

## Appendix: Degrees of similarity between Romanian and related languages

In the *D* (with diacritics) and *ND* (without diacritics) columns we provide the average degrees of similarity (*AVG*) and the values obtained for the three metrics used: the edit distance (*EDIT*), the longest common subsequence ratio (*LCSR*) and the rank distance (*RD*). For languages for which we determined cognate pairs (besides etymons), we report both versions of the results, before and after cognate identification. In the *%words* column we provide the percentage of related words and the number of lemmas, type words and token words having an etymon or a cognate pair in each language. The results are ordered according to the ranking of similarity for the corpus comprising the parliamentary debates, after identifying cognates and with diacritics included.

Language	Metric	Parliament				Eminescu				Chronicles				RVR										
		#words		D		ND		#words		D		ND		#words		D		ND						
French	AVG	70.6%	45.5	46.0	48.3	48.8	57.2%	35.2	36.1	37.2	38.2	36.7%	20.3	21.1	22.3	23.1	50.6%	30.3	31.4	32.2	33.3			
	EDIT	17,757	44.7	45.2	47.6	48.2	9,277	34.5	35.4	36.7	37.6	1,801	19.9	20.7	22.0	22.7	1,136	29.6	30.6	31.6	32.6			
	LCSR	76,148	45.5	46.0	48.4	48.9	30,080	35.2	36.2	37.4	38.4	5,613	20.4	21.2	22.5	23.3	1,246	30.5	31.5	32.4	33.5			
	RD	15,858,140	46.4	46.9	48.8	49.3	497,742	35.8	36.7	37.6	38.6	93,085	20.7	21.5	22.5	23.4	1,246	30.9	32.0	32.5	33.7			
Latin	AVG	63.7%	40.2		42.0		59.9%		34.6		36.6		44.9%		24.2		25.7		56.5%		34.0		37.3	
	EDIT	9,292	39.9		41.8		6,891		34.2		36.3		2,037		23.7		25.2		1,246		32.0		33.7	
	LCSR	48,320	40.3		42.2		25,367		34.7		36.8		7,145		24.3		25.8		1,391		34.2		37.6	
	RD	14,302,847	40.3		42.1		521,455		34.8		36.8		114,031		24.7		26.1		1,391		34.3		37.4	
Italian	AVG	48.5%	28.1	33.4	29.1	34.5	44.7%	26.9	30.2	27.9	31.2	31.7%	19.6	20.3	20.7	21.4	41.4%	23.4	26.2	25.2	28.0			
	EDIT	6,768	27.9	33.2	28.9	34.3	4,668	26.6	29.9	27.6	31.0	1,344	19.4	20.1	20.5	21.2	922	23.0	25.7	24.8	27.6			
	LCSR	35,581	28.2	33.5	29.2	34.6	17,462	27.0	30.3	28.0	31.4	4,765	19.8	20.5	20.9	21.6	1,019	23.6	26.4	24.8	28.3			
	RD	10,895,298	28.2	33.5	29.2	34.5	389,439	27.0	30.3	28.0	31.3	80,376	19.7	20.4	20.8	21.5	1,019	23.7	26.5	25.4	28.2			
Spanish	AVG	40.2%	9.2	24.9	10.7	27.0	38.1%	10.9	21.2	12.9	23.7	29.7%	11.9	15.1	13.9	17.2	32.5%	9.0	19.5	9.9	21.0			
	EDIT	4,549	9.1	24.6	10.5	26.7	3,429	10.5	20.8	12.6	23.3	996	11.4	14.6	13.5	16.8	721	8.4	18.8	9.3	20.4			
	LCSR	25,402	9.4	25.1	10.9	27.2	13,254	11.2	21.5	13.2	24.0	3,671	12.1	15.3	14.2	17.5	802	9.2	19.8	10.1	21.3			
	RD	9,028,235	9.2	25.1	10.6	27.1	331,733	11.0	21.4	12.9	23.8	75,278	12.1	15.3	14.1	17.4	802	9.3	20.0	10.2	21.4			
Portuguese	AVG	35%	8.3	22.1	9.5	24.0	31.3%	9.6	18.5	11.3	21.0	28.3%	12.2	16.3	13.9	18.4	29.3%	8.6	17.4	9.4	18.9			
	EDIT	4,471	8.0	21.6	9.2	23.5	3,116	9.3	18.1	11.0	20.6	959	11.9	15.9	13.6	18.1	641	8.1	16.7	8.9	18.3			
	LCSR	24,599	8.4	22.3	9.7	24.3	12,108	9.8	18.7	11.5	21.1	3,654	12.4	16.4	14.1	18.6	722	8.9	17.8	9.6	19.2			
	RD	7,855,639	8.4	22.4	9.6	24.3	272,217	9.8	18.8	11.4	21.2	71,746	12.3	16.5	14.0	18.6	722	8.9	17.8	9.6	19.2			
English	AVG	22.1%	2.2	14.0	2.2	14.2	18.8%	1.1	9.9	1.2	10.1	11.3%	1.3	5.9	1.3	6.2	14.3%	1.6	10.3	1.6	10.4			
	EDIT	3,544	2.2	13.7	2.2	13.9	2,086	1.1	9.7	1.2	9.9	435	1.2	5.7	1.2	6.0	330	1.5	10.0	1.5	10.1			
	LCSR	17,182	2.2	14.1	2.2	14.3	7,685	1.2	10.0	1.2	10.2	1,154	1.4	6.0	1.4	6.3	352	1.6	10.4	1.6	10.5			
	RD	4,972,898	2.2	14.1	2.2	14.4	163,705	1.1	9.9	1.1	10.1	28,765	1.4	6.0	1.4	6.2	352	1.6	10.4	1.6	10.5			
Provençal	AVG	17.7%	9.6		9.8		20.7%		11.3		11.6		21.8%		13.0		13.4		16.8%		9.7		10.5	
	EDIT	857	9.5		9.7		769		11.2		11.5		620		12.9		13.3		345		9.4		10.2	
	LCSR	5,858	9.7		10.0		3,818		11.4		11.8		2,892		13.1		13.5		414		9.8		10.6	
	RD	3,974,366	9.5		9.7		180,291		11.2		11.5		55,210		13.0		13.4		414		10.0		10.6	
German	AVG	9.2%	5.8		5.9		6.9%		4.5		4.6		4.9%		2.4		2.4		10.2%		6.3		6.6	
	EDIT	2,103	5.8		5.9		1,244		4.5		4.6		398		2.3		2.3		239		6.2		6.5	
	LCSR	8,757	5.8		6.0		3,892		4.5		4.7		1,051		2.4		2.5		251		6.4		6.7	
	RD	2,063,769	5.7		5.9		59,655		4.5		4.6		12,432		2.4		2.5		251		6.3		6.6	
Turkish	AVG	7.7%	0.9	5.4	0.9	5.6	6.6%	1.7	4.5	1.7	4.7	5.6%	2.9	3.7	3.1	3.9	7.4%	1.6	5.0	1.8	5.3			
	EDIT	2,253	0.9	5.4	0.9	5.6	1,308	1.7	4.5	1.7	4.7	492	2.9	3.6	3.0	3.8	180	1.6	4.9	1.8	5.3			
	LCSR	9,047	0.9	5.4	0.9	5.6	3,838	1.7	4.5	1.8	4.7	1,120	2.9	3.7	3.1	3.9	182	1.6	5.0	1.8	5.3			
	RD	1,733,432	0.8	5.4	0.9	5.6	57,185	1.7	4.5	1.7	4.7	14,317	2.9	3.7	3.1	3.9	182	1.6	5.0	1.8	5.3			
Russian	AVG	5.9%	3.7		4.0		6.5%		4.0		4.4		7.5%		4.3		4.9		9.0%		5.4		6.2	
	EDIT	1,263	3.7		4.0		857		4.0		4.4		509		4.2		4.9		207		5.3		6.2	
	LCSR	5,469	3.7		4.0		2,898		4.0		4.4		1,556		4.3		4.9		221		5.4		6.3	
	RD	1,320,594	3.7		4.0		56,533		3.9		4.3		19,076		4.3		4.8		221		5.4		6.2	
Catalan	AVG	5.9%	3.3		3.4		9%		4.8		5.1		11.2%		5.9		6.4		8.4%		4.6		4.9	
	EDIT	401	3.1		3.2		368		4.6		4.9		297		5.7		6.2		170		4.4		4.7	
	LCSR	2,740	3.3		3.4		1,841		4.8		5.1		1,455		5.9		6.5		206		4.7		5.0	
	RD	1,323,697	3.4		3.5		78,538		4.9		5.2		28,544		6.1		6.5		206		4.8		5.0	
Greek	AVG	4.8%	2.9		3.0		6%		3.6		3.7		4.5%		2.6		2.7		4.6%		2.5		2.6	
	EDIT	1,730	2.9		2.9		909		3.6		3.7		367		2.6		2.7		108		2.4		2.6	
	LCSR	5,920	2.9		3.0		2,536		3.6		3.7		847		2.6		2.7		113		2.5		2.6	
	RD	1,074,583	2.9		3.0		52,674		3.6		3.7		11,310		2.6		2.7		113		2.5		2.6	
Albanian	AVG	4.8%	2.6		3.0		6.7%		3.7		4.0		9.1%		4.9		5.3		8.4%		4.2		4.8	
	EDIT	668	2.5		2.9		525		3.6		3.9		389		4.8		5.2		188		4.1		4.7	
	LCSR	3,334	2.6		3.0		1,933		3.7		4.0		1,418		4.9		5.3		208		4.3		4.9	
	RD	1,083,260	2.6		3.0		58,720		3.7		4.0		23,205		4.9		5.3		208		4.3		4.9	
Bulgarian	AVG	4%	2.6		3.0		7.4%		4.7		5.5		10.6%		6.8		7.8		11.8%		7.2		8.4	
	EDIT	1,386	2.6		3.0		982		4.6		5.5		695		6.7		7.7		279		7.0		8.2	
	LCSR	5,522	2.6		3.0		3,060		4.7		5.5		2,248		6.9		7.9		290		7.3		8.4	
	RD	894,931	2.6		3.1		64,361		4.7		5.5		26,973		6.9		7.8		290		7.3		8.5	
Slavic	AVG	4.9%	2.3		2.5		6.6%		3.4		3.8		12.1%		6.5		7.7		9.8%		5.0		5.7	
	EDIT	995	2.3		2.5		755		3.3		3.7		615		6.4		7.6		228		4.8		5.5	
	LCSR	4,633	2.4		2.5		2,681		3.5		3.8		2,184		6.6		7.7		242		5.1		5.8	
	RD	1,096,845	2.3		2.5		57,820		3.4		3.8		30,700		6.5		7.7		242		5.1		5.9	
Old Slavic	AVG	3.8%	2.2		2.7		6.1%		3.3		4.3		11.9%		6.8		8.7		9.5%		5.2		6.0	
	EDIT	1,017	2.1		2.6		819		3.2		4.2		681		6.6		8.4		220		5.0		5.8	
	LCSR	4,670	2.2		2.7		2,791		3.4		4.4		2,539		7.0		8.8		234		5.3		6.1	
	RD	864,626	2.2		2.7		53,368		3.4		4.3		30,246		6.9		8.8		234		5.3		6.2	
Hungarian	AVG	2.9%	1.8		2.0		5.1%		2.9		3.3		7.5%		4.3		4.7		7.4%		3.7			

Polish	AVG	1.3%	0.7	0.8	2.2%	1.2	1.5	4.3%	2.2	2.6	4.3%	2.5	2.8
	EDIT	507	0.7	0.8	352	1.2	1.4	286	2.2	2.5	101	2.4	2.7
	LCSR	1,965	0.7	0.8	1,054	1.3	1.5	781	2.2	2.6	106	2.5	2.8
	RD	300,428	0.7	0.8	19,372	1.2	1.5	10,857	2.2	2.6	106	2.5	2.8
Neo-Greek	AVG	1.4%	0.7	0.9	2.5%	1.2	1.7	2.7%	1.1	2.0	2.4%	1.2	1.5
	EDIT	459	0.7	0.9	339	1.2	1.7	182	1.0	2.0	57	1.2	1.5
	LCSR	1,742	0.7	1.0	1,036	1.2	1.7	502	1.1	2.0	58	1.2	1.5
	RD	322,644	0.7	0.9	21,613	1.3	1.7	6,823	1.2	2.0	58	1.2	1.5
Venetian	AVG	0.8%	0.4	0.4	1%	0.5	0.5	0.6%	0.3	0.3	0.6%	0.3	0.4
	EDIT	36	0.3	0.4	36	0.4	0.5	20	0.3	0.3	14	0.3	0.4
	LCSR	198	0.4	0.4	137	0.5	0.5	68	0.3	0.3	16	0.3	0.4
	RD	168,666	0.4	0.4	8,284	0.5	0.5	1,524	0.3	0.4	16	0.4	0.4
Galician	AVG	0.7%	0.3	0.3	0.6%	0.2	0.3	0.9%	0.4	0.5	0.7%	0.3	0.4
	EDIT	46	0.3	0.3	40	0.2	0.2	34	0.4	0.5	17	0.3	0.4
	LCSR	224	0.3	0.3	170	0.3	0.3	127	0.4	0.5	17	0.3	0.4
	RD	157,247	0.3	0.3	5,056	0.2	0.3	2,199	0.4	0.5	17	0.3	0.4
American	AVG	0.2%	0.2	0.2	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0
	EDIT	31	0.2	0.2	8	0.0	0.0	4	0.0	0.0	1	0.0	0.0
	LCSR	97	0.2	0.2	30	0.0	0.0	8	0.0	0.0	1	0.0	0.0
	RD	52,829	0.2	0.2	389	0.0	0.0	24	0.0	0.0	1	0.0	0.0
Serbo-Croatian	AVG	0.3%	0.1	0.1	0.6%	0.3	0.3	1.1%	0.5	0.5	1.6%	0.8	0.9
	EDIT	172	0.1	0.1	104	0.2	0.3	69	0.4	0.5	40	0.7	0.8
	LCSR	673	0.1	0.1	334	0.3	0.3	223	0.5	0.5	40	0.8	0.9
	RD	66,821	0.1	0.1	5,382	0.3	0.3	2,831	0.5	0.5	40	0.8	0.9
Croatian	AVG	0.2%	0.1	0.1	0.4%	0.2	0.2	0.6%	0.4	0.4	1.2%	0.7	0.7
	EDIT	126	0.1	0.1	89	0.2	0.2	78	0.3	0.4	29	0.7	0.7
	LCSR	562	0.1	0.1	316	0.2	0.2	261	0.4	0.4	29	0.7	0.7
	RD	51,098	0.1	0.1	3,422	0.2	0.2	1,490	0.4	0.4	29	0.7	0.7
Chzech	AVG	0.1%	0.1	0.1	0.1%	0.1	0.1	0.6%	0.2	0.3	0.7%	0.4	0.5
	EDIT	111	0.1	0.1	78	0.1	0.1	55	0.2	0.3	16	0.4	0.5
	LCSR	410	0.1	0.1	179	0.1	0.1	166	0.2	0.3	17	0.4	0.5
	RD	22,484	0.1	0.1	987	0.1	0.1	1,480	0.2	0.3	17	0.4	0.5
Friulian	AVG	0.1%	0.1	0.1	0.5%	0.2	0.2	0.6%	0.2	0.3	0.2%	0.1	0.1
	EDIT	25	0.1	0.1	23	0.2	0.2	19	0.2	0.2	6	0.1	0.1
	LCSR	132	0.1	0.1	90	0.2	0.2	79	0.2	0.3	6	0.1	0.1
	RD	33,392	0.1	0.1	4,609	0.2	0.2	1,517	0.2	0.3	6	0.1	0.1
Persian	AVG	0.1%	0.1	0.1	0.1%	0.1	0.1	0.3%	0.2	0.2	0.4%	0.2	0.3
	EDIT	124	0.1	0.1	72	0.1	0.1	52	0.2	0.2	11	0.2	0.3
	LCSR	419	0.1	0.1	174	0.1	0.1	115	0.2	0.2	11	0.2	0.3
	RD	20,524	0.1	0.1	1,172	0.1	0.1	694	0.2	0.2	11	0.2	0.3
Slovenian	AVG	0.2%	0.1	0.1	0.2%	0.2	0.2	0.3%	0.2	0.2	1.2%	0.7	0.8
	EDIT	91	0.1	0.1	63	0.2	0.2	56	0.2	0.2	29	0.7	0.8
	LCSR	392	0.1	0.1	218	0.2	0.2	144	0.2	0.2	30	0.7	0.8
	RD	34,747	0.1	0.1	2,025	0.2	0.2	646	0.2	0.2	30	0.7	0.8
Dalmatian	AVG	0.1%	0.1	0.1	0.1%	0.1	0.1	0.3%	0.1	0.1	0.1%	0.1	0.1
	EDIT	6	0.1	0.1	6	0.1	0.1	4	0.1	0.1	3	0.1	0.1
	LCSR	48	0.1	0.1	29	0.1	0.1	28	0.1	0.1	3	0.1	0.1
	RD	24,058	0.0	0.0	1,114	0.1	0.1	684	0.1	0.1	3	0.1	0.1
Arabic	AVG	0.1%	0.0	0.1	0.2%	0.1	0.1	0.6%	0.4	0.4	0.5%	0.2	0.2
	EDIT	171	0.0	0.1	135	0.1	0.1	75	0.3	0.4	11	0.2	0.2
	LCSR	582	0.0	0.1	328	0.1	0.1	198	0.4	0.4	12	0.2	0.2
	RD	26,038	0.1	0.1	2,042	0.1	0.1	1,396	0.4	0.4	12	0.2	0.2
Ukrainian	AVG	0%	0.0	0.0	0.1%	0.0	0.0	0.6%	0.3	0.3	0.4%	0.3	0.3
	EDIT	133	0.0	0.0	76	0.0	0.0	75	0.3	0.3	10	0.3	0.3
	LCSR	364	0.0	0.0	171	0.0	0.0	206	0.3	0.3	11	0.3	0.3
	RD	6,619	0.0	0.0	472	0.0	0.0	1,432	0.3	0.3	11	0.2	0.3
Indo-European	AVG	0.1%	0.0	0.0	0.1%	0.0	0.0	0.2%	0.1	0.1	0.3%	0.2	0.2
	EDIT	16	0.0	0.0	12	0.0	0.0	13	0.1	0.1	8	0.2	0.2
	LCSR	96	0.0	0.0	48	0.0	0.0	46	0.1	0.1	8	0.2	0.2
	RD	22,475	0.0	0.0	1,304	0.0	0.0	407	0.1	0.1	8	0.2	0.2
Cuman	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0.1%	0.1	0.1
	EDIT	15	0.0	0.0	11	0.0	0.0	12	0.0	0.0	3	0.1	0.1
	LCSR	49	0.0	0.0	27	0.0	0.0	28	0.0	0.0	3	0.1	0.1
	RD	1,341	0.0	0.0	134	0.0	0.0	103	0.0	0.0	3	0.1	0.1
Tatar	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0
	EDIT	10	0.0	0.0	8	0.0	0.0	7	0.0	0.0	1	0.0	0.0
	LCSR	30	0.0	0.0	18	0.0	0.0	22	0.0	0.0	1	0.0	0.0
	RD	228	0.0	0.0	68	0.0	0.0	96	0.0	0.0	1	0.0	0.0
Hindu	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0.1%	0.0	0.0
	EDIT	6	0.0	0.0	4	0.0	0.0	5	0.0	0.0	2	0.0	0.0
	LCSR	26	0.0	0.0	11	0.0	0.0	10	0.0	0.0	2	0.0	0.0
	RD	730	0.0	0.0	62	0.0	0.0	70	0.0	0.0	2	0.0	0.0
Latvian	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0
	EDIT	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0
	LCSR	9	0.0	0.0	8	0.0	0.0	5	0.0	0.0	1	0.0	0.0
	RD	515	0.0	0.0	99	0.0	0.0	36	0.0	0.0	1	0.0	0.0
Judeo-Spanish	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0.1%	0.1	0.1
	EDIT	13	0.0	0.0	11	0.0	0.0	4	0.0	0.0	2	0.1	0.1
	LCSR	53	0.0	0.0	35	0.0	0.0	8	0.0	0.0	3	0.1	0.1
	RD	719	0.0	0.0	170	0.0	0.0	24	0.0	0.0	3	0.1	0.1
Judeo-Portuguese	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0
	EDIT	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0
	LCSR	8	0.0	0.0	6	0.0	0.0	4	0.0	0.0	1	0.0	0.0
	RD	571	0.0	0.0	62	0.0	0.0	41	0.0	0.0	1	0.0	0.0
Armenian	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0
	EDIT	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0
	LCSR	8	0.0	0.0	4	0.0	0.0	2	0.0	0.0	1	0.0	0.0
	RD	235	0.0	0.0	15	0.0	0.0	9	0.0	0.0	1	0.0	0.0
Hebrew	AVG	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0	0%	0.0	0.0
	EDIT	4	0.0	0.0	6	0.0	0.0	2	0.0	0.0	0	0.0	0.0
	LCSR	13	0.0	0.0	15	0.0	0.0	2	0.0	0.0	0	0.0	0.0
	RD	65	0.0	0.0	34	0.0	0.0	2	0.0	0.0	0	0.0	0.0