Conversion of Redwoods

"Conversion of wood is whereby the log/tree is altered to create planks, timbers, or other desired elements"

Angelina Ivanova

March 28, 2012

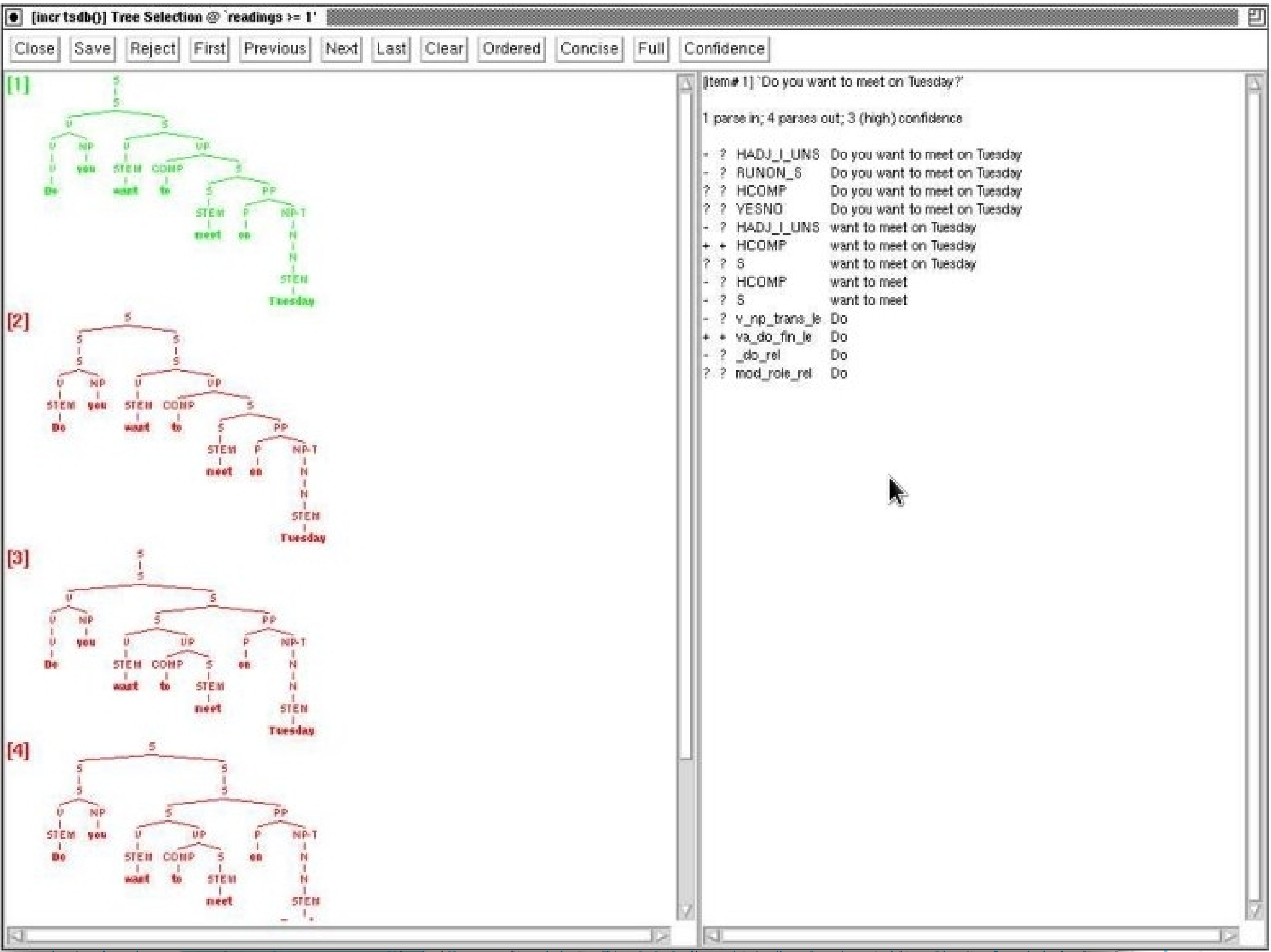
The LinGO Redwoods Treebank

- Language: **English**
- Linguistic approach: **HPSG**
- Grammar: **LinGO ERG**
- Annotation style:
  - syntactic trees,
  - syntactic dependencies,
  - POS

The LinGO Redwoods Treebank

Data:

- Verbmobil and e-commerce corpora
- LOGON Norwegian-English MT corpus
- English Wikipedia (from WeScience)
- Brown corpus (SemCor)
- other


[incr tsdb()] Tree Selection @ `readings >= 1'
Close
Save
Reject
First
Previous
Next
Last
Clear
Ordered
Concise
Full
Confidence
[1]
[2]
[3]
[4]
[Item# 1] `Do you want to meet on Tuesday?'
1 parse in; 4 parses out; 3 (high) confidence
- ? HADJ_I_UNS Do you want to meet on Tuesday
- ? RUNON_S Do you want to meet on Tuesday
? ? HCOMP Do you want to meet on Tuesday
? ? YESNO Do you want to meet on Tuesday
- ? HADJ_I_UNS want to meet on Tuesday
+ + HCOMP want to meet on Tuesday
? ? S want to meet on Tuesday
- ? HCOMP want to meet
- ? S want to meet
- ? v_np_trans_le Do
+ + va_do_fin_le Do
- ? _do_rel Do
? ? mod_role_rel Do

DELPH-IN Syntactic Derivation Tree

- ERG grammar
- HPSG framework
- Fully independent of the PTB
- Only unary and binary branches

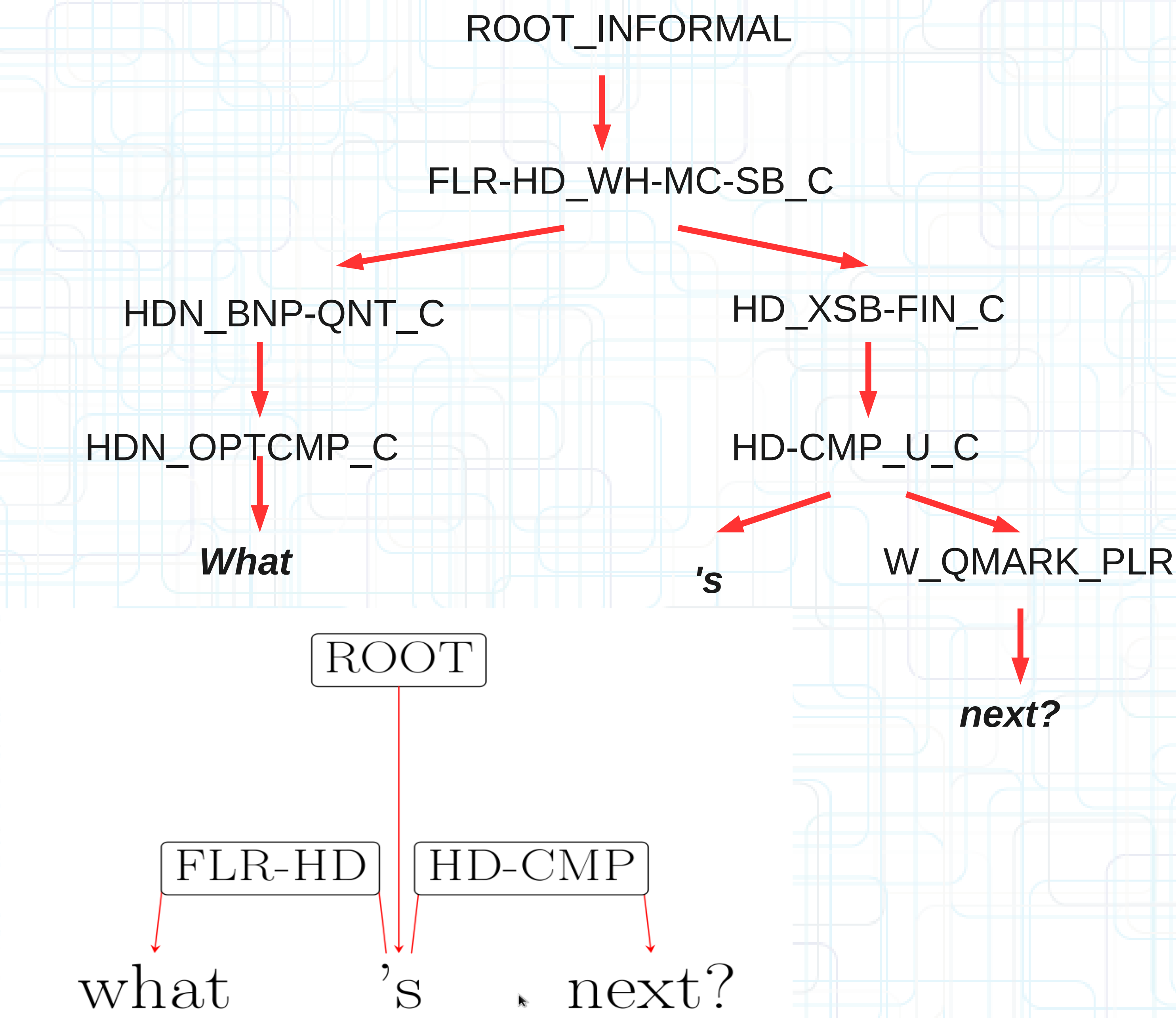
DELPH-IN Minimal Recursion Semantics

- MSR proper utilizes predicate calculus
- We work with MRS reduced to Elementary Dependency Structures

Elementary Dependency Structures

- Predicates that correspond to groups of words
- Some words are semantically empty
- EDS doesn't form a tree

DERIVATION TREE CONVERSION



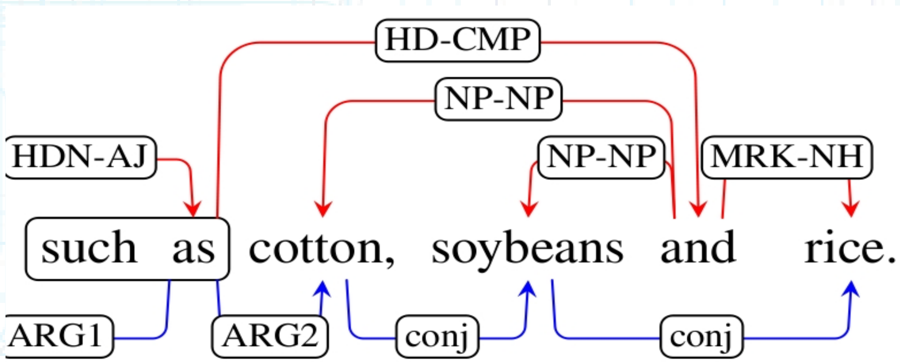
EDS conversion

<u>Difficult cases:</u>

- contracted negation
- punctuation

The dog ***couldn't*** *bark.*

- multiword expressions



EDS conversion

‘Special’ classes of relations:

- transparent
- relational
- redundant

Transparent relations

The relation is equated with one of its arguments

```
[transparent]
nominalization ARG1
implicit_conj L-HNDL
implicit_conj L-INDEX
/_c$/ L-HNDL
/_c$/ L-INDEX
```

Transparent relations

Root

e9 e2

The dog arrived and barked.

```
{e2:
 _1:_the_q<0:3>[BV x5]
 x5:_dog_n_1<4:7>[]
 e9:_arrive_v_1<8:15>[ARG1 x5]
 e2:_and_c<16:19>[L-INDEX e9, R-INDEX e11,
L-HNDL e9, R-HNDL e11]
 e11:_bark_v_1<20:27>[ARG1 x5]
}
```

Relational predicates

For some predicate symbols that have two arguments, the first is the head and the second is its dependent in the word-to-word relation.

```
/_c$/ L-HNDL R-HNDL

/_c$/ L-INDEX R-INDEX

of_p ARG2 ARG1

part_of ARG0 ARG1

poss ARG2 ARG1
```

Relational predicates

x12 x10

Browne arrived on Tuesday morning.

```
{e2:
 _1:proper_q<0:6>[BV x5]
 x5:named<0:6>("Browne")[]
 e2:_arrive_v_1<7:14>[ARG1 x5]
 e9:_on_p_temp<15:17>[ARG1 e2, ARG2 x10]
 x12:dofw<18:25>("Tue")[]
 _2:def_explicit_q<18:25>[BV x10]
 e17:of_p<18:25>[ARG1 x10, ARG2 x12]
 _3:def_implicit_q<18:25>[BV x12]
 x10:_morning_n_of<26:34>[]
}
```

Redundant relations

The predicate has an argument that should be purged.

```
[redundant]
/.*/ L-HNDL L-INDEX
/.*/ R-HNDL R-INDEX
```

Redundant relations

e9 e11

The dog arrived and barked.

```
{e2:
 _1:_the_q<0:3>[BV x5]
 x5:_dog_n_1<4:7>[]
 e9:_arrive_v_1<8:15>[ARG1 x5]
 e2:_and_c<16:19>[L-INDEX e9, R-INDEX e11,
L-HNDL e9, R-HNDL e11]
 e11:_bark_v_1<20:27>[ARG1 x5]
}
```

Lexical relations

Relations that start with an underscore but yet can be associated with a surface token

```
[lexical]

/^_.*/

named

card

thing

numbered_hour

person

pron

time
```

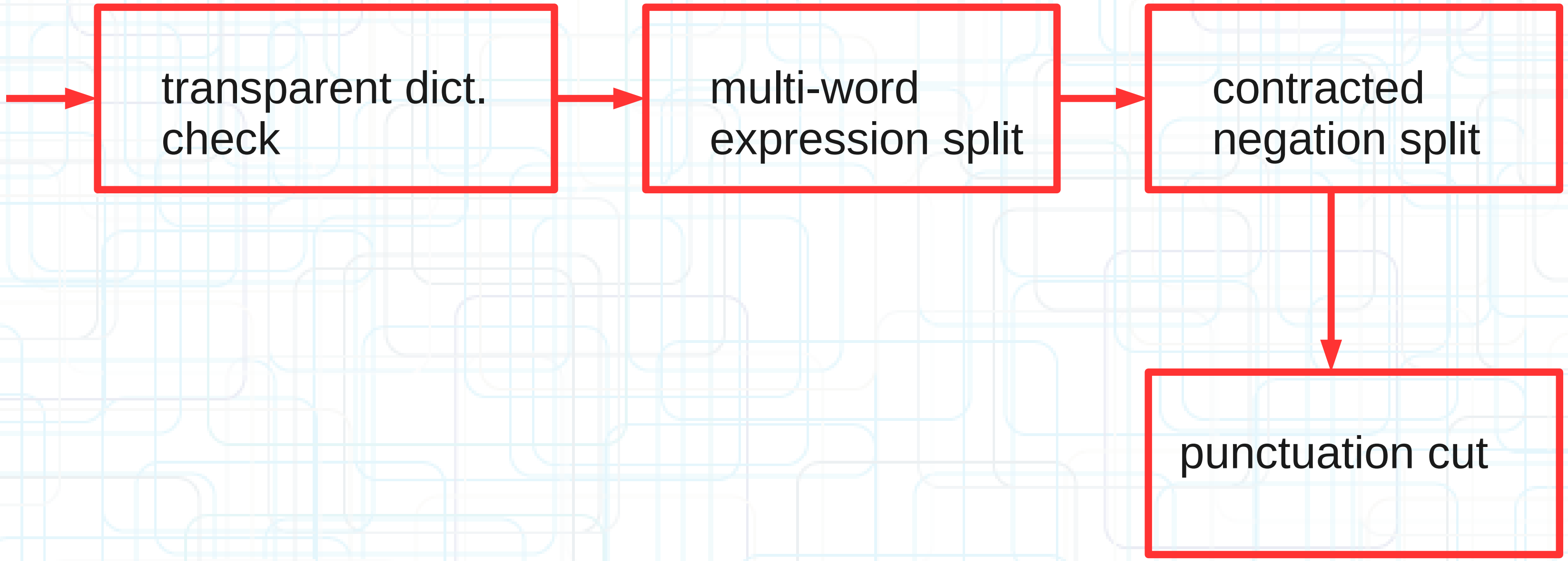
Lexical relations

i18 e10 i16 x4 e2

Two hundred twenty dogs bark.

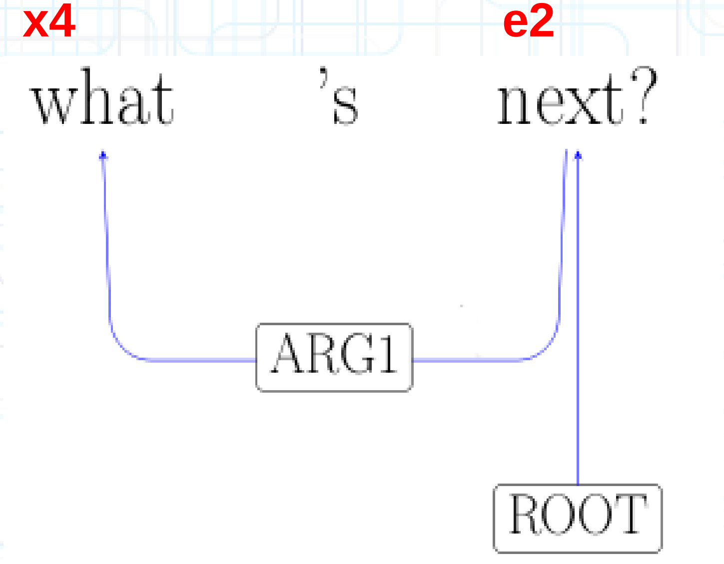
```
{e2:
 _1:udef_q<0:18>[BV x4]
 i8:card<0:3>("2")[ARG1 x4]
 e10:card<4:11>("100")[ARG1 x4]
 i14:plus<4:11>[ARG1 x4, ARG2 i15, ARG3 i16]
 i15:times<4:11>[ARG1 x4, ARG2 i8, ARG3 e10]
 i16:card<12:18>("20")[ARG1 x4]
 x4:_dog_n_1<19:23>[]
 e2:_bark_v_1<24:29>[ARG1 x4]
}
```

EDS conversion

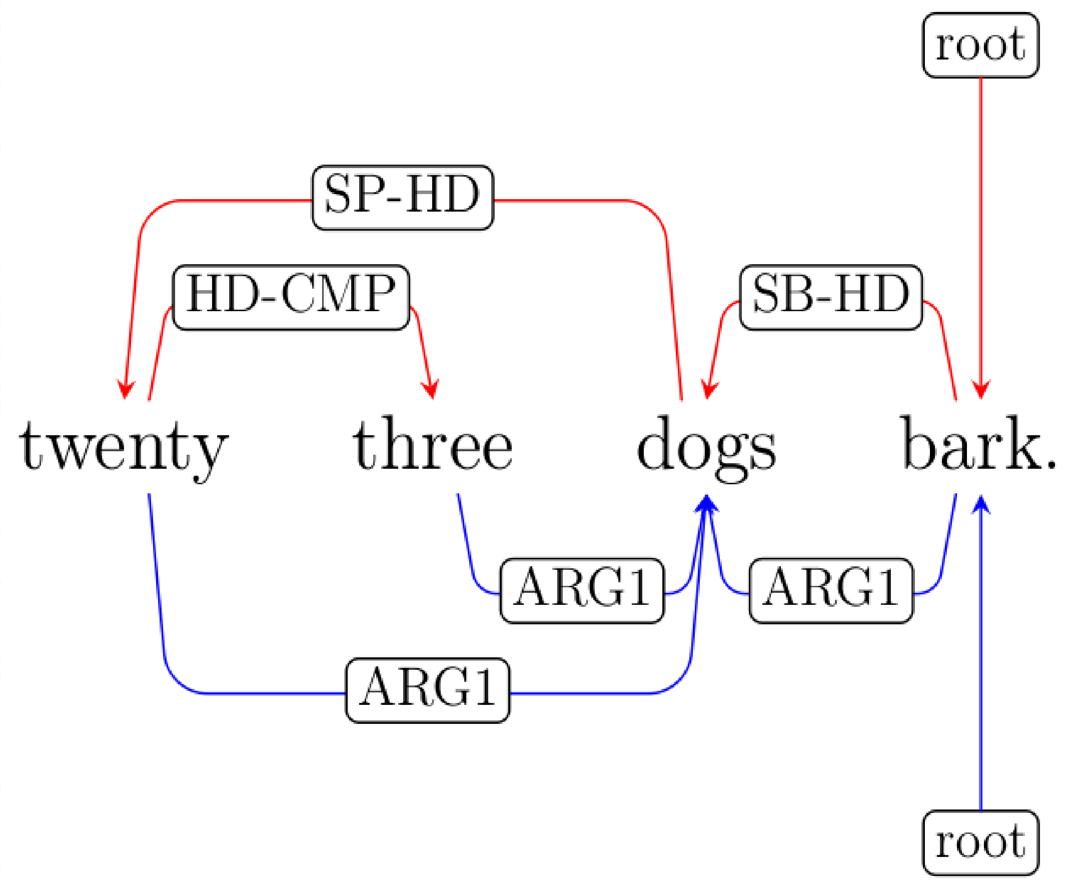


EDS conversion

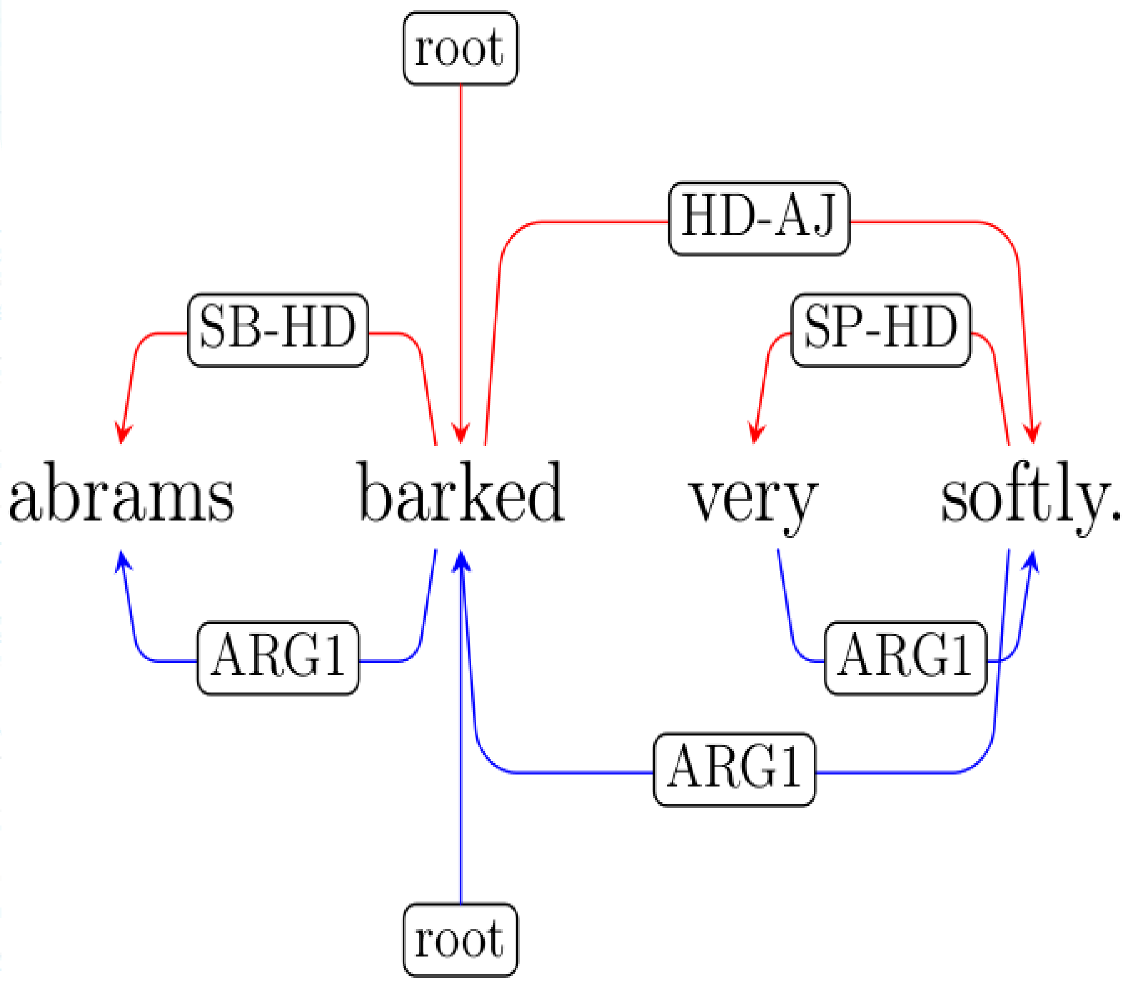
```
{e2:
 x4:thing<0:4>[]
 _1:which_q<0:4>[BV x4]
 e2:_next_a_1<7:12>[ARG1 x4]
}
```



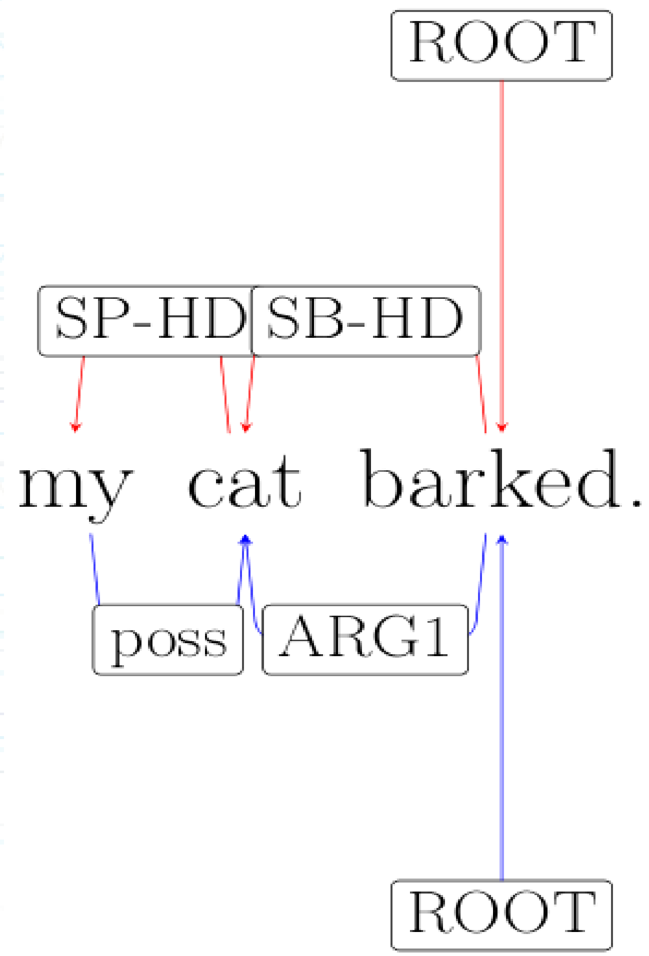
Gold standard MRS-ERG collection conversion



Gold standard MRS-ERG collection conversion



Gold standard MRS-ERG collection conversion



PEST corpus

Language: **English**

Two sets: **10 sentences** and **15 sentences**

Formats:

- CoNLL Syntactic Dependencies
- CoNLL PropBank Semantics
- Stanford basic
- Stanford collapsed dependencies
- Enju predicate – argument structures

CoNLL Syntactic Dependencies

- PTB trees converted with the PennConverter software
- Functional heads
- The dependency graph is a directed tree:
  - every token in the sentence is a node in the graph
  - the graph is connected
  - every node in the graph has at most one head
  - the graph is acyclic

CoNLL PropBank Semantics

- The PropBank and NomBank annotations ‘on top’ of the PTB syntax
- Lexical heads
- Arguments with several syntactic heads
- Discontinuous arguments
- Empty categories

Stanford Basic Dependencies

- Converted from PTB phrase trees
- Lexical heads
- The dependency graph is a directed tree:
  - every token in the sentence is a node in the graph
  - the graph is connected
  - every node in the graph has at most one head
  - the graph is acyclic

Stanford Standard Dependencies

- Functional heads
- Dependency graph is not a tree:
  - Semantically empty words
  - Multiple heads
  - Cycles

Enju Predicate – Argument Structures (EP)

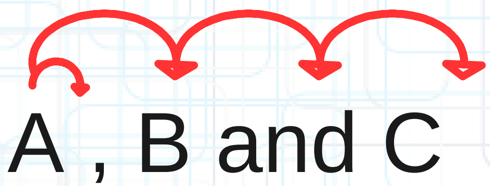
- Lexical heads
- Semi-automatical HPSG-conversion from PTB
- Dependency graph is not a tree

Root choice

A similar technique is almost impossible to apply to other crops, such as cotton, soybeans and rice.

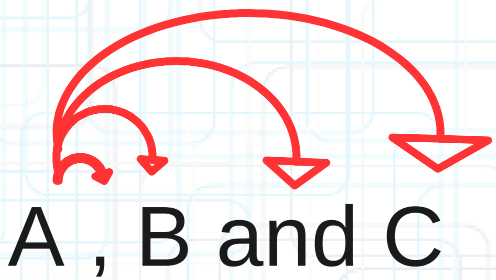
CoNLL Syntactic: **is**
CoNLL PropBank: -
Stanford Basic: **impossible**
Stanford Standard: **impossible**
Enju Predicate-Argument Structures: **is**
DELHP-IN Derivation Tree: **is**
DELPH-IN MRS: **almost**

Conjunction

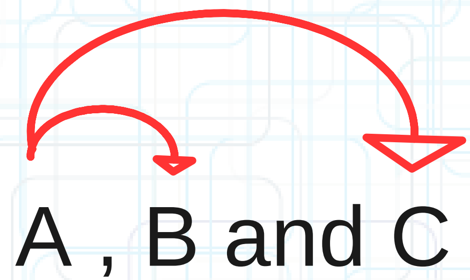


CoNLL Syntactic Dependencies

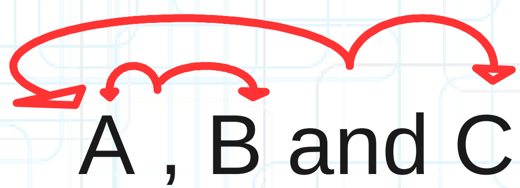
A , B and C
CoNLL PropBank Dependencies



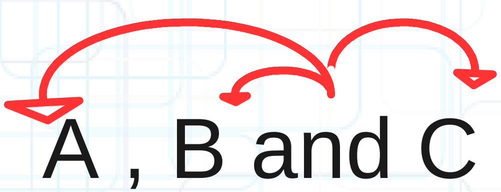
Stanford Basic Dependencies



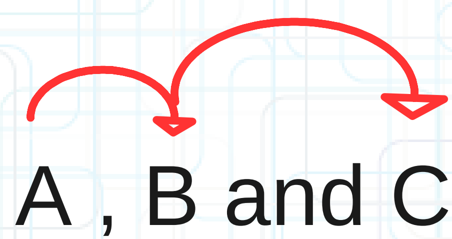
Stanford Standard Dependencies



Enju PAS



DELHP-IN Derivation Tree



DELHP-IN MRS

Infinitive

- CoNLL syntactic
- Enju PAS
- DELPH-IN Derivation Tree

to apply

- *Stanford Basic*
- *Stanford Standard*

to apply

- *CoNLL PropBank*
- *DELPH-IN MRS*

-

Article

- CoNLL Syntactic
- Stanford Basic
- Stanford Standard
- DELPH-IN Derivation Tree

a technique

- Enju PAS
- DELPH-IN MRS

a technique

- CoNLL PropBank

Adjective

- CoNLL Syntactic
- Stanford Basic
- Stanford Standard
- DELPH-IN Derivation Tree

similar technique

- Enju PAS
- DELPH-IN MRS

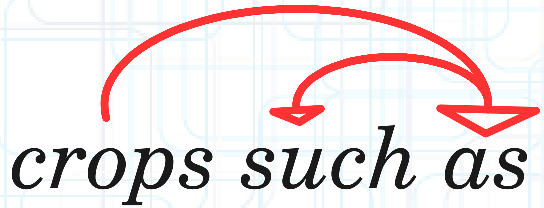
similar technique

- CoNLL PropBank

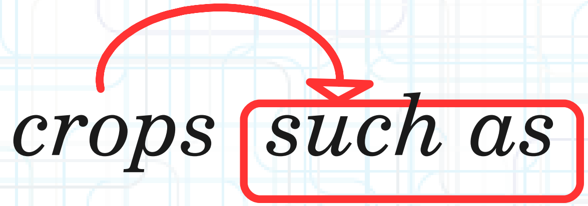
-

Prepositions

- CoNLL Syntactic
- Stanford Basic



- DELPH-IN Derevation Tree



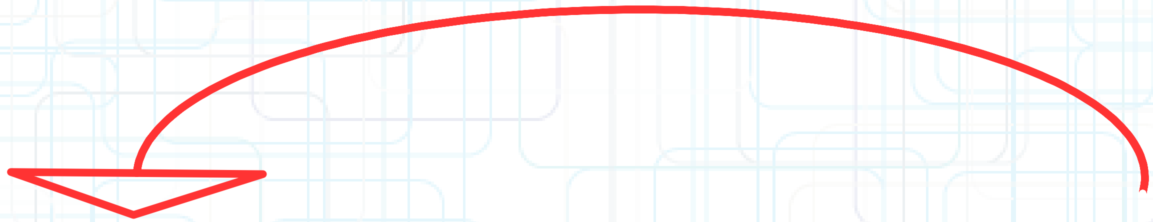
- Enju PAS

- DELPH-IN MRS

- CoNLL PropBank
- Stanford Standard

-

Tough adjective



A similar ***technique*** *is almost impossible to* ***apply***

A long-distant dependency is detected only in:

- CoNLL PropBank
- Enju PAS
- DELPH-IN MRS

Pairwise unlabeled dependency arcs overlap

	CD	CP	SB	SD	EP	DT	DM
CD	*19*	1	12	5	6	12	2
CP	1	*2*	1	0	1	1	1
SB	12	1	*19*	10	4	7	3
SD	5	0	10	*14*	2	4	3
EP	6	1	4	2	*20*	6	8
DT	12	1	7	4	6	*15*	0
DM	2	1	3	3	8	0	*11*

Pairwise Jaccard Similarity

	CD	CP	SB	SD	EP	DT	DM
CD		.171	.427	.248	.187	.488	.115
CP	.171		.171	.177	.122	.158	.173
SB	.427	.171		.541	.123	.319	.147
SD	.248	.177	.541		.14	.264	.144
EP	.187	.122	.123	.14		.192	.462
DT	.488	.158	.319	.264	.192		.13
DM	.115	.173	.147	.144	.462	.13	

Conclusions

- Large variability across formats
- **DELPH-IN Derivation Trees** are closer to **CoNLL Syntactic Dependencies**
- **DELPH-IN MRS** are closer to **Enju PAS**

Next steps

- Finalize the converter
- Transform Redwoods to word-to-word dependencies
- Use Redwoods data for parsing experiments

THANK YOU!