:University of Oslo

Who Did What to Whom?

[A Contrastive Study of Syntacto-Semantic Dependencies]
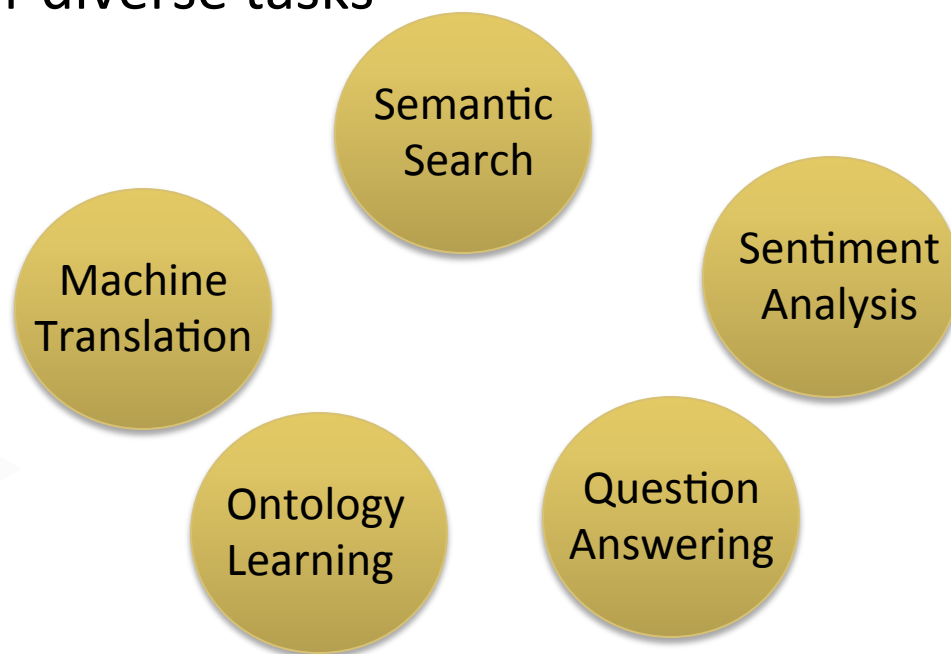
**Angelina Ivanova, Stephan Oepen, Lilja Øvrelid, and
Dan Flickinger (Stanford University)**

Introduction

Dependency representations:

➢ Useful for diverse tasks

Semantic Search

Sentiment Analysis

Machine Translation
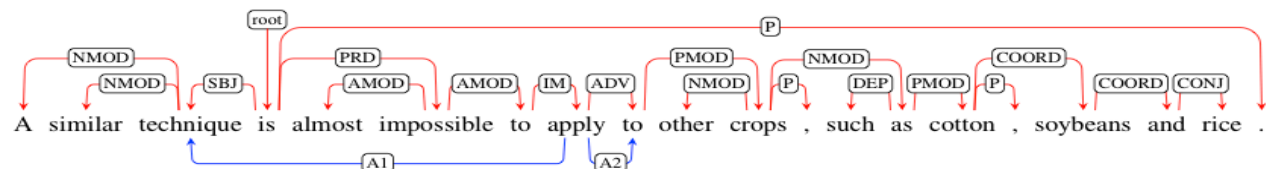
Ontology Learning

Question Answering
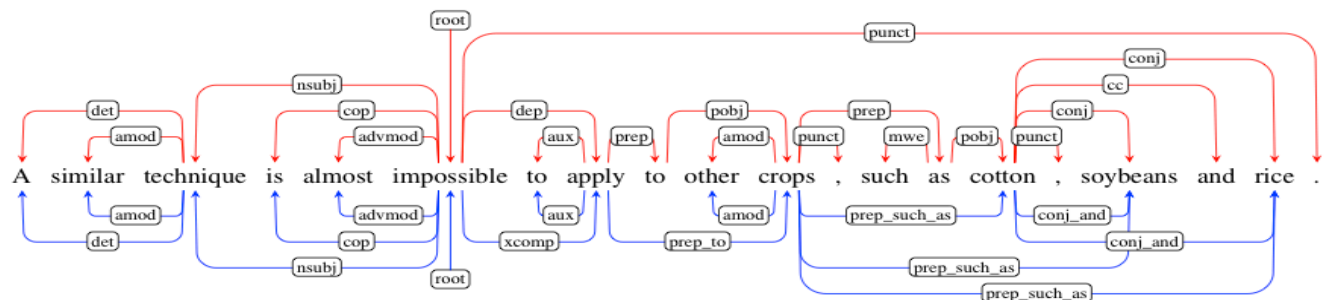
➢ Can be obtained automatically

Motivation

Variety of incompatible representation formats that challenges the task of parser evaluation.
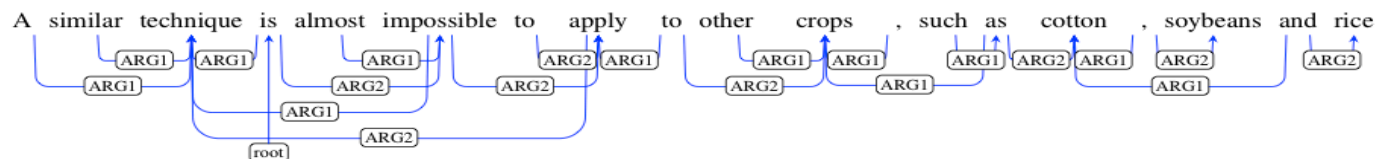
<u>Example:</u> A similar technique is almost impossible to apply to other crops, such as cotton, soybeans and rice.

(a) CoNLL 2008 *syntactic dependencies* (CD; top) and *propositional semantics* (CP; bottom).

(b) Stanford Dependencies, in the so-called *basic* (SB; top) and *collapsed & propagated* (SD; bottom) variants.

(c) Enju *predicate-argument structures* (EP).

Goals

Theoretical

Commonalities and differences between a broad range of dependency formats

Practical

Making LinGO Redwoods Treebank accessible for a   broader range of users

Dependency formats overview

- <u>PEST corpus</u>:

  Language: **English**

  Two sets: **10 sentences** and **15 sentences** from Wall Street Journal

  CoNLL Syntactic Dependencies (CD)

  CoNLL PropBank Semantics (CP)

  Stanford Basic Dependencies (SB)

  Stanford Collapsed Dependencies (SD)

  Enju Predicate-Argument Structures (EP)

- <u>Conversion from LinGO ERG:</u>

  DELPH-IN Syntactic Derivation Tree (DT)

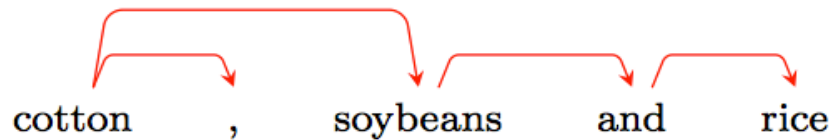 DELPH-IN MRS-derived Dependencies (DM)

Summary of dependency formats

| Abbr.	Format	Head status	Is the structure an acyclic tree?	Are the tokens connected?
CD	CoNLL Syntactic Dependencies	Functional	+	+
CP	CoNLL PropBank Semantics	Substantive	-	-
SB	Stanford Basic Dependencies	Substantive	+	+
SD	Stanford Collapsed Dependencies	Substantive	-	-
EP	Enju Predicate-Argument Structures	Substantive	-	+
DT	DELPH-IN Syntactic Derivation Tree	Functional	+	+
DM	DELPH-IN MRS-derived Dependencies	Substantive	-	-

Root choice

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

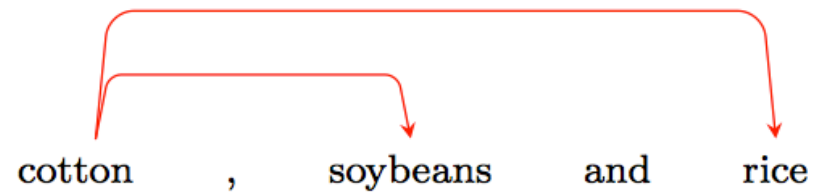|		
CD	CoNLL Syntactic Dependencies	**is**
CP	CoNLL PropBank Semantics	**-**
SB	Stanford Basic Dependencies	**impossible**
SD	Stanford Collapsed Dependencies	**impossible**
EP	Enju Predicate-Argument Structures	**is**
DT	DELHP-IN Syntactic Derivation Tree	**is**
DM	DELPH-IN MRS-derived Dependencies	**almost**

Conjunction

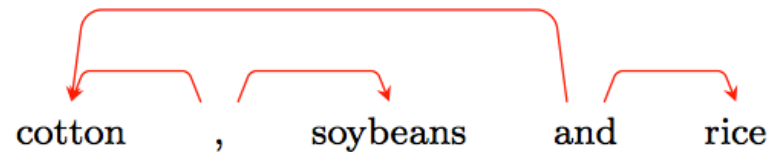**CoNLL Syntactic Dependencies (CD)**
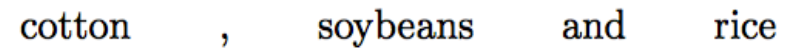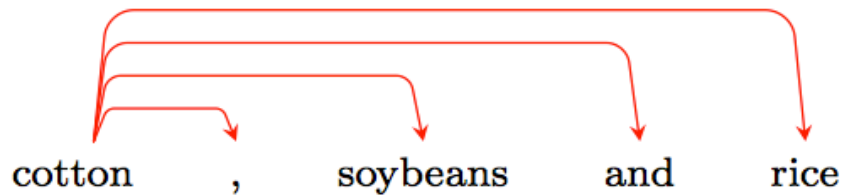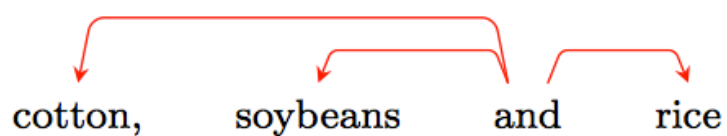
CoNLL PropBank Semantics (CP)
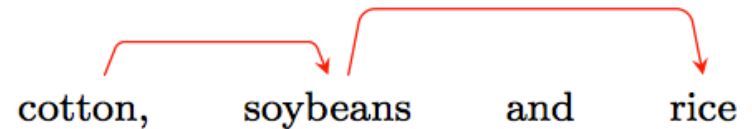
Stanford Basic Dependencies (SB)

Stanford Collapsed Dependencies (SD)

Enju Predicate-Argument Structures (EP)
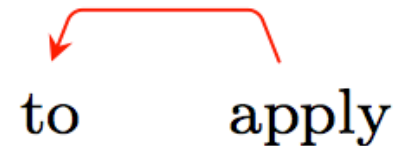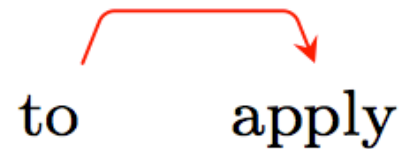
DELPH-IN Derivation Tree (DT)

DELPH-IN MRS-derived Dependencies (DM)

Infinitive

CoNLL Syntactic Dependencies (CD)
Enju Predicate-Argument Structures (EP)
DELPH-IN Syntactic Derivation Tree (DT)

to   apply

Stanford Basic Dependencies (SB)
Stanford Collapsed Dependencies (SD)

to   apply

CoNLL PropBank Semantics (CP)
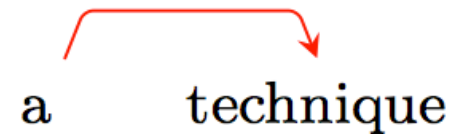DELPH-IN MRS-derived Dependencies (DM)

___

Article

CoNLL Syntactic Dependencies (CD)
Stanford Basic Dependencies (SB)
Stanford Collapsed Dependencies (SD)
DELPH-IN Syntactic Derivation Tree (DT)

Enju Predicate-Argument Structures (EP)
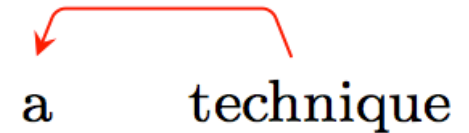DELPH-IN MRS-derived Dependencies (DM)

CoNLL PropBank Semantics (CP)

Adjective

CoNLL Syntactic Dependencies (CD)
Stanford Basic Dependencies (SB)
Stanford Collapsed Dependencies (SD)
DELPH-IN Syntactic Derivation Tree (DT)

similar ← technique

Enju Predicate-Argument Structures (EP)
DELPH-IN MRS-derived Dependencies (DM)

similar → technique

CoNLL PropBank Semantics (CP)

—

Preposition

CoNLL Syntactic Dependencies (CD)
Stanford Basic Dependencies (SB)

DELPH-IN Syntactic Derivation Tree (DT)

Enju Predicate-Argument Structures (EP)

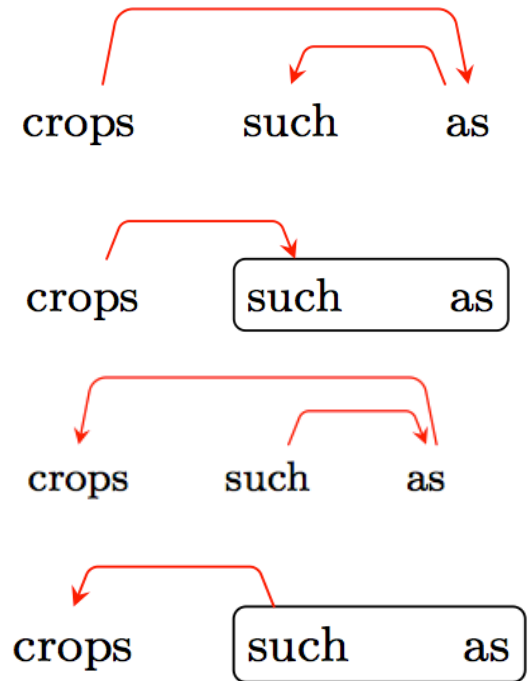DELPH-IN MRS-derived Dependencies (DM)

CoNLL PropBank Semantics (CP)
Stanford Collapsed Dependencies (SD)

Tough adjective

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

The long-distance dependency is detected only in:

CoNLL PropBank Semantics (CP)
Enju Predicate-Argument Structures (EP)
DELPH-IN MRS-derived Dependencies (DM)

Pairwise Jaccard similarity on PEST

	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	

DELPH-IN Syntactic Derivation Tree (DT) format is closer to **CoNLL Syntactic Dependencies (CD)**

DELPH-IN MRS-derived Dependencies (DM) are closer to **Enju Predicate-Argument Structures (EP)**

Goals

Theoretical

Commonalities and differences between a broad range of dependency formats

Practical

Making LinGO Redwoods Treebank accessible for a   broader range of users

The LinGO Redwoods Treebank

- Language: **English**
- Size: **45 000 utterances**
- Linguistic approach: **HPSG**
- Grammar: **LinGO ERG**
- Data:
  - Verbmobil and e-commerce corpora
  - LOGON Norwegian-English MT corpus
  - English Wikipedia (from WeScience)
  - Brown corpus (SemCor)
  - other

DELPH-IN Syntactic Derivation Tree

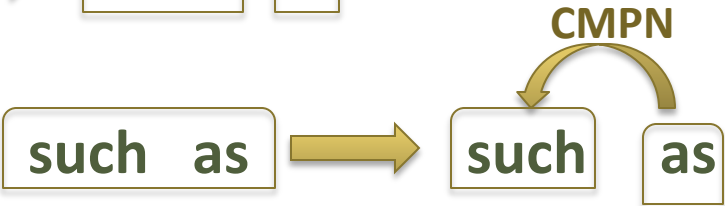**DELPH-IN Syntactic Derivation Tree representation**
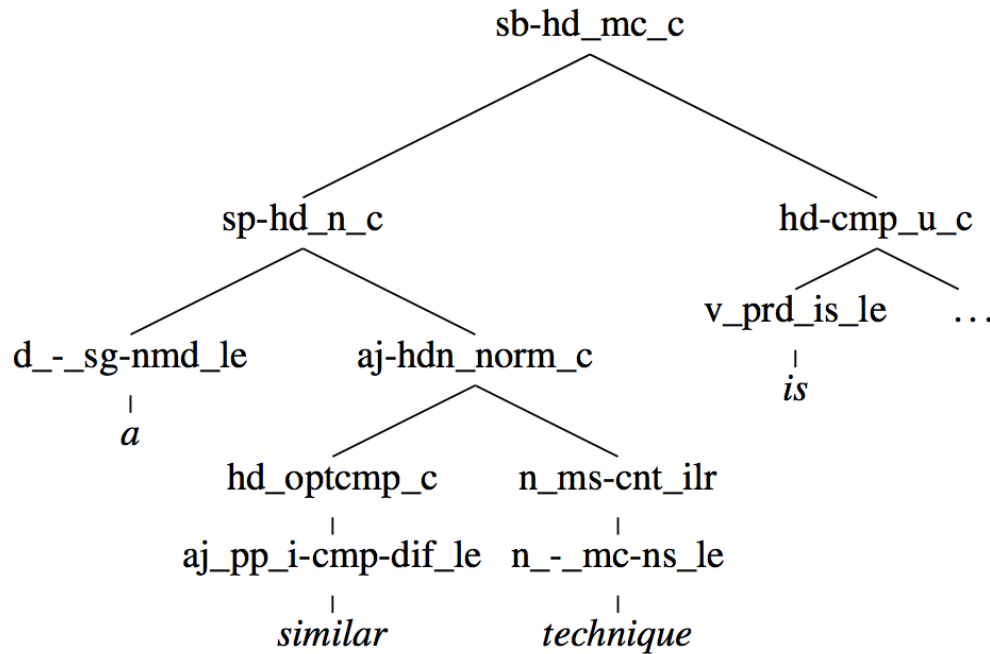
**Conversion to bilexical dependencies**
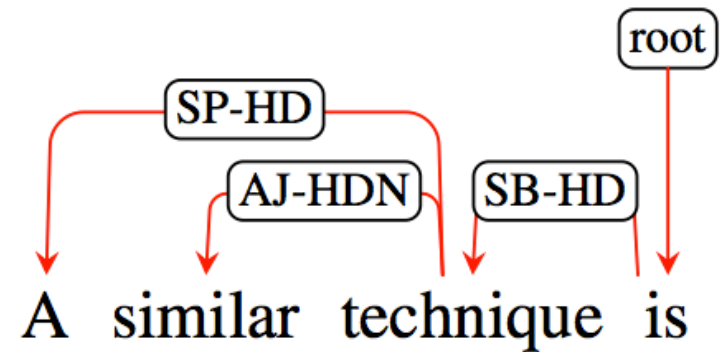
Special cases during the conversion

Special cases:

- Contracted negation: **doesn't** → **does** **n't** (NEG)

- Punctuation: **bark.** → **bark** **.** (PUNCT)

- Multiword expressions: **such as** → **such** **as** (CMPN)

- Hyphenated words: **end-** **state** → **end-state**

Elementary Dependency Structure

{ e12

_1:_a_q(BV x6)

e9:_similar_a_to(ARG1 x6)

x6:_technique_n_1

e12:_almost_a_1(ARG1 e3)

e3:_impossible_a_for(ARG1 e18)

e18:_apply_v_to(ARG2 x6, ARG3 x19)

_2:udef_q(BV x19)

e25:_other_a_1(ARG1 x19)

x19:_crop_n_1

…}

A similar technique is almost impossible to apply to other crops
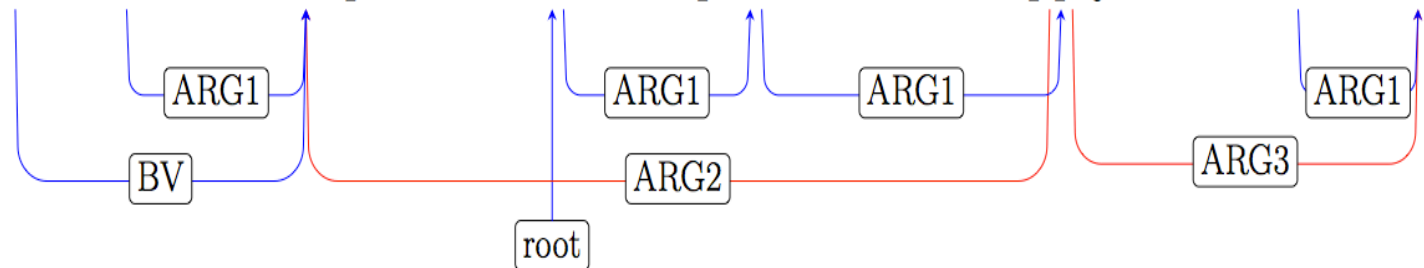
Elementary Dependency Structure

{...

x33:_cotton_n_1

_5:udef_q(BV i38)

x27:implicit_conj(L-INDEX x33, R-INDEX i38)

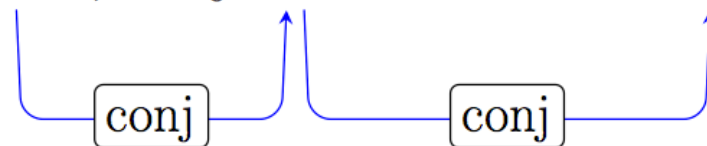_6:udef_q(BV x43)

x43:_soybeans/nns_u_unknown

i38:_and_c(L-INDEX x43, R-INDEX x47)

_7:udef_q(BV x47)

x47:_rice_n_1

}

cotton, soybeans and rice.

Conclusions

- Qualitative and quantitative comparison of various dependency formats

- Automatic mapping from HPSG representations to syntactic and semantic dependencies

Future work

- Release of the converted Redwoods treebanks and conversion software

- Modification of DELPH-IN MRS-derived Dependencies conversion into a dependency tree

- Training parsers on DELPH-IN Syntactic Derivation Tree and DELPH-IN MRS-derived Dependencies formats

- Experimentation in domain adaptation for parsing on Redwoods treebanks

Thank you for your attention!

Questions?