ISO	Name	Family	Size	bg	cs	da	de	el	en	es	fa	fi	fr	he	hi	hu	id	it	ja	ko	ms	nl	no	pl	pt	ru	tr	uk	vi	zh	Total
ar	**Arabic**	**Arabic**	**196**	3.0	3.9	2.7	7.5	3.3	6.5	10.0	3.1	2.7	-	2.2	1.4	2.7	4.1	5.8	5.0	2.5	1.5	5.1	2.5	4.5	6.7	9.2	5.5	1.5	4.2	5.4	**112.3**
bg	**Bulgarian**	**Slavic**	**68**	-	6.1	3.7	9.9	4.3	3.7	10.7	2.3	3.6	11.4	2.1	1.5	3.8	3.8	7.4	5.7	2.8	1.3	6.9	3.0	7.2	7.5	17.4	5.8	2.3	4.4	5.0	**146.5**
cs	**Czech**	**Slavic**	**303**	-	-	5.9	18.3	5.4	9.8	15.5	2.9	6.1	17.3	3.1	2.0	6.1	5.3	11.2	8.0	4.0	2.0	11.6	4.9	13.2	10.7	18.1	8.6	2.6	6.0	7.0	**215.8**
da	**Danish**	**Germanic**	**109**	-	-	-	12.6	3.8	4.5	-	2.0	4.8	12.0	2.3	1.5	3.7	3.9	7.3	5.6	2.9	1.4	9.5	9.6	6.5	7.4	9.2	5.7	1.5	4.2	4.9	**139.2**
de	**German**	**Germanic**	**1728**	-	-	-	-	9.8	67.3	-	4.8	11.3	50.0	5.6	3.2	11.0	9.6	29.5	11.6	6.2	3.5	33.2	10.4	20.5	23.4	29.3	15.5	3.8	9.7	11.8	**429.5**
el	**Greek**	**Hellenic**	**144**	-	-	-	-	-	5.6	12.2	2.2	3.6	12.9	2.3	1.4	3.7	3.7	8.5	5.2	2.6	1.4	6.9	3.0	6.2	8.4	9.9	5.6	1.7	4.2	4.7	**142.7**
en	**English**	**Germanic**	**8677**	-	-	-	-	-	-	86.3	2.5	4.1	94.1	1.5	0.7	3.6	13.4	31.3	33.7	7.2	0.8	23.8	3.8	16.0	33.1	72.4	26.8	1.6	18.5	17.6	**590.4**
es	**Spanish**	**Romance**	**1534**	-	-	-	-	-	-	-	5.5	9.7	-	5.9	3.2	9.5	12.4	44.3	-	6.2	-	23.3	8.8	19.6	59.4	32.4	15.2	4.0	11.9	13.2	**419.3**
fa	**Farsi**	**Iranian**	**192**	-	-	-	-	-	-	-	-	2.0	5.5	1.7	1.2	1.9	3.1	3.6	3.5	2.0	1.3	3.6	1.9	3.2	4.1	5.6	4.9	1.1	3.3	3.4	**82.3**
fi	**Finnish**	**Uralic**	**132**	-	-	-	-	-	-	-	-	-	11.1	2.2	1.4	4.2	3.8	7.1	6.2	3.0	1.4	8.1	4.1	6.8	7.1	9.9	6.2	1.7	4.4	5.2	**142.0**
fr	**French**	**Romance**	**1869**	-	-	-	-	-	-	-	-	-	-	6.8	3.5	10.3	11.9	-	12.6	6.9	4.2	32.1	9.9	21.1	37.9	31.9	17.4	4.2	12.5	14.0	**451.2**
he	**Hebrew**	**Semitic**	**70**	-	-	-	-	-	-	-	-	-	-	-	1.2	1.9	2.8	4.0	5.3	2.5	1.1	4.2	2.0	3.6	4.3	6.4	4.4	1.2	3.6	3.6	**87.8**
hi	**Hindi**	**Indo-Aryan**	**48**	-	-	-	-	-	-	-	-	-	-	-	-	1.3	1.9	2.3	2.7	1.6	0.9	2.4	1.4	2.1	2.6	3.4	3.2	0.8	1.9	2.4	**53.0**
hu	**Hungarian**	**Uralic**	**148**	-	-	-	-	-	-	-	-	-	-	-	-	-	3.2	7.0	5.2	2.6	1.3	7.1	3.0	7.1	6.8	9.6	5.6	1.7	3.7	4.6	**132.2**
id	**Indonesian**	**Malayo-Polynesian**	**366**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7.4	5.9	3.5	4.4	7.6	3.7	6.0	9.1	9.9	8.1	1.7	7.9	6.3	**164.4**
it	**Italian**	**Romance**	**686**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8.9	4.7	2.5	16.6	6.1	14.7	25.4	20.5	10.5	2.8	8.0	8.6	**306.1**
ja	**Japanese**	**Japonic**	**2944**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3.3	8.9	5.1	7.7	9.1	11.6	12.1	2.8	6.5	13.5	**205.8**
ko	**Korean**	**Koreanic**	**778**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.9	4.8	2.6	4.0	4.9	6.0	8.4	1.4	5.2	6.3	**106.6**
ms	**Malay**	**Malayo-Polynesian**	**25**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.6	1.3	2.3	2.8	3.7	3.4	0.8	3.2	2.8	**57.1**
nl	**Dutch**	**Germanic**	**510**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7.8	12.9	15.5	17.7	11.0	2.7	7.2	8.4	**301.3**
no	**Norwegian**	**Germanic**	**109**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5.5	6.4	8.1	5.2	1.4	3.9	4.3	**130.0**
pl	**Polish**	**Slavic**	**505**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13.5	22.9	9.1	3.4	6.5	7.1	**253.2**
pt	**Portuguese**	**Romance**	**729**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20.9	11.0	3.0	8.8	9.5	**359.4**
ru	**Russian**	**Slavic**	**3047**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	31.2	10.4	13.0	**440.7**
tr	**Turkish**	**Turkic**	**1382**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.5	10.4	10.0	**232.0**
uk	**Ukrainian**	**Slavic**	**110**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.2	2.2	**85.8**
vi	**Vietnamese**	**Vietic**	**1172**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9.1	**179.6**
zh	**Chinese**	**Chinese**	**2512**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**203.9**

Table 1: CCMatrix: size of mined sentences **(in millions)** for each language pair.