

POWER LANGUAGE INDEX (RESULTS, METHODOLOGY & INDICATORS)

RESULTS

The Power Language Index (PLI) is an assessment of the influence of a language on the global stage.

TABLE 1 below lists 124 languages on their overall importance, as well as their strengths in opening the opportunities of geography, economy, communication, knowledge & media, and diplomacy.

Column 2 below is the PLI score (expressed to 3 decimal places), which ranges from 0 (least powerful) to 1 (most powerful). Column 4 is the number of native speakers of a language in millions. Columns 5 through 9 are the rank orderings of the languages with respect to the five opportunities. Note that only 9 languages are used in (high-level) global diplomacy!

TABLE 1: POWER LANGUAGE INDEX RESULTS (ALL CHINESE & HINDUSTAN LANGUAGES)

RANK	SCORE	LANGUAGE	NATIVE	GEOGRAPHY	ECONOMY	COMM.	K&M	DIPLOMACY
1	0.889	English	446.0	1	1	1	1	1
2	0.411	Mandarin	960.0	6	2	2	3	6
3	0.337	French	80.0	2	6	5	5	1
4	0.329	Spanish	470.0	3	5	3	7	3
5	0.273	Arabic	295.0	4	9	6	18	4
6	0.244	Russian	150.0	5	12	10	9	5
7	0.191	German	92.5	8	3	7	4	8
8	0.133	Japanese	125.0	27	4	22	6	7
9	0.119	Portuguese	215.0	7	19	13	12	9
10	0.122	Hindi	376.0	13	16	8	2	10
11	0.116	Cantonese	80.0	21	11	4	13	10
12	0.108	Italian	64.0	10	8	19	8	10
13	0.084	Dutch	22.0	16	7	24	11	10
14	0.077	Malay	77.0	9	17	21	22	10
15	0.055	Polish	40.0	23	22	23	15	10
16	0.053	Korean	80.0	22	14	37	10	10
17	0.053	Shanghainese	80.0	100	43	9	28	10
18	0.047	Turkish	75.0	11	24	38	16	10
19	0.047	Shanxinese	48.0	89	59	11	28	10
20	0.046	Hunnanese	38.0	85	66	14	28	10
21	0.046	Hokkien	47.0	104	63	12	28	10
22	0.046	Gan Chinese	22.0	88	58	16	28	10
23	0.046	Romanian	24.0	26	37	20	41	10
24	0.045	Northern Min	10.9	104	54	17	28	10
25	0.045	Hakka	31.0	104	63	15	28	10
26	0.045	Eastern Min	9.5	104	54	18	28	10
27	0.043	Norwegian	5.0	43	10	106	25	10
28	0.041	Swedish	9.2	36	13	72	14	10
29	0.040	Persian	52.5	14	36	30	17	10
30	0.040	Urdu	66.0	24	44	25	43	10
31	0.034	Danish	5.5	42	15	84	20	10
32	0.034	Hebrew	4.4	58	23	36	21	10
33	0.033	Czech	10.6	37	21	45	27	10
34	0.033	Kazakh	11.0	15	32	57	56	10
35	0.032	Thai	56.0	17	33	65	37	10
36	0.032	Finnish	5.4	46	18	90	19	10
37	0.032	Ukrainian	30.0	18	50	27	59	10
38	0.030	Tamil	70.0	35	34	31	57	10
39	0.029	Bengali	210.0	71	74	26	36	10
40	0.029	Greek	13.0	20	28	89	24	10

Kai L. Chan, PhD

Distinguished Fellow, INSEAD

E: Kai.Chan@INSEAD.edu W: www.KaiLChan.ca M: +971 (0)50 258-5317

Power Language Index (May 2016)

Kai L. Chan, PhD

RANK	SCORE	LANGUAGE	NATIVE	GEOGRAPHY	ECONOMY	COMM.	K&M	DIPLOMACY
41	0.028	Serbo-Croatian	19.0	19	45	47	48	10
42	0.028	Slovak	5.2	48	26	53	49	10
43	0.027	Slovene	2.5	64	27	52	40	10
44	0.027	Maltese	0.5	72	25	48	65	10
45	0.026	Hungarian	13.0	33	30	71	38	10
46	0.026	Swahili	10.0	12	79	59	70	10
47	0.026	Javanese	82.0	31	40	32	70	10
48	0.024	Icelandic	0.3	70	20	119	53	10
49	0.023	Bulgarian	9.0	44	39	49	45	10
50	0.023	Latvian	7.8	67	35	54	58	10
51	0.020	Belarusian	7.6	66	38	56	65	10
52	0.019	Vietnamese	76.0	34	49	70	23	10
53	0.019	Azerbaijani	26.0	61	41	64	50	10
54	0.019	Punjabi	100.0	38	60	41	70	10
55	0.019	Tagalog	28.0	45	53	61	26	10
56	0.017	Estonian	1.2	57	31	120	51	10
57	0.017	Lithuanian	3.0	62	29	117	60	10
58	0.017	Turkmen	8.0	52	42	76	70	10
59	0.016	Zulu	12.0	40	46	79	46	10
60	0.016	Macedonian	2.0	77	48	60	63	10
61	0.015	Xhosa	8.2	40	46	94	46	10
62	0.015	Pashto	50.0	28	73	69	70	10
63	0.016	Awadhi	38.0	104	110	28	70	10
64	0.014	Sindhi	75.0	63	70	50	70	10
65	0.016	Chhattisgarhi	18.0	91	100	29	70	10
66	0.014	Amharic	25.0	55	97	42	70	10
67	0.014	Uyghur	10.4	30	68	75	70	10
68	0.013	Tigrinya	6.9	73	116	46	39	10
69	0.013	Mongolian	10.0	25	52	116	53	10
70	0.013	Odia	33.0	90	95	34	70	10
71	0.013	Uzbek	27.0	47	72	68	70	10
72	0.013	Bhojpuri	40.0	104	96	33	70	10
73	0.012	Telugu	76.0	83	92	43	70	10
74	0.012	Maithili	30.0	104	110	35	70	10
75	0.012	Sinhalese	16.0	69	51	78	70	10
76	0.012	Assamese	15.0	93	99	39	70	10
77	0.011	Magahi	14.0	104	110	40	70	10
78	0.011	Malagasy	18.0	49	115	62	70	10
79	0.011	Sylheti	11.0	98	117	44	70	10
80	0.011	Burmese	33.0	39	75	87	43	10
81	0.011	Sundanese	38.0	94	78	55	70	10
82	0.010	Kannada	38.1	87	93	51	70	10
83	0.010	Kyrgyz	4.3	53	86	73	65	10
84	0.009	Georgian	4.3	51	57	110	52	10
85	0.009	Malayalam	38.0	97	98	58	70	10
86	0.009	Madurese	15.0	101	85	63	70	10
87	0.009	Marathi	73.0	82	91	66	70	10
88	0.009	Lao	3.3	54	76	88	64	10
89	0.009	Kurdish	25.0	92	69	74	70	10
90	0.008	Zhuang	16.0	104	54	92	70	10
91	0.007	Somali	17.0	32	122	97	70	10
92	0.007	Saraiki	20.0	104	110	67	70	10
93	0.007	Bambara	4.0	29	107	112	65	10
94	0.007	Armenian	10.0	74	62	111	61	10

Kai L. Chan, PhD

Distinguished Fellow, INSEAD

E: Kai.Chan@INSEAD.edu W: www.KaiLChan.ca M: +971 (0)50 258-5317

Power Language Index (May 2016)
Kai L. Chan, PhD

RANK	SCORE	LANGUAGE	NATIVE	GEOGRAPHY	ECONOMY	COMM.	K&M	DIPLOMACY
95	0.007	Khmer	16.0	50	81	108	53	10
96	0.007	Balochi	7.6	104	77	80	70	10
97	0.007	Quechua	8.9	56	71	115	70	10
98	0.006	Gujarati	50.0	86	94	77	70	10
99	0.006	Bhutanese	0.2	78	67	124	42	10
100	0.006	Fijian	0.3	76	61	123	65	10
101	0.006	Nepali	25.0	65	89	99	70	10
102	0.005	Marwari	22.0	104	82	85	70	10
103	0.005	Dakhini	11.0	84	105	82	70	10
104	0.005	Chewa	12.0	68	118	101	70	10
105	0.005	Cebuano	21.0	99	80	91	70	10
106	0.005	Hmong	8.4	104	63	121	70	10
107	0.005	Wolof	4.2	60	90	113	61	10
108	0.005	Kinyarwanda	9.8	75	106	102	70	10
109	0.004	Yoruba	28.0	104	87	93	70	10
110	0.004	Konkani	7.4	102	109	86	70	10
111	0.004	Igbo	25.0	104	87	96	70	10
112	0.004	Ilocano	9.1	104	82	98	70	10
113	0.004	Hiligaynon	8.2	104	82	100	70	10
114	0.004	Hausa	34.0	104	124	81	70	10
115	0.004	Kirundi	8.8	80	121	103	70	10
116	0.004	Oromo	38.0	104	123	83	70	10
117	0.004	Dhundhari	9.6	96	110	95	70	10
118	0.004	Mossi	7.6	59	104	122	70	10
119	0.004	Fula	24.0	81	119	105	70	10
120	0.003	Haitian Creole	9.6	79	103	118	70	10
121	0.003	Haryanvi	14.0	95	101	109	70	10
122	0.003	Shona	8.3	104	120	104	70	10
123	0.002	Chittagonian	16.0	103	108	107	70	10
124	0.002	Akan	11.0	104	102	114	70	10

Dealing with the Chinese languages poses certain difficulties. For one, some would argue that the collection of Chinese languages are indeed just one language which are better viewed as dialects. Given the mutual intelligibility of the variants of Chinese most linguist would disagree. Nevertheless, none of the Chinese languages other than Mandarin Chinese and Cantonese have some kind of official status – even in the case of Cantonese, its status as official in Hong Kong is peculiar as Hong Kong itself has become a special administrative region (SAR) of China and was never a member of the UN. Moreover, China’s policy of promoting Mandarin Chinese (*Putonghua*), even at the expense of the other variants of Chinese, detracts from much of the efficacy of the other Chinese languages. For historical reasons and because of the large Cantonese diaspora and the prominence of Hong Kong, Cantonese is typically viewed outside of China as the alternative variant of Chinese. Moreover, Cantonese has a rich number of valid indicators to measure its efficacy, whereas the other variants of Chinese do not.

Similarly, Hindi and Urdu are essentially the same languages that use different scripts and have since their split (the partition of India and Pakistan) developed their own unique features. Moreover, there are various dialects of Hindi which are often classified as languages. Putting these “languages” under the umbrella of Hindi would significantly raise its number of native speakers and extend its scope.

TABLE 2 below presents the results of the PLI if: (1) The Chinese languages are all lumped into one (“Chinese”), with the exception of Cantonese; and (2) The Hindustan languages (Indo-Aryan Central Zone) are all grouped as one (“Hindi”). This change reduces the count of languages in the PLI to 113.

Kai L. Chan, PhD

Distinguished Fellow, INSEAD

E: Kai.Chan@INSEAD.edu W: www.KaiLChan.ca M: +971 (0)50 258-5317

Power Language Index (May 2016)
Kai L. Chan, PhD

TABLE 2: POWER LANGUAGE INDEX (MERGE CHINESE & HINDUSTAN LANGUAGES)

RANK	SCORE	LANGUAGE	NATIVE	GEOGRAPHY	ECONOMY	COMM.	K&M	DIPLOMACY
1	0.889	English	446.0	1	1	1	1	1
2	0.411	Mandarin	960.0	6	2	2	3	6
3	0.337	French	80.0	2	6	5	5	1
4	0.329	Spanish	470.0	3	5	3	7	3
5	0.273	Arabic	295.0	4	9	6	18	4
6	0.244	Russian	150.0	5	13	10	9	5
7	0.191	German	92.5	8	3	7	4	8
8	0.134	Hindi	376.0	10	12	8	2	10
9	0.133	Japanese	125.0	26	4	22	6	7
10	0.119	Portuguese	215.0	7	19	13	12	9
11	0.116	Cantonese	80.0	21	11	4	13	10
12	0.108	Italian	64.0	11	8	19	8	10
13	0.084	Dutch	22.0	16	7	24	11	10
14	0.077	Malay	77.0	9	17	21	22	10
15	0.055	Polish	40.0	23	22	23	15	10
16	0.053	Korean	80.0	22	15	37	10	10
17	0.047	Turkish	75.0	12	24	38	16	10
18	0.046	Romanian	24.0	25	37	20	41	10
19	0.043	Norwegian	5.0	42	10	106	25	10
20	0.041	Swedish	9.2	35	14	72	14	10
21	0.040	Persian	52.5	14	36	30	17	10
22	0.034	Danish	5.5	41	16	84	20	10
23	0.034	Hebrew	4.4	57	23	36	21	10
24	0.033	Czech	10.6	36	21	45	27	10
25	0.033	Kazakh	11.0	15	32	57	55	10
26	0.032	Thai	56.0	17	33	65	37	10
27	0.032	Finnish	5.4	45	18	90	19	10
28	0.032	Ukrainian	30.0	18	49	27	58	10
29	0.030	Tamil	70.0	34	34	31	56	10
30	0.029	Bengali	210.0	70	73	26	36	10
31	0.029	Greek	13.0	20	28	89	24	10
32	0.028	Serbo-Croatian	19.0	19	44	47	47	10
33	0.028	Slovak	5.2	47	26	53	48	10
34	0.027	Slovene	2.5	63	27	52	40	10
35	0.027	Maltese	0.5	71	25	48	64	10
36	0.026	Hungarian	13.0	32	30	71	38	10
37	0.026	Swahili	10.0	13	78	59	69	10
38	0.026	Javanese	82.0	30	40	32	69	10
39	0.024	Icelandic	0.3	69	20	119	52	10
40	0.023	Bulgarian	9.0	43	39	49	44	10
41	0.023	Latvian	7.8	66	35	54	57	10
42	0.020	Belarusian	7.6	65	38	56	64	10
43	0.019	Vietnamese	76.0	33	48	70	23	10
44	0.019	Azerbaijani	26.0	60	41	64	49	10
45	0.019	Punjabi	100.0	37	59	41	69	10
46	0.019	Tagalog	28.0	44	52	61	26	10
47	0.017	Estonian	1.2	56	31	120	50	10
48	0.017	Lithuanian	3.0	61	29	117	59	10
49	0.017	Turkmen	8.0	51	42	76	69	10
50	0.016	Zulu	12.0	39	45	79	45	10
51	0.016	Macedonian	2.0	76	47	60	62	10
52	0.015	Xhosa	8.2	39	45	94	45	10
53	0.015	Pashto	50.0	27	72	69	69	10

Kai L. Chan, PhD

Distinguished Fellow, INSEAD

E: Kai.Chan@INSEAD.edu W: www.KaiLChan.ca M: +971 (0)50 258-5317

Power Language Index (May 2016)

Kai L. Chan, PhD

RANK	SCORE	LANGUAGE	NATIVE	GEOGRAPHY	ECONOMY	COMM.	K&M	DIPLOMACY
54	0.014	Sindhi	75.0	62	69	50	69	10
55	0.014	Amharic	25.0	54	96	42	69	10
56	0.014	Uyghur	10.4	29	67	75	69	10
57	0.013	Tigrinya	6.9	72	115	46	39	10
58	0.013	Mongolian	10.0	24	51	116	52	10
59	0.013	Odia	33.0	89	94	34	69	10
60	0.013	Uzbek	27.0	46	71	68	69	10
61	0.013	Bhojpuri	40.0	103	95	33	69	10
62	0.012	Telugu	76.0	82	91	43	69	10
63	0.012	Maithili	30.0	103	109	35	69	10
64	0.012	Sinhalese	16.0	68	50	78	69	10
65	0.012	Assamese	15.0	92	98	39	69	10
66	0.011	Magahi	14.0	103	109	40	69	10
67	0.011	Malagasy	18.0	48	114	62	69	10
68	0.011	Sylheti	11.0	97	116	44	69	10
69	0.011	Burmese	33.0	38	74	87	43	10
70	0.011	Sundanese	38.0	93	77	55	69	10
71	0.010	Kannada	38.1	86	92	51	69	10
72	0.010	Kyrgyz	4.3	52	85	73	64	10
73	0.009	Georgian	4.3	50	56	110	51	10
74	0.009	Malayalam	38.0	96	97	58	69	10
75	0.009	Madurese	15.0	100	84	63	69	10
76	0.009	Marathi	73.0	81	90	66	69	10
77	0.009	Lao	3.3	53	75	88	63	10
78	0.009	Kurdish	25.0	91	68	74	69	10
79	0.008	Zhuang	16.0	103	53	92	69	10
80	0.007	Somali	17.0	31	121	97	69	10
81	0.007	Saraiki	20.0	103	109	67	69	10
82	0.007	Bambara	4.0	28	106	112	64	10
83	0.007	Armenian	10.0	73	61	111	60	10
84	0.007	Khmer	16.0	49	80	108	52	10
85	0.007	Balochi	7.6	103	76	80	69	10
86	0.007	Quechua	8.9	55	70	115	69	10
87	0.006	Gujarati	50.0	85	93	77	69	10
88	0.006	Bhutanese	0.2	77	66	124	42	10
89	0.006	Fijian	0.3	75	60	123	64	10
90	0.006	Nepali	25.0	64	88	99	69	10
91	0.005	Marwari	22.0	103	81	85	69	10
92	0.005	Dakhini	11.0	83	104	82	69	10
93	0.005	Chewa	12.0	67	117	101	69	10
94	0.005	Cebuano	21.0	98	79	91	69	10
95	0.005	Hmong	8.4	103	62	121	69	10
96	0.005	Wolof	4.2	59	89	113	60	10
97	0.005	Kinyarwanda	9.8	74	105	102	69	10
98	0.004	Yoruba	28.0	103	86	93	69	10
99	0.004	Konkani	7.4	101	108	86	69	10
100	0.004	Igbo	25.0	103	86	96	69	10
101	0.004	Ilocano	9.1	103	81	98	69	10
102	0.004	Hiligaynon	8.2	103	81	100	69	10
103	0.004	Hausa	34.0	103	123	81	69	10
104	0.004	Kirundi	8.8	79	120	103	69	10
105	0.004	Oromo	38.0	103	122	83	69	10
106	0.004	Dhundhari	9.6	95	109	95	69	10
107	0.004	Mossi	7.6	58	103	122	69	10

Kai L. Chan, PhD

Distinguished Fellow, INSEAD

E: Kai.Chan@INSEAD.edu W: www.KaiLChan.ca M: +971 (0)50 258-5317

Power Language Index (May 2016)

Kai L. Chan, PhD

RANK	SCORE	LANGUAGE	NATIVE	GEOGRAPHY	ECONOMY	COMM.	K&M	DIPLOMACY
108	0.004	Fula	24.0	80	118	105	69	10
109	0.003	Haitian Creole	9.6	78	102	118	69	10
110	0.003	Haryanvi	14.0	94	100	109	69	10
111	0.003	Shona	8.3	103	119	104	69	10
112	0.002	Chittagonian	16.0	102	107	107	69	10
113	0.002	Akan	11.0	103	101	114	69	10

* Note that lumping all the variants of Chinese into one (with the exception of Cantonese) also changes the scores of the other languages in the index since the transformed indicator scores are derived by dividing by the maximum value in the sample. (The maximum number of native speakers of any given language – Chinese – increases from 960 million to 1,246.4 million.)

With the Chinese (except Cantonese) and Hindi languages each taken collectively, the rank ordering of languages catapults Hindi from 10th to 8th place. But the higher count of native speakers of Chinese does not change its rank, as there is a big gap between 2nd and 1st ranking. Cantonese, the lone “other” Chinese, places 11th just ahead of Italian.

METHODOLOGY

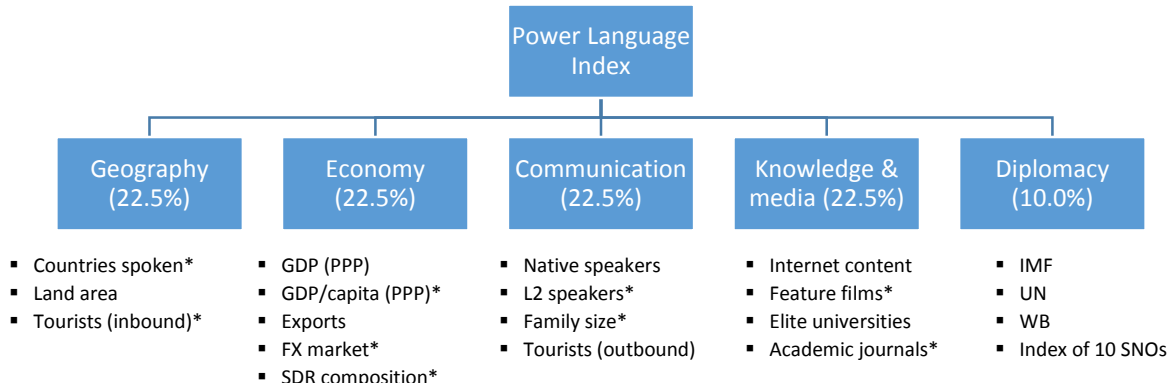
The Power Language Index (PLI) consists of 20 indicators grouped into 5 categories (“opportunities”): geography, economy, communication, knowledge & media, and diplomacy. The index is constructed so that each of the first four opportunities are equally weighted at 22.5 per cent apiece of the final score. The final pillar thus constitutes 10.0 per cent of the index score. Scores of all indicators are mapped into the [0,1] interval by expressing them as a ratio of the maximum indicator value. Indicators in the diplomacy opportunity are indicator variables that take on a value of 1 when condition is true and 0 otherwise. Note that the final index score is itself cardinal and in the range [0,1].

The 124 languages were chosen based on a multi-stage process: (1) Compile a list of the top 100 languages by native speakers; (2) Add languages associated with the 20 indicators if language has significant number of native speakers and/or has official status (though the latter is not sufficient); (3) Add languages associated with UN member states if they are “significant”.

GROUPING & WEIGHTS

Within each opportunity the contribution of each indicator is inversely proportional to the number of indicators within that group. However, some indicators are assigned a half weight within their respective opportunity (denoted with an asterisk (*)) in FIGURE 1 below). This is done to take into account that some variables logically seem less relevant than their peers (within the same opportunity). For example, native speakers has full weight while the second language speakers (L2) is assigned a half weight.

FIGURE 1: POWER LANGUAGE INDEX STRUCTURE



Kai L. Chan, PhD

Distinguished Fellow, INSEAD

E: Kai.Chan@INSEAD.edu W: www.KaiLChan.ca M: +971 (0)50 258-5317

MAPPING

Many of the indicators used in the index are not directly linked with a language. Rather the bulk are associated with nation states/economies. To map indicator values associated with nation states to languages it is required to define within a country its dominant and minority languages.

A “dominant” language is one that is widely spoken and, for the most part, the majority mother tongue of a country. It is the (or a) *de facto* working language of a country and often has some form of official status. A “minority” language is one subordinate to the dominant language but which has a significant presence (minimum threshold of 10-15 per cent native speakers, but is also dependent on its geographical concentration); it often also some kind of official status or political recognition (“minority”). Having status as an official language of a country is neither necessary nor sufficient to be classified as either dominant or minority.

Then for each indicator associated with a nation-state, the value is mapped to its dominant language(s) by a factor of 1; for minority language(s) it is mapped by a factor of ½.

For example, in Canada English spoken by roughly ¾ of the population and is geographically widespread; whereas French is the mother tongue of about ¼ of the population and is spoken primarily in the provinces of Quebec and New Brunswick. Both languages have official (and equal) status in Canada (at the federal level). Here English would take on a value of 1, while French would count as ½. So the full GDP of Canada is mapped to English, while French gets recognised for half that value.

The number of countries for which the indicators in the index are assessed are the 193 countries recognized by the UN, as well as a few large states outside of the UN charter (e.g. Kosovo and Taiwan) and exceptional cases (e.g. Hong Kong).

INDICATORS

GEOGRAPHY

Knowledge of a language enables the ability to travel and explore geography.

Countries spoken:* This indicator captures the number of countries in which a language is spoken – typically as a native tongue, but sometimes also as a second language. Note that language spoken can be, and is often the case, highly different than the official language(s) of a country. For example, in some African nations French is the official language even though very few actually speak it as a mother tongue or at home. Nevertheless, in the case of many post-French-colonial African nations French acts as the effective *lingua franca* (typically spoken at L2 level) given the otherwise disparate and numerous varieties of tribal languages.

As per the mapping rule, if a language is deemed “dominant” it counts full, whereas if it is “minority” it is counted as half.

Source: UN; national censuses (2016 or most recent)

Land area: The total land and inland water area within the recognised borders of a country, not including territorial claim on the seas. (This indicator corrects for the fact that not all countries are equal in size.) Knowledge of a language enables the exploration of a country. No adjustment is made on quality of land – every square kilometer of land is exactly interchangeable with any other in the world.

Source: CIA Factbook; national sources (2016)

Tourists (inbound):* The number of international inbound tourist (overnight visitors) whose main purpose is other than business. This indicator captures the desirability of a given land and hence of its associated language(s). It partially corrects for the land area indicator as it, in a sense, is a proxy for the desirability of land.

Source: World Bank (2014 or most recent)

ECONOMY

Language enables participation in the economic life of a country where it is spoken. However, when international business transactions are conducted they are often done in a third language.

GDP (PPP): Gross domestic product (GDP) – i.e. the monetary value of goods and services produced within an economy in a given year – measured in purchasing power parity (PPP) international dollars. PPP is generally preferred to the nominal measure of GDP when making cross-country comparisons.

Knowledge of a language enables one to participate in economies where that language is used.

Source: IMF; World Bank (2015 or most recent)

GDP per capita (PPP):* Gross domestic product (GDP) – i.e. the monetary value of goods and services produced within an economy in a given year – measured in purchasing power parity (PPP) international dollars divided by total population.

Large countries might be on a whole rich, but poor on a per capita basis. GDP per capita is a proxy for the average standard of living in a country and thus also the economic attractiveness of a country.

Source: IMF; World Bank (2015 or most recent)

Exports: The US dollar (f.o.b.) value of goods sold by a country into foreign markets in a calendar year.

Exports from a country are a means by which countries (and hence languages) exert their influence on the international stage. Importing nations receive both goods and soft culture (e.g. language) from their trading partners.

Source: CIA Factbook (2014 or most recent)

Foreign exchange market:* The share of foreign exchange transactions (by value) in the global foreign exchange market. Note that the sum is 200% because each trade is a two-way transaction.

This indicator is a proxy for the global economic heft of countries, as expressed by the share of market turnover by currency in the foreign exchange market. Currencies are mapped to languages based on the “presumed” home country of the currency. In some cases, a country might not have its own sovereign currency (e.g. dollarisation). In this case, English (in the case of dollarization) is over-represented.

Source: Bank for International Settlements (BIS); Investopedia (2015)

Special drawing rights (SDR) composition:* The foreign exchange assets maintained by the IMF.

This indicator is a reflection of the global financial power structure. SDRs are pseudo-currency of foreign exchange reserves managed by the IMF. It is the unit of account for the IMF. Its composition is based on 4 currencies as defined in the table below:

Kai L. Chan, PhD

Distinguished Fellow, INSEAD

E: Kai.Chan@INSEAD.edu W: www.KaiLChan.ca M: +971 (0)50 258-5317

TABLE 3: SDR COMPOSITION

SDR BASKET	USD (USA)	EUR (Eurozone)	CNY (China)	JPY (Japan)	GBP (UK)
2016-2020	41.73%	30.93%	10.92%	8.33%	8.09%
Language	English	Various*	Mandarin	Japanese	English

Mapping the EUR to languages is based on the share of Eurozone GDP of the respective countries. Where multiple official (and de facto) languages of communication exist in a single country, then the GDP is split equally across the languages. (This works as countries where this condition hold are generally split evenly in terms of population and economic size by language.)

Source: IMF; national censuses (2016)

COMMUNICATION

Language is a means of interacting and communicating with others. The communication opportunity captures how language facilitates social interactions with other humans.

Native speakers: The total number of native speakers of a language.

It may include people who are (fully) bilingual and thus may double count people across languages.

Source: National censuses (most recent)

Second language (L2) speakers:* The total number of people who speak a language as a functional second language.

This excludes, for example, students who (casually) study a language in an academic setting. It is primarily a measure of people who regularly communicate in a second language. For example, native Russian speakers who live in Ukraine and speak Ukrainian at a proficient level (typically in interacting with native Ukrainian speakers within their country).

Source: Wikipedia

Family size:* The total number of native speakers from a given language family (e.g. Romance). It excludes, for a given language, its own count within its family in order to avoid double counting.

Knowledge of a particular language enables speakers to pick up the language of similar/related languages. For example, Italian and Spanish speakers are able to understand much of each other (spoken and written) even without formal training in the other language. Likewise, a native Cantonese speaker (who is literate in Cantonese) is able to read a Mandarin-language newspaper.

The family grouping is usually chosen at a level where there is some mutual intelligibility. Often this can be found 2 or 3 levels below the standard family grouping. For example, Polish belongs to the Indo-European family, which further branches out to Balto-Slavic, then Slavic, then West Slavic, then Lechitic, then Polish. For Polish, Slavic is considered the family.

This variable also takes into account cases where languages may share a common but unique script. For instance, many Chinese characters are used in Japanese (known as kanji). Because kanji is one of three script systems in the Japanese language, the Chinese language family is given just 1/3 weight in Japanese (and vice versa).

Source: Wikipedia

KNOWLEDGE & MEDIA (K&M)

One of the primary purposes of a language is to enable the consumption of media, and to share ideas and knowledge.

Internet content: The share of Internet sites and traffic displayed in a given language.

Much of the media and knowledge we consume is on the Internet. Although there are services that translate web content, the translations are often not great and browsers seldom will surf in languages they do not know.

Source: W3Techs.com (2015)

Feature films:* The total number of feature films produced in a country.

It is a reflection of the soft power of a language and the extent to which the language is consumed and shared as a medium of entertainment.

Source: Unesco Institute for Statistics (2014)

Top-500 universities: The number of top-500 universities as defined by ARWU (Shanghai JT University).

Note that the language of instruction for a university might not necessarily correspond with the dominant, minority or official language of a country. This would tend to undercount the influence of English, which is often the medium of instruction even in countries where it is not widely spoken.

Source: Shanghai Jia Tong University – Academic Ranking of World Universities (2015)

Academic journals:* The number of peer-reviewed academic journals produced in a given language.

There may be cases where a journal has more than one language of publication (e.g. some Canadian journals that are bilingual English-French) so the total count by languages may exceed the total population of journals.

Source: Wikipedia

DIPLOMACY

The diplomacy opportunity reflects which languages are used in global settings at supranational organisations (SNOs). Most such institutions have a combination of official and working languages (i.e. languages for official communication and work). Many SNOs have multiple official languages but typically have a single *de facto* working language (almost always English).

IMF: This is an indicator variable that takes on the value 1 if a language is an official language of the International Monetary Fund or for which official communication is translated in a given language, and 0 otherwise. The IMF has 7 official/working languages: English (official), Arabic, Chinese (Mandarin), French, Japanese, Russian and Spanish.

UN: This is an indicator variable that takes on the value 1 if a language is an official language of the United Nations and 0 otherwise. The UN has 6 official/working languages: Arabic, Chinese (Mandarin), English, French, Russian and Spanish.

WB: This is an indicator variable that takes on the value 1 if a language is an official language of the World Bank and 0 otherwise. The WB has 6 official/working languages: Arabic, Chinese (Mandarin), English, French, Russian and Spanish.

Index of 10 organisations: This is an index comprised of 10 indicator variables. Each of these takes on the value 1 if a language is an official/working language of the institution and 0 otherwise. The index is thus created by summing these values and dividing by the total number of institution (10), thus creating an index indicator whose value ranges from 0 to 1. The 10 organisations in the index are:

1. Bureau International des Expositions (BIE)
2. Fédération Internationale de Football Association (FIFA)
3. International Criminal Court (ICC)
4. International Labour Organisation (ILO)
5. International Olympic Committee (IOC)
6. Inter-Parliamentary Union (IPU)
7. International Telecommunication Union (ITU)
8. Organisation for Economic Cooperation and Development (OECD)
9. Universal Postal Union (UPU)
10. World Trade Organisation (WTO)

MISSING VALUES

Where data are missing they are taken to be zero. This is generally inconsequential as data which are missing are generally from sources where the likely true values are indeed zero or close to zero. Indeed, for the indicators in the index, missing data typically affect least developed countries which generally have no or low values for the associated variables.

OMITTED VARIABLES

The index strives to be parsimonious and included indicators were specifically chosen to be cardinal variables (so that the index itself is a cardinal measure). For sure there are many indicators not included whose inclusion would help make the index more robust. For example, size and geographic reach of expat and diaspora communities, etc. An obvious limitation is the availability of such data.

TRANSFORMATION

All 20 of the indicators in the index are non-negative cardinal-valued variables. They are all “positive” in that higher values are associated with more language efficacy. They are transformed to a score in the [0,1] range by dividing the raw indicator value by the maximum value in the sample: $s_j = x_j / x_{\max}$. (So for each indicator there will be at least one language that takes on the maximal score of one.)

AGGREGTION

The transformed scores of the 20 indicators are aggregated into an index score for a country by summing the weighted scores of the indicators: $I = \sum_j \omega_j s_j$. Since each transformed score takes a value in the range [0,1] it follows that the aggregated index score also falls in the range [0,1].

Note that the aggregated final index score is a cardinal measure. This is because all the inputs are themselves cardinal and transformed in a way that preserves cardinality.