



UPPSALA
UNIVERSITET

Universal Dependency Parsing

Joakim Nivre

Uppsala University
Department of Linguistics and Philology

The Grand Vision

The Grand Vision

The problem

- Why 90% parsing accuracy for English but only 80% for Finnish?
- Are some languages intrinsically harder to parse?
- Not just morphological richness – many typological parameters

The Grand Vision

The problem

- Why 90% parsing accuracy for English but only 80% for Finnish?
- Are some languages intrinsically harder to parse?
- Not just morphological richness – many typological parameters

An ideal solution

- A universal parser for all languages
- Linguistic universals are hard-coded
- Typological parameters are learned from data

The Grand Vision

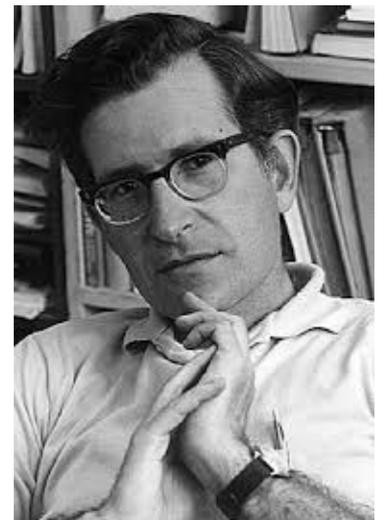
The problem

- Why 90% parsing accuracy for English but only 80% for Finnish?
- Are some languages intrinsically harder to parse?
- Not just morphological richness – many typological parameters

An ideal solution

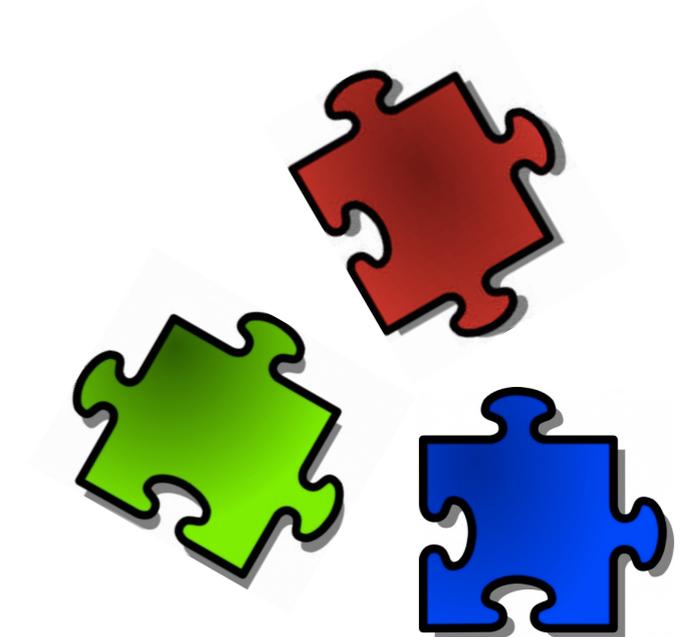
- A universal parser for all languages
- Linguistic universals are hard-coded
- Typological parameters are learned from data

Gee!
What a neat idea!



Pieces of the Puzzle

1. Morphosyntactic disambiguation
2. Universal dependency annotation
3. Parsing with universal dependencies



Part I

Morphosyntactic Disambiguation



Joint work with Bernd Bohnet, Igor Boguslavsky,
Richárd Farkas, Filip Ginter and Jan Hajič

Background

Parsing accuracy for morphologically rich languages tends to be lower than for languages like English (Nivre et al., 2007)

Hypothesized explanations (Tsarfaty et al. 2010, 2013):

- Strict separation morphology-syntax
- Data sparsity due to high type-token ratio

Suggested remedies

- Joint morphological and syntactic analysis (Lee et al., 2011)
- Lexical resources (Hajič, 2000; Goldberg and Elhadad, 2013)

This Study

Parsing techniques:

- Transition-based model for joint morphological and syntactic analysis
- Lexical resources integrated as hard or soft constraints

Evaluation on five morphologically rich languages:

- Czech, Finnish, German, Hungarian, Russian
- New state of the art in dependency parsing for all languages

Bernd Bohnet, Joakim Nivre, Igor Boguslavsky, Richárd Farkas, Filip Ginter, and Jan Hajič. 2013. Joint Morphological and Syntactic Analysis for Richly Inflected Languages. *Transactions of the Association for Computational Linguistics* 1:415–428

Representations



Representations

For each word of a sentence w_1, \dots, w_n :

Ein Haus hat er in Ulm gebaut .

Representations

For each word of a sentence w_1, \dots, w_n :

- A part-of-speech tag $p \in P$

Ein	Haus	hat	er	in	Ulm	gebaut	.
ART	NN	VAFIN	PPER	APPR	NE	VVPP	\$.

Representations

For each word of a sentence w_1, \dots, w_n :

- A part-of-speech tag $p \in P$
- A morphological feature bundle $m \in M$

Ein	Haus	hat	er	in	Ulm	gebaut	.
ART	NN	VAFIN	PPER	APPR	NE	VVPP	\$.
acclsglneut	acclsglneut	sgl3 preslind	nomlsglmascl3	—	datlsglneut	—	—

Representations

For each word of a sentence w_1, \dots, w_n :

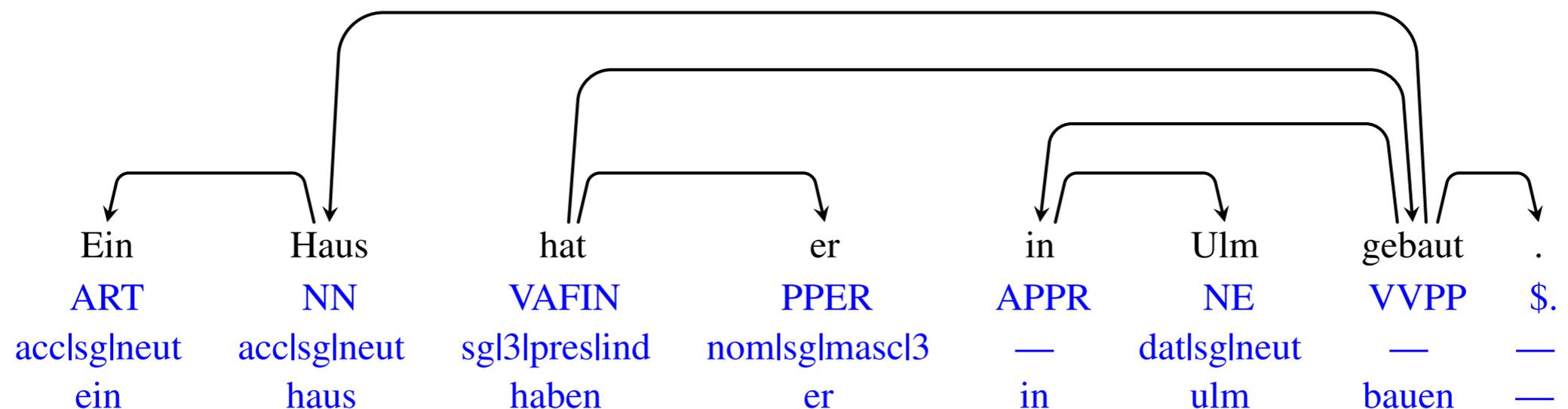
- A part-of-speech tag $p \in P$
- A morphological feature bundle $m \in M$
- A lemma $l \in Z^*$

Ein	Haus	hat	er	in	Ulm	gebaut	.
ART	NN	VAFIN	PPER	APPR	NE	VVPP	\$.
acclsglneut	acclsglneut	sgl3 preslind	nomlsglmascl3	—	datlsglneut	—	—
ein	haus	haben	er	in	ulm	bauen	—

Representations

For each word of a sentence w_1, \dots, w_n :

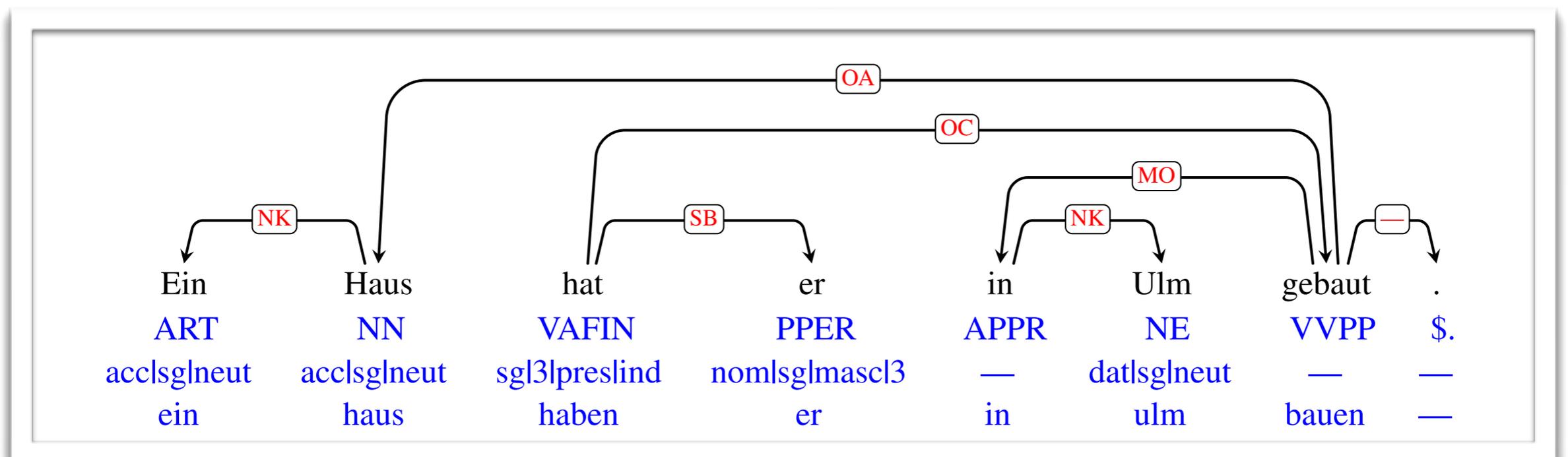
- A part-of-speech tag $p \in P$
- A morphological feature bundle $m \in M$
- A lemma $l \in Z^*$
- A syntactic head $h \in \{0, w_i, \dots, w_n\}$



Representations

For each word of a sentence w_1, \dots, w_n :

- A part-of-speech tag $p \in P$
- A morphological feature bundle $m \in M$
- A lemma $l \in Z^*$
- A syntactic head $h \in \{o, w_i, \dots, w_n\}$
- A dependency label $d \in D$



Parsing Framework

Transition system (Bohnet and Nivre, 2012):

- Arc-standard system with online reordering (Nivre, 2009)
- Select morphology when shifting words to the stack

Beam search and structured learning (Zhang and Clark, 2008)

Preprocessing (at learning and parsing time)

- Tagger assigns **k** best tags and feature bundles
- Parser can only select analyses licensed by preprocessor

Experiment I

Pipeline

Preprocessor assigns a single tag and feature bundle per word

Beam = 40 distinct trees

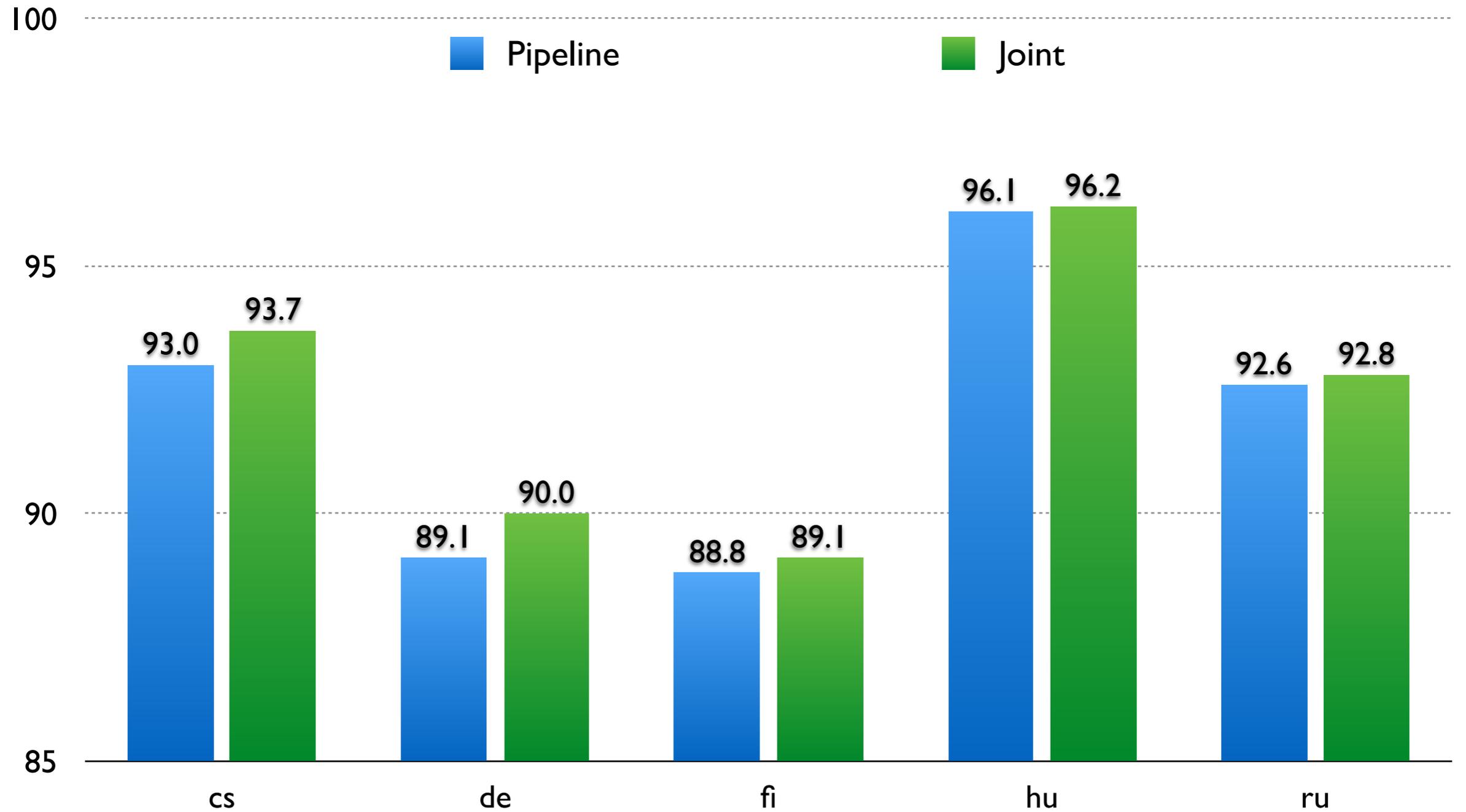
Joint

Preprocessor assigns up to 2 tags and feature bundles per word

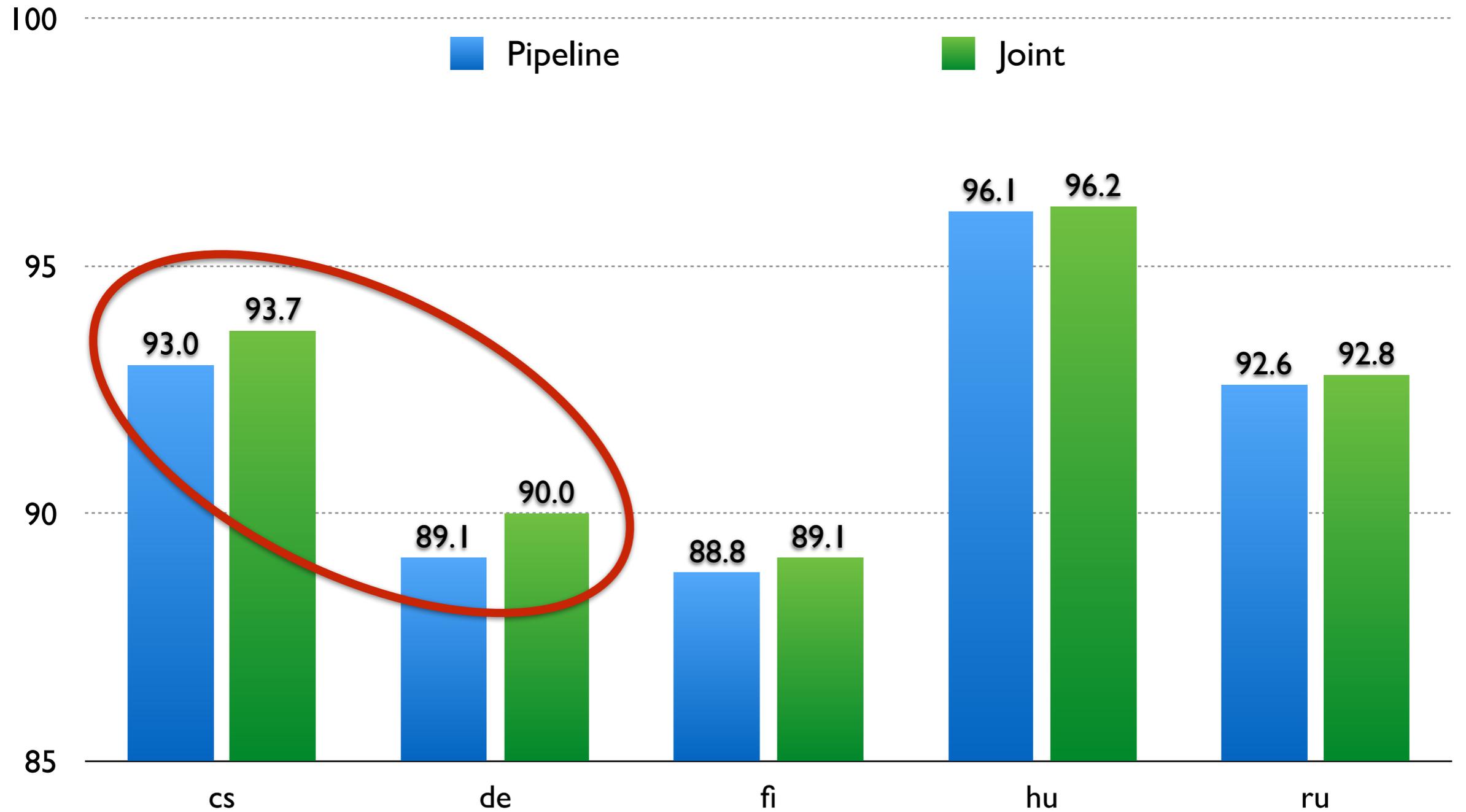
Beam = 40 distinct trees + 8 tag variants + 8 feature variants

	Treebank	Train	Dev	Test	P	M	D
cs	PDT (Hajič et al., 2001)	1249K	88K	70K	12	1851	49
de	Tiger (Brants et al., 2002)	648K	32K	32K	54	257	43
fi	TDT (Haverinen et al., 2013)	183K		21K	12	1917	47
hu	Szeged (Farkas et al., 2012)	1101K	210K	171K	22	1105	33
ru	SynTagRus (Boguslavsky et al., 2000)	575K	73K	72K	14	454	78

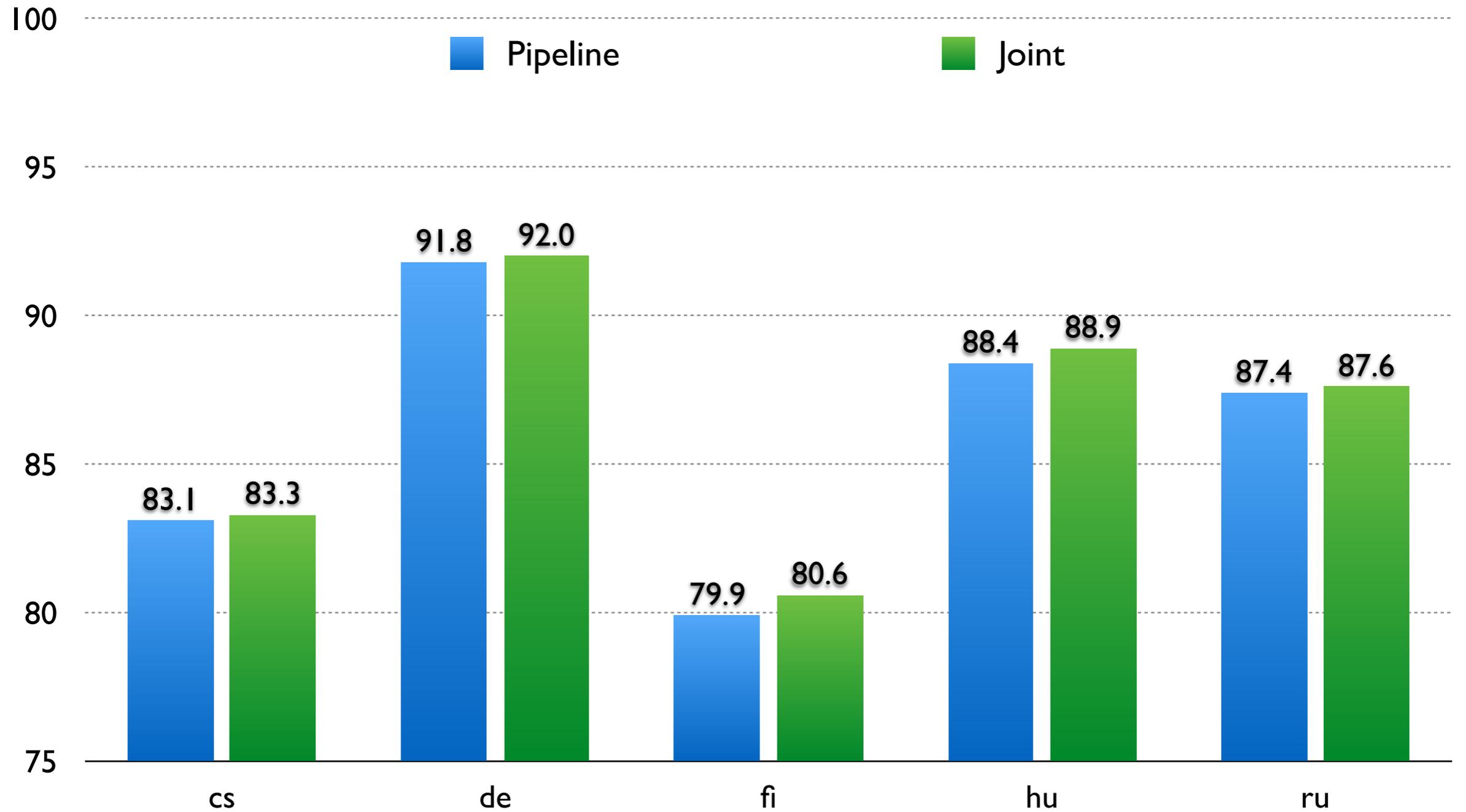
Morphological Accuracy



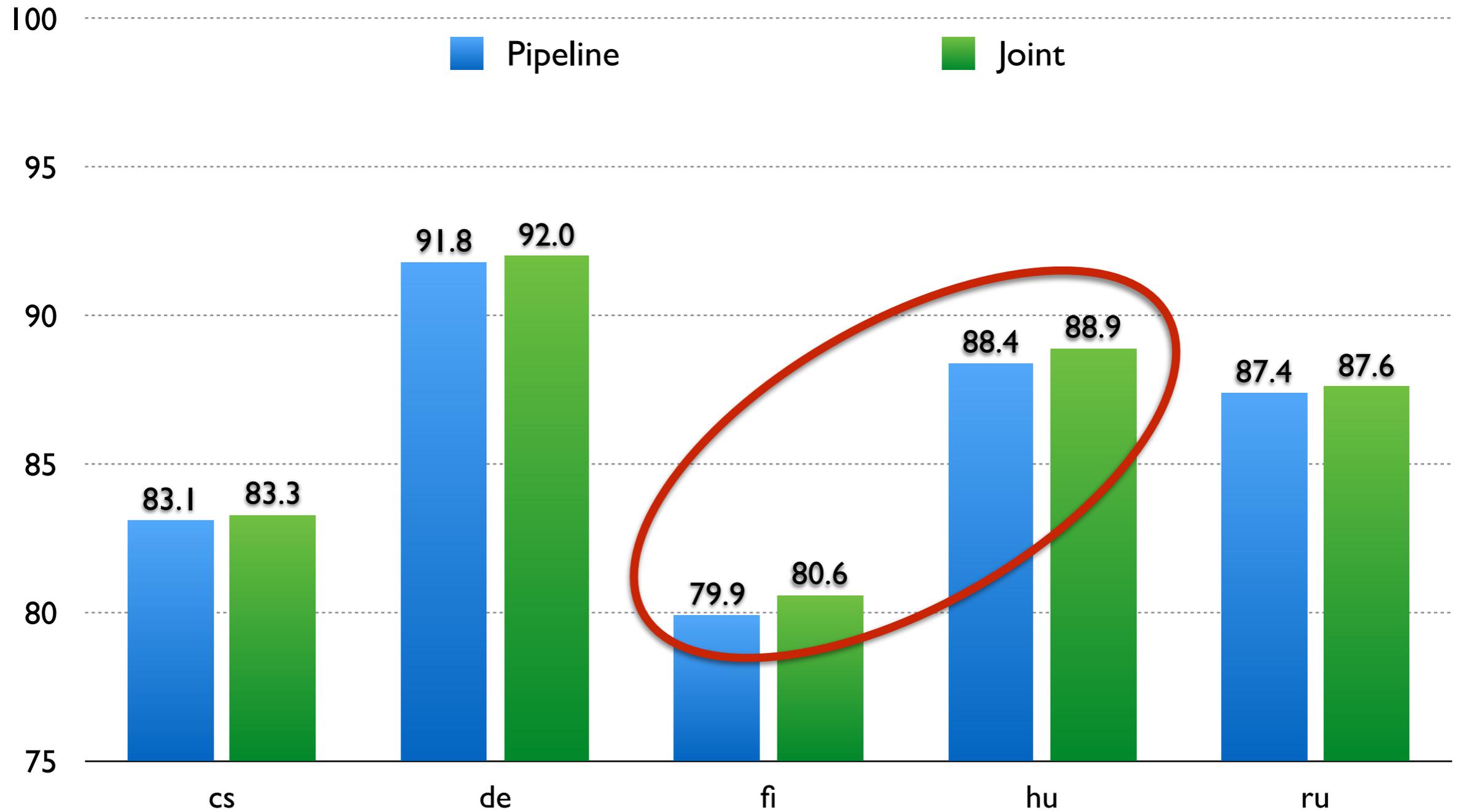
Morphological Accuracy



Syntactic Accuracy



Syntactic Accuracy

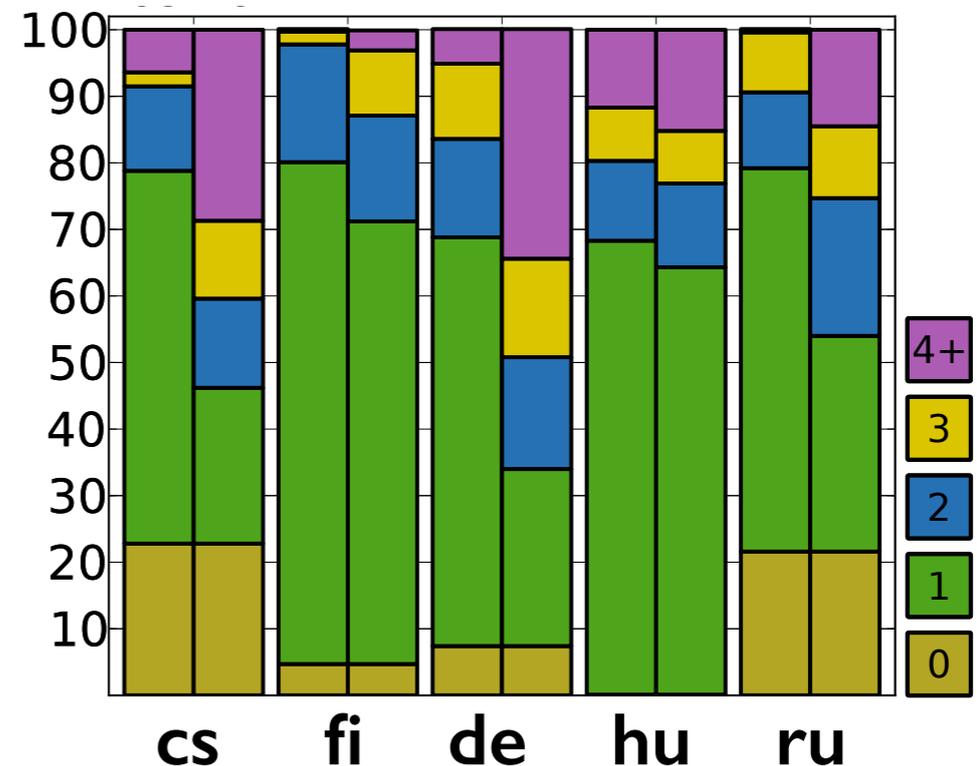


Experiment 2

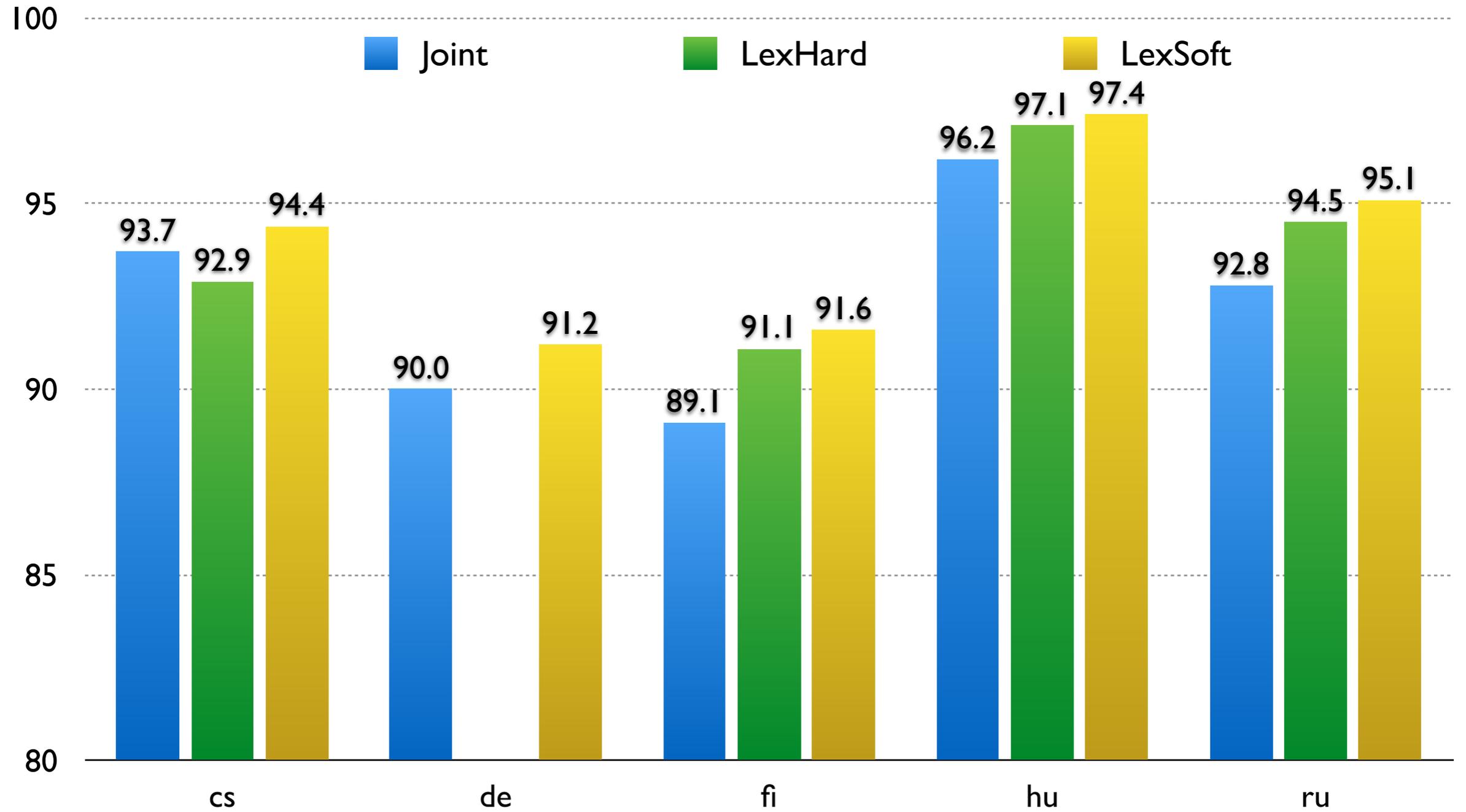
Morphological lexicons

- Lookup of tag, feature bundle and lemma
- Added as **hard** or **soft** constraints to preprocessor and parser
- Lemma selected deterministically from form + tag + feature bundle

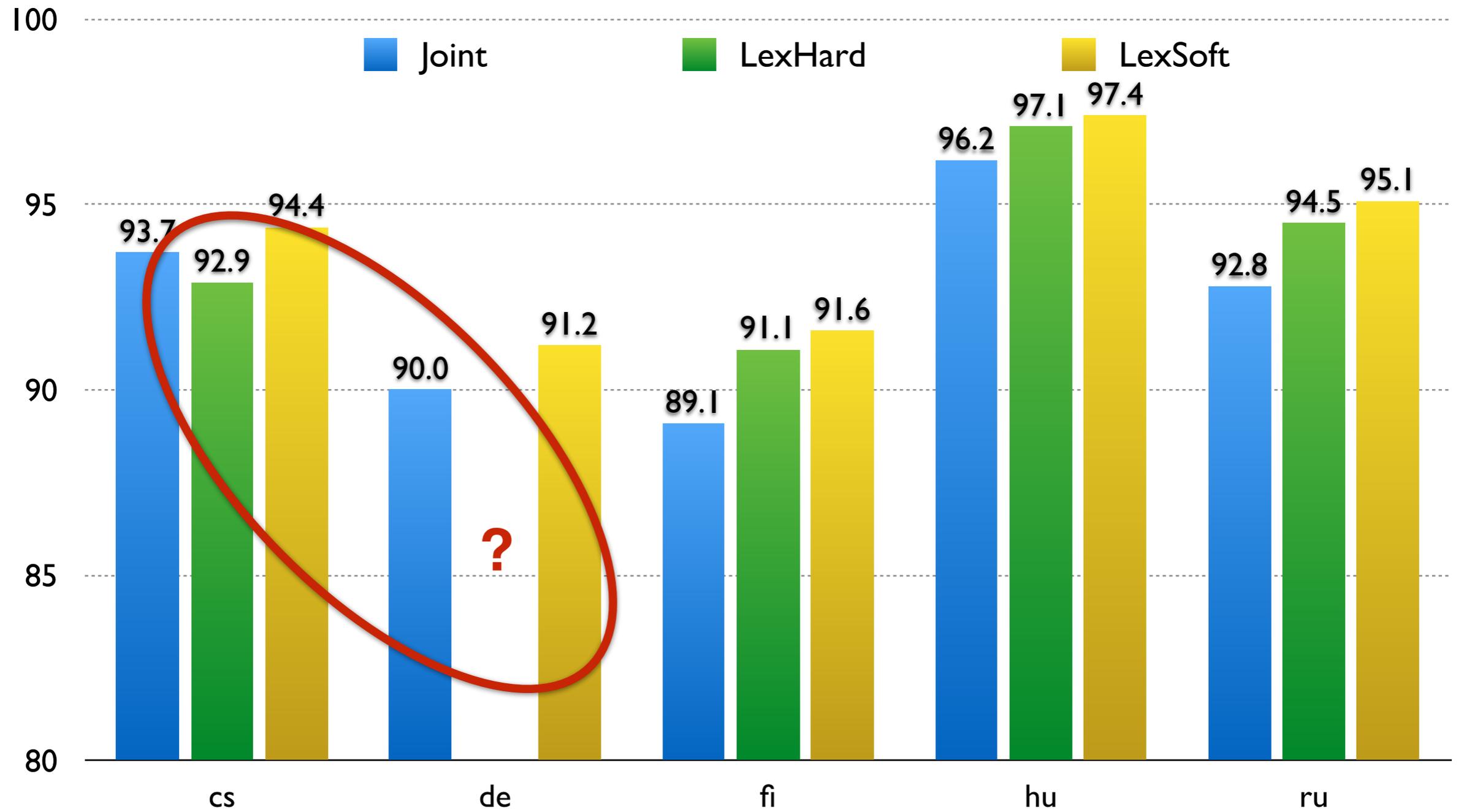
	Lexicon
cs	PDT (Hajič and Hladká, 1998)
de	SMOR (Schmid et al., 2004)
fi	OMorFi (Pirinen, 2011)
hu	morphdb.hu (Trón et al., 2006)
ru	ETAP-3 (Apresian et al., 2003)



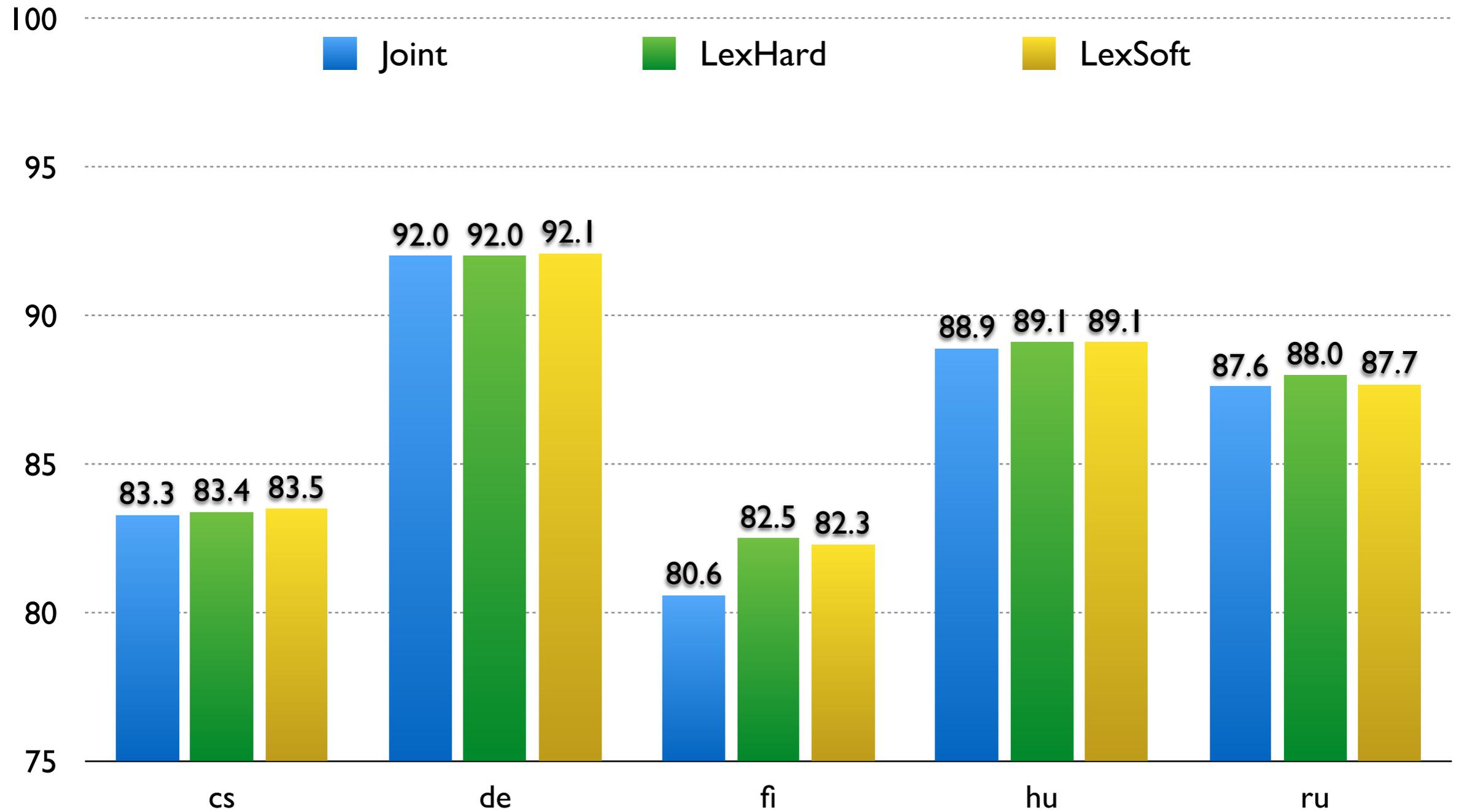
Morphological Accuracy



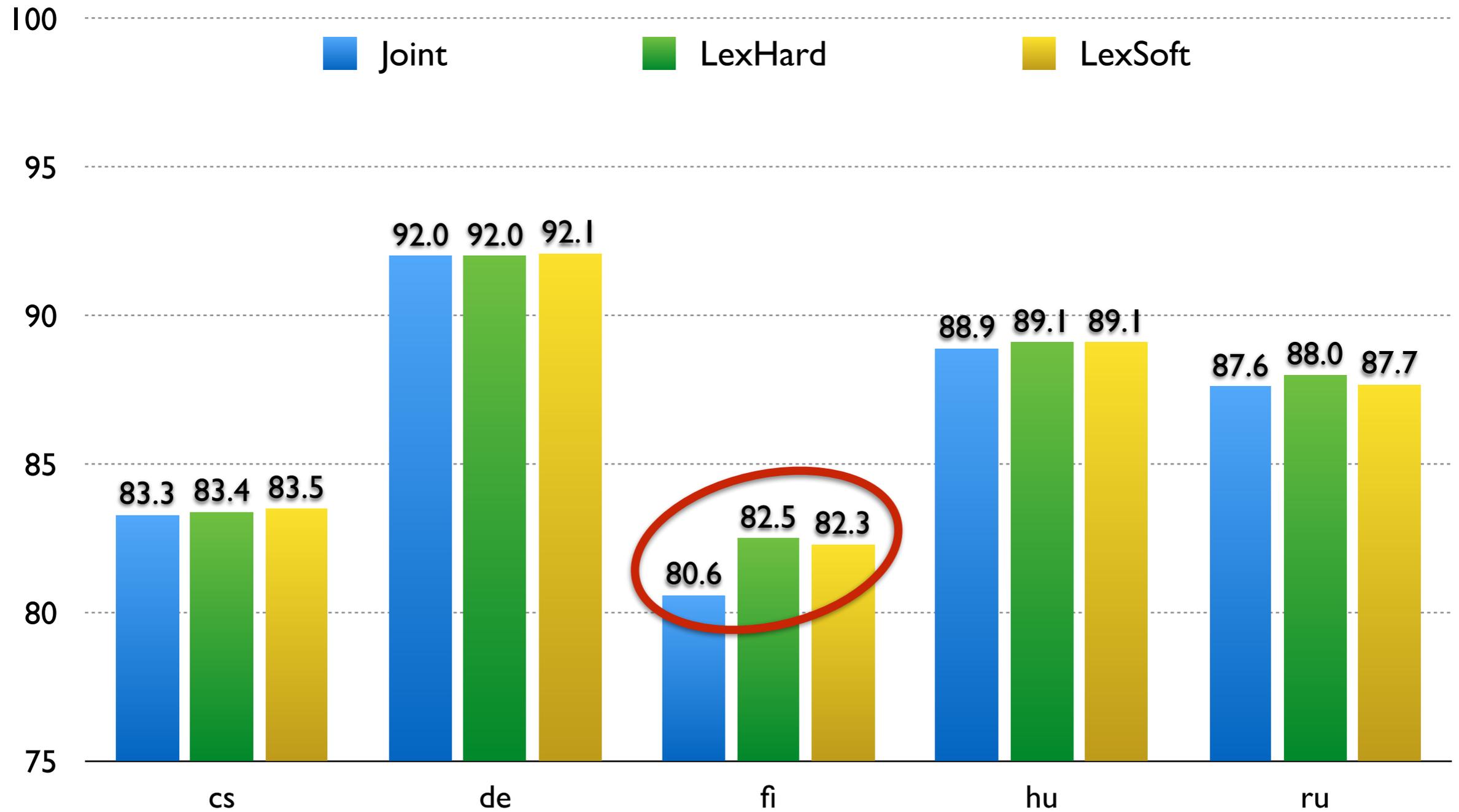
Morphological Accuracy



Syntactic Accuracy



Syntactic Accuracy



Discussion

Joint inference benefits both morphology and syntax

- Strongest effect on morphology for Czech and German – syncretism?
- Strongest effect on syntax for Finnish and Hungarian – why?

Lexical resources mitigate data sparseness

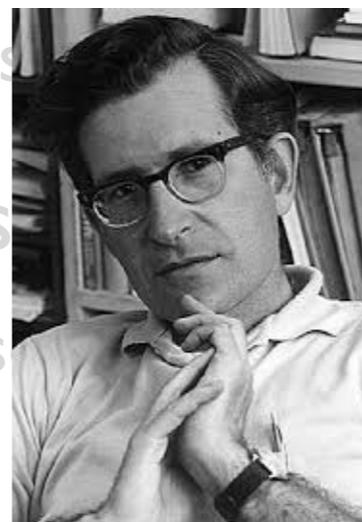
- Strongest effect on morphology – soft constraints always best?
- Weaker effect on syntax – relevant errors fixed by joint inference?
- Strong effect on syntax for Finnish – sparse data?

Discussion

Joint inference benefits both morphology and syntax

- Strongest effect on morphology for Czech and German – syncretism?
- Strongest effect on syntax for Finnish – why?

Nice!
But why only 82%
for Finnish?



Lexical resources mitigate data sparsity

- Strongest effect on morphology – syntactic resources always best?
- Weaker effect on syntax – mitigated by joint inference?
- Strong effect on syntax for Czech – sparse data?

Can we
even compare the
numbers?

Part 2

Universal Dependency Annotation



Joint work with Ryan McDonald, Slav Petrov, Chris Manning, Marie de Marneffe, Jinho Choi, Filip Ginter, Yoav Goldberg, Jan Hajič and Reut Tsarfaty

Apples and Oranges

Apples and Oranges

Treebank annotation schemes vary across languages

- Hard to compare parsing results across languages (Nivre et al., 2007)
- Hard to evaluate cross-lingual learning (McDonald et al., 2013)
- Hard to make progress towards a universal parser?

Apples and Oranges

Treebank annotation schemes vary across languages

- Hard to compare parsing results across languages (Nivre et al., 2007)
- Hard to evaluate cross-lingual learning (McDonald et al., 2013)
- Hard to make progress towards a universal parser?

Recent initiatives:

- **HamleDT**: Conversion of 29 existing treebanks to a PDT-like annotation scheme (Zeman et al., 2012)
- **Universal Dependency Treebank Project**: New annotation, conversion and harmonization to Google universal PoS tags and Stanford dependencies (McDonald et al., 2013)

Case Study

Delexicalized transfer parsing with universal PoS tags

Source Training Language	Target Test Language											
	Unlabeled Attachment Score (UAS)						Labeled Attachment Score (LAS)					
	Germanic			Romance			Germanic			Romance		
	DE	EN	SV	ES	FR	KO	DE	EN	SV	ES	FR	KO
DE	74.86	55.05	65.89	60.65	62.18	40.59	64.84	47.09	53.57	48.14	49.59	27.73
EN	58.50	83.33	70.56	68.07	70.14	42.37	48.11	78.54	57.04	56.86	58.20	26.65
SV	61.25	61.20	80.01	67.50	67.69	36.95	52.19	49.71	70.90	54.72	54.96	19.64
ES	55.39	58.56	66.84	78.46	75.12	30.25	45.52	47.87	53.09	70.29	63.65	16.54
FR	55.05	59.02	65.05	72.30	81.44	35.79	45.96	47.41	52.25	62.56	73.37	20.84
KO	33.04	32.20	27.62	26.91	29.35	71.22	26.36	21.81	18.12	18.63	19.52	55.85

Ryan McDonald, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency Annotation for Multilingual Parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 92–97.

Case Study

Delexicalized transfer parsing with universal PoS tags

First evaluation of labeled accuracy

Source Training Language	Target Test Language											
	Unlabeled Attachment Score (UAS)						Labeled Attachment Score (LAS)					
	Germanic			Romance			Germanic			Romance		
	DE	EN	SV	ES	FR	KO	DE	EN	SV	ES	FR	KO
DE	74.86	55.05	65.89	60.65	62.18	40.59	64.84	47.09	53.57	48.14	49.59	27.73
EN	58.50	83.33	70.56	68.07	70.14	42.37	48.11	78.54	57.04	56.86	58.20	26.65
SV	61.25	61.20	80.01	67.50	67.69	36.95	52.19	49.71	70.90	54.72	54.96	19.64
ES	55.39	58.56	66.84	78.46	75.12	30.25	45.52	47.87	53.09	70.29	63.65	16.54
FR	55.05	59.02	65.05	72.30	81.44	35.79	45.96	47.41	52.25	62.56	73.37	20.84
KO	33.04	32.20	27.62	26.91	29.35	71.22	26.36	21.81	18.12	18.63	19.52	55.85

Ryan McDonald, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency Annotation for Multilingual Parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 92–97.

Case Study

Delexicalized transfer parsing with universal PoS tags

Results make typological sense First evaluation of labeled accuracy

Source Training Language	Target Test Language											
	Unlabeled Attachment Score (UAS)						Labeled Attachment Score (LAS)					
	Germanic			Romance			Germanic			Romance		
	DE	EN	SV	ES	FR	KO	DE	EN	SV	ES	FR	KO
DE	74.86	55.05	65.89	60.65	62.18	40.59	64.84	47.09	53.57	48.14	49.59	27.73
EN	58.50	83.33	70.56	68.07	70.14	42.37	48.11	78.54	57.04	56.86	58.20	26.65
SV	61.25	61.20	80.01	67.50	67.69	36.95	52.19	49.71	70.90	54.72	54.96	19.64
ES	55.39	58.56	66.84	78.46	75.12	30.25	45.52	47.87	53.09	70.29	63.65	16.54
FR	55.05	59.02	65.05	72.30	81.44	35.79	45.96	47.41	52.25	62.56	73.37	20.84
KO	33.04	32.20	27.62	26.91	29.35	71.22	26.36	21.81	18.12	18.63	19.52	55.85

Ryan McDonald, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency Annotation for Multilingual Parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 92–97.

Too Many Standards?

Too Many Standards?



Too Many Standards?

English SD

de Marneffe et al. (2006)

Google SD

McDonald et al. (2013)

Too Many Standards?

English SD

de Marneffe et al. (2006)

Google SD

McDonald et al. (2013)

Finnish SD

Haverinen et al. (2013)

Hebrew SD

Tsarfaty (2013)

Too Many Standards?

CLEAR

Choi (2012)

English SD

de Marneffe et al. (2006)

HamleDT

Zeman et al. (2012)

Google SD

McDonald et al. (2013)

Finnish SD

Haverinen et al. (2013)

Hebrew SD

Tsarfaty (2013)

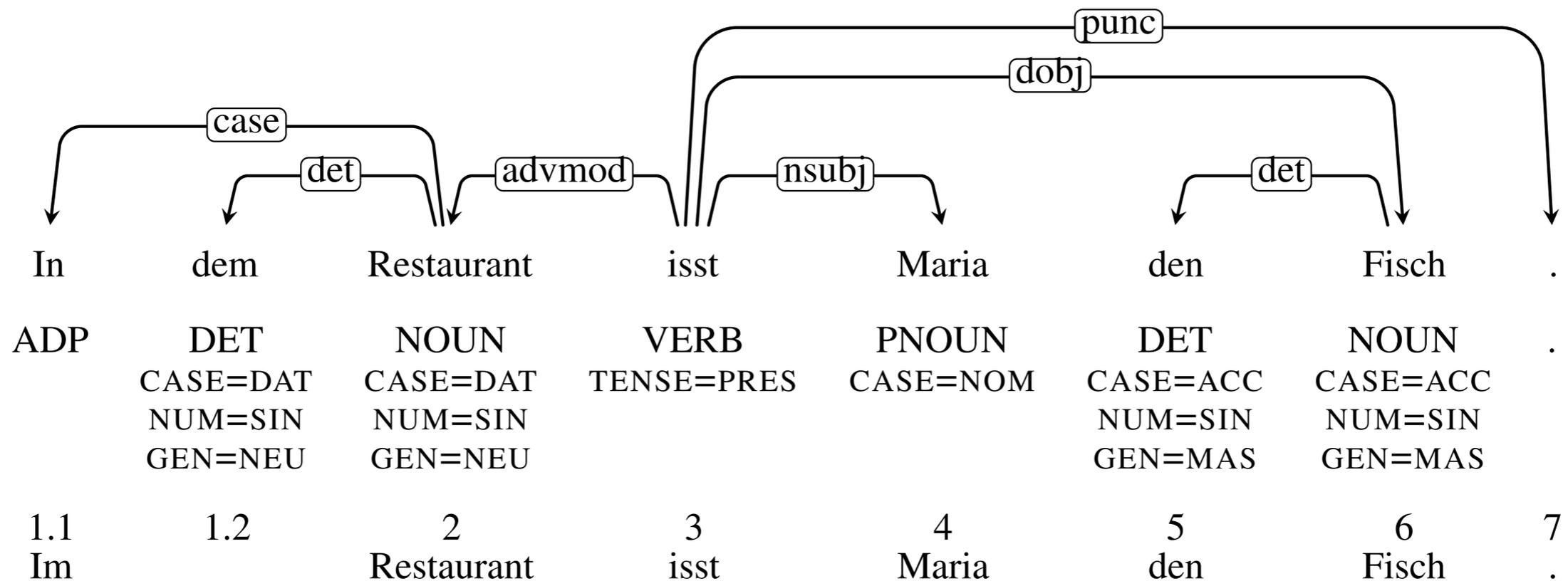
Too Many Standards?

Universal Dependencies

Marie-Catherine De Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre and Christopher D. Manning. 2014. Universal Stanford Dependencies: a Cross-Linguistic Typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*.

Universal Dependencies

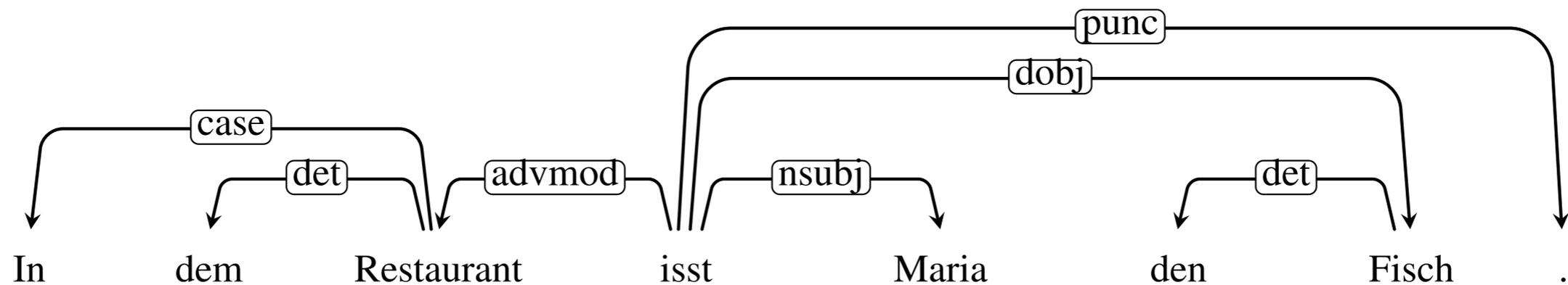
<http://universaldependencies.github.io/docs/>



Universal Dependencies

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

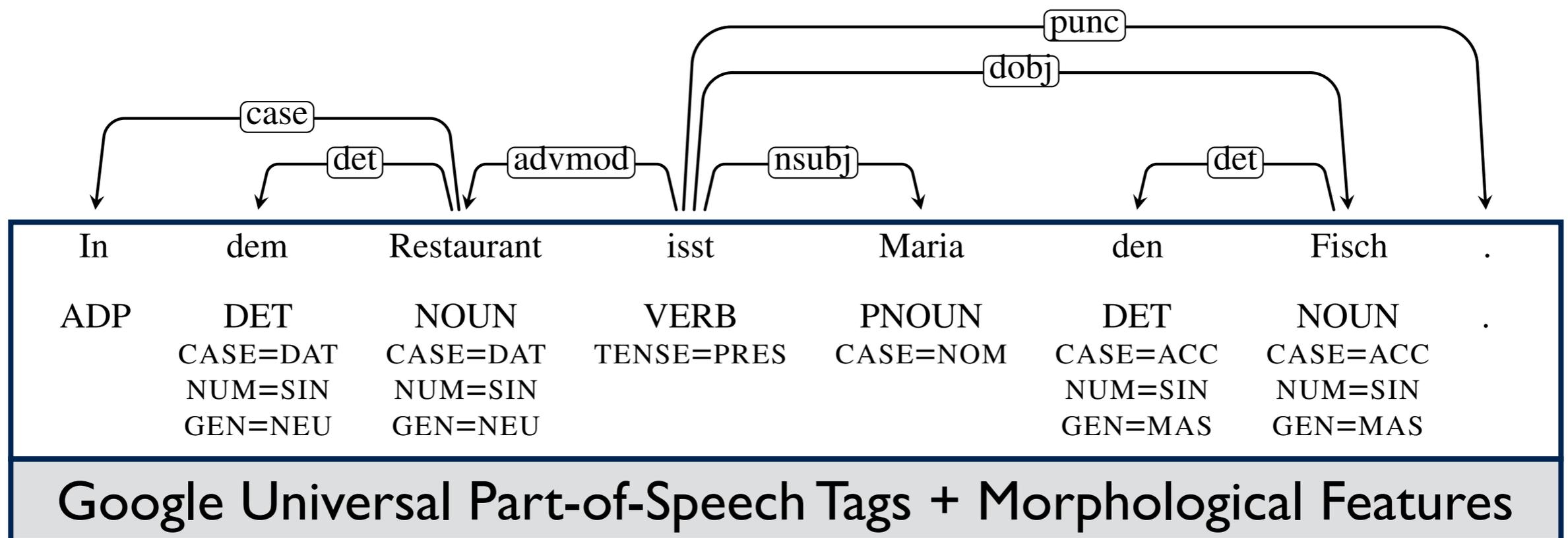
Stanford Universal Dependencies



ADP	DET	NOUN	VERB	PNOUN	DET	NOUN	.
	CASE=DAT NUM=SIN GEN=NEU	CASE=DAT NUM=SIN GEN=NEU	TENSE=PRES	CASE=NOM	CASE=ACC NUM=SIN GEN=MAS	CASE=ACC NUM=SIN GEN=MAS	
1.1	1.2	2	3	4	5	6	7
Im		Restaurant	isst	Maria	den	Fisch	.

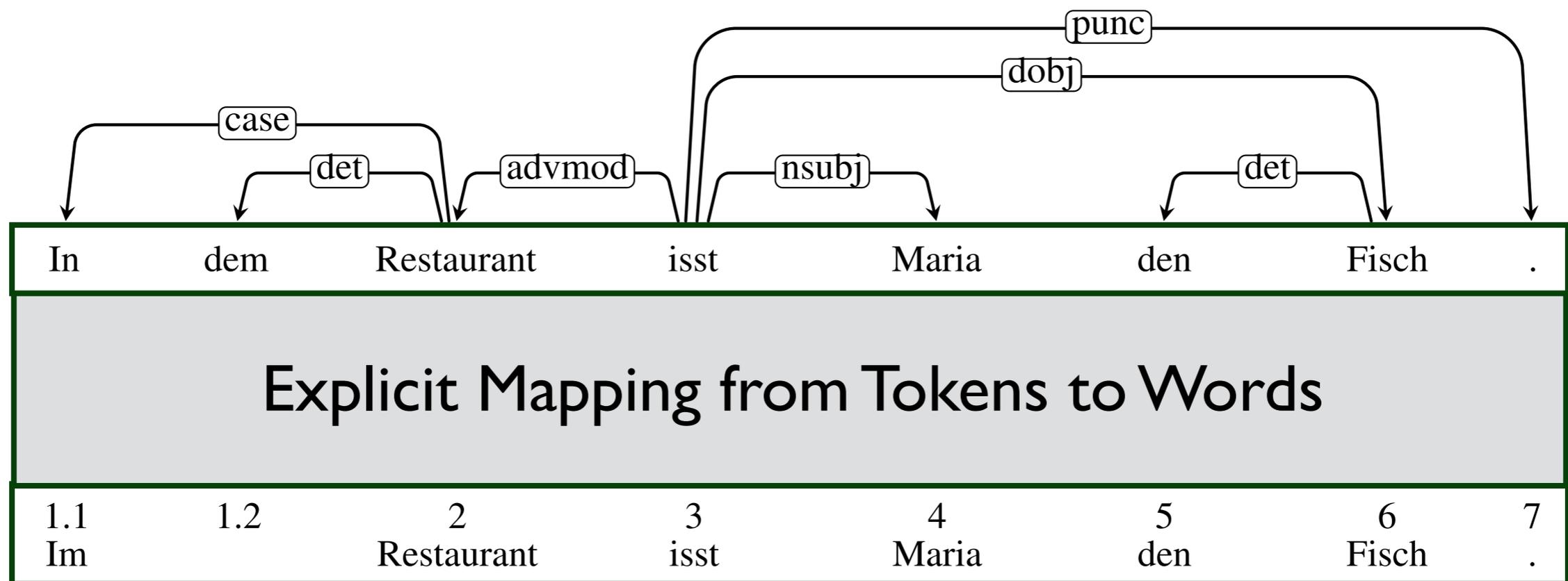
Universal Dependencies

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



Universal Dependencies

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



Guiding Principles

Guiding Principles

Maximize parallelism

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

Guiding Principles

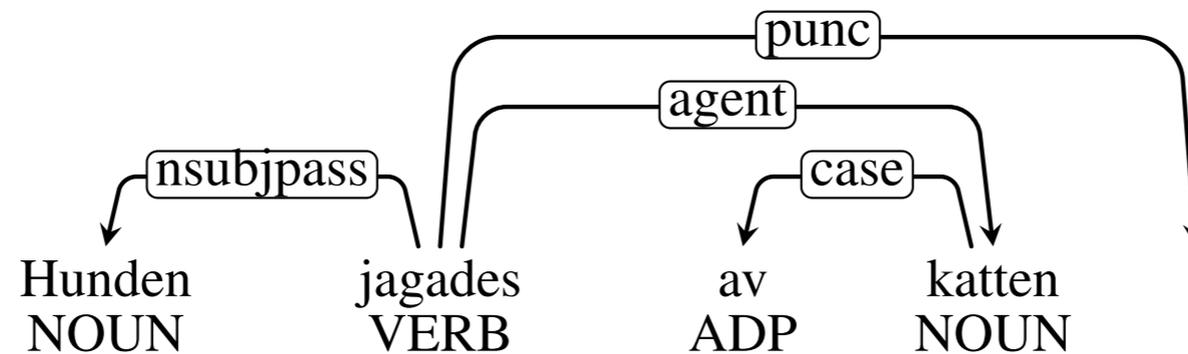
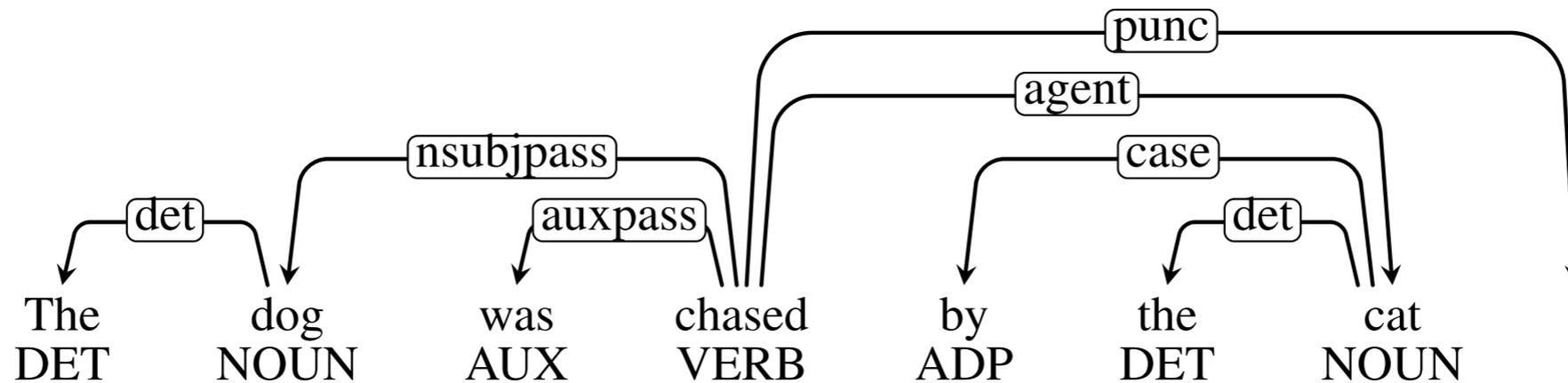
Maximize parallelism

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

But don't overdo it

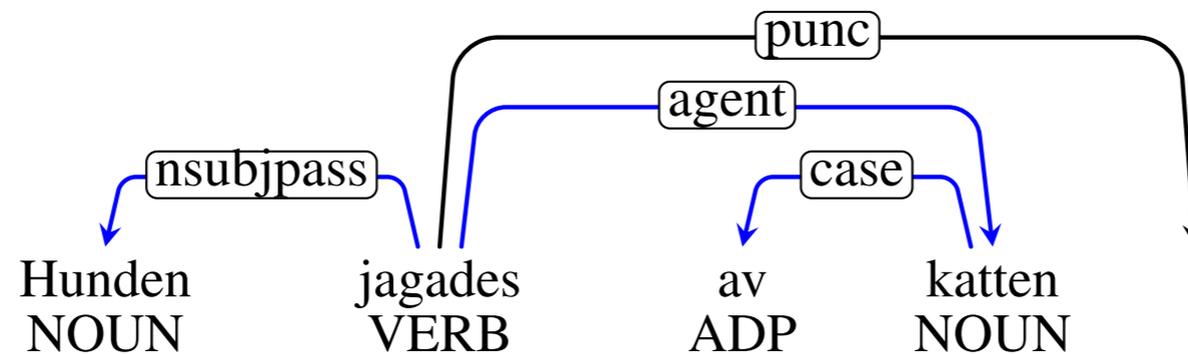
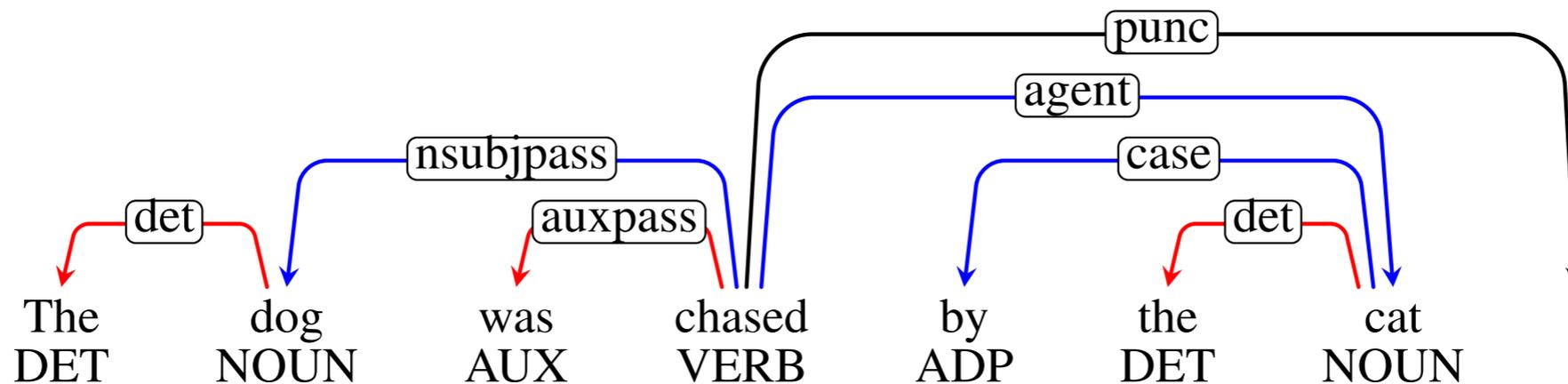
- Don't annotate things that are not there
- Languages select from a universal pool of categories
- Allow language-specific extensions

Dependency Structure



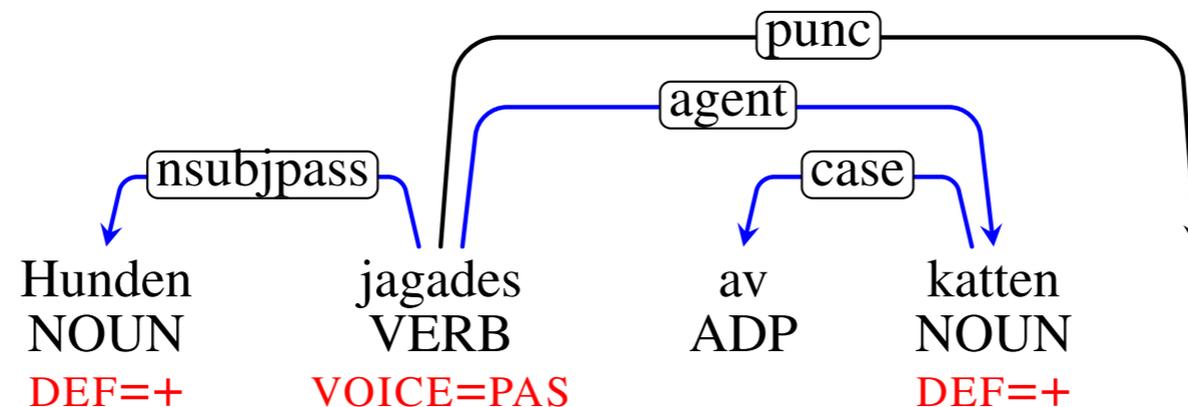
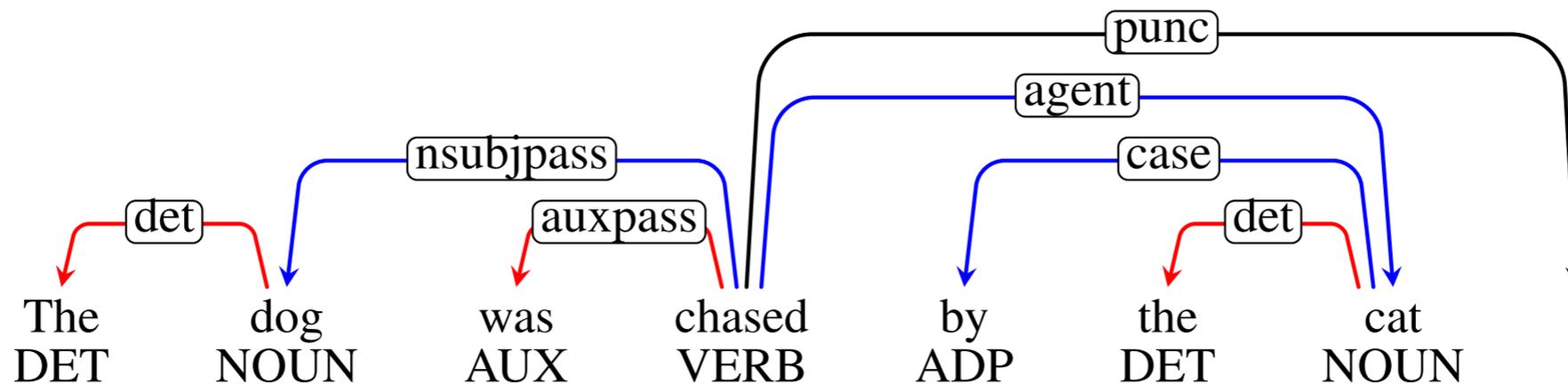
- Keeping content words as heads promotes parallelism
- Function words often correlate with morphology

Dependency Structure



- Keeping content words as heads promotes parallelism
- Function words often correlate with morphology

Dependency Structure



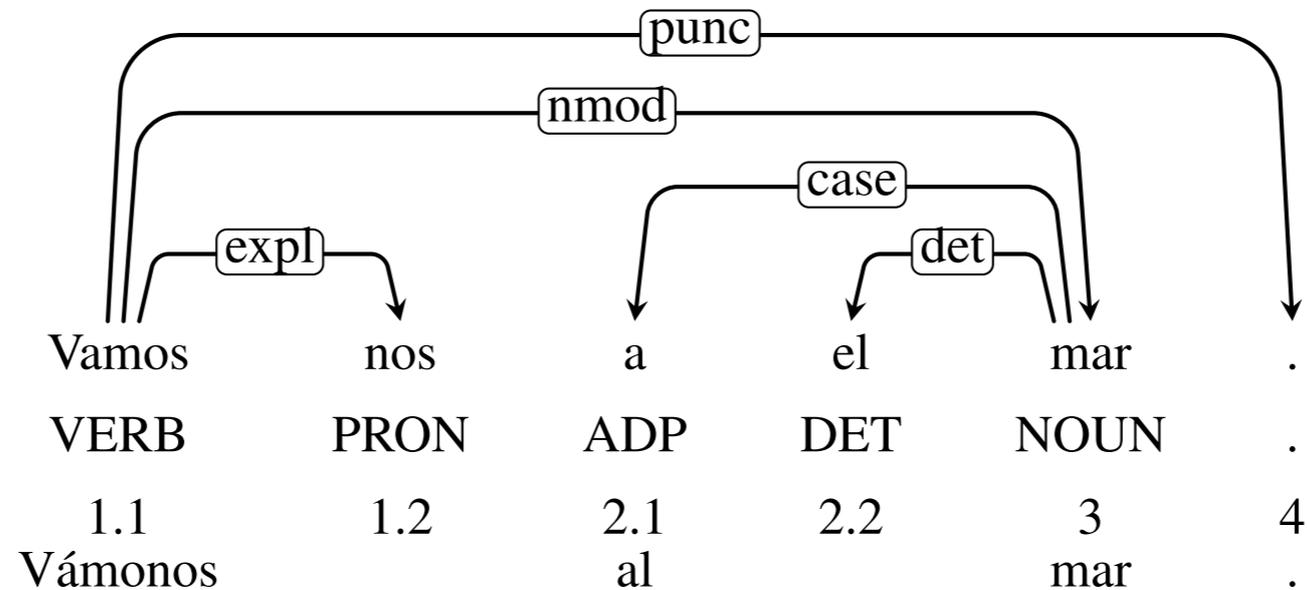
- Keeping content words as heads promotes parallelism
- Function words often correlate with morphology

Morphology

ADJ ADP ADV AUX CONJ DET
INTJ NOUN NUM PNOUN PRON
PRT PUNCT SCONJ SYM VERB X

- The lexicalist hypothesis
 - Grammatical relations hold between words (including clitics)
 - Morphological categories are properties of words
- Morphological annotation
 - Revised Google Universal Part-of-Speech Tags (Petrov et al., 2012)
 - Universal inventory of morphological features (under construction)

Tokens and Words

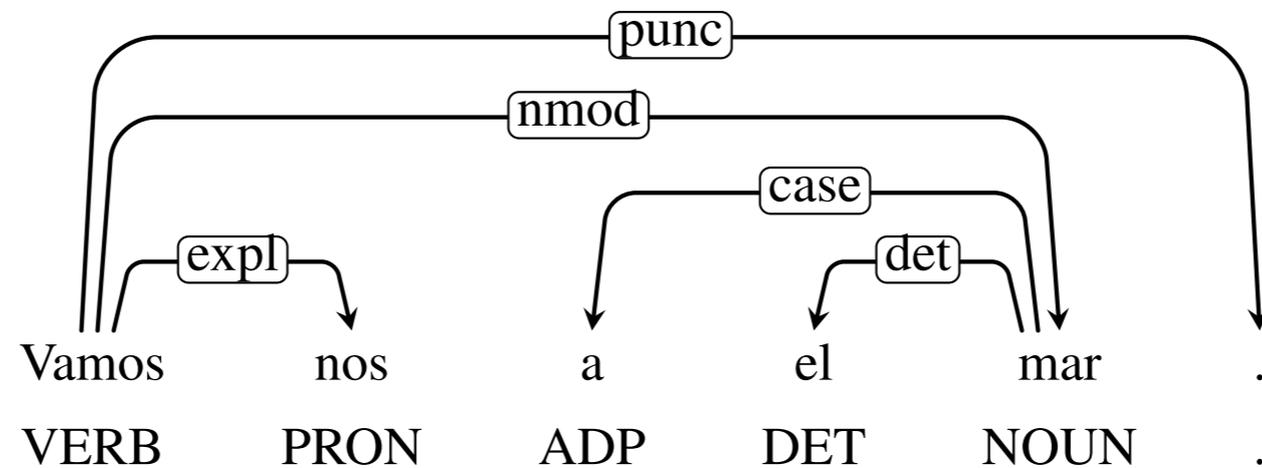


Principle of recoverability

- Clitics and contractions are split to allow meaningful annotation
- Mapping from basic tokenization is explicitly represented
- Heuristic mapping of annotation to basic tokens is provided

Encoded in revised CoNLL format (CoNLL-U)

Tokens and Words

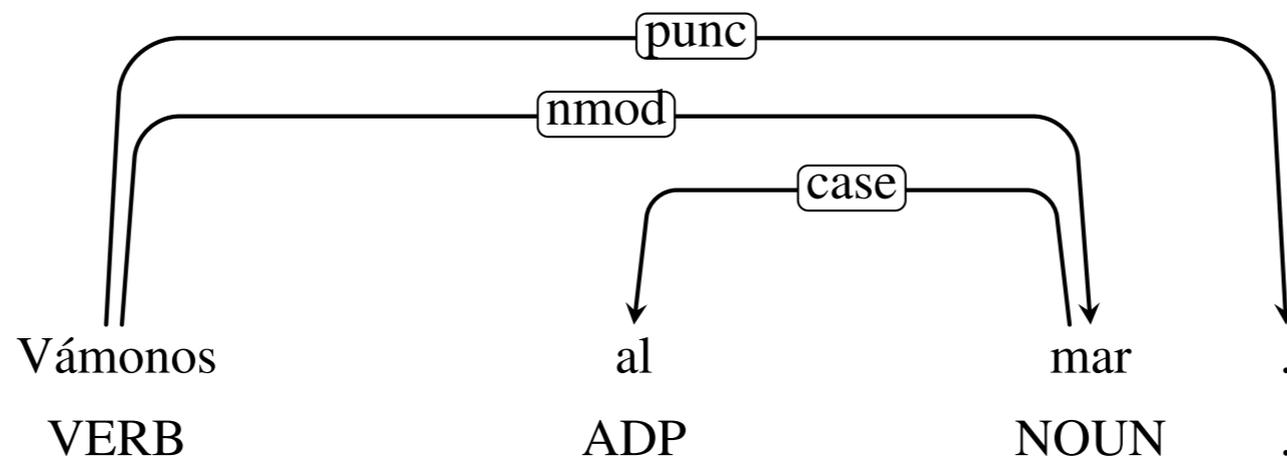


Principle of recoverability

- Clitics and contractions are split to allow meaningful annotation
- Mapping from basic tokenization is explicitly represented
- Heuristic mapping of annotation to basic tokens is provided

Encoded in revised CoNLL format (CoNLL-U)

Tokens and Words

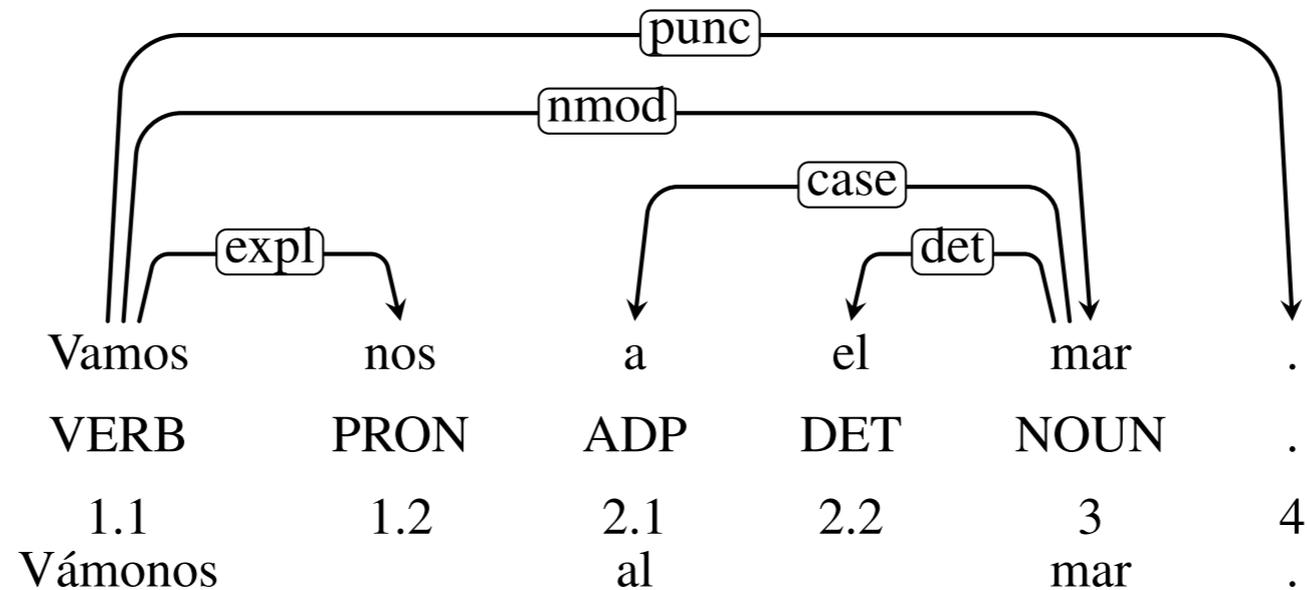


Principle of recoverability

- Clitics and contractions are split to allow meaningful annotation
- Mapping from basic tokenization is explicitly represented
- Heuristic mapping of annotation to basic tokens is provided

Encoded in revised CoNLL format (CoNLL-U)

Tokens and Words



Principle of recoverability

- Clitics and contractions are split to allow meaningful annotation
- Mapping from basic tokenization is explicitly represented
- Heuristic mapping of annotation to basic tokens is provided

Encoded in revised CoNLL format (CoNLL-U)

Work in Progress

Current time plan:

- Stable annotation guidelines by end of September
- First release of data sets before the end of 2014

Follow our progress and give feedback:

- Universal Dependencies: <http://universaldependencies.github.io/docs/>

Check out old releases:

- Uni-Dep-TB: <https://code.google.com/p/uni-dep-tb/>

Let's work together!

Work in Progress

Current time plan:

- Stable annotation guidelines by end of September
- First release of Uni-Dep by the end of 2014

Follow our progress

- Universal Dependencies: <http://universaldependencies.github.io/docs/>

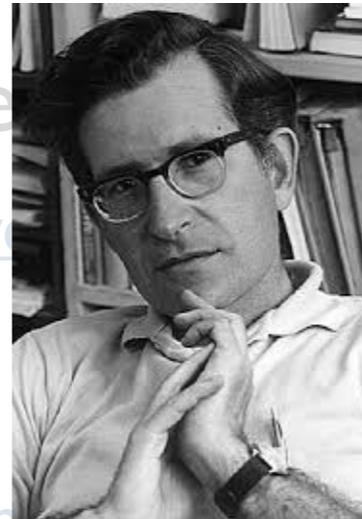
Check out old releases

- Uni-Dep-TB: <http://universaldependencies.github.io/p/uni-dep-tb/>

Let's work together!

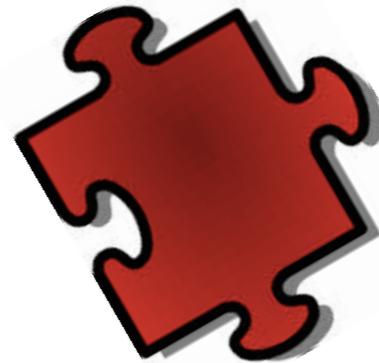
Hey, wait
a minute!

How is this
going to improve
parsing?



Part 3

Parsing with Universal Dependencies



Confessions of a converted dependency parser

Taking Stock

Taking Stock

Motivation for universal dependencies

- Improve comparability of parsing results across languages
- Facilitate development of multilingual systems
- Enable typological studies of syntactic structure

Taking Stock

Motivation for universal dependencies

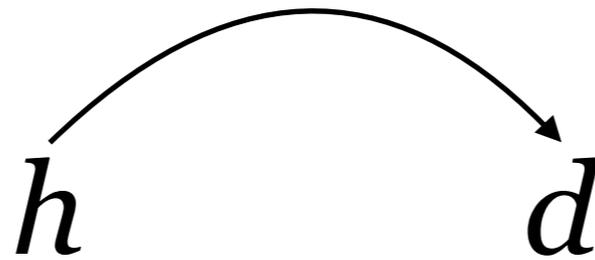
- Improve comparability of parsing results across languages
- Facilitate development of multilingual systems
- Enable typological studies of syntactic structure

What about parsing?

- Not likely to improve parsing accuracy with existing parsers
- Parsers tend to prefer function words as heads (Schwarz et al., 2012)
- We risk bringing English down to 80% instead of Finnish up to 90%

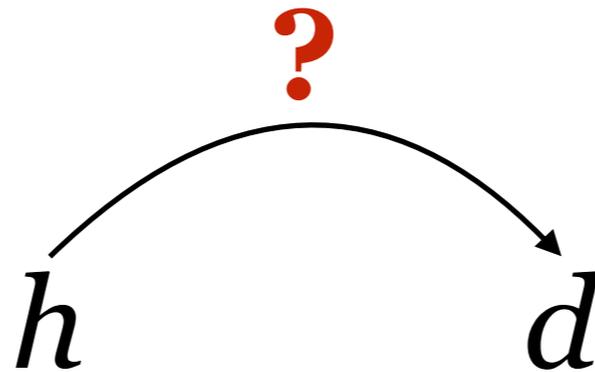
What's the Problem?

What's the Problem?



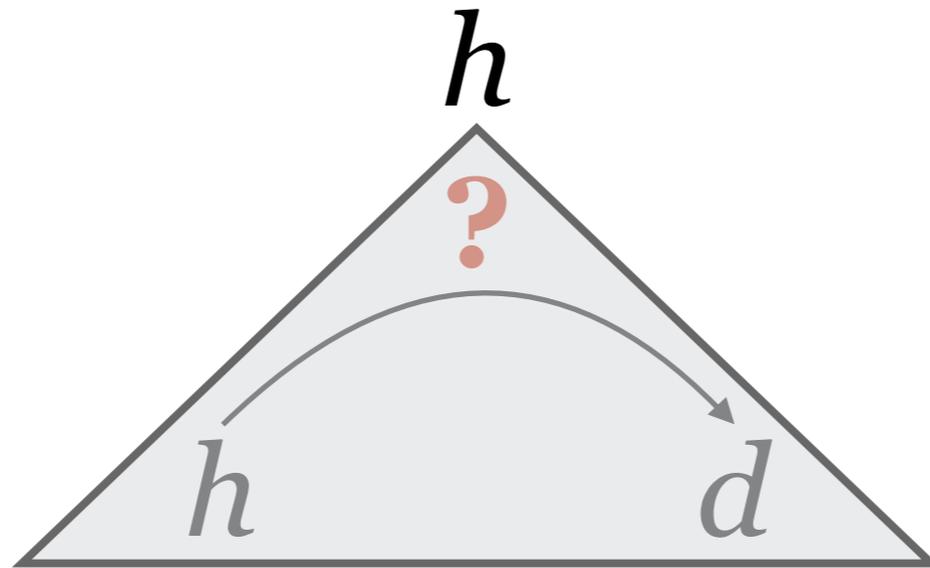
- Dependency parsers know only one syntactic relation

What's the Problem?



- Dependency parsers know only one syntactic relation
- They do not interpret dependency labels

What's the Problem?



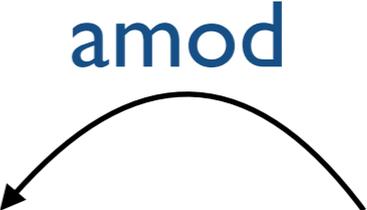
- Dependency parsers know only one syntactic relation
- They do not interpret dependency labels
- They represent a construction primarily by its head

A Simple Case

The dog was chased by the *black cat*.

A Simple Case

The dog was chased by the *black* *cat* .

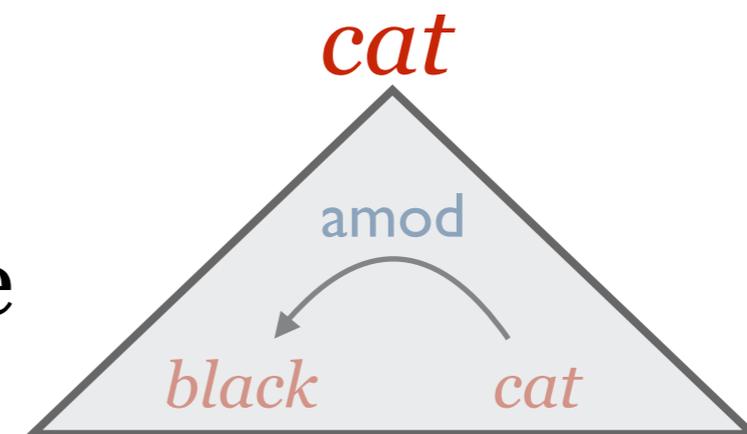


The diagram illustrates a dependency arc between the words 'black' and 'cat'. A curved arrow labeled 'amod' (adpositional modifier) originates from the word 'cat' and points to the word 'black', indicating that 'black' is a modifier of 'cat'.

- All criteria point to *cat* being the head

A Simple Case

The dog was chased by the



- All criteria point to *cat* being the head
- Little (syntactic) information is lost by dropping *black*

A Tricky Case

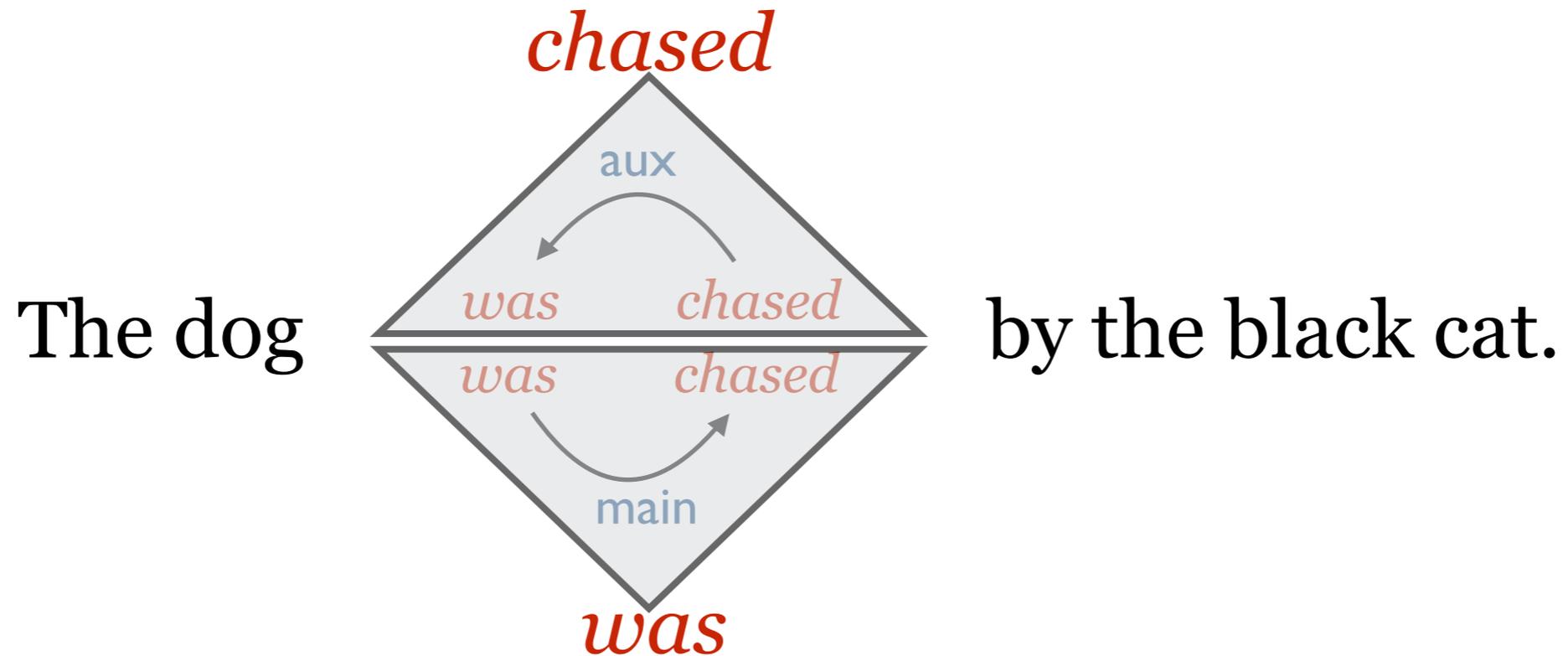
The dog *was chased* by the black cat.

A Tricky Case

The dog *was* *chased* by the black cat.

- Some criteria point to *was*, others to *chased* as the head

A Tricky Case

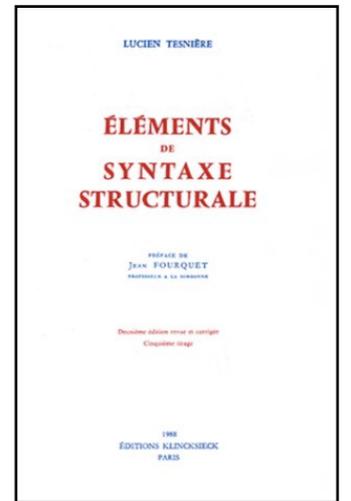
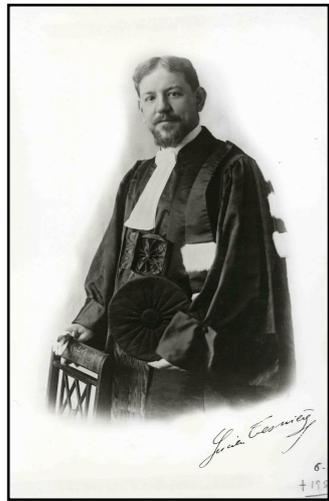


- Some criteria point to *was*, others to *chased* as the head
- Neither word can represent the whole

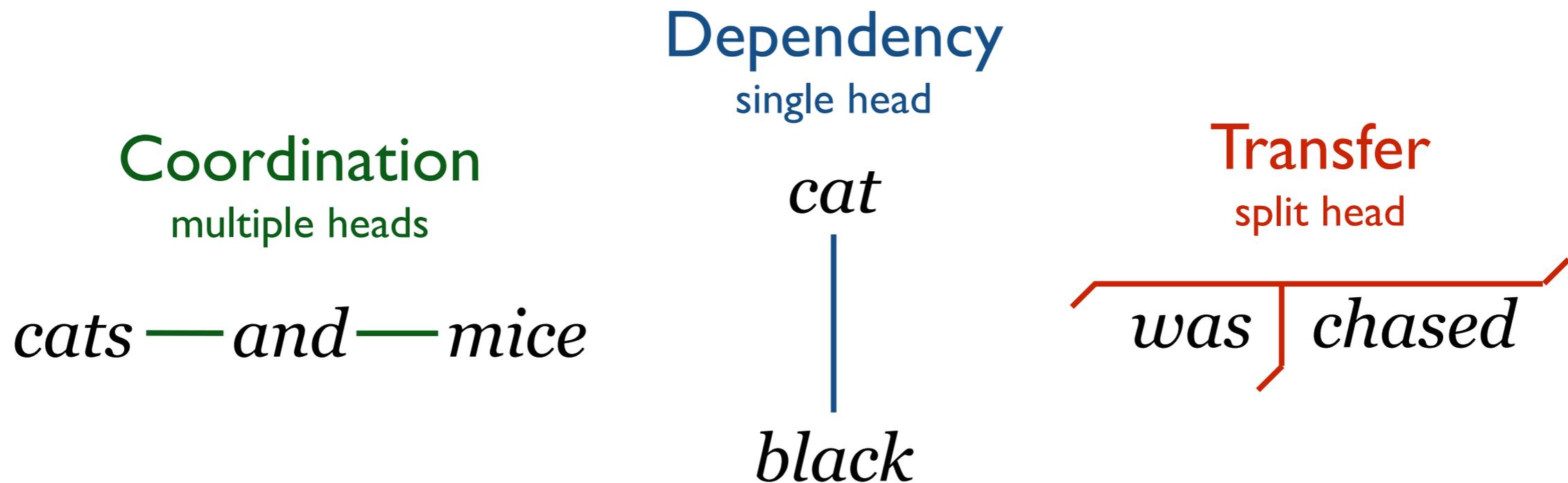
A Tricky Case

The dog was chased *by* the black *cat*.

- Some criteria point to *was*, others to *chased* as the head
- Neither word can represent the whole
- The same problem arises with *by* (the black *cat*)



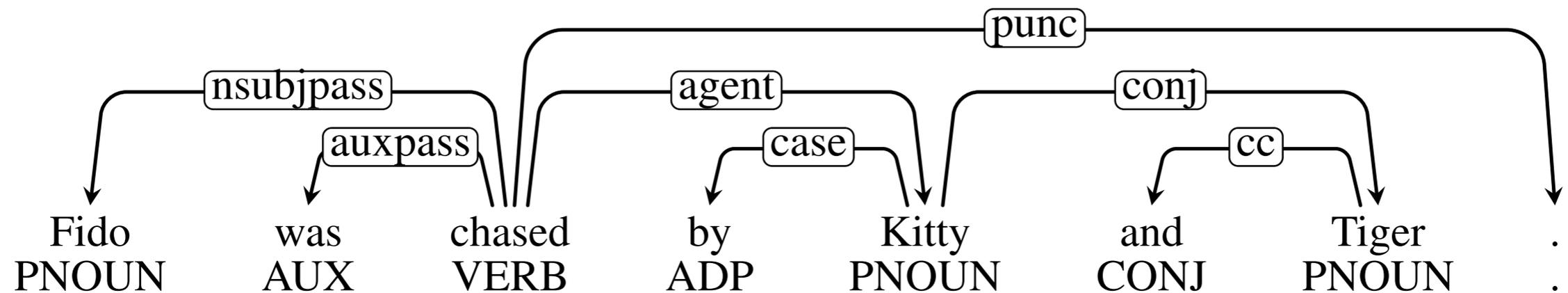
Back to the Roots



- Three fundamental syntactic relations (Tesnière, 1959)
- Tesnière-style dependency treebank (Sangati and Mazza, 2009)

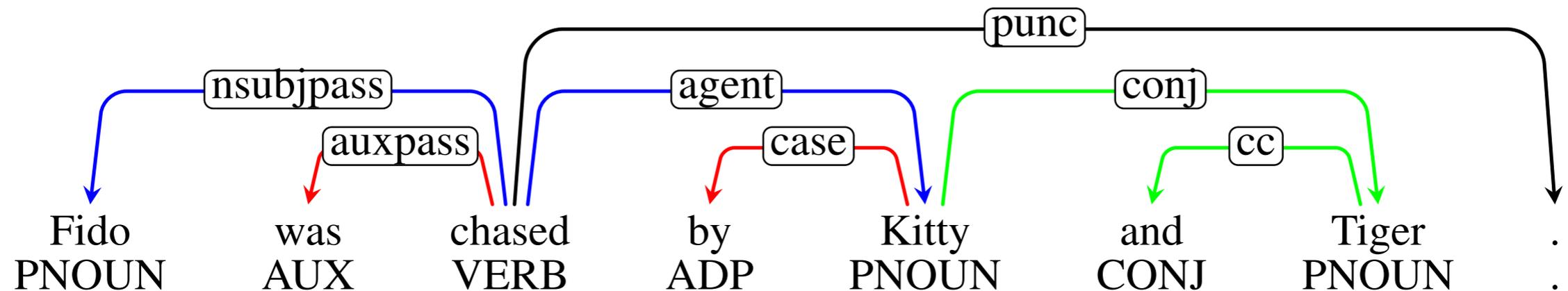
Universal Dependencies

Universal Dependencies



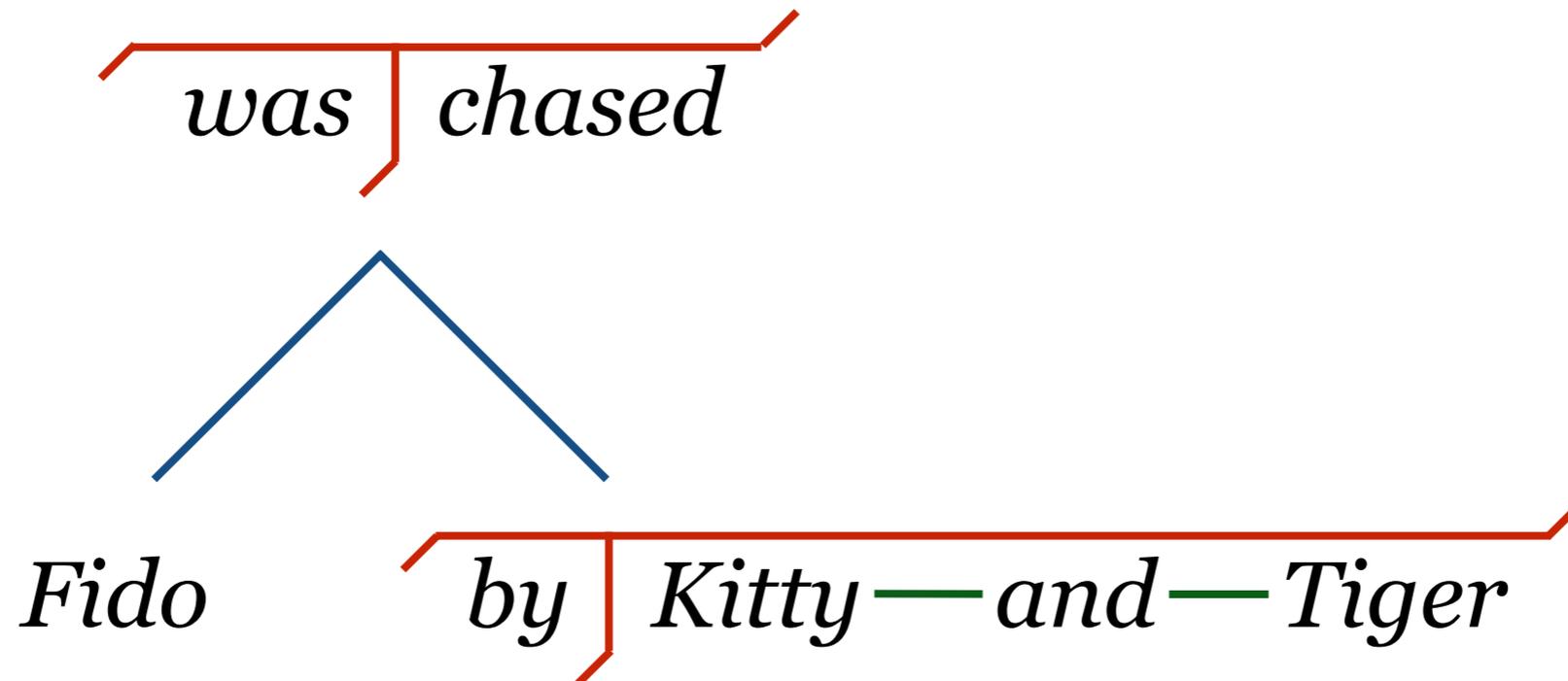
- The UD representation is a simple dependency tree

Universal Dependencies



- The UD representation is a simple dependency tree
- But labels are universal and can be interpreted

Universal Dependencies



- The UD representation is a simple dependency tree
- But labels are universal and can be interpreted
- Therefore, we can put more knowledge into the parser

Final Remarks

Final Remarks

Constituency parsing – largely driven by PTB

- Perhaps too much emphasis on English (until recently)
- But deep analysis of categories and representations

Final Remarks

Constituency parsing – largely driven by PTB

- Perhaps too much emphasis on English (until recently)
- But deep analysis of categories and representations

Dependency parsing – largely driven by CoNLL data

- More attention to typological diversity
- But ignorance and blind faith with respect to representations

Final Remarks

Constituency parsing – largely driven by PTB

- Perhaps too much emphasis on English (until recently)
- But deep analysis of categories and representations

Dependency parsing – largely driven by CoNLL data

- More attention to typological diversity
- But ignorance and blind faith with respect to representations

Can UD give us the best of both worlds?



UPPSALA
UNIVERSITET

Thanks for Your Attention!

Questions?



<http://stp.lingfil.uu.se/~nivre/>