# Universal Dependencies
## A Framework for Morphosyntactic Annotation

Joakim Nivre

Uppsala University
Department of Linguistics and Philology

# Why?
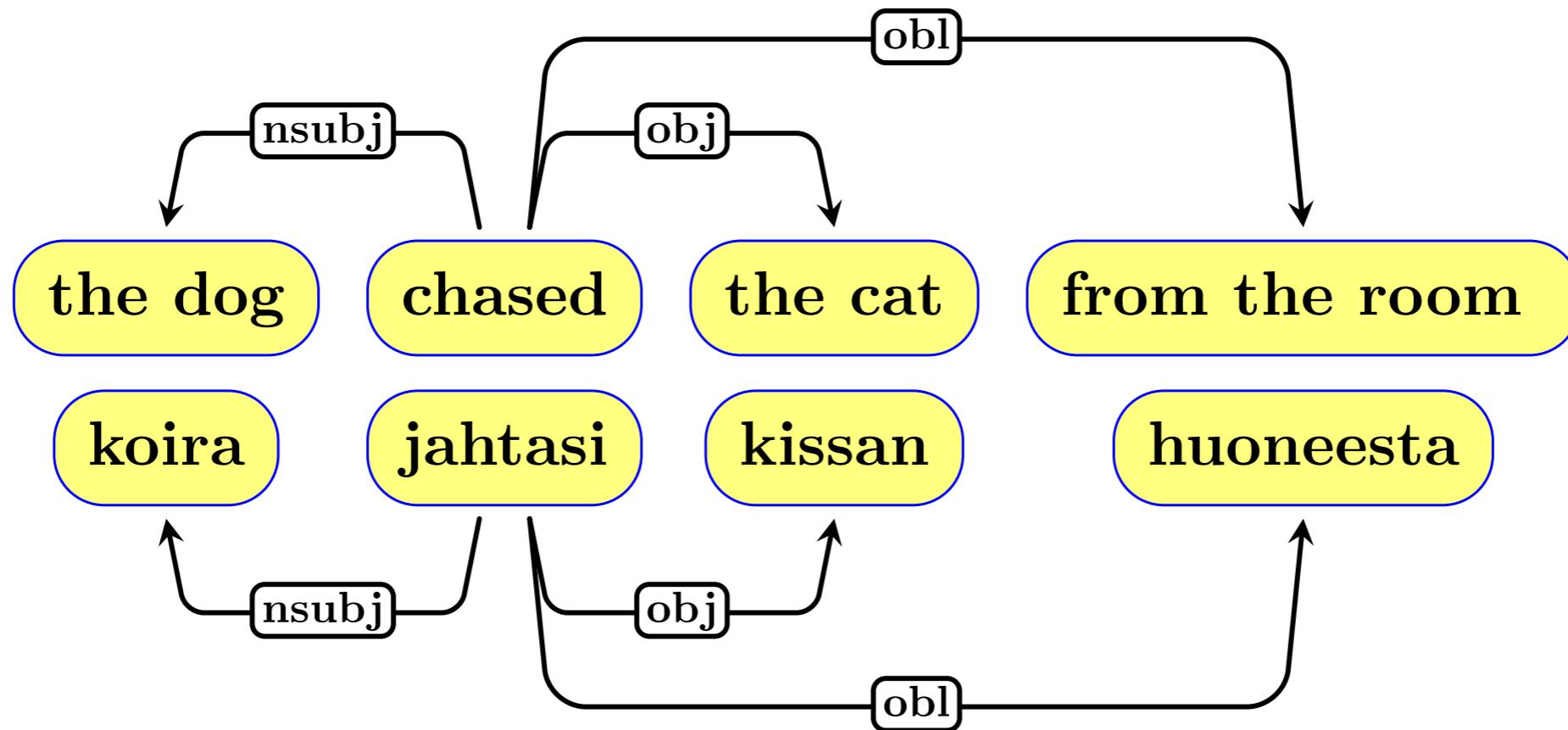
Cross-linguistically consistent morphosyntactic annotation

Facilitate multilingual research in NLP and linguistics

- Meaningful linguistic analysis across languages
- Syntactic parsing in cross-lingual settings
- NLP systems for multiple languages
- Facilitate resource-building for new languages

Complement – not replace – language-specific schemes

# How?



Focus on grammatical relations between (content) words

# Who?
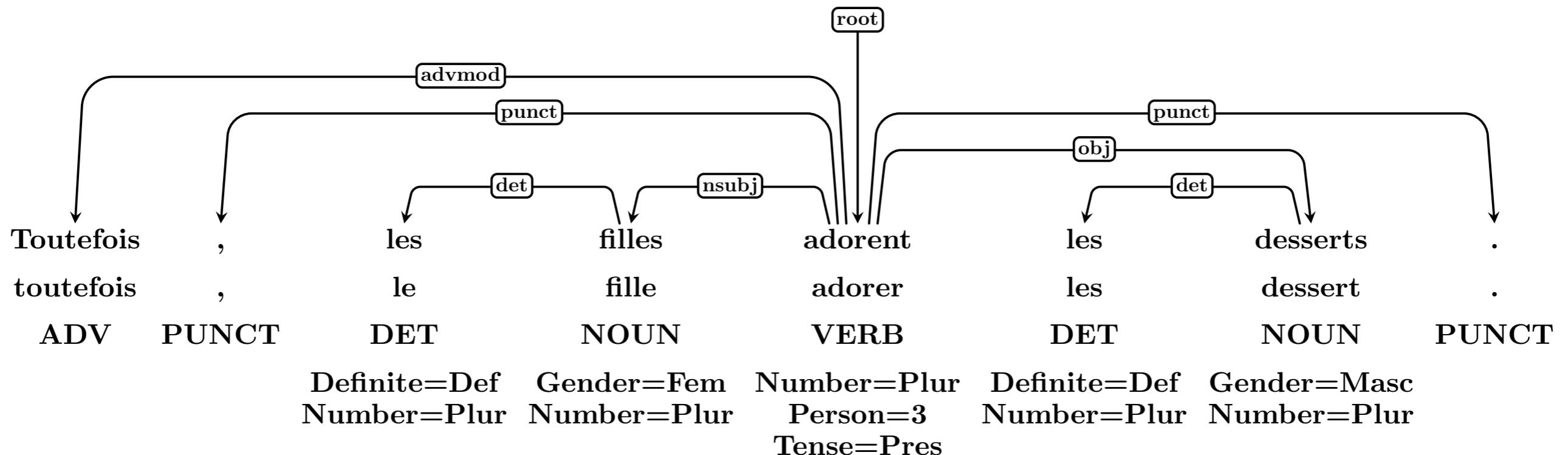


Open community effort – a big tent

UD v2.4: 83 languages, 146 treebanks, 323 contributors

Come join us at http://universaldependencies.org
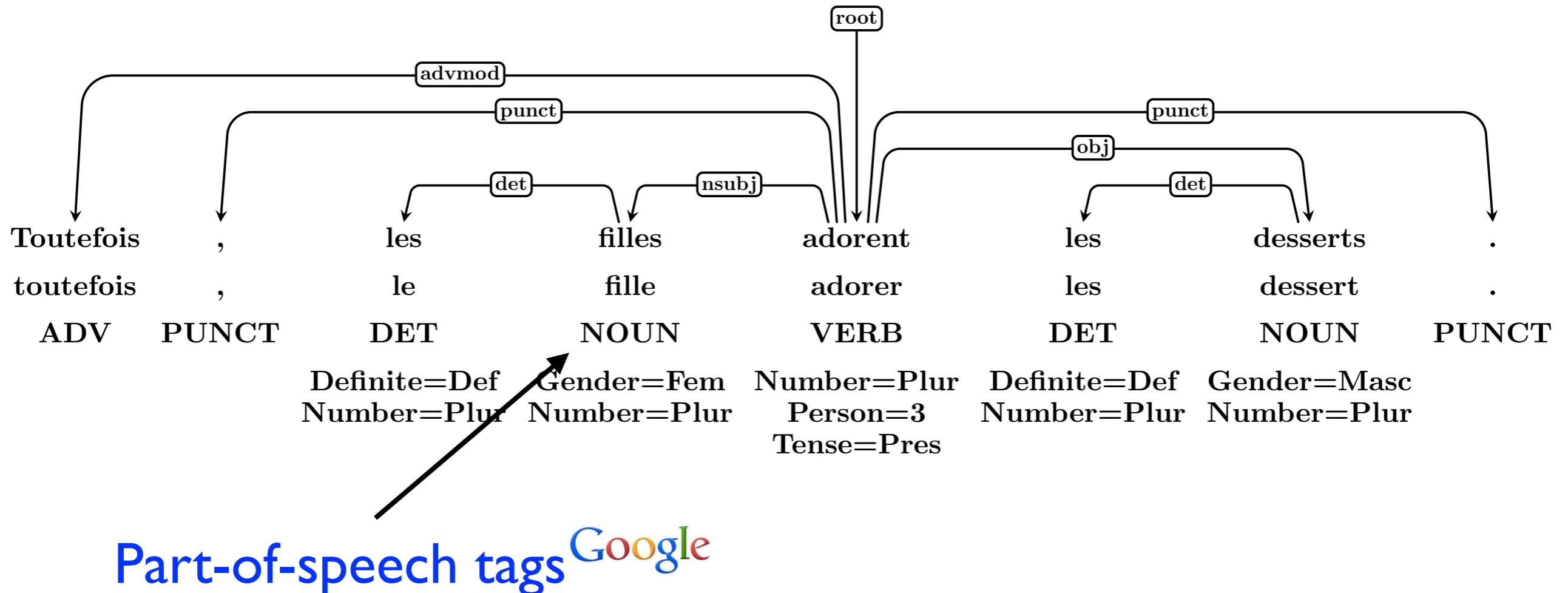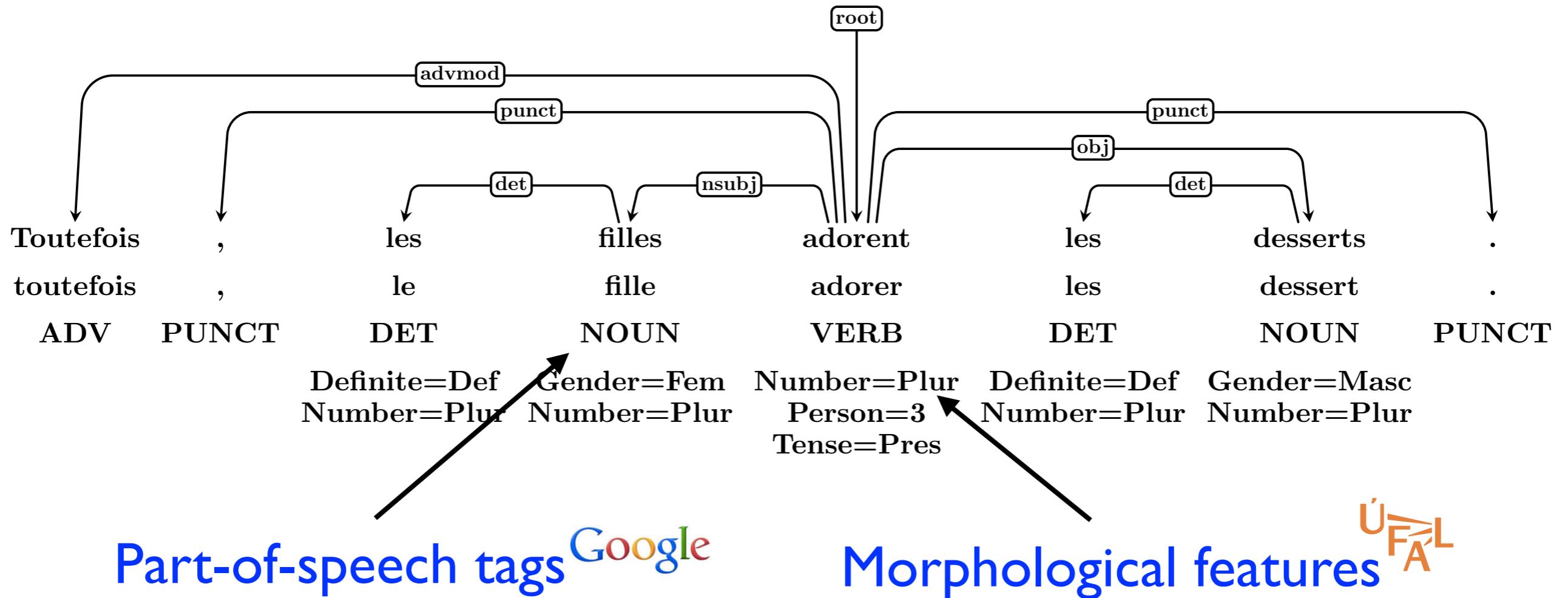
# Universal Dependencies

# Universal Dependencies

http://universaldependencies.org



| Toutefois | , | les | filles | adorent | les | desserts | . |
|---|---|---|---|---|---|---|---|
| toutefois | , | le | fille | adorer | les | dessert | . |
| ADV | PUNCT | DET | NOUN | VERB | DET | NOUN | PUNCT |
| | | Definite=Def Number=Plur | Gender=Fem Number=Plur | Number=Plur Person=3 Tense=Pres | Definite=Def Number=Plur | Gender=Masc Number=Plur | |

Part-of-speech tags Google

# Universal Dependencies

| Toutefois | , | les | filles | adorent | les | desserts | . |
|-----------|---|-----|--------|---------|-----|----------|---|
| toutefois | , | le | fille | adorer | les | dessert | . |
| ADV | PUNCT | DET | NOUN | VERB | DET | NOUN | PUNCT |
| | | Definite=Def Number=Plur | Gender=Fem Number=Plur | Number=Plur Person=3 Tense=Pres | Definite=Def Number=Plur | Gender=Masc Number=Plur | |

Part-of-speech tags Google

Morphological features

# Universal Dependencies

http://universaldependencies.org

Syntactic relations

Part-of-speech tags

Morphological features

| Toutefois | , | les | filles | adorent | les | desserts | . |
|-----------|---|-----|--------|---------|-----|----------|---|
| toutefois | , | le | fille | adorer | les | dessert | . |
| ADV | PUNCT | DET | NOUN | VERB | DET | NOUN | PUNCT |

| | | Definite=Def Number=Plur | Gender=Fem Number=Plur | Number=Plur Person=3 Tense=Pres | Definite=Def Number=Plur | Gender=Masc Number=Plur | |

advmod
punct
det
nsubj
root
obj
det
punct

# Morphology

Le          chat          chasse          les          chiens          .

# Morphology

| Le | chat | chasse | les | chiens | . |
|----|------|--------|-----|--------|---|
| le | chat | chasser | le | chien | . |

- Lemma representing the semantic content of the word

# Morphology

| Le | chat | chasse | les | chiens | . |
|----|------|--------|-----|--------|---|
| le | chat | chasser | le | chien | . |
| **DET** | **NOUN** | **VERB** | **DET** | **NOUN** | **PUNCT** |

- Lemma representing the semantic content of the word

- Part-of-speech tag representing its grammatical class

# Morphology

Le                     hiens        .
le                    hien        .
**DET**        **N**               **OUN**     **PUNCT**

| Open | Closed | Other |
|------|--------|-------|
| ADJ | ADP | PUNCT |
| ADV | AUX | SYM |
| INTJ | CCONJ | X |
| NOUN | DET | |
| PROPN | NUM | |
| VERB | PART | |
| | PRON | |
| | SCONJ | |

- Lemma rep                        of the word

- Part-of-speech tag representing its grammatical class

# Morphology

| Le | chat | chasse | les | chiens | . |
|----|------|--------|-----|--------|---|
| le | chat | chasser | le | chien | . |
| **DET** | **NOUN** | **VERB** | **DET** | **NOUN** | **PUNCT** |

| Definite=Def | Gender=Masc | Mood=Ind | Definite=Def | Gender=Masc | |
| Gender=Masc | Number=Sing | Number=Sing | Gender=Masc | Number=Plur | |
| Number=Sing | | Person=3 | Number=Plur | | |
| | | Tense=Pres | | | |
| | | VerbForm=Fin | | | |

- Lemma representing the semantic content of the word

- Part-of-speech tag representing its grammatical class

- Features representing lexical and grammatical properties of the lemma or the particular word form

# Morphology

Le      chiens    .
le      chien    .
**DET**     **N**      **OUN**     **PUNCT**

<span style="color:red">Definite=Def   Gend...    ...er=Masc
Gender=Masc   Num...    ...er=Plur
Number=Sing</span>
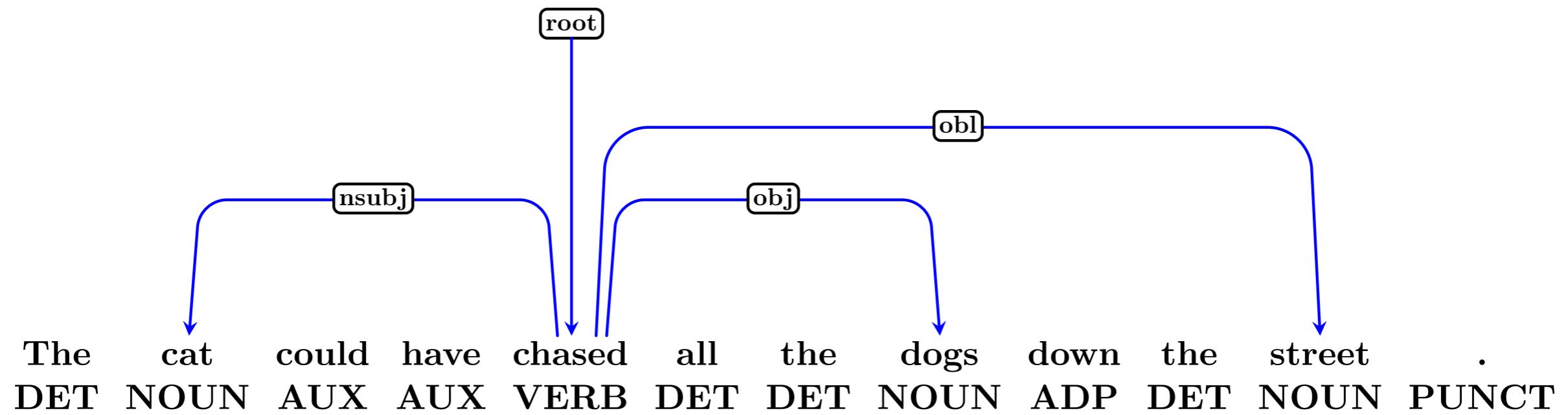
| Lexical | Inflectional Nominal | Inflectional Verbal |
|---------|----------------------|---------------------|
| PronType | Gender | VerbForm |
| NumType | Animacy | Mood |
| Poss | Number | Tense |
| Reflex | Case | Aspect |
| Foreign | Definite | Voice |
| Abbr | Degree | Evident |
|  |  | Polarity |
|  |  | Person |
|  |  | Polite |

- Lemma rep... of the word

- Part-of-spe... tical class

- Features representing lexical and grammatical properties of the lemma or the particular word form

# Syntax
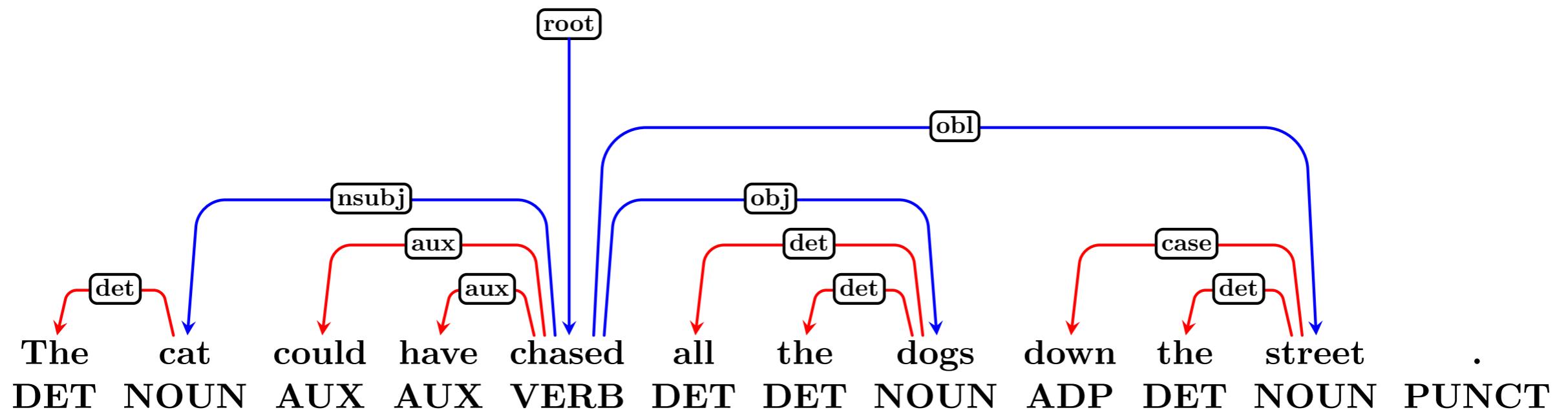
The cat could have chased all the dogs down the street .
DET NOUN AUX AUX VERB DET DET NOUN ADP DET NOUN PUNCT

# Syntax



The cat could have chased all the dogs down the street .
DET NOUN AUX AUX VERB DET DET NOUN ADP DET NOUN PUNCT
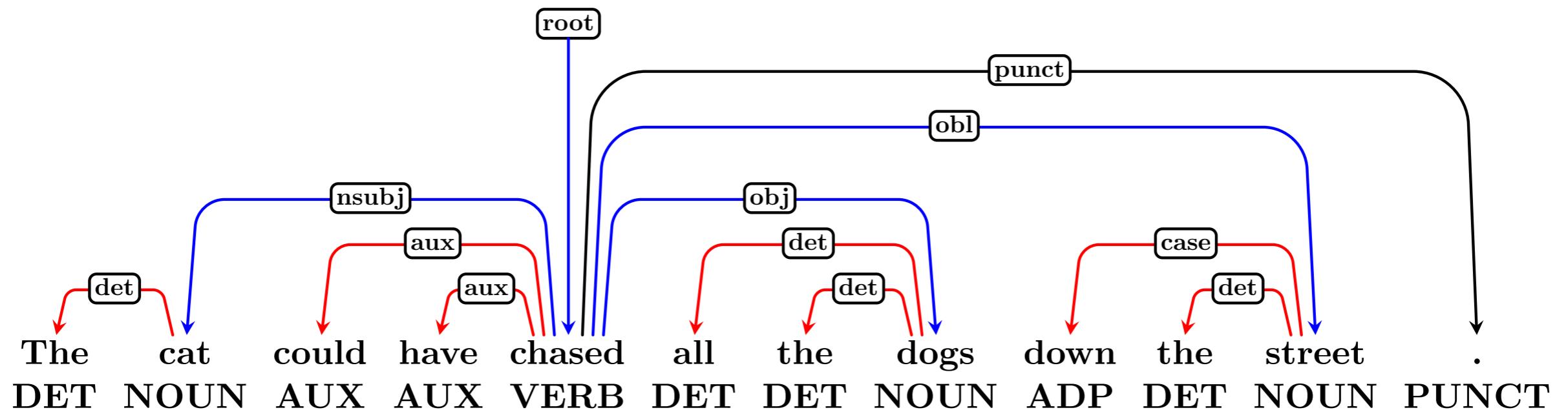
- Content words are linked by grammatical relations

# Syntax
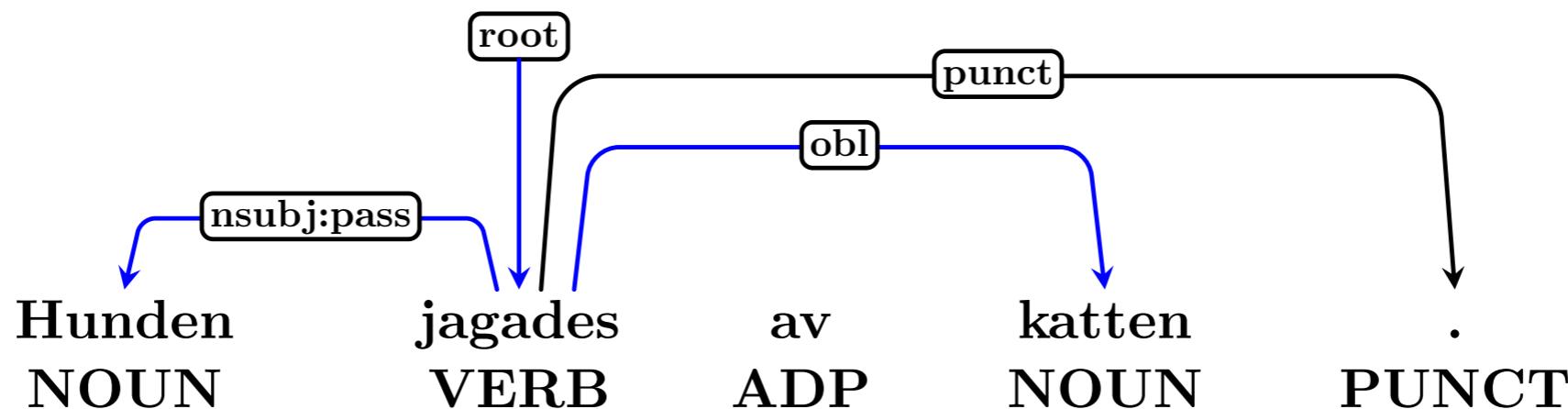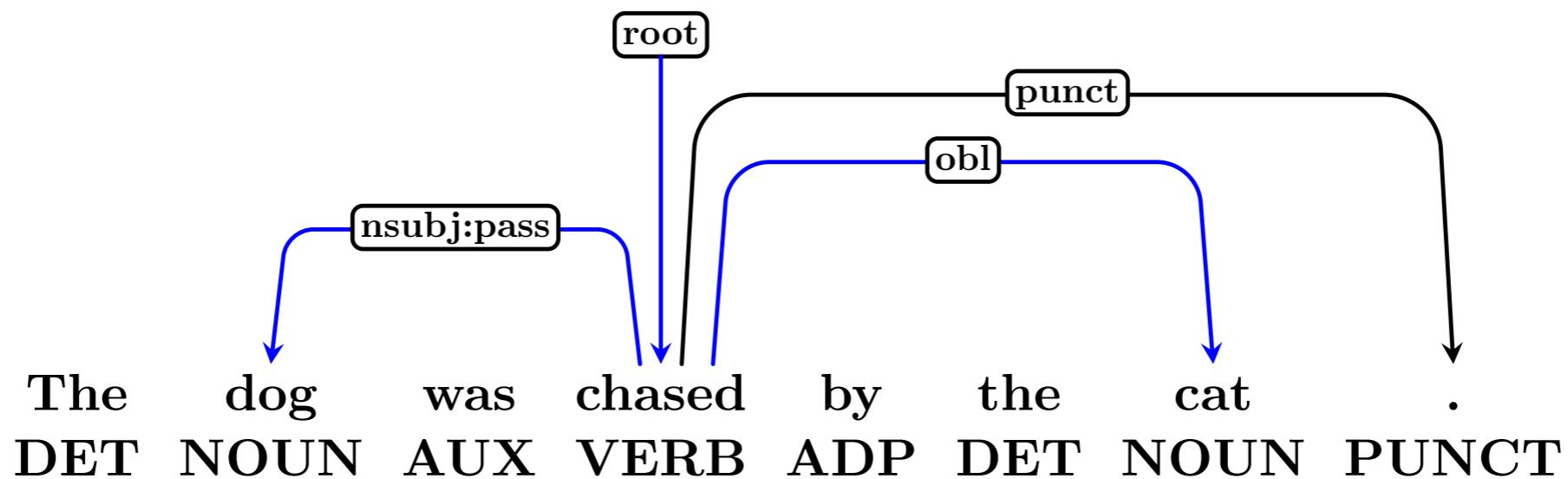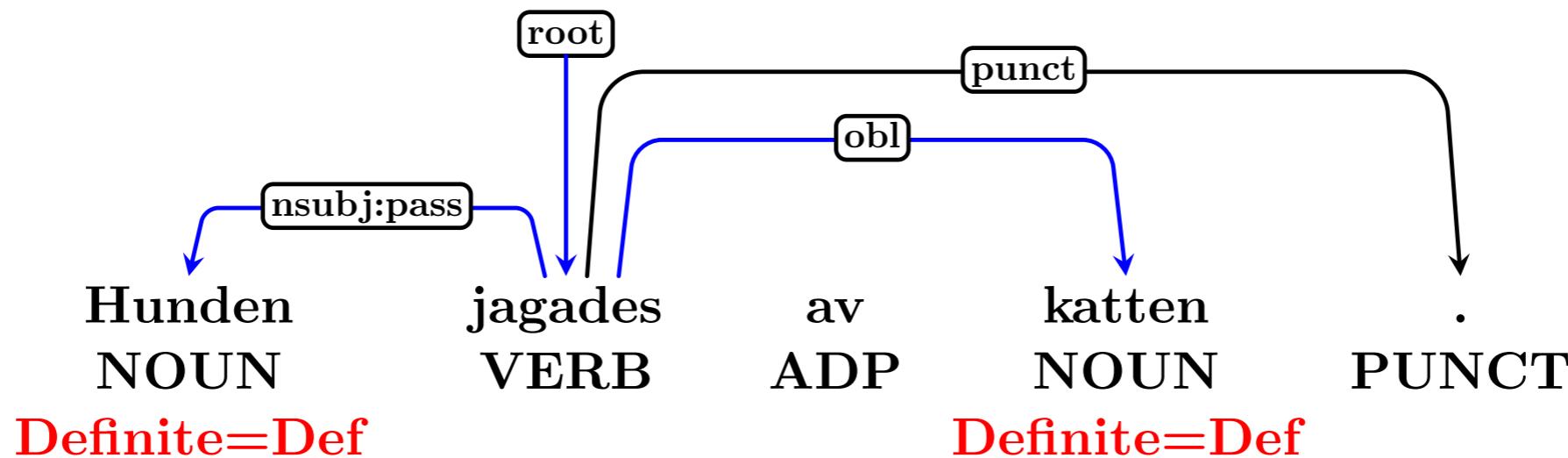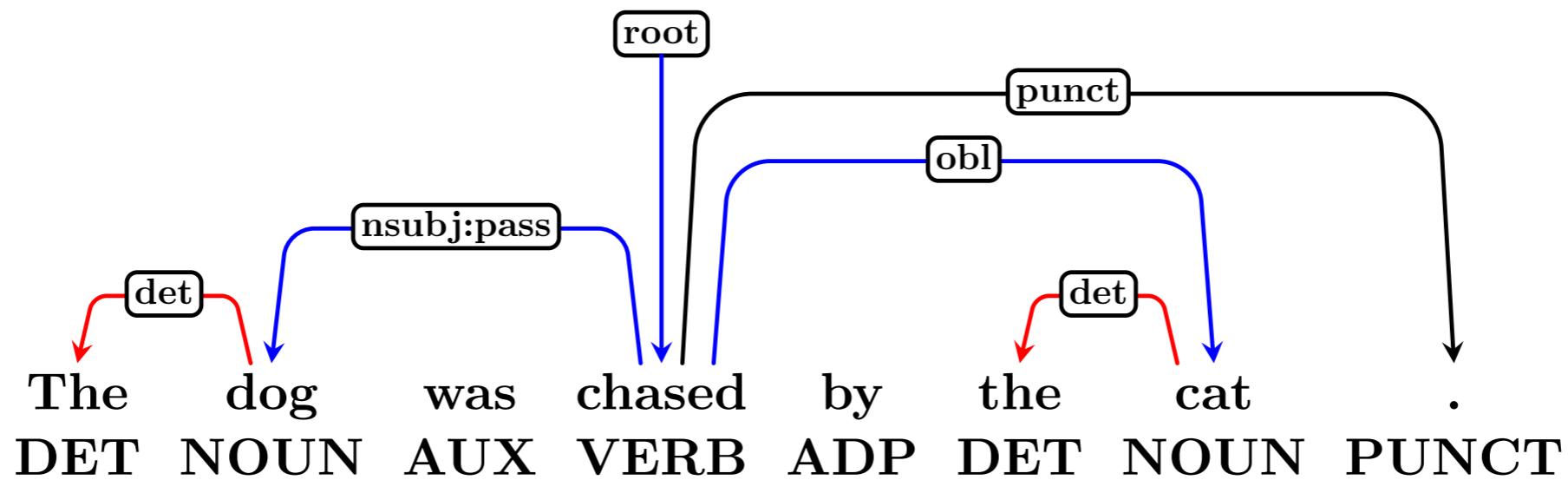


- Content words are linked by grammatical relations
- Function words attach to the content word they modify
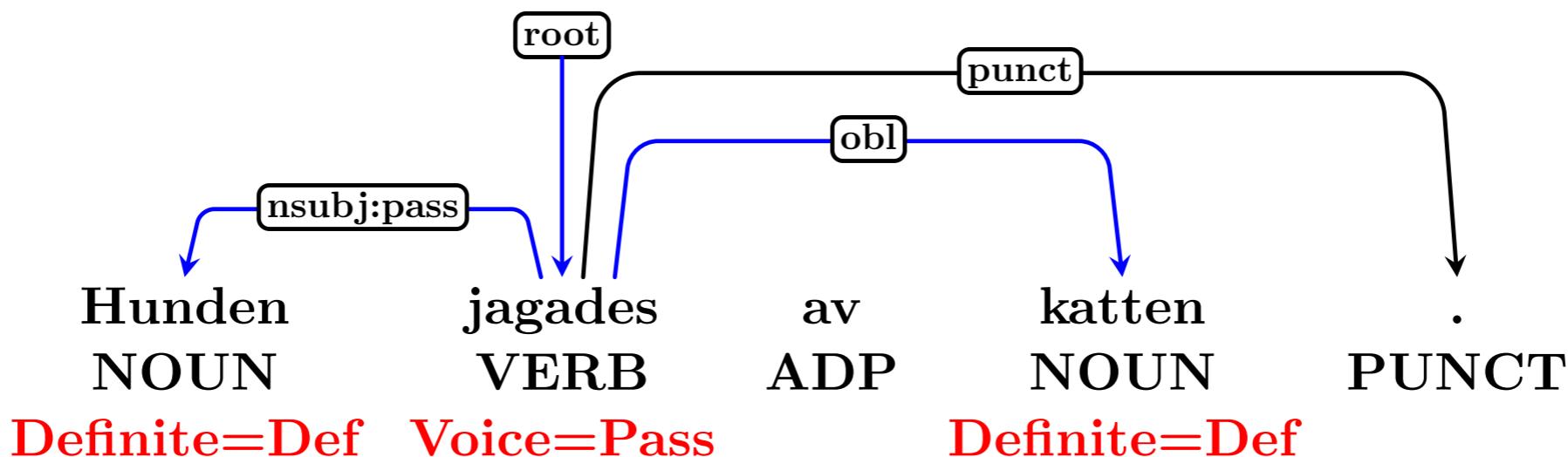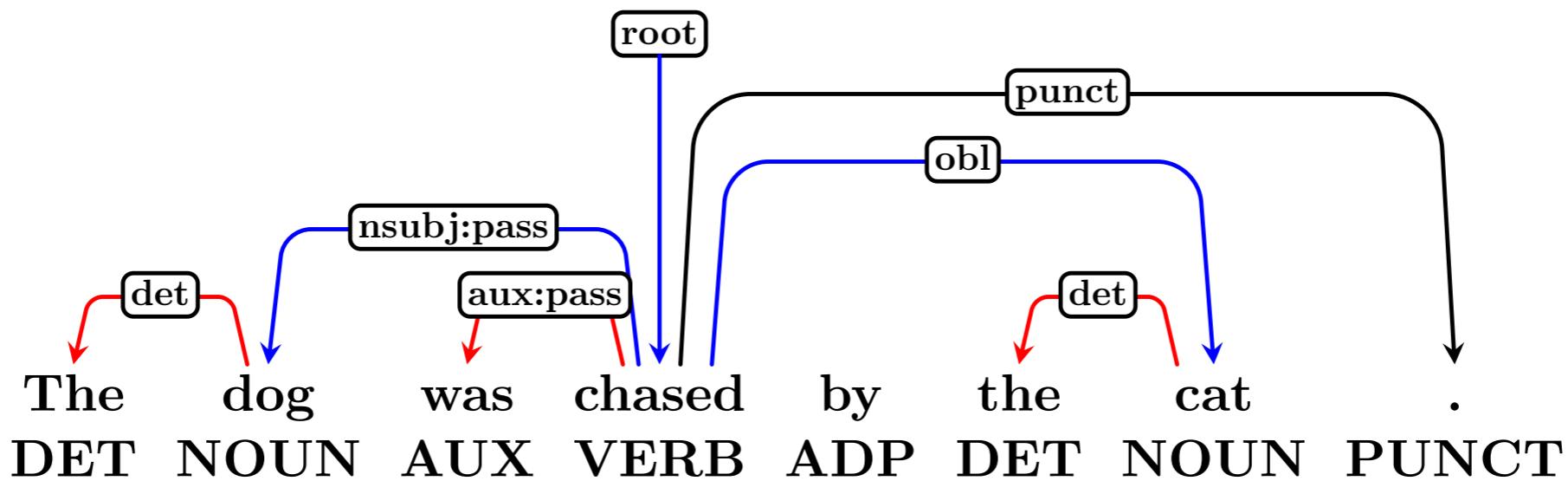
# Syntax
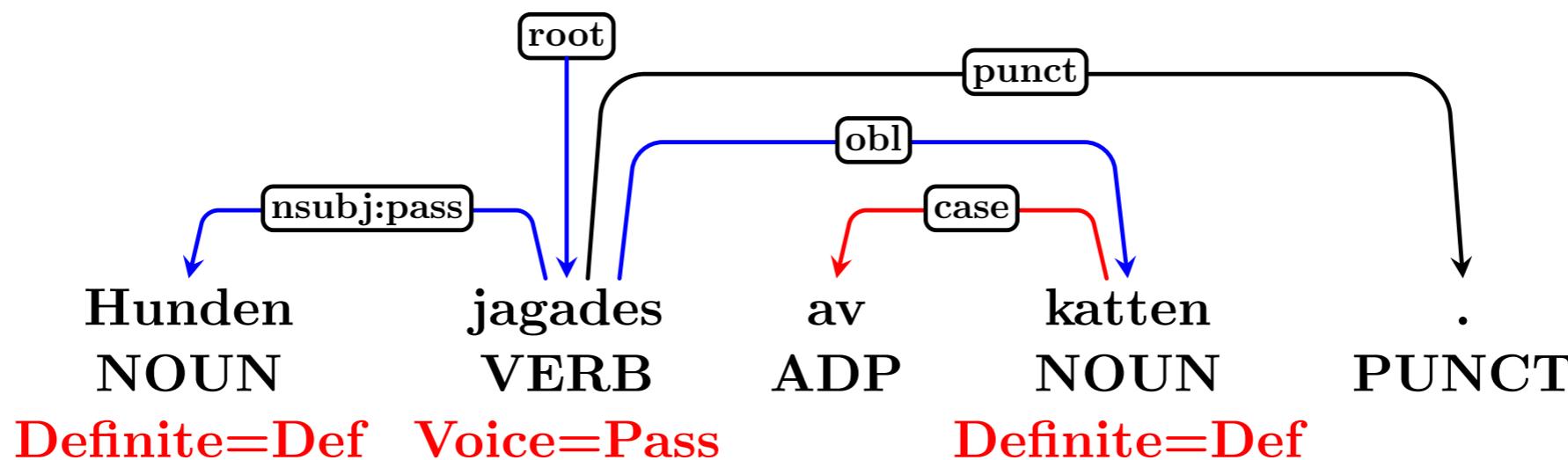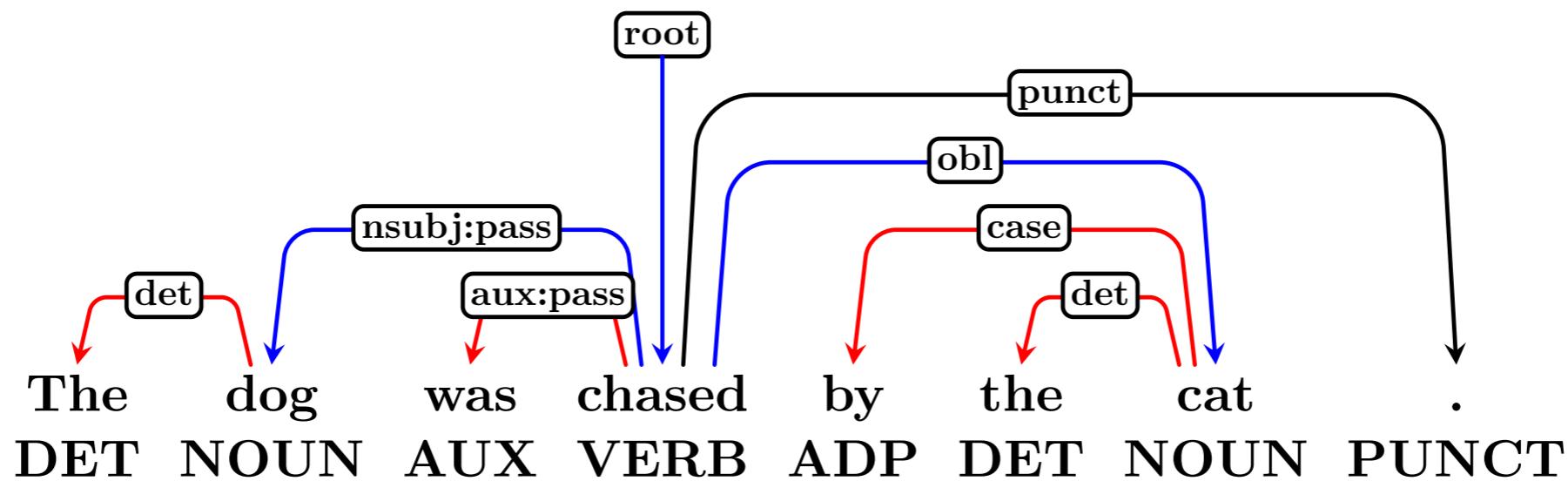


- Content words are linked by grammatical relations

- Function words attach to the content word they modify

- Punctuation attach to head of phrase or clause

**The dog was chased by the cat .**

DET NOUN AUX VERB ADP DET NOUN PUNCT

- root → chased
- nsubj:pass: chased → dog
- obl: chased → cat
- punct: chased → .

**Hunden jagades av katten .**

NOUN VERB ADP NOUN PUNCT

- root → jagades
- nsubj:pass: jagades → Hunden
- obl: jagades → katten
- punct: jagades → .

The / DET — dog / NOUN — was / AUX — chased / VERB — by / ADP — the / DET — cat / NOUN — . / PUNCT

det, nsubj:pass, root, punct, obl, det

Hunden / NOUN — jagades / VERB — av / ADP — katten / NOUN — . / PUNCT

Definite=Def     Definite=Def

nsubj:pass, root, punct, obl

**The dog was chased by the cat .**

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

**Hunden jagades av katten .**

- Hunden — NOUN — nsubj:pass — Definite=Def
- jagades — VERB — root — Voice=Pass
- av — ADP
- katten — NOUN — obl — Definite=Def
- . — PUNCT — punct

**The dog was chased by the cat .**

DET NOUN AUX VERB ADP DET NOUN PUNCT

- root → chased
- det: The → dog
- nsubj:pass: dog → chased
- aux:pass: was → chased
- obl: chased → cat
- case: by → cat
- det: the → cat
- punct: chased → .

**Hunden jagades av katten .**

NOUN VERB ADP NOUN PUNCT

Definite=Def    Voice=Pass    Definite=Def

- root → jagades
- nsubj:pass: Hunden → jagades
- obl: jagades → katten
- case: av → katten
- punct: jagades → .

# Syntactic Relations

# Syntactic Relations

Taxonomy of 37 universal syntactic relations

- Three types of structures: nominals, clauses, modifiers

- Core arguments vs. obliques (not arguments vs. adjuncts)

- Language-specific subtypes

# Syntactic Relations

Taxonomy of 37 universal syntactic relations

- Three types of structures: nominals, clauses, modifiers

- Core arguments vs. obliques (not arguments vs. adjuncts)

- Language-specific subtypes

Basic and enhanced representations

- Basic representation forms a (possibly non-projective) tree

- Additional relations in the enhanced representation

# 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 |

# Papers

Nivre, J. (2015) Towards a Universal Grammar of Natural Language Processing. In *Proceedings of CICLing*, 3–16.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R. and Zeman, D. (2016) Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Nivre, J., Marongiu, P., Ginter, F., Kanerva, J., Montemagni, S., Schuster, S. and Simi, M. (2018) Enhancing Universal Dependency Treebanks: A Case Study. In *Proceedings of the Universal Dependencies Workshop (UDW)*.

Tiedemann, J. (2015) Cross-Lingual Dependency Parsing with Universal Dependencies and Predicted PoS Labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 340–349.

Straka, M., Hajic, J. and Straková, J. (2016) UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Reddy, S., Täckström, O., Petrov, S., Steedman, M. and Lapata, M. (2017) Universal Semantic Parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 89-101.

Futrell, R., Mahowald, K. and Gibson, E. (2015) Quantifying Word Order Freedom in Dependency Corpora. In *Proceedings of Depling*, 91–100.

Berzak, Y., Kenney, J., Spadine, C., Wang, J.-X., Lam, L., Mori, K.S., Garza, S. and Katz, B. (2016) Universal Dependencies for Learner English. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 737-746.

Dyer, W. (2018) Integration Complexity and the Order of Cosisters. In *Proceedings of the Universal Dependencies Workshop*.

# Papers

Nivre, J. (2015) Towards a Universal Grammar of Natural Language Processing. In *Proceedings of CICLing*, 3–16.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R. and Zeman, D. (2016) Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Nivre, J., Marongiu, P., Ginter, F., Kanerva, J., Montemagni, S., Schuster, S. and Simi, M. (2018) Enhancing Universal Dependency Treebanks: A Case Study. In *Proceedings of the Universal Dependencies Workshop (UDW)*.

Tiedemann, J. (2015) Cross-Lingual Dependency Parsing with Universal Dependencies and Predicted PoS Labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 340–349.

Straka, M., Hajic, J. and Straková, J. (2016) UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Reddy, S., Täckström, O., Petrov, S., Steedman, M. and Lapata, M. (2017) Universal Semantic Parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 89-101.

Futrell, R., Mahowald, K. and Gibson, E. (2015) Quantifying Word Order Freedom in Dependency Corpora. In *Proceedings of Depling*, 91–100.

Berzak, Y., Kenney, J., Spadine, C., Wang, J.-X., Lam, L., Mori, K.S., Garza, S. and Katz, B. (2016) Universal Dependencies for Learner English. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 737-746.

Dyer, W. (2018) Integration Complexity and the Order of Cosisters. In *Proceedings of the Universal Dependencies Workshop*.

# Papers

Nivre, J. (2015) Towards a Universal Grammar of Natural Language Processing. In *Proceedings of CICLing*, 3–16.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R. and Zeman, D. (2016) Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Nivre, J., Marongiu, P., Ginter, F., Kanerva, J., Montemagni, S., Schuster, S. and Simi, M. (2018) Enhancing Universal Dependency Treebanks: A Case Study. In *Proceedings of the Universal Dependencies Workshop (UDW)*.

Tiedemann, J. (2015) Cross-Lingual Dependency Parsing with Universal Dependencies and Predicted PoS Labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 340–349.

Straka, M., Hajic, J. and Straková, J. (2016) UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Reddy, S., Täckström, O., Petrov, S., Steedman, M. and Lapata, M. (2017) Universal Semantic Parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 89-101.

Futrell, R., Mahowald, K. and Gibson, E. (2015) Quantifying Word Order Freedom in Dependency Corpora. In *Proceedings of Depling*, 91–100.

Berzak, Y., Kenney, J., Spadine, C., Wang, J.-X., Lam, L., Mori, K.S., Garza, S. and Katz, B. (2016) Universal Dependencies for Learner English. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 737-746.

Dyer, W. (2018) Integration Complexity and the Order of Cosisters. In *Proceedings of the Universal Dependencies Workshop*.

# Papers

Nivre, J. (2015) Towards a Universal Grammar of Natural Language Processing. In *Proceedings of CICLing*, 3–16.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R. and Zeman, D. (2016) Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Nivre, J., Marongiu, P., Ginter, F., Kanerva, J., Montemagni, S., Schuster, S. and Simi, M. (2018) Enhancing Universal Dependency Treebanks: A Case Study. In *Proceedings of the Universal Dependencies Workshop (UDW)*.

Tiedemann, J. (2015) Cross-Lingual Dependency Parsing with Universal Dependencies and Predicted PoS Labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 340–349.

Straka, M., Hajic, J. and Straková, J. (2016) UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Reddy, S., Täckström, O., Petrov, S., Steedman, M. and Lapata, M. (2017) Universal Semantic Parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 89-101.

Futrell, R., Mahowald, K. and Gibson, E. (2015) Quantifying Word Order Freedom in Dependency Corpora. In *Proceedings of Depling*, 91–100.

Berzak, Y., Kenney, J., Spadine, C., Wang, J.-X., Lam, L., Mori, K.S., Garza, S. and Katz, B. (2016) Universal Dependencies for Learner English. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 737-746.

Dyer, W. (2018) Integration Complexity and the Order of Cosisters. In *Proceedings of the Universal Dependencies Workshop*.

# Papers

Nivre, J. (2015) Towards a Universal Grammar of Natural Language Processing. In *Proceedings of CICLing*, 3–16.

Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R. and Zeman, D. (2016) Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Nivre, J., Marongiu, P., Ginter, F., Kanerva, J., Montemagni, S., Schuster, S. and Simi, M. (2018) Enhancing Universal Dependency Treebanks: A Case Study. In *Proceedings of the Universal Dependencies Workshop (UDW)*.

Tiedemann, J. (2015) Cross-Lingual Dependency Parsing with Universal Dependencies and Predicted PoS Labels. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 340–349.

Straka, M., Hajic, J. and Straková, J. (2016) UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Reddy, S., Täckström, O., Petrov, S., Steedman, M. and Lapata, M. (2017) Universal Semantic Parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 89-101.

Futrell, R., Mahowald, K. and Gibson, E. (2015) Quantifying Word Order Freedom in Dependency Corpora. In *Proceedings of Depling*, 91–100.

Berzak, Y., Kenney, J., Spadine, C., Wang, J.-X., Lam, L., Mori, K.S., Garza, S. and Katz, B. (2016) Universal Dependencies for Learner English. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 737-746.

Dyer, W. (2018) Integration Complexity and the Order of Cosisters. In *Proceedings of the Universal Dependencies Workshop*.

# Research Problems

## Annotation

- Produce guidelines and/or annotation for language X
- Study the annotation of construction Y across languages

## Parsing

- Develop and/or evaluate a parser for language X
- Study cross-lingual transfer learning and/or annotation projection

## Typology

- Study word order patterns in language X
- Compare the realisation of construction Y across languages

# Questions?