

Out-of-Domain Evaluation of Finnish Dependency Parsing

Jenna Kanerva and Filip Ginter

TurkuNLP, Department of Computing
University of Turku, Finland

Language Resources and Evaluation Conference (LREC) 2022



/ Motivation

- The state-of-the-art is approaching human performance
 - ...when tested on data set aside from the training corpus
 - ...test dataset shares the same properties
- May not be the case in many real world applications
 - The data may very substantially differ from the characteristics of the training data
- What happens with domains not seen during the training?

/ Objective of the work

- New out-of-domain corpus for Finnish dependency parsing
 - New text domain absent from the previous treebanks, some differ very substantially from any text in the training data
- several Finnish out-of-domain parsing experiment
 - Test on the new out-of-domain data
 - Section-level metadata in order to carry out section level performance evaluation using the existing data

UD Finnish-OOD



TURKUNLP
.ORG



UNIVERSITY
OF TURKU

/ UD Finnish-OOD

- Full manual annotation
 - Segmentation, POS, morphology, lemmatization, dependency relations
 - Universal Dependencies
- Data sampled from 5 text different sources

/ Medical – clinical nursing narratives

- The patient's visit in the hospital is recorded in a free text narrative
 - Intensive care unit patients, anonymized
 - Domain terminology, abbreviations, frequent misspellings, telegraphic writing style and non-standard syntactic structures (Laippala et al. 2014)
- 2 documents (939 sentences)

/ Medical – clinical nursing narratives

- The patient's visit in the narrative
 - Intensive care unit p
 - Frequent misspelling telegraphic writing s (Laippala et al. 2014)
- 2 documents (939 se

Example:

Breathing: in the beginning vm 40%, co2 rising, switched to 28% with saturation falling ad 84 and co2 still rising, nasal cannula with saturation 90-91. Breathing shallow and raspy. Hemodynamics: BP high, occasionally > 200 syst. SR, rare arrhythmias. Diuresis ok. Sodium has started to correct with saline. K also corrected with K cl supplementation.



/ Poetry – poems and song lyrics

- Web documents manually identified to include poems or song lyrics
 - FinCORE corpus (Laippala et al., 2019) – a random sample of Finnish web crawled documents manually labeled for their text register (lyrical label – poems and song lyrics)
- 6 documents (144 sentences)

/ Social — tweets

- Finnish tweets downloaded using the Twitter streaming API
 - First filtered using Twitter's language recognition, later manually curated to remove remaining non-Finnish tweets
- 130 randomly sampled tweets (192 sentences)

/ Social – discussion forum messages

- Social network messages from the Suomi24 online discussion forum
 - Historically, one of the largest social network forums in Finland with a broad range of topics
- Random sample of 51 documents (263 sentences)

/ Web – sample of the Internet crawl

- A random sample of web crawl documents
 - Finnish Internet Parsebank (Luotolahti et al., 2015)
 - Post-processing: machine translated removed, each document truncated after 25 sentences, unnatural repetition removed (e.g. repeating quotations blocks)
- 30 documents (584 sentences)

/ Data sources - summary

- Genres: medical, poetry, social and web
 - (1) clinical nursing narratives of hospital patients (medical)
 - (2) poems or song lyrics (poetry)
 - (3) discussion forum messages (social)
 - (4) tweets (social)
 - (5) randomly sampled web crawl documents

Compare: Wikipedia articles, online fiction, JRC-Acquis legislation, online blogs, EuroParl speeches, grammar examples, Wikinews, university news, economy news, online newswire, student magazine articles

/ Corpus statistics

- A single annotator with a long-term experience in Finnish UD treebanking
- Released as test set only

	Doc.	Test Sent.	Words
Finnish-OOD			
Web documents	30	584	6,906
Clinical	2	939	5,330
Online discussions	51	263	3,071
Tweets	130	192	2,070
Poetry	6	144	2,005
Total	218	2,122	19,382

Section	Train			Dev			Test		
	Doc.	Sent.	Words	Doc.	Sent.	Words	Doc.	Sent.	Words
Finnish-TDT									
Wikipedia	160	1,799	25,109	20	200	2,890	20	270	3,936
Fiction	51	2,202	26,342	7	221	2,785	7	316	3,732
Legal	23	914	19,130	3	85	1,938	3	142	2,892
Blogs	61	1,356	16,773	8	259	3,348	8	166	2,219
EuroParl	64	872	16,298	8	94	1,674	8	116	1,986
Grammar examples	—	1,601	13,608	—	200	1,623	—	201	1,771
Wikinews	80	921	11,953	10	92	1,086	10	107	1,256
University news	40	765	10,644	5	86	1,342	5	91	1,243
Economy news	40	854	10,499	5	63	821	5	85	1,136
Student magazines	19	933	12,668	2	64	823	2	61	928
Total	—	12,217	163,024	—	1,364	18,330	—	1,555	21,099
Finnish-PUD									
Wikipedia							251	625	9,901
News							146	375	5,916
Total							397	1,000	15,817
Finnish-OOD									
Web documents							30	584	6,906
Clinical							2	939	5,330
Online discussions							51	263	3,071
Tweets							130	192	2,070
Poetry							6	144	2,005
Total							218	2,122	19,382



Out-of-domain parsing



TURKUNLP
.ORG



UNIVERSITY
OF TURKU

/ Out-of-domain parsing

- Turku neural parser pipeline including FinBERT pre-trained language model
 - Virtanen et al. (2019) reported state-of-the-art performance
- 1. Treebank-level results
- 2. Section-level results
- 3. In-domain vs. out-of-domain parser comparison

/ Out-of-domain parsing – treebank-level

- Train: TDT
 - Wikipedia, fiction, legislation, blogs, EuroParl, grammar examples, news, magazine articles
- Test: TDT, PUD and OOD

/ Out-of-domain parsing – treebank-level

- Train: TDT
 - Wikipedia, fiction, legislation, blogs, EuroParl, grammar examples, news, magazine articles
- Test: TDT, PUD and OOD

Treebank	Tokens	Sent.	Words	UPOS	UFeats	Lemmas	UAS	LAS
TDT	99.6	87.2	99.6	97.9	96.7	95.8	93.0	91.0
PUD	99.6	91.3	99.6	98.0	97.1	95.3	94.0	92.1
OOD	97.6	65.5	97.5	92.5	91.9	91.1	81.6	77.5



/ Out-of-domain parsing – treebank-level

- Train: TDT
 - Wikipedia, fiction, legislation, blogs, Eu examples, news, magazine articles
- Test: TDT, PUD and OOD

1. No drop
2. Not a real “out-of-domain” – wikipedia and news
3. Other studies also found PUD to be easier

Treebank	Tokens	Sent.	Words	UPOS	UFeats	Lemmas	UAS	LAS
TDT	99.6	87.2	99.6	97.9	96.7	95.8	93.0	91.0
PUD	99.6	91.3	99.6	98.0	97.1	95.3	94.0	92.1
OOD	97.6	65.5	97.5	92.5	91.9	91.1	81.6	77.5

/ Out-of-domain parsing – treebank-level

- Train: TDT
 - Wikipedia, fiction, legislation, blogs, EuroParl, grammar examples, news, magazine articles
- Test: TDT, PUD and OOD

Treebank	Tokens	Sent.	Words	UPOS	UFeats	Lemmas	UAS	LAS
TDT	99.6	87.2	99.6	97.9	96.7	95.8	93.0	91.0
PUD	99.6	91.3	99.6	98.0	97.1	95.3	94.0	92.1
OOD	97.6	65.5	97.5	92.5	91.9	91.1	81.6	77.5

1. Clear decrease in performance



/ Out-of-domain parsing – treebank-level

- Train: TDT
 - Wikipedia, fiction, legislation, blogs, EuroParl, grammar examples, news, magazine articles
- Test: TDT, PUD and OOD

Treebank	Tokens	Sent.	Words	UPOS	UFeats	Lemmas	UAS	LAS
TDT	99.6	87.2	99.6	97.9	96.7	95.8	93.0	91.0
PUD	99.6	91.3	99.6	98.0	97.1	95.3	94.0	92.1
OOD	97.6	65.5	97.5	92.5	91.9	91.1	81.6	77.5

/ Out-of-domain parsing

- 1. Treebank-level results
- 2. Section-level results
- 3. In-domain vs. out-of-domain parser comparison

/ Out-of-domain parsing – section-level

- Train: TDT (leave-section-out)
- Test: TDT, PUD and OOD sections
 - All of the results are out-of-domain
 - Indicates how difficult and distinct the sections are in general

/ Out-of-domain parsing – section-level

- Train: TDT (leave-section-out)
- Test: TDT, PUD and OOD sections

Source	Domain	Tokens	Sentences	Words	UPOS	UFeats	Lemmas	UAS	LAS
TDT	University news	99.9	81.3	99.9	98.5	98.1	94.1	95.6	93.8
TDT	Student magazines	100.0	100.0	100.0	98.1	96.9	96.3	95.8	93.2
PUD	News	99.8	94.1	99.8	98.1	96.4	95.8	94.0	92.0
TDT	EuroParl	99.9	94.9	99.9	98.5	98.0	98.3	93.7	91.9
PUD	Wikipedia	99.5	87.9	99.5	97.6	96.8	94.9	93.2	91.2
TDT	Economy news	99.9	75.8	99.9	98.1	97.8	97.3	91.7	89.8
TDT	Wikipedia	99.1	89.2	99.1	96.8	96.2	93.4	91.8	89.5
TDT	Wikinews	99.6	81.0	99.6	98.6	96.3	92.6	91.5	89.5
TDT	Blogs	98.9	83.0	98.9	96.8	94.6	94.5	91.5	89.4
TDT	Fiction	99.7	93.2	99.6	96.5	94.3	93.3	90.8	88.4
OOD	Web documents	99.3	80.3	99.3	96.3	95.2	94.3	89.1	86.4
TDT	Legal	99.0	45.2	99.0	97.2	95.8	95.5	88.1	85.9
TDT	Grammar examples	99.8	71.4	99.8	96.2	93.9	94.8	88.5	85.7
OOD	Online discussions	98.1	86.2	98.1	94.0	93.8	93.0	87.9	83.9
OOD	Poetry	99.6	55.5	99.6	95.2	94.7	94.6	80.3	76.1
OOD	Tweets	92.2	57.8	92.1	83.5	82.6	81.4	73.9	69.5
OOD	Clinical	96.4	53.1	96.4	89.2	89.2	88.5	72.0	66.1
AVERAGE		98.9	78.2	98.8	95.8	94.7	93.7	88.8	86.0

/ Out-of-domain parsing – section-level

- Train: TDT (leave-section-out)
- Test: TDT, PUD and OOD sections

Source	Domain	Tokens	Sentences	Words	UPOS	UFeats	Lemmas	UAS	LAS
TDT	University news	99.9	81.3	99.9	98.5	98.1	94.1	95.6	93.8
TDT	Student magazines	100.0	100.0	100.0	98.1	96.9	96.3	95.8	93.2
PUD	News	99.8	94.1	99.8	98.1	96.4	95.8	94.0	92.0
TDT	EuroParl	99.9	94.9	99.9	98.5	98.0	98.3	93.7	91.9
PUD	Wikipedia	99.5	87.9	99.5	97.6	96.8	94.9	93.2	91.2
TDT	Economy news	99.9	75.8	99.9	98.1	97.8	97.3	91.7	89.8
TDT	Wikipedia	99.1	89.2	99.1	96.8	96.2	93.4	91.8	89.5
TDT	Wikinews	99.6	81.0	99.6	98.6	96.3	92.6	91.5	89.5
TDT	Blogs	98.9	83.0	98.9	96.8	94.6	94.5	91.5	89.4
TDT	Fiction	99.7	93.2	99.6	96.5	94.3	93.3	90.8	88.4
OOD	Web documents	99.3	80.3	99.3	96.3	95.2	94.3	89.1	86.4
TDT	Legal	99.0	45.2	99.0	97.2	95.8	95.5	88.1	85.9
TDT	Grammar examples	99.8	71.4	99.8	96.2	93.9	94.8	88.5	85.7
OOD	Online discussions	98.1	86.2	98.1	94.0	93.8	93.0	87.9	83.9
OOD	Poetry	99.6	55.5	99.6	95.2	94.7	94.6	80.3	76.1
OOD	Tweets	92.2	57.8	92.1	83.5	82.6	81.4	73.9	69.5
OOD	Clinical	96.4	53.1	96.4	89.2	89.2	88.5	72.0	66.1
AVERAGE		98.9	78.2	98.8	95.8	94.7	93.7	88.8	86.0

/ Out-of-domain parsing – section-level

- Train: TDT (leave-section-out)
- Test: TDT, PUD and OOD sections

Source	Domain	Tokens	Sentences	Words	UPOS	UFeats	Lemmas	UAS	LAS
TDT	University news	99.9	81.3	99.9	98.5	98.1	94.1	95.6	93.8
TDT	Student magazines	100.0	100.0	100.0	98.1	96.9	96.3	95.8	93.2
PUD	News	99.8	94.1	99.8	98.1	96.4	95.8	94.0	92.0
TDT	EuroParl	99.9	94.9	99.9	98.5	98.0	98.3	93.7	91.9
PUD	Wikipedia	99.5	87.9	99.5	97.6	96.8	94.9	93.2	91.2
TDT	Economy news	99.9	75.8	99.9	98.1	97.8	97.3	91.7	89.8
TDT	Wikipedia	99.1	89.2	99.1	96.8	96.2	93.4	91.8	89.5
TDT	Wikinews	99.6	81.0	99.6	98.6	96.3	92.6	91.5	89.5
TDT	Blogs	98.9	83.0	98.9	96.8	94.6	94.5	91.5	89.4
TDT	Fiction	99.7	93.2	99.6	96.5	94.3	93.3	90.8	88.4
OOD	Web documents	99.3	80.3	99.3	96.3	95.2	94.3	89.1	86.4
TDT	Legal	99.0	45.2	99.0	97.2	95.8	95.5	88.1	85.9
TDT	Grammar examples	99.8	71.4	99.8	96.2	93.9	94.8	88.5	85.7
OOD	Online discussions	98.1	86.2	98.1	94.0	93.8	93.0	87.9	83.9
OOD	Poetry	99.6	55.5	99.6	95.2	94.7	94.6	80.3	76.1
OOD	Tweets	92.2	57.8	92.1	83.5	82.6	81.4	73.9	69.5
OOD	Clinical	96.4	53.1	96.4	89.2	89.2	88.5	72.0	66.1
AVERAGE		98.9	78.2	98.8	95.8	94.7	93.7	88.8	86.0

/ Out-of-domain parsing

- 1. Treebank-level results
- 2. Section-level results
- 3. In-domain vs. out-of-domain parser comparison

/ In-domain vs. out-of-domain

- Train: TDT (full) / TDT (leave-section-out)
- Test: TDT sections
 - Indicates the out-of-domain effect

/ In-domain vs. out-of-domain

	F1	Tokens	Sentences	Words	UPOS	UFeats	Lemmas	UAS	LAS
Student magazines	Δ	0.0	0.0	0.0	-0.6	-0.2	-0.1	0.8	0.5
	OOD	100.0	100.0	100.0	98.1	96.9	96.3	95.8	93.2
	ID	100.0	100.0	100.0	98.7	97.1	96.4	95.0	92.7
Blogs	Δ	-0.8	-1.2	-0.7	-1.1	-1.4	-1.8	0.0	0.1
	OOD	98.9	83.0	98.9	96.8	94.6	94.5	91.5	89.4
	ID	99.7	84.2	99.7	97.9	96.1	96.4	91.5	89.2
University news	Δ	-0.1	-6.3	-0.1	0.0	0.0	-2.5	0.1	0.1
	OOD	99.9	81.3	99.9	98.5	98.1	94.1	95.6	93.8
	ID	100.0	87.6	100.0	98.5	98.2	96.5	95.5	93.7
EuroParl	Δ	-0.1	0.0	-0.1	-0.1	0.0	0.1	0.1	-0.6
	OOD	99.9	94.9	99.9	98.5	98.0	98.3	93.7	91.9
	ID	99.9	94.9	99.9	98.6	98.0	98.2	93.6	92.5
Fiction	Δ	0.1	2.3	0.1	-0.8	-1.2	-1.9	-0.8	-1.0
	OOD	99.7	93.2	99.6	96.5	94.3	93.3	90.8	88.4
	ID	99.6	90.9	99.5	97.3	95.5	95.2	91.6	89.4
Economy news	Δ	0.0	-0.6	0.0	-0.2	1.0	-0.4	-0.7	-1.2
	OOD	99.9	75.8	99.9	98.1	97.8	97.3	91.7	89.8
	ID	99.9	76.4	99.9	98.3	96.8	97.8	92.4	91.0
Wikipedia	Δ	-0.1	-5.2	-0.1	-0.4	-0.5	0.2	-1.1	-1.2
	OOD	99.1	89.2	99.1	96.8	96.2	93.4	91.8	89.5
	ID	99.2	94.4	99.2	97.2	96.7	93.2	92.9	90.7
Wikinews	Δ	0.4	-2.9	0.4	0.5	-0.4	-0.6	-1.0	-1.2
	OOD	99.6	81.0	99.6	98.6	96.3	92.6	91.5	89.5
	ID	99.2	83.9	99.2	98.1	96.7	93.2	92.5	90.7
Grammar examples	Δ	-0.3	-11.2	-0.3	-0.8	-1.4	-1.6	-3.3	-3.5
	OOD	99.8	71.4	99.8	96.2	93.9	94.8	88.5	85.7
	ID	100.0	82.5	100.0	97.0	95.3	96.4	91.9	89.2
Legal	Δ	-0.7	-27.4	-0.7	-1.9	-2.5	-1.8	-6.8	-7.5
	OOD	99.0	45.2	99.0	97.2	95.8	95.5	88.1	85.9
	ID	99.7	72.6	99.7	99.1	98.3	97.3	94.9	93.4

/ In-domain vs. out-of-domain

	F1	Tokens	Sentences	Words	UPOS	UFeats	Lemmas	POS	LAS
Student magazines	Δ	0.0	0.0	0.0	-0.6	-0.2	-0.1	0.8	0.5
	OOD	100.0	100.0	100.0	98.1	96.9	96.3	95.8	93.2
	ID	100.0	100.0	100.0	98.7	97.1	96.4	95.0	92.7
Blogs	Δ	-0.8	-1.2	-0.7	-1.1	-1.4	-1.8	0.0	0.1
	OOD	98.9	83.0	98.9	96.8	94.6	94.5	91.5	89.4
	ID	99.7	84.2	99.7	97.9	96.1	96.4	91.5	89.2
University news	Δ	-0.1	-6.3	-0.1	0.0	0.0	-2.5	0.1	0.1
	OOD	99.9	81.3	99.9	98.5	98.1	94.1	95.6	93.8
	ID	100.0	87.6	100.0	98.5	98.2	96.5	95.5	93.7
EuroParl	Δ	-0.1	0.0	-0.1	-0.1	0.0	0.1	0.1	-0.6
	OOD	99.9	94.9	99.9	98.5	98.0	98.3	93.7	91.9
	ID	99.9	94.9	99.9	98.6	98.0	98.2	93.6	92.5
Fiction	Δ	0.1	2.3	0.1	-0.6	-1.2	-1.3	-0.6	-1.0
	OOD	99.7	93.2	99.6	96.5	94.3	93.3	90.8	88.4
	ID	99.6	90.9	99.5	97.3	95.5	95.2	91.6	89.4
Economy news	Δ	0.0	-0.6	0.0	-0.2	1.0	-0.4	-0.7	-1.2
	OOD	99.9	75.8	99.9	98.1	97.8	97.3	91.7	89.8
	ID	99.9	76.4	99.9	98.3	96.8	97.8	92.4	91.0
Wikipedia	Δ	-0.1	-5.2	-0.1	-0.4	-0.5	0.2	-1.1	-1.2
	OOD	99.1	89.2	99.1	96.8	96.2	93.4	91.8	89.5
	ID	99.2	94.4	99.2	97.2	96.7	93.2	92.9	90.7
Wikinews	Δ	0.4	-2.9	0.4	0.5	-0.4	-0.6	-1.0	-1.2
	OOD	99.6	81.0	99.6	98.6	96.3	92.6	91.5	89.5
	ID	99.2	83.9	99.2	98.1	96.7	93.2	92.5	90.7
Grammar examples	Δ	-0.3	-11.2	-0.3	-0.8	-1.4	-1.6	-3.3	-3.5
	OOD	99.8	71.4	99.8	96.2	93.9	94.8	88.5	85.7
	ID	100.0	82.5	100.0	97.0	95.3	96.4	91.9	89.2
Legal	Δ	-0.7	-27.4	-0.7	-1.9	-2.5	-1.8	-6.8	-7.5
	OOD	99.0	45.2	99.0	97.2	95.8	95.5	88.1	85.9
	ID	99.7	72.6	99.7	99.1	98.3	97.3	94.9	93.4

/ In-domain vs. out-of-domain

	F1	Tokens	Sentences	Words	UPOS	UFeats	Lemmas	UAS	LAS
Student magazines	Δ	0.0	0.0	0.0	-0.6	-0.2	-0.1	0.8	0.5
	OOD	100.0	100.0	100.0	98.1	96.9	96.3	95.8	93.2
	ID	100.0	100.0	100.0	98.7	97.1	96.4	95.0	92.7
Blogs	Δ	-0.8	-1.2	-0.7	-1.1	-1.4	-1.8	0.0	0.1
	OOD	98.9	83.0	98.9	96.8	94.6	94.5	91.5	89.4
	ID	99.7	84.2	99.7	97.9	96.1	96.4	91.5	89.2
University news	Δ	-0.1	-6.3	-0.1	0.0	0.0	-2.5	0.1	0.1
	OOD	99.9	81.3	99.9	98.5	98.1	94.1	95.6	93.8
	ID	100.0	87.6	100.0	98.5	98.2	96.5	95.5	93.7
EuroParl	Δ	-0.1	0.0	-0.1	-0.1	0.0	0.1	0.1	-0.6
	OOD	99.9	94.9	99.9	98.5	98.0	98.3	93.7	91.9
	ID	99.9	94.9	99.9	98.6	98.0	98.2	93.6	92.5
Fiction	Δ	0.1	2.3	0.1	-0.8	-1.2	-1.9	-0.8	-1.0
	OOD	99.7	93.2	99.6	96.5	94.3	93.3	90.8	88.4
	ID	99.6	90.9	99.5	97.3	95.5	95.2	91.6	89.4
Economy news	Δ	0.0	-0.6	0.0	-0.2	1.0	-0.4	-0.7	-1.2
	OOD	99.9	75.8	99.9	98.1	97.8	97.3	91.7	89.8
	ID	99.9	76.4	99.9	98.3	96.8	97.8	92.4	91.0
Wikipedia	Δ	-0.1	-5.2	-0.1	-0.4	-0.5	0.2	-1.1	-1.2
	OOD	99.1	89.2	99.1	96.8	96.2	93.4	91.8	89.5
	ID	99.2	94.4	99.2	97.2	96.7	93.2	92.9	90.7
Wikinews	Δ	0.4	-2.9	0.4	0.5	-0.4	-0.6	-1.0	-1.2
	OOD	99.6	81.0	99.6	98.6	96.3	92.6	91.5	89.5
	ID	99.2	83.9	99.2	98.1	96.7	93.2	92.5	90.7
Grammar examples	Δ	-0.3	-11.2	-0.3	-0.8	-1.4	-1.6	-3.3	-3.5
	OOD	99.8	71.4	99.8	96.2	93.9	94.8	88.5	85.7
	ID	100.0	82.5	100.0	97.0	95.3	96.4	91.9	89.2
Legal	Δ	-0.7	-27.4	-0.7	-1.9	-2.5	-1.8	-6.8	-7.5
	OOD	99.0	45.2	99.0	97.2	95.8	95.5	88.1	85.9
	ID	99.7	72.6	99.7	99.1	98.3	97.3	94.9	93.4

/ In-domain vs. out-of-domain

	F1	Tokens	Sentences	Words	UPOS	UFeats	Lemmas	UAS	LAS
Student magazines	Δ	0.0	0.0	0.0	-0.6	-0.2	-0.1	0.8	0.5
	OOD	100.0	100.0	100.0	98.1	96.9	96.3	95.8	93.2
	ID	100.0	100.0	100.0	98.7	97.1	96.4	95.0	92.7
Blogs	Δ	-0.8	-1.2	-0.7	-1.1	-1.4	-1.8	0.0	0.1
	OOD	98.9	83.0	98.9	96.8	94.6	94.5	91.5	89.4
	ID	99.7	84.2	99.7	97.9	96.1	96.4	91.5	89.2
University news	Δ	-0.1	-6.3	-0.1	0.0	0.0	-2.5	0.1	0.1
	OOD	99.9	81.3	99.9	98.5	98.1	94.1	95.6	93.8
	ID	100.0	87.6	100.0	98.5	98.2	96.5	95.5	93.7
EuroParl	Δ	-0.1	0.0	-0.1	-0.1	0.0	0.1	0.1	-0.6
	OOD	99.9	94.9	99.9	98.5	98.0	98.3	93.7	91.9
	ID	99.9	94.9	99.9	98.6	98.0	98.2	93.6	92.5
Fiction	Δ	0.1	2.3	0.1	-0.8	-1.2	-1.9	-0.8	-1.0
	OOD	99.7	93.2	99.6	96.5	94.3	93.3	90.8	88.4
	ID	99.6	90.9	99.5	97.3	95.5	95.2	91.6	89.4
Economy news	Δ	0.0	-0.6	0.0	-0.2	1.0	-0.4	-0.7	-1.2
	OOD	99.9	75.8	99.9	98.1	97.8	97.3	91.7	89.8
	ID	99.9	76.4	99.9	98.3	96.8	97.8	92.4	91.0
Wikipedia	Δ	-0.1	-5.2	-0.1	-0.4	-0.5	0.2	-1.1	-1.2
	OOD	99.1	89.2	99.1	96.8	96.2	93.4	91.8	89.5
	ID	99.2	94.4	99.2	97.2	96.7	93.2	92.9	90.7
Wikinews	Δ	0.4	-2.9	0.4	0.5	-0.4	-0.6	-1.0	-1.2
	OOD	99.6	81.0	99.6	98.6	96.3	92.6	91.5	89.5
	ID	99.2	83.9	99.2	98.1	96.7	93.2	92.5	90.7
Grammar examples	Δ	-0.3	-11.2	-0.3	-0.8	-1.4	-1.6	-3.3	-3.5
	OOD	99.8	71.4	99.8	96.2	93.9	94.8	88.5	85.7
	ID	100.0	82.5	100.0	97.0	95.3	96.4	91.9	89.2
Legal	Δ	-0.7	-27.4	-0.7	-1.9	-2.5	-1.8	-6.8	-7.5
	OOD	99.0	45.2	99.0	97.2	95.8	95.5	88.1	85.9
	ID	99.7	72.6	99.7	99.1	98.3	97.3	94.9	93.4

/ Conclusions

- New out-of-domain evaluation set for Finnish dependency parsing
- Syntactic parsing performance can substantially degrade on several domains
 - The effect was strongest when measuring LAS
 - Notably less on tagging and lemmatization

/ Thank you!

References:

Virtanen, A., Kanerva, J., Ilo, R., Luoma, J., Luotolahti, J., Salakoski, T., Ginter, F., and Pyysalo, S. (2019). *Multilingual is not enough: BERT for Finnish*. arXiv preprint arXiv:1912.07076.

Laippala, V., Viljanen, T., Airola, A., Kanerva, J., Salanterä, S., Salakoski, T., and Ginter, F. (2014). *Statistical parsing of varieties of clinical Finnish*. Artificial Intelligence in Medicine, Text Mining and Information Analysis of Health Documents.

Laippala, V., Kyllönen, R., Egbert, J., Biber, D., and Pyysalo, S. (2019). *Toward multilingual identification of online registers*. In Proceedings of the 22nd Nordic Conference on Computational Linguistics, Linköping University Electronic Press.

Aller Media Ltd. (2020). *The Suomi24 Corpus 2001-2017*, VRT version 1.1 [text corpus]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:lb-2020021801>

Luotolahti, J., Kanerva, J., Laippala, V., Pyysalo, S., and Ginter, F. (2015). *Towards universal web parsebanks*. In Proceedings of the International Conference on Dependency Linguistics (Depling'15).

