



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

HDT-UD: A VERY LARGE UNIVERSAL DEPENDENCIES TREEBANK FOR GERMAN

EMANUEL BORGES VÖLKER, MAXIMILIAN WENDT, FELIX HENNIG AND ARNE KÖHN

THE HDT: A LARGE DEPENDENCY TREEBANK

- **OVER 260.000 SENTENCES (4.7 MILLION TOKENS)**
- **SOURCED FROM HEISE.DE**
→ **TECH NEWS ARTICLES**
- **TEXT PUBLISHED BETWEEN 1996 AND 2001**
- **AVERAGE SENTENCE LENGTH OF 18.4**

THE HDT: A LARGE DEPENDENCY TREEBANK

PART A

- 101 999 SENTENCES
- MANUALLY ANNOTATED
- CHECKED FOR CONSISTENCY WITH DECCA

PART B

- 104 795 SENTENCES
- MANUALLY ANNOTATED

PART C

- 55 027 SENTENCES
- AUTOMATICALLY PARSED

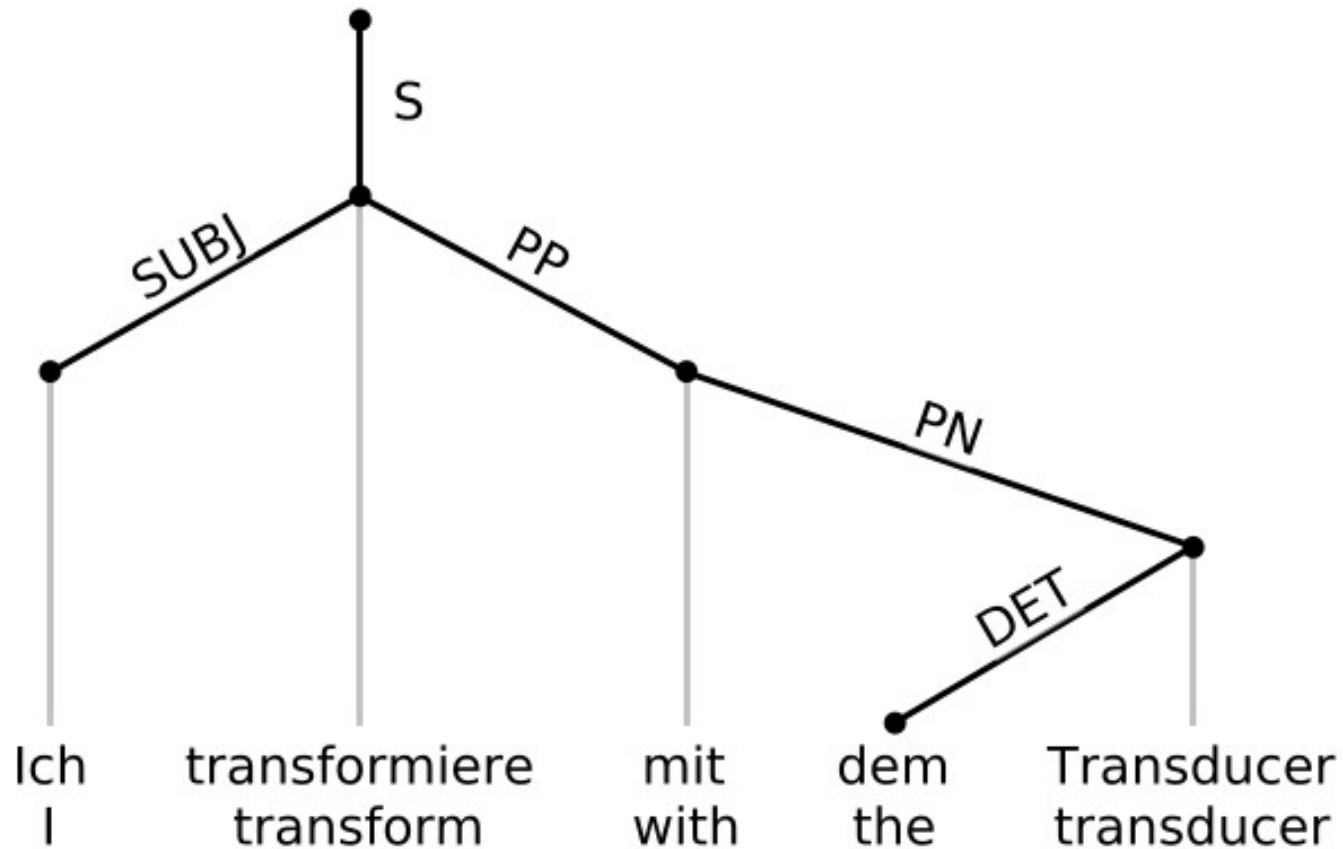


CONVERTED TO UNIVERSAL DEPENDENCIES

- 187 585 SENTENCES

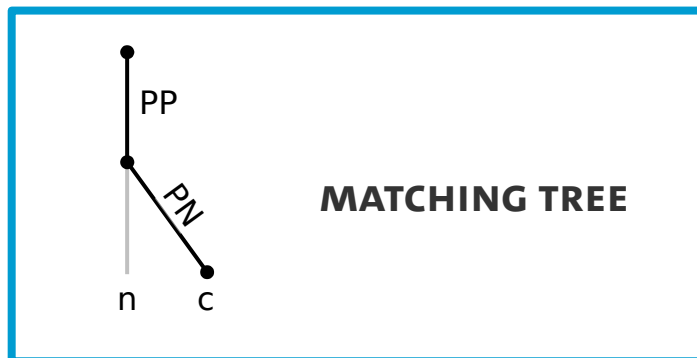
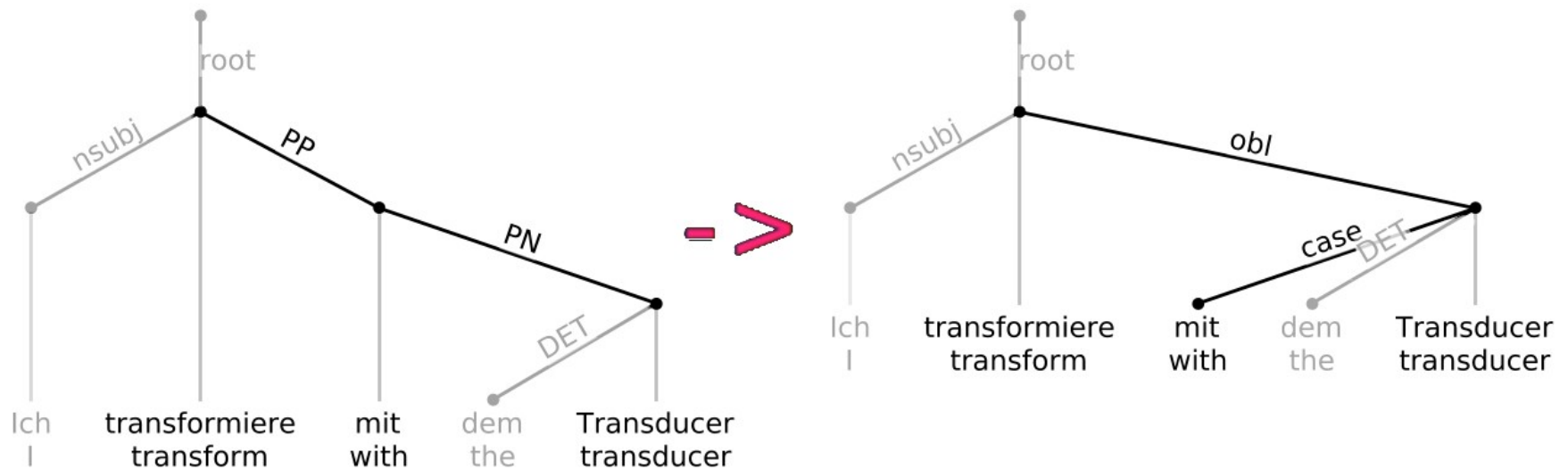
TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017

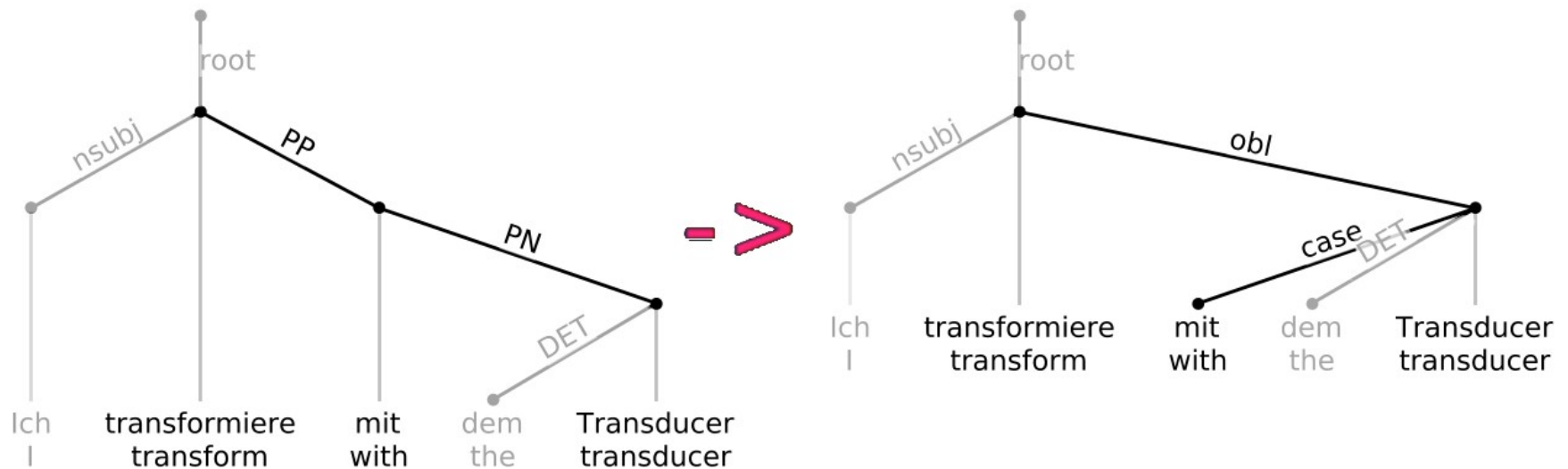


->



TREE TRANSDUCING WITH TRUDUCER

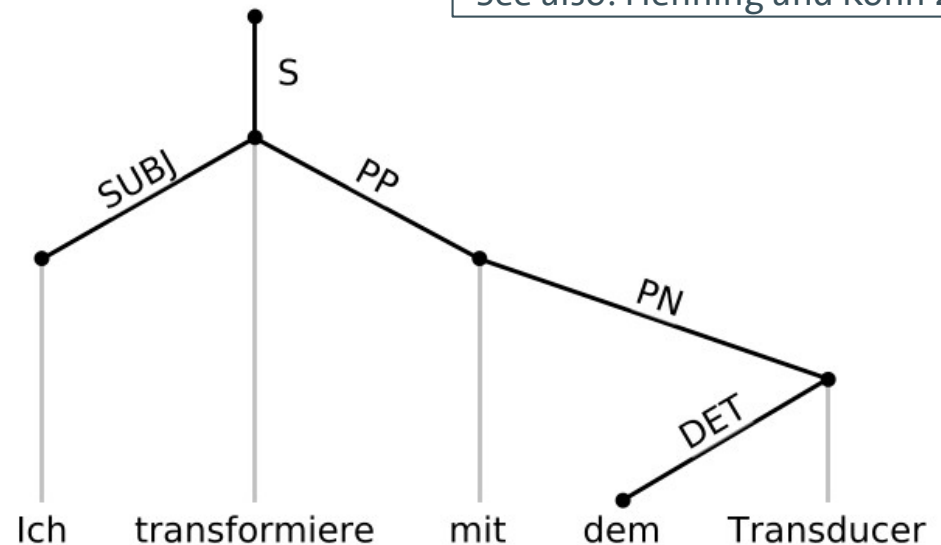
See also: Henning and Köhn 2017



$n:PP(c:PN()) \rightarrow c:obl(n:case()) ;$

TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



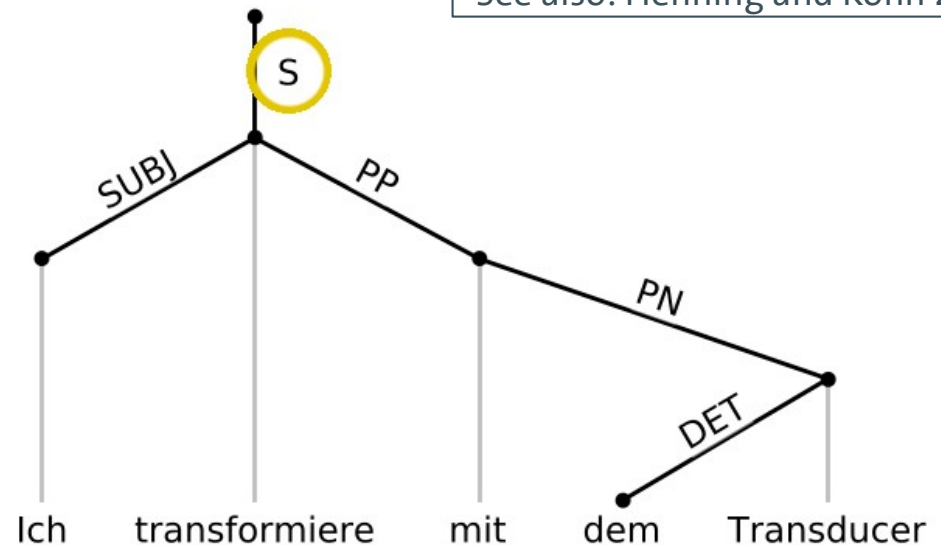
conversion_rules.txt ✕

```

1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
  
```

TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



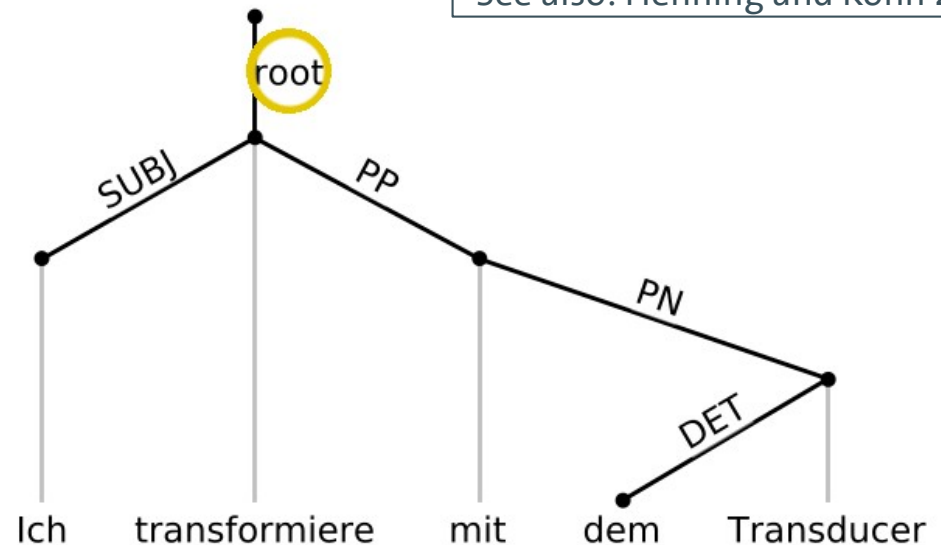
conversion_rules.txt ✕

```

1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
  
```


TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



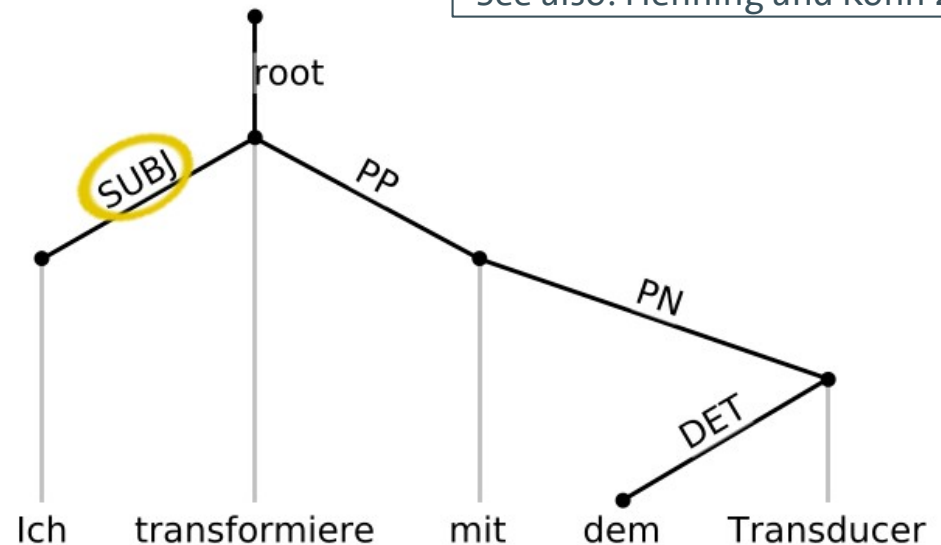
conversion_rules.txt ✕

```

1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
  
```

TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



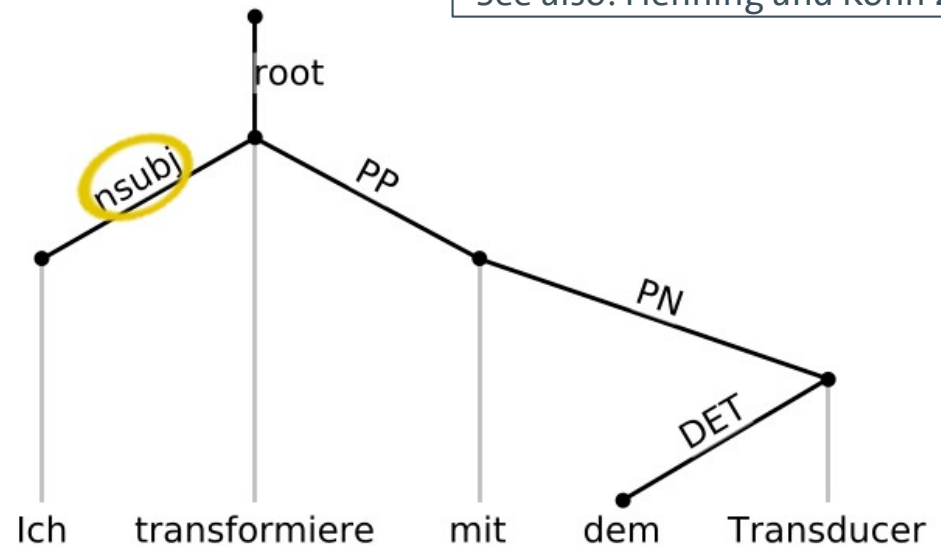
conversion_rules.txt ✕

```

1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
  
```

TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



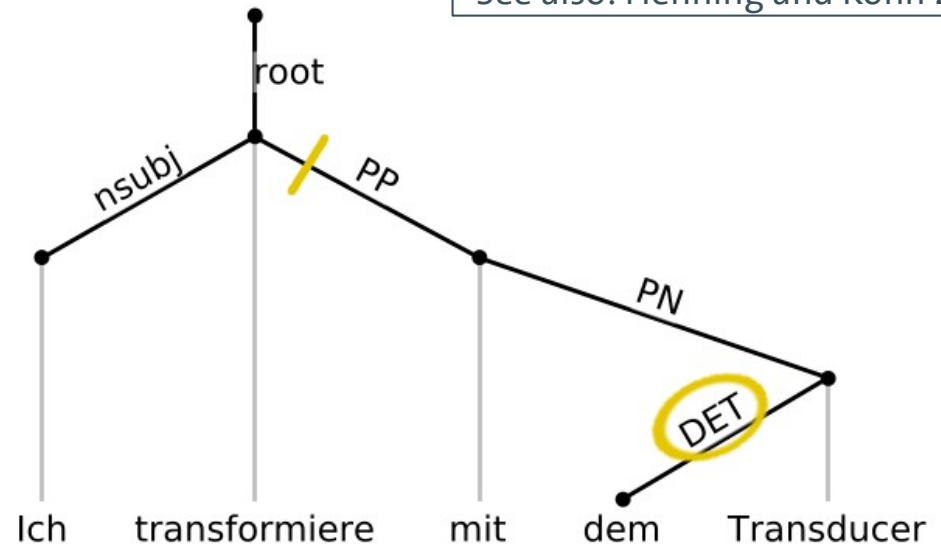
conversion_rules.txt ✕

```

1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
  
```

TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017

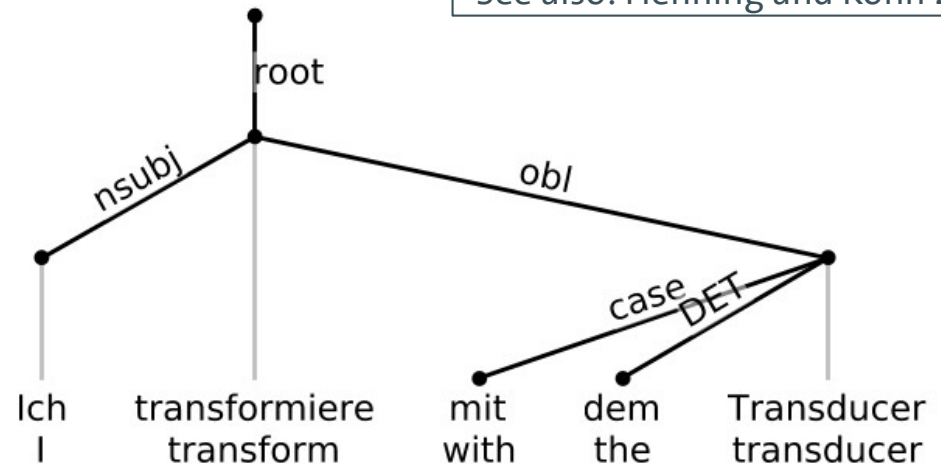


conversion_rules.txt ✕

```
1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
```

TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



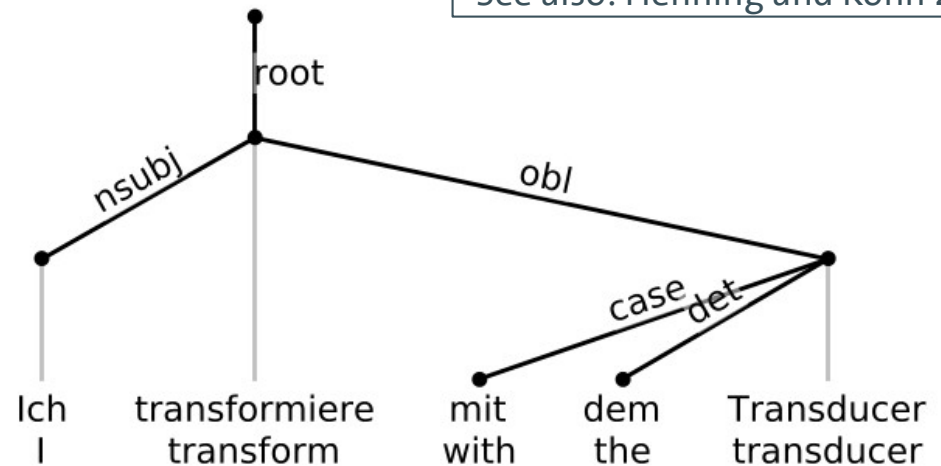
conversion_rules.txt ✕

```

1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
  
```

TREE TRANSDUCING WITH TRUDUCER

See also: Henning and Köhn 2017



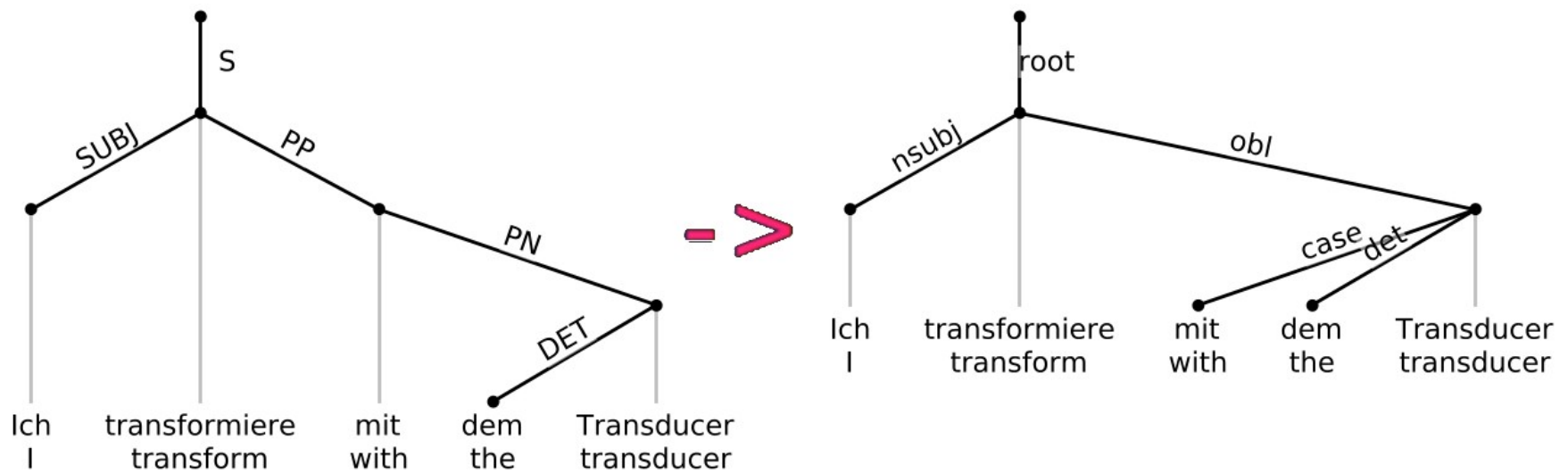
conversion_rules.txt ✕

```

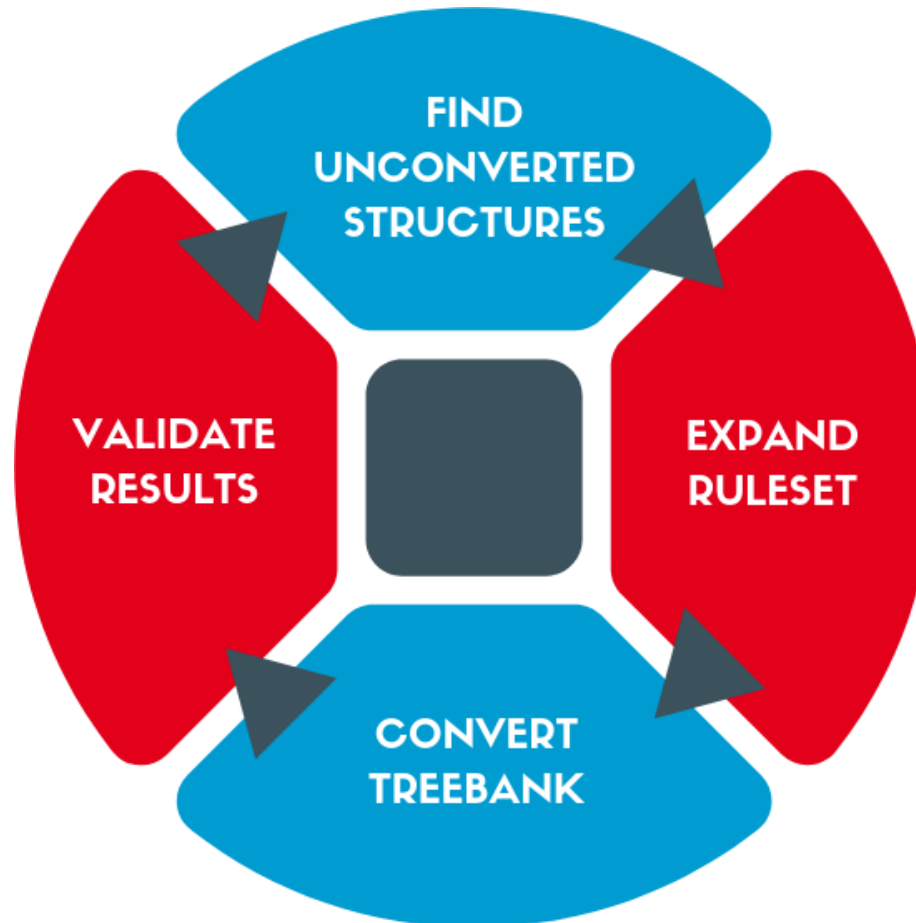
1 n:S() -> n:root();      # convert root node
2 n:SUBJ() -> n:nsubj();  # convert subject
3 n:DET() -> n:det();     # convert determiner
4 # convert prepositional phrases
5 n:PP( c:PN() ) -> c:obl( n:case() );
  
```


TREE TRANSDUCING WITH TRUDUCER

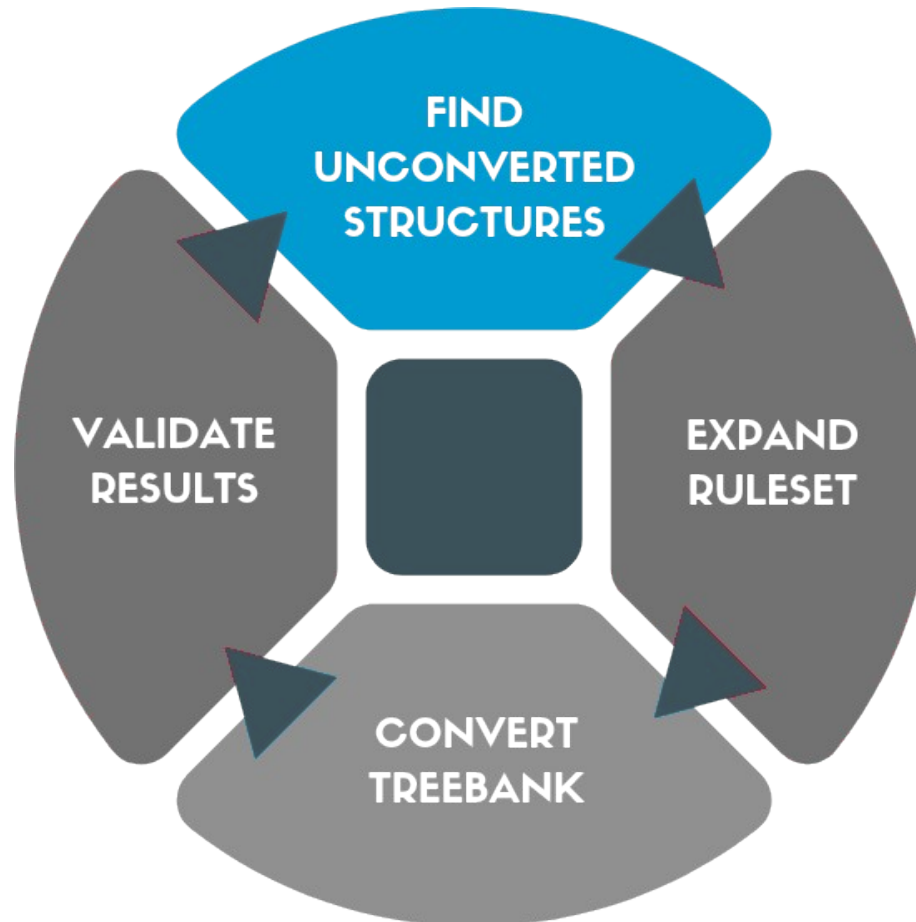
See also: Henning and Köhn 2017



TREEBANK CONVERSION WORKFLOW

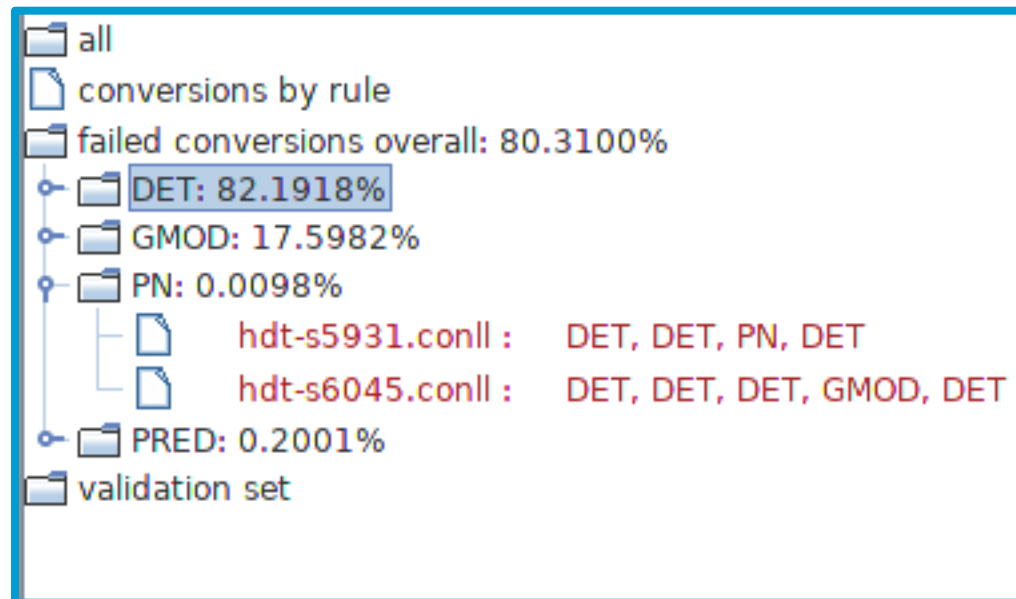


TREEBANK CONVERSION WORKFLOW



FIND UNCONVERTED STRUCTURES

LIST AND ORDER UNCONVERTED SENTENCES



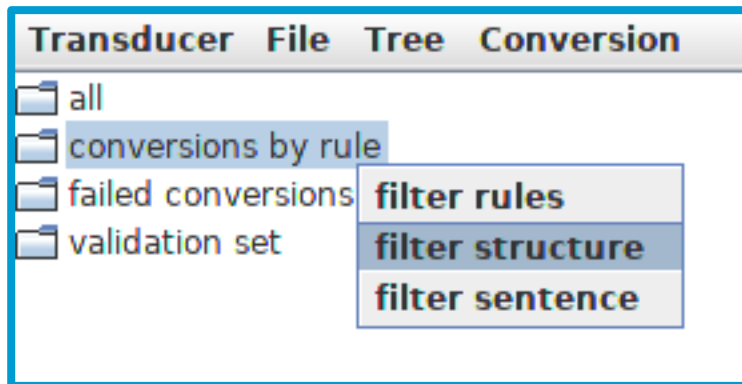
```

all
├── conversions by rule
├── failed conversions overall: 80.3100%
├── DET: 82.1918%
├── GMOD: 17.5982%
├── PN: 0.0098%
├── PRED: 0.2001%
└── validation set
  
```

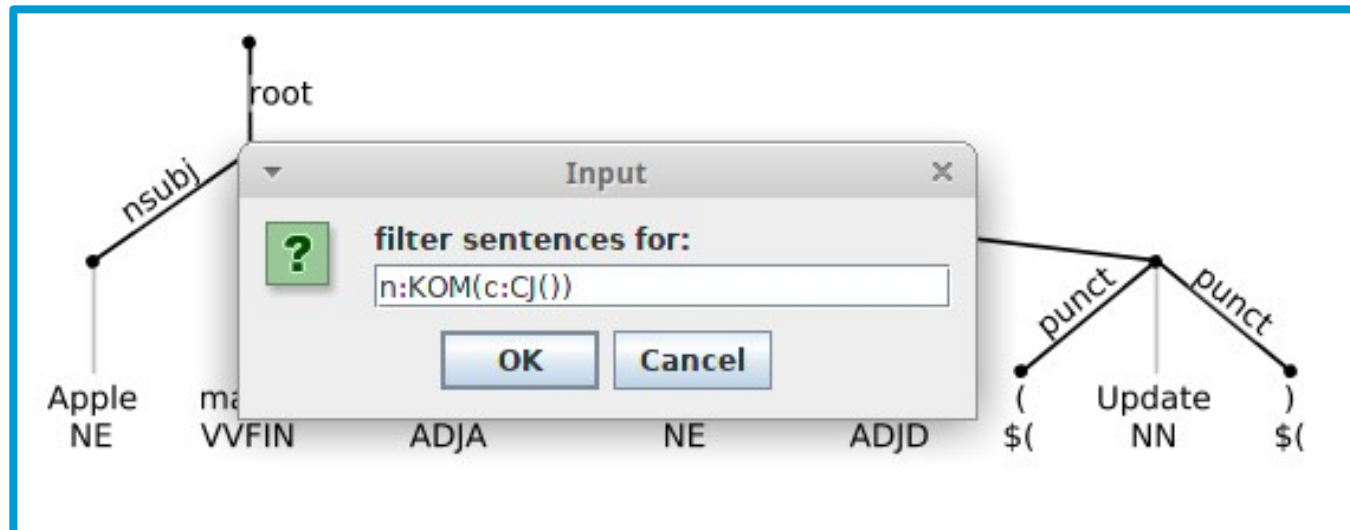
hdt-s5931.conll : DET, DET, PN, DET

hdt-s6045.conll : DET, DET, DET, GMOD, DET

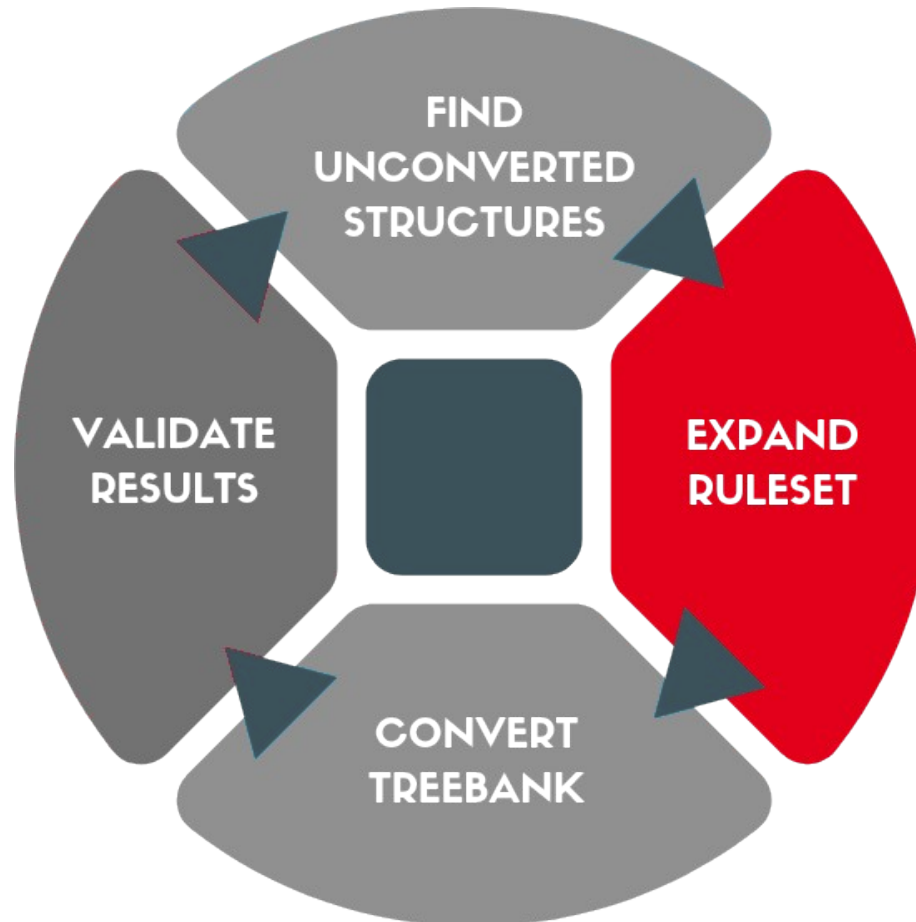
FIND UNCONVERTED STRUCTURES



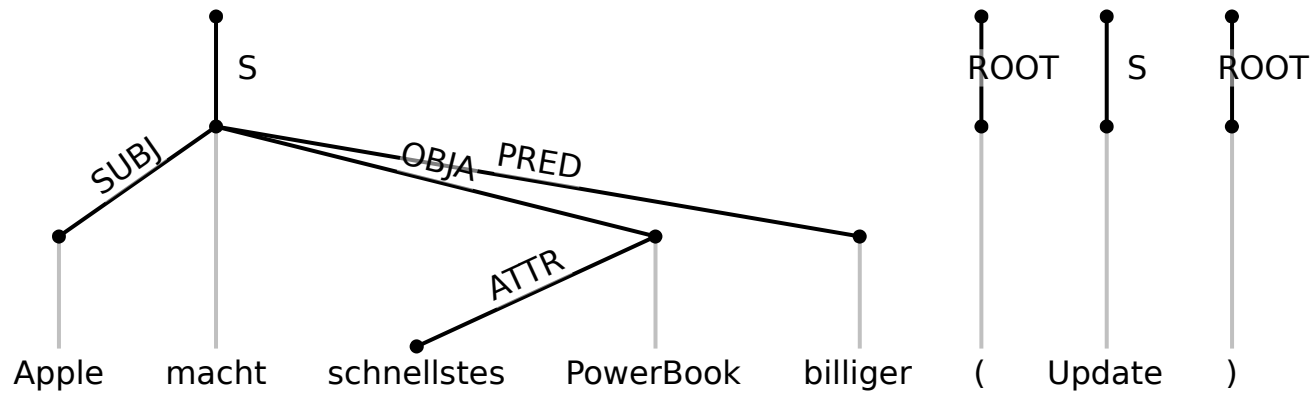
**SORT AND STRUCTURE
DEPENDENCY TREEBANK**



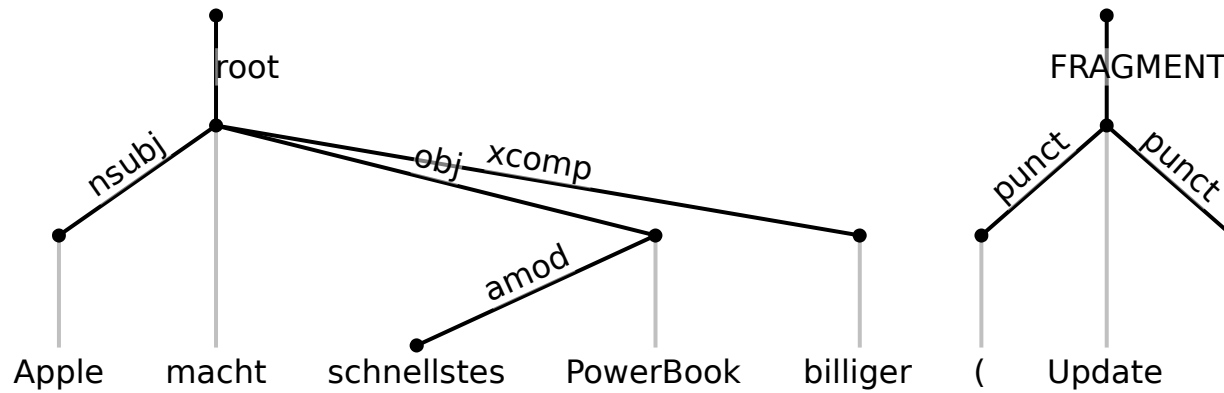
TREEBANK CONVERSION WORKFLOW



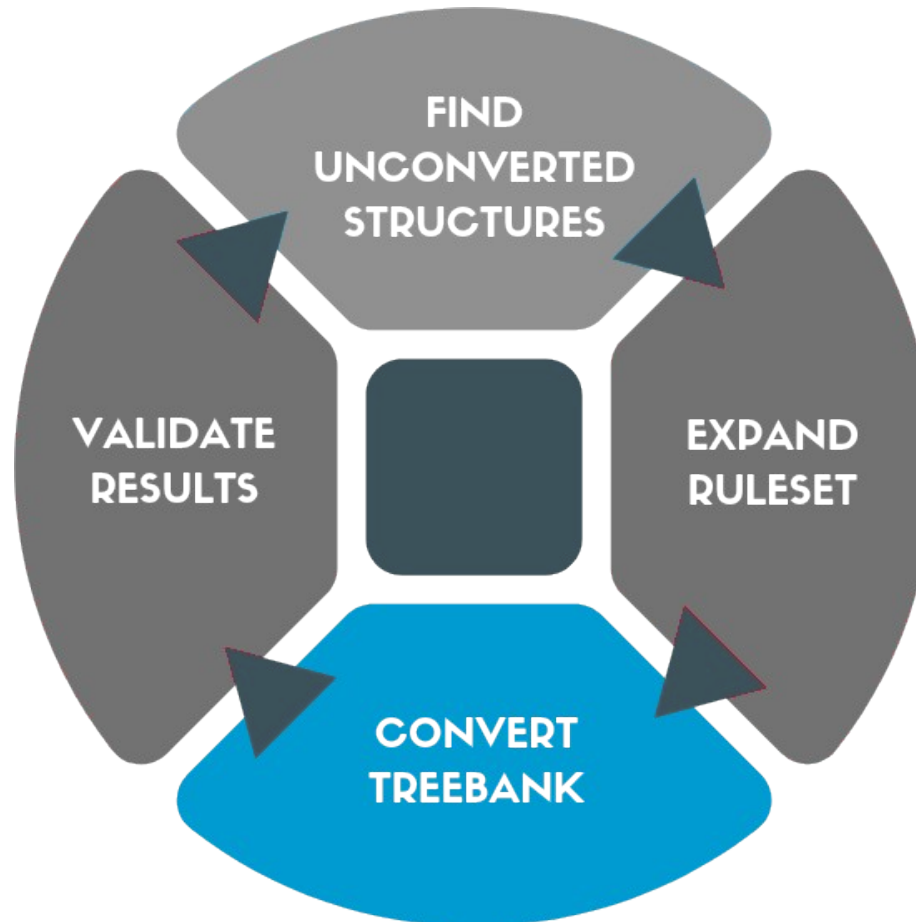
EXPAND RULESET



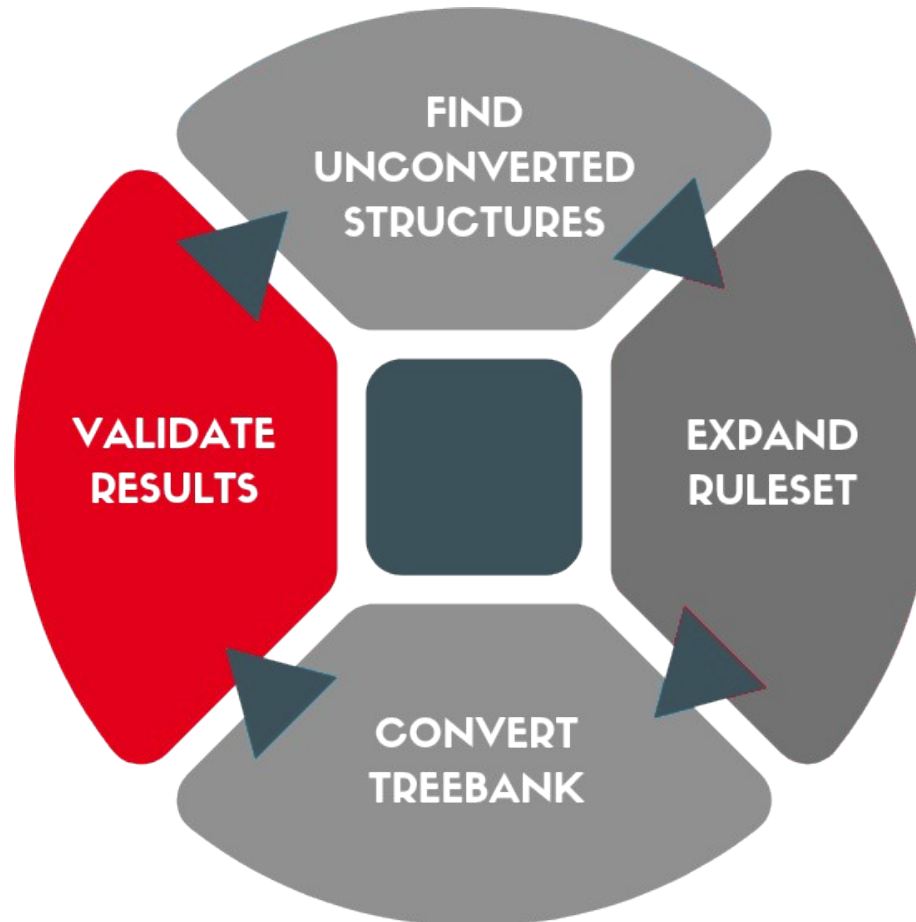
EXPAND RULESET



TREEBANK CONVERSION WORKFLOW



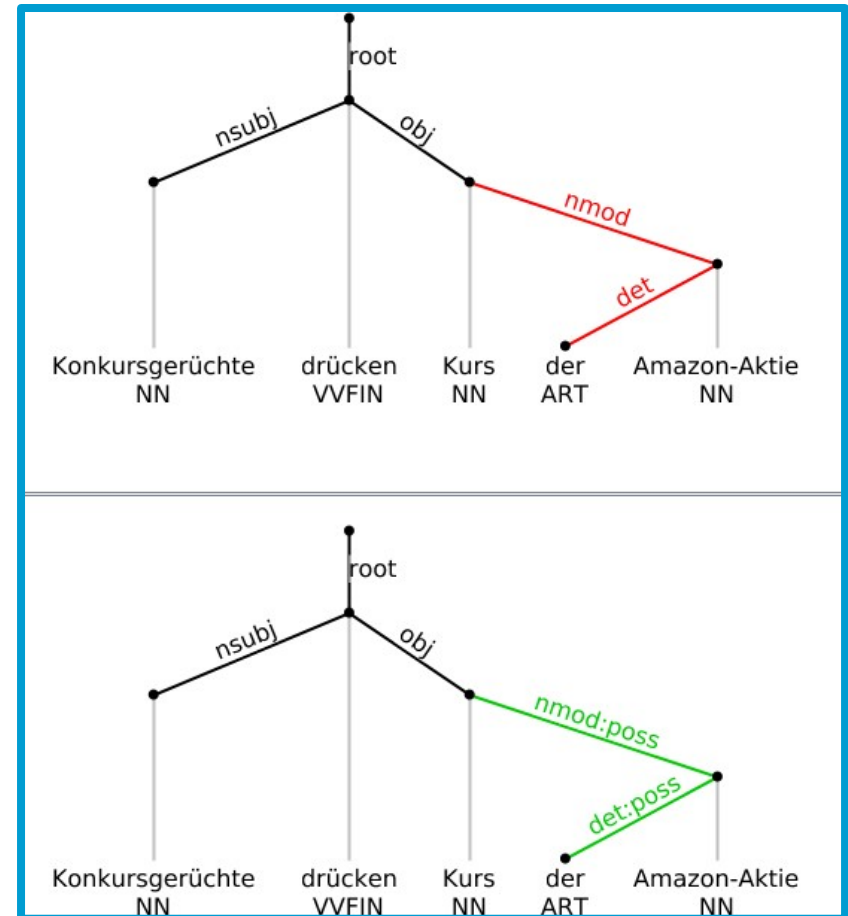
TREEBANK CONVERSION WORKFLOW



VALIDATE RESULTS

TEST RESULTS FOR ERRORS AND INCONSISTENCIES

Transducer	File	Tree	Conversion
validation set			<input checked="" type="checkbox"/> remember rules used
	hdt-s1.conll :		convert all
	hdt-s1.conll :		validate
	hdt-s10.conll :		show rule statistics



INHERENTLY REFLEXIVE VERBS

sputen

sputen (Deutsch) [Bearbeiten]

Verb [Bearbeiten]

Worttrennung:

spu·ten, Präteritum: spu·te·te, Partizip II: ge·spu·tet

Aussprache:

IPA: [ˈʃpu:tən]

Hörbeispiele:  sputen [\(Info\)](#),  sputen [\(Info\)](#)

Reime: -u:tən

Bedeutungen:

[1] *reflexiv*: sich beeilen

INHERENTLY REFLEXIVE VERBS

öffnen

öffnen (Deutsch) [Bearbeiten]

Verb [Bearbeiten]

Worttrennung:

öff·nen, Präteritum: öff·ne·te, Partizip II: ge·öff·net

Aussprache:

IPA: ['œfnən]

Hörbeispiele: öffnen (Info), öffnen (Österreich) (Info)

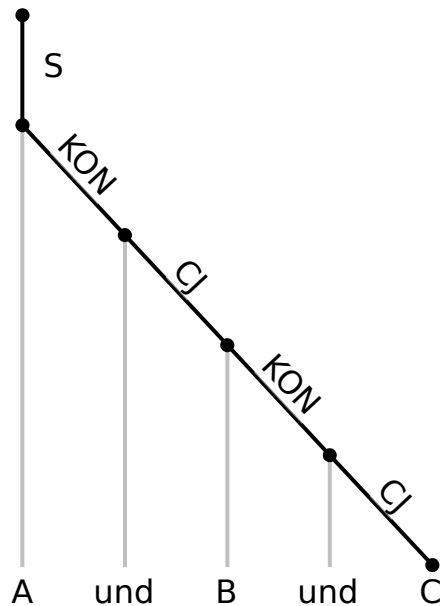
Bedeutungen:

[1] etwas **aufmachen**; **offen** machen

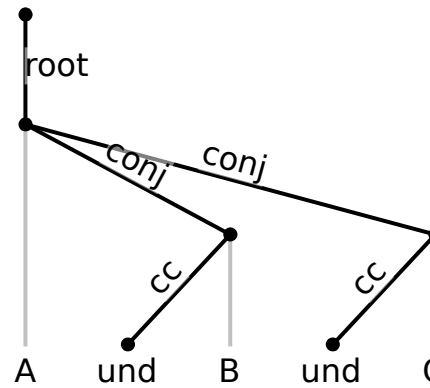
[2] *sich* **öffnen**: **aufgehen**; offen werden

[3] *sich* (etwas) **öffnen**: einer Sache nicht mehr **widerstrebend gegenüberstehen**

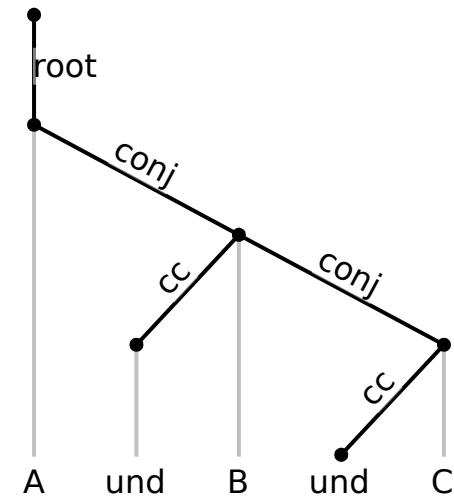
NESTED COORDINATION



HDT conjunction



standard UD conjunction



explicitly nested UD
conjunction

- **187 585 CONVERTED SENTENCES**
- **ACCURACY OF 97,3% COMPARED TO GOLD STANDARD**
- **NEW AUTOMATIC CONVERSION TOOL**



**THANK YOU FOR
LISTENING!**