Corpus	Language	Training set		Test set			
		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%
Both	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		
Alpino (6K)	Dutch	6000	65839	386	5585	76.97%	72.53%

Corpus	total	# of sentences with... "	( )	:	-	% of sentences with... "	( )	:	-	LAS for sentences with... "	( )	:	-	Δ LAS "	( )	:	-	Δ LAS x presence "	( )	:	-
DDT	322	93	2	28	14	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	389	0	18	11	26	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	1450	3	377	45	2	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	100	0	21	2	0	0.0%	21.0%	2.0%	0.0%		77.49%	87.76%		0.00	-0.22	10.04	0.00	0	5	0	0
Newspaper	100	14	7	11	9	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
Both	200	14	28	13	9	7.0%	14.0%	6.5%	4.5%	71.25%	76.39%	82.40%	77.56%	-5.34	-0.20	5.81	0.97	2	3	1	1
Cast3lb	206	22	0	11	41	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)	357	58	0	22	8	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)	398	38	21	19	43	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	402	193	1	10	35	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	365	20	25	16	42	5.5%	6.8%	4.4%	11.5%					0.00	0.00	0.00	0.00	5	7	4	12
Metu	623	40	4	38	15	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
Alpino (6K)	386	2	36	32	0	0.5%	9.3%	8.3%	0.0%	74.07%	72.80%	76.15%		1.54	0.26	3.62	0.00	0	3	2	0

Experiments

Corpus	Augmentation applied	Sentence focus	Sentences		Baseline scores		System scores		Improvement		Error reduction	
					UAS	LAS	UAS	LAS	Δ UAS	Δ LAS	LA	p-value
DDT	Bullet + QuotedPhrase + LemmaOnly	overall			88.12%	82.89%	88.72%	83.77%	0.60	0.87	5.09%	0.0068
DDT	Lemma	overall			88.12%	82.89%	87.37%	82.23%	-0.75	-0.67	-3.90%	0.0235
DDT	LemmaOnly	overall			88.12%	82.89%	88.72%	83.68%	0.60	0.79	4.60%	0.0117
DDT	Inanimacy	overall			88.12%	82.89%	88.31%	82.95%	0.19	0.05	0.30%	
		affected		0%					0.00	0.00	0.00%	
		unaffected		0%					0.00	0.00	0.00%	
DDT	Animacy	overall			88.12%	82.89%	88.35%	83.15%	0.22	0.26	1.50%	0.2098
		affected	83	26%	87.27%	81.60%	86.92%	81.50%	-0.35	-0.10	-0.54%	0.4329
		unaffected	239	74%	88.57%	83.56%	89.08%	84.00%	0.51	0.44	2.68%	0.1331
DDT	Animacy + Inanimacy	overall			88.12%	82.89%	88.33%	83.13%	0.21	0.24	1.38%	0.1715
		affected		0%					0.00	0.00	0.00%	
		unaffected		0%					0.00	0.00	0.00%	
DDT	QuotedPhrase	overall			88.12%	82.89%	88.88%	83.65%	0.75	0.75	4.40%	0.0014
		affected	43	13%	86.10%	81.17%	87.31%	82.38%	1.21	1.21	6.43%	0.0123
		unaffected	279	87%	88.62%	83.32%	89.26%	83.96%	0.64	0.64	3.84%	0.0106
DDT	InnerQuotedPhrase	overall			88.12%	82.89%	88.52%	83.25%	0.40	0.36	2.08%	0.1033
		affected		0%			86.48%	81.49%	0.00	0.00	0.00%	0.3143
		unaffected		0%			89.10%	83.76%	0.00	0.00	0.00%	0.1119
DDT	InnerQuotedPhrase + QuotationMark	overall			88.12%	82.89%	88.57%	83.27%	0.44	0.38	2.20%	
		affected		0%					0.00	0.00	0.00%	
		unaffected		0%					0.00	0.00	0.00%	
DDT	QuotedPhrase (POS+CPOS)	overall			88.12%	82.89%	88.21%	82.76%	0.09	-0.14	-0.80%	0.3473
		affected		0%			85.58%	80.05%	0.00	0.00	0.00%	0.1679
		unaffected		0%			88.86%	83.43%	0.00	0.00	0.00%	0.4027
DDT	QuotedPhrase (POS)	overall			88.12%	82.89%	88.57%	83.13%	0.44	0.24	1.40%	0.2322
		affected		0%			87.13%	81.26%	0.00	0.00	0.00%	0.4567
		unaffected		0%			88.92%	83.60%	0.00	0.00	0.00%	0.2219
DDT	QuotationMark	overall			88.12%	82.89%	88.55%	83.56%	0.43	0.67	3.90%	0.0066
		affected	93	29%	87.41%	82.62%	88.01%	83.32%	0.60	0.70	4.03%	0.0681
		unaffected	229	71%	88.54%	83.05%	88.86%	83.70%	0.32	0.65	3.83%	0.0260
DDT	QuotedPhrase + QuotationMark	overall			88.12%	82.89%	88.74%	83.56%	0.62	0.67	3.90%	0.0125
		affected	93	29%	87.41%	82.62%	88.01%	83.04%	0.60	0.42	2.42%	0.2230
		unaffected	229	71%	88.54%	83.05%	89.16%	83.86%	0.62	0.81	4.78%	0.0133
DDT	NumericalBullet	overall			88.12%	82.89%	88.53%	83.19%	0.41	0.30	1.73%	0.1426
		affected	6	2%	83.78%	76.58%	90.99%	85.59%	7.21	9.01	38.47%	0.0287
		unaffected	316	98%	88.21%	83.02%	88.49%	83.14%	0.28	0.12	0.71%	0.3276
DDT	Bullet	overall			88.12%	82.89%	88.55%	83.58%	0.43	0.69	4.01%	0.0094
		affected	9	3%	81.82%	76.22%	87.41%	80.42%	5.59	4.20	17.66%	0.1579
		unaffected	313	97%	88.28%	83.06%	88.58%	83.66%	0.30	0.60	3.54%	0.0203
DDT	Colon	overall			88.12%	82.89%	88.64%	83.27%	0.51	0.38	2.20%	0.1048
		affected	24	7%	82.78%	76.78%	84.36%	78.20%	1.58	1.42	6.12%	0.1265
		unaffected	298	93%	88.77%	83.64%	89.16%	83.89%	0.39	0.25	1.53%	0.2164
BIO	Parenthesis	overall			91.71%	91.47%	91.82%	91.56%	0.11	0.09	1.03%	0.1813
		affected	377	26%	89.04%	88.77%	89.37%	89.07%	0.33	0.30	2.67%	0.0656
		unaffected	1,073	74%	93.03%	92.81%	93.03%	92.79%	0.00	-0.02	-0.28%	0.4064
Talbanken	Parenthesis	overall			87.27%	81.08%	86.99%	81.15%	-0.28	0.07	0.37%	0.4202
		affected	18	5%	78.60%	72.67%	82.20%	75.42%	3.60	2.75	10.06%	0.1335
		unaffected	371	95%	88.06%	81.85%	87.42%	81.67%	-0.64	-0.18	-0.99%	0.3614
Talbanken	Colon	overall			87.27%	81.08%	87.41%	81.29%	0.14	0.21	1.11%	0.2620
		affected	5	1%	63.77%	53.58%	65.28%	55.85%	1.51	2.27	4.89%	0.3777
		unaffected	384	99%	88.43%	82.42%	88.50%	82.54%	0.07	0.12	0.68%	0.3612
Cast3lb	CommaApposition	overall			82.19%	78.42%	82.28%	78.42%	0.09	0.00	0.02%	0.4937
		affected	8	4%	80.29%	77.78%	82.08%	79.93%	1.79	2.15	9.68%	0.0657
		unaffected	198	96%	82.29%	78.45%	82.29%	78.34%	0.00	-0.11	-0.51%	0.3836
Cast3lb	DashApposition	overall			82.19%	78.42%	82.56%	78.70%	0.37	0.28	1.32%	0.2228
		affected	10	5%	86.79%	84.08%	85.29%	82.28%	-1.50	-1.80	-11.31%	0.1897
		unaffected	196	95%	81.91%	78.06%	82.39%	78.47%	0.48	0.41	1.87%	0.1354
Cast3lb	Apposition (both)	overall			82.19%	78.42%	82.40%	78.52%	0.21	0.10	0.48%	0.3782
		affected	17	8%	83.45%	81.03%	83.79%	81.21%	0.34	0.18	0.95%	0.4895
		unaffected	189	92%	82.05%	78.12%	82.24%	78.22%	0.19	0.10	0.46%	0.3920
Cast3lb	Colon	overall			82.19%	78.42%	82.24%	78.57%	0.05	0.16	0.73%	0.3503
		affected	11	5%	78.21%	72.91%	79.05%	74.02%	0.84	1.11	4.10%	0.2305
		unaffected	195	95%	82.46%	78.79%	82.46%	78.88%	0.00	0.09	0.42%	0.4231
Newspaper	QuotedPhrase	overall			85.53%	75.14%	85.87%	75.31%	0.34	0.17	0.69%	0.3701
		affected	14	14%	83.00%	70.50%	84.00%	70.00%	1.00	-0.50	-1.69%	0.3098
		unaffected	86	86%	86.05%	76.09%	86.25%	76.39%	0.20	0.30	1.25%	0.2732
Newspaper	QuotationMark	overall			85.53%	75.14%	86.00%	75.48%	0.47	0.34	1.38%	0.2359
		affected	14	14%	83.00%	70.50%	83.00%	71.25%	0.00	0.75	2.54%	0.2289
		unaffected	86	86%	86.05%	76.09%	86.61%	76.34%	0.56	0.25	1.05%	0.3193
Newspaper	Parenthesis	overall			85.53%	75.14%	85.87%	75.27%	0.34	0.13	0.51%	0.3873
		affected	7	7%	76.44%	69.78%	79.56%	72.00%	3.12	2.22	7.35%	0.2546
		unaffected	93	93%	86.49%	75.70%	86.54%	75.61%	0.05	-0.09	-0.37%	0.4135
Law	Parenthesis	overall			91.45%	77.71%	91.98%	78.21%	0.54	0.50	2.24%	0.1441
		affected	21	21%	90.03%	77.49%	90.19%	77.49%	0.16	0.00	0.00%	0.4402
		unaffected	79	79%	91.89%	77.78%	92.54%	78.44%	0.65	0.66	2.97%	0.1210
Both	Colon	overall			88.86%	76.59%	89.00%	76.69%	0.14	0.10	0.43%	0.3503
		affected	0	0%	0.00%	0.00%	0.00%	0.00%	0.00	0.00	0.00%	
		unaffected	200	100%	88.86%	76.59%	89.00%	76.69%	0.14	0.10	0.44%	0.3503
Both	Parenthesis	overall			88.86%	76.59%	89.24%	76.89%	0.38	0.30	1.29%	0.1758
		affected	28	14%	86.66%	76.39%	86.89%	76.27%	0.23	-0.12	-0.51%	0.4922
		unaffected	172	86%	89.31%	76.63%	89.73%	77.02%	0.42	0.39	1.67%	0.1391
Both	Parenthesis + QuotationMark	overall			88.86%	76.59%	88.84%	76.73%	-0.02	0.14	0.59%	0.3581
Both	Parenthesis + QuotedPhrase	overall			88.86%	76.59%	89.34%	77.10%	0.48	0.50	2.15%	0.0744
Both	LemmaOnly	overall			88.86%	76.59%	89.56%	77.32%	0.71	0.73	3.10%	0.0412
Newspaper	OpenClose	overall			85.53%	75.14%	85.62%	74.88%	0.08	-0.25	-1.02%	0.2749
		affected	7	7%	76.44%	69.78%	76.89%	69.63%	0.45	-0.15	-0.50%	0.4064
		unaffected	93	93%	86.49%	75.70%	86.54%	75.47%	0.05	-0.23	-0.95%	0.2894
Newspaper	OpenClose + Parenthesis	overall			85.53%	75.14%	85.96%	75.10%	0.42	-0.04	-0.17%	0.4452
		affected	7	7%	76.44%	69.78%	77.33%	70.22%	0.89	0.44	1.46%	0.5096
		unaffected	93	93%	86.49%	75.70%	86.87%	75.61%	0.38	-0.09	-0.37%	0.4206
Newspaper	Parenthetical (parentheses + dashes)	overall			85.53%	75.14%	85.19%	74.54%	-0.34	-0.59	-2.39%	0.1111
		affected	9	9%	76.29%	69.76%	75.95%	68.38%	-0.34	-1.38	-4.56%	0.1249
		unaffected	91	91%	86.83%	75.90%	86.50%	75.41%	-0.33	-0.49	-2.03%	0.1888
Bul (6K)	Parenthesis	overall			89.91%	85.24%	89.92%	85.24%	0.02	-0.00	-0.00%	0.4951
		affected	21	5%	87.63%	84.09%	86.11%	82.32%	-1.52	-1.77	-11.13%	0.1359
		unaffected	377	95%	90.07%	85.32%	90.20%	85.45%	0.13	0.13	0.89%	0.3472
Bul (6K)	QuotedPhrase	overall			89.91%	85.24%	89.96%	85.07%	0.05	-0.17	-1.16%	0.3201
		affected	38	10%	88.62%	83.50%	88.49%	83.25%	-0.13	-0.25	-1.52%	0.3982
		unaffected	360	90%	90.10%	85.50%	90.18%	85.35%	0.08	-0.15	-1.03%	0.3447
Bul (6K)	QuotationMark	overall			89.91%	85.24%	90.12%	85.59%	0.22	0.35	2.38%	0.1629
		affected	38	10%	88.62%	83.50%	89.00%	84.40%	0.38	0.90	5.45%	0.1891
		unaffected	260	65%	90.10%	85.50%	90.30%	85.77%	0.20	0.27	1.86%	0.2463
Bul (6K)	Enumeration	overall			89.91%	85.24%	90.09%	85.20%	0.18	-0.04	-0.27%	0.4508
		affected	9	2%	85.16%	80.00%	93.55%	86.45%	8.39	6.45	32.25%	0.0622
		unaffected	389	98%	90.03%	85.38%	90.00%	85.17%	-0.03	-0.21	-1.44%	0.2590
Bul (6K)	Enumeration1	overall			89.91%	85.24%	90.07%	85.27%	0.16	0.03	0.20%	0.4702
		affected	25	6%	88.72%	84.32%	90.44%	85.09%	1.72	0.77	4.91%	0.2468
		unaffected	373	94%	90.02%	85.33%	90.04%	85.29%	0.02	-0.04	-0.27%	0.4541
Bul (6K)	N-Enumeration	overall			89.91%	85.24%	89.75%	84.97%	-0.15	-0.27	-1.84%	0.1654
		affected	7	2%	85.82%	78.57%	92.86%	84.82%	7.04	6.25	29.16%	0.1241
		unaffected	391	98%	90.00%	85.37%	89.69%	84.97%	-0.31	-0.40	-2.73%	0.0780
Bul (6K)	DashApposition	overall			89.91%	85.24%	90.09%	85.29%	0.19	0.05	0.33%	0.4318
		affected	3	1%	88.89%	77.78%	81.48%	77.78%	-7.41	0.00	0.00%	1.0000
		unaffected	395	99%	89.91%	85.27%	90.13%	85.32%	0.22	0.05	0.34%	0.4488
Bul (6K)	PosEnumeration	overall			89.91%	85.24%	90.12%	85.52%	0.22	0.28	1.92%	0.1715
		affected	9	2%	85.16%	80.00%	92.90%	90.97%	7.74	10.97	54.85%	0.0319
		unaffected	389	98%	90.03%	85.38%	90.05%	85.38%	0.02	0.00	0.00%	0.4909
Tiger (6K)	QuotedPhrase	overall			85.34%	81.03%	85.30%	81.00%	-0.04	-0.04	-0.19%	
		affected		0%					0.00	0.00	0.00%	
		unaffected		0%					0.00	0.00	0.00%	
Tiger (6K)	QuotationMark	overall			85.34%	81.03%	85.58%	81.03%	0.25	0.00	1.32%	
		affected		0%					0.00	0.00	0.00%	
		unaffected		0%					0.00	0.00	0.00%	
SDT	QuotedPhrase	overall			78.56%	69.59%	79.39%	70.49%	0.83	0.89	2.93%	0.0070
		affected	44	11%	75.63%	63.37%	78.13%	66.27%	2.50	2.90	7.92%	0.0260
		unaffected	358	89%	78.96%	70.43%	79.56%	71.05%	0.60	0.62	2.10%	0.0474
SDT	QuotationMark	overall			78.56%	69.59%	79.11%	70.11%	0.55	0.52	1.70%	0.0909
		affected	193	48%	72.41%	62.52%	73.28%	63.36%	0.87	0.84	2.24%	0.1231
		unaffected	209	52%	83.21%	74.94%	83.51%	75.21%	0.30	0.27	1.08%	0.2488
SDT	DashApposition	overall			78.56%	69.59%			-78.56	-69.59	-228.87%	
		affected		0%					0.00	0.00	0.00%	
		unaffected		0%					0.00	0.00	0.00%	
SDT	Colon	overall			78.56%	69.59%	78.75%	69.97%	0.19	0.38	1.24%	0.1531
		affected	10	2%	63.19%	50.69%	63.19%	52.08%	0.00	1.39	2.82%	0.2552
		unaffected	392	98%	79.29%	70.49%	79.48%	70.81%	0.19	0.32	1.08%	0.1964
SDT	Parenthesis	overall			78.56%	69.59%	78.70%		0.14	-0.02	-0.05%	0.4794
		affected		0%					0.00	0.00	0.00%	
		unaffected		0%					0.00	0.00	0.00%	
Corpus	Feature set	Sentences			UAS	LAS	UAS	LAS	Δ UAS	Δ LAS	LA	p-value
			Sentences		Baseline scores		System scores		Improvement		Error reduction	

	QuotedPhr.		Bullet		Colon		Parenthesis		Comma		Dash		LemmaOnly	
	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All
DA	*6.4%*	**4.4%**	*17.7%*	**4.0%**	6.1%	*2.2%*								**4.6%**
SE					*4.9%*	*1.1%*	*10.1%*	0.4%						
IT					0.0%	0.4%	0.0%	*1.3%*						*3.1%*
ES					*4.1%*	0.7%			9.7%	0.0%	0.0%	*1.3%*		
DE														
BL							0.0%	*-11.0%*						
SL														
CZ														
TU														
DU														