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%
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		

65839

386

6000

Alpino (6K)

Dutch

Test set

76.97%

5585

72.53%

Training set

# of sentences with				% of sentences with				LAS for sentences with				ΔLAS				Δ LAS x presence			ce		
Corpus	total	"	()	:	-	"	()	:	-	"	()	:	-	"	()	:	-	"	()	:	-
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

6.4% 0.6% **6.1%** 2.4% 59.84% 61.00% 65.32% 59.26%

0 0.5% 9.3% 8.3% 0.0% 74.07% 72.80% 76.15%

Presence

623

386

Metu

Alpino (6K)

38

36

Difficulties

-9.21

1.54

-8.05

0.26

-3.73

3.62

-9.79

0.00

0

Indicator of potential

2

0

Corpus	Augmentation applied	Sentence focus	Experime Sentences	Baseline UAS	scores	System UAS	scores LAS	Improve Δ UAS	ement Δ LAS	Error redu LA	uction p-valu
DDT	Bullet + QuotedPhrase + LemmaOnly	overall		88.12%	82.89%	88.72%	83.77%	0.60	0.87	5.09%	0.00
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> 0.60	-0.67 0.79	-3.90% <i>4.60</i> %	0.02
DDT	Inanimacy	overall affected	0%	88.12%	82.89%	88.31%	82.95%	0.19 0.00	0.05 0.00	0.30% 0.00%	
		unaffected overall	0%	88.12%	82.89%	88.35%	83.15%	0.00 0.22	0.00 0.26	0.00% 1.50%	0.20
DDT	Animacy	affected unaffected	83 26% 239 74%	87.27% 88.57%	81.60% 83.56%	86.92% 89.08%	81.50% 84.00%	-0.35 0.51	-0.10 0.44	-0.54% 2.68%	0.43 <i>0.13</i>
DDT	Animacy + Inanimacy	overall affected	0%	88.12%	82.89%	88.33%	83.13%	0.21 0.00	0.24 0.00	1.38% 0.00%	0.17
		unaffected overall	0%	88.12%	82.89%	88.60%	83.24%	0.00 0.48	0.00	0.00% 2.00%	0.10
DDT	DashApposition	affected unaffected	1 0% 321 100%	90.91%	81.82% 82.90%	90.91%	81.82% 83.24%	0.00	0.00	0.00% 1.99%	1.00
	Overted Physics	overall		88.12%	82.89%	88.88%	83.65%	0.75	0.75	4.40%	0.0
DDT	QuotedPhrase	affected unaffected	43 13% 279 87%	86.10% 88.62%	81.17% 83.32%	87.31% 89.26%	82.38% 83.96%	1.21 0.64	1.21 0.64	6.43% 3.84%	0.0
DDT	InnerQuotedPhrase	overall affected	0%	88.12%	82.89%	88.52% 86.48%	83.25% 81.49%	0.40	0.36	2.08% 0.00%	0.1
	InnerQuotedPhrase +	unaffected overall	0%	88.12%	82.89%	89.10% 88.57%	83.76% 83.27%	0.00	0.00	0.00% 2.20%	0.1
DDT	QuotationMark	affected unaffected	0%					0.00	0.00	0.00% 0.00%	
DDT	QuotedPhrase (POS+CPOS)	overall affected	0%	88.12%	82.89%	88.21% 85.58%	82.76% 80.05%	0.09	- <i>0.14</i> 0.00	-0.80% 0.00%	0.3 0.1
		unaffected overall	0%	88.12%	82.89%	88.86% 88.57%	83.43% 83.13%	0.00	0.00 0.24	0.00% 1.40%	0.4
DDT	QuotedPhrase (POS)	affected unaffected	0% 0%			87.13% 88.92%	81.26% 83.60%	0.00	0.00	0.00%	0.4 0.2
DDT	QuotationMark	overall affected	93 29%	88.12% 87.41%	82.89% 82.62%	88.55% 88.01%	83.56% 83.32%	0.43 0.60	0.67 0.70	3.90% 4.03%	0.0
DDI	Quotationiviaix	unaffected	229 71%	88.54%	83.05%	88.86%	83.70%	0.32	0.65	3.83%	0.0
DDT	QuotedPhrase + QuotationMark	overall affected	93 29%	88.12% 87.41%	82.89% 82.62%	88.74% 88.01%	83.56% 83.04%	0.62	0.67	3.90% 2.42%	0.0
		unaffected overall	229 71%	88.54% 88.12%	83.05% 82.89%	89.16% 88.53%	83.86% 83.19%	0.62	0.81	4.78% 1.73%	0.0
DDT	NumericalBullet	affected unaffected	6 2% 316 98%	83.78% 88.21%	76.58% 83.02%	90.99% 88.49%	85.59% 83.14%	7.21 0.28	9.01 0.12	<i>38.47</i> % 0.71%	0.0
DDT	Bullet	overall affected	9 3%	88.12% 81.82%	82.89% 76.22%	88.55% 87.41%	83.58% 80.42%	0.43 5.59	0.69 4.20	4.01% 17.66%	0.0
וטטו	Dullet	unaffected	313 97%	88.28%	83.06%	88.58%	83.66%	0.30	0.60	3.54%	0.0
DDT	Colon	overall affected	24 7%	88.12% 82.78%	82.89% 76.78%	88.64% 84.36%	83.27% 78.20%	0.51 1.58	0.38 1.42	2.20% 6.12%	0.1
		unaffected overall	298 93%	88.77% 91.71%	83.64% 91.47%	89.16% 91.82%	83.89% 91.56%	0.39	0.25	1.53% 1.03%	0.2
BIO	Parenthesis	affected unaffected	377 26% 1,073 74%	89.04% 93.03%	88.77% 92.81%	89.37% 93.03%	89.07% 92.79%	0.33 0.00	0.30 -0.02	2.67% -0.28%	0.0 0.4
Falbanken	Parenthesis	overall affected	18 5%	87.27% 78.60%	81.08% 72.67%	86.99% 82.20%	81.15% 75.42%	- <i>0.28</i> 3.60	0.07 2.75	0.37% 10.06%	0.4
		unaffected 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	-0.99% -1.11%	0.3
Falbanken	Colon	affected unaffected	5 1% 384 99%	87.27% 63.77% 88.43%	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%	0.2 0.3 0.3
		overall		82.19%	78.42%	82.28%	78.42%	0.09	0.00	0.02%	0.4
Cast3lb	CommaApposition	affected unaffected	8 4% 198 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% -0.51%	0.0
Cast3lb	DashApposition	overall affected	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	<i>1.32%</i> -11.31%	0.2
		unaffected overall	196 95%	81.91% 82.19%	78.06% 78.42%	82.39% 82.40%	78.47% 78.52%	0.48	0.41	1.87% 0.48%	0.3 0.3
Cast3lb	Apposition (both)	affected unaffected	17 8% 189 92%	83.45% 82.05%	81.03% 78.12%	83.79% 82.24%	81.21% 78.22%	0.34 0.19	0.18 0.10	0.95% 0.46%	0.4
Coatalla	Colon	overall affected	11 5%	82.19% 78.21%	78.42% 72.91%	82.24% 79.05%	78.57% 74.02%	0.05 0.84	0.16	0.73%	0.3
Cast3lb	COIOII	unaffected	195 95%	82.46%	78.79%	82.46%	78.88%	0.00	0.09	4.10% 0.42%	0.4
lewspaper	QuotedPhrase	overall affected	14 14%	85.53% 83.00%	75.14% 70.50%	85.87% 84.00%	75.31% 70.00%	0.34 1.00	0.17 -0.50	0.69% -1.69%	0.3
		unaffected overall	86 86%	86.05% 85.53%	76.09% 75.14%	86.25% 86.00%	76.39% 75.48%	0.20	0.30 0.34	1.25% 1.38%	0.2
lewspaper	QuotationMark	affected unaffected	14 14% 86 86%	83.00% 86.05%	70.50% 76.09%	83.00% 86.61%	71.25% 76.34%	0.00 0.56	0.75 0.25	2.54% 1.05%	0.2
lewspaper	Parenthesis	overall affected	7 7%	85.53% 76.44%	75.14% 69.78%	85.87% 79.56%	75.27% 72.00%	0.34 3.12	0.13 2.22	0.51% 7.35%	0.3
. Спораро.		unaffected overall	93 93%	86.49% 91.45%	75.70% 77.71%	86.54% 91.98%	75.61% 78.21%	0.05	-0.09 0.50	-0.37% 2.24%	0.4
Law	Parenthesis	affected unaffected	21 21% 79 79%	90.03%	77.49% 77.78%	90.19%	77.49% 78.44%	0.16 0.65	0.00	0.00%	0.4
		overall		88.86%	76.59%	89.00%	76.69%	0.14	0.10	0.43%	0.3
Both	Colon	affected unaffected	0 0% 200 100%	0.00% 88.86%	0.00% 76.59%	0.00% 89.00%	0.00% 76.69%	0.00 0.14	0.00	0.00% 0.44%	0.3
Both	Parenthesis	overall affected	28 14%	88.86% 86.66%	76.59% 76.39%	89.24% 86.89%	76.89% 76.27%	0.38 0.23	0.30 -0.12	1.29% -0.51%	0.1 0.4
Both	Parenthesis + QuotationMark	unaffected overall	172 86%	89.31% 88.86%	76.63% 76.59%	89.73% 88.84%	77.02% 76.73%	0.42 -0.02	0.39	1.67% 0.59%	<i>0.</i>
Both	Parenthesis + QuotedPhrase	overall		88.86%	76.59%	89.34%	77.10%	0.48	0.50	2.15%	0.0
Both	LemmaOnly	overall overall		88.86% 85.53%	76.59% 75.14%	89.56% 85.62%	77.32% 74.88%	0.71	0.73 -0.25	3.10% -1.02%	0.2
lewspaper	OpenClose	affected unaffected	7 7% 93 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	-0.50% -0.95%	0.4
lewspaper	OpenClose + Parenthesis	overall affected	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	-0.17% <i>1.4</i> 6%	0.4 0.5
	· ·	unaffected 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	-0.37% -2.39%	0.4 0.1
lewspaper	Parenthetical (parentheses + dashes)	affected unaffected	9 9% 91 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	-4.56% -2.03%	0. 0.
D 1 (010)	Color	overall		89.91%	85.24%	89.74%	84.88%	-0.17	-0.36	-2.41%	0.
Bul (6K)	Colon	affected unaffected	6 2% 392 98%	85.85% 89.98%	81.13% 85.31%	83.96% 89.84%	81.13% 84.95%	-1.89 -0.14	0.00 -0.36	0.00%	0.4
Bul (6K)	Parenthesis	overall affected	21 5%	89.91% 87.63%	85.24% 84.09%	89.92% 86.11%	85.24% 82.32%	0.02 -1.52	-0.00 -1.77	-0.02% -11.13%	0.4
		unaffected overall	377 95%	90.07% 89.91%	85.32% 85.24%	90.20% 89.96%	85.45% 85.07%	0.13 0.05	0.13 -0.17	0.89%	0.3
Bul (6K)	QuotedPhrase	affected unaffected	38 10% 360 90%	88.62% 90.10%	83.50% 85.50%	88.49% 90.18%	83.25% 85.35%	<i>-0.13</i> 0.08	-0.25 -0.15	-1.52% -1.03%	0.0
Bul (6K)	QuotationMark	overall affected	38 10%	89.91% 88.62%	85.24% 83.50%	90.12% 89.00%	85.59% 84.40%	0.22 0.38	0.35 0.90	2.38% 5.45%	0.
(3.3)		unaffected	260 65%	90.10%	85.50% 85.24%	90.30%	85.77% 85.20%	0.20	0.27	1.86% -0.27%	0.2
Bul (6K)	Enumeration	affected unaffected	9 2% 389 98%	89.91% 85.16% 90.03%	85.24% 80.00% 85.38%	90.09% 93.55% 90.00%	85.20% 86.45% 85.17%	8.39	-0.04 6.45	-0.27% 32.25% -1.44%	0.4 0.0 0.2
B. 1 /2-	F	overall		89.91%	85.24%	90.07%	85.27%	-0.03 0.16	-0.21 0.03	0.20%	0.4
Bul (6K)	Enumeration1	affected unaffected	25 6% 373 94%	88.72% 90.02%	84.32% 85.33%	90.44% 90.04%	85.09% 85.29%	1.72 0.02	0.77 -0.04	4.91% -0.27%	0.2
Bul (6K)	N-Enumeration	overall affected	7 2%	89.91% 85.82%	85.24% 78.57%	89.75% 92.86%	84.97% 84.82%	-0.15 7.04	-0.27 6.25	-1.84% 29.16%	0. · 0.
		unaffected overall	391 98%	90.00% 89.91%	85.37% 85.24%	89.69% 90.09%	84.97% 85.29%	- <i>0.31</i>	- <i>0.40</i> 0.05	-2.73% 0.33%	0.0
Bul (6K)	DashApposition	affected unaffected	3 1% 395 99%	88.89% 89.91%	77.78% 85.27%	81.48% 90.13%	77.78% 85.32%	-7.41 0.22	0.00	0.00% 0.34%	1.0
Rul (GV)	PosEnumeration	overall affected	9 2%	89.91% 85.16%	85.24% 80.00%	90.12%	85.52% 90.97%	0.22	0.28	1.92%	0
Bul (6K)	i osenumeration	unaffected overall	389 98%	90.03% 85.34%	80.00% 85.38% 81.03%	92.90% 90.05% 85.30%	90.97% 85.38% 81.00%	0.02	0.00	54.85% 0.00% -0.19%	0.4
Tiger (6K)	QuotedPhrase	affected	0%	JJ.J4%	J1.U3%	JJ.JU%	01.00%	- <i>0.04</i> 0.00	- <i>0.04</i> 0.00	0.00%	
Ti (0	Outstall 11	unaffected overall	0%	85.34%	81.03%	85.58%	81.03%	0.00	0.00	0.00%	
Tiger (6K)	QuotationMark	affected unaffected	0%					0.00	0.00	0.00%	
SDT	QuotedPhrase	overall affected	44 11%	78.56% 75.63%	69.59% 63.37%	79.39% 78.13%	70.49% 66.27%	0.83 2.50	0.89 2.90	2.93% 7.92%	0.0
		unaffected overall	358 89%	78.96% 78.56%	70.43% 69.59%	79.56% 79.11%	71.05% 70.11%	0.60 0.55	0.62 0.52	2.10% 1.70%	0.0 0.0
SDT	QuotationMark	affected unaffected	193 48% 209 52%	72.41% 83.21%	62.52% 74.94%	73.28% 83.51%	63.36% 75.21%	0.87 0.30	0.84 0.27	2.24% 1.08%	<i>0.</i> 0.2
SDT	DashApposition	overall affected	8 2%	78.56% 65.28%	69.59% 56.48%	78.83% 68.52%	69.51% 58.33%	0.27	-0.08 1.85	-0.26% <i>4.25</i> %	0.4
	1-1-3	unaffected overall	394 98%	79.02% 78.56%	70.05% 69.59%	79.19% 78.75%	69.91% 69.97%	0.17	-0.14 0.38	-0.47% 1.24%	0.3
SDT	Colon	affected unaffected	10 2% 392 98%	63.19% 79.29%	50.69% 70.49%	63.19% 79.48%	52.08% 70.81%	0.19	1.39	2.82% 1.08%	0.2
		overall	392 98%	79.29% 78.56%	69.59%	79.48%	69.58%	0.19	-0.02 0.00	-0.05% 0.00%	0.4
SDT	Parenthesis	affected	70					0.00	0.00	5.50 /0	

												• •						
	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All	Affec.	All
DA	6.4%	4.4%			6.1%	2.2%							17.7%	4.0%				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%						

CommaApp.

DashApp.

0.0% 0.3%

0.3%

0.0%

Parenthesis

Bullet

Enumeration

54.9%

1.9%

LemmaOnly

Quot. Phrase

7.9% 2.9%

BL

SL

Quot. Mark

2.2%

Colon

2.8% 1.2%

-1.5% -1.2% 5.5% 2.4% 0.0% -2.4% -11.0% -0.0%

1.7%

	% of sentences with										
Corpus	"	()	:	_							
Law	0.0%	21.0%	2.0%	0.09							

7.0% 14.0%

14.0%

Newspaper

Both

Presence

7.0% 11.0%

6.5%

9.0%

4.5%