

Dictionary Reversal

*Building a German-Japanese Dictionary from Japanese-German Data,
Using Vocabulary Analysis, User Cooperation and
Approaches from Natural Language Processing*

Ulrich Apel

Information Systems Architecture Science Research Division

National Institute of Informatics, Tokyo

The Presentation

- * *Personal background*
- * *The WaDoku Project*
 - *example applications of the data*
 - *data structure and user collaboration*
- * *Dictionary reversal*
 - *preliminary project and problems*
 - *data preparation, additional data and expected results*
- * *Conclusion*

Personal Background

- * *Study of japanology (Japanese studies), sociology and ethnology in Munich, Germany*
- * *Doctoral thesis on Japanese futures studies (futurology), research at Ōsaka university*
- * *Lack of up-to-date Japanese-German dictionary lead to WaDoku Dictionary Project*
 - *“newest” dictionary from 1980*
 - *comprehensive dictionary from 1952 (original version from 1937 – old kanji and kana-zukai)*
 - *situation of Japanese-English hardly better*

WaDoku Project

- * *Very comprehensive Japanese-German electronic dictionary*
- * *FileMaker database with more than 100,000 headwords, about 250,000 records*
- * *With assumed 40 headwords per page, size corresponds to 2,500 printed pages*
- * *Highly flexible and easily accessible data*
- * *Look-up of German is possible (Japanese entry with pronunciation)*

User Access to the Dictionary

- * 1999
 - *FileMaker runtime as download*
 - *online version on server at Ōsaka*
- * *Web page <WaDoku.de>, better user collaboration*
 - *2,000–4,000 visits/day*
 - *25,000–45,000 page views/day*
- * *Data used in JMDict, Papillon Project, Reading Tutor, PopJisyo, W3dict etc.*
- * *Other formats: edict, EPWING, Firefox sidebar etc.*

EPWING format in Jamming

すべての辞書・一網打尽

鳥居 前 後 条 複 ク 熟語 191

研5 鳥居

日(単) とりい【鳥居】 [名]

日(単) とりい【鳥居】 [名]

和 鳥居

英 鳥居【とりい】

英 鳥居【とりい】

英 鳥居【とりい】

日(単) とりいがおか【鳥居】

35 front guard frame as

英 鳥居つかまり【とりい】

難 鳥居の水練【トリイノス】

三 とりい=の数(カズ)が重な

広 ○鳥居を越す

三 とりい=を越(コ)・す

広 ○鳥居を越す

難 鳥居ケ奥川【トリイガオ】

novaprije 鳥居つかまり

35 guard frame deck

novaprije 鳥居デッキ

英 鳥居デッキ【とりいでっ】

英 鳥居デッキ【とりいでっ】

とりい【鳥居】 (名)

{Rel., Archit.} Torii {n} (wörtl. etwa „Vogelsitz“; Tor vor Shintō-Heiligtümern; gebildet aus zwei senkrechten Holzpfählen, die durch zwei Querbalken verbunden sind).

鳥居前町 (とりい・まえ・まち) um bzw. vor einem Schrein entstandene Stadt (wie z.B. Ise).

"Myōjin-torii"



kasagi (笠木)

shimagi (島木)

daiwa (台輪)

kusabi (楔)

hashira (柱)

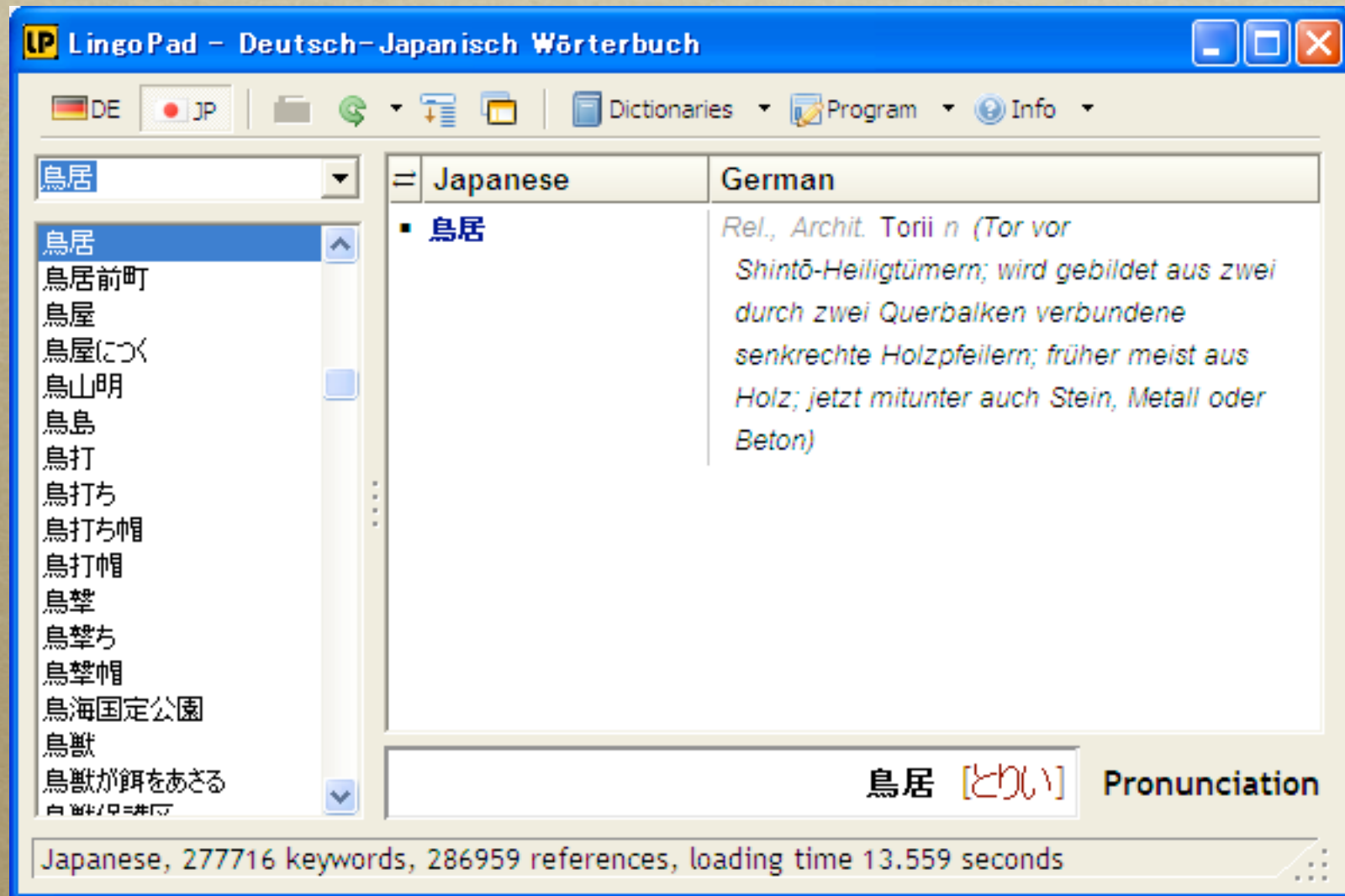
kamebara (亀腹)

gakuzuka (額束)

nuki (貫)

korobi (転び)

LingoPad on Windows



Moji Sidebar for Firefox

Moji

言 Wort 漢 Kanji Mehr...

Search Word

Meaning

鳥居	Torii
百科事典	Enzyklopädie
神域	Schreinbezirk
区画	Bezirk
入口	Eingang

鳥居 [とりい]
{Rel., Archit.}; Torii; (wörtl. etwa "Vogelsitz"; Tor vor Shintô-Heiligtümern; gebildet aus zwei senkrechten Holzpfeilern, die durch zwei Querbalken verbunden sind)



ウィキペディア
フリー百科事典

ナビゲーション

- [メインページ](#)
- [コミュニティ・ポータル](#)
- [最近の出来事](#)
- [最近更新したページ](#)
- [おまかせ表示](#)
- [アップロード \(ウィキメディア・コモンズ\)](#)
- [ウィキペディアに関するお問い合わせ](#)

ヘルプ

- [ヘルプ](#)

本文 ノート 編集 履歴

鳥居

出典: フリー百科事典『ウィキペディア』

鳥居 (とりい) は、神社などにおいて、「門」である。御陵や寺院に建てられ、鳥居を図案化したものになっている。

目次 [非表示]

1 数え方

2 起源

3 形式

3.1 神明鳥居

3.2 明神鳥居

4 種類 (材料)

POPjisyo/POP辞書

鳥居

出典: フリー百科事典『ウィキペディア (Wikipedia)』

鳥居（とりい）は、**神社**などにおいて、神域と人間が住む俗界を区画するもの（**結界**）である。鳥居の入口を示すもの。一種の「門」である。御蔭や寺院に建てられていることもあるが、神社の地図記号は鳥居を図案化したものになる。

鳥居

とりい

Rel., Archit. Torii (wörtl. etwa "Vogelsitz", Tor vor Shintô-Heiligtümern, gebildet aus zwei senkrechten Holzpfählern, die durch zwei Querbalken verbunden sind)

1

2

3

Char	Stroke	Mandarin	Cantonese	Jp On	Jp Kun
Definition					
鳥	11	niao3 diao3 dao3 que4	niu5	チョウ	とり
bird; KangXi radical 196					
居	8	ju1 ji1	geui1	キヨ、キ	いる、おる、おく
live, dwell, reside, sit					

4

5

6

7

wadokujt.w3dict.com

日本語 English Deutsch *

wadokujt. w3dict.com

Lesehilfe Japanisch-Deutsch

examples

about

contact

examples

dictionary-in-text examples.
Move mouse pointer over the text.

Article 9, Constitution of Japan

1. 日本国民は、正義と秩序を基調とする
国際平和を誠実に希求し、国権の発動たる
国際平和
威嚇又は武力の行使
は、主権の保全に
必要最小限度の
手段として、
許さるるのみ。
2. 陸海空軍その他の
戦力は、これを
保持しない。国の
交戦権は、これを認めない。

wadoku.de

和独辞典

Japanisch-Deutsches Wörterbuch

Suchbegriff:

鳥居

suche

[日本語](#) [Hilfe](#)

Umlaute zum Reinkopieren: Ä ä Ü ü Ö ö ß

Fragezeichen als Platzhalter für genau ein Zeichen (z.B. 大?院)

Sternchen als Platzhalter für mehrere Zeichen (z.B. 大*院)

2 items found, displaying all items.

Nr.	Japanisch	Lesung	Deutsch
1	鳥居	とりい	{Rel., Archit.} Torii {n} (Tor vor Shintō-Heiligtümern; wird gebildet aus zwei durch zwei Querbalken verbundene senkrechte Holzpfählen; früher meist aus Holz; jetzt mitunter auch Stein, Metall oder Beton).
2	鳥居前町	とりいまえまち	um bzw. vor einem Schrein entstandene Stadt {f} (wie z.B. Ise).

User collaboration

- * *Comments on entries on the web page*
- * *Proposals for new entries and corrections*
- * *Corrections through editors*
- * *Forum*
- * *Wiki*
- * *Maintenance of web page by volunteers*
- * *NPO WaDoku e.V.*
- * *Good basis for organic growth of the project*

Data and Possible Representation

DaTimeS...	DaJapanisch	DaLesung	DaDeutsch	DaWortart	DaEintragsebene	DaUE_Art	DaUntereintrMidashigo	DaRomajiBe
11.04.2007 19:15:18	飲酒 [1]	いんしゅ [1]	(<POS: N.>) Trinken<Gen.: n>	名	HE			いん'しゅ
	飲酒する	いんしゅする	trinken (<Expl.: Alkohol>).	サ変自	飲酒 [1]	Abl. mit <Umschr. >	<飲酒>する	<いんしゅ>
12.06.2007 20:13:32	飲酒運転者	いんしゅうんてんしゃ	unter Alkoholeinfluss stehender	名	飲酒 [1]	Komp. Anf.	<飲酒>運転者	<いんしゅ>・うんてんしゃ
	飲酒狂	いんしゅきょう	[1] Alkoholismus<Gen.: m>.	名	飲酒 [1]	Komp. Anf.	<飲酒>狂	<いんしゅ>・きょう
	飲酒恐怖症	いんしゅきょうふしよ	{<Dom.: Med.>} Dipsophobie<Gen.: f>.	名	飲酒 [1]	Komp. Anf.	<飲酒>恐怖症	<いんしゅ>・きょうふしよ
	飲酒検査	いんしゅけんさ	Alkoholtest<Gen.: m>.	名	飲酒 [1]	Komp. Anf.	<飲酒>検査	<いんしゅ>・けんさ
	飲酒検知器	いんしゅけんちき	Alkohol-detektor<Gen.: m>.	名	飲酒 [1]	Komp. Anf.	<飲酒>検知器	<いんしゅ>・けんちき
	飲酒癖	いんしゅへき	Trunksucht<Gen.: f>; Hang<Gen.: m>.	名	飲酒 [1]	Komp. Anf.	<飲酒>癖	<いんしゅ>・へき
	飲酒量	いんしゅりょう	konsumierte Alkoholmenge<Gen.: f>.	名	飲酒 [1]	Komp. Anf.	<飲酒>量	<いんしゅ>・りょう
	飲酒にふける	いんしゅにふける	sich dem Trinken hingeben; trinken.	下一他	飲酒 [1]	Verwendung shaisiolei	<飲酒>にふける	<いんしゅ>
	飲酒を慎む	いんしゅをつつしむ	mäßig trinken; beim Trinken Maß	五他	飲酒 [1]	Verwendung shaisiolei	<飲酒>を慎む	<いんしゅ>

inshu 飲酒 (N.) Trinken *n* (von Alkohol).

➡ ~する trinken (Alkohol).

△ ~癖 ~*heki* Trunksucht *f*, Hang *m* zum Alkoholismus. || ~家 ~*ka* (N.) Trinker *m*; Zecher *m*; Säufer *m*.

|| ~検知器 ~*kenchiki* Alkoholdetektor *m*. || ~検査 ~*kensa* Alkoholtest *m*. || ~狂 ~*kyō*

① Alkoholismus *m*. ② Alkoholiker *m*. || ~恐怖症 ~*kyōfushō* MED. Dipsophobie *f*, Angst vor dem Trinken. || ~量 ~*ryō* konsumierte Alkoholmenge *f*. || ~運転 ~*unten* (N.) Trunkenheit *f* am Steuer; Fahren *n* unter Alkoholeinwirkung. || ~運転者 ~*unten-sha* unter Alkoholeinfluss stehender Fahrer *m*; betrunkenener Fahrer *m*.

☆ ~にふける sich dem Trinken hingeben; trinken. || ~を慎む ~*o tsutsushimu* mäßig trinken; beim Trinken Maß halten.

Highly Flexible Data

- * *Mostly automated Romanization in several flavours (segmentation, mark-up for Japanese pitch accent, nasalisation of velar plosive and devocalization, capitalisation, joshi は wa and へ he)*
- * *Parts of speech, conjugation, gender of nouns*
- * *Domains, definitions, comments on usage, etymology, references to synonyms and antonyms*
- * *Main entries and corresponding compounds, derivations, examples of usage & example sentences*

Dictionary for *MekiMekiDoitsugo*

- * *Project for “Germany in Japan 2005/06”*
- * *German course for mobile phones*
- * *Automated German pronunciation in katakana*
- * *Simple reversible dictionary from Japanese-German word pairs for 10,000 most common German words*
- * *Added data on pronunciation, part of speech etc.*
- * *Short comings of this approach became obvious (uncommon Japanese writings, obsolete entries, missing information for Japanese users etc.)*

Problems for Dictionary Reversal

- * *Grammatical distance — no word-to-word correspondence (noun translated as adjective)*
- * *Cultural distance — no semantical one-to-one correspondence (definition instead of translation)*
- * *Different needs of German and Japanese users*
 - *information about cultural background*
 - *no need for Romaji transcription, or gender of German entries and inflection*
 - *Research necessary, e.g. questionnaires*
- * *Complex structure of WaDoku data*

Problems of Data and Structure

- * *Many orthographical variants Japanese entries*
- * *Appropriate translation even on expense of the quality of the German (e.g. many nominalisation or not typical German)*
- * *Explanations on cultural background are difficult to process*
- * *Important information for Japanese users is missing — German pronunciation, gender, no information on inflection or valency of verbs etc.*

Preparation of WaDoku Data

- * *Mark-up for basic meanings of Japanese entry*
- * *Mark-up for main meanings of Japanese entry
(likely to be easily reversible)*
- * *Mark-up for reversible translations —
“harmonic dictionary (Kumiko Tanaka-Ishii)*
- * *Mark-up for non-reversible translations*

Building a Raw Reversed Dictionary

- * Create records of one “official” Japanese writing, one reading and one German translation*
- * Delete cultural and grammatical explanation on Japanese entry (POS, definitions, explanations etc.)*
- * Replace actual German word forms with dictionary form if necessary (e.g. singular instead of plural)*
- * Add available information about German entry (pronunciation, inflection, gender etc.)*

Improvement of the Raw Dictionary

- * *User cooperation*

- *give input about their needs for the dictionary*
- *rank translation candidates according to suitability and group sub-meanings*
- *add and correct translations*

- * *Analyses of the entries and NLP*

- *domains with good or bad one-to-one correspondence (modern sciences vs. religion)*
- *mark up unlikely candidates (e.g. archaic usage)*
- *probabilities through e.g. frequencies in corpuses*

Expected Results

- * *Raw dictionary*
 - *usable dictionary in relatively short time*
 - *decent size but relatively mediocre quality*
- * *Improvements through user cooperation*
 - *good experience with WaDoku.de*
 - *several iterations necessary*
- * *Improvements through analyses and NLP*
 - *in parts very promising*
 - *in parts uncertain*

Conclusion

- * *WaDoku dictionary one of the most important electronic tools for German research on Japan*
- * *DokuWa dictionary might play a similar role for Japanese research on Germany*
- * *Developed method could be applied to other dictionary projects*
- * *Data probably basis for “harmonic dictionary”*
- * *Relatively big benefit through relatively small effort*

Thank You for Your Attention!

