Language	ISO 639-3	Family	Type	# Speakers
Chinese Mandarin	CMN	Sino-Tibetan	Isolating	1.116 B
Welsh	CYM	IE: Celtic	Fusional	0.7 M
English	ENG	IE: Germanic	Fusional	1.132 B
Estonian	EST	Uralic	Agglutinative	1.1 M
Finnish	FIN	Uralic	Agglutinative	5.4 M
French	FRA	IE: Romance	Fusional	280 M
Hebrew	HEB	Afro-Asiatic	Introflexive	9 M
Polish	POL	IE: Slavic	Fusional	50 M
Russian	RUS	IE: Slavic	Fusional	260 M
Spanish	SPA	IE: Romance	Fusional	534.3 M
Kiswahili	SWA	Niger-Congo	Agglutinative	98 M
Yue Chinese	YUE	Sino-Tibetan	Isolating	73.5 M

Table 1: The list of 12 languages in the Multi-SimLex multilingual suite along with their corresponding language family (IE = Indo-European), broad morphological type, and their ISO 639-3 code. The number of speakers is based on the total count of L1 and L2 speakers, according to ethnologue.com.

5.1 Word Pair Translation

Translators for each target language were instructed to find direct or approximate translations for the 1,888 word pairs that satisfy the following rules. (1) All pairs in the translated set must be unique (i.e., no duplicate pairs); (2) Translating two words from the same English pair into the same word in the target language is not allowed (e.g., it is not allowed to translate *car* and *automobile* to the same Spanish word *coche*). (3) The translated pairs must preserve the semantic relations between the two words when possible. This means that, when multiple translations are possible, the translation that best conveys the semantic relation between the two words found in the original English pair is selected. (4) If it is not possible to use a single-word translation in the target language, then a multi-word expression (MWE) can be used to convey the nearest possible semantics given the above points (e.g., the English word *homework* is translated into the Polish MWE *praca domowa*).

Satisfying the above rules when finding appropriate translations for each pair–while keeping to the spirit of the intended semantic relation in the English version–is not always straightforward. For instance, kinship terminology in Sinitic languages (Mandarin and Yue) uses different terms depending on whether the family member is older or younger, and whether the family member comes from the mother's side or the father's side. In Mandarin, *brother* has no direct translation and can be translated as either: 哥哥(*older brother*) or 弟弟(*younger brother*). Therefore, in such cases, the translators are asked to choose the best option given the semantic context (relation) expressed by the pair in English, otherwise select one of the translations arbitrarily. This is also used to remove duplicate pairs in the translated set, by differentiating the duplicates using a variant at each instance. Further, many translation instances were resolved using near-synonymous terms in the translation. For example, the words in the pair: *wood – timber* can only be directly translated in Estonian to *puit*, and are not distinguishable. Therefore, the translators approximated the translation for *timber* to the compound noun *puitmaterjal* (literally: *wood material*) in order to produce a valid pair in the target language. In some cases, a direct transliteration from English is used. For example, the pair: *physician* and