

Chinese on the Internet:

Wikipedia and the Chinese Language

HINO Yoshihiro

Translation: Philip Flavin

Introduction

An encyclopaedia is a printed and bound collection of knowledge and information from all over the world. The internet attempts to summarise this same information as it appears. From this point of view, it is possible to interpret the internet as a gigantic encyclopaedia. This to say, the structure of an encyclopaedia is applicable in the computerisation of modern society; however, in spite of stating that both are encyclopaedias, there are substantial differences in the structure of the physical version and that which is found on the internet. Naturally, there are limits to both entries and the amount of information that can be included in the paper version while there are no restrictions for the internet. There are few differences in conducting a search through the use of key words for definitions and examples of these items; however, with the internet, there is an inexhaustible supply of possible objects. The physical encyclopaedia is fixed and cannot be changed; the internet, however, is fluid. The physical encyclopaedia is a completed object and the information contained within is already in the past and nothing that can be done to prevent the deterioration of the knowledge contained. The internet, however, gives continual birth to new information. Nonetheless, despite the continual revision of information, it important to bear in mind that the internet retains outdated information; it also has incorrect and imprecise information. With the physical encyclopaedia, there is an index which can be used to access the different types of information. The internet, on the other hand, initially was organised manually, and with each entry, it was possible to access the necessary information from the entered data; this process, however, gradually transformed from manual searches to automated searches through the entering of key

犬(Jp.), the Chinese word *gǒu* 狗(Ch.), and the English word “dog”, all represent the same idea or concept regardless of whether or not it is a standard dictionary or Wikipedia. Nonetheless, for those people whose native language is Japanese, the initial image of “dog” may differ from those people whose native language is English. Of course, the same thing can be said for those people whose native language is Chinese. It can then be supposed that the descriptions provided in the entries for “dog” in each will then have subtle differences that arise from linguistic divergence. It is also possible to argue that the internet will facilitate the transmission of ideas or concepts that are linguistically or geographically specific.

Words that have yet to be translated, events that have yet to be interpreted are presented in other languages with the translation of related terms and vocabulary. In looking up the Wikipedia entry for “Tokugawa Ieyasu”, for example, the contents and the related terms collected for links within these articles in the Chinese Wikipedia have increased exponentially over the last few years. When tracing the changes in these articles, it is possible to understand the process by which the Wikipedia users create entries, include articles, paste links, and create articles for new entries. It is also possible to understand whether or not their recognition or awareness is the same. This process itself is also a type of translation, a translation of culture, which reflects another insight to this process. With Wikipedia, the number of entries in different languages changes according to the universality and importance of an article. For example, there are entries for “tea” in eighty-eight different languages; for Lu Yu 陸羽 and *The Classic of Tea*, however, there are articles in only seven languages. Of course, the precision and accuracy of the contents changes according to the language in question. The articles in English and Chinese on *The Classic of Tea*, for example, are extremely detailed; in Japanese, however, only the titles of the chapters are given. This is a possible criteria for measuring the degree of interest in the different language groups for the subjects in question. As for the object of this study, the Chinese Wikipedia, it is possible to designate the presence or existence of negotiation even within the Chinese language group. Written Chinese, according to the region, can be divided into either those regions that make use of the classical characters or those that make use of the simplified characters. There are also differences in vocabulary and grammar. The disparity between mainland China, Hong Kong, Taiwan, Singapore, and Malaysia is particularly pronounced. When users from these various regions submit articles, there is the possibility that there is some

form of negotiation undertaken. This is the premise for this article on the Chinese Wikipedia.

1. Defining Wikipedia

Wikipedia is an internet-based “Free Encyclopaedia” sponsored by Mediawiki that allows for reading, writing and editing. Its greatest feature is that anybody can use it and the articles can be duplicated and changed at will. The word itself, “Wikipedia”, is a compound created by combing “wiki”¹⁾ with the second half of “encyclopedia”. The English version was established in May 2001, which then served as a model for the Wikipedias in other languages, and there are presently over 260 versions in different languages. There is, however, no clear linguistic standard to which the different versions are expected to adhere. With Chinese, for example, along with the “Chinese” version used throughout the Chinese language region, there are also versions for different Chinese dialects – Cantonese, the Jiangxi language and the Wu language being examples. There is also a version that makes use of literary Chinese and thus there are problems encountered in saying there is “one” Chinese Wikipedia; this, however, shall not be addressed in this essay. At the time this article was written, December 2008, the English version of Wikipedia had 2,400,000 entries, the Japanese Wikipedia had 500,000 entries and the Chinese Wikipedia had 190,000 entries, which in turn reflects the level of awareness of Wikipedia for each of these languages. There are, however, some versions of Wikipedia with only 1,000 entries, even some with a mere 100 entries, and it cannot be said that the development of Wikipedia is uniform. There are also numerous related projects such as Wiktionary, Wiki-books, Wikiquote and Wikisource,²⁾ each of which is structured along similar lines.

Group endeavours based on the same premise that make use of wiki engines can be seen throughout the internet.³⁾ None of them can be considered “com-

1) One of the steps in the creation of a web site is the development of the reading page and the editing page in the same location. All that is required is the mastery of syntax basic for the creation of a web page; knowledge of HTML is unnecessary.

2) With the Chinese version, *Weiji wenku* (Weiji Books) provides a wide collection of electronic texts, from the Chinese classics to contemporary novels or the basics of Chinese pronunciation through *pinyin* 拼音 (Ch.).

3) For an immediate example, with the Yahoo Japan news articles, there is a link for related articles.

pleted” as they are constantly being updated through continual editing, writing and corrections. Within this process of multiple operators, there is the possibility of this advancing in positive as well as negative directions. There is also the possibility of strife between those authors with differing principles and positions. Nonetheless, the ideal is that with the numerous people reading and writing the articles, which are in turn reread and rewritten by others, the level of accuracy for the articles in question should be extremely high. This is something unthinkable with a physical encyclopaedia. From the point of possible secondary use, rather than the establishment of analogous close sites or the establishment of derivative sites that collect in compact form the contents of entries and articles, there are examples of the original articles themselves are being used derivatively as linguistic materials.⁴⁾

With the surrounding conditions of Wikipedia as such, I would like to now consider Wikipedia itself. In the English, Japanese and Chinese versions of Wikipedia, the top page for each clearly marked with the respective translations of “The Free Encyclopedia” – *furi-hyakka jiten* フリー百科事典 (Jp.) and *ziyou de baikeshu* 自由的百科全书 (Ch) allowing for no error in understanding the nature of the site. The general definition of encyclopaedia is that one can locate information on a given subject through a word that is arranged according to the alphabet or, in the Japanese instance, the Japanese syllabary. Nevertheless, accessing the comprehensive information on all of the different versions of Wikipedia on the internet requires using the search engine located in a readily viewable spot on the page. Recently, in external searches, the results now commonly include Wikipedia entries in the top results, which people also utilise. Wikipedia, however, also has a “paper”-like component with an index or category function. If one focuses on the index accessible from the top page, the English, Japanese, and Chinese versions are all very different. The index and category index for the English version of Wikipedia is arranged A-Z, the Japanese version has an index and category

In Chinese, there is a link *Beicilang* that leads to a joint edited Chinese-Japanese/Japanese-Chinese dictionary project. There is a Wikipedia derived on line encyclopaedia known as *Baidu baike*.

4) HASEBE, Yoichiro. 2006. Method for using Wikipedia as Japanese Corpus. *Doshisha studies in language and culture* 9 (2) pp.373-403

HINO Yoshihiro. 2008. Chinese Language Wikipedia as a Potential Linguistic Object. *Journal of East Asian cultural interaction studies* 1 pp.337-345

can be said to embody a similar concept inherent in the original Wikipedia.

One thing that should not be forgotten with Wikipedia is the unique browsing that can be done by progressing from one article to the next through the links found on each page. Certainly, anybody who has used Wikipedia has had this experience. These links simultaneously function as requests or commissions for the creation of new articles by authors and readers, this function being one of the characteristics of Wikipedia.

2. The Conversion Function for the Chinese Wikipedia⁵⁾

I would like to turn to the Chinese Wikipedia. I selected the Chinese Wikipedia because the environment for the Chinese Wikipedia differs from the Japanese and English environments. When the Chinese Wikipedia was established in October 2002, there were two versions: that which made use of the simplified characters, the zh-cn version, and the zh-tw version which made use of the traditional characters. The table would be partitioned after being automatically read according to the access location. The regions that make use of the simplified characters included the Chinese mainland, Malaysia, and Singapore, while Taiwan, Hong Kong and Macao were categorised as those areas that operated with the use of the traditional characters. This can be interpreted that these areas have differences in grammar, vocabulary, and character usage despite their all being the same Chinese language. To illustrate the differences at the character level, in Japanese, for example, the characters for “postage stamp” are *kitte* 切手 (Jp.) however, in Chinese with the simplified characters, “postage stamp” is rendered *yóupiào* 邮票 (Ch.). When using the classical characters, it is rendered 郵票. To illustrate the differences at the vocabulary level, on the Chinese mainland, the word “printer” is rendered 打印機 (*dǎyìnjī*), while in Taiwan, it is rendered as *yinbiǎojī* 印表機 (Ch.). As can be seen with these two examples, the linguistic differences in the Chinese language, both written and in vocabulary, can lead to confusion in the development of something like Wikipedia as these differences would give rise to multiple entries for the same object. In October 2004, a conversion function was

5) Wikipedia: 字词转换请求或候选

<http://zh.wikipedia.org/w/index.php?title=Wikipedia:%E5%AD%A7%E8%AF%8D%E8%BD%AC%E6%8D%A2%E8%AF%B7%E6%B1%82%E6%88%96%E5%80%99%E9%80%89&variant=zh-cn>

installed that then allowed for the integration of the differing versions of the Chinese Wikipedia as the users could toggle between the two writing systems. In September 2008, several improvements were effected and the conversion function was expanded to include the following differences: no change, simplified characters, traditional characters, simplified characters used in mainland China, the traditional characters used in Hong Kong and Macao, the Malaysian simplified characters, and the traditional characters used in Taiwan – a total of seven different regions. From this point, it is possible to have a Chinese version of Wikipedia independent from regional differences, one field in which articles can be created and edited. I would like to show the functions here.

(不转换) [打印机] (No change)

打印机是一种電腦输出设备，可以電腦內儲存的数据按照文字或图形的方式永久的输出到纸张或者透明胶片上。

(简体) [打印机] (Simplified)

打印机是一种电脑输出设备，可以电脑內儲存的数据按照文字或图形的方式永久的输出到纸张或者透明胶片上。

(繁體) [印表機] (Traditional)

印表機是一種電腦輸出設備，可以電腦內儲存的數據按照文字或圖形的方式永久的輸出到紙張或者透明膠片上。

(大陆简体) [打印机] (Mainland simplified)

打印机是一种电脑输出设备可以电脑內儲存的数据按照文字或图形的方式永久的输出到纸张或者透明胶片上。

(马新简体) [打印机] (Malaysian simplified)

打印机是一种电脑输出设备，可以电脑內儲存的数据按照文字或图形的方式永久的输出到纸张或者透明胶片上。

(港澳繁體) [打印機] (Hong Kong traditional)

打印機是一種電腦輸出設備，可以電腦內儲存的數據按照文字或圖形的方式永久的輸出到紙張或者透明膠片上。

(台灣正體) [印表機] (Taiwan traditional)

印表機是一種電腦輸出設備，可以電腦內儲存的數據按照文字或圖形的方式永久的輸出到紙張或者透明膠片上。

(Wikipedia: excerpt from November 2008)

As can be seen, with the non-conversion example, there is a mixture of simplified characters and traditional characters. This mixture arose as users from both writing systems effected their entries. At this time, however, regional differences in vocabulary were not accounted for. “Printer”, for example, in Taiwan is 印表機, but in Hong Kong and Macao is 打印機, and one can see that there are problems that extend beyond traditional versus simplified characters.

Next, I would like to show the following source data.⁶⁾

```

{{noteTA7)
|T=zh:打印机; zh-hans:打印机; zh-hant:印表機; zh-hk:打印機;
|dt=中国大陆: 打印機; 台灣: 印表機; 香港: 打印機
|l=zh:打印机; zh-hans:打印机; zh-hant:印表機; zh-hk:打印機;
|d1=中国大陆: 打印機; 台灣: 印表機; 香港: 打印機
|2=zh:激光; zh-hans:激光; zh-hant:雷射; zh-hk:鐳射;
|d2=中国大陆: 激光; 台灣: 雷射; 香港: 鐳射
}}
[[Image:Hp business inkjet 3000.jpg|thumbnail|right|300px|惠
普business inkjet3000工作组级喷墨打印机]]
'''打印机'''是一种[[電腦]]输出设备，可以電腦內儲存的數據按照
文字或圖形的方式永久的輸出到紙張或者透明膠片上。

```

The information enclosed in the brackets `{{ }}` is meta-data. It is in these sections where the instruction for automatic conversion occurs. This is followed by the instructions to convert vocabulary in both the title and the article. The

6) The following is an excerpt from the original text for the “no change”.

7) Template:noteTA

<http://zh.wikipedia.org/wiki/NoteTA>

information contained within the brackets [[]] indicates that there are links leading to entries related to the word in question. I will now turn to an explanation of who creates the meta-code and at what step this occurs.

[*Yáshàng de bōniū*]

A user in Taiwan created an article on 2008 April 13th. When this article was created the following information was included:

《懸崖上的波妞》（日文：崖の上のポニョ，2008年）是由宮崎駿導演，編劇，吉卜力工作室推出的一部動畫電影。預定2008年7月19日在日本上映

From the source information, *The Internet Movie Database*,⁸⁾ an American website that was the source of the quote, is the only thing cited in this simple description. After this, a request for an expansion of the article was effected through the use of the dialogue function [对话], immediately after which, a different user in Taiwan as well as a Chinese user located in England undertook the expansion of the article. It is possible to determine the existence of a period of intense rewriting through the entries as part of the process.

懸崖上的波妞（2008. 4. 13）→哈尔移动城堡（2008. 8. 5）→崖上的波妞（2008. 8. 5）→悬崖上的金鱼公主（2008. 8. 5）→崖上的波妞（2008. 8. 7）→悬崖上的金鱼公主（2008. 11. 22）

On 5 August 2008, there were multiple rewritings, the result of which was, that within that day, the meta-data shows the following addition of conversion information for vocabulary as shown below.

|3=zh-hans:悬崖上的金鱼公主; zh-hant:崖上的波妞; zh-tw:崖上的波妞;

In simplified characters, the title has been rendered 悬崖上的金鱼公主; however, in the traditional characters, specifically those used in Taiwan, it has been ren-

8) <http://www.imdb.com/title/tt0876563/>

dered 崖上的波妞. After this, the following detailed information was added.

```
|T=zh-hans:悬崖上的金鱼公主; zh-hant:崖上的波妞; zh-tw:崖上的波妞; zh-hk:崖上の波兒;
|1=zh-hans:悬崖上的金鱼公主; zh-hant:崖上的波妞; zh-tw:崖上的波妞; zh-hk:崖上の波兒;
```

For the conversion function, the final entry to be made was for Hong Kong.

```
{{noteTA
|T=zh-hans:悬崖上的金鱼公主; zh-hant:崖上的波妞; zh-tw:崖上的波妞; zh-hk:崖上の波兒;
|1=zh-hans:悬崖上的金鱼公主; zh-hant:崖上的波妞; zh-tw:崖上的波妞; zh-hk:崖上の波兒;
|2=zh-hans:哈尔移动城堡; zh-hant:霍爾的移動城堡; zh-tw:霍爾的移動城堡; zh-hk:哈爾移動城堡;
|3=zh-hans:千与千寻; zh-hant:神隱少女; zh-tw:神隱少女; zh-hk:千與千尋;
|4=zh-hans:波妞; zh-hant:波妞; zh-tw:波妞; zh-hk:波兒;
}}
```

In the latest version, it is not just the conversion of the entry [T], but a list for the following words [1–4] that appear in the article was added.

There is, however, a slight difference in the circumstances surrounding the necessity of converting ‘postage stamp’ in the same fashion. The meta-data for ‘postage stamp’ has been given below.

```
{{noteTA
|1=zh-hans:朝鲜; zh-hant:北韓;
|2=zh-hans:名著; zh-hant:名著;
|3=zh-hans:著名; zh-hant:著名;
}}
{{featured article}}
{{四大平面藏品}}
```

[[Image:Stamp China 1946 2000 on 5 ovpt. jpg|thumb|200px|[[中国抗日战争|抗战]]胜利后[[中国]]发行的邮票]]

[[Image:Chang-Kai-Shiek-Stamp. JPG|thumb|200px|蔣總統像臺北版郵票]]

''' 邮票 ''' 是郵政機關發行，供寄遞郵件貼用的[[郵資]]憑證。发送者为邮政服务付费的一种形式。发送者会将邮票贴在信件上，再由邮局盖章消值，以用于在[[郵件]]被寄出前，證明寄郵人已支付全部或部分傳遞費用。為方便使用，郵票四周一般會打上[[齒孔]]，背面則加上一層[[背膠]]。世界上的第一枚郵票是英国的[[黑便士]]。[…以下省略]

Here, the simplified character 邮 can be changed to the classical 郵, but distinctions of this sort between characters is not recorded in the meta-data, [1=zh-hans:朝鮮; zh-hant:北韓;] only the differences in the word base, as shown in this example, have been recorded. From this, it is possible to understand both the character exchange function included in the page's meta-data and the configuration for the definition of the common exchange. This is [Mediawiki 系统转换表] ([Mediawiki systems replacement graph]. On this page, the entries have been organised according to pre-conversion and post-conversion, to create an exchange table. This is followed by a corresponding graph with the simplified characters used in the mainland writing system from which I have provided an excerpt.⁹⁾ It can be seen that to a certain degree, this page contains fixed or pre-determined conversions. With the conversions, this accounts for the presence or existence of multiple steps for character conversion that are only used on specified pages that have a commonly shared conversion used throughout the Chinese version of Wikipedia. Below is an excerpt from the [字詞轉換處理/公共轉換組] (Chinese character and vocabulary conversion process common conversion entries). The conversion for the glossary or character base, and, as can be seen, covers a variety of genres.

*Examples of mainland simplified conversion:*¹⁰⁾

9) <http://zh.wikipedia.org/wiki/MediaWiki:Conversiontable/zh-cn>

10) MediaWiki:Conversiontable/zh-cn

<http://zh.wikipedia.org/wiki/MediaWiki:Conversiontable/zh-cn>

地名 Geographic Names

- # 諾曼第⇒诺曼底 //Normandie (Normandy);
- # 土魯斯⇒图卢兹 //Toulouse;
- # 坎城⇒戛纳 //Cannes;
- # 羅亞爾⇒卢瓦尔 //Loire;

电脑与信息学 Computer and Information Science

- # 頻寬⇒带宽;
- # 新聞群組⇒新闻组;
- # 數位相機⇒数码相机;
- # 數位單眼相機⇒数码单反相机;

- # 水準⇒水平;
- # 形上學⇒形而上学;
- # 乡愿⇒鄉愿;

度量衡 Measurement

- #公尺⇒米;
- #公分⇒厘米;
- #公升⇒升;

* The characters to the left of the symbol ⇒ are classic, while those to the right are simplified.

3. The Formation Process of the Compiled Articles

Here we can see with the Chinese Wikipedia what sorts of entries have been established. With this, we can now examine how articles are constructed with the two following examples of “Allah” (安拉) and “spouse” (配偶者).

安拉

阿拉 (الله), 阿拉伯語的「神」, 相當於英文中的「The God」(上

Wikipedia:字詞轉換處理/公共轉換組

<http://zh.wikipedia.org/wiki/Wikipedia:%E5%AD%97%E8%A9%9E%E8%BD%89%E6%8F%9B%E8%99%95%E7%90%86/%E5%85%AC%E5%85%B1%E8%BD%89%E6%8F%9B%E7%B5%84>

帝), 中文音譯為安拉或阿拉。穆斯林, 即伊斯蘭教的信徒, 只信奉安拉為自己的主。雖然猶太教, 基督教, 伊斯蘭教三種宗教教義不同, 甚至會彼此攻訐仇視, 但基於一神教的屬性, 還有伊斯蘭教將摩西, 亞倫, 耶穌, 穆罕默德均視為神派到人世間的先知的這種觀點, 可以確定伊斯蘭教說的安拉和猶太教, 基督教的耶和華, 其實指的都是同一個神。這三種宗教均被認為屬於沙漠一神諸教。…[以下省略]/2008年12月3日版

分类: 伊斯兰教 | 上帝

From the creation of this article on 6 January 2004 through 3 December 2008, there have been a total of seventy-three corrections and expansions by forty-two people. The editors are diverse, located in Peking, Guangdong, Taiwan, the United States and more, and comprise an even distribution of both classical and simplified characters users participating in the editing.

The entry for “Allah” was first created by an unknown user on 6 May 2004 and consisted of the following simple text: ==大字标题==安拉 安拉, 伊斯兰教的真主, 伊斯兰教的信徒只信奉安拉为自己的主. 伊斯兰的信徒也称作“穆斯林”. On 9 May 2004, the article was expanded to include a chart tracing the origins of the Islamic language; it also provides multiple links to entries in other languages. After a given period of time, with mechanical collection, the number of links to other languages increased automatically, and a category for entries on the Islamic religion was established on 15 March 2005. On 21 May 2005, the passage 安拉, 伊斯兰教的真主 was rewritten as 安拉, 伊斯兰教的神, in which the translation of “Allah” (安拉) as 真主 *zhēnzhǔ* (the true lord), which is a translation of the meaning for Allah rather than a transliteration, was replaced with 神, or “god”. An entry explaining the meaning of *zhēnzhǔ* was created by the same user on 21 May 2005. This was followed on 9 July 2005 with information on the Islamic religion being added, after which, on 25 July 2005, the largest number of articles were inserted. But, because of the presence of specialised vocabulary in the articles, links leading to other articles explaining these terms were entered on 25 July 2005. After this, there was a brief period during which the characters used in the transliteration of “Allah” were repeatedly altered from 安拉 *ānlā* to 阿拉 *ālā*. Nonetheless, the article had more or less reached its tentative final form.

On one hand, on 25 May 2005, three months after the creation of the entry, a different user provided the following explication for the term *zhēnzhǔ*: 伊斯兰

教中的神（安拉，阿拉，胡大）的意译。This was later simplified by another user.

The next example is the term *àirén* (爱人/愛人) (Ch.). According to the *Xiàndài hànyǔ cídiǎn* (现代汉语词典), the following definitions are provided:

1. 指丈夫或妻子 (Indicates husband or wife)
2. 指恋爱中的男女的一方 (indicates either a lover of either gender)

This indicates that on the Chinese mainland *àirén* can be used to mean either ‘spouse’ or ‘lover’ and the Japanese meaning of *aijin* 愛人 (Jp.), which makes use of the same Chinese characters, is not present in mainland Chinese. In the Chinese used in Taiwan, however, it has the same meaning as the Japanese, and does not have the meaning of ‘spouse’ as it does in the mainland. This is to say that even though the same characters are used, regardless of their rendering in simplified or classical characters, the meaning of the word is different according to the geographic local. If, for example, the term *àirén* given in the Chinese Wikipedia is examined, the following explanation is provided.

爱人是在一个时期中国大陆对婚姻配偶的称呼。在其他使用中文的地区，这个词作为名词使用得很少，而通常是动宾词组，如《论语》中“仁者爱人”。

“In mainland China, the term àirén has come to mean “spouse”. In the other Chinese languages spoken in different regions, this term is seldom used as a noun, and is more commonly used as a verb followed by a direct object (“to love a person”). For example, in the Lùnyǔ (論語), the following sentence is found: 仁者愛人 “The virtuous man loves humanity.”

After the creation of this simple explanation, the following titles were created: *qǐyuán* 起源 (Ch.) “linguistic origins”, *fāzhǎn* 发展 (Ch.) “development”, *xiànzhuàng* 现状 (Ch.) “present state”, *qítā yányǔ* 其他语言 (Ch.) “other languages”. The difference in meaning with Japanese is briefly discussed in the “other languages” entry, the difference between Taiwanese Chinese and mainland Chinese is not discussed, and the present meaning of *àirén* as used on the mainland is introduced. After this, a user of classical Chinese characters made the following

entry:

'''爱人'''本是'''戀人'''，'''情人'''的意思，是在一个时期中国大陆对婚姻配偶的称呼。其他使用中文的地區則為未婚情侶稱呼對方或夫妻稱呼配偶所通用。

"The original meaning of àirén was 'lover' or 'sweetheart' or 'mistress', but on the Chinese mainland, this term has a new meaning of 'spouse'. In the Chinese languages used in other regions, it has the meaning of an unmarried lover or spouse."

In Taiwan, *àirén* is not used to indicate 'spouse' and this point is emphasised (20 November 2006). On the following day, a mainland user made the following entry:

在共产党控制的地区，如延安，成了被广泛使用的中性词，男女均可使用。中華人民共和國成立後，爱人是中國大陸唯一可以使用的正式称呼。

"In Yan'an (a city in Shanxi) where the influence of communist ideology is very strong, the term àirén is gender free and can be used to indicate either a male or female lover. With the establishment of the People's Republic of China, this is the established usage."

This was later commented on by a Hong Kong user who states that the classic meaning of *àirén* is the same as Japanese. On 18 December 2008, this comment was then followed by a comment from a user in Peking in which includes the following:

在日文中，日文汉字“爱人”是情人的意思。在一些受日本文化影响较深的中文地区，这种使用也相当广泛。...

"The Japanese term aijin means qingrén. Because of this, those Chinese speaking regions that have been influenced by Japanese culture make use of the Japanese meaning."

Overall, with the changes, the explanations have been simplified, and there has been a change from a mainland Chinese-centric view to a more balanced selection

of articles, and it is possible to see the inclusion of synonyms/antonyms and a user vocabulary/lexicon. To summarise the different stages as of February 2009, *àirén* was a term first used in the literature of the New Cultural Movement in the 1910s. Around 1930, *àirén* began to be used to indicate “spouse” as part of the Communist Party’s ideology to establish equality between men and women. Part of this ideology included the discarding of gender oriented terms such as *xiānsheng* 先生 (Mr.), *tàitai* 太太 (Mrs.), which replaced with the neutral *àirén* 爱人. Outside of the mainland, the above terms as well as other words such as *lǎopō* 老婆, *lǎogōng* 老公, *lǎobànr* 老伴儿, *qīzi* 妻子, *nánren* 男人, *nǚrén* 女人, *zhàngfu* 丈夫, *xiānsheng* 先生 and *tàitai* 太太 continued to be used as terms indicating “spouse”. In the entry for usage in other languages is the explanation of the Japanese meaning and use of the term *aijin*, which indicates an illicit lover. In confirming the links with other languages and the Chinese *àirén*, there are links to the Japanese and English. Of course, from the meaning of the simplified character for *àirén*, there is no link that leads to the Japanese definition of “spouse”, but the Japanese *aijin* and the English “love” each have links. With the presence of an article, however, that explains “those regions that have been heavily influenced by Japanese culture share the same meaning for *àirén*”, there is the possibility that the Japanese meaning and use of *aijin* will spread on the mainland through the use of Wikipedia.

With the above examples, we can see and understand one aspect of the process from the creation of the entries to the creation of the articles. A simple explanation is provided for a recently established entry, and if necessary, a list of the contents is also created. At the same time, there is a dialogue function for the users by which they can expand and broaden the articles. With this, users from different geographic locations can include explanations should one be necessary. As for the methods of including explanations, translated terms can be included in the glossary; transliterations, free translations, the synonyms associated can be included, footnotes can be inserted into texts written by other users should there be points that are unclear, and it is also possible to create entries for all these words. Through this process, authors in different locations can contribute equally to the article despite differences or discrepancies in data from their respective regions.

4. The Relationship of Wikipedia Vocabulary

The vocabulary used in Wikipedia is mixed according to whether or not the articles in question are recorded entries in the language of the culture in question with those articles that are translations from other languages. That is to say, Wikipedia functions as an encyclopaedia as well as a translation dictionary. The results reflect the different personalities of the vocabulary used for the entries and it is necessary to be wary when making use of it.

I would like to examine one a general noun that is universal in its value and pursue the differences.

In perusing the differences in the articles and the headings for the entry “dog” in Japanese (*inu*), Chinese (*gǒu, quǎn*) and English, the composition of each differs according to the language, which reflects the extent of linguistic interest towards the term/word in question as well as trends.

Japanese	Chinese	English
犬 (<i>inu</i>)	狗/犬 (<i>gǒu, quǎn</i>)	dog
1 概要	1 概要	Contents
2 生態的・形態的特徴	2 食物	1 Etymology and related terminology
3 イヌに与えてはいけな い食べ物	3 功能	2 Taxonomy and evolution
4 イヌの起源	4 名稱及術語	3 Biology
5 人間社会との関わり	5 狗類的共性	4 Behavior and intelligence
6 イヌと歴史・文化	6 解剖學結構	5 Differences from wolves
7 歴史に名を残したイヌ	7 狗與人類的關係	6 See also
8 障害犬	8 分類	7 References
9 イヌの登場することわ ぎ・故事成語	9 狗的本能	8 Further reading
10 名前にイヌを持つもの	10 狗的近親	9 External links
11 その他 イヌについて	11 關於狗的文化	
12 関連項目	12 参考書目	
	13 参看	
	14 外部链接	

(Excerpted from the relevant Wikipedia pages)

The English provides an excellent overall balance for “dog”. In Japanese, there is an entry for those foods that should not be given to dogs, which emphasises the training of and interaction with dogs. In Chinese, under the heading for “nomenclature/technical language” there is an in depth explanation of the character for “dog”. Through these variations, it is possible to determine what information is deemed important for the contents and its linguistic background. There are

similar differences that can be seen in Japanese/Chinese/English entries for “tea” and “coffee”.

There is also the following example: in focusing on the key word *nōson*, the Japanese term for farming community/village, and pursuing this through the “other languages” link, Wikipedia comes up with the following:

Japanese	Chinese	English
農村 = 村落	镇	town
田舍	农村	rural

The contents of the language link are not necessarily correct. In general use dictionaries, the following entries are rendered as: *machi* 町 (Jp.) = *chéngshì* 城市 (Ch.) = town, *inaka* 田舍 (Jp.) = *xiāngxià* 乡下 (Ch.) = country, *nōson* 農村 (Jp.) = *nóngcūn* 农村 (Ch.) = farm village. With this, we can see that the correspondence of the Wikipedia entries is inaccurate. With examples such as this, provided one has the requisite linguistic knowledge, it is possible to recognise the errors; however, without this knowledge, particularly of those languages that make use of Chinese characters such as Chinese and Japanese, there is the possibility that the reader will interpret the data presented as correct.

Compared with general nouns, it is clear that there are both differences as well as uniformity in the value attached to recently coined terms, proper nouns and others. In these instances, there is no error in the entries in other languages. For example, there is no difference in the importance attached to the entry for the Chinese State Council 国务院 organisation in either the Japanese or English entries. According to the level of interest, the amount of explanation changes; fundamentally, the data presented at the point of departure is faithfully maintained. The same holds true with vocabulary such as *hanyu pinyin*, *tsaijing*, *chajing*. However, with translations from other languages, there is a tendency to generalise the meanings according to the language of translation. A few examples include the Chinese *Sānguó yǎnyì* 三国演义 (Romance of the Three Kingdoms) and the Japanese *Sangokushi engi* 三国志演義 (the Japanese translation). Doubtless, the entry for *Chángchéng* (The Great Wall) is similar.

Then, there are also differences in the characters used according to the geographic location within the Chinese Wikipedia, for example, new terms or phrases

such as 悬崖上的金鱼公主-崖上的波妞, or proper nouns which are principally used as translation terminology in which there are differences that can be given as examples.

With the above, not all of the entries are necessarily equally balanced; should the way in which they came about differ, from other languages or even with Chinese, the standpoint will change.

According to the Wikipedia entries, the contents of the articles will reflect the author's viewpoints, and as the degree of interest by the users determines the maturation of the article, there are unmistakably many problems. But, generally speaking, as a function, group editing, when one entry is created, the editors contribute to the maturation of the article, and along with constructing relations with entries in other languages, it is nonetheless a very effective process. It is, however, necessary to be cognizant of the awareness of readers for the Chinese Wikipedia. What degree of awareness do the users bring with them? Do they, for example, have an awareness of the grammar and the vocabulary they use? It is only possible to proceed with this in mind.

5. Conclusion

This concludes the general outline of the Chinese Wikipedia. To explain the fundamental differences in Wikipedia from the original dictionary/encyclopaedia, I have given the three examples of the change in the concept of encyclopaedia, the clarification of the operations, the unspecified editors.

With the production of a physical encyclopaedia, the arrangement or disposition of the entries, what is to be written, the placement of the collected words, the inclusion of references, the positioning of all these must be decided in the creation process. With Wikipedia, however, this is unnecessary. Included in the classification and index, through the wiki engine, it is extremely easy to create links between words, and add to entries with equal ease. For example, in the process of creating an article, it is possible to indicate the need for an explanation by somebody else by surrounding the term in question with brackets [[]]. This action alone creates a new entry, and there is the possibility that somebody with the desired knowledge will see this and provide the article. This ability is completely different from the original encyclopaedia. With the use of this function, the limitations in terms of space found in the physical encyclopaedia have been eliminated. The linguistic limitation of the original encyclopaedia to one language

has also been overcome as Wikipedia is able to confirm what writings there are for entries in other languages as well as confirm what the suitable realisation of a word is in another language. Wikipedia thus embodies the function of an encyclopaedia as well as a translation dictionary. With the existence of exact correspondences in multiple languages, vocabulary and entries spread more effectively. This is the change in concept of encyclopaedia embodied by Wikipedia.

The next point is that the editing process with physical encyclopaedia is invisible. With Wikipedia, it is possible to view all of the edited versions of an article from the inception of the article, and this includes a profile of the changes, those who made the modifications and their debates. This is not only for the process of creating an article, but also for determining the process of deciding upon terms to be used in a translation from another language. Also, with the clarification of the operations, those problems encountered in the editing process are clarified and it possible to see how an article reached its present state, which in turn provides a means for ascertaining whether or not the article reflects authorial bias.

With the unspecified editors, there is the question of whether or not this weakens the reliability of Wikipedia as a resource, but as there are numerous people from different standpoints participating in the creation of these entries, it is important to remember that there is the possibility of this being the creation of a diverse encyclopaedia. With Chinese and the coexistence of simplified and classical characters, because of the regional differences in language and expression, encyclopaedias that are created by an unspecified number of people in the same way as Wikipedia will doubtless will bring accusations of there being a lack of homogeneity. Nonetheless, in compiling the disparate articles into one unified Chinese Wikipedia, it is possible that there will be users separated by geographic locale and linguistic differences such as vocabulary, expression and meaning. This in turn will assuredly change the way in which new terms in Chinese are transmitted.

As for this field known as the internet, the borders of cultural exchange exist in multiple layers. From the point of Wikipedia, the articles, the editing tools, the authors and the readers exist as intermediaries. Above all, as a practical problem, it is necessary to observe the shape of the present advances made by Wikipedia. I hope to see further research from various directions.