Corpus	Language	Sentences	Tokens	Sentences	Tokens	UAS	LAS	
DDT	Danish	5190	94386	322	5852	88.12%	82.89%	
Talbanken	Swedish	11042	191467	389	5656	87.27%	81.08%	
BIO	English	7482	197724	1450	38120	91.71%	91.47%	
Law	Italian	1001	25423	100	2607	91.45%	77.71%	
Newspaper	Italian	1001	28233	100	2357	85.53%	75.14%	
Law+Newspaper	Italian	2002	53656	200	4964	88.86%	76.59%	
Cast3lb	Spanish	3306	89334	206	5694	82.19%	78.42%	
Tiger	German	39216	699610	357	5694			
Tiger (6K)	German	6000	109035	357	5694	85.34%	81.03%	
Bul	Bulgarian	12823	190217	398	5934			
Bul (6K)	Bulgarian	6000	80905	398	5934	89.91%	85.24%	
SDT	Slovene	1534	28750	402	6390	78.56%	69.59%	
PDT	Czech	72703	1249408	365	5853			
Metu	Turkish	4997	57510	623	7547	79.89%	69.05%	
Alpino	Dutch	13349	195069	386	5585			

65839

386

5585

76.97%

6000

Test set

72.53%

Training set

Dutch

Alpino (6K)

			Experir		D !!		0		•		F	J 12
Corpus	Annotation applied	Focus	Sente	ences	Baseline UAS	LAS	System UAS	LAS	Improve Δ UAS	ement Δ LAS	Error red LA	p-value
DDT DDT	Lemma LemmaOnly	overall overall			88.12% 88.12%	82.89% 82.89%	87.37% 88.72%	82.23% 83.68%	- <i>0.75</i>	- <i>0.67</i> 0.79	0.00% 4.60%	0.0235 0.0117
DDT	Inanimacy	overall affected sentences		0%	88.12%	82.89%	88.31%	82.95%	0.19 0.00	0.05 0.00	0.30% 0.00%	
55.		unaffected sentences overall		0%	88.12%	82.89%	88.35%	83.15%	0.00	0.00	0.00%	0.2098
DDT	Animacy	affected sentences	83		87.27%	81.60%	86.92%	81.50%	-0.35	-0.10	1.50% 0.00%	0.4329
		unaffected sentences overall	239		88.57% 88.12%	83.56% 82.89%	89.08% 88.33%	84.00% 83.13%	0.51	0.44 0.24	2.68% 1.38%	0.1331 0.1715
DDT	Animacy + Inanimacy	affected sentences unaffected sentences		0% 0%					0.00	0.00	0.00% 0.00%	
DDT	QuotedPhrase	overall affected sentences	43	13%	88.12% 86.10%	82.89% 81.17%	88.88% 87.31%	83.65% 82.38%	0.75 1.21	0.75 1.21	4.40% 6.43%	0.0014 0.0123
		unaffected sentences overall	279		88.62% 88.12%	83.32% 82.89%	89.26% 88.52%	83.96% 83.25%	0.64	0.64 0.36	3.84% 2.08%	0.0106
DDT	InnerQuotedPhrase	affected sentences		0%	00.1270	02.0970	86.48%	81.49%	0.00	0.00	0.00%	0.3143
	InnerQuotedPhrase +	unaffected sentences overall		0%	88.12%	82.89%	89.10% 88.57%	83.76% 83.27%	0.00	0.00	0.00% 2.20%	0.1119
DDT	QuotationMark	affected sentences unaffected sentences		0% 0%					0.00	0.00	0.00% 0.00%	
DDT	QuotedPhrase (POS+CPOS)	overall affected sentences		0%	88.12%	82.89%	88.21% 85.58%	82.76% 80.05%	0.09	<i>-0.14</i> 0.00	0.00% 0.00%	0.3473 0.1679
	, , , , , , , , , , , , , , , , , , ,	unaffected sentences overall		0%	88.12%	82.89%	88.86% 88.57%	83.43% 83.13%	0.00 0.44	0.00 0.24	0.00% 1.40%	0.4027 0.2322
DDT	QuotedPhrase (POS)	affected sentences unaffected sentences		0% 0%	00.1270	02.0970	87.13% 88.92%	81.26% 83.60%	0.00	0.00	0.00% 0.00%	0.4567 0.2219
		overall			88.12%	82.89%	88.55%	83.56%	0.43	0.67	3.90%	0.0066
DDT	QuotationMark	affected sentences unaffected sentences	93 229		87.41% 88.54%	82.62% 83.05%	88.01% 88.86%	83.32% 83.70%	0.60 0.32	0.70 0.65	4.03% 3.83%	0.0681 0.0260
DDT	QuotedPhrase + QuotationMark	overall affected sentences	93	29%	88.12% 87.41%	82.89% 82.62%	88.74% 88.01%	83.56% 83.04%	0.62 0.60	0.67 0.42	3.90% 2.42%	0.0125 0.2230
		unaffected sentences overall	229	71%	88.54% 88.12%	83.05% 82.89%	89.16% 88.53%	83.86% 83.19%	0.62 0.41	0.81 0.30	4.78% 1.73%	0.0133 0.1426
DDT	NumericalBullet	affected sentences	6 316		83.78%	76.58%	90.99%	85.59%	7.21	9.01 0.12	38.47% 0.71%	0.0287
	D. II /	unaffected sentences overall			88.21% 88.12%	83.02% 82.89%	88.49% 88.55%	83.14% 83.58%	0.28	0.69	4.01%	0.3276
DDT	Bullet (any)	affected sentences unaffected sentences	9 313		81.82% 88.28%	76.22% 83.06%	87.41% 88.58%	80.42% 83.66%	5.59 0.30	4.20 0.60	17.66% 3.54%	0.1579 0.0203
Talbanken	Parenthesis	overall affected sentences	18	5%	87.27% 78.60%	81.08% 72.67%	86.99% 82.20%	81.15% 75.42%	-0.28 3.60	0.07 2.75	0.37% 10.06%	0.4202 <i>0.1335</i>
		unaffected sentences overall	371	95%	88.06% 87.27%	81.85% 81.08%	87.42% 87.41%	81.67% 81.29%	-0.64 0.14	-0.18 0.21	0.00%	0.3614 0.2620
Talbanken	Colon	affected sentences	5 384	1% 99%	87.27% 63.77% 88.43%	81.08% 53.58% 82.42%	87.41% 65.28% 88.50%	81.29% 55.85% 82.54%	0.14 1.51 0.07	0.21 2.27 0.12	1.11% 4.89% 0.68%	0.3777
		unaffected sentences overall	384	99%	88.43%	78.42%	82.28%	78.42%	0.07	0.12	0.68%	0.3612 0.4937
Cast3lb	CommaApposition	affected sentences unaffected sentences	8 198	4% 96%	80.29% 82.29%	77.78% 78.45%	82.08% 82.29%	79.93% 78.34%	1.79 0.00	2.15 -0.11	9.68% 0.00%	0.0657 0.3836
Cast3lb	DashApposition	overall affected sentences	10	5%	82.19% 86.79%	78.42% 84.08%	82.56% 85.29%	78.70% 82.28%	0.37 -1.50	0.28 -1.80	1.32% 0.00%	0.2228 0.1897
		unaffected sentences overall	196		81.91% 82.19%	78.06% 78.42%	82.39% 82.40%	78.47% 78.52%	0.48	0.41	1.87% 0.48%	0.1354 0.3782
Cast3lb	Apposition (both)	affected sentences	17	8%	83.45%	81.03%	83.79%	81.21%	0.34	0.18	0.95%	0.4895
		unaffected sentences overall	189	92%	82.05% 85.53%	78.12% 75.14%	82.24% 85.87%	78.22% 75.31%	0.19 0.34	0.10	0.46% 0.69%	0.3920 0.3701
Newspaper	QuotedPhrase	affected sentences unaffected sentences	14 86		83.00% 86.05%	70.50% 76.09%	84.00% 86.25%	70.00% 76.39%	1.00 0.20	- <i>0.50</i> 0.30	0.00% 1.25%	0.3098 0.2732
Newspaper	QuotationMark	overall affected sentences	14	14%	85.53% 83.00%	75.14% 70.50%	86.00% 83.00%	75.48% 71.25%	0.47 0.00	0.34 0.75	1.38% 2.54%	0.2359 0.2289
. попорарог		unaffected sentences	86		86.05%	76.09% 75.14%	86.61% 85.87%	76.34% 75.27%	0.56 0.34	0.25 0.13	1.05% 0.51%	0.3193
Newspaper	Parenthesis	overall affected sentences	7		85.53% 76.44%	69.78%	79.56%	72.00%	3.12	2.22	7.35%	0.2546
		unaffected sentences overall	93	93%	86.49% 91.45%	75.70% 77.71%	86.54% 91.98%	75.61% 78.21%	0.05 0.54	- <i>0.09</i> 0.50	0.00% 2.24%	0.4135 <i>0.1441</i>
Law	Parenthesis	affected sentences unaffected sentences	21 79	21% 79%	90.03% 91.89%	77.49% 77.78%	90.19% 92.54%	77.49% 78.44%	0.16 0.65	0.00 0.66	0.00% 2.97%	0.4402 <i>0.1210</i>
Law+Newspaper	Colon	overall affected sentences	0	0%	88.86% 0.00%	76.59% 0.00%	89.00% 0.00%	76.69% 0.00%	0.14 0.00	0.10	0.43% 0.00%	0.3503
Zaw Monopapor		unaffected sentences		100%	88.86%	76.59%	89.00%	76.69%	0.14	0.10	0.44%	0.3503
Law+Newspaper	Parenthesis	overall affected sentences	28		88.86% 86.66%	76.59% 76.39%	89.24% 86.89%	76.89% 76.27%	0.38	0.30 -0.12	1.29% 0.00%	0.1758 0.4922
Law+Newspaper	Parenthesis + QuotationMark	unaffected sentences overall	172	86%	89.31% 88.86%	76.63% 76.59%	89.73% 88.84%	77.02% 76.73%	-0.02	0.39 0.14	1.67% 0.59%	0.1391 0.3581
Law+Newspaper Law+Newspaper	Parenthesis + QuotedPhrase LemmaOnly	overall overall			88.86% 88.86%	76.59% 76.59%	89.34% 89.56%	77.10% 77.32%	0.48 0.71	0.50 0.73	2.15% 3.10%	0.0744 0.0412
		overall	7	70/	85.53%	75.14%	85.62%	74.88%	0.08	-0.25	0.00%	0.2749
Newspaper	OpenClose	affected sentences unaffected sentences	93	7% 93%	76.44% 86.49%	69.78% 75.70%	76.89% 86.54%	69.63% 75.47%	0.45 0.05	-0.15 -0.23	0.00%	0.4064 0.2894
Newspaper	OpenClose + Parenthesis	overall affected sentences	7	7%	85.53% 76.44%	75.14% 69.78%	85.96% 77.33%	75.10% 70.22%	0.42 0.89	-0.04 0.44	0.00% 1.46%	0.4452 0.5096
		unaffected sentences overall	93	93%	86.49% 85.53%	75.70% 75.14%	86.87% 85.19%	75.61% 74.54%	0.38 -0.34	-0.09 -0.59	0.00%	0.4206 <i>0.1111</i>
Newspaper	Parenthetical (parentheses + dashes)	affected sentences unaffected sentences	9	9% 91%	76.29% 86.83%	69.76% 75.90%	75.95% 86.50%	68.38% 75.41%	-0.34 -0.33	-1.38 -0.49	0.00%	<i>0.1249</i> 0.1888
	D	overall			89.91%	85.24%	89.92%	85.24%	0.02	-0.00	0.00%	0.4951
Bul (6K)	Parenthesis	affected sentences unaffected sentences	21 377	5% 95%	87.63%	84.09%	86.11%	82.32%	-1.52 0.00	-1.77 0.00	0.00% 0.00%	0.1359
Bul (6K)	QuotedPhrase	overall affected sentences	38		89.91% 88.62%	85.24% 83.50%	89.96% 88.49%	85.07% 83.25%	0.05 -0.13	-0.17 -0.25	0.00% 0.00%	0.3201 0.3982
		unaffected sentences overall	360	90%	90.10% 89.91%	85.50% 85.24%	90.18% 90.12%	85.35% 85.59%	0.08	-0.15 0.35	0.00% 2.38%	0.3447 0.1629
Bul (6K)	QuotationMark	affected sentences unaffected sentences	38 260		88.62% 90.10%	83.50% 85.50%	89.00% 90.30%	84.40% 85.77%	0.38	0.90 0.27	5.45% 1.86%	0.1891 0.2463
B. I (617)	Enumeration	overall			89.91%	85.24%	90.09%	85.20%	0.18	-0.04	0.00%	0.4508
Bul (6K)	Enumeration	affected sentences unaffected sentences	389		85.16% 90.03%	80.00% 85.38%	93.55%	86.45% 85.17%	8.39 -0.03	6.45 -0.21	32.25% 0.00%	0.0622 0.2590
Bul (6K)	Enumeration1	overall affected sentences	25	6%	89.91% 88.72%	85.24% 84.32%	90.07% 90.44%	85.27% 85.09%	0.16 1.72	0.03 0.77	0.20% <i>4.91</i> %	0.4702 0.2468
		unaffected sentences overall	373	94%	90.02% 89.91%	85.33% 85.24%	90.04% 89.75%	85.29% 84.97%	0.02 -0.15	-0.04 -0.27	0.00% 0.00%	0.4541 0.1654
Bul (6K)	N-Enumeration	affected sentences unaffected sentences	7 391	2% 98%	85.82% 90.00%	78.57% 85.37%	92.86% 89.69%	84.82% 84.97%	7.04	6.25	29.16% 0.00%	0.1241
D.J /0/0	Enumeration (restart)	overall	301		89.91%	85.24%	89.94%	85.14%	0.03	-0.10	0.00%	.5700
Bul (6K)	Enumeration (rooted)	affected sentences unaffected sentences		0% 0%					0.00	0.00	0.00%	
Bul (6K)	PosEnumeration	overall affected sentences	9		89.91% 85.16%	85.24% 80.00%	90.12% 92.90%	85.52% 90.97%	0.22 7.74	0.28 10.97	1.92% 54.85%	0.1715 <i>0.0319</i>
		unaffected sentences overall	389	98%	90.03% 85.34%	85.38% 81.03%	90.05% 85.30%	85.38% 81.00%	0.02 -0.04	0.00 -0.04	0.00% 0.00%	0.4909
Tiger (6K)	QuotedPhrase	affected sentences unaffected sentences		0% 0%					0.00	0.00	0.00% 0.00%	
Tiger (6K)	QuotationMark	overall affected sentences		0%	85.34%	81.03%	85.58%	81.03%	0.25	0.00	0.00%	
J. (514)		unaffected sentences overall		0%	78.56%	69.59%	79.39%	70.49%	0.00	0.00	0.00%	0.0070
SDT	QuotedPhrase	affected sentences	44 358		78.56% 75.63% 78.96%	63.37%	79.39% 78.13% 79.56%	66.27% 71.05%	2.50 0.60	2.90 0.62	7.92%	0.0260
ODT	OustationMark	unaffected sentences overall	358		78.56%	70.43% 69.59%	79.11%	70.11%	0.55	0.52	2.10% 1.70%	0.0474
SDT	QuotationMark	affected sentences unaffected sentences	193 209		72.41% 83.21%	62.52% 74.94%	73.28% 83.51%	63.36% 75.21%	0.87	0.84 0.27	2.24% 1.08%	0.1231 0.2488
		overall affected sentences							0.00	0.00	0.00%	
		unaffected sentences overall							0.00	0.00	0.00%	
		affected sentences unaffected sentences							0.00 0.00	0.00	0.00% 0.00%	
		overall affected sentences							0.00	0.00	0.00%	
		unaffected sentences overall							0.00	0.00	0.00%	
		affected sentences unaffected sentences							0.00 0.00	0.00	0.00% 0.00%	
)							2.33	5.50	/ 0	
Corpus	Feature set	Sentences			UAS	LAS	UAS	LAS	ΔUAS	ΔLAS	LA	p-value
			Sente	ences	Baseline	scores	System	scores	Improve	ement	Error rec	uction

					· · · · · · · · · · · · · · · · · · ·											
	% (of sente	nces wit	th	LAS for sentences with					Δ LAS x presence						
Corpus	"	()	:	-	"	()	:	-	"	()	:	-	"	()	:	-
DDT	28.9%	0.6%	8.7%	4.3%	82.62%	84.75%	76.84%	77.42%	-0.27	1.85	-6.06	-5.48	5	0	2	1
Talbanken	0.0%	4.6%	2.8%	6.7%		72.67%	56.47%	68.67%	0.00	-8.41	-24.61	-12.41	0	1	1	2
BIO	0.2%	26.0%	3.1%	0.1%	78.23%	88.77%	85.45%	92.05%	-13.24	-2.70	-6.02	0.58	0	3	0	0
Law	0.0%	21.0%	2.0%	0.0%		77.49%	87.76%		0.00	-0.22	10.04	0.00	0	5	0	0
Law2	0.0%	21.0%	2.0%	0.0%		86.33%	93.88%		0.00	1.21	8.76	0.00	0	3	0	0
Newspaper	14.0%	7.0%	11.0%	9.0%	70.50%	69.78%	81.59%	73.40%	-4.64	-5.36	6.45	-1.74	4	2	2	2
Law+News	14.0%	28.0%	13.0%	9.0%	71.25%	76.39%	82.40%	77.56%	-6.46	-1.33	4.69	-0.15	4	7	2	2
Cast3lb	10.7%	0.0%	5.3%	19.9%	77.63%		72.91%	79.73%	-0.79	0.00	-5.51	1.31	2	0	1	4
Tiger (6K)	16.2%	0.0%	6.2%	2.2%	78.78%		80.68%	82.91%	-2.25	0.00	-0.35	1.88	3	0	1	0
Bul (6K)	9.5%	5.3%	4.8%	10.8%	83.50%	84.09%	90.69%	83.96%	-1.74	-1.15	5.45	-1.28	2	1	0	2
SDT	48.0%	0.2%	2.5%	8.7%	62.52%	56.52%	50.69%	54.36%	-7.07	-13.07	-18.90	-15.23	18	0	1	4
PDT	5.5%	6.8%	4.4%	11.5%					0.00	0.00	0.00	0.00	5	7	4	12
Metu	6.4%	0.6%	6.1%	2.4%	59.84%	61.00%	65.32%	59.26%	-9.21	-8.05	-3.73	-9.79	3	0	2	1

Difficulties

0.26

1.54

3.62

0.00

Indicator of potential

Presence

0.5% **9.3% 8.3%** 0.0% 74.07% 72.80% 76.15%

Alpino (6K)