漢(株) 日中韓辭典研究所

The CJK Dictionary Institute, Inc.

本研究所の紹介

本研究所簡介

Brief Introduction

日中韓辭典研究所所長

日中韓辭典研究所所长

President of The CJKI

春遍雀來

Jack Halpern

Comprehensive CJK Lexical Database
包括的日中韓辞書データベース
中日韩辞典综合数据库

Description	Japanese	Chinese	
		Simplified	Traditional
General vocabulary	390,000	250,000	250,000
Katakana loanwords	50,000	-	-
Companies and organizations	600,000	55,000	55,000
Personal names	570,000	650,000	650,000
Place names	90,000	170,000	170,000
Famous people's names	30,000	60,000	-
Computer terminology	100,000	100,000	100,000
Technical terminology	1,200,000	100,000	100,000
Neologisms	25,000	15,000	-
Single characters	17,000	18,000	14,000
Orthographic variants	80,000	-	-
Bilingual English	120,000	85,000	85,000
Others	-	160,000	120,000
Total	3,272,000	1,663,000	1,544,000

Named Entity Contextual Clues

Headword	Reading	Example
センター	せんたー	国民生活**センター**
ホテル	ほてる	**ホテル**シオノ
駅	えき	朝霞**駅**
協会	きょうかい	日本ユニセフ**協会**

Multilingual Database of Place Names

English	Japanese	SC	LO	TC	Korean
Azerbaijan	アゼルバイジャン	阿塞拜疆	L	亞塞拜然	아제르바이잔
Caracas	カラカス	加拉加斯	L	卡拉卡斯	카라카스
Cairo	カイロ	开罗	O	開羅	카이로
Chad	チャド	乍得	L	查德	차드
New Zealand	ニュージーランド	新西兰	L	紐西蘭	뉴질랜드
Seoul	ソウル	首尔	O	首爾	서울
Seoul	ソウル	汉城	O	漢城	서울
Yemen	イエメン	也门	L	葉門	예멘

????

丝绸之路	sīchóuzhīlù	silk road
机器翻译	jīqifānyì	machine translation
爱国主义	àiguózhǔyì	patriotism
录像带	lùxiàngdài	video cassette
新西兰	Xīnxīlán	New Zealand
临阵磨枪	línzhènmóqiāng	start to prepare at the last moment

????

北京日本人学校	multiword lexemic
北京+日本人+学校	lexemic
北京+日本+人+学校	sublexemic
北京+[日本+人] [学+校]	morphemic
[北+京] [日+本+人] [学+校]	submorphemic

????

电脑迷	diànnǎomí	cyberphile
电子商务	diànzǐshāngwù	e-commerce
追车族	zhuīchēzú	auto fan

Code Conversion

SC	TC1	TC2	TC3	TC4	Remarks
门	們				one-to-one
汤	湯				one-to-one
发	發	髮			one-to-many
暗	暗	闇			one-to-many
干	幹	乾	干	榦	one-to-many

Orthographic Code Conversion

English	SC	TC1	TC2	Incorrect
Telephone	电话	電話		
Dry	干燥	乾燥		干燥　幹燥　斡燥
	阴干	陰乾	陰干	

Lexemic Conversion

English	SC	Taiwan TC	HK TC	Incorrect TC
Software	软件	軟體	軟件	軟件
Taxi	出租汽车	計程車	的士	出租汽車
Osama	奥萨马	奧薩瑪賓拉登	奧薩瑪賓拉丹	奧薩馬本拉登
Bin Laden	本拉登	歐胡島		瓦胡島

TC Variants

Var. 1	Var. 2	English	Comment
裏	裡	Inside	100% interchangeable
著	着	Particle	variant 2 not in Big5
沉	沈	sink; surname	partially interchangeable

Okurigana Variants

HEADWORD	READING	NORMALIZED
書き著す	かきあらわす	書き著す
書き著わす	かきあらわす	書き著す
書著す	かきあらわす	書き著す
書著わす	かきあらわす	書き著す

Cross-Script Variants

Kanji vs. Hiragana	大勢　おおぜい
Kanji vs. Katakana	硫黄　イオウ
Kanji vs. hiragana vs. katakana	猫　ねこ　ネコ
Katakana vs. hybrid	ワイシャツ　Yシャツ
Kanji vs. katakana vs. hybrid	皮膚　ヒフ　皮フ
Kanji vs. hybrid	彗星　すい星
Hiragana vs. katakana	ぴかぴか　ピカピカ

Hangul Variants

cake	케이크 (*keikeu*)	케잌
yellow	옐로우 (*yelrou*)	(*keik*)
Mao Zedong	마오쩌뚱 (*maojjeottung*	옐로 (*yelro*)
Clinton	클린턴 (*keulrinteon*)	모택동 (*motaekdong*)

Cross-Script Variation

Type of Variation	English	Var. 1	Var. 2	Var. 3
Hanja vs. hangul	many people	大勢 (*daese*)	대세 (*daese*)	
Hangul vs. hybrid	shirt	와이셔츠 (*wai-syeacheu*)	Y셔츠 (*wai-syeacheu*)	
Hangul vs. numeral vs. hanja	one o'clock	한시 (*hansi*)	1시 (*hansi*	一時 (*hansi*
English vs. hangul	sex	sex		

Hit Distribution for 人参 'carrot' ninjin

ID	Keyword	Normalized	Google Hits
A	人参	人参	67,500
B	にんじん	人参	66,200
C	ニンジン	人参	58,000

Kana Variants

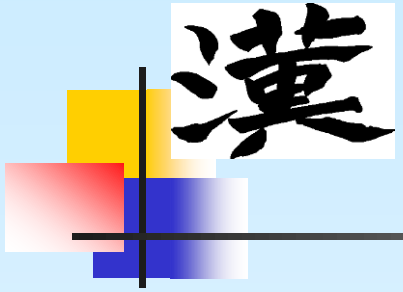
Type	English	Standard	Variants
Macron	computer	コンピュータ	コンピューター
Long vowels	maid	メード	メイド
Multiple kana	team	チーム	ティーム
Traditional	big	おおきい	おうきい
づ vs. ず		つづく	つずく

Kana Variants???

HEADWORD	NORMALIZED	English
アーキテクチャ	アーキテクチャー	Architecture
アーキテクチャー	アーキテクチャー	Architecture
アーキテクチュア	アーキテクチャー	Architecture

Orthographic Normalization Table

HEADWORD	READING	NORMALIZED
空き缶	あきかん	空き缶
空缶	あきかん	空き缶
明き罐	あきかん	空き缶
あき缶	あきかん	空き缶
あき罐	あきかん	空き缶
空きかん	あきかん	空き缶
空き カン	あきかん	空き缶
空き罐	あきかん	空き缶
空罐	あきかん	空き缶
空き鑵	あきかん	空き缶
空鑵	あきかん	空き缶



The CJK Dictionary Institute, Inc.

Visit the CJKI website at

http://www.cjk.org