
Calculating the Distance Between Near-Synonyms Across Languages: A Case Study on Chinese and Japanese

Shi Jianjun

Institute of Linguistics, Shanghai International Studies University, Shanghai, China

Email address:

jianjunshi_cn@126.com

To cite this article:

Shi Jianjun. Calculating the Distance Between Near-Synonyms Across Languages: A Case Study on Chinese and Japanese. *International Journal of Language and Linguistics*. Vol. 7, No. 2, 2019, pp. 85-92. doi: 10.11648/j.ijll.20190702.15

Received: March 19, 2019; **Accepted:** April 26, 2019; **Published:** May 15, 2019

Abstract: The study of differences between near-synonyms across languages has always been a hot area of research in foreign language teaching and cross-language comparison. A linguistic phenomenon received special attention, Chinese-Japanese homographs often confuse learners since they have same forms yet differ slightly in terms of their meanings and usages. Traditional studies were unfolded mainly in two aspects: case studies on homograph discrimination and division of homographs according to meaning & usage distance. Researches concerning the latter aspect tend to distinguish near-synonyms between languages by means of a three-way classification, i.e. synonyms, heteronyms, and near-synonyms. However, this classification is far from satisfactory in that they cannot measure “near-synonymy” in an accurate and gradable manner since the term “near-synonymy” per se is too broad a term to define. This paper proposes a statistical method for calculating near-synonyms across languages by means of parallel corpus, where *translation ratio*, *F-measure* and *inter-translation ratio* are taken into account as parameters. By means of the F-value, this highly productive method is not only applicable in classifying synonyms, heteronyms, and near-synonyms between Chinese and Japanese, but also allows measuring the meanings & usage distance between cross-linguistic near-synonyms. To prove the effectiveness of this method, around 1900 pairs of Chinese and Japanese near-synonyms are compared and has gained good effects.

Keywords: Near-Synonyms, Meaning & Usage Distance, Inter-Translation Ratio, F-measure

1. Introduction

The issue of near-synonyms has always been a hot topic in the fields of linguistics and Natural Language Processing. [1] On the one hand, many resources were developed from the perspective of traditional linguistics, e.g. “Tongyici Cilin” (*Synonymy Thesaurus*) [2]. On the other hand, however, the NLP approach proposed building algorithm which is based on the word embedding model extracting from large-scale corpus to calculate semantic similarity by measuring the distance between word vectors. Researches mentioned above aim at depicting and calculating semantic similarity of language [3]. Cross-language lexical similarity, therefore, has both academic and practical value on cross-language information processing, comparative studies of languages and Foreign Language Teaching (FLT). However, more related issues need to be further studied [4]. Among them are such question as how to measure the meaning & usage distance and how to classify a large number of cross-linguistic near-synonyms

according to this measurement. In this paper, the author focuses on the near-synonyms between Chinese and Japanese and provides a statistical method that measures the meaning & usage distance between them.

In 1972, the Agency for Cultural Affairs, which affiliates to the Japanese government, asked the Language Education Institute of Waseda University to carry out a research project, titled “中国語と対応する漢語について” (*On Kanji and their logographic Chinese counterparts*). The theme of this project is to select some commonly-used Chinese-Japanese homographs and to classify them as near-synonyms according to word meanings. The outcome of this project is a book titled “中国語と対応する漢語” (*Kanji and their logographic Chinese counterparts*) [5], which was published by the Agency for Cultural Affairs. In this book, those Kanji collected were divided into three categories, namely, *Same* (S), *Overlap* (O), and *Different* (D), which were based on the

morphological and semantic similarities with their Chinese counterparts. The book claims that most Chinese-Japanese homographs share the same or very similar meanings, that is, they are in the category S. There are about 80 homographs with overlapping yet different meaning & uses, which account for about 4% of the total words collected.

In terms of the above study, different opinions [6, 7] can be summarized as follows: firstly, there are problems in defining what counts as homographs (see Shi [8] for a detailed discussion); secondly, the criteria used in homographs classification in the book is not without a question as well; thirdly, it's difficult to determine which category a certain word belong in under specific conditions. The latter two issues involve cross-linguistic near-synonyms. Japanese scholar Otsuka [9] believes that cross-linguistic near-synonyms between Chinese and Japanese should be categorized according to the following criteria: whether there exists a translatable relationship between cross-linguistic near-synonyms. That is, if cross-linguistic near-synonyms can be used as synonyms in the target language, they can be regarded as synonyms (i.e. category S); If cross-linguistic near-synonyms can sometimes be translated into synonyms in the target language, and sometimes not, they are near-synonyms (i.e. category O); if they are not inter-translatable at all, the cross-linguistic near-synonyms are heteronyms (i.e. category D). Based on this criterion, Otsuka classified near-synonyms of function words in Chinese and Japanese (including *Modern Chinese 800 Words*) and found that the classification results were quite different from those in *Kanji and their logographic Chinese counterparts*.

Otsuka's revision to *Kanji and their logographic Chinese counterparts* was carried out on the premise of recognizing the feasibility of three-way classifications of cross-linguistic near-synonyms. Otsuka raises no objection to the classification method per se. However, does the three-way classification of Chinese and Japanese near-synonyms proposed by the Institute of Language Education of Waseda University sticks to linguistics facts? Can the three-way classification really describe the difference between Chinese and Japanese near-synonyms? How can we precisely measure the meaning & usage gap between Chinese and Japanese near-synonyms? This paper will explore these issues in further details.

2. Near-Synonyms Between Chinese and Japanese Are Some Continuums

The common method adopted to understand language phenomena is classification. Therefore, as mentioned above, classifying near-synonyms by means of their meaning & usage is used to study Chinese and Japanese near-synonyms. These near-synonyms are formed in the process of long-term language contact between Chinese and Japanese. The differences between the meaning & usage of them emerges along with the respective evolvement of the two languages. Roughly speaking, Chinese and Japanese near-synonyms can

be divided into three categories according to their different meaning & usage: *Same* (S), *Overlap* (O), and *Different* (D). However, since the emergence of different meaning & usage between Chinese and Japanese near-synonyms is a dynamic process, which means it is by no means a clear-cut matter. For example, The difference between the meanings of the words “椅子（椅子）” is relatively small, because both words refer to “a piece of furniture for one person to sit on, with a back, a seat and four legs” in Chinese and Japanese, and their usages are basically the same. Another case is that both of the words “话题（话题）” refer to the concept of “the subject of conversation (topic)”, i.e. with the same meaning and syntactic function, in Chinese the word “话题” hardly modifies a noun as an attributive while in Japanese it is common for the word “話題” to modify a noun as an attributive, such as “話題の人物、話題の商品”. Therefore, the meaning & usage distance with regards to the word pair “椅子（椅子）” is larger than that of “话题（话题）”. Still another example is the word “人选（人選）”, the annotation of which in *The Contemporary Chinese Dictionary* [10] is “Persons selected for a certain purpose”. Its Japanese counterpart, quoted from *Gakken Japanese Dictionary* [11] notes “その仕事をするにふさわしい人を選ぶこと” which means “to select people fit for that job”. From the annotation in dictionaries, the pair of words “人选（人選）” seem to have roughly the same lexical meaning in Chinese and Japanese, but further analyses reveal that there are in fact some differences. In this case, Chinese focuses on “persons”, while Japanese focuses on the “selection of people”. From the perspective of grammatical function, in Chinese “人选” can only be used as nouns, while in Japanese “人選” can be used as either nouns or verbs. The gap between Chinese and Japanese in the case of “人选（人選）” is larger than that of “椅子（椅子）” and “话题（话题）”. To sum up, according to the classification of *Kanji and their logographic Chinese counterparts*, the two words are near-synonyms because they have overlapping meaning and usage. However, these cannot account for the subtle differences between these two words, with regards to word pairs that may have a smaller or larger gap. So how to quantify the differences?

Although the three-way classification of synonyms, heteronyms and near-synonyms proposed in *Kanji and their logographic Chinese counterparts* can provide a useful guide in distinguishing the differences Chinese and Japanese near-synonyms, but there is also a need to describe the minute differences in a more accurate manner. Synonyms and heteronyms are two polar opposites from the perspective of meaning & usage of near-synonyms. There is no doubt that there exist these two kinds of words in different languages. The key point, however, is that we need to account for the grey areas in-between, which cannot be easily subject to a very general term of “near-synonyms”. This is because the gap between Chinese and Japanese near-synonyms in terms of meaning & usage can be described in many aspects. In addition, such discrepancies in meaning and usage can be seen as a continuous data axis with synonyms at one end and heteronyms at the other. The discrepancies, large or small, spreads all over this data axis. We can therefore resort to a

quantitative way in measuring the differences between Chinese and Japanese near-synonyms with regards to their meaning & usage.

3. The Inter-Translation Ratio for Near-Synonyms

According to the foregoing discussion, the differences between near-synonyms vary, in which synonyms have the smallest one, and heteronyms the greatest. However, those near-synonyms varies wildly so much so that the traditional classification method is very much imprecise. Here, we define the differences between Chinese and Japanese near-synonyms as against the distance between Chinese and Japanese near-synonyms. If we could find a method to measure the distance, we can accurately calculate and describe them, which also serves a practical purpose for automatic classification of Chinese-Japanese near-synonyms, Natural language processing for semantic analyses, as well as Japanese-Chinese teaching.

3.1. Inspiration from Traditional Studies

Japanese scholar Ootsuka proposed his own classification method in view of the classification errors of some specific words in “*Kanji and their logographic Chinese counterparts*”. He claimed that the classification standard for judging Chinese-Japanese near-synonyms should be whether they are inter-translatable and reclassified the function words in “*Xiandai Hanyu Babai Ci*” (*Modern Chinese 800 Words*) [12] with this method, which is effective and feasible for manual classification of a small number of near-synonyms. For example, we can make a correct judgment of pairs of near-synonyms like “简单(简单)” with overlapping meaning & usage in Chinese and Japanese. However, this method also has its own weakness. On the one hand, there are not only Chinese and Japanese near-synonyms like “贵重(贵重)”, in which the meaning & usage of one word cover the other, but also Chinese and Japanese near-synonyms like “深刻(深刻)”, in which there is only a small crossover between them. Specifically, while the word “贵重” in Chinese can be translated into the word “贵重” in Japanese, the word “贵重” in Japanese may not always be translated into the word “贵重” in Chinese in every case. That is, the standard of “whether inter-translatable or not” cannot accurately describe the meaning & usage distance of Chinese and Japanese near-synonyms such as “贵重(贵重)”. The criterion “whether inter-translatable or not” is a rigid one, but the meaning & usage distance of Chinese and Japanese near-synonyms is a continuum, which requires a more complex method to describe. On the other hand, the classification criterion put forward by Ootsuka is designed mainly for manual work in classifying Chinese-Japanese near-synonyms. Whether they are inter-translatable largely depends on the judgement of those who undertake the classification work. This entails those people who make such judgments to have a high language proficiency in both Chinese and Japanese. Otherwise, they

cannot accurately capture the nuances between Chinese and Japanese near-synonyms. Although Ootsuka’s method has taken a step forward from the traditional classification method, it is still not satisfactory in describing the meaning & usage distance of synonyms in detail. Nevertheless, the practice of Ootsuka has inspired us, that is, whether we can use the mathematical parameters reflecting the inter-translatability of Chinese and Japanese near-synonyms to measure the meaning & usage distance of Chinese and Japanese near-synonyms.

3.2. The Inter-Translation Ratio of Chinese-Japanese Near-Synonyms

One mathematical parameter we can easily think of is the inter-translation ratio of near-synonyms. We assume that, within a pair of Chinese and Japanese near-synonyms, the Japanese translation of the Chinese word W_{ch} is its Japanese near-synonym W_{ja} , and the Chinese translation of the Japanese word W_{ja} is its Chinese near-synonym W_{ch} . Then the inter-translation ratio of Chinese and Japanese near-synonyms is the ratio of the Chinese word W_{ch} translated into the Japanese word W_{ja} and that of the Japanese word W_{ja} translated into the Chinese word W_{ch} in the case of actual translation. This ratio can be obtained from a large-scale Chinese-Japanese parallel corpus. Assuming that the frequency of source language W_{ch} used in Chinese-Japanese parallel corpus is F_{SC} and the frequency of W_{ch} translated into W_{ja} in corpus is F_{TJ} , the ratio of Chinese word W_{ch} translated into Japanese near-synonym W_{ja} in Chinese-Japanese parallel corpus (F_{CJ}) can be calculated by the following formula:

$$F_{CJ}=F_{TJ}/F_{SC} \quad (1)$$

Similarly, assuming that the frequency of the source language W_{ja} used in corpus is F_{SJ} and that the frequency of W_{ja} translated into W_{ch} in the corpus is F_{TC} , the ratio of Japanese word W_{ja} translated into Chinese near-synonym W_{ch} in the Chinese-Japanese parallel corpus (F_{JC}) can be calculated by the following formula:

$$F_{JC}=F_{TC}/F_{SJ} \quad (2)$$

3.3. Features of the Inter-Translation Ratio of Chinese-Japanese Near-Synonyms

Assuming that the Chinese-Japanese parallel corpus we use is large enough and that the translation of each pair of near-synonyms in the corpus are from different sources, the translation of near-synonyms in the corpus would be less affected by personal factors, so that the translation ratio of the near-synonyms mentioned above can truly reflect the translation of near-synonym. If this assumption holds, then F_{CJ} and F_{JC} should have the following properties:

- 1) Within the same pair of near-synonyms, when F_{CJ} and F_{JC} are close to 1 at the same time, it indicates that it is highly possible that the word be translated into the near-synonyms in the target language in the corpus. That is to say, the meaning & usage distance of this pair of Chinese and Japanese near-synonyms is very close,

which corresponds to the synonyms in the traditional sense. To illustrate this point, we carried out the statistical analysis using the “Chinese-Japanese Parallel Corpus” developed by Beijing Center for Japanese

Studies [13]. The following table shows the frequencies and translations of examples of pairs of Chinese-Japanese near-synonyms whose inter-translation ratios are close to 1 at the same time.

Table 1. F_{CJ} and F_{JC} close to 1 simultaneously.

Chinese	Japanese	Numbers of Chinese texts	Numbers of Chinese-Japanese translation	Chinese-Japanese translation ratio	Numbers of Japanese text	Numbers of Japanese-Chinese translation	Japanese-Chinese translation ratio
政治	政治	2385	2344	0.98	198	182	0.92
政策	政策	1733	1708	0.99	130	118	0.91
全国	全国	1246	1217	0.98	114	107	0.94
矛盾	矛盾	625	595	0.95	73	69	0.95
科学	科学	524	513	0.98	90	85	0.94
企业	企業	422	418	0.99	250	234	0.94
艺术	芸術	211	199	0.94	68	67	0.99
世纪	世紀	186	182	0.98	124	117	0.94

2) Within the same pair of near-synonyms, when F_{CJ} and F_{JC} are close to 0 at the same time, it indicates that it is very unlikely that the word be translated into the near-synonyms in the target language in the corpus. In this case, it can be deduced that this pair of near-synonyms can be regarded as heteronyms in the

traditional classification since its meaning & usage distance is quite far. For example, words like “结构 (結構)、迷惑 (迷惑)、恰好 (恰好)” have totally different meaning & usage between Chinese and Japanese. In the corpus we use, the inter-translation ratio is 0. The details are presented in the table below.

Table 2. F_{CJ} and F_{JC} close to 0 simultaneously.

Chinese	Japanese	Numbers of Chinese texts	Numbers of Chinese-Japanese translation	Chinese-Japanese translation ratio	Numbers of Japanese text	Numbers of Japanese-Chinese translation	Japanese-Chinese translation ratio
迷惑	迷惑	50	0	0	109	0	0
结构	結構	83	0	0	97	0	0
用事	用事	14	0	0	86	0	0
气味	気味	46	0	0	78	0	0
下手	下手	54	0	0	76	0	0
大抵	大抵	23	0	0	66	0	0
恰好	恰好	44	0	0	66	0	0

3) Within the same pair of near-synonyms, if $F_{CJ} > F_{JC}$, that is, in Chinese-Japanese translation, it's more likely that the Chinese words are translated with their corresponding Japanese near-synonyms, but in Japanese-Chinese translation, it's less likely the case that Japanese words are translated into their corresponding Chinese

near-synonyms. This shows that there are overlaps between Chinese and Japanese near-synonyms in terms of meaning & usage, but the meaning & usage of Chinese is less than that of Japanese, that is, the meaning & usage of Japanese is more than that of Chinese. Some near-synonyms of this kind are listed in the table below.

Table 3. $F_{CJ} > F_{JC}$.

Chinese	Japanese	Numbers of Chinese texts	Numbers of Chinese-Japanese translation	Chinese-Japanese translation ratio	Numbers of Japanese text	Numbers of Japanese-Chinese translation	Japanese-Chinese translation ratio
说明	説明	362	159	0.44	301	89	0.30
文字	文字	60	26	0.43	133	36	0.27
发达	発達	159	63	0.40	125	34	0.27
姿势	姿勢	38	16	0.42	114	34	0.30
使用	使用	239	84	0.35	77	23	0.30
真实	真実	129	47	0.36	77	22	0.29

4) Within the same pair of Japanese synonyms, if $F_{CJ} < F_{JC}$, in other words, in Japanese-Chinese translation, it's highly possible that the Japanese words are translated with their corresponding Chinese near-synonyms while in Chinese-Japanese translation it's less possible that Chinese words are translated into their corresponding

Japanese near-synonyms. This indicates that there are overlaps between Chinese and Japanese near-synonyms in terms of meaning & usage, but the meaning & usage of Chinese is less than that of Japanese, that is to say, the meaning & usage of Chinese is more than that of Japanese. Some of these words are listed in the table below.

Table 4. $F_{CJ} < F_{JC}$.

Chinese	Japanese	Numbers of Chinese texts	Numbers of Chinese-Japanese translation	Chinese-Japanese translation ratio	Numbers of Japanese text	Numbers of Japanese-Chinese translation	Japanese-Chinese translation ratio
机会	機会	454	144	0.32	199	144	0.72
努力	努力	670	388	0.58	191	144	0.75
不可	不可	1080	181	0.17	181	78	0.43
一般	一般	697	332	0.48	171	101	0.59
以前	以前	468	128	0.27	152	76	0.5
比较	比較	695	245	0.35	150	95	0.63
希望	希望	985	365	0.37	150	87	0.58

From the aforementioned features, we can see that, when the Chinese-Japanese translation ratio and the Japanese-Chinese translation ratio are close to 1 at the same time, the meaning & usage distance of this pair of near-synonyms is the closest and can be regarded as synonyms. When these two parameters are close to 0 at the same time, the meaning & usage distance of this pair of near-synonyms is the farthest and can be regarded as heteronyms. If these two translation ratios do not fit into any one of these scenarios, then the meaning & usage distance of this pair of near-synonyms falls in between the above situation. In traditional sense, these words belong to near-synonyms. From the above analysis, we can also see that the meaning & usage distance of these Chinese and Japanese near-synonyms varies and the situation is quite complex, which is reflected in the ratio value and the interrelation of the two translation ratios. So the inter-translation ratio can reflect the inter-translation situation of Chinese and Japanese near-synonyms and observe the meaning & usage distance of Chinese and Japanese near-synonyms. Since translation is directional, the translation of near-synonyms must be measured by Chinese-Japanese translation ratio as well as Japanese-Chinese translation ratio. Those near-synonyms with the closest meaning & usage distance must have two translation ratios approaching 1 at the same time, not a single one of these conditions can be dispensed with in depicting its synonymity. Take the pair of near-synonyms “处分（処分）” for example, in our parallel corpus, there are 51 cases of “处分” in Chinese, 47 cases of which have been translated into Japanese “処分”. The Chinese-Japanese translation ratio is 92%, which is close to 1. But there are 37 cases of “処分” in Japanese, and only 10 cases of them have been translated into “处分” in Chinese, with a translation ratio of 29%. The Japanese-Chinese translation ratio of the near-synonym “注意” is 90%, while its Chinese-Japanese translation ratio is only 49%. Although one of the translation ratios of the two pairs of near-synonyms (Chinese-Japanese or Japanese-Chinese) is close to 1, the meaning & usage distance between these two words is still quite large. Therefore, if we use translation ratio to measure the meaning & usage distance between Chinese and Japanese near-synonyms, we must take the two translation ratios into consideration at the same time. Otherwise, it cannot accurately reflect the actual situation of the difference in meaning & usage between Chinese and Japanese

near-synonyms.

4. Using F-measure as the Inter-Translation Coefficient

According to the above analysis, inter-translation ratio can be used to measure the meaning & usage distance between the meaning & usage of Chinese and Japanese synonyms. However, it is necessary to consider both the Chinese-Japanese translation ratio as well as Japanese-Chinese translation ratio. Metaphorically speaking, two rulers are needed to measure the inter-translation ratio, which is effective and feasible for studying synonyms and heteronyms. However, the situation of Chinese-Japanese near-synonyms is very complicated. It can be said that Chinese-Japanese synonyms are spread in between synonyms and heteronyms, forming a continuous line between these two extremes.

The meaning & usage distance of these Chinese-Japanese near-synonyms varies. Those near-synonyms having smaller meaning & usage distance are closer to synonyms, while those having larger meaning & usage distance are closer to heteronyms. However, this measurement of meaning & usage distance is inconvenient since both Chinese-Japanese translation ratio as well as Japanese-Chinese translation ratio are taken into account, in addition, applying this method to the large-scale text processing would be quite complicated and troublesome. Therefore, if we want to measure and compare the meaning & usage distance between different pairs of Chinese-Japanese near-synonyms and to visualize the result by putting these near-synonyms along the line aforementioned according to the meaning & usage distance between Chinese and Japanese, it is more preferable to use only one parameter to calculate the distance. This parameter must be taken into account for both Chinese-Japanese translation ratio and Japanese-Chinese translation ratio, so that the size of this parameter is proportional and the meaning & usage distance between Chinese and Japanese near-synonyms should be in a relation of this proportion (either direct or inverse). That is to say that if both translation ratios value are large at the same time within one pair of near-synonyms, the value of this parameter should also be large (or vice versa), which indicates that the meaning & usage distance of this pair of near-synonyms are small, i.e. they have similar meanings. If one translation ratio is large while the other is small, it means that the meaning & usage of the near-synonyms has an overlap

but there is still a gap, and the parameter should not be large. If both translation ratios are very small, it shows that this pair of near-synonyms has large meaning & usage distance, it is likely to be a heteronym, then this parameter should be very small (or vice versa).

4.1. Constructing F-measure Using the Inter-Translation Ratio of Chinese-Japanese Near-Synonyms

In natural language processing, F-measure is often used to evaluate the outcomes of natural language processing. A good result should be seen as all the goals needed being listed in the output, with as less irrelevant information as possible. The former requirement is generally measured by recall rate, while the second requirement is measured by precision rate. To evaluate the quality of computer output, it is necessary to comprehensively evaluate the recall rate and the correctness rate of output at the same time. For example, if output contains required goals, that is, the precision rate is very high, but most of the required goals as a whole do not appear, that is, the recall rate is very low. In this case the results should not be regarded as good. For good output, both the recall rate and the precision rate must be high. Natural language processing research concerning this issue usually uses F-measure to synthesize the accuracy and recall rate to evaluate the output.

From here we see that when measuring the usage distance of Chinese-Japanese near-synonyms by inter-translation ratio, we must also take into account both Chinese-Japanese translation ratio and Japanese-Chinese translation ratio. Those pairs of near-synonyms with high values on these two parameters have the smallest meaning & usage distance. This is consistent with the evaluation of information retrieval results by F-measure. Therefore, we can use F-measure as a parameter to describe the meaning & usage distance of Chinese-Japanese near-synonyms (F value for short). This parameter is constructed by inter-translation ratio of Chinese-Japanese near-synonyms.

Assuming that within a pair of Chinese-Japanese near-synonyms whose Chinese-Japanese translation ratio and Japanese-Chinese translation ratio are F_{CJ} and F_{JC} respectively, then the parameter F-measure reflecting the usage distance between this pair of Chinese-Japanese near-synonyms can be calculated by the following formula:

$$F \text{ measure} = \frac{2 \times F_{CJ} \times F_{JC}}{F_{CJ} + F_{JC}} \quad (3)$$

Formula (3) shows that the value of F-measure is proportional to the product of the Chinese-Japanese translation ratio F_{CJ} and the Japanese-Chinese translation ratio F_{JC} . According to the previous analysis, we can know that if both of these two values are very large, the meaning & usage distance of this pair of near-synonyms are relatively small. From formula (3), it can be deduced that the value of F-measure is also very large at this time. If only one of these two ratios is large while the other is relatively small, which signifies a large meaning & usage distance, as can be seen from formula (3). Certainly, the value of F-measure wouldn't be large at this time. If one of these two translation ratios is 0, which suggests that the word is heteronym in most cases, and the meaning & usage distance between Chinese and Japanese is the greatest. It can be inferred from the formula that the value of F-measure at this time is 0. Thus, the F-measure reflects both the Chinese-Japanese translation ratio and the Japanese-Chinese translation ratio of Chinese-Japanese near-synonyms, and integrates these two ratios into a numerical value, which can be used to describe the meaning & usage distance between Chinese-Japanese near-synonyms.

4.2. Validation of F-measure to Measure the Meaning & Usage Distance of Near-Synonyms

In order to test the validity of F-measure in describing and distinguishing the meaning & usage distance of Chinese-Japanese near-synonyms, a statistical analysis was conducted on about 1900 pairs of frequently used Chinese-Japanese near-synonyms using Chinese-Japanese parallel corpus, and the F-measure of each pair of near-synonyms were calculated according to the two translation ratios. Then these near-synonyms are arrayed according to the value of F-measure in descending order. As a result, these 1900 pairs of near-synonyms have gradually transitioned from synonyms to heteronyms, forming a nearly continuous line. Due to the limitation of space, only three cases are listed here: F value greater than 0.9, F value near 0.5 and meaningless F value. From these three tables below, it proves that the meaning & usage distance of Chinese-Japanese near-synonyms gradually increases as that of F value gradually decreases.

Table 5. F value > 0.9 (Synonyms).

Chinese	Japanese	Numbers of Chinese text	Numbers of Chinese-Japanese translation	Numbers of Japanese text	Numbers of Japanese-Chinese translation	F value
改革	改革	1005	993	675	624	0.96
政策	政策	2029	2008	1050	963	0.95
原则	原則	852	845	233	209	0.94
全国	全国	1338	1310	1102	988	0.94
分析	分析	377	356	494	456	0.93
技术	技術	493	494	1077	924	0.92
社会	社会	1427	1551	2984	2388	0.92
统一	統一	804	758	332	298	0.92
教授	教授	241	221	476	435	0.92
解决	解決	1620	1513	538	474	0.91
文化	文化	770	731	1035	898	0.91

Chinese	Japanese	Numbers of Chinese text	Numbers of Chinese-Japanese translation	Numbers of Japanese text	Numbers of Japanese-Chinese translation	F value
目标	目標	354	327	573	508	0.90
企业	企業	535	556	2672	2139	0.90
具体	具体	619	591	543	464	0.90
干部	幹部	1626	1615	482	397	0.90

Table 6. $F \text{ value} \approx 0.5$ (Synonyms).

Chinese	Japanese	Numbers of Chinese text	Numbers of Chinese-Japanese translation	Numbers of Japanese text	Numbers of Japanese-Chinese translation	F value
重大	重大	327	168	242	137	0.54
利用	利用	313	249	1040	419	0.53
发表	発表	348	273	2251	878	0.52
时间	時間	1171	585	2849	1529	0.52
主要	主要	1174	458	360	274	0.52
倾向	傾向	276	179	400	168	0.51
时期	時期	885	587	610	252	0.51
完全	完全	1166	482	362	220	0.49
认识	認識	858	502	507	210	0.49
一般	一般	686	323	633	306	0.48
必要	必要	496	477	2695	854	0.48
注意	注意	933	428	452	223	0.48
共同	共同	425	164	588	347	0.47
希望	希望	1013	352	528	374	0.47
以来	以来	412	143	624	439	0.46
部分	部分	281	175	648	236	0.46
使用	使用	244	92	637	373	0.46
状况	狀況	235	135	1475	555	0.45

Table 7. Meaningless F value (heteronyms).

Chinese	Japanese	Numbers of Chinese text	Numbers of Chinese-Japanese translation	Numbers of Japanese text	Numbers of Japanese-Chinese translation	F value
结构	結構	87	0	124	0	divisor is equal to zero
迷惑	迷惑	46	0	179	0	divisor is equal to zero
下手	下手	41	0	87	0	divisor is equal to zero
气味	氣味	48	0	62	0	divisor is equal to zero
恰好	恰好	34	0	59	0	divisor is equal to zero
暴乱	乱暴	15	0	58	0	divisor is equal to zero
上手	上手	13	0	124	0	divisor is equal to zero
素质	質素	69	0	14	0	divisor is equal to zero
脱出	脱出	12	0	67	0	divisor is equal to zero
心肝	肝心	12	0	59	0	divisor is equal to zero
作法	作法	12	0	48	0	divisor is equal to zero
喧哗	喧嘩	7	0	158	0	divisor is equal to zero
贫乏	貧乏	6	0	62	0	divisor is equal to zero
皮肉	皮肉	6	0	50	0	divisor is equal to zero
见地	見地	5	0	23	0	divisor is equal to zero

The case where F-measure is meaningless is that both the Chinese-Japanese translation ratio F_{CJ} and Japanese-Chinese translation F_{JC} are 0 within a pair of Chinese-Japanese near-synonyms, so the divisor $F_{CJ} + F_{JC}$ in formula (3) is 0, which makes the F-measure value impossible to calculate. In this case, Chinese-Japanese near-synonyms are not inter-translatable, that is, there is no overlap between Chinese and English in terms of meaning & usage. In fact, such words are Chinese-Japanese heteronyms. To sum up, when F value is meaningless, the meaning & usage distance between these pairs of Chinese-Japanese heteronyms is infinite, that is, they are heteronyms.

5. Concluding Remarks

The differences of meaning & usage between Chinese and Japanese is not an either-or issue, but a complicated one. If we use the meaning & usage distance between Chinese and Japanese near-synonyms to describe the difference between Chinese and Japanese, then the collection of the distance between Chinese and Japanese near-synonyms will approximately form a continuous line with synonyms and heteronyms as two endpoints. It is this fact that renders the traditional three-way classification method of synonyms, heteronyms and near-synonyms impossible to accurately and

objectively describe the differences in meaning & usage between Chinese-Japanese near-synonyms. In order to solve this problem, a large-scale Chinese-Japanese parallel corpus was adopted to calculate the inter-translation ratio of Chinese and Japanese near-synonyms, and the F-measure, based on inter-translation ratio of Chinese and Japanese near-synonyms was conducted to measure the meaning & usage distance between Chinese and Japanese near-synonyms. Since translations provided in the corpus are indeed authoritative, it can objectively reflect the translation situation of Chinese-Japanese near-synonyms as well as the gap between them. Therefore, this method can not only avoid the influence of personal factors in judging the differences of Chinese-Japanese near-synonyms in virtue of translator's wisdom, but also describe the subtle differences between each pair of Chinese-Japanese near-synonyms in detail. Taking the Chinese-Japanese parallel corpus as the knowledge base, the meaning & usage distance of about 1900 commonly used Chinese-Japanese near-synonyms were calculated by applying the F-measure parameter and were arrayed according to the value of F-measure in descending order. We see that synonyms are words whose F value are close to 1; heteronyms are words whose F value close to 0; near-synonyms are those words lie in between these two extremes. The meaning & usage distance of Chinese-Japanese near-synonyms is also reflected by the F value, which can be calculated accurately, thus verifying the feasibility of this method.

In this paper, the term Chinese-Japanese near-synonyms is a very broad concept, which includes synonyms and heteronyms. In fact, synonyms and heteronyms are regarded as special types of near-synonyms. When the meaning & usage distance of Chinese-Japanese near-synonyms (F value) is large, they become heteronyms. Conversely, when the meaning & usage distance of Chinese-Japanese near-synonyms (F value) is small, they become synonyms. Another issue worth mentioning is the so-called meaning & usage of near-synonyms. Strictly speaking, it should include two aspects: "meaning" and "usage", while the "meaning" aspect of Chinese-Japanese near-synonyms should include "lexical meaning" and "grammatical meaning". The F value proposed in this paper is an integrated embodiment of the meaning & usage distance of Chinese-Japanese near-synonyms. As for how to measure the differences between Chinese-Japanese near-synonyms from lexical, grammatical and usage aspects is a new issue, which awaits further studies. This paper is based on quantitative studies of meaning & usage distance of Chinese-Japanese Homograph. [14]

References

- [1] Shi J. J, Hong J. (2013) On a General Methodology for Contrastive Analysis of Chinese-Japanese Homographs. *Foreign Language Teaching and Research*, 45, pp 531-542.
- [2] Duan Y. G, Liu Y, Yu Sh. W. (2018) An Embedded Representation for "Tongyici Cilin" and Its Evaluation on Tasks, *Journal of Xiamen University (Natural Science)*, 57(06), pp 867-875.
- [3] Li X. T, You Sh. J, Chen W. (2019), An Algorithm of Semantic Similarity between Words Based on Word Single-Meaning Embedding Model, *Acta Automatica Sinica*, DOI: 10.16383/j.aas.c180312.
- [4] Shi J. J. (2013) The Current Situation and Subjects of Synchronic Comparative Studies of Chinese-Japanese Homographs, *Foreign Language Research in Northeast Asia*, 1, pp 4-9.
- [5] The Agency for Cultural Affairs (1978) Kanji and their logographic Chinese counterparts, *Research Materials for Japanese Language Education*, pp 85-143.
- [6] Arakawa (1979) Chinese and Kanji: A Book Review on "Kanji and their logographic Chinese counterparts", *Aichi Forum on Language and Literature*, 62, pp 388-361.
- [7] Oukouchi (1992) Homographs of Japanese and Chinese, *Collection of Comparative Studies of Japanese and Chinese Languages*, pp 179-215.
- [8] Shi J. J, Xu X. H. (2014) On Homographs in Chinese and Japanese Languages: a revisit. *Journal of PLA University of Foreign Languages*, 37, pp 132-139.
- [9] Ootsuka (1990) On Chinese-Japanese Homographs. *Collection of Studies in foreign language education*, 12, 327-337.
- [10] Institute of Linguistics CASS (2005) *The Contemporary Chinese Dictionary* 5th ed., Beijing: The Commercial Press.
- [11] Kindaichi, Ikeda (1988) *Gakken Japanese Dictionary*, 2nd ed, Tokyo: Gakken.
- [12] Lv Sh. X. (1980) *Modern Chinese 800 Words*, Beijing: The Commercial Press.
- [13] Beijing Center for Japanese Studies (2003) *Chinese-Japanese Parallel Corpus*, Beijing: Beijing Center for Japanese Studies. (Fruition of the Project: The Development and Application of Chinese-Japanese Parallel Corpus, No. 99BYY007).
- [14] Shi J. J, Qiao Y. (2016) Measuring the Semantic Distance of Chinese-Japanese Homographs: Approached from the F-measure Constructed on the Basis of Translation Ratio. *Journal of PLA University of Foreign Languages*, 39, pp 76-84+160.