

The CJK Dictionary Institute

日中韓辭典研究所

*The world's prime source of
CJK and Arabic dictionary data*



www.cjk.org

The CJK Dictionary Institute

日中韓辭典研究所

The CJK Dictionary Institute, Inc. (CJKI) specializes in CJK and Arabic computational lexicography. The institute creates and maintains CJK (Chinese, Japanese and Korean) and Arabic lexical databases currently covering over 24 million entries. Located in Saitama, Japan, CJKI is headed by Jack Halpern, editor-in-chief of the world-renowned New Japanese-English Character Dictionary and of various other CJK dictionaries.

CJKI plays a leading role in helping the IT industry penetrate the lucrative East Asian market by providing software developers with high quality dictionary data. This includes comprehensive databases of general vocabulary, proper nouns and technical terms for CJK languages, including Chinese dialects such as Cantonese and Hakka. CJKI also maintains databases and romanization systems of Arabic proper nouns, a large-scale Spanish-English dictionary, and various multilingual databases of proper nouns and geographic data.

CJKI has become one of the world's prime sources for CJK lexical resources. It is contributing to CJK and Arabic information processing technology by providing high-quality lexical resources and professional consulting services to some of the world's leading software developers and IT companies, including Fujitsu, Sharp, Sony, IBM, Google, Microsoft, Yahoo, Amazon and Baidu.

How CJKI's resources are used

CJKI's team of professional editors and software engineers use advanced computational lexicography methods to compile and maintain comprehensive lexical databases and dictionaries that include a variety of features for a broad gamut of applications, such as:

- Natural language processing applications such as information retrieval tools, search engine technology and morphological analyzers.
- Anti-money laundering and fraud detection.
- Security applications such as criminal watch lists.
- CJK input method editors (front-end processors).
- Machine translation and online translation tools.
- Speech technology applications, both text-to-speech and automatic speech recognition.
- Geographical data for multilingual maps, machine translation and tokenization.
- Conversion between Simplified and Traditional Chinese.
- Electronic dictionaries for desktop and mobile platforms .
- Pedagogical, linguistic and computational lexicography research.
- Transcription and transliteration applications.
- Data cleansing.

Doing business with CJKI

CJKI has a flexible business model that is decided on a case-by-case basis to suit the convenience of the customer. We are not "resellers," nor are we "data vendors" -- we are a linguistic institute, and create the data ourselves based on several decades of experience and extensive knowhow of CJK and Arabic lexicography.

Our fundamental policy is to customize our databases to the specific requirements of the customer at no extra charge. To achieve this, we study our customers' needs in-depth and prepare a data package that meets the customer's precise needs. We also build custom databases from scratch. We have extensive experience in putting together teams to compile large-scale dictionaries in a short period of time, using our sophisticated tools for automating the compilation process, which significantly reduces costs to the customer.

It is important to note that the benefits of working with CJKI go well beyond cost. We are flexible in matters of format, delivery dates and business model, work hard to gain an in-depth understanding of the customer's needs, and provide excellent service that includes a reasonable amount of free technical and linguistic consulting as well as free minor upgrades. Licensing data from CJKI is not merely "buying" data -- it is entering into a close relationship that ensures constant advice, technical/linguistic support, upgrades, and reasonable fees.

CJKI's Lexical Resources

CJKI's extensive CJKI and Arabic lexical resources currently cover over 24 million entries, used by major portals and software developers in a wide variety of applications. Our main resources include:

- Bilingual dictionaries
- Multilingual dictionaries
- Arabic personal names
- Proper nouns and geographical data
- Technical terminology
- Monolingual lexical databases
- Phonetic and phonological databases
- Mapping tables for Chinese conversion
- Morphological databases
- Lexical databases

In addition to the resources described above, CJKI has developed resources containing millions of more entries for the following:

- Arabic transcription and vocalization systems
- Arabic, Japanese, and Spanish-English full-form lexicons*
- Arabic Place names
- Databases for input method editors
- Frequency statistics based on web and corpora
- Database for CJK IMEs

* "Full form lexicon" refers to a comprehensive lexical database that includes every single inflected form (verb conjugations, plurals, etc.) and declined forms (case endings) of a language. Each full form lexicon contains millions of entries accompanied by a rich set of grammatical attributes.

Chinese Lexical Resources

汉语词汇资源

CJKI's comprehensive Chinese lexical resources currently include nearly 10 million entries, covering general vocabulary, technical terminology, proper nouns, company and organization names, and others, in both **Simplified Chinese** (SC) and **Traditional Chinese** (TC), used in such applications as machine translation (MT), information retrieval (IR) and input method editors (IME). They includes a rich set of grammatical, phonological and semantic attributes, including *pinyin* and *zhuyin* readings, part-of-speech codes, frequency of occurrence statistics, and others.

Below is a description of CJKI's principal Chinese lexical resources.

Principal Resources

- Simplified Chinese-English Dictionary
- English-Simplified Chinese Dictionary
- Chinese-English Database of Proper Nouns
- Database of Chinese Name Variants
- Chinese-English Technical Terms
- Chinese-Japanese Technical Terms
- Chinese Dictionary of Computer Terms
- Chinese Lexical Database
- Chinese Pinyin Database
- Chinese to Chinese Conversion
- Hanzi-Pinyin Transcription System
- Chinese Morphological Database
- Chinese Lexical Frequency Statistics
- English-Traditional Chinese Dictionary
- Chinese IME Databases

Chinese-English Dictionary

汉英词典简体版

CJKI's **Simplified Chinese-English Dictionary** (SCED) is the most comprehensive Chinese dictionary available today. Covering over 700,000 entries of general vocabulary, technical terms, important proper nouns and example sentences, SCED was compiled in collaboration with lexicographers from a leading Chinese university on the basis of the most authoritative dictionaries published in China. This dictionary, which is without peer, has undergone extensive proofreading and validation by a team of native Chinese editors. It is ideally suited for:

- Machine translation dictionaries
- Cross-language information retrieval
- Handheld electronic dictionaries
- Mobile device applications

Simplified Chinese-English Dictionary

CHINESE	POS	PINYIN	ENGLISH
国丧	N	guósāng	national mourning
国乐	N	guóyuè	traditional Chinese music
国书	N	guóshū	letter of credence; credentials; letter of commission
国书付本	N	guóshū fùběn	copy of credentials
国产设备	N	guóchǎn shèbèi	home equipment
国产货	N	guóchǎnhuò	domestic products; domestic goods
国产轿车	N	guóchǎn jiàochē	domestic-made cars
国产影片	N	guóchǎn yǐngpiàn	Chinese film
国产税	N	guóchǎnshuì	excise duties
国产装备	N	guóchǎn zhuāngbèi	Chinese-made equipment
国产品	N	guóchǎnpǐn	home products; national products; domestic products
国优产品	N	guóyōu chǎnpǐn	national quality product
国债	N	guózhài	national debt; government loan; public debt; national bonds

English-Chinese Dictionary

英 汉 词 典 简 体 版

CJKI's **English-Simplified Chinese Dictionary** (ESCD) covers about 100,000 entries including general vocabulary and important proper names. Optimized for the convenience of users of electronic dictionaries, ESCD has just the right amount of detail: enough equivalents to give an in-depth understanding, yet short enough not to clutter up the screen. ESCD is being used in such well-known translation tools like Babylon and Quicktionary, as well as on various mobile platforms around the world including TangoTown in Japan and Australia.

English-Simplified Chinese Dictionary

ENGLISH	POS	CHINESE
canoe	v. tr.	用独木舟载运
cay	n.	岩礁, 沙洲, 珊瑚礁
cheddar	n.	切德干酪
clad	v. tr.	电镀
codger	n.	怪人; 有怪癖的人
Comoros	NP	科摩罗
congruence	n.	适合, 相合性, 一致
couch	v. intr.	躺着; 埋伏; 蹲着
crew cut	n.	平头发式
cutthroat	adj.	残酷的; 杀人的
decimal	n.	小数
depersonalize	v. tr.	使失去人性
dial	v. tr.	调; 拨; 收听, 收视; 打电话给
discrepancy	n.	相差; 矛盾; 差异
divide	n.	分歧, 不和; 分水岭
dramatize	v. intr.	戏剧化; 可改编成剧本; 举止夸张
emotional	adj.	情绪的; 情感的

An English-Traditional Chinese Dictionary is also available.

Chinese↔English Proper Nouns

汉英英汉专有名词数据库

CJKI's **Chinese↔English Database of Proper Nouns** (CEP) is very comprehensive, covering millions of entries in both Simplified and Traditional Chinese. It includes various data fields such as pinyin, zhuyin, frequency rankings, classification codes, locale codes and English equivalents. Included are a large variety of both Chinese and non-Chinese name types, such as:

- Place names
- Personal names (surnames and given names)
- Companies and organizations
- Facilities and points of interest
- Western personal and place names
- Miscellaneous, such as periodicals and abbreviations

Chinese Personal Names

ID	TYPE	SIMPLIFIED CHINESE	TRADITIONAL CHINESE	PINYIN
N045347	G	桂花	桂花	guìhuā
N002382	S	鄂	鄂	è
N070261	G	尔和	爾和	ěrhé
N000370	S	戚	慼	qī
N129913	G	联谊	聯誼	liányì
N018980	G	亚军	亞軍	yàjūn
N044924	G	军营	軍營	jūnyíng
N001532	GS	耽	耽	dān
N001141	G	嗣	嗣	sì
N001079	G	耕耘	耕耘	gēngyún
N061246	G	庄稼	莊稼	zhuāngjia
N132248	G	和全	和全	héquán
N002101	S	依	儂	nóng
N065668	G	之遥	之遙	zhīyáo
N139754	S	刁	刁	diāo
N065194	S	司马	司馬	sīmā
N001576	G	津津	津津	jīnjīn

Chinese and non-Chinese Place Names

ID	ENGLISH	SC	L/O	TC	PINYIN	ZHUYIN
N002657	Aruba	阿鲁巴	L	阿盧巴	ālǔbā	ㄚㄌㄨˇㄅㄚ
N001635	Azerbaijan	阿塞拜疆	L	亞塞拜然	āsāibàijiāng	ㄚㄧㄝˋㄇㄟㄘㄟㄣˊ
N081006	Brasilia	巴西利亚	O	巴西利亞	bāxīliya	ㄅㄚㄒㄧˊㄌㄧㄚˋ
N016658	Caracas	加拉加斯	L	卡拉卡斯	jiālājīāsī	ㄐㄚㄌㄚˊㄐㄚㄙㄧ
N014214	Cairo	开罗	O	開羅	kāiluó	ㄎㄞㄌㄨㄛˊ
N058842	Chad	乍得	L	查德	zhàdé	ㄓㄚˋㄉㄟˊ
N078960	Fukuoka	福岡	O	福岡	dōngyángshì	ㄉㄨㄥˊㄧㄤˊㄕㄟˋ
N047517	Georgia	乔治亚	O	喬治亞	fúgāng	ㄈㄨˊㄍㄤ
N023778	Guinea	几内亚	O	幾內亞	qiáozhìyà	ㄑㄧㄠˊㄓㄧˋㄚˋ
N031561	Haiyan	海盐	O	海鹽	jǐnèiyà	ㄐㄞˋㄋㄟˋㄚˋ
N036150	Hanyang	汉阳	O	漢陽	hǎiyán	ㄏㄞˋㄧㄢˊ
N032756	Heshan	鹤山	O	鶴山	hànyáng	ㄏㄚˋㄧㄤˊ
N032307	Huailai	怀来	O	懷來	hèshān	ㄏㄟˋㄕㄞˊ
N000617	Ireland	爱尔兰	O	愛爾蘭	huáilái	ㄏㄞˋㄌㄞˊ

L : lexemic mapping O : orthographic mapping
(see **Chinese Dictionary of Computer Terms** for detail)

Chinese Name Variants

汉语人名罗马字异形数据库

The number of Chinese personal names and their variants is very large -- in the millions -- which makes it difficult to identify them and process them. Named Entity Recognition (NER) technology is a hot topic in computational linguistics. To enhance NER technology, CJKI maintains databases of several million CJK and Arabic name variants in all major and most minor romanization systems.

There are several well-established systems for romanizing Chinese, such as Hanyu Pinyin, Wade-Giles, Yale, and Tongyong Pinyin, as well as various popular ones and many older ones that have fallen out of use. Chinese has seven major dialect groups, and another four minor ones. The CJKI **Database of Chinese Name Variants** (CNV) includes Chinese personal names in all the standard and dialectal Chinese romanization systems, covering all the major dialects, including Cantonese, Hakka and Hokkien and including classification codes and frequency of occurrence statistics.

Chinese Name Variants

CHINESE	PINYIN	ZHUYIN	ENGLISH	TONGYONG	YALE	WADE-GILES	VARIANTS
百欣	bǎixīn	ㄅㄞˋ ㄒㄩㄣ	Baixin	Baisin	Baisyin	Paihsin	Paisin
白	bái	ㄅㄞˊ	Bai	Bai	Bai	Pai	
北强	běiqiáng	ㄅㄟˋ ㄑㄩㄤˊ	Beiqiang	Beiciang	Beichyang	Peich'iang	Peits'iêng Peichiang Peitsiêng
炳章	bǐngzhāng	ㄅㄩㄥˋ ㄓㄤ	Bingzhang	Bingjhang	Bingjang	Pingchang	
宝程	bǎochéng	ㄅㄞˇ ㄔㄥˊ	Baocheng	Baocheng	Baucheng	Paoch'eng	Paocheng
爱华	àihuá	ㄞˋ ㄏㄨㄚˊ	Aihua	Aihua	Aihwa	Aihua	Ngaihua
伯芝	bózhī	ㄅㄛˊ ㄓ	Bozhi	Bojhih	Bwojr	Pochih	
长流	chángliú	ㄔㄤˊ ㄌㄩˊ	Changliu	Changliou	Changlyou	Ch'angliu	
邦达	bāngdá	ㄅㄤ ㄉㄚˊ	Bangda	Bangda	Bangda	Pangta	
曹	cáo	ㄘㄠˊ	Cao	Cao	Tsau	Ts'ao	
冰晓	bīngxǎo	ㄅㄩㄥ ㄒㄠˇ	Bingxiao	Bingsiao	Bingsyau	Pinghsiao	Pingsiao
百成	bǎichéng	ㄅㄞˇ ㄔㄥˊ	Baicheng	Baicheng	Baicheng	Paich'eng	Paicheng

Chinese↔English Computer Terms

汉英英汉计算机术语词典

CJKI's **Chinese↔English Dictionary of Computer Terms** (ECCT) is an English-Chinese Chinese-English dictionary containing about 100,000 Simplified Chinese (SC) and 100,000 Traditional Chinese (TC) entries, including acronyms. This dictionary covers both SC, used in The People's Republic of China and Singapore, and TC, used in Taiwan, Hong Kong and among overseas Chinese. It has several features that distinguish it from any other Chinese computer dictionary available today.

- Covers about 100,00 entries selected on the basis of frequency statistics.
- Constantly updated and expanded to include recent terms.
- Contains more than 10,000 acronyms cross-referenced to the expanded forms.
- Linguistically accurate TC equivalents (explained below).

The above features make this dictionary an invaluable tool for translators and for use in various IT applications such as information retrieval, machine translation, and input method editors.

The TC in ECCT is not merely a code-conversion of the SC, but has been carefully proofread to ensure accuracy both on the orthographic and lexemic levels. An example of orthographic conversion, marked "O" in the table below, is 目录 'directory' converted to 目錄. An example of lexemic conversion, marked "L" in the table below, is 计算机 converted to 電腦 in TC.

Chinese-English Computer Terms

ENGLISH	SIMPLIFIED	TRADITIONAL	TYPE
file	文件	檔案	L
Internet	因特网	網際網路	L
program	程序	程式	L
CD-ROM	光盘	光碟	L
information	信息	資訊	L
computer network	计算机网络	電腦網路	L
modulator/demodulator	调制解调器	調變解調器, 數據機	L
modem	调制解调器	調變解調器, 數據機	L
computer software	计算机软件	電腦軟體	L
database	数据库	資料庫	L
flowcharting	流程图编制	繪製流程圖	L
expert system	专家系统	專家系統	O
directory	目录	目錄	O

Chinese↔English Technical Terms

汉英英汉科技术语词典

CJKI's **Chinese↔English Dictionary of Technical Terms** (CE-TERM) covers approximately five million terms from over 90 domains, including computers/IT, mechanical engineering, biotechnology, chemistry, and medicine.

Because of the dramatic growth of the Chinese economy, there is an urgent need for Chinese-English technical dictionaries in electronic form. This has prompted us to greatly expand our Chinese-English Dictionary of Technical Terms, based on our extensive experience in compiling our Japanese-English Dictionary of Technical Terms over the last decade.

Our technical terminology dictionaries are being used in a wide variety of applications, including:

- Machine translation dictionaries.
- Information retrieval systems.
- NLP tools like morphological analyzers and tokenizers.
- Dictionaries on handheld electronic dictionaries and CD-ROMs.
- Dictionaries for smartphone platforms such as iPhone and Android.

Chinese-English Technical Terms

DOMAIN	CHINESE	ENGLISH	PINYIN
化学	盐酸普马嗪	promazine hydrochloride	yánsuānpǔmǎqín
医学	类天花	pseudovariola	lèitiānhuā
建設	净水厂	purification plant	jìngshuǐchǎng
化学	盐酸奎宁	quinine hydrochloride	yánsuānkúíning
機械	兰金循环	Rankine cycle	lánjīnxúnhuán
電気	换向	rectify	huànxìàng
機械	紧固	reinforce	jǐngù
建設	刚性梁	reinforcing beam	gāngxíngliáng
医学	肾清除率	renal clearance	shènníngchúlǜ
医学	肾性佝偻病	renal rickets	shènxìnggōulóubìng
生物	环形染色体	ring chromosome	huánxíngǎnsètǐ
機械	环形扳手	ring spanner	huánxíngbǎnshǒu
医学	类丹毒	Rosenbach's disease	lèidāndú
金属	氧化铁抛光粉	rouge	yǎnghuàtiě pāoguāngfěn
化学	盐酸橡胶	rubber hydrochloride	yánsuānxiàngjiāo
生物	环血病	scurvy	huánxiěbìng

Chinese↔Japanese Technical Terms

汉日日汉科技术语词典

CJKI's **Chinese↔Japanese Dictionary of Technical Terms** (CJ-TERM) covers approximately 750,000 terms from 20 domains, including computers/IT, mechanical engineering, biotechnology, chemistry, and medicine.

Because of the rapidly growing trade relations between China and Japan, there has been a growing need for Chinese-Japanese technical dictionaries in electronic form. This has prompted us to greatly expand our Chinese-Japanese Dictionary of Technical Terms) based on our extensive experience in compiling our Japanese-English Dictionary of Technical Terms over the last decade.

Our technical terms dictionaries are being used in a wide variety of applications, including:

- Machine translation dictionaries.
- Information retrieval systems.
- NLP tools like morphological analyzers and tokenizers.
- Dictionaries on handheld electronic dictionaries and CD-ROMs.
- Dictionaries for smartphone platforms such as the iPhone and Android.

Chinese-Japanese Technical Terms

DOMAIN	CHINESE	JAPANESE	PINYIN	HIRAGANA
電気	传真	ファックス	chuánzhēn	ふあつくす
生物	丝心蛋白	フィブロイン	sīxīndànbái	ふいぶろいん
生物	丝状噬菌体	線状ファージ	sīzhuàngshìjūntǐ	せんじょうふあーじ
医学	类丹毒	類丹毒	lèidāndú	るいたんどく
化学	粒化	粒状化する	lìhuà	りゅうじょうかする
金属	卤化物	ハロゲン化物	lǔhuàwù	はろげんかぶつ
金属	卤素	ハロゲン	lǔsù	はろげん
金属	卤化物	ハロゲン化物	lǔhuàwù	はろげんかぶつ
建設	凿岩机	ハンマードリル	záoyánjī	はんまーどりる
化学	临氢重整	ハイドロホーミング	línqīngchóngzhěng	はいどろほーみんぐ
化学	亲水胶体	親水コロイド	qīnshuǐjiāotǐ	しんすいころいど
生物	亚氨基酸	イミノ酸	yànjīsuān	いみのさん

Chinese-English Technical Terms

DOMAIN	ENTRIES
Science (technical)	554,569
Medical	537,173
Science (general)	533,412
Chemical Engineering	218,869
Computer	195,487
Shipping	189,979
Shipbuilding	176,115
Aerospace	157,273
Energy	156,096
Mechanical	134,487
Business	133,321
Exports	128,450
Electric	117,986
Other	1,561,915
Total	4,795,132

Chinese-Japanese Technical Terms

DOMAIN	ENTRIES
Computer/IT	108,637
Mechanical Engineering	56,200
Chemistry	43,467
Medical	41,269
Life Sciences	35,554
Electricity	19,319
Electronics	19,141
Geology	17,937
Physics	16,214
Metallurgy	14,635
Other	376,630
Total	749,003

Chinese Lexical Database

汉语词汇数据库

The **CJKI Chinese Lexical Database** (CLD) is a comprehensive monolingual lexical database of Chinese consisting of the Simplified Chinese Lexical Database (CLD-SC) and the Traditional Chinese Lexical Database (CLD-TC) modules. Developed by CJKI's team of experienced Chinese editors and linguists over many years, the CLD is a significant contribution to the field of Chinese lexicography. CLD is especially suitable for applications in the fields of information retrieval, morphological analysis, machine translation and various natural language processing (NLP) applications, and is being used by various IT companies to enhance their Chinese segmentation technology.

Chinese Lexical Database

ID	POS	TYPE	CHINESE	PINYIN	RANK	WEB RANK
S0185240Aa	NP	G	东霞	dōngxiá	C	000205863
S0185241Aa	NP	P	东会村	dōnghuìcūn	C	000331481
S0000306Ac	NC		东海	dōnghǎi	A	000009255
S0000306Aa	NP	G	东海	dōnghǎi	A	000009255
S0000306Ab	NP	P	东海	dōnghǎi	A	000009255
S0185242Aa	NP	P	东海县	dōnghǎixiàn	C	000078031
S0000307Aa	E		东海扬尘	dōnghǎiyángchén	C	000263750
S0000308Aa	E		东海捞针	dōnghǎilāozhēn	C	000124028
S0430527Ax	U		东海舰队	dōnghǎijiànduì		000064698
S0000309Aa	E		东海桑田	dōnghǎisāngtián	C	000090763
S0000310Aa	NP	Oe	东海大学	dōnghǎidàxué	C	000069472
S0185243Aa	NP	P	东外大街	dōngwàidàjiē	C	000166158
S0000311Ab	NC		东郭	dōngguō	C	000069927
S0000311Aa	NP	S	东郭	dōngguō	C	000069927
S0000312Aa	E		东郭先生	dōngguōxiānshēng	C	000101330
S0000313Aa	NC		东郭履	dōngguōlǚ	C	000267748
S0185244Aa	NP	P	东革新里	dōnggéxīnlǐ	C	000234655
S0000314Ab	NC		东岳	dōngyuè	C	000065236
S0000314Aa	NP	G	东岳	dōngyuè	C	000065236

Chinese Pinyin Database

汉语拼音数据库

The CJKI Chinese Pinyin Database (CPD) contains several million Simplified Chinese (SC) and Traditional Chinese (TC) headwords covering general vocabulary, technical terms, and proper nouns. Each lexeme is accompanied by pinyin readings for SC and both pinyin and zhuyin (not shown here) for TC. What is especially noteworthy is that the pinyin/zhuyin readings take into account the differences in pronunciation between Taiwan and the PRC, as shown in the table below. Even highly educated native Chinese speakers are often surprised to discover that such differences exist.

An important feature of this database is its high accuracy, and explicit indication of the neutral tone, which is often ignored by conventional dictionaries. The data can be provided in all the major transcription systems such as Yale, Wade-Giles, and Tongyong Pinyin. An IPA edition, especially useful for speech technology applications such as TTS, is now under development.

The Diff field below indicates whether pairs of SC-TC equivalents have identical pinyin. "D" indicates that pinyin is different; "S" indicates that pinyin is the same.

Chinese Pinyin Database

ID	DIFF	SC HANZE	SC FREQUENCY	SC PINYIN	TC HANZI	TC FREQUENCY	TC PINYIN
I00001	D	临期	0000029000	línqī	臨期	0000028800	línqí
I00003	D	企业	0163000000	qīyè	企業	0102000000	qiyè
I00004	D	倬雄	0000000167	zhuōxióng	倬雄	0000000167	zhuóxióng
G00018	S	咖啡豆	0000779000	kāfēidòu	咖啡豆	0000779000	kāfēidòu
I00006	D	危险	0022400000	wēixiǎn	危險	0003080000	wéixiǎn
I00008	D	埧城	0000000411	lièchéng	埧城	0000000411	lèchéng
I00009	D	夕日	0002020000	xīrì	夕日	0002020000	xìrì
I00010	D	大期	0000061500	dàqī	大期	0000061500	dàqí
I00012	D	帆柱	0000030600	fānzhù	帆柱	0000030600	fānzhù
I00014	D	微笑	0018400000	wēixiào	微笑	0018400000	wéixiào
I00016	S	无着	0000265000	wúzuó	無著	0000265000	wúzuó
G00027	D	咖喱粉	0000087400	gālífěn	咖喱粉	0000087400	kālífěn
I00017	D	昔日	0004880000	xīrì	昔日	0004880000	xìrì
I00018	D	显微镜	0003390000	xiǎnwēijìng	顯微鏡	0000228000	xiǎnwéijìng
I00019	D	期待	0059100000	qídài	期待	0059100000	qídài
G04357	D	咖喱饭	0000122000	gālífàn	咖喱飯	0000122000	kālífàn
I00020	D	池穴	0000059400	chíxué	池穴	0000059400	chíxuè
I00021	D	理发	0002170000	lǐfà	理髮	0000495000	lǐfà

Chinese-to-Chinese Conversion

中文簡繁轉換

A common fallacy is that there is a straightforward correspondence between Simplified Chinese (SC) and Traditional Chinese (TC), and that conversion between the two merely requires mapping from one character set to another. In fact, code-conversion from SC to TC will often lead to errors both on the *orthographic* and *lexemic* levels. An example of orthographic conversion is 头发 'hair' converted to 頭髮, in which 頭 and 髮 are the traditional equivalents of 头 and 发 respectively. An example of lexemic conversion is SC 激光 'laser' converted to 雷射 in TC, a distinct word of identical meaning.

CJKI ranks among the world's foremost experts on Simplified to/from Traditional Chinese conversion, and has in-depth knowledge of Chinese segmentation issues, having collaborated with Chinese universities such as Beijing Language and Culture University. Our comprehensive SC to/from TC mapping tables, developed over a period of about 12 years, have several million entries, the largest in existence.

The table below illustrates lexemic mappings of computer terms between SC and TC.

Chinese to Chinese Conversion Technology

ENGLISH	SIMPLIFIED	TRADITIONAL
File	文件	檔案
CD-ROM	光盘	光碟
Data	数据	資料
Compatibility	兼容性	相容性
Information	信息	資訊
Software	软件	軟體
Message	消息	訊息
Camera	摄像机	攝影機
Recording/Burning	刻录	錄製
Drive	驱动器	光碟機
Audio frequency	音频	音訊
Memory	存储	儲存
Video frequency	视频	視訊
Compatible	兼容	相容
Rewritable	可擦写	可重寫
Optical drive	光驱	燒錄機

Spanish Full Form Lexicon

CJKI's **Spanish Full Form Lexicon** (S-FULEX) is part of an eight-year effort by CJKI's team of lexicographers and software engineers to develop full form lexicons for Spanish, Arabic and Japanese. To this end, we have analyzed the grammar and morphology of these languages in great depth, to a degree well beyond found in comprehensive descriptive grammars for these languages.

S-FULEX contains over 26 million records, and is designed to cover general vocabulary. A distinctive feature of S-FULEX is that it is fully bilingual, so that every inflected form is normally given multiple equivalents in the target language. Other important features include:

- All inflected and declined wordforms are included.
- Rich set of attributes, such as conjugation patterns and orthographic variants.
- Over 46 data fields with supplementary information for each entry.
- Detailed part-of-speech and other grammatical codes.

S-FULEX is extremely comprehensive. Our mission was to create the largest Spanish-English lexicon for general vocabulary in existence. S-FULEX surpasses the coverage of the most comprehensive Spanish dictionaries ever published, including the prestigious *Diccionario de la Lengua Española* published by the Real Academia Española.

Spanish Full Form Lexicon

SOURCE	TARGET	POS
(yo) hablo	(I) speak	V
(yo) hablo	(I) am speaking	V
(yo) hablo	(I) spoke	V
(yo) hablo	(I) will speak	V
(yo) hablo	(I) shall speak	V
(yo) hablo	(I) would have spoken	V
(yo) hablo	(I) am going to speak	V
(yo) hablo	(I) talk	V
(yo) hablo	(I) am talking	V
(yo) hablo	(I) will talk	V
(yo) hablo	(I) shall talk	V
(yo) hablo	(I) would have talked	V
(yo) hablo	(I) am going to talk	V

The above 13 entries comprise a very small part of the 3,429 entries for *hablar* contained in S-FULEX.

Multilingual Technical Terms

多语言术语词典

Because of the rapidly growing trade relations between China, Japan and the English speaking world, there is an urgent need for Chinese-Japanese-English technical dictionaries in electronic form. CJKI's Multilingual Dictionary of Technical Terms (MDT) is a comprehensive trilingual dictionary of technical terms in Simplified Chinese, Japanese and English covering about 300,000 entries in all major fields of science and technology. A sister edition, the Chinese-English Dictionary of Technical Terms, is under development in collaboration with Chinese institutions and is expected to cover several million entries.

Multilingual Technical Terms

CHINESE	JAPANESE	ENGLISH	PINYIN	HIRAGANA
加油	給油	fueling, lubrication	jiāyóu	きゅうゆ
加油	注油	lubrication, oiling	jiāyóu	ちゅうゆ
加油车	給油車	refueling truck	jiāyóuchē	きゅうゆしゃ
加油器	給油器	feeder, lubricator, oiler	jiāyóuqì	きゅうゆき
加油器	油差し	lubricator, oil can, oil feeder, syringe	jiāyóuqì	あぶらさし
加油孔	油穴	lubrication hole, oil hole	jiāyóukǒng	あぶらあな
加油站	給油所	filling station, service station	jiāyóuzhàn	きゅうゆしょ
加油站	フィリングスタンド	filling stand	jiāyóuzhàn	ふいりんぐすたんど
加硫	加硫	sulfurization, thionation, vulcanization, curing	jiāliú	かりゅう
加亮	強調	highlighting	jiāliàng	きょうちょう
加亮	強調表示	highlighting	jiāliàng	きょうちょうひょうじ
加亮	高輝度表示	highlighting, brightening	jiāliàng	こうきどひょうじ
加亮	ハイライト表示	highlighting	jiāliàng	はいらいとひょうじ
加力燃烧室	アフターバーナー	afterburner	jiālìránshā oshì	あふたーばーなー
加和性	加成性	additive property, additivity	jiāhéxìng	かせいせい
加气剂	空気連行剤	air entraining agent	jiāqìjì	くうきれんこうざい
加聚物	付加重合体	addition polymer, addition resin	jiājùwù	ふかじゅうごうたい
加勒金法	ガレルキン法	Galerkin method	jiālèjīnfǎ	がれるきんほう

Multilingual Proper Nouns

多言語固有名詞データベース

CJKI maintains comprehensive databases of CJK and Arabic personal names and place names, including various kinds of geographical data covering millions of entries. These databases are used by some of the world's major IT companies for a wide variety of applications such as online multilingual maps, named entity recognition, machine translation and information retrieval. The Multilingual Database of Proper Nouns (MPN) brings together six languages -- Simplified Chinese (SC), Traditional Chinese (TC), Japanese, Korean and English, and Arabic -- in a multidirectional format.

Key Features

- The database includes various data fields (many not shown here), such as readings in pinyin and zhuyin, hiragana, romanization in all major and most important minor romanization systems, semantic classification codes and frequency rankings, locale codes, and other useful information.
- A unique feature is that it's important to note that the TC place names are not merely a code-conversion equivalent of the SC names, but are accurate on both the orthographic and the lexemic levels ("O" and "L" in the tables below). For example, New Zealand in SC is 新西兰 *Xīnxīlán* but in TC it is 紐西蘭 *Nǚxīlán*.
- Another unique feature is that SC and TC readings are distinguished. Thus the pinyin for SC 期荣 is *qíróng*, but for the TC 期榮 it is *qíróng*.

CJKEA Database of Place Names

ENGLISH	JAPANESE	SC	LO	TC	KOREAN	ARABIC
Aruba	アルーバ	阿魯巴	L	阿盧巴	아루바섬	أروبا
Brasilia	ブラジリア	巴西利亞	O	巴西利亞	브라질리아	برازيليا
Caracas	カラカス	加拉加斯	L	卡拉卡斯	카라카스	كراكاس
Cairo	カイロ	开罗	O	開羅	카이로	القاهرة
Chad	チャド	乍得	L	查德	차드	تشاد
Georgia	ジョージア	喬治亞	O	喬治亞	조지아	جورجيا
Ireland	アイルランド	爱尔兰	O	愛爾蘭	아일랜드	أيرلندا
Seoul	ソウル	首尔	O	首爾	서울	سيول
Seoul	ソウル	汉城	O	漢城	서울	سيول
Tel Aviv	テルアビブ	特拉維夫	O	特拉維夫	텔아비브	تل أبيب
Yemen	イエメン	也門	L	葉門	예멘	اليمن

Phonemic Transcriptions of CJKE Place Names 1

ENGLISH	JAPANESE	SC	TC	KOREAN	ARABIC
Aruba	あるーば	ālǔbāā	ālúbā	arupasǝm	aruba
Brasilia	ぶらじりあ	bāxīlìyà	bāxīlìyà	pŭrachilria	burazilia
Caracas	からかす	jiālājīāsī	kālākāsī	k'arak'asŭ	karakasu
Cairo	かいろ	kāiluó	kāiluó	k'airo	al-qahirah
Chad	ちゃど	zhàdé	chádé	ch'atŭ	tshad
Georgia	じょーじあ	qiáozhìyà	qiáozhìyà	chochia	jurjia
Ireland	あいるらんど	àiěrlán	àiěrlán	ailraentŭ	ayirlanda
Seoul	そうる	shǒuěr	shǒuěr	sŏul	siwul
Seoul	そうる	hànychéng	hànychéng	sŏul	siwul
Tel Aviv	てるあびぶ	tèlāwéifū	tèlāwéifū	t'elapipŭ	tallu-abib
Yemen	いえめん	yěmén	yěmén	yemen	al-yaman

CJKE Multilingual Database of Personal Names

ENG	JPN	SC	TC	KOR	LO	HIRAGANA	SC PIN	TC PIN	MOE
Abba	アッバ	阿巴	亞伯	아바	L	あっぱ	ābā	yàbó	apa
Abbas	アッバース	阿巴斯	阿巴斯	아바스	O	あっぱーす	ābāsī	ābāsī	apasū
Alberto	アルベルト	阿尔韦托	阿爾韋托	알베르토	O	あるべると	āěrwéituō	āěrwéituō	alperūt'o
Qirong	期榮	期榮	期榮	치롱	O	きえい	qīróng	qíróng	ch'irung
Akiko	暁子	晓子	曉子	아키코	O	あきこ	xiǎozǐ	xiǎozǐ	ak'ik'o
Akiko	顕子	显子	顯子	아키코	O	あきこ	xiǎnzǐ	xiǎnzǐ	ak'ik'o
Akiko	昭子	昭子	昭子	아키코	O	あきこ	zhāozǐ	zhāozǐ	ak'ik'o
Akira	明	明	明	아키라	O	あきら	míng	míng	ak'ira
Deng	登	登	登	덩	O	とう	dēng	dēng	tōng
Einstein	アインSTEIN	爱因斯坦	愛因斯坦	아인슈타인	O	あいんすたいん	àiyīnsītān	àiyīnsītān	ainsyut'ain
Ernest	アーネスト	欧内斯特	歐尼斯特	어니스트	L	あーねすと	ōunèisītè	ōunísītè	ōnisūt'ū
Gregg	グレッグ	格雷格	葛瑞格	그레그	L	ぐれっぐ	géléigé	gěruìgé	kūrekū
Greg	グレッグ	格雷格	葛瑞格	그레그	L	ぐれっぐ	géléigé	gěruìgé	kūrekū
Haiyang	海洋	海洋	海洋	하이양	O	かいよう	hǎiyáng	hǎiyáng	haiyang
Huaiyang	懷陽	怀阳	懷陽	화이양	O	かいよう	huáiyáng	huáiyáng	hwaiyang
Jack	ジャック	杰克	傑克	잭	O	じゃっく	jiékè	jiékè	chaek
Jackie	ジャッキー	杰基	傑基	재키	O	じゃっきー	jiéjī	jiéjī	chaek'I
Kennedy	ケネディ	肯尼迪	甘迺迪	케네디	L	けねでい	kěnnídí	gānnǎidí	k'eneti
Kaiyang	開陽	开阳	開陽	카이양	O	かいよう	kāiyáng	kāiyáng	k'aiyang
Nakajima	中島	中岛	中島	나카지마	O	なかじま	zhōngdǎo	zhōngdǎo	nak'achima
William	ウィリアム	威廉	威廉	빌리엄	O	ういりあむ	wēilián	wēilián	pilriam
Zhang	張	张	張	장	O	ちょう	zhāng	zhāng	chang

Meet Our CEO

Jack Halpern (春遍雀來), CEO of **The CJK Dictionary Institute**, is a lexicographer by profession. For sixteen years was engaged in the compilation of the *New Japanese-English Character Dictionary*, and as a research fellow at Showa Women's University (Tokyo), he was editor-in-chief of several kanji dictionaries for learners, which have become standard reference works.

Jack Halpern, who has lived in Japan over 35 years, was born in Germany and has lived in six countries including France, Brazil, Japan and the United States. An avid polyglot who specializes in Japanese and Chinese lexicography, he has studied 14 languages (speaks nine fluently) and has devoted several decades to the study of linguistics and lexicography.

Jack Halpern has published over twenty books and dozens of articles and academic papers, mostly on the Japanese writing system and CJK information processing, has given over 600 public lectures on Japanese language and culture, and has presented dozens of papers at international conferences.

On a lighter note, Jack Halpern loves the sport of unicycling. Founder and long-time president of the International Unicycling Federation, he has promoted the sport worldwide and is a director of the Japan Unicycling Association. Currently, his passion is playing the quena and improving his Chinese, Esperanto and Arabic.



Contact Information

The CJK Dictionary Institute, Inc.

Komine Building
34-14, 2-chome, Tohoku, Niiza-shi
Saitama 352-0001
JAPAN

Phone: +81-48-473-3508

Fax: +81-48-486-5032

Email: jack@cjk.org

Web: www.cjk.org