Extracting Conceptual Differences between Translation Pairs Using Multilingual WordNet

Ikkyu Nishimura[1,2], Yohei Murakami[1,3], Mondheera Pituxcoosuvarn[1,4],
[1] Faculty of Information Science and Engineering,
Ritsumeikan University, Ibaraki, Osaka, Japan
[2] is0368xk@ed.ritsumei.ac.jp [3] yohei@fc.ritsumei.ac.jp [4] mond-p@fc.ritsumei.ac.jp

Abstract

The concepts expressed by words and their translations in different languages do not always align. Despite advancements in machine translation, such differences can still lead to misunderstandings. Therefore, we proposed a method to extract conceptual differences in translation pairs from WordNet and Open Multilingual WordNet. We applied the proposed method to Japanese, Chinese, and Indonesian WordNets to investigate how many translation pairs with conceptual differences they have. Furthermore, we validated the extracted conceptual differences using human evaluators.

1 Introduction

Improvements in machine translation (hereinafter, this is called "MT") are reducing language barriers in multilingual communication. However, since languages evolve with their unique cultures and histories, their conceptual systems can differ, leading to misunderstandings. In particular, Asian languages tend to have lower similarity to English compared to European languages (Chiswick and Miller, 2005). This may result in a correct translation, but with differences in the concepts expressed between words. In this study, we propose a method to quantify the concepts represented by words using WordNet and Open Multilingual WordNet (hereinafter, this is called "OMW") to identify such conceptual differences across languages. We focus on Japanese, Chinese, and Indonesian, detecting conceptual differences across the three languages based on WordNet Synsets.

The following two challenges can be identified:

Quantification of the range of concepts
It is challenging to articulate the meanings and ranges of concepts associated with words because concepts are abstract. Therefore, we extract the linked concepts of each word as a set using a conceptual dictionary, allowing us to quantify the concepts of each word.

Detection of conceptual differences
Due to the unique conceptual systems of each language, comparing concepts expressed by words across languages is challenging. To address this, we utilize WordNet for the English concept system and OMW for aligning other languages, allowing us to compare and extract conceptual differences based on the English system.

2 Conceptual Differences in Multilingual Communication

2.1 Conceptual Difference

The advent of MT is making communication with people from different cultures and languages easier. However, communication through MT carries the potential for misunderstandings due to differences in the concepts expressed by words. For example, Yamashita and Ishida (2006) analyzed communication between speakers of different languages, observing that in conversations using MT, dialogues sometimes broke down when the meaning of a polysemous word changed during translation. In that case, the difference in concepts arose because the word had multiple meanings, but even words with the same meaning can express different concepts. For example, in the Japanese and English translation pair "団子 (dango)" and "dumpling", the latter broadly refers to boiled, ball-shaped foods, while Japanese distinguishes "団子 (dango)" as items without fillings, categorizing stuffed foods like "餃子 (gyoza)" and "小籠包 (xiao long bao)" differently. In translation, there is no exact equivalent for the word "団子 (dango)" in the English conceptual system, so the translation result becomes "dumpling".

This issue is also present in the construction of Multilingual WordNet. Fellbaum and Vossen

Table 1: Conceptual differences between English and other languages

Lang	Total Word Pair	Concept Difference	Ratio
ja	477,031	33,190	7.0%
zh	199,724	12,327	6.1%
id	220,128	20,587	9.4%
nl	245,068	15,857	6.5%
fr	313,473	14,974	4.8%
es	165,224	8,705	5.3%
it	228,055	144,36	6.3%

※ja:Japanese, zh:Chinese, id:Indonesian, nl:Dutch, fr:French, es:Spanish, it:Italian

(2012) pointed out that even words deemed synonymous between different languages often only partially overlap in their meanings and concepts. Multilingual WordNet is created based on the synsets of the English WordNet. As a result, in non-English languages, a word may not correspond to a single synset but may instead be linked to multiple adjacent synsets (hypernyms and hyponyms). This happens because the conceptual system of the target language may differ from that of English. Table.1 illustrates the number of translation pairs where non-English words appear across adjacent synsets. It reveals that less than 10% of word pairs exhibit such differences. In this study, we define such variations in the concepts expressed by words as conceptual differences and focus on detecting these differences.

2.2 Related Works

2.2.1 Cross-Language WordNet

In addition to the initiatives related to the Japanese, Chinese, and Indonesian WordNets, which are the focus of this study, we introduce research efforts on cross-lingual WordNets.

Fellbaum and Vossen (2012) analyzed the challenges of aligning WordNets across different languages from a linguistic perspective. Isahara et al. (2008) developed the Japanese WordNet by translating words within WordNet into Japanese. Wang and Bond (2013) aligned Chinese with WordNet. Additionally, Choi et al. (2004) developed a cross-linguistic WordNet that aligns Korean, Japanese, and Chinese. Putra et al. (2008) proposed the development of the Indonesian WordNet through manual annotation by human annotators. Additionally, Open WordNet Bahasa, which aligns with both Indonesian and Malay, was developed by Noor et al. (2011). Rudnicka et al. conducted a study aimed at mapping the Polish WordNet to the English Princeton WordNet. In addition, they proposed a system to detect gaps and mismatches between the Polish and English WordNets (Rudnicka et al., 2023).

These studies focused on the development of WordNets in various languages and their alignment with other languages.

2.2.2 Conceptual Difference Detection

Next, we discuss research on detecting conceptual differences between languages caused by cultural and linguistic variations.

Yoshino et al. (2015) proposed a method to detect conceptual differences between Japanese and Chinese using Wikipedia category information and descriptions. We detected conceptual differences between languages due to cultural variations, using image similarity and an optimized threshold for automatic detection (Pituxcoosuvarn et al.; Nishimura et al., 2020). Li et al. (2019) conducted research on measuring semantic similarity between texts in different languages. In this study, they calculated the similarity between Chinese and Lao texts using WordNet. Stoyanova et al. (2013) proposed a method for identifying relationships between English and Bulgarian concepts by using word similarity within WordNet.

These studies detect interpretation differences among speakers of different languages. This research aims to automatically identify conceptual differences among Japanese, Chinese, and Indonesian, using WordNet and OMW based on the English conceptual system.

3 Extraction Method of Conceptual Difference

3.1 Quantification of the Concept Range

To extract the conceptual differences between translation pairs, quantifying the range of concepts for each word is necessary. We propose a method for quantifying the conceptual range of words using WordNet and OMW[1]. Our method's feature is that by leveraging the structure of OMW, which links various languages to the conceptual system represented by English WordNet, it allows for the comparison of word concepts based on WordNet

[1]OMW version 1.4

synsets across different languages with distinct conceptual systems (Bond and Foster, 2013).

We target the languages Japanese (ja), Chinese (zh), and Indonesian (id) to extract word pairs with conceptual differences across these three languages. Asian languages, in comparison to European languages, tend to show lower similarity to English. Therefore, in WordNet, which is based on the English conceptual framework, these languages are mapped to a different conceptual system. By comparing Asian languages using WordNet as the reference point, we can detect a greater number of conceptual differences.

The first step of our proposal is to quantify the range of concepts for each word. The synsets in the WordNet and their superordinate-subordinate relationships are represented by a graph G, as in Equation 1.

$$G = (V, E) \tag{1}$$

A graph G has a set V of synsets and a set E whose edges are is-a relations between synsets.

$$V = \{v_1, v_2, v_3, \cdots, v_l\} \tag{2}$$

Also, a graph G is a directed graph using an is-a relation called hypernym, which identifies upper-level synsets from lower-level synsets.

If v_i be the upper synset and v_j the lower synset, the edges are represented as (v_j, v_i). Due to the structure of WordNet, each synset has a word representing its concept because synsets are defined by a collection of synonymous words, and there is a set of words W as in Equation 3.

$$W = \{w_1, w_2, w_3, \cdots, w_n\} \tag{3}$$

Some synset v_i as in Equation 4 is a set of words as in Equation 5.

$$v_i \in V \tag{4}$$

$$v_i \subset W, w_k \in v_i \tag{5}$$

We define the set of synsets C_{w_k} that is the range of the concept of a word w_k. First, C_{w_k} has v_i ($v_i \in C_{w_k}$) because $w_k \in v_i$. For any synset $v \in C_{w_k}$, the parent node v_p, child node v_c, and sibling node v_b are used to obtein the set of synsets associated with the word w_k (Equation 6). In this paper, this set C_{w_k} is the concept range of word w_k.

$$\begin{aligned} C_{w_k} = C_{w_k} \cup \\ \{v_p \mid v \in C_{w_k}, (v, v_p) \in E, w_k \in v_p\} \cup \\ \{v_c \mid v \in C_{w_k}, (v_c, v) \in E, w_i \in v_c\} \cup \\ \{v_b \mid v \in C_{w_k}, (v, v') \in E, (v_b, v') \in E, \\ w_k \in v_b\} \end{aligned} \tag{6}$$

3.2 Extraction of Conceptual Difference

The second step of our proposal is to extract word pairs with different conceptual ranges between words in one language and its translation in another. Specifically, we compare the concept ranges $C_{w_k^{l_1}}$ and $C_{w_k^{l_2}}$ for words $w_k^{l_1}$ in language l_1 and $w_k^{l_2}$ in different language l_2 contained on the same synset. If $C_{w_k^{l_1}} \neq C_{w_k^{l_2}}$ then $w_k^{l_1}$ and $w_k^{l_2}$ are as word pairs with conceptual difference. We get concept-differential word pairs in three languages: Japanese, Chinese and Indonesian.

There are seven patterns of conceptual differences obtained between the two languages depending on how the conceptual ranges are combined. First, the common set (ComSet) of concept ranges of words ($w_k^{l_1}$, $w_k^{l_2}$) in the two languages (l_1, l_2) is Equation 7 .

$$ComSet(w_k^{l_1}, w_k^{l_2}) = C_{w_k^{l_1}} \cap C_{w_k^{l_2}} \tag{7}$$

The symmetric difference set (DiffSet) of concept ranges of words ($w_k^{l_1}$, $w_k^{l_2}$) in the two languages (l_1, l_2) is Equation 8. For word pairs with conceptual difference $C_{w_i^{l1}} \neq C_{w_i^{l2}}$, so $DiffSet \neq \emptyset$.

$$\begin{aligned} DiffSet(w_k^{l_1}, w_k^{l_2}) = \\ (C_{w_k^{l_1}} - C_{w_k^{l_2}}) \cup (C_{w_k^{l_2}} - C_{w_k^{l_1}}) \end{aligned} \tag{8}$$

We classify as horizontal types the cases in which the synsets in the common set and the synsets in the symmetric difference set are siblings, as in Equation 9.

$$\begin{aligned} \forall v_x \forall v_y ((v_x \in ComSet(w_k^{l_1}, w_k^{l_2}) \\ \wedge v_y \in DiffSet(w_k^{l_1}, w_k^{l_2})) \Rightarrow \\ \exists v (v \in V \wedge (v_x, v) \in E \wedge (v_y, v) \in E))) \end{aligned} \tag{9}$$

In addition, we define a horizontal (part common) type as one that satisfies the 10 equation.

$$((C_{w_k^{l_1}} - C_{w_k^{l_2}}) \neq \emptyset) \wedge ((C_{w_k^{l_2}} - C_{w_k^{l_1}}) \neq \emptyset) \tag{10}$$

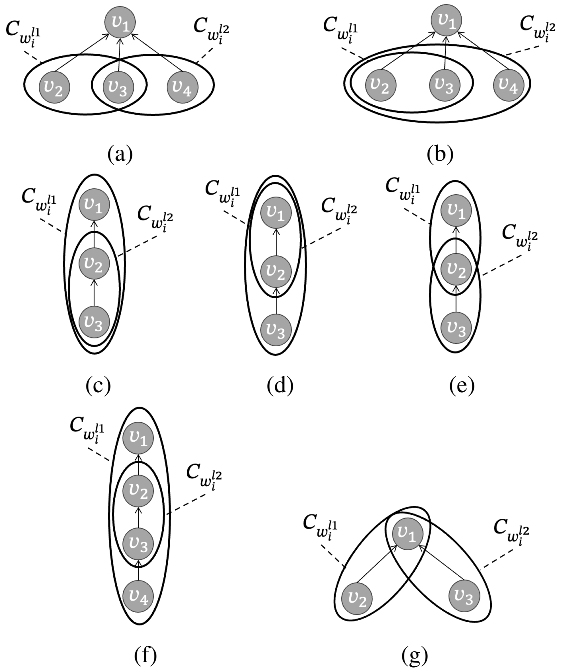

Figure 1: Type of conceptual difference

Figure 1a is an example of the horizontal-(part common) type. In Figure 1a, the nodes v_1, v_2, v_3 and v_4 are each synset and the edges between the nodes are Hypernym. The conceptual range $C_{w_k^{l_1}}$ of a word $w_k^{l_1}$ in language l_1 is the circle enclosing node v_2 and node v_3, the conceptual range $C_{w_k^{l_2}}$ of a word $w_k^{l_2}$ in language l_2 is the circle enclosing node v_3 and node v_4, which is the circle enclosing v_4. $C_{w_k^{l_1}} = \{v_2, v_3\}$ and $C_{w_k^{l_2}} = \{v_3, v_4\}$ as shown in Figure 1a, the common set is $ComSet(w_k^{l_1}, w_k^{l_2}) = \{v_3\}$, the symmetric difference set is $DiffSet(w_k^{l_1}, w_k^{l_2}) = \{v_2, v_4\}$. They satisfy the Equation 9 and Equation 10. The two concept ranges are partially common because $w_k^{l_1}$ and $w_k^{l_2}$ overlap the concept of node v_3 and have concepts (nodes v_2 and v_4) that can only be represented by their respective concept ranges.

On the other hand, among the horizontal types, those satisfying the Equation 11 are considered horizontal (inclusive). Figure 1b is an example of the horizontal (inclusive) type, the concept ranges of w1 and w2 are $C_{w_k^{l_1}} = \{v_2, v_3\}$ and $C_{w_k^{l_2}} = \{v_2, v_3, v_4\}$. Since $C_{w_k^{l_2}}$ encompasses $C_{w_k^{l_1}}$, $w_k^{l_2}$ expresses a broader meaning than $w_k^{l_1}$ and is an inclusion relation.

$$((C_{w_k^{l_1}} - C_{w_k^{l_2}}) = \emptyset) \vee ((C_{w_k^{l_2}} - C_{w_k^{l_1}}) = \emptyset) \tag{11}$$

As in the Equation 12, we classify the conceptual difference as vertical type if there exists at least one parent-child relationship between the synset in the common set and the synset in the symmetric difference set.

$$\exists v_x \exists v_y ((v_x \in ComSet(w_k^{l_1}, w_k^{l_2}) \wedge v_y \in DiffSet(w_k^{l_1}, w_k^{l_2})) \Rightarrow ((v_x, v_y) \in E \vee (v_y, v_x) \in E) \tag{12}$$

Among the vertical types, the vertical (lower inclusive) type is the one that satisfies the Equation 13. As in the example in Figure 1c, the vertical (lower inclusive) type is the case where the synset in the symmetric difference set is the upper synset of the synset in the common set.

$$\forall v_x \in ComSet(w_k^{l_1}, w_k^{l_2}), \forall v_y \forall v_z (((v_x, v_y) \in E \wedge (v_z, v_x) \in E) \Rightarrow (v_y \in (ComSet(w_k^{l_1}, w_k^{l_2}) \cup DiffSet(w_k^{l_1}, w_k^{l_2})) \wedge v_z \notin DiffSet(w_k^{l_1}, w_k^{l_2})) \tag{13}$$

There are vertical types that satisfy the Equation 14. The synset in these symmetric difference sets is a subsynset of the synset in the common set. As shown in Figure 1d, the one satisfying the Equation 10 and the Equation 14 is the vertical (upper common) type. The others are the vertical (upper inclusive) type (as shown in Figure 1g).

$$\forall v_x \in ComSet(w_k^{l_1}, w_k^{l_2}), \forall v_y \forall v_z (((v_x, v_y) \in E \wedge (v_z, v_x) \in E) \Rightarrow (v_y \notin DiffSet(w_k^{l_1}, w_k^{l_2}) \wedge v_z \in (ComSet(w_k^{l_1}, w_k^{l_2}) \cup DiffSet(w_k^{l_1}, w_k^{l_2})) \tag{14}$$

As in the Equation 15, there is at least one parental relationship and at least one child relationship between the synset of the symmetric difference set and the synset of the common set here. Among them, those satisfying the Equation 10 are the vertical (part common) type, and those satisfying the Equation 11 are the vertical (inclusive) type. An example of the vertical (part common) type is shown in Figure 1e and an example of the vertical (inclusive) type is shown in Figure 1f.

In these inclusion relation, the conceptual range of one word is broader than that of the other, so there can be a discrepancy in communication in one direction, as in dumpling and dumpling, but

Table 2: Number of word pairs extracted

Lang	Word Pair	Concept Difference
ja - zh	236,590	27,005
ja - id	398,048	60,581
zh - id	111,450	14,175
Word Triplets	**11,571**	**1,375**

※ja:Japanese, zh:Chinese, id:Indonesian

not in the other. On the other hand, partial or superordinate commonality may lead to discrepancies in conversation in either orientation.

$$\exists v_w \, \exists v_x \, \exists v_y \, \exists v_z ((v_w, v_x \in ComSet(w_k^{l_1}, w_k^{l_2}) \wedge v_y, v_z \in DiffSet(w_k^{l_1}, w_k^{l_2})) \Rightarrow ((v_w, v_y) \in E \wedge (v_z, v_x) \in E)) \quad (15)$$

4 Human Judgements of Conceptual Differences

The creation of the correct labels is the result of manually determining the conceptual differences between words using a questionnaire.

4.1 Creating Word Triplets

In this section, we describe a method for manually creating the correct labels in order to evaluate the accuracy of the proposed method. Our proposal could find word pairs with conceptual differences between the Japanese, Chinese and Indonesian. As a result, 27,005 word pairs were obtained out of 236,590 total word pairs in ja-zh, 60,581 out of 398,048 word pairs in ja-id and 14,175 out of 111,450 word pairs in zh- id (Table 2) Word triplets are Japanese-Chinese-Indonesian word combinations that are filtered in two steps for word pairs, and they are created from the remaining word pairs (Table 2). The first filtering is to remove abstract concepts that are difficult to determine conceptual differences. The top-level concepts of synsets on WordNet: physical entity, abstraction and thing, which are located one level below entity, and word pairs extracted from the lower-level concepts of abstraction are excluded. The second filtering is to check the bilingual relationship between word pairs. In order to target word pairs that are most likely to appear as translation results in MT, word pairs where neither translation result was the other when each word was translated by MT are deleted. In the word pairs $w_k^{l_1}$ and $w_k^{l_2}$, word pairs where the Google MT of $w_k^{l_1}$ does not result in $w_k^{l_2}$ and the translation of $w_k^{l_2}$ does not result in $w_k^{l_1}$ are removed.

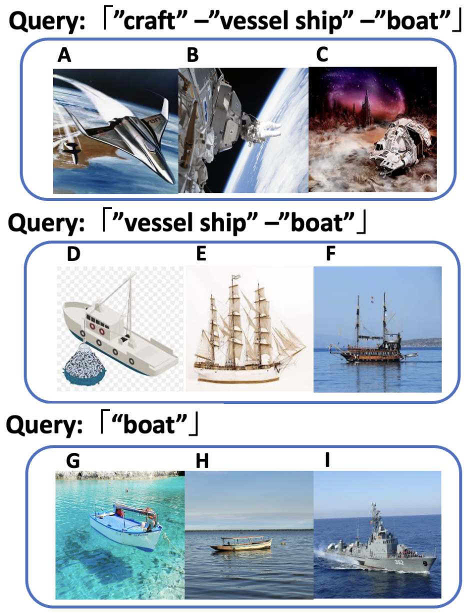

Figure 2: Images and queries in questionnaires

The next step is to create word triplets from the filtered Japanese-Chinese, Japanese-Indonesian and Chinese-Indonesian word pairs. Word pairs with words in common between any two languages are retrieved. If the three words in the two word pairs have a common part of the concept range, the three words are a word triplet. For example, there is a word pair w_k^{ja} and w_k^{zh} in ja-zh, and a word pair w_k^{ja} and w_k^{id} in ja-id. If the concept range $C_{w_k^{ja}} \wedge C_{w_k^{zh}} \wedge C_{w_k^{id}} \neq \emptyset$, then $(w_k^{ja}, w_k^{zh}, w_k^{id})$ is a word triplet. There are 11,571 word triplets created from word pairs between all languages (Table 2). Of these, 1,375 are from word pairs with conceptual differences.

4.2 Design of the Questionnaire and Collection of Images

In one task of the questionnaire, respondents were asked to select images that matched a word in their mother tongue from a group of images corresponding to the word triplet. The subjects were Japanese, Chinese and Indonesian, and we checked whether there were differences in responses between languages.

The image groups used in the questionnaire are the images obtained by image retrieval using the synsets in each concept range, if the word triplet of words are ja:w_k^{ja}, zh:w_k^{zh} and id:w_k^{id}. We make the query for Image Search from an English word

Table 3: Sample word triplet and image selection results for Japanese (ボート), Chinese (船), and Indonesian (kapal)

	Answer				
	1	2	3	4	5
Japanese	A,B,G	C,G	Non	D,F	G,H,I
Chinese	G,H,I	H,I	I	G,H,I	G,H

Table 4: The tallyed table generated from the query 「"boat"」 in Table 3

	matching image found	no matching image found
Japanese	3	2
Chinese	5	0

that is randomly selected from a synset in $C_{w_k^{ja}} \cup C_{w_k^{zh}} \cup C_{w_k^{id}}$. It is necessary to differ between higher and lower level synsets. the query was created so that it was an AND search with the higher synset words minus the lower concept words. The query is an AND search based on the relationship between synsets in $C_{w_k^{ja}} \cup C_{w_k^{zh}} \cup C_{w_k^{id}}$, where the higher concept words are subtracted from the lower concept words. For example, there is a word triplet (ボート (bo-to), 船 (chuan), kapal). If union of the concept range of these words has $\{v_1, v_2, v_3\}$, we get one word at random from the English words in $v_1, v_2 and v_3$: (v_1:craft, v_2:vessel ship, v_3:boat). If the parent relation $((v_2, v_1) \in E) \wedge ((v_3, v_2) \in E)$ holds in the order v_1, v_2, v_3, then query for v_1 is 「"craft"-"vessel ship" -"boat"」, the query for v_2 is 「"vessel ship" - "boat"」 and the query for v_3 is 「"boat"」. The "-" in the query means NOT and words with "-" are deleted. We use the top three images (Figure 2) that are got from Image Search with these query in the questionnaire.

4.3 Judgements of the Conceptual Difference

Judgement of the conceptual difference is to totaling up of the questionnaire answer by language and is to check if there is a difference in the answer between languages. If there is a difference in the aggregated answers, it is judged as having a conceptual difference; if there is no difference in the answers, it is judged as having no conceptual difference. This process involves evaluating each tally table generated from the query by applying Fisher's exact test to identify conceptual differences. Tasks with significant differences in any of the tallied tables are a conceptual difference. As an example, to judge the conceptual differences between (ボート (bo-to), 船 (chuan), kapal) in Japanese and Chinese, responses from Japanese and Chinese speakers will be collected based on the following categories: 「"craft" -"vessel ship" -"boat"」, 「"vessel ship" -"boat"」, and 「"boat"」. Fisher's exact test will be performed on each of the three generated contingency tables to determine if there is a statistically significant difference between the responses of Japanese and Chinese speakers. The contingency tables will be created based on whether the respondents judged that one or more images matched the keyword. For each language, the tables will count the number of people who determined 'matching image found' and 'no matching image found.' In the case of the contingency table for 「"boat"」 from the (ボート (bo-to), 船 (chuan), kapal) set, selecting one or more images from Image G, Image H, or Image I (Figure 2) will be classified as 'matching image found', while selecting none will be classified as 'no matching image found.' For example, in the (ボート (bo-to), 船 (chuan), kapal) task, if the responses follow Table 3, the contingency table for 「"boat"」 appears as in Table 4. I will create similar contingency tables for other queries, and if any show statistically significant differences, I will conclude that this task reflects a conceptual difference. We used the 1,375 word triplets identified by the proposed method as having conceptual differences (Table 2) as analysis data. Fifteen participants (5 Japanese, 5 Chinese, and 5 Indonesian) manually judged the conceptual differences. The results showed 528 triplets with differences between Japanese and Chinese, 476 between Japanese and Indonesian, and 471 between Chinese and Indonesian.

Table 5: Assessment results

	ja-zh	ja-id	zh-id
Accuracy	0.81	0.70	0.69
Precision	0.43	0.44	0.32
Recall	0.87	0.80	0.71
F-Score	0.58	0.57	0.44

※ja:Japanese, zh:Chinese, id:Indonesian

Table 6: Classification of conceptual differences and breakdown of data

Type of Conceptual Difference	ja-zh		ja-id		zh-id	
	Pair	Ratio of the Conceptual Difference	Pair	Ratio of the Conceptual Difference	Pair	Ratio of the Conceptual Difference
horizontal (part common)	43	**(+) 72.0%**	105	**(+) 49.5%**	81	**(+) 56.8%**
horizontal (inclusive)	637	35.5%	758	34.4%	758	33.2%
vertical (lower inclusive)	280	39.6%	248	40.8%	278	35.6%
vertical (upper inclusive)	634	34.7%	575	32.7%	621	35.3%
vertical (part common)	26	26.9%	62	38.7%	59	44.1%
vertical (inclusive)	27	**(+) 63.0%**	52	25.0%	89	**(-) 11.2%**
vertical (upper common)	62	**(+) 51.6%**	111	**(+) 45.0%**	109	29.4%
Equivalence	24	33.3%	13	23.1%	17	29.4%

5 Evaluation and Analysis

5.1 Evaluation

We evaluated the proposed method using four metrics: accuracy, precision, recall, and F-score. As the evaluation dataset, we use 100 randomly selected triplets from the total of 11,571 word triplets in Table 2. These are assigned as correct labels based on the results of human judgments of conceptual differences. Table 5 presents the results. In the evaluation dataset between Japanese and Chinese (85 cases without conceptual difference, 15 with), the proposed method classified 30 out of 100 data points as having a conceptual difference. For Japanese and Indonesian (75 cases without, 25 with), 45 out of 100 were determined to have a conceptual difference. In the Chinese and Indonesian comparison (83 cases without, 17 with), 38 out of 100 were classified as having a conceptual difference. The proposed method demonstrates a consistently high recall across all languages, with a precision of approximately 43%. This indicates that the method tends to over-detect conceptual differences. Consequently, the F-score hovers around 50%. However, since conceptual differences can serve as a root cause of misunderstanding, over-detection is arguably less problematic than failing to detect such differences. Therefore, the relatively low precision and F-score of the proposed method are not considered critical shortcomings.

5.2 Discussion

Based on the classification of conceptual differences, we categorize the analysis data and analyze the detection accuracy of conceptual differences for each type.

Table 6 shows the breakdown of data by classification in the analysis dataset. The "Equivalence" classification refers to data where no conceptual difference was found between the languages, though differences were observed between other language pairs. Conversely, the remaining seven types are cases where the proposed method detected conceptual differences. We examined the independence of classification and detection accuracy for each language using the Chi-squared test. To perform the test, a contingency table was created based on the breakdown of data by the seven classifications, and the test was conducted. The results indicated that for all language pairs, the p-value was less than the significance level of 0.01, suggesting that there was a significant difference in detection accuracy based on the classification of conceptual differences. Further residual analysis was conducted to identify which classifications showed significant differences. (+) or (-) marks indicate where the adjusted residuals exceeded an absolute value of 1.96. The results of the residual analysis suggest that for all language pairs, the detection accuracy was high for the horizontal (part common) type of conceptual differences. On the other hand, for the zh-id pair, the detection accuracy was low for the vertical (inclusion) type. An example of the horizontal (part common) type is the triplet (容器, 容器, bekas). The Japanese term "容器" refers to small containers used in daily life, while the Chinese term "容器" encompasses containers in general. As a result, in the survey, Japanese respondents chose images of items like bowls and garbage bins, whereas Chinese respondents selected images of not only bottles but also large containers used on container ships. This indicates that, in the case of the horizontal (part common) type, there are differences in conceptual range at the same level, making it easier for human raters to perceive conceptual differences.

6 Conclusion

Translation pairs can have differing concepts expressed by words across languages, potentially leading to misunderstandings in multilingual communication. This study used WordNet and OMW to quantify concepts and extract conceptual differences by comparing ranges across language pairs. From the three languages of Japanese, Chinese, and Indonesian, we formed word pairs based on the same synset, detecting conceptual differences in 27,005 pairs (Japanese-Chinese), 60,581 pairs (Japanese-Indonesian), and 14,175 pairs (Chinese-Indonesian) out of the total 236,590, 398,048, and 111,450 pairs, respectively. We also classified the differences in conceptual ranges topologically and found significant differences in detection accuracy based on classification type among the three languages.

Acknowledgments

This research was partially supported by a Grant-in-Aid for Scientific Research (B) (23K21730,2024) from the Japan Society for the Promotion of Sciences (JSPS).

References

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362.

Barry R Chiswick and Paul W Miller. 2005. Linguistic distance: A quantitative measure of the distance between english and other languages. *Journal of Multilingual and Multicultural Development*, 26(1):1–11.

Key Sun Choi, Hee Sook Bae, Wonseok Kang, Juho Lee, Eunhe Kim, Hekyeong Kim, Donghee Kim, Youngbin Song, and Hyosik Shin. 2004. Korean-chinese-japanese multilingual wordnet with shared semantic hierarchy. In *4th International Conference on Language Resources and Evaluation, LREC 2004*, pages 1131–1134.

Christiane Fellbaum and Piek Vossen. 2012. Challenges for a multilingual wordnet. *Language Resources and Evaluation*, 46:313–326.

Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. 2008. Development of the japanese wordnet. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*.

Sizhuo Li, Lanjiang Zhou, Jianan Zhang, Feng Zhou, Jianyi Guo, and Wenjie Huo. 2019. Chinese-lao cross-language test similarity computing based on wordnet. In *Proceedings of International Conference on Mechatronics and Intelligent Robotics (ICMIR2018)*, pages 459–464.

Ikkyu Nishimura, Yohei Murakami, and Mondheera Pituxcoosuvarn. 2020. Image-based detection criteria for cultural differences in translation. In *International Conference on Collaboration Technologies and Social Computing*, pages 81–95.

Nurril Hirfana Bte Mohamed Noor, Suerya Sapuan, and Francis Bond. 2011. Creating the open wordnet bahasa. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, pages 255–264.

Mondheera Pituxcoosuvarn, Donghui Lin, and Toru Ishida. A method for automated detection of cultural difference based on image similarity.

Desmond Darma Putra, Abdul Arfan, and Ruli Manurung. 2008. Building an indonesian wordnet. In *Proceedings of the 2nd International MALINDO Workshop*, pages 12–13.

Ewa Rudnicka, Łukasz Grabowski, Maciej Piasecki, and Tomasz Naskręt. 2023. In search of gaps between languages and wordnets: the case of polish-english wordnet. *International Journal of Lexicography*, 36(1):68–92.

Ewa Rudnicka, Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. A strategy of mapping polish wordnet onto princeton wordnet.

Ivelina Stoyanova, Svetla Koeva, and Svetlozara Leseva. 2013. Wordnet-based cross-language identification of semantic relations. In *Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing*, pages 119–128.

Shan Wang and Francis Bond. 2013. Building the chinese open wordnet (cow): Starting from core synsets. In *Proceedings of the 11th Workshop on Asian Language Resources*, pages 10–18.

Naomi Yamashita and Toru Ishida. 2006. Automatic prediction of misconceptions in multilingual computer-mediated communication. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, pages 62–69.

Takashi Yoshino, Mai Miyabe, and Tomohiro Suwa. 2015. A proposed cultural difference detection method using data from japanese and chinese wikipedia. In *Proceeding of 2015 International Conference on Culture and Computing (Culture Computing)*, pages 159–166.