Enhanced Universal Dependencies

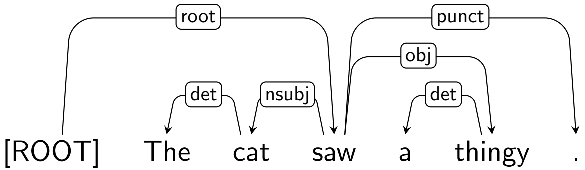
Introduction and WIP

Adam Ek

Centre for Linguistic Theory and Studies in Probability,
Department of Philosophy, Linguistics and Theory of Science,
University of Gothenburg,
adam.ek@gu.se

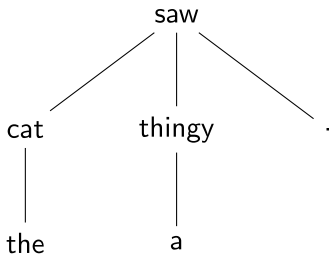
Dependency Grammar

- Syntactic structures realized as dependency relations
- A dependency relation is a directed arc between two lexical items, the *head* and the *dependant*
- Each dependency arc is labeled with a grammatical function



Dependency Structures

- Constraints of dependency structures:
  - Dependency structures are complete
  - Dependency structures are hierarchical
  - Every lexical item has at most one head
- Dependency structures are trees



The cat saw a thingy.

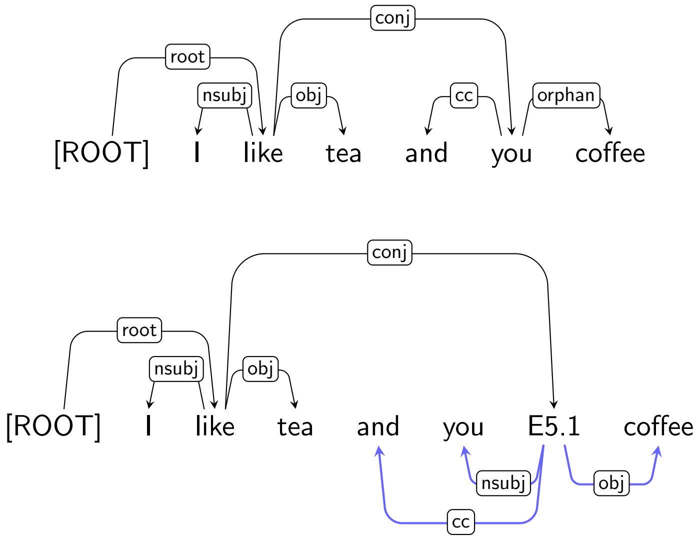
Some issues with dependency grammar in downstream tasks

- Dependency structures are great, but...
- In practice they can be impractical to use:
- Some arguments are not directly connected to predicates
- Some dependency relations are uninformative (`nmod/obl`)
- In general: Can UD be extended to be more practical for semantic tasks?

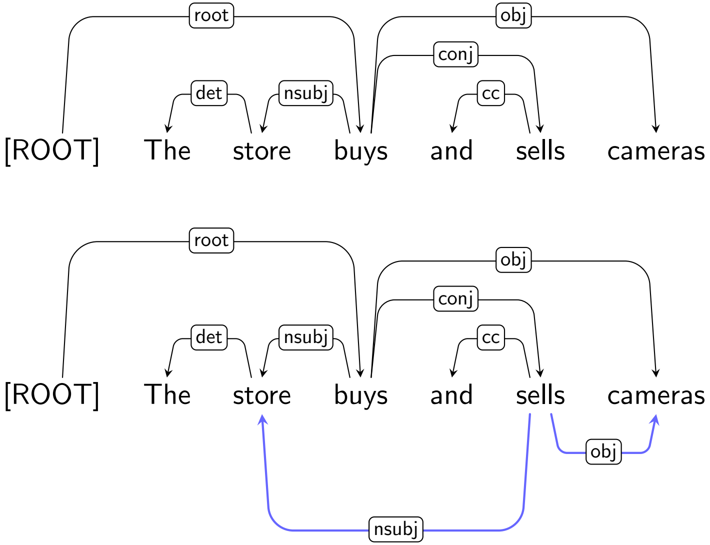
Enhanced Universal Dependencies

- Enhanced Universal Dependencies (EUD) is an extension of the universal dependencies annotation schema
- EUD is aimed at providing easily accessible semantic interpretations of dependency structures:
  - Includes annotations of ellipsis
  - Propagation of conjuncts
  - Controlled/raised subjects
  - Coreference in relative clauses
  - Case-marking information

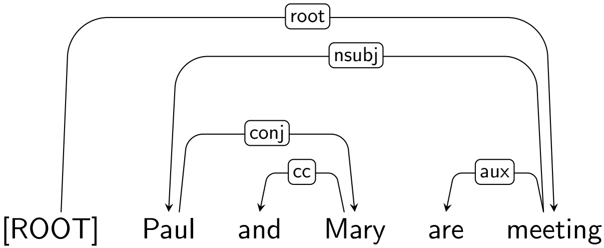
Ellipsis

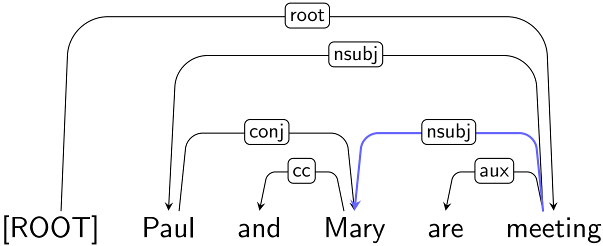

conj
root
nsubj
obj
cc
orphan
[ROOT]
I
like
tea
and
you
coffee
conj
root
nsubj
obj
[ROOT]
I
like
tea
and
you
E5.1
coffee
nsubj
obj
cc

Propagation of conjuncts: verbs and phrases

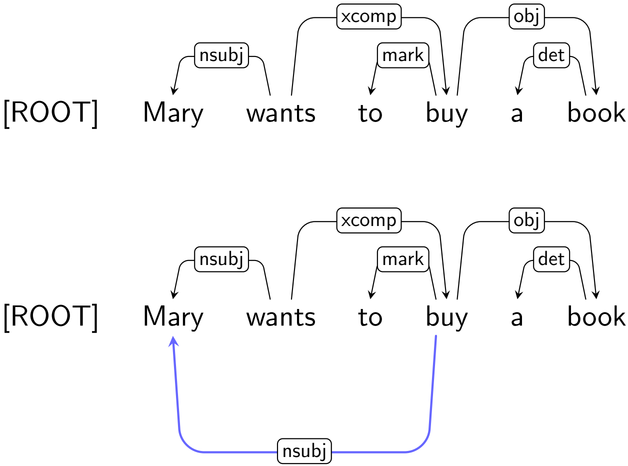


Propagation of conjuncts: subjects

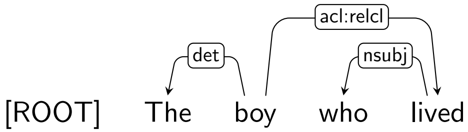


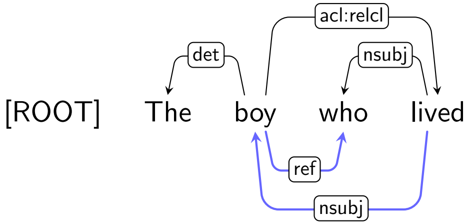


Controlled/raised subjects

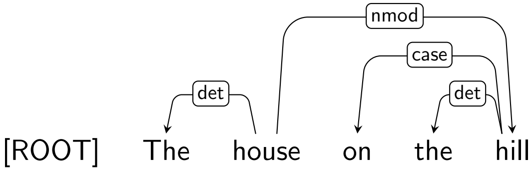


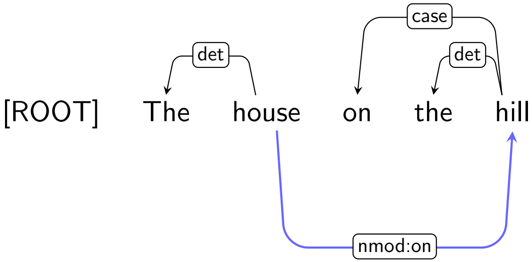
Relative clauses





Case information





Parsing EUD (1)

What types of changes have been introduced?

- Additional dummy nodes
- New arcs between lexical items
- Many new labels (in english, the number of arc labels is 389)

Parsing EUD (2)

- As a consequence of the additional arcs, we now deal with graphs
- Main problem for systems: a lexical item can now have two heads instead of one
- Transition models for UD assume one head and have trouble dealing with the new formalism
- Graph parsers should still work well with some modifications
- We develop a graph-based parsing model based on Pointer networks (Vinyals et al. 2017, Ma et al. 2018)

But why...

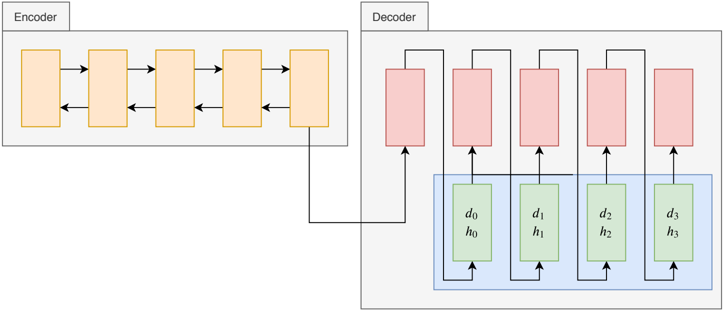
- We want to use syntactic information to derive semantic information
- More specifically: predicate-argument structures

Tasks:

- QA
- Inference

Model Outline

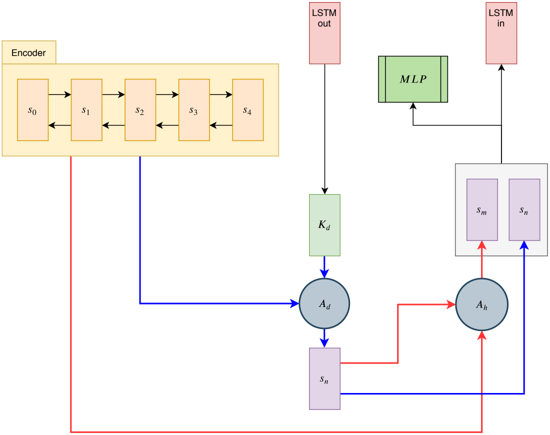
- "Standard" sequence2sequence architecture
- With a twist...



Encoder

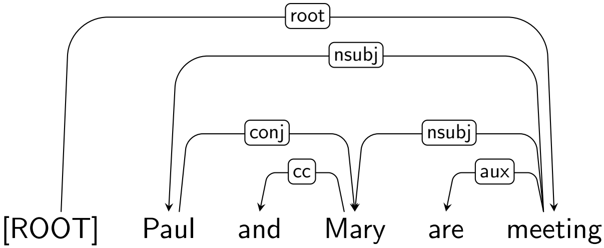
- We initialize word and part-of-speech embeddings randomly
- The input to the LSTM is the word and part-of-speech embeddings concatenated
- The LSTM is bidirectional and use 3 layers

Decoder



Dependent attention

- A_d selects the dependent and operate as a "step-function", traversing the sentence from right to left
- A_d is a "pointer", it selects an index from the source sequence and returns that hidden state
- 0, 1, 2, 3, 3, 4, 5



Dependent attention module

Dependent attention (local monotonic-ish additive attention)

- $v_a = lmat(K_d, V)$
- $e = V_{p:p+3} W_0 + K_d W_1 + b$
- $a = softmax(eW_2 \in \mathbb{R}^1)$
- $att_d = V_{argsmax(a)}$

Head attention (multiplicative attention)

- $v_a = ma(att_d, V)$
- $e = att_d W_0 V_{0:n}$
- $a = softmax(eW_1 \in \mathbb{R}^1)$
- $att_h = V_{argsmax(a)}$

LSTM input $= [att_d; att_h]$

Objectives

We are optimizing 3 objectives:

- Selecting a dependent
- Selecting a head for the dependent
- Assigning a label to the dependent and the head
- $\mathcal{L} = \mathcal{L}_{dep} + \mathcal{L}_{head} + \mathcal{L}_{label}$

Data

- IWPT shared task
- Data from 16 languages (20 treebanks from UD)

Treebank	Language	Train	Dev	Treebank	Language	Train	Dev
ar_padt	arabic	6076	909	nl_alpino	dutch	12265	718
bg_btb	bulgarian	8909	1115	nl_lassysmall	dutch	5788	676
cs_cac	czech	23479	603	pl_lfg	polish	13775	1745
cs_fictree	czech	10161	1309	pl_pdb	polish	17723	2215
cs_pdt	czech	68496	9270	ru_syntagrus	russian	48815	6584
en_ewt	english	12544	2002	sk_snk	slovak	8484	1060
fi_tdt	finnish	12218	1364	sv_talbanken	swedish	4304	504
fr_sequoia	french	2232	412	ta_ttb	tamil	401	81
it_isdt	italian	13122	564	uk_iu	ukranian	5497	672
lt_alksnis	lithuanian	2342	618	lv_lvtb	latvian	10157	1664

Table: Sentences in the training and development set from 20 treebanks.

Training setup

Word embedding	224
POS embedding	32
Batch size	16
Encoder LSTM features	256
Decoder features	1024
Loss	Cross Entropy
Optimizer	Adam
Learning rate	0.002
Weight decay	0.1

Dropout: 25% dropout before each prediction module (dependent, head, and label)

Preliminary results (latest)

Treebank	GA	LGA	b-GA	b-LGA
ar_padt	**.650**	**.588**	.225	.114
bg_btb	**.813**	**.764**	.675	.578
cs_cac	.668	.601	**.788**	**.716**
cs_fictree	.781	**.717**	**.799**	.674
cs_pdt	.759	**.695**	**.821**	.662
en_ewt	.798	**.751**	**.830**	.721
fi_tdt	.640	.562	**.796**	**.643**
fr_sequoia	.625	.575	**.780**	**.692**
it_isdt	**.769**	**.722**	.456	.372
lt_alksnis	.503	.420	-	-
lv_lvtb	.657	.599	**.799**	**.651**

Table: Results on 20 treebanks on the development set. GA is the graph accuracy. LGA is the labeled graph accuracy. b-* is the baseline (Stanford Dependency Parser) score for the treebank.

Updated (super latest)

Treebank	GA	LGA	b-GA	b-LGA	Treebank	GA	LGA	b-GA	b-LGA
ar_padt	**.650**	**.588**	.225	.114	nl_alpino	**.798**	**.738**	.797	.657
bg_btb	**.813**	**.764**	.675	.578	nl_lassysmall	**.752**	**.698**	.740	.600
cs_cac	.668	.601	**.788**	**.716**	pl_lfg	.902	**.861**	**.936**	.853
cs_fictree	.781	**.717**	**.799**	.674	pl_pdb	**.740**	**.661**	.649	.485
cs_pdt	.759	**.695**	**.821**	.662	ru_syntagrus	.787	**.731**	**.819**	.666
en_ewt	.798	**.751**	**.830**	.721	sk_snk	**.755**	**.668**	.840	.670
fi_tdt	.640	.562	**.796**	**.643**	sv_talbanken	.631	.583	**.817**	**.685**
fr_sequoia	.625	.575	**.780**	**.692**	ta_ttb	.551	.361	-	-
it_isdt	**.769**	**.722**	.456	.372	uk_iu	.648	.561	.817	.653
lt_alksnis	.503	.419	-	-	lv_lvtb	.657	.599	.799	.651

Table: Results on 20 treebanks on the development set. GA is the graph accuracy. LGA is the labeled graph accuracy. b-* is the baseline (Stanford Dependency Parser) score for the treebank.