ISO	Name	Language Family	size	az	ba	be	bg	bn	bs	ca	cs	da	de	el	en	eo	es	et	eu	fa	fi	fr	gl	he	hi	hr	hu	id	is	it	ja	kk	ko	lt	mk	ml	mr	ne	nl	no	oc	pl	pt	ro	ru	sh	si	sk	sl	sq	sr	sv	sw	ta	te	tl	tr	tt	uk	vi	zh	total
ar	Arabic	Arabic	**6516**	17	15	11	54	40	34	94	67	53	99	66	999	37	174	40	24	58	53	163	50	68	38	38	60	90	18	123	83	11	48	33	52	32	32	12	73	58	9	74	157	71	125	35	32	32	39	30	49	58	13	27	27	18	69	14	70	93	86	**4435**
az	Azerbaijani	Turkic	**1873**		7	3	14	8	9	15	17	11	34	10	71	9	31	10	7	16	14	29	10	12	9	10	15	17	4	25	23	6	11	9	8	5	4	2	22	13	2	22	25	14	47	9	3	9	8	9	10	19	5	11	5	2	42	8	19	10	19	**885**
ba	Bashkir	Turkic	**536**			2	14	4	9	17	16	13	27	10	28	9	28	8	6	9	12	29	12	9	4	10	12	12	3	26	13	4	5	7	9	1	1	1	21	15	3	19	24	15	42	10	1	10	10	7	10	20	4	3	1	1	11	8	15	7	10	**697**
be	Belarusian	Slavic	**1690**				16	4	7	16	14	9	20	8	33	9	28	7	5	5	10	24	9	10	3	9	11	9	3	24	12	2	5	7	8	2	2	1	14	10	2	19	23	13	161	8	2	8	7	5	10	16	2	3	2	1	8	2	80	8	8	**803**
bg	Bulgarian	Slavic	**3327**					38	34	76	79	53	132	62	357	40	122	43	25	37	61	117	43	58	30	47	68	60	17	102	71	11	38	42	86	29	35	13	84	58	9	96	114	69	270	41	31	43	46	26	65	63	12	21	23	21	56	12	126	60	60	**3770**
bn	Bengali	Indo-Aryan	**1412**						21	41	47	33	70	36	280	27	81	26	14	20	37	68	27	34	21	23	41	36	8	64	38	4	20	21	23	7	8	3	50	35	4	52	76	46	62	20	8	25	26	17	25	54	6	12	8	4	33		37	31	31	**1991**
bs	Bosnian	Slavic	**1060**							44	43	32	71	33	210	24	70	25	16	20	36	60	32	30	16	164	39	38	11	52	36	6	22	23	39	19	20	8	45	36	6	48	62	37	59	178	16	25	34	19	130	38	7	15	12	13	33	8	39	38	31	**2348**
ca	Catalan	Romance	**8332**								100	86	180	90	1205	81	1580	54	77	44	83	490	268	84	37	57	92	107	23	316	103	12	52	45	61	45	56	17	144	102	57	121	358	110	169	52	52	50	57	34	67	102	14	30	35	31	77	16	98	106	90	**8047**
cs	Czech	Slavic	**7634**									75	233	70	519	75	181	62	36	45	95	185	54	72	38	63	105	78	23	161	105	15	53	55	51	36	41	15	139	86	11	176	153	82	186	50	37	474	64	30	63	97	15	31	34	25	75	14	104	74	80	**5325**
da	Danish	Germanic	**2881**										180	54	436	39	140	45	26	29	75	142	44	55	25	43	69	63	20	115	76	9	37	35	39	30	35	12	110	303	8	89	123	70	109	37	32	39	40	23	43	168	11	20	21	23	55	11	62	68	57	**4006**
de	German	Germanic	**50944**											95	1573	186	418	106	53	66	163	626	80	109	57	87	192	107	34	388	217	23	82	78	64	51	58	21	472	207	17	285	294	129	368	68	50	94	106	51	81	216	20	58	57	32	127	23	165	107	134	**9692**
el	Greek	Hellenic	**3211**												620	39	145	41	23	35	55	137	48	56	26	43	64	73	15	119	69	6	35	34	52	27	32	9	76	60	8	77	144	78	114	38	31	35	46	28	52	62	11	16	18	20	56	9	68	75	62	**3728**
en	English	Germanic	**134431**													298	3377	243	119	303	375	2757	446	545	231	259	488	1019	85	2126	851	20	306	157	395	71	124	15	796	636	37	668	2461	631	1661	224	115	178	318	180	395	546	51	95	91	75	477	32	681	1073	786	**34321**
eo	Esperanto	constructed	**2371**														149	31	25	23	46	134	46	39	22	29	57	46	15	101	48	7	26	28	30	28	28	9	81	47	9	77	91	43	81	26	28	41	32	19	36	53	8	16	19	17	37	8	50	42	39	**2817**
es	Spanish	Romance	**25202**															89	154	83	155	905	610	153	71	94	167	198	42	671	219	26	108	76	92	65	98	25	272	181	35	235	923	183	393	81	84	81	93	53	107	181	21	57	71	48	147	26	187	206	174	**14980**
et	Estonian	Uralic	**2303**																22	24	70	85	32	39	20	33	56	41	14	75	57	6	29	35	32	20	21	8	72	49	7	73	76	48	96	27	19	34	35	18	34	58	9	16	16	15	43	7	56	40	44	**2511**
eu	Basque	Isolate	**2259**																	14	33	65	43	25	13	21	35	27	10	54	33	5	18	19	19	11	10	5	44	29	6	43	58	30	47	19	9	20	20	12	21	38	6	13	11	8	29	5	30	25	23	**1699**
fa	Farsi	Iranian	**3954**																		34	71	25	36	20	24	39	46	9	64	46	8	26	20	27	11	10	6	49	32	4	50	77	40	72	21	9	21	24	17	30	42	8	21	12	7	42	9	41	38	42	**2122**
fi	Finnish	Uralic	**7428**																			156	47	64	28	48	90	63	22	131	87	9	43	47	40	29	30	10	126	86	10	119	131	69	139	39	27	50	46	25	46	126	12	23	24	21	72	10	76	60	64	**3998**
fr	French	Romance	**34494**																				154	136	60	85	164	161	38	744	214	24	89	71	83	62	83	25	331	166	124	255	558	206	410	72	74	83	86	48	92	186	19	56	65	42	130	26	170	165	157	**12441**
gl	Galician	Romance	**2221**																					41	21	33	50	56	14	120	50	9	28	27	35	29	39	11	66	52	17	65	227	56	84	30	36	29	33	20	39	54	8	15	17	22	43	12	51	58	46	**3969**
he	Hebrew	Semitic	**6962**																						28	41	65	63	17	121	82	7	43	35	42	26	25	9	86	64	8	84	133	67	131	35	21	36	38	23	45	67	10	21	25	13	54	8	73	66	62	**3666**
hi	Hindi	Indo-Aryan	**1353**																							21	33	31	7	56	35	3	18	16	24	6	11	12	40	27	4	44	63	35	56	17	6	18	21	14	22	40	7	13	18	3	29	4	33	26	30	**1781**
hr	Croatian	Slavic	**2229**																								58	47	14	80	48	8	27	31	52	24	24	10	65	48	7	71	85	51	85	666	19	35	53	21	205	56	9	16	16	17	42	9	55	46	42	**3629**
hu	Hungarian	Uralic	**7702**																									70	20	146	99	11	49	48	47	27	28	10	121	75	10	126	148	87	149	46	26	56	55	27	53	88	13	29	30	20	75	9	83	74	75	**4357**
id	Indonesian	Malayo-Polynesian	**3899**																										16	146	77	8	45	33	55	25	23	10	101	83	8	93	204	94	127	43	23	37	46	32	56	79	13	24	19	21	79	11	73	146	83	**4743**
is	Icelandic	Germanic	**488**																											31	18	2	9	12	12	6	5	3	27	22	3	27	35	20	30	13	4	13	13	9	13	28	4	7	5	5	16	3	18	16	14	**995**
it	Italian	Romance	**21025**																												179	24	83	62	73	58	78	24	240	150	20	219	480	161	303	67	68	72	81	48	83	153	19	44	58	41	112	23	144	143	137	**10084**
ja	Japanese	Japonic	**31614**																													14	222	47	48	21	23	8	123	81	9	128	175	79	196	40	19	48	50	28	51	96	12	37	31	12	84	12	92	75	267	**5382**
kk	Kazakh	Turkic	**1684**																														6	5	6	2	2	1	18	11	1	17	22	12	32	8	1	8	7	6	8	16	3	3	1	1	10	4	14	6	9	**581**
ko	Korean	Koreanic	**5400**																															23	26	10	10	4	56	41	5	63	93	47	89	23	9	25	26	17	29	51	7	17	13	5	47	5	48	49	57	**2539**
lt	Lithuanian	Baltic	**2031**																																28	16	16	6	57	39	6	70	64	39	107	25	15	30	30	16	29	46	8	13	11	12	36	6	57	33	35	**2091**
mk	Macedonian	Slavic	**1487**																																	21	22	9	53	46	6	56	93	56	88	52	19	29	39	25	106	48	9	14	15	16	43	8	57	57	45	**2761**
ml	Malayalam	Dravidian	**272**																																		3	1	41	32	7	41	58	35	46	21	2	22	21	13	20	44	6	4	3	1	16	2	26	15	17	**1388**
mr	Marathi	Indo-Aryan	**503**																																			1	46	35	5	47	86	47	50	22	3	25	25	13	19	56	4	5	3	1	17	1	26	14	17	**1648**
ne	Nepali	Indo-Aryan	**235**																																				17	13	4	17	21	14	19	10	1	10	10	6	9	17	3	1	1		7	1	11	6	6	**584**
nl	Dutch	Germanic	**14091**																																					133	13	177	218	96	199	53	42	66	64	34	61	151	16	37	35	29	90	18	104	84	88	**6364**
no	Norwegian	Germanic	**1364**																																						9	103	161	74	121	42	28	43	51	26	47	270	12	24	23	24	61	13	69	79	63	**4771**
oc	Occitan	Romance	**472**																																							13	24	11	14	6	5	6	6	4	7	12	2	3	3	3	7	3	9	8	7	**706**
pl	Polish	Slavic	**16270**																																								200	97	285	56	40	81	68	35	69	121	16	39	42	28	92	16	172	84	92	**6035**
pt	Portuguese	Romance	**13354**																																									177	312	74	76	71	85	47	101	155	20	42	54	45	140	23	156	213	165	**10932**
ro	Romanian	Romance	**3504**																																										136	44	43	42	49	30	58	75	15	23	27	29	72	13	82	96	74	**4515**
ru	Russian	Slavic	**32537**																																											70	42	85	78	44	114	140	17	54	55	29	119	25	2486	122	148	**11311**
sh	Serbo-Croatian languages	South Slavic	**2069**																																												17	27	46	19	373	46	8	14	13	17	35	9	45	42	36	**3337**
si	Sinhala	Indo-Aryan	**320**																																													22	22	10	15	48	4	4	4	2	15	1	21	14	16	**1464**
sk	Slovak	Slavic	**1904**																																														36	17	34	51	8	14	15	16	37	9	51	38	38	**2726**
sl	Slovenian	Slavic	**1836**																																															19	47	50	9	15	16	17	39	8	50	48	42	**2623**
sq	Albanian	Albanian	**740**																																																25	31	6	13	10	10	27		29	30	24	**1515**
sr	Serbian	Slavic	**4226**																																																	51	9	19	18	14	43	8	71	56	45	**3510**
sv	Swedish	Germanic	**17906**																																																		16	33	39	35	72	17	82	74	73	**5107**
sw	Swahili	Niger-Congo	**235**																																																			6	3	4	12		12	10	11	**652**
ta	Tamil	Dravidian	**1629**																																																				6	2	29	5	30	19	27	**1315**
te	Telugu	Dravidian	**1509**																																																					1	21	1	30	16	20	**1315**
tl	Tagalog	Malayo-Polynesian	**312**																																																						12	1	16	17	10	**1015**
tr	Turkish	Turkic	**4067**																																																							10	67	77	69	**3698**
tt	Tatar	Turkic	**449**																																																								11	7	10	**603**
uk	Ukrainian	Slavic	**12871**																																																										73	72	**7043**
vi	Vietnamese	Vietic	**6727**																																																											89	**4660**
zh	Chinese	Chinese	**12306**																																																													**4415**

Table 1: WikiMatrix: size of mined sentences **(in thousands)** for each langauge pair.