Multilingual Conceptual Access to Lexicon based on Shared Orthography: An ontology-driven study of Chinese and Japanese

Chu-Ren Huang
Institute of Linguistics,
Academia Sinica
Nanking, Taipei,
Taiwan 115
churen@sinica.edu.tw

Ya-Min Chou
Ming Chuan University
250 Zhong Shan N. Rd., Sec. 5,
Taipei 111, Taiwan
milesymchou@yahoo.com.tw

Chiyo Hotani
Department of Linguistics
University of Tuebingen
Wilhelmstr. 19
72074 Tübingen, Deutschland
chiyo.hotani@student.uni-tuebingen.de

Sheng-Yi Chen
Institute of Linguistics,
Academia Sinica
Nanking, Taipei,
Taiwan 115
eagles@gate.sinica.edu.tw

Wan-Ying Lin
Institute of Linguistics,
Academia Sinica
Nanking, Taipei,
Taiwan 115
waiin@gate.sinica.edu.tw

Abstract

In this paper we propose a model for conceptual access to multilingual lexicon based on shared orthography. Our proposal relies crucially on two facts: That both Chinese and Japanese conventionally use Chinese orthography in their respective writing systems, and that the Chinese orthography is anchored on a system of radical parts which encodes basic concepts. Each orthographic unit, called hanzi and kanji respectively, contains a radical which indicates the broad semantic class of the meaning of that unit. Our study utilizes the homomorphism between the Chinese hanzi and Japanese kanji systems to ide[1]ntify bilingual word correspondences. We use bilingual dictionaries, including WordNet, to verify semantic relation between the cross-lingual pairs. These bilingual pairs are then mapped to an ontology constructed based on relations to the relation between the meaning of each character and the basic concept of their radical parts. The conceptual structure of the radical ontology is proposed as a model for simultaneous conceptual access to both languages. A study based on words containing characters composed of the "口(mouth)" radical is given to illustrate the proposal and the actual model. The fact that this model works for two typologically very different languages and that the model contains generative lexicon like coersive links suggests that this model has the conceptual robustness to be applied to other languages.

1 Motivation

Computational conceptual access to multilingual lexicon can be achieved through the use of ontology or WordNet as interlingual links. Some languages do conventionally encode semantic classification information, such as the linguistic system of classifiers or the orthographic system of characters. We attempt to make use of these implicitly encoded linguistic knowledge for conceptual access to lexical information.

On the other hand, even though ontology seems to be a natural choice for conceptual framework to access multilingual lexical information, there is no large-scale implementation nor is there any direct evidence for psychological reality of the frameworks of ontology. Hence, we hope that using a conventionalized semantic classification system will mitigate some of the problems and provide the constructed ontology some motivation since they are the shared and implicit conceptual systems.

2 Background

2.1. Hanzi and kanji: Shared Orthography of Two Typologically Different Languages

Chinese and Japanese are two typologically different languages sharing the same orthography since they both use Chinese characters in written text. What makes this sharing of orthography unique among languages in the world is that Chinese characters (*kanji* in Japanese and *hanzi* in Chinese) explicitly encode information of semantic classification (Xyu 121, Chou and Huang 2005). This partially explains the process of Japanese adopting Chinese orthography even though the two languages are not related. The adaptation is supposed to be based on meaning and not on cognates sharing some linguistic forms. However, this meaning-based view of kanji/hanzi orthography faces a great challenge given the fact that Japanese and Chinese form-meaning pair do not have strict one-to-one mapping. There are meanings instantiated with different forms, as well as same forms representing different meanings. The character 湯 is one of most famous *faux amis*. It stands for 'hot soup' in Chinese and 'hot spring' in Japanese. In sum, these are two languages where their forms are supposed to be organized according to meanings, but show inconsistencies.

It is important to note that WordNet and the Chinese character orthography are not so different as they appear. WordNet assumes that there are some generalizations in how concepts are clustered and lexically organized in languages and propose an explicit lexical level representation framework which can be applied to all languages in the world. Chinese character orthography intuited that there are some conceptual bases for how meanings are lexicalized and organized, hence devised a sub-lexical level representation to represent semantic clusters. Based on this observation, the study of cross-lingual homo-forms between Japanese and Chinese in the context of WordNet offers an unique window for different approaches to lexical conceptualization. Since Japanese and Chinese use the same character set with the same semantic primitives (i.e. radicals), we can compare their conceptual systems with the same atoms when there are variations in meanings of the same word-forms. When this is overlaid over WordNet, we get to compare the ontology of the two represent systems.

2.2. Hantology and the Ontologization of the Semantic Classification of the Radicals

The design of Hantology differs from other word-based ontology. A typical word-based ontology is WordNet which describes the different relations among synonyms. All of the relations among synonyms are based on the senses of words. Therefore, WordNet only needs to take senses into consideration. Hantology is more complicated than WordNet because it describes orthographic forms, pronunciations, senses, variants, lexicalization, the spread of Chinese characters and Japanese kanji.This approach can systematically illustrate the development of Chinese writing system (Chou et al. 2007).

Hantology also provides mapping with Sinica BOW(Academia Sinica Bilingual Ontological WordNet). Sinica BOW is a Chinese-English Ontology and have mapping with WordNet. Therefore, character-based and word-based ontologies are integrated to provide resources from character to word for Chinese language processing.

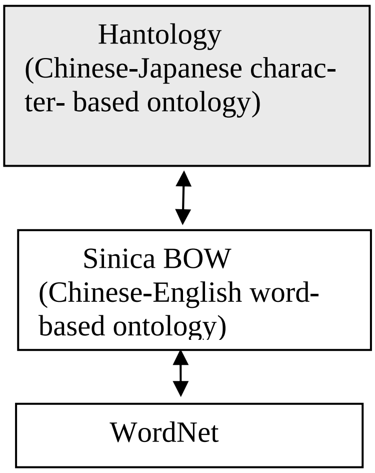


Figure 1. The Mapping among Hantology, Sinica BOW and WordNet

The structure of Hantology is divided into three parts: orthography, pronunciation, and lexicalization.

The orthographic part of Hantology describes the structure of characters, the principles of formatting characters, the evolution of script, glyph expression, the relation of variant and the spread of Chinese characters.

(1) The structure of characters describes the components of each hanzi/kanji, including semantic and phonetic symbols.

(2) The principles of formatting Chinese characters encode the classification of the relation used to compose the character from its components: The pictographic characters were formed by reformatting the pictures of concrete objects. The ideographic (zhi3shi4, refer-event) characters are formed by abstract representation of an concept. The compound ideographic characters are formed by combining two (ore more) semantic symbols. The semantic-phonetic (xing2sheng1) characters, representing over 90 percent of Chinese character, are formed by combining a semantic symbol and a phonetic symbol.

(3) The evolution of script illustrates the different scripts of Chinese characters. The script is a kind of writing style. Because Chinese characters have been used for thousands years, the scripts have changed.The orthographic forms do not change with different scripts. Hantology provides Bronze, Lesser Seal, Kaishu scripts to illustrate evolution of Chinese scripts used from 3000 years ago.

(4) Variants are the characters with different orthographic forms with identical pronunciation and meaning. For example, Chinese characters台and 臺are variants. Variants relations are an important feature in Hantology, similar to WordNet synset relations.

(5) The contrasts between kanji and hanzi glyphs are also encoded. The Japanese language continues to evolve and change after the adoption of Chinese characters. Hence the kanji system includes both historical changes and cross-lingual variations. The kanji system has its own variants which are not necessarily the same set of variants in the hanzi system. Most of Chinese characters adopted by simplified kanji are the variants already used in Chinese. For example, '国' is a simplified kanji of traditional kanji '國'. In addition, Chinese character '国' is also the variant of Chinese character'國'. So, '國'and'国' both are variants in Chinese and Japanese. But, some simplified kanji are not variants used in Chinese. For example, new kanji '欠' is the variant of old kanji '缺' in Japan. However, '欠' is not the variant of '缺' in Chinese.

The second reason of the kanji orthographic form to to be changed is that Japanese not only adopted Chinese characters but also have created hundreds kanji known as Kokuji (国字). Most Kokuji characters have only Japanese pronunciations. Some of Kokuji have been adopted in Chinese. For example, Kokuji '癌'is also borrowed by Chinese. The meaning of '癌' is the same both in Japanese and Chinese.

3. Preliminaries: Orthography based Mapping of Chinese and Japanese Words

3.1 EDR Japanese-English Dictionary

The Japanese-English dictionary of *EDR Electronic Dictionary* is a machine-tractable dictionary that contains the lexical knowledge of Japanese and English.[1]It contains list of 165,695 Japanese words (jwd) and each of their related information.

In this experiment, the English synset, definition and the Part-of-Speech category (POS) of each jwd are used to determine the semantic relations.

We assume that the concept, synonyms, near-synonyms, and paraphrases are the synset of each jwd.In the case when there is no English definition for the word, we assume that there is no equivalent term in English, therefore we use the concept definition of the jwd as its definition.

3.2 SinicaBow

In the previous experiment, the CWN, which contains a list of 8,624 Chinese word (cwd) entries, was used as the cwd data, however since the number of cwds was too small, many jwds were not mapped, even when there is actually a corresponding J-C word pairs exists.

This time we adopt SinicaBow, which contains 9,9642 entries, hoping to find more valid corresponding J-C word pairs.In SinicaBow, each entry is a definition and it contains one or more cwds corresponds to the definition.

In this experiment, the English synset, definition and the POS of each cwd are used to determine the semantic relations.

3.3List of Kanji Variants

List of 125 pairs of manually matched Chinese and Japanese characters with variant glyph forms provided by Kyoto University.

[1] http://www2.nict.go.jp/r/r312/EDR/index.html

Some Japanese kanji and Chinese hanzi have identical property but have different font and Unicode.This resource contains list of Japanese kanji and Chinese hanzi pairs that the kanji properties are exactly the same but the forms and the Unicode are different.

During the mapping procedure, whenever a Japanese kanji and a Chinese hanzi being compared are in the variant list and are the variants of each other, they are considered to be the identical hanzi.

3.4 Procedure

3.4.1Kanji Mapping

Each jwd is mapped to the corresponding cwd according to their kanji similarity.Such mapping pairs are divided in to the following three groups:

(1) *Identical Kanji Sequence Pairs,* where the numbers of kanji in the jwd and cwd are identical and the nth characters in the two words are also identical.

E.g. 頭, 歌手

(2) *Different Kanji Order Pairs,* where the numbers of kanji in the jwd and cwd are identical, and the kanji appear in the two words are identical, but the order is different.

E.g.	Japanese	Chinese
	制限	限制
	律法	法律

(3) *Partially Identical Pairs,* where at least half kanji in the shorter word matches with the part of the longer word.In the case when the shorter word has 4 or less kanji, 2 of the kanji have to be in the longer word.In the case when the shorter word is only 1 kanji, the pair is not considered.jwd matches with a kanji in the cwd.

E.g.,	Japanese	Chinese
	浅黄色	棕黄色
		蛋黄色的
		黄色的
	宇宙飛行体	飛行
		飛行的
		etc…

In the case no corresponding pair relation (one of the three groups explained above) is found for a jwd or a cwd, each word is classified to one of the following group

(4) unmapped jwd is classified to an independent Japanese

(5) unmapped cwd is classified to an independent Chinese

J-C word pairs in such mapping groups are classified in the following manner: (1) A jwd and a cwd are compared.If the words are identical, then they are an identical kanji sequence pair.(2) If the pair is found to be not an identical kanji sequence pair, check if the pair has identical kanji in different order (equal length).If so, then they are a different kanji order pair.(3) If the pair is found to be not a different kanji order pair, then check the partial identity of the pair.Meanwhile, if they are partially identical (according to the characteristics of partially identical pairs described above), the pair is classified to a partially identical pair.

After the mapping process, if the jwd is not mapped to any of the cwd, the jwd is classified to (4) independent Japanese group. If a cwd is not mapped by any of the jwd, it is classified to (5) independent Chinese group.

The number of Japanese kanji- Chinese hanzi pairs' similarity distribution is shown in Table1.

	Number of Words	Number of J-C Word Pairs
(1) Identical hanzi Sequence Pairs	2815 jwds	20199
(2) Different hanzi Order Pairs	204 jwds	473
(3) Partly Identical Pairs	264917 jwds	8438099
(4) Independent Japanese	57518 jwds	-
(5) Independent Chinese	851 cwds	-

Table1. J-C Hanzi Similarity Distribution (Huang et al. 2008).

3.4.2Finding Synonymous Relation (Word Relation)

After the kanji mapping, each of (1) identical kanji sequence pairs, (2) different kanji order pairs and (3) partially identical pairs is divided into three subgroups;

(1-1, 2-1, 3-1) Synonym pairs with identical POS: words in a pair are synonym with identical POS.

E.g. (1-1) 歌手: singer (noun)
(2-1) 藍紫色 (Japanese) and
紫藍色 (Chinese):
blue-violet color (noun)

(3-1) 赤砂糖 (Japanese) and
紅砂糖 (Chinese):
brown sugar (noun)

(1-2, 2-2, 3-2) Synonym pairs with unmatched POS: words in a pair are synonym with different POS or POS of at least one of the words in the pair is missing.

E.g. (1-2) 包:
(Japanese) action of wrapping (noun)
(Chinese) to wrap (verb)
(2-2) 嗽咳 (Japanese): a cough (noun)
咳嗽 (Chinese): cough (verb)

(1-3, 2-3, 3-3) Relation Unidentified: the relation is not determinable by machine processing with the given information at this point.

E.g.	Japanese	Chinese
(1-3)	湯: hot spring (noun)	湯: soup (noun)
(2-3)	生花: arrangement (noun)	花生: flower peanut (noun)
(3-3)	青葡萄: blue grapes (noun)	葡萄牙: Portugal (noun)

In order to find the semantic relation of J-C word pairs by machine analysis, the jwd and the cwd in a pair are compared according to the following information:

Jwd: English synset (jsyn), definition (jdef) and POS

Cwd: English synset (csyn), definition (cdef) and POS

The process of checking the synonymy of each pair is done in the following manner:

If any of the following conditions meets, we assume that the pair is a synonym pair:

at least any one of the synonym from each of jsyn and csyn are identical

at least one of the word definition contains a synonym of the other word

If any synonym pair was found, check if the POS are identical.If the POS are identical, the pair is classified to a synonym pair with identical POS.Otherwise the pair is classified to a synonym pair with non-identical POS.If the pair is not a synonym pair then they are classified to a relation-unidentified pair.

After the process, each of the subgroups is manually examined to check the actual semantic relations of each word pair.

4. Result

4.1 Word Family as Domain Ontology Headed by a Basic Concept

Chinese radical (*yi4fu2*, ideographs; semantic symbols) system offers a unique opportunity for systematic and comprehensive comparison between formal and linguistic ontologies. Chou and Huang (2005) suggests that the family of Chinese characters sharing the same radical can be linked to a basic concept by Qualia relations. Based on Pustejovsky's Quilia Structure [Pustejovsky, 1995] and the original analysis of "ShuoWen-JieXi"[Xyu, 121], each radical group can be as domain ontology headed by one basic concept.

Chou and Huang (2005) assume that 540 radicals in "ShuoWenJieXi" can each represent a basic concept and that all derivative characters are conceptually dependent on that basic concept. Also, they hypothesis that a radical can be classified into six main types: formal, constitutive, telic, participating, descriptive (state, manner) and agentive. Modes of conceptual extension capture the generative nature of radical creativity. All derived characters are conceptually dependent on the basic concept. In their preliminary studies, word family could be headed by a basic concept and also could be represented ontologies in OWL format.

4.2Data Analysis: Japanese and Chinese Words with Identical Orthography

4.2.1 Kanji Mapping

We present our study over Japanese and Chinese lexical semantic relation based on the kanji sequences and their semantic relations.We compared Japanese-English dictionary of Electric Dictionary Research (EDR) with the SinicaBow in order to examine the nature of cross-lingual lexical semantic relations.

	Identical	Different Order	Part Identical
Synonym (Identical POS)	(1-1) 13610 pairs	(2-1) 567 pairs	(3-1) 37466 pairs
Synonym (Unmatched POS)	(1-2) 2265 pairs	(2-2) 214 pairs	(3-2) 22734 pairs
Relation Unidentified	(1-3) 21154 pairs	(2-3) 2336 pairs	(3-3) 1116141 pairs
Total	(1) 37029 pairs	(2) 3117 pairs	(3) 1176341 pairs
	16950 jwds	1497 jwds	39821 jwds

(4) Unmapped Japanese: 107427 jwds

(5) Unmapped Chinese: 41417 entries
Table 1.J-C Kanji Similarity Distribution

The next step is to find Synonymous Relation. (Word Relation).

	Number of 1-to-1 Form-Meaning Pairs Found by Machine Analysis	% in (1)
(1-1) Synonym (Identical POS)	13610	36.8%
(1-2) Synonym (Unmatched POS)	2265	6.1%
(1-3) Relation Unidentified	21154	57.1%

Table 2. Identical Kanji Sequence Pairs (37029 pairs) Synonymous Relation Distribution

	Number of 1-to-1 Form-Meaning Pairs Found by Machine Analysis	% in (2)
(2-1) Synonym (Identical POS)	567	18.2%
(2-2) Synonym (Unmatched POS)	214	6.9%
(2-3) Relation Unidentified	2336	74.9%

Table 3.Identical Kanji But Different Order Pairs (3117 pairs) Synonymous Relation Distribution

	Number of 1-to-1 Form-Meaning Pairs Found by Machine Processing	% in (3)
(3-1) Synonym (Identical POS)	37466	3.2%
(3-2) Synonym (Unmatched POS)	22734	1.9%
(3-3) Relation Unidentified	1116141	94.9%

Table 4. Partially Identical Pairs (1176341 pairs) Synonymous Relation Distribution

The following tables are summarized tables showing the Japanese-Chinese form-meaning relation distribution examined in our preliminary study.

	Pairs Found to be Synonym	% in (1)	Relation Unidentified	% in (1)
Machine Analysis	15875	42.9 %	21154	57.1%

Table 5. Identical kanji Sequence Pairs (37029 pairs) Lexical Semantic Relation

	Pairs Found to be Synonym	% in (2)	Relation Unidentified	% in (2)
Machine Analysis	781	25.1%	2336	74.9%

Table 6. Identical kanji But Different Order Pairs (3117 pairs) Lexical Semantic Relation

	Pairs Found to be Synonym	% in (3)	Relation Unidentified	% in (3)
Machine Analysis	60200	5.1%	1116141	94.9%

Table7. Partially Identical Pairs (1176341 pairs) Lexical Semantic Relation

Since each entry in SinicaBow corresponds to a definition and each jwd has at least a definition or a concept definition, no pairs with insufficient information to check the semantic relation was found.The data shows that as the word forms of the two languages are closer, the more synonyms are found.In order to confirm this observation and to see the actual semantic relation of each pairs, we will continue with more detailed analysis.In addition, in order to pursue the further details of the Japanese-Chinese words relation, we will also analyze the semantic relations (not only synonymous relation) of the relation-unidentified pairs.

4.2.2 "口(mouth)"Analysis Procedure:

In our experiment, we select the identical kanji Sequence Pairs (POS) as our main resources. Characters with the radical"口(mouth)"are selected. In addition, if any character of the words owns the radical "口(mouth)", then it would be included here for anaylysing the detailed semantic relation between jwd and cwd..

Second, we would like to define the semantic relations of J-C word pairs in more details. We examined the actual semantic relation of J-C word pairs by by classifying into 8 semantic relations and marked the relation into [ ] remark.

1.[SYN](Synonym)
2.[NSN](Near-Synonym)
3.[HYP](Hypernym)
4.[HPO](Hyponym)
5.[HOL](Holonym)
6.[MER](Meronym)
7.[/](No Corresponding Semantic Relation)
8.[??](unable to decide)
The pattern is as follows.
[(JWD>jsyn>詞類>jdef>)-[Semantic Relation]-(CWD)>csyn>詞類>cdef]]

Sample:
[(J)-[HYP]-(C)]@
(J is the hypernym of C)

The examples are shown here. In each pair, we define the semantic relation between the jwd and the cwd. The mapping process would be as follows.

E.g
1. [(唖> JWD0028646> N> a condition of being incapable of speaking using the voice>)-[SYN]-(啞> 10137481N> N> paralysis of the vocal cords resulting in an inability to speak> alalia,)]@
2. [(嘴> JWD0378514> N> of a bird, a bill> bill)-[SYN]-(嘴> 01278388N> N> horny projecting jaws of a bird> nib,neb,bill,beak,)]@
3. [(咽喉> JWD0161758> N> part of an animal called a throat>)-[SYN]-(咽喉> 04296952N> N> the passage to the stomach and lungs; in the front part of the neck below the chin and above the collarbone> pharynx,throat,)]@
4. [(啄木鳥> JWD0398785> N> a bird that is related to the picidae, called woodpecker> woodpecker)-[SYN]-(啄木鳥> 01355454N> N> bird with strong claws and a stiff tail adapted for climbing and a hard chisel-like bill for boring into wood for insects> woodpecker,)]@
5. [(人工呼吸器> JWD0401642> N> a medical instrument with which a patient can breathe artificially> respirator)-[SYN]-(人工呼吸器> 03233384N> N> a device for administering long-term artificial respiration> inhalator,respirator,)]@

According to our observation, we notice that most of the Japanese kanji can get their synonyms or near-synonyms in Chinese hanzi and the percentage for this relation is about 63 % in characters with the radical“口(mouth) selected from Identical Synonym POS data. Please refer to table1. The distributions of Semantic Relations comparing jwd to cwd in characters with the radical“口(mouth) chosen from Identical Syno POSare as follows.

Semantic Relations between J-C word	Distribution in Characters with the radical□(mouth)	% in Characters with the Radical □(mouth), 486 total pairs
[SYN]	190	39%
[NSN]	129	27%
[HYP]	16	4%
[HPO]	7	2%
[HOL]	11	3%
[MER]	12	3%
[/]	118	25%
[??]	1	1%

Table8. Semantic Relation Distribution in Characters with the radical“口 Mouth”

4.3 Conceptual Access: A Preliminary Model

In this part, we try to apply dimension of conceptual extension of “□(mouth)” radical into the data we have chosen from the Identical Synonym POS data comparing with Japanese kanji and Chinese hanzi.(Please refer to the Appendix A.) A study based on words containing characters composed of the “□(mouth) ” radical is given for illustration in this preliminary study. It shows that the conceptual robustness can also be applied to other languages, such as Japanese kanji.

Categories in “口 mouth conceptual extension”	Examples in “口 mouth conceptual extension”	Japanese kanji-Chinese hanziExample
Formal -Sense-Vision&Size	嗶	
Formal -Sense-Hearing	叫	
Constitutive	吻、嚨、喉	吻、口吻、嘴、咽喉、喉頭、喉頭炎、喉頭鏡
Descriptive-Active	吐、叫	嘔吐
Descriptive-State	含	含量、含意、含糊、嗜好
Participating-Action	咳、啞、呼、吸	啞、咳嗽、吸血鬼、呼吸、吸盤
Participating-others	哼、嚏	
Participating-instrument	右	左右、右側、右手，周到
Metaphor	启	入口、門口、出入口、出口
TELIC- Subordinate Concept1& Subordinate Concept2		
Subordinate Concept1(Speaking)		
Formal-Property	唐	
Formal-Sense-Hearing	呷	
Constitutive	名、吾	匿名、名詞、名言、名人、物質名詞
Descriptive-Active	吃、哽	吃、吃水線
Participator	吠、喔	狗吠、唯我論、唯

		心論
Participating-Action-Way	呻、吟	唱歌
Participating-others	君，命	君、命令、革命、生命、命運
Subordinate Concept2 (Eating)		
Formal-Sense-Taste	味、啜	味、趣味
Descriptive-Active	噎	
Participating-Action	啜	
Participating-State	嚵	
Participator	啄	啄木鳥、啄木鳥目

Table 9.Jwd Correspondence to“☐(mouth) Conceptual Extension” Graph (☐(mouth), Basic Concept: the body part which used mainly in Language & Food)

5. Conclusion

The result of the experiment comparing the Japanese and Chinese words is to see their form-meaning similarities.Since the Japanese and the Chinese writing system (kanji) and its semantic meanings are near-related, analyzing such relation may contribute to the future research related to Hantology.In this paper, we examine and analyze the form of kanji and the semantic relations between Japanese and Chinese.This paper describes the structure of Hantology which is a character-based bilingual ontology for Chinese and Japanese. Hantology represents orthographic forms, pronunciations, senses, variants, lexicalization, the spread and relation between Chinese characters and Japanese kanji. The results show Hantology has two implications. First, Hantology provides the resources needed by Chinese language processing for computers.Second, Hantology provides a platform to analyze the variation and comparison of Chinese characters and kanji use.

References

Chou, Ya-Min and Chu-Ren Huang. 2005. *Hantology: An Ontology based on Conventionalized Conceptualization.* Proceedings of the Fourth OntoLex Workshop. A workshop held in conjunction with the second IJCNLP. October 15. Jeju, Korea.

Chou,Ya-Min, Shu-Kai Hsieh and Chu-Ren Huang. 2007. HanziGrid: Toward a knowledge infrastructure for Chinese characters-based cultures. In: Ishida, T., Fussell, S.R., Vossen, P.T.J.M. Eds.: Intercultural Collaboration I. Lecture Notes in Computer Science, State-of-the-Art Survey. Springer-Verlag

Fellbaum Christiane.1998. *WordNet: An Electronic Lexical Database*. Cambridge : MIT Press.

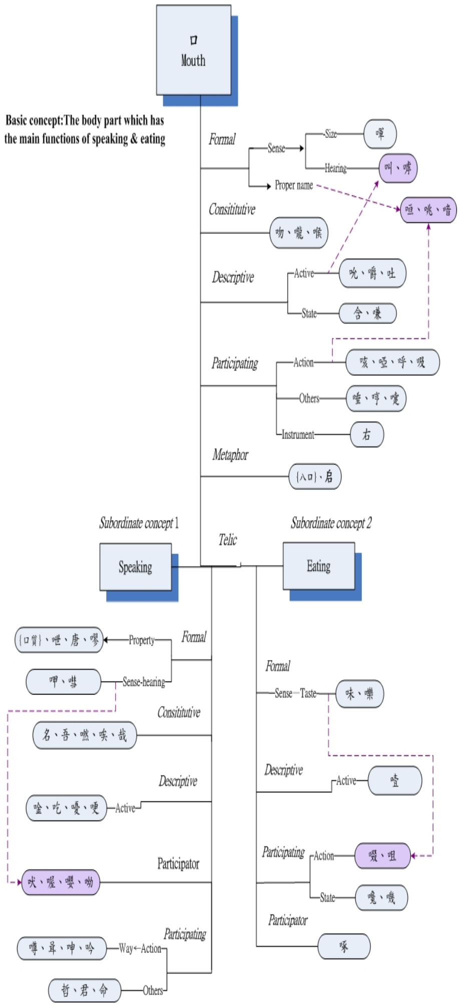
Hsieh, Ching-Chun and Lin, Shih. *A Survey of Full-text Data Bases and Related Techniques for Chinese Ancient Documents in Academia Sinica*, International Journal of Computational Linguistics and Chinese Language Processing, Vol. 2, No. 1, Feb. 1997. (in Chinese)

Huang, Chu-Ren, Chiyo Hotani, Tzu-Yi Kuo, I-Li Su, and Shu-kai Hsieh. 2008. WordNet-anchored Comparison of Chinese-Japanese kanji Word. Proceedings of the 4th Global WordNet Conference. Szeged, Hungary. January 22-25

Pustejovsky, James. 1995. *The Generative Lexicon*, The MIT Press.

Xyu, Sheng. 121/2004. '*The Explanation of Words and the Parsing of Characters' ShuoWenJieZi*. This edition. Beijing: ZhongHua.

Appendix A. The Dimension of “☐ (mouth) Conceptual extension”.



口
Mouth
Basic concept:The body part which has the main functions of speaking & eating
Formal
Sense
Size
嘩
Hearing
叫、嘯
Proper name
咺、咷、嗜
Consititutive
吻、嚨、喉
Descriptive
Active
吮、嚼、吐
State
含、嗛
Participating
Action
咳、啞、呼、吸
Others
唾、哼、噯
Instrument
右
Metaphor
{八口}、启
Subordinate concept 1
Telic
Subordinate concept 2
Speaking
Eating
Formal
{口質}、呭、唐、嘐
Property
呷、嘒
Sense-hearing
Consititutive
名、吾、嘫、唉、哉
Descriptive
唫、吃、嘎、哽
Active
Participator
吠、喔、嚶、呦
Participating
嘩、咠、呻、吟
Way←Action
哲、君、命
Others
Formal
Sense
Taste
味、嚛
Descriptive
Active
噎
Participating
Action
啜、咀
State
嚵、嘰
Participator
啄