The Case for Universal Dependencies

Christopher Manning

@stanfordnlp

Stanford University

Depling 2015, Uppsala

The Case for Universal Dependencies

You've been hearing a lot about why UD is bad … here's why it's good … for at least some things

1. NLP motivation (brief!)

2. What is UD?

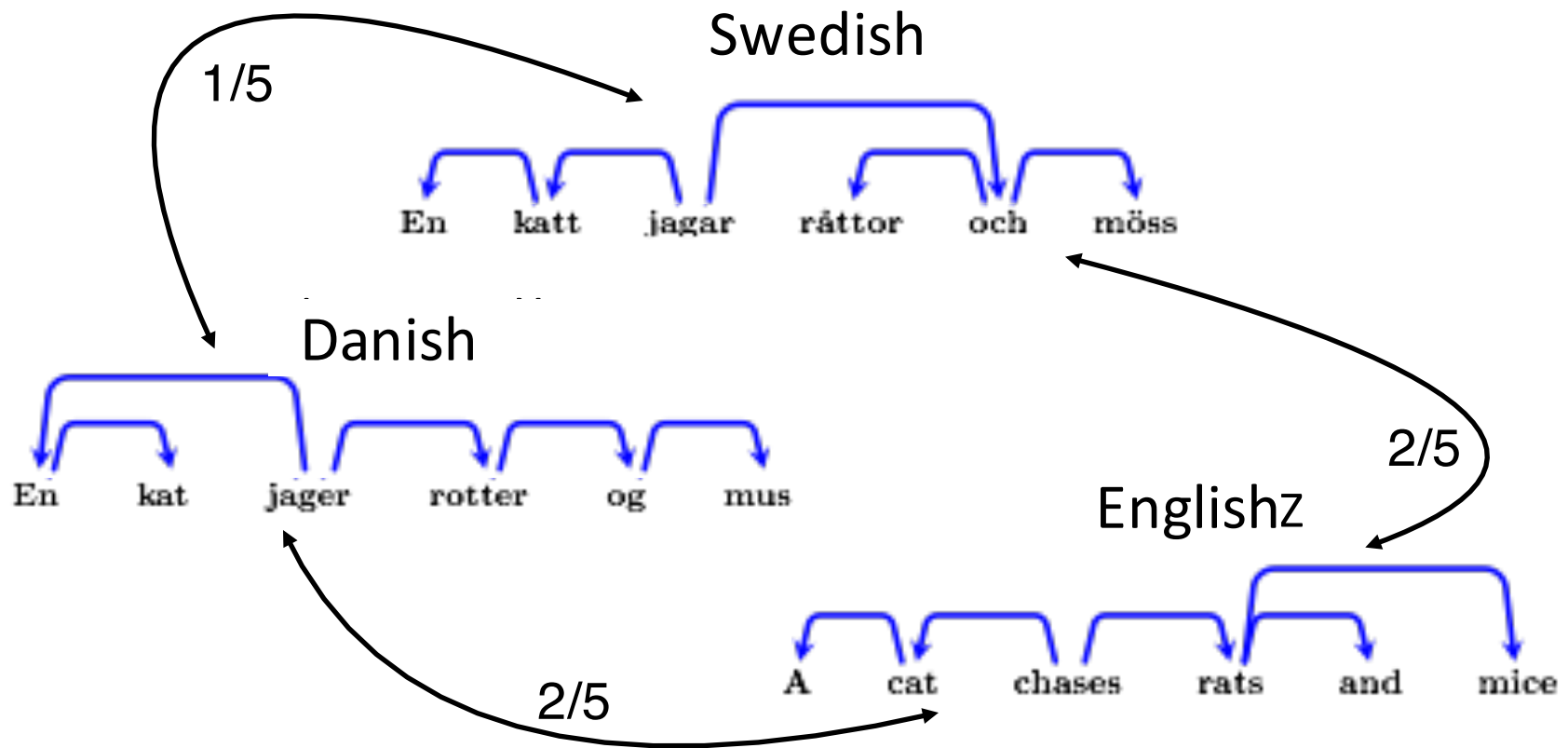3. Why is UD as it is?

4. Some more personal reflections on grammar

Thanks to many collaborators!

Core UD group: Timothy Dozat, Filip Ginter, Yoav Goldberg, Jan Hajič, **Joakim Nivre**, Ryan McDonald, Natalia Silveira, Marie-Catherine de Marneffe, Slav Petrov, Sampo Pyysalo, Reut Tsarfaty, Dan Zeman

UD contributors: Željko Agić, Maria Jesus Aranzabe, Masayuki Asahara, Aitziber Atutxa, Cristina Bosco, Jinho Choi, Kaja Dobrovoljc, Binyam Ephrem, Tomaž Erjavec, Richárd Farkas, Jennifer Foster, Koldo Gojenola, Iakes Goenaga, Anders Trærup Johannsen, Hiroshi Kanayama, Jenna Kanerva, Simon Krek, Veronika Laippala, Alessandro Lenci, Nikola Ljubešić, Teresa Lynn, Aibek Makazhanov, Catalina Maranduc, Héctor Martínez Alonso, Anna Missilä, Verginica Mititelu, Yusuke Miyao, Simonetta Montemagni, Shinsuke Mori, Petya Osenova, Barbara Plank, Prokopis Prokopidis, Wolfgang Seeker, Mojgan Seraji, Maria Simi, Kiril Simov, Aaron Smith, Takaaki Tanaka, Sumire Uematsu, Veronika Vincze, Rob Voigt

1. NLP motivations (multilingual)

Swedish

1/5

En   katt   jagar   råttor   och   möss

Danish

En   kat   jager   rotter   og   mus

2/5

Englishz

2/5

A   cat   chases   rats   and   mice

Which languages are most closely related?

NLP motivations (multilingual) McDonald, Petrov, and Hall (2011)

Target \ Source	da	el	en	it	pt	sv
da (Danish)	**79.2**	44.0	45.9	48.6	48.1	47.8
el (Greek)	33.3	**77.5**	63.9	59.3	58.6	47.5
en (English)	34.4	45.7	**82.5**	38.6	42.3	43.7
it (Italian)	44.8	66.8	57.7	**79.3**	69.1	50.9
pt (Portuguese)	42.5	66.6	69.2	74.7	**84.6**	52.1
sv (Swedish)	44.5	57.8	58.3	53.4	66.8	**84.8**

# Why is annotation variety a problem?

- Hard to compare empirical results across languages

- Hard to usefully do cross-lingual structure transfer

- Hard to evaluate cross-lingual learning

- Hard to build and maintain multilingual systems

- Hard to make progress towards a universal parser

- Hard to do valid linguistic typology

We need to bring some order into the chaos

NLP motivations (monolingual)

A single level of typed dependencies can give a simple, human-friendly representation of sentence structure and meaning
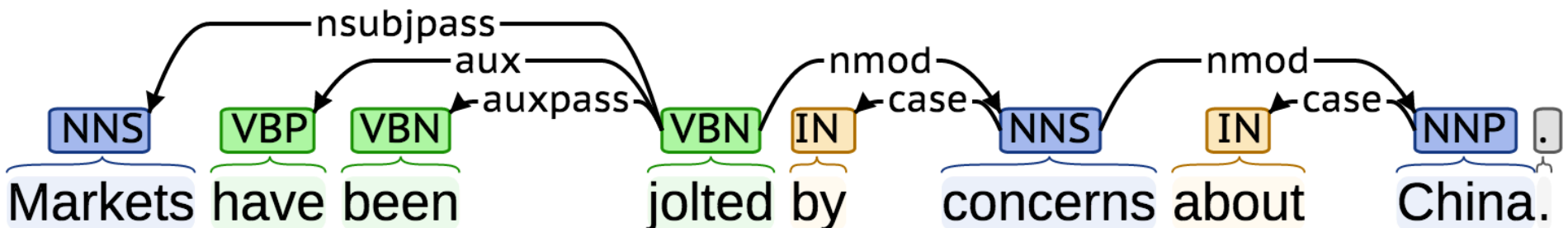
Better than a constituency tree for machine interpretation – it's almost a semantic network

UD aims to be **linguistically better** than earlier, common, simple NLP representations, such as CoNLL dependencies
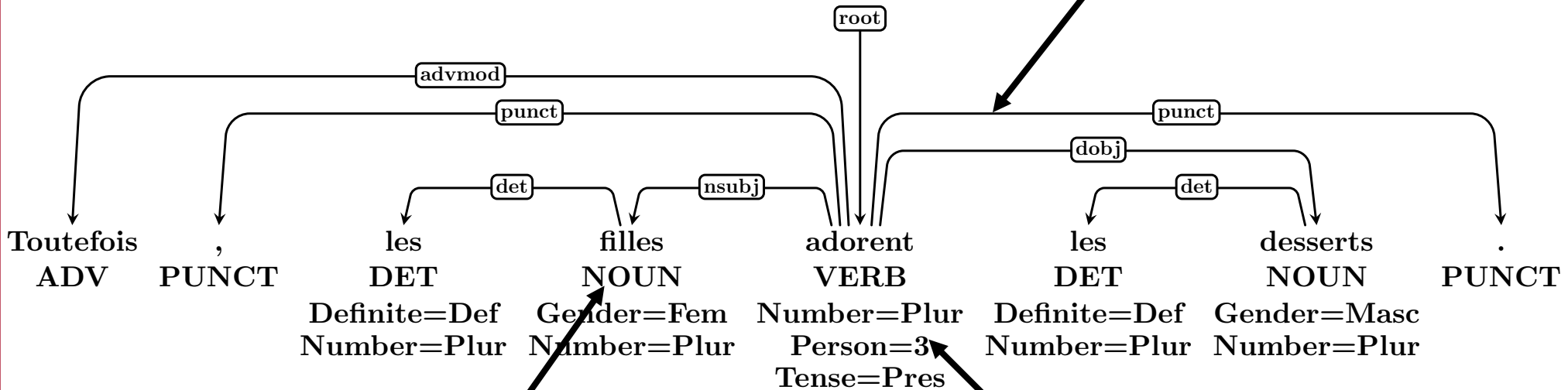
2. Universal Dependencies

http://universaldependencies.github.io/docs/

Dependency relations

root

advmod

punct

det

nsubj

punct

dobj

det

Toutefois	,	les	filles	adorent	les	desserts	.
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

Morphological features

Universal Dependencies

http://universaldependencies.github.io/docs/

Stanford
Depe...

...gle universal

Universal Dependencies
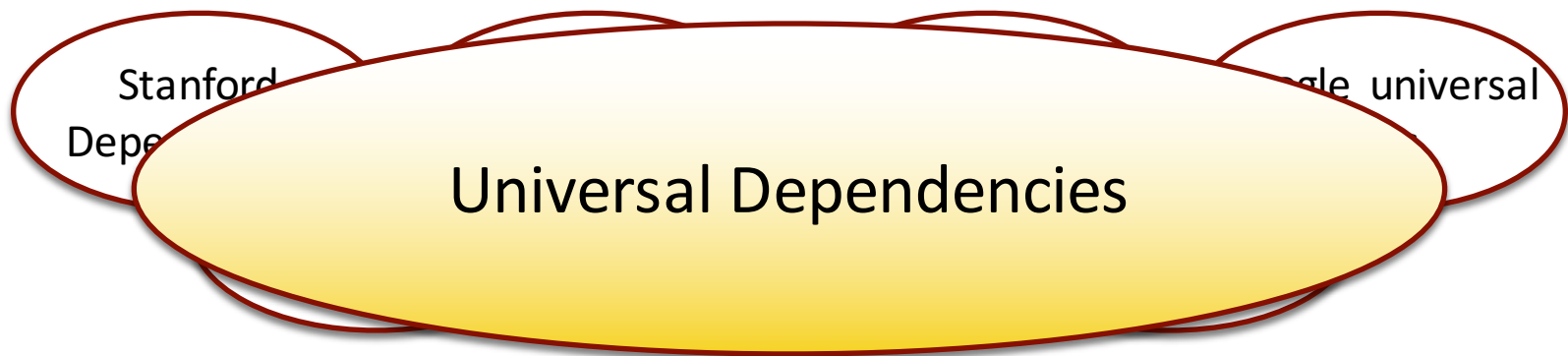
Milestones:

- Kick-off meeting at EACL in Gothenburg, April 2014

- Release of annotation guidelines, Version 1, October 2014

- Release of treebanks for 10 languages, January 2015

- Release of treebanks for 18 languages, May 2015

Morphology

Toutefois	,	les	filles	adorent	les	desserts	.
toutefois	,	le	fille	adorer	le	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	

- Lemma representing the semantic content of the word
- Part-of-speech tag representing the abstract lexical category associated with the word
- Features representing lexical and grammatical properties associated with the lemma or the particular word form

Part-of-Speech Tags

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

- Taxonomy of 17 universal part-of-speech tags, expanding on the Google Universal Tagset (Petrov et al., 2012)
- All languages use the same inventory, but not all tags have to be used by all languages
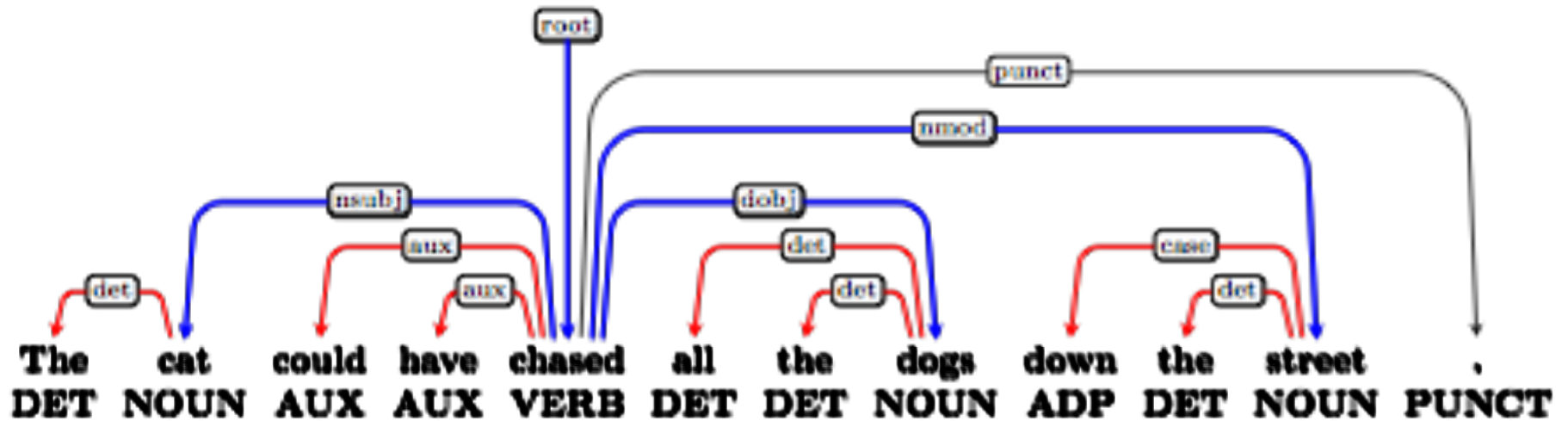
Features

Lexical	Inflectional Nominal	Inflectional Verbal
PronType	Gender	VerbForm
NumType	Animacy	Mood
Poss	Number	Tense
Reflex	Case	Aspect
	Definite	Voice
	Degree	Person
		Negative

- Standardized inventory of morphological features, based on the Interset system (Zeman, 2008)

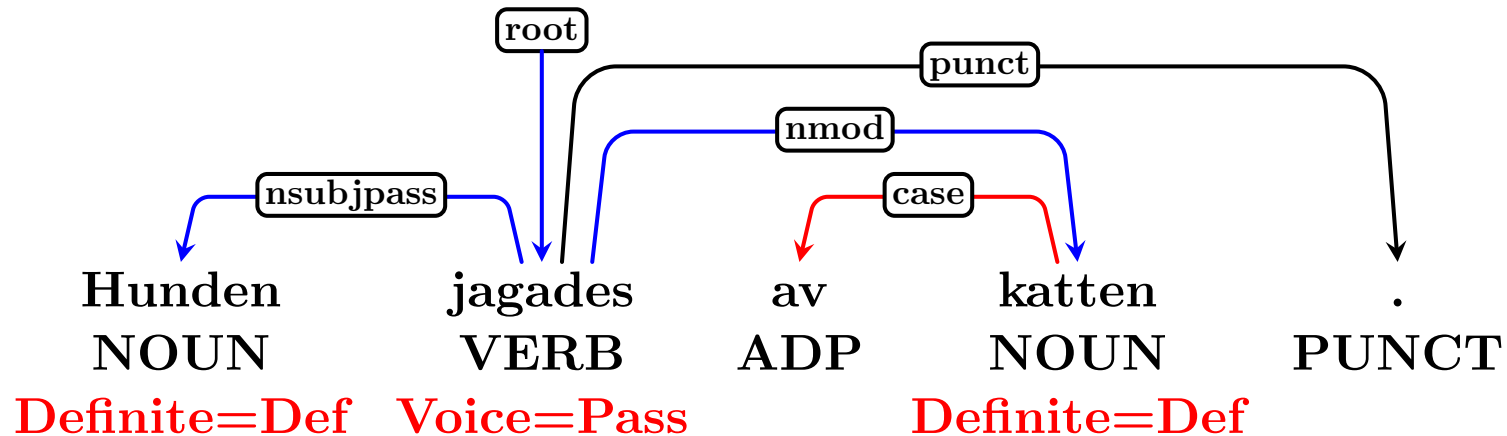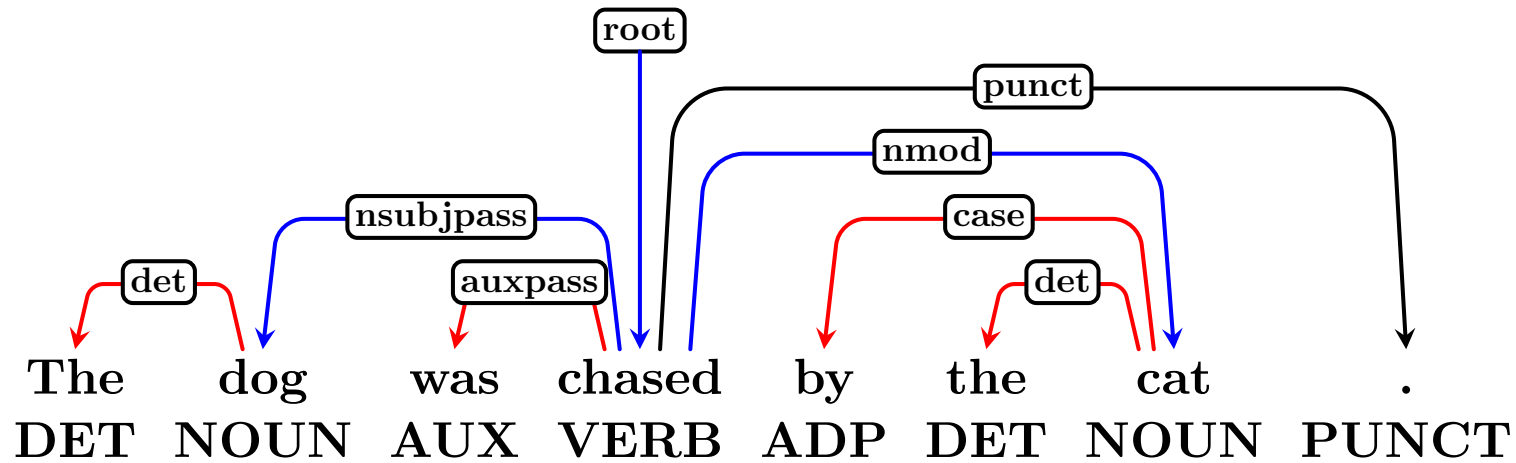- Languages select relevant features and can add language-specific features or values with documentation

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 related by dependency relations
- Function words attach to the content word they further specify
- Punctuation attaches to head of phrase or clause

Substantive parallelism is achieved

root
punct
nmod
nsubjpass
auxpass
case
det
det

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

root
punct
nmod
nsubjpass
case

Hunden jagades av katten .
NOUN VERB ADP NOUN PUNCT
Definite=Def Voice=Pass Definite=Def

Dependency Relations

- Taxonomy of 40 universal grammatical relations, broadly attested in language typology (de Marneffe et al., 2014)
  - Language-specific subtypes may be added

- Organizing principles
  - Three types of structures: nominals, clauses, modifiers
  - Core arguments vs. other dependents (not complements vs. adjuncts)
  - Differentiate compounding vs. phrasal modification

Dependents of Clausal Predicates

	Nominal	Clausal	Other
Core	nsubj nsubjpass dobj iobj	csubj csubjpass ccomp xcomp	
Non-Core	nmod vocative discourse expl	advcl	advmod neg aux auxpass cop mark punct

Dependents of Nominals

Nominal	Clausal	Other
nummod appos nmod	acl	amod det case

Cairo , the lovely capital of Egypt

PROPN PUNCT DET ADJ NOUN ADP PROPN

appos — mark — det — amod — nmod — case
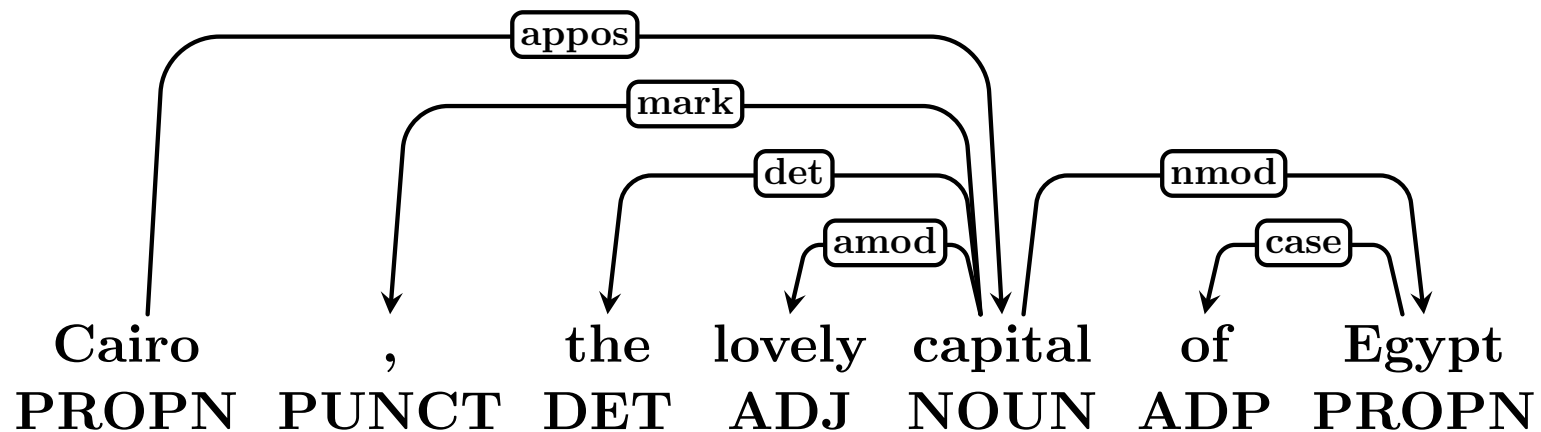
Coordination

Coordination
conj
cc
(punct)

Huey , Dewey and Louie
PROPN PUNCT PROPN CONJ PROPN
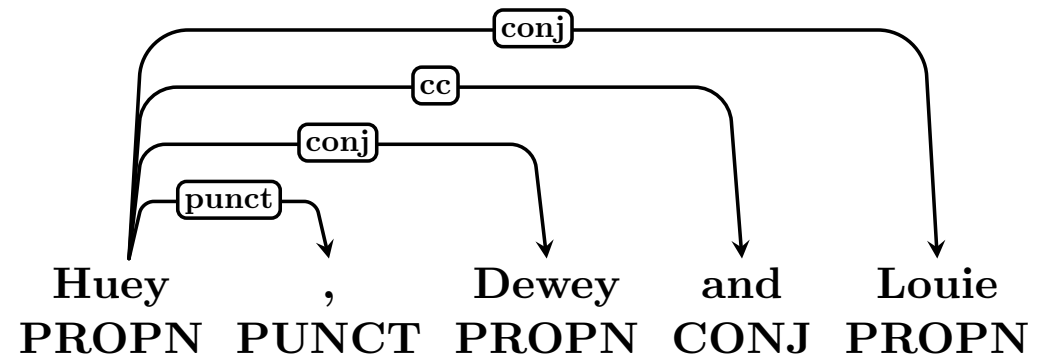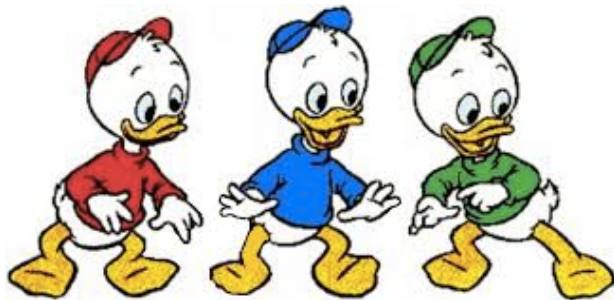
- Coordinate structures are headed by the first conjunct
  - Subsequent conjuncts depend on it via the conj relation
  - Conjunctions depend on it via the cc relation
  - Punctuation marks depend on it via the punct relation

Multiword Expressions

Relation	Examples
mwe	*in spite of, as well as, ad hoc*
name	*Roger Bacon, Carl XVI Gustaf, New York*
compound	*phone book, four thousand, dress up*
goeswith	*notwith standing, with out*

- UD annotation does/did not permit "words with spaces"
  - Multiword expressions are analysed using special relations
  - The mwe, name and goeswith relations are always head-initial
  - The compound relation reflects the internal structure

Other Relations

Relation	Explanation
dislocated	Preposed or postposed peripheral elements of clause
parataxis	Loosely linked clauses of same rank
list	Lists without syntactic structure
remnant	Orphans in ellipsis linked to parallel elements
reparandum	Disfluency linked to (speech) repair
foreign	Elements within opaque stretches of code switching
dep	Unspecified dependency
root	Syntactically root element of sentence (or fragment)

Language-Specific Relations

- Language-specific relations are subtypes of universal relations added to capture important phenomena

- Subtyping permits us to "back off" to universal relations

Relation	Explanation
acl:relcl	Relative clause
compound:prt	Verb particle (*dress up*)
nmod:poss	Genitive nominal (*Mary's book*)
nmod:agent	Agent in passive (*saved by the bell*)
cc:preconj	Preconjunction (*both ... and*)
det:predet	Predeterminer (*all those ...*)

Where are we now?

- Universal Dependencies, Version 1
  - Guidelines released October 2014
  - Latest treebank release May 2015 (v1.1):

    Basque, Bulgarian, Croatian, Czech, Danish, English, Finnish, French, German, Greek, Hebrew, Hungarian, Indonesian, Irish, Italian, Persian, Spanish, Swedish

- Future plans:
  - New releases every six months (May, November)
    - Next release: November 2015
  - Revision of guidelines as needed

- Have a look at http://universaldependencies.github.io/docs/
  - Open community effort – anyone can contribute!

The goal is a cross-linguistically workable universal grammar suitable for the common person

Universal Dependencies: User-centered design
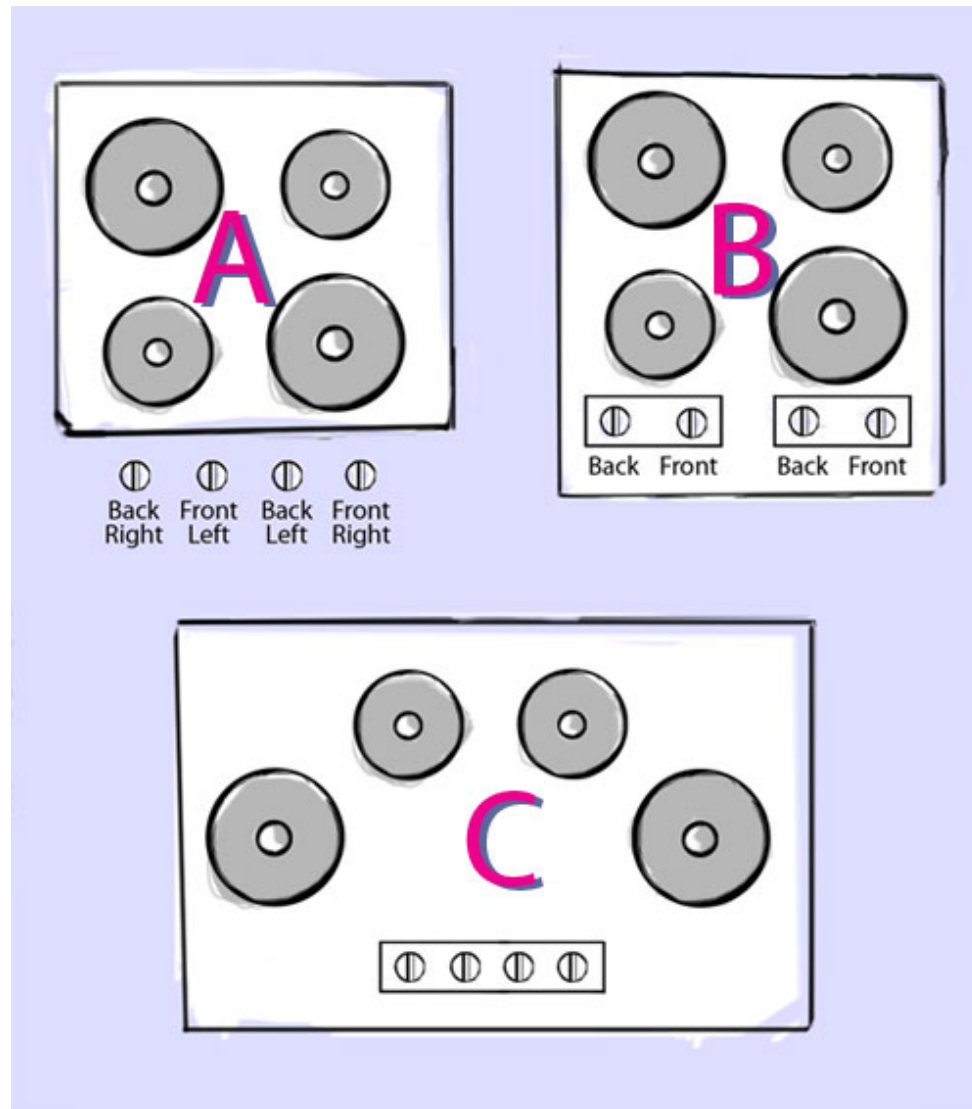
The design problem

- Building UD is a **design problem**!

- UD is a quite subtle trade-off between several competing goals:
  - an analysis which is reasonably satisfactory on linguistic grounds – a journeyman's **universal grammar**
  - an analysis that is reasonably comprehensible to non-linguist users – a **habitable** design
  - an analysis which can be automatically applied with good accuracy (i.e., **computer parsing**)
  - an analysis which enables **language understanding** tasks, such as relation extraction, machine translation, …

Design principles
Updated from [de Marneffe & Manning 2008]

1. Only have binary relations between words

2. Relations should be semantically contentful and draw on traditional grammar for easier comprehension by users

3. Relations should be useful for applications and automatically assignable

4. Maximize parallelism, but don't overdo it

5. Relations between content words, not function words

6. Adopt a lexicalist analysis

7. Less can be better

1. Binary relations between words

det(boy-3, The-1)

amod(boy-3, little-2)

nsubj(jumped-4, boy-3)

root(ROOT-0, jumped-4)

case(fence-7, over-5)

det(fence-7, the-6)

nmod(jumped-4, fence-7)
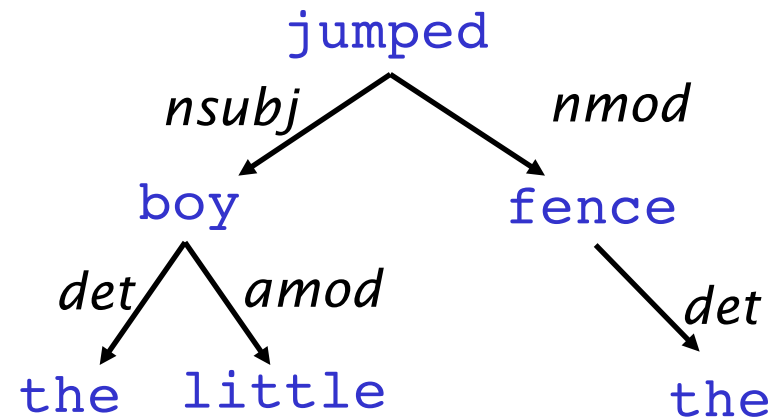
It's a natural format for users that maps easily on to:

- graph representations (nodes and labeled arcs)
- semantic notations like RDF triples, etc.

2. Semantically contentful relations using notions of traditional grammar

"I feel like a kid," says a gleeful Alex de Castro, a car salesman, who has stopped by a workout of the Suns to slip six Campaneris cards to the Great Man Himself

SD	**Typical**
num(cards, six)	*adjunct*(cards, six)
appos(Castro, salesman)	*adjunct*(Castro, salesman)
amod(Castro, gleeful)	*adjunct*(Castro, gleeful)
advcl(stop, slip)	*adjunct*(stop, slip)
nmod:of(workout, Suns)	*adjunct*(workout, Suns)

3. Relations should be useful for applications and automatically assignable

Advantage of grammatical relations over semantic roles:

1. Grammatical relations can be correctly assigned by our current NLP systems with far greater accuracy than semantic roles

2. Grammatical relations coupled with a few features (such as voice, case) are sufficient for most language understanding applications

3. Grammatical relations (subject, object) are familiar to many more users than semantic roles (experiencer, theme)

4. Maximize parallelism but don't overdo it

Maximize parallelism

- Don't annotate the same thing in different ways
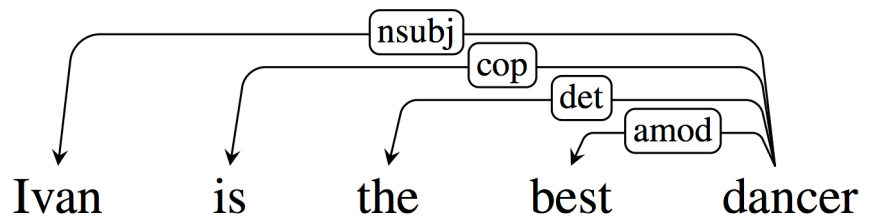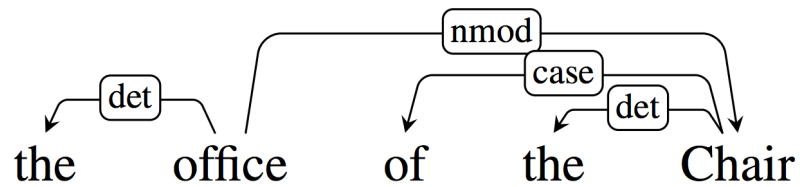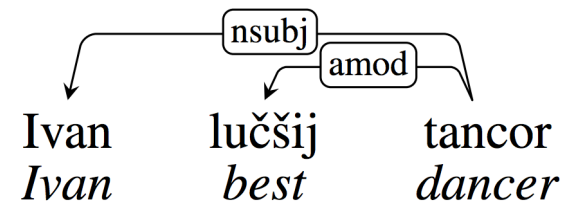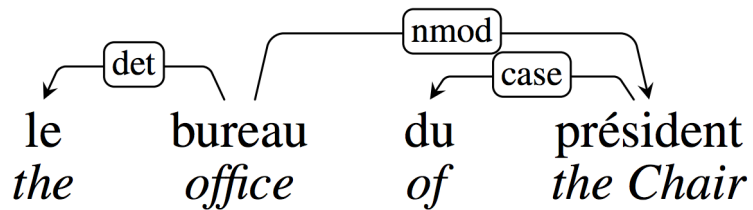- Don't make different things look the same

But don't overdo it

- Don't annotate things that are not there

So differences between languages should be the genuine examples of "metataxis" – how ideas are expressed differently across languages

Making content words heads maximizes cross-linguistic parallelism
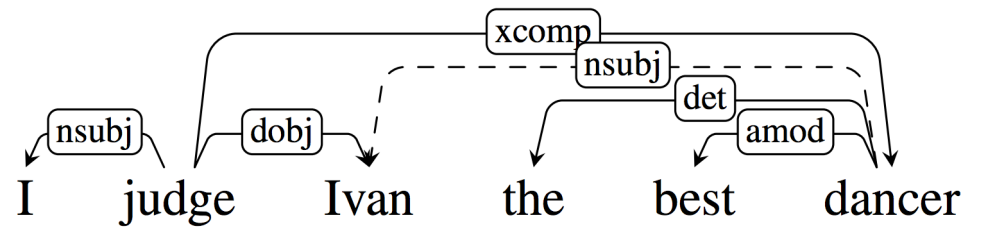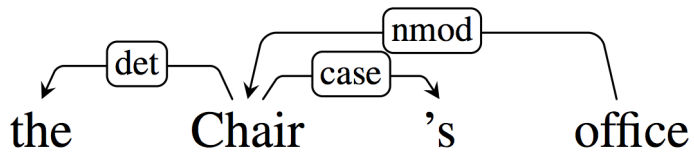
5. Relations between content words

The filings required under the proposed rules will be effective.

***nsubj*(effective, filings)**

cop(effective, be)

aux(effective, will)

This promotes:

- cross-lingual parallelism
- parallelism within a language across tense, aspect, etc.
- easy extraction of relations (predicate and its arguments)

Relation Extraction
Extracting protein-protein interactions

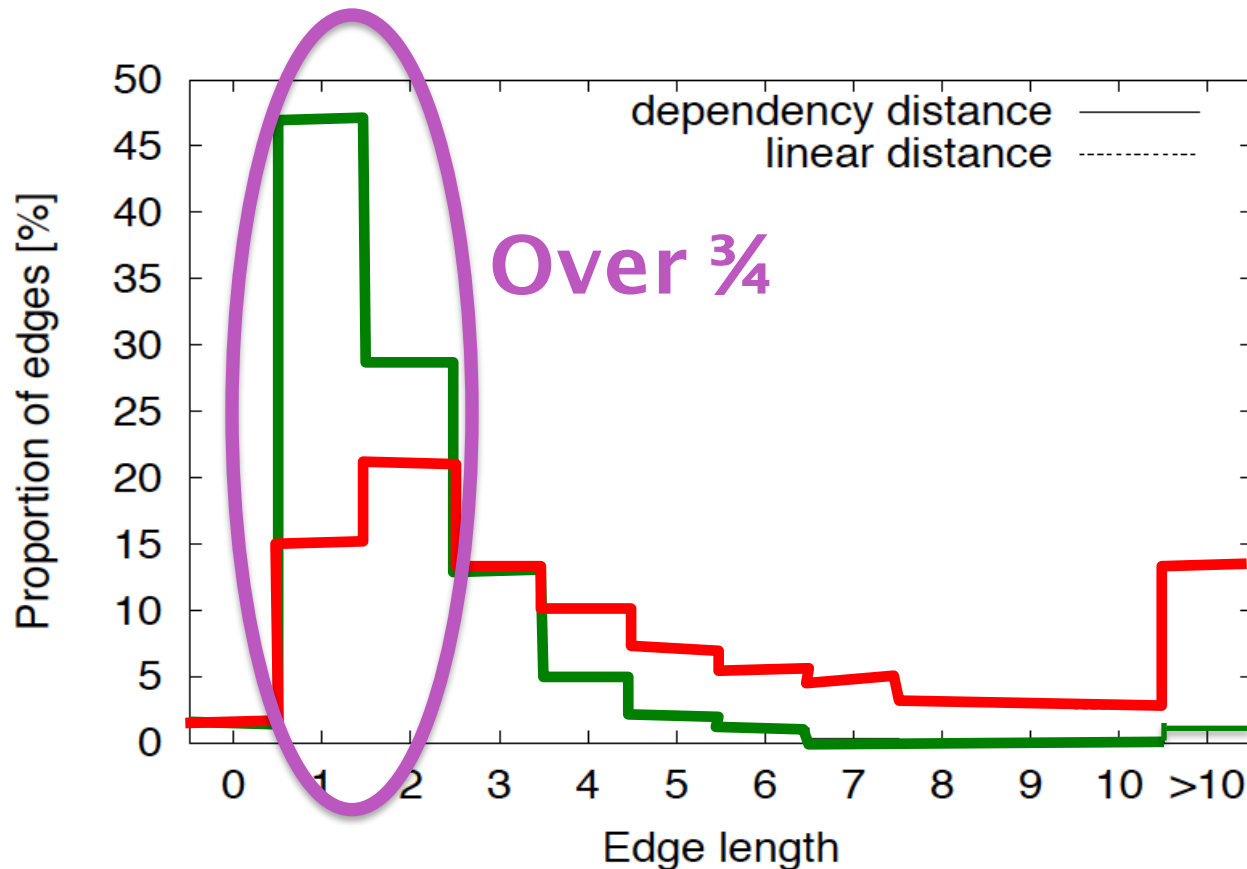$$\text{PROTEIN} \xleftarrow{\textit{nsubj}} \text{interact} \xrightarrow{\textit{nmod:with}} \text{PROTEIN}$$

$$\text{PROTEIN} \xleftarrow{\textit{nmod:of}} \text{reduction} \xrightarrow{\textit{nmod:to}} \text{PROTEIN}$$

$$\text{PROTEIN} \xleftarrow{\textit{nsubj}} \text{encode} \xrightarrow{\textit{dobj}} \text{PROTEIN}$$

Universal/Stanford Dependencies as a representation for relation extraction

- Stanford Dependencies produce short paths between related content words; widely used in BioNLP relation extraction tasks

Over ¾

Björne et al. 2009

6. A lexicalist analysis

- "Lexical integrity principle"
  [Chomsky 1970, Bresnan and Mchombo 1995, Aronoff 2007]
  - the word is a fundamental unit
  - the morphological processes that build up words are fundamentally different from those of syntax

- Grammatical relations should be between whole words
- Basic annotation units are words – syntactic words
  - We do separate off **clitics**
- Words have morphological properties
- Words enter into syntactic relations

7. Less can be better

- An analysis which you can't fit on a slide probably isn't right for many purposes
    - Cf. HPSG, minimalism

- Empty elements, polygraphs, more fine-grained relations, etc., etc. all have their appeal, but also have a cost….

7. Less can be better

- In general, the approach of linguistics has been to represent **everything**
  - Find some new distinction in behavior, add some new feature or tree structure or functional head to account for it.
  - But the result is incomprehensible to the common person
- In contrast, in school, we are taught that English has 8 parts of speech:
  - Noun, Verb, Adjective, Adverb, Pronoun, Preposition, Interjection, Conjunction
- This isn't really true, but it has enough fidelity, enough simplicity, and enough tradition to satisfy most people

37

4. Some personal reflections on grammar

This part's for Tim!

A big tent

The amazing – and great! – thing is that many people from disparate backgrounds have found UD congenial enough that they feel able and motivated to use it

So what exactly is UD?

- ## A new linguistic theory?

  No, but we like to think it is well-informed by linguistic theory and potentially useful also for linguistic studies

- ## A better parsing framework?

  Maybe not, since there is some evidence that parsers prefer function words as heads, so we may have to tweak the representations for parsing

- ## The ultimate annotation scheme?

  Not quite, more like a lingua franca f   ef
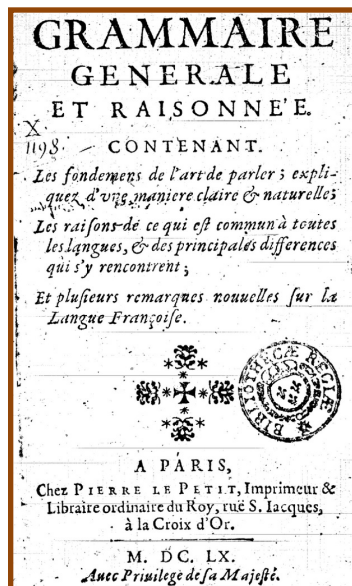 some annotation projects

  Well, who knows?

- ## A universal grammar?

  Not in the Chomskyan sense, but hopefully in the more practical sen
  multilingual NLP by bringing a little order into the chaos
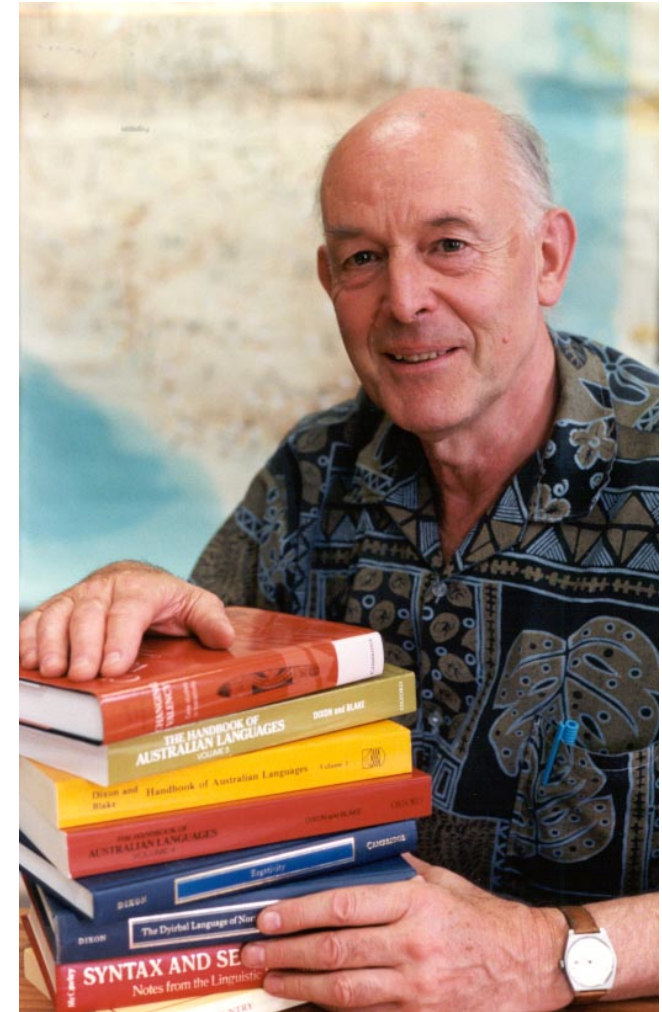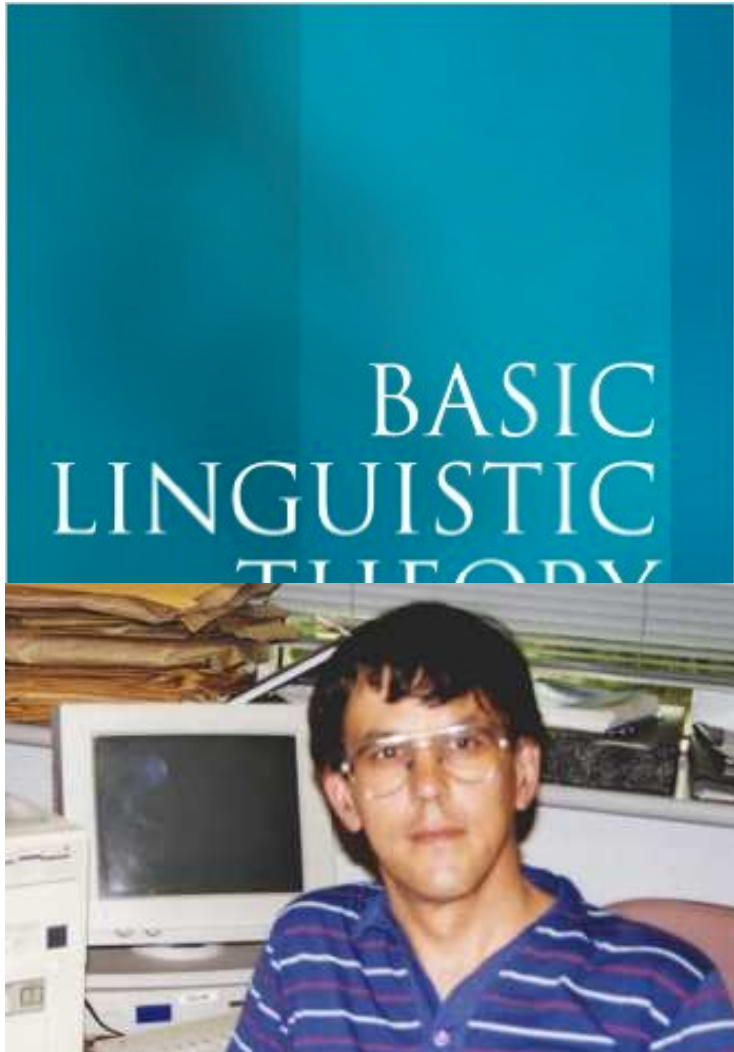
"In its substance, grammar is one and the same in all languages, even if it accidentally varies."
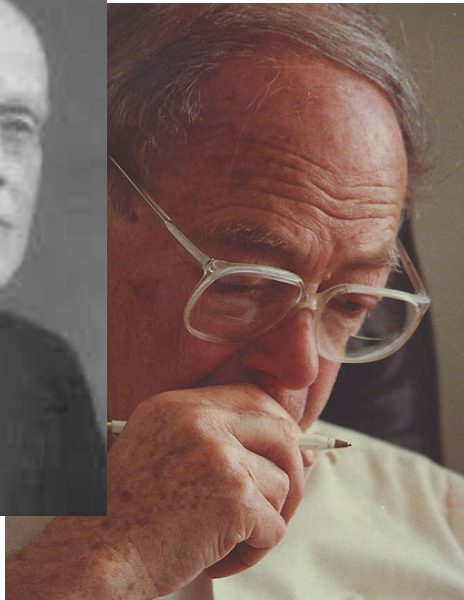
My origin in linguistics at the Australian National University: Bob Dixon

Academic Genealogy

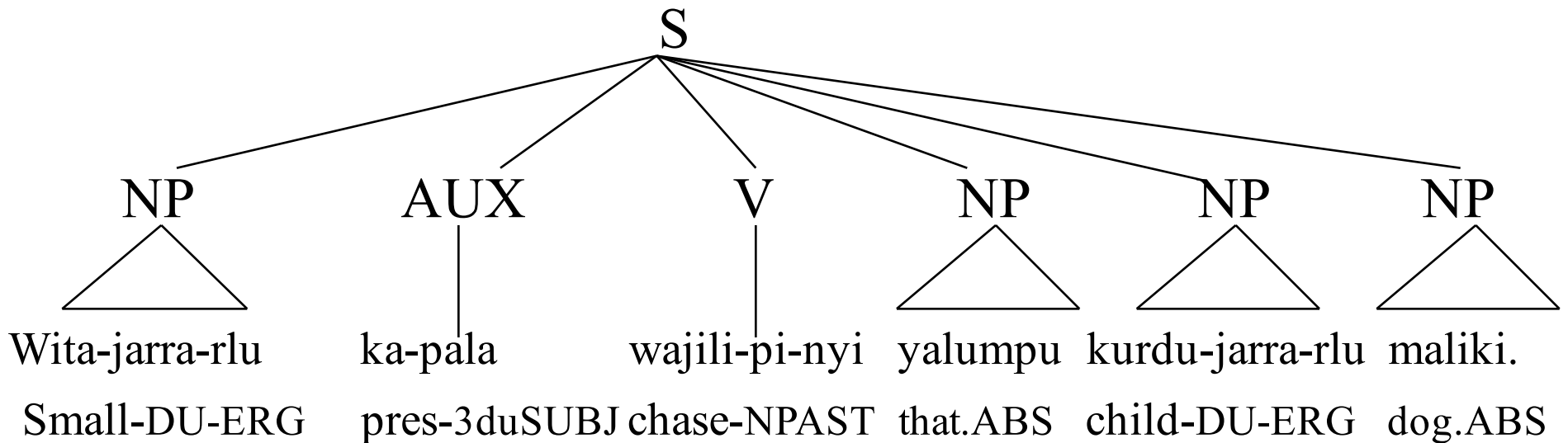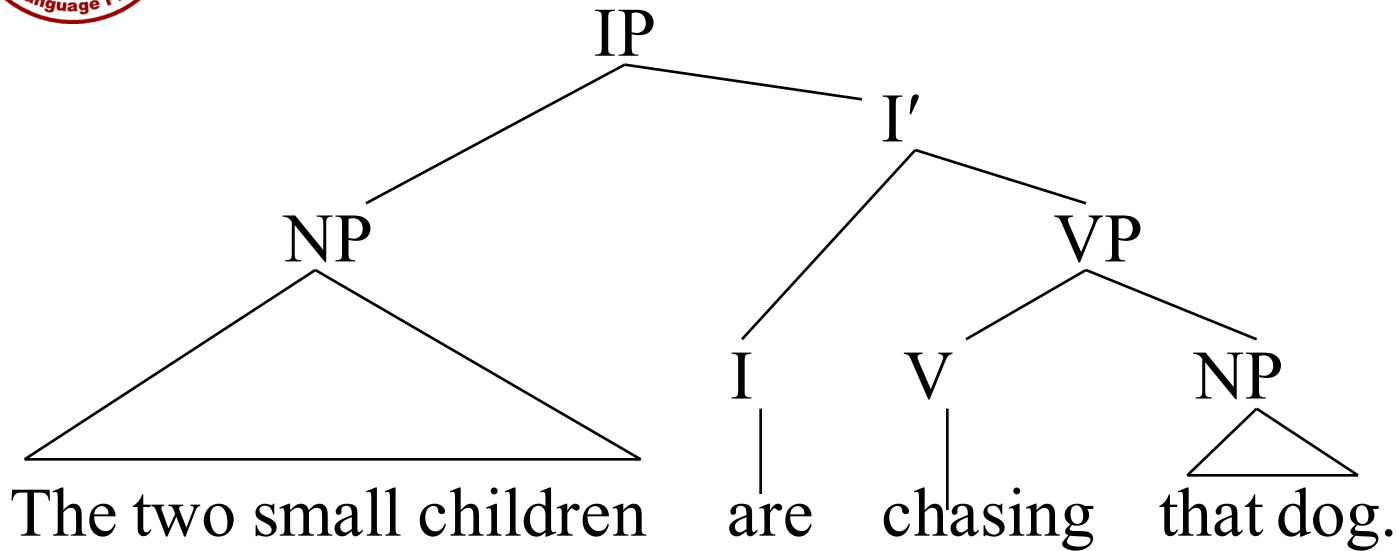# LFG's universalist view of language

- There is "a common organizing structure of all languages that underlies their superficial variations in modes of expression" (Bresnan 2011, *Lexical Functional Syntax*)

- Much of this structure is revealed at the level of **grammatical relations**
  - E.g., Passives that look very different in different languages can be described by a universal passive mapping.

Some differences between English and Warlpiri

The two small children are chasing that dog.

NP	AUX	V	NP	NP	NP
Wita-jarra-rlu	ka-pala	wajili-pi-nyi	yalumpu	kurdu-jarra-rlu	maliki.
Small-DU-ERG	pres-3duSUBJ	chase-NPAST	that.ABS	child-DU-ERG	dog.ABS

Surface evidence for constituency

Possible word orders in Warlpiri that are not possible in English:

- *The two small are chasing that children dog.
- *The two small are dog chasing that children.
- *Chasing are the two small that dog children.
- *That are children chasing the two small dog.

Word orders not possible in Warlpiri:

- The two small that dog are children chasing.
- The two small children chasing are that dog.

Something that English and Warlpiri have in common

- Lucy is hitting herself.
- *Herself is hitting Lucy.

- Napaljarri-rli   ka-nyanu paka-rni
  Napaljarri-ERG   PRES-REFL hit-NONPAST

  "Napaljarri is hitting herself."

- *Napaljarri   ka-nyanu paka-rni
  Napaljarri.ABS PRES-REFL hit-NONPAST

  "Herself is hitting Napaljarri."

What English and Warlpiri have in common according to LFG

- Same grammatical relations (and semantic roles)
  - SUBJECT: the two small children: AGENT
  - PREDICATE: are chasing
  - OBJECT: that dog: PATIENT
- Different codings of grammatical relations:
  - English subject: NP immediately under S
  - Warlpiri subject: Ergative case marked NP (if verb is transitive)

Levels of Representation in LFG

[s [np The bear] [vp ate [np a sandwich]]]   **constituent structure**

Grammatical encoding
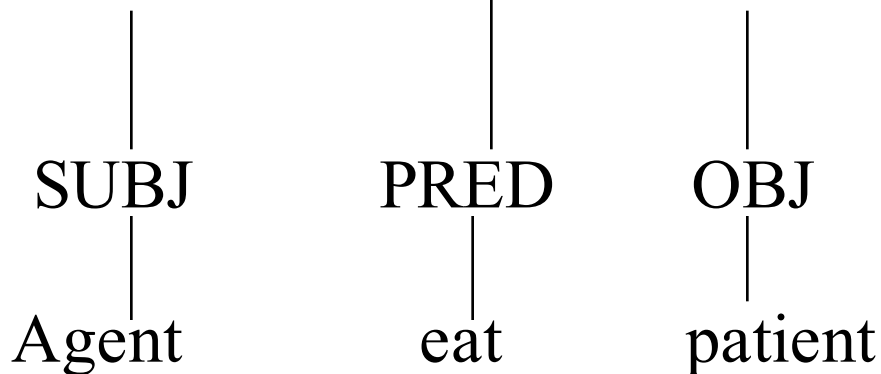
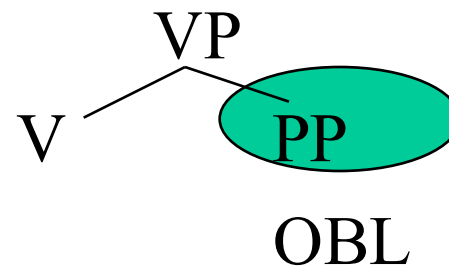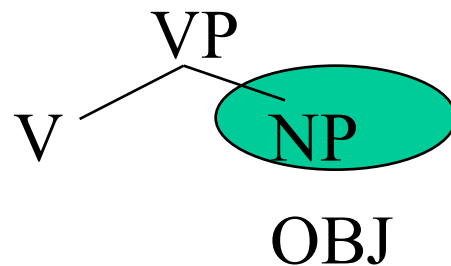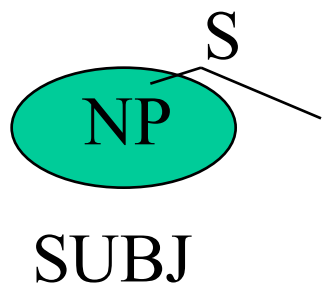SUBJ   PRED   OBJ   **functional structure**

Lexical mapping

Agent   eat   patient   **thematic roles**

Eat < agent   patient >   lexical mapping

SUBJ   OBJ

Grammatical
Encoding

For English!!!

SUBJ   OBJ   OBL

Syntactic Tree

```
                    S
              ┌─────┴─────┐
             NP           VP
              │      ┌─────┴─────┐
              N      V        VP-bar
              │      │      ┌────┴────┐
                   INF        VP
                    │     ┌────┴────┐
                    V         PP
                          ┌────┴────┐
                          P        NP
                                ┌───┴───┐
                              DET       N
```

Lions seem to live in the forest

F-structure
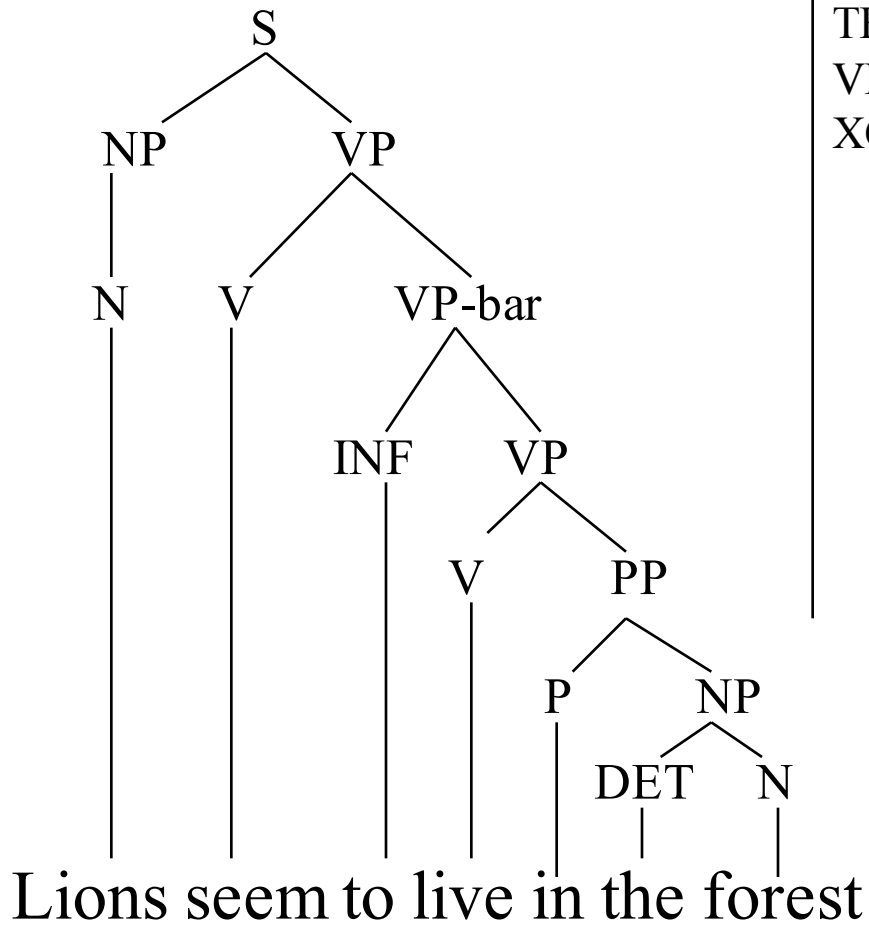
SUBJ		PRED	'lion'
		NUM	pl
		PERS	3
PRED	'seem < theme > SUBJ'		
	XCOMP		
TENSE	pres		
VFORM	fin		
XCOMP	SUBJ [ ]		
	VFORM INF		
	PRED 'live< theme loc >'		
		SUBJ	OBL-loc
	OBL-loc	CASE	OBL-loc
		PRED	'in<OBJ>'
		OBJ	PRED 'forest'
			NUM sg
			PERS 3
			DEF +

Organization of grammar

There are many differences, but on the other hand, many frameworks have levels that are similar….

Theory	Constituency	Grammatical Relations	Argument Structure
LFG	c-structure	f-structure	a-structure
HPSG	DTR lists	valence lists	ARG-ST
GB	S-structure / PF	S-structure configurations	D-structure / θ-theory
Relational Grammar		final relations	initial relations
Functional Generative Desc.	analytical level		tectogrammatical level
Basic Linguistic Theory		A, S, O	

Does UD still have flaws? Plenty.

- It's hard to see how to get a satisfactory treatment of ellipsis without introducing empty nodes … which we're reluctant to do

- To what extent should universal part of speech labels reflect an inherent lexical or formal category vs. function in a clause?

- The *name* relation for proper names wasn't necessarily a good idea, and making them always left headed may not be right

- At present with use of *nmod* for modifying nominals and predicates, you can't tell apart predicate and nominal modification when you have a predicate nominal

- Treatment of light verb constructions and complex predicates

- …

52

A victory for linguistics

and the average user

The End