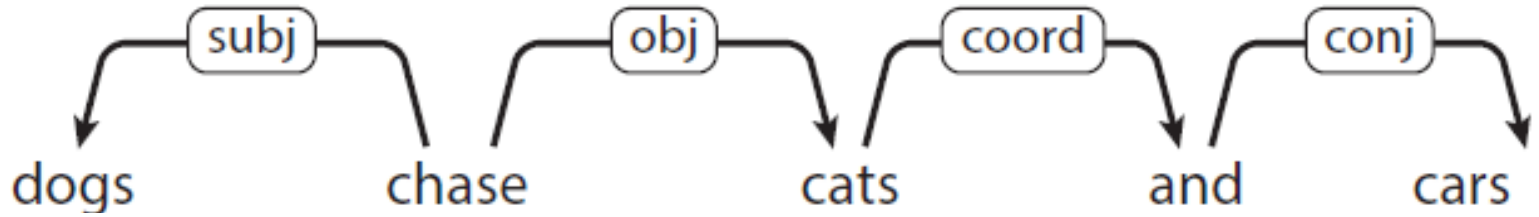
Why does UD do it this way?

- Generalization across languages: some languages use functional words, where other languages use morphology

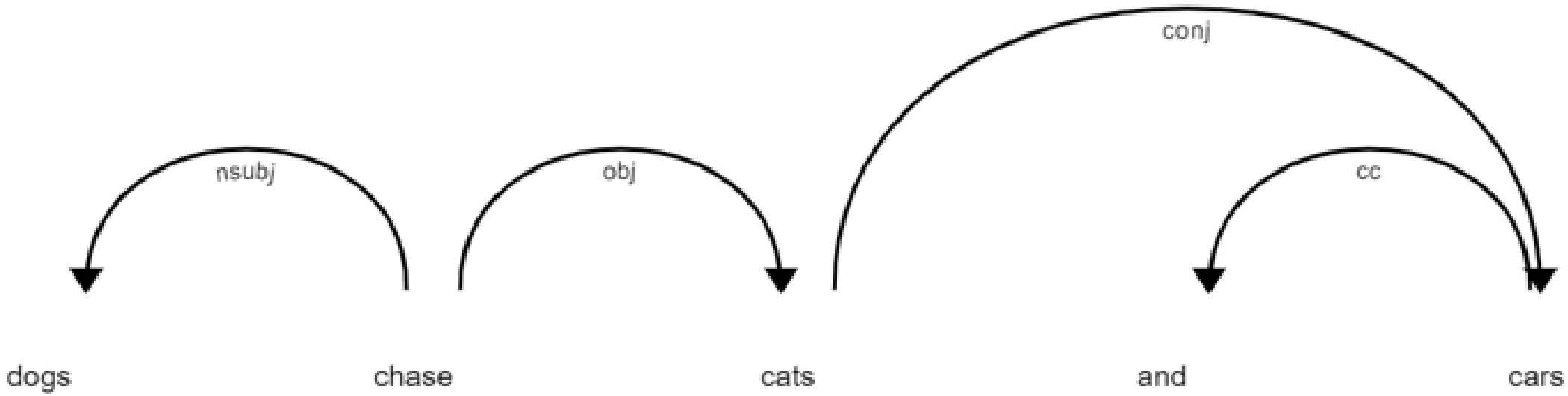


Comparing UD and MTM

- Coordination
- MTM:



- UD:



- In UD, subsequent conjuncts are linked to the first conjunct using the *conj* relation.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Advantages of dependencies in general

Not specifically to UD, but to all Dependency-based frameworks:

- 1. easier generalization for free word-order
  - Dependency trees are not sensitive to the word-order, so they can capture generalizations in languages with flexible word order, in contrast to constituency trees.

- 2. easier to parse
  - Dependency parsing is simply linking existing words in a sentence together, whereas constituency parsing has to infer additional higher-level nodes (phrase-level nodes).

- 3. transparent encoding of Predicate-Argument structure, which supports semantic interpretation.
  - Dependency trees are therefore useful for downstream applications in NLP such as relation extraction and question answering

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Advantages of UD specifically

UD specifically:

- 1. consistent basic annotation across a diverse set of languages
  - the Predicate-Argument structure can be annotated consistently across languages, regardless of functional words vs. morphology
- 2. allowing relation subtypes to further capture language-specific constructions.
- 3. plenty of data
  - 259 UD treebanks for 148 languages
  - treebank size ranges from 1,000 to over 3 million tokens

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Limitation of dependencies

Not specifically to UD, but to all Dependency-based frameworks:

- Having no constituency would make it hard to distinguish different scope of modifiers.
  - (young men) and women vs. young (men and women)

- To solve this, we must introduce some kind of constituency:
  - MTM allows bracketing for these cases
  - UD does not distinguish them

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Tools and resources for UD

- The *Universal Dependencies* corpora
- An online UD parser from Stanza: http://stanza.run/
- Some relevant libraries:
  - UDPipe
  - Stanza: includes a UD parser for 50 languages
  - spacy-stanza: a package wraps the Stanza library in the spaCy pipeline
- An online course for UD parsing in python, using spacy-stanza: https://applied-language-technology.mooc.fi/html/notebooks/part_iii/02_universal_dependencies.html

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Literature

- De Marneffe, M. C., & Nivre, J. (2019). Dependency grammar. *Annual Review of Linguistics*, *5*, 197-218.
- De Marneffe, M. C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal dependencies. *Computational linguistics*, *47*(2), 255-308.
- Mel'čuk, I. (2009). Dependency in natural language. *Dependency in linguistic description*, *111*(1).
- Mel'čuk, I. (2015). *Semantics: From meaning to text* (Vol. 3). John Benjamins Publishing Company.
- Nivre, J., De Marneffe, M. C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., ... & Zeman, D. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. *arXiv preprint arXiv:2004.10643*.

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Thank you!