

## *A Book Review Analysis on Haruki Murakami in Four Countries*

Yukari Shiota, Gakushuin University, Faculty of Economics, Japan  
Yuriko Yano, Gakushuin University, Japan  
Takako Hashimoto, Chiba University of Commerce, Japan

The Asian Conference on Cultural Studies 2017  
Official Conference Proceedings

### **Abstract**

Haruki Murakami novels have obtained a worldwide reputation. We analyze how his novels are read in foreign countries via their book reviews, using the Amazon book reviews on his novels in Japan, US, UK, and Canada. If possibly we would like to identify the cultural differences among the countries. The reasons why we selected Haruki Murakami are first that the number of reviews is so large and secondly that Mr Murakami has a good command of English so the translation from Japanese has done quite well owing to his sense of English. We set the target novels *Norwegian Wood* and *Kafka On The Shore* because types and themes of the two novels are different.

The core technology of this analysis is text mining. The morphological analysis tool for Japanese named “MeCab” and one for English named “TreeTagger” are used. In addition, by using “word2vec”, vector representation models of reviews have been made. The word2vec also offers the high measurement quality of representations in a word similarity level. We would like to analyze the change of the word representations in the vector space among the four countries’ reviews. From the results, we spotted a notable difference in reviews of *Norwegian Wood* between Japanese and the Western countries, although we found no distinction in reviews of *Kafka On The Shore*.

Keywords: text mining, morphological analysis, word2vec, book reviews, Murakami Haruki, Norwegian Wood, Kafka on the Shore.

**iafor**

The International Academic Forum  
[www.iafor.org](http://www.iafor.org)

## Introduction

Haruki Murakami novels have obtained a worldwide reputation. We analyze how his novels are read in foreign countries via their book reviews, using the Amazon book reviews on his novels in Japan, US, UK, and Canada. If possibly we would like to identify the cultural differences among the countries. The reasons why we selected Haruki Murakami are first that the number of reviews is so large and secondly that Murakami has a good command of English so the translation from Japanese has done quite well owing to his sense of English. The translations are also wonderful. We think that the excellent translation keeps the essence of the novels as the original Japanese ones. For example, Jay Rubin, an honorary professor of Harvard University on Japanese literature is one of the notable translators and there are many ingenuity in his translation(Rubin 2005). Among the Murakami novels, we select the target novels *Norwegian Wood* (Murakami 2000) and *Kafka On The Shore* (Murakami 2005), because *Norwegian Wood* was sold explosively around the world and because types and themes of the two novels are different.

The core technology of this analysis method is text mining. The morphological analysis tool for Japanese named “MeCab” and one for English named “TreeTagger” are used. In addition, by using “word2vec”, vector representation models of reviews have been made. The word2vec outputs the principal component vectors by which the vector space coordinates are represented. The number of the principal component vectors is about 100. The word2vec also offers the high measurement quality of representations in a word similarity level. We shall analyze the change of the word representations in the vector space among the four countries’ reviews.

In the next section, we shall survey the worldwide reputation on Murakami novels and the features. Then the analysis methods are explained. In Section 4, the analysis results are shown. Finally we conclude the paper with discussions.

## Survey of Murakami Novels Features

Murakami’s novels are translated in many languages and they have caused a boom that we did not expect around the world. In the section, we survey how Murakami works are read in the world and its features. As a generally approved descriptions on Murakami novels, we selected *The CAMBRIDGE COMPANION TO Modern Japanese Culture*(Sugimoto 2009), because we think that would offer impartial evaluations. The Murakami novels are described as follows:

- ❑ non-Japanese: Storylines can take place anywhere in the world and place no emphasis on uniquely Japanese attributes (by Yoshio Sugimoto in “An Overview”)
- ❑ Characters enjoy encountering others in cozy settings.(by Takeshi Inoguchi in Chapter 9)
- ❑ The leitmotif of his novels preceded the neuroscientific discovery of the “monkey-see, monkey-do” theory in which the monkey sees a human subject’s movement and its memory is activated causing its muscles to move. (by Takeshi Inoguchi in Chapter 9)

- ❑ His characters are all individualist. They are interested in the inner itself, but they are also open and associated with others. (by Takeshi Inoguchi in Chapter 9)
- ❑ He has created new models free from modernist values. He suggests that they are ways to live through this age, whether in Japan or elsewhere, as long as people gracefully abandon norms that governed them in the past. (by Toshiko Ellis, in Chapter 11)

Concerning Murakami novels wide popularity, Shozo Fujii analyzes as follows: *The young middle class in the Chinese-speaking world is reading Murakami as a manual of urban consumption culture such as a bar, a single apartment house, a travel, jazz, etc., which was rapidly spread along with economic growth* (Fujii 2009). Fujii says that the Murakami boom had moved as the order Taiwan→Hong-Kong→Shanghai→Beijing and the boom occurred at a time when their high-rate economic growth nearly halves (Fujii 2007). The Murakami boom may be a phenomenon related to cultural maturity.

Christian Rudder analyzes SNS messages on the online dating web system and concerning *Norwegian Wood*, he asks “Why do Asian people like *Norwegian Wood* ?” because the word *Norwegian Wood* appears as most typical words for Asian men/women’s messages; in other groups that are (1) white men/women, (2) black men/women, and (3) Latinos/Latinas, there is no the same phenomena (Rudder 2014a, 2014b). We would like to find a difference concerning *Norwegian Wood* between Japan and Western countries in our analysis.

We have selected two Murakami bestsellers as our target; they are *Norwegian Wood* and *Kafka On The Shore*. For those who did not read the novels, we cite the editorial reviews from Amazon.com in the following:

- In 1987, when *Norwegian Wood* was first published in Japan, it promptly sold more than 4 million copies and transformed Haruki Murakami into a pop-culture icon.
- Joining the rich literature of runaways, *Kafka On The Shore* follows the solitary, self-disciplined schoolboy Kafka Tamura as he hops a bus from Tokyo to the randomly chosen town of Takamatsu, reminding himself at each step that he has to be “the world's toughest fifteen-year-old.”... Meanwhile, in a second, wilder narrative spiral, an elderly Tokyo man named Nakata veers from his calm routine by murdering a stranger. An unforgettable character, beautifully delineated by Murakami, Nakata can speak with cats but cannot read or write, nor explain the forces drawing him toward Takamatsu and the other characters.

The novel *Norwegian Wood* has especially boomed in Asian countries such as China and Thailand. As shown in the above Rudder’s comment, even in USA, the distinct concern concerning *Norwegian Wood* can be found between Asian and other groups. Therefore we selected *Norwegian Wood* as our target. In addition, as a non-love story, we selected *Kafka On The Shore* to conduct the comparison.

## Our Analysis Method and Data

In this section, we will describe our research method and data we used. The data we used in the analysis are reviews of Amazon; the USA data are from Amazon.com, the UK data are from Amazon.co.uk, the Canada data are from Amazon.ca, the Australia data are from Amazon.com.au, and the Japan data are from Amazon.co.jp. The number of collected data is shown in Table 1. The text size of the Japanese texts is 1,714,599 Japanese characters including Chinese character Kanji, with the retrieval keyword “Haruki Murakami”. With the keyword “Norwegian Wood”, the number of characters is 1,202,571. From the figures, we found that *Norwegian Wood* is much more popular than others.

Concerning the UK, USA, Canada, and Australia, the given retrieval keyword is “Haruki Murakami” only. The collected data size is 1,301,842 words in the UK reviews. Although we expected the number of reviews in USA was bigger than others, contrary to our expectations, the number of UK reviews is the biggest. The second one is the USA and the third is the Canada data. In the Australia Amazon, we could not collect enough reviews.

Table 1: Collected data from 5 country review sites.

Country	Language	Retrieval Keyword	Text Size (Japanese characters, English words)
Japan	Japanese	Haruki Murakami	1714599
Japan	Japanese	Norwegian Wood	1202571
UK	English	Haruki Murakami	1301842
USA	English	Haruki Murakami	972965
Canada	English	Haruki Murakami	324793
Australia	English	Haruki Murakami	6611

We used TreeTagger<sup>1</sup> as the morphological analysis tool for segmentation of the English reviews. As a word unit, we use a unigram that is one word. In the same way, we conducted text mining in Japanese reviews. As the messages are written in Japanese, we used MeCab<sup>2</sup> (Yet Another Part-of-Speech and Morphological Analyzer) as the morphological analyzer. As a word unit, we use a bigram such as “恋愛”+”小説” which means “love” + ”story” in English.

We used the word2vec<sup>3</sup> for the analysis (Mikolov et al. 2013). Word2vec, based on the two-layer neural network, makes the word vectors which correspond to principal components so that they can represent words in the vector space. In advance, we make a vector representation model from the input text file. The output is the vector representation model (vector space) in which each word is represented as a vector. Let us show you an example in Figure 1. There, the first to sixth vector element values are shown.

<sup>1</sup> <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

<sup>2</sup> <http://taku910.github.io/mecab/>

<sup>3</sup> <https://code.google.com/archive/p/word2vec/>

```

> word2vec_model[["Japanese"]]
A VectorSpaceModel object of 1 words and 100 vectors
      V1      V2      V3      V4      V5      V6
0.009571 -0.221544  0.340363  0.147666 -0.384261 -0.048622

```

Figure 1: A vector representation example of the word “Japanese”.

Using the vectors, we can calculate the similarity level between two words. Then, given a word, we can extract the similar word list to the word. Then we can get the subset of the vector space of the extracted word list. For the subset vector space, we shall conduct visualization using t-SNE (t-Distributed Stochastic Neighbor Embedding) as shown in Figure 2. The t-SNE conducts dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets (Maaten 2014). In the paper, the number of dimensions are reduced from 100 to two. The two axis are the two principal component directions that t-SNE extracted. In the following of the paper, we shall use the same framing of the t-SNE without axis names.

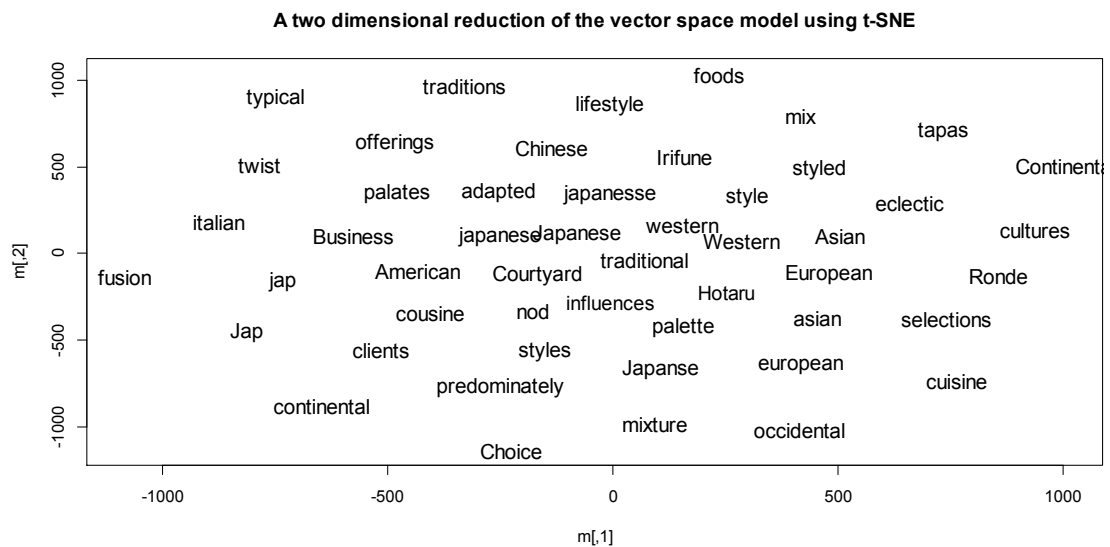


Figure 2: A two dimensional reduction of the vector space model using t-SNE.

### Result Analysis

In the section, we present the results from word2vec and t-SNE and do the comparison.

First let us show the UK results. The vector model here is generated from the reviews with retrieval work “Haruki Murakami”. Namely every Murakami related reviews are input. Then given a work such as “Murakami” against the vector model, a similar word list can be obtained and we shall visualize the word list using t-SNE.

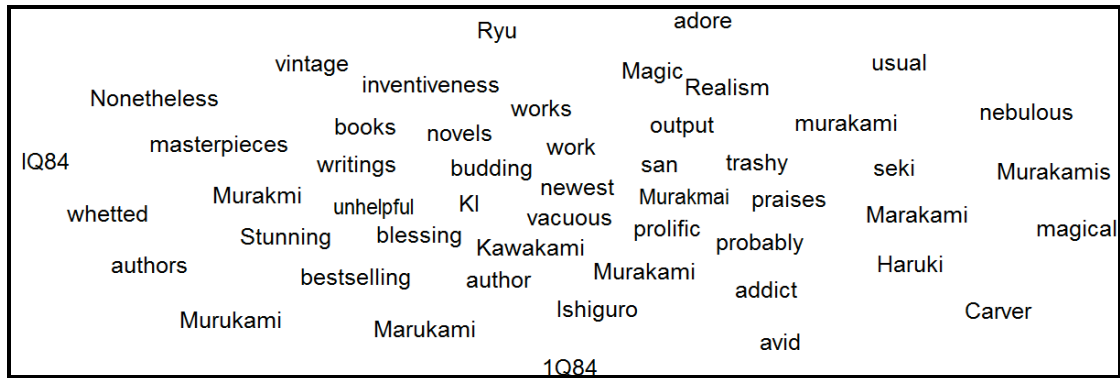


Figure 3: The vector model similar to “Murakami” in the UK data.

Figure 3 shows the entire reputation in UK. The positive word are *whetted*, *stunning*, *blessing*, *inventiveness*, *budding*, *adore*, *avid*, *praises*, *newest*, *magical*, *bestselling*, *masterpiece*, and *vintage*. As negative words, we can see *trashy*, *vacuous*, and *usual*. As another popular writer names, we found “Ryo (Murakami)” and “(Kazuo) Ishiguro”.

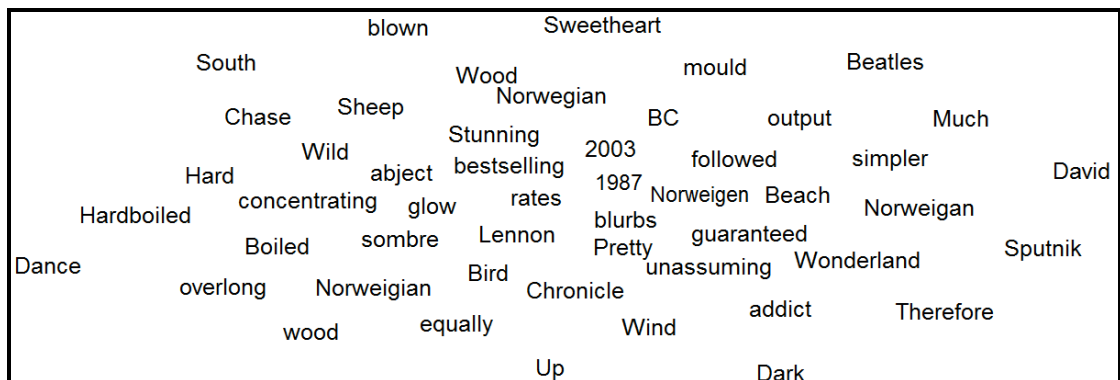


Figure 4: The vector model similar to “Norwegian (Wood)” in the UK data.

Next let us show the similar words of “Norwegian” in Figure 4. The extracted words are considered to be tightly connected to *Norwegian Wood*. The positive words are *bestselling*, *stunning*, *guaranteed*, *concentration*, and *addict*. On the other hand, we found as negative words, *abject* (failure) and *mould*. Words in parentheses are complementary by the authors. Contrary to our expectations, there is no word “love story” in the UK result. When *Norwegian Wood* was sold explosively in Japan at the beginning, it seems that young people read and sympathize more than aged persons. We think as follows; When *Norwegian Wood* was translated into English as a popular novel and read by UK readers, it might have been read from the beginning with the stance of a best-selling novel. As a result, it seems that the reader became more aged, and the spirit of deeply reading the philosophical theme of the novel, like life and death, was strong.

The similar words to “Kafka” are shown in Figure 5. The word “converges” in Murakami works has an impressive meaning which are found in the reviews, for example, as follows:

- the two stories *converge* with the truth coming out in the last few chapters.

- causing their lives to *converge* with devastating effect
- there are parallel narratives which *converge*

In *Kafka On The Shore*, different two stories converge. The words “converge”, “Oedipal (complex)”, and other character names appear there, from which we can find that reviewers discussed the theme and story of the novel. In addition, we extract similar words to “Nakata, Kafka” (See Figure 6). There, Mr. Nakata is also lovable in UK as in Japan. Concerning *Kafka On The Shore*, we can hardly find the notable difference between UK and Japan.

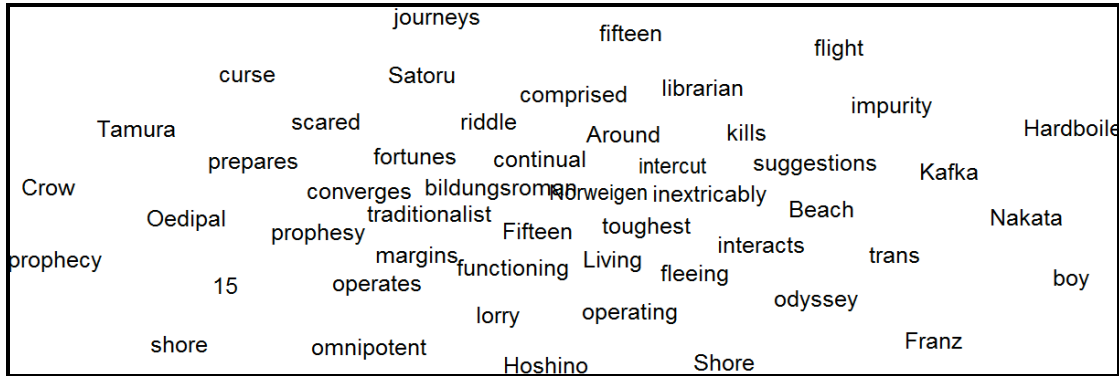


Figure 5: The vector model similar to “Kafka (On The Shore)” in the UK data.

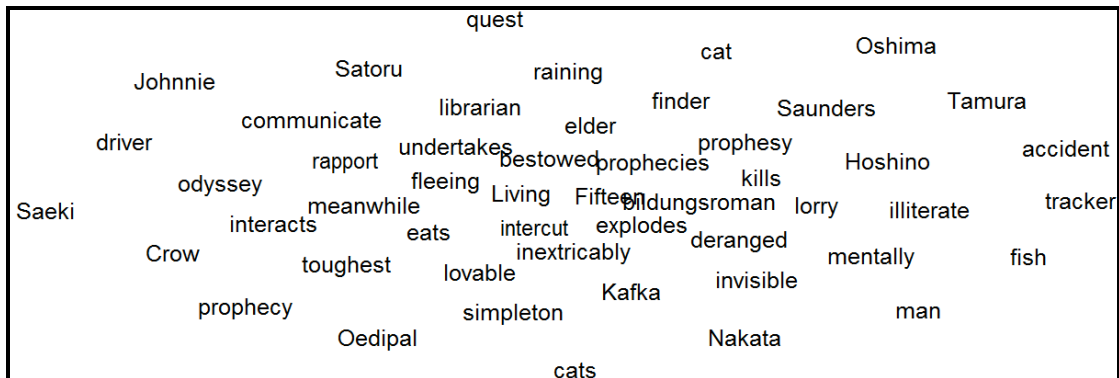


Figure 6: The vector model similar to “Nakata, Kafka” in the UK data.

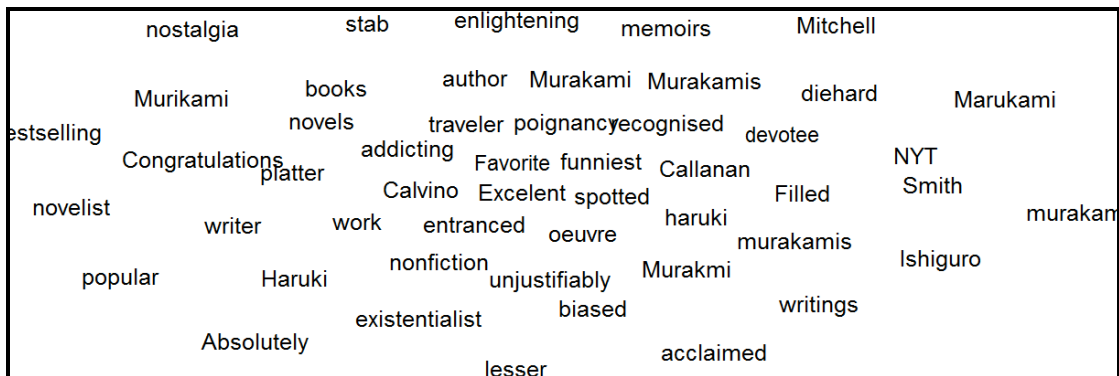


Figure 7: The vector model similar to “(Haruki) Murakami” in the USA data.

Next, let us describe the USA reviews. Concerning the word “Murakami”, the similar words are shown in Figure 7. In the United States, *Colorless Tsukuru Tazaki and His Years of Pilgrimage* was admired in the Sunday Book Review of The New York Times (Smith August 5th, 2014) and *Kafka On The Shore* was also reviewed in The New York Times (Miller February 6th, 2005), which made Murakami popularity still higher. The word “NYT” in Figure 7 is an abbreviation of New York Times. In Figure 7, positive words are *nostalgia, stab, enlightening, entranced, stab, acclaimed, excellent, favorite, funniest, recognized, addicting, devotee, and poignancy*. The negative word *biased* is found; this is for the newspaper book review. The word list similar to “Norwegian” is shown in Figure 8. The positive words are *polished, vein, generous, pyrotechnics, stab, irresistible, and moody*. The negative words are *pales* and *tamer*. There is no word like a love story.

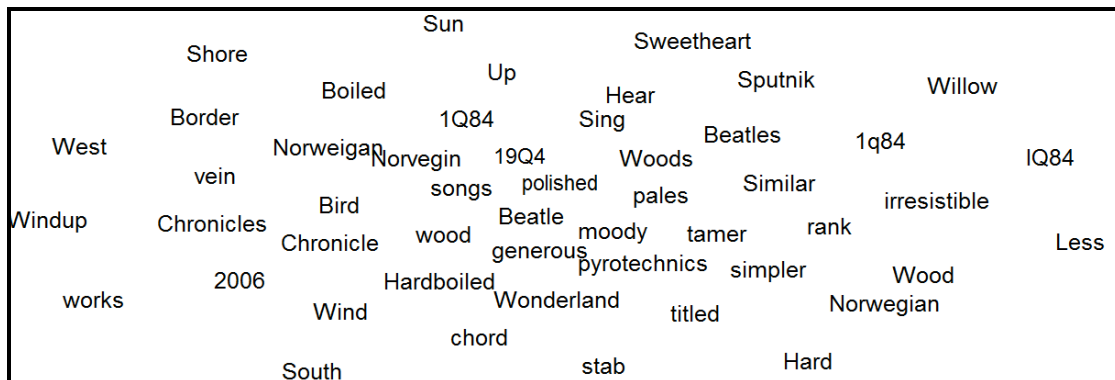


Figure 8: The vector model similar to “Norwegian (Wood)” in the USA data.

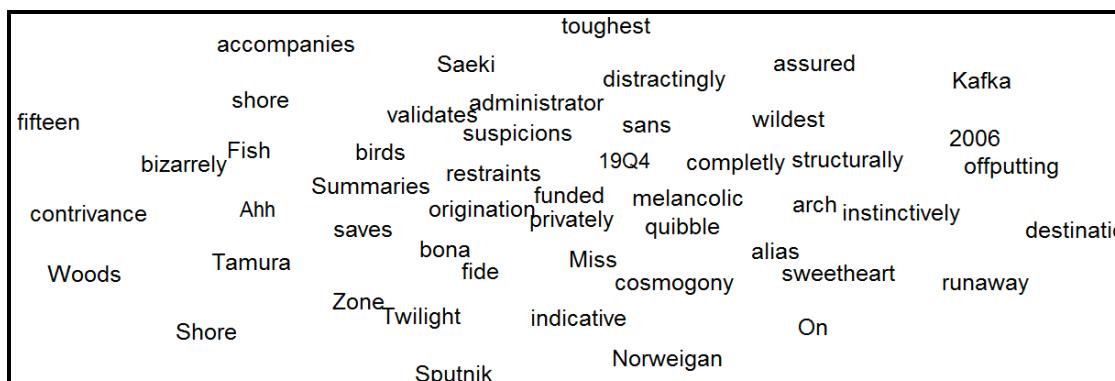


Figure 9: The vector model similar to “Kafka (On The Shore)” in the USA data.

Figure shows the word list similar to “Kafka”. There as other countries reviews, we found the term “Oedipal” (complex).

Next let us see the Canadian results. We found a subtle distinction concerning *Norwegian Wood* (See Figure 10). There are many praise for *Norwegian Wood*, but some reviews say that it is *convoluted, complicated, skewed, hesitatingly*. This is an impression that it was not in US, UK, and Japan. On the other hand, concerning *Kafka On The Shore*, there is no distinction; the word “Oedipal” was found in Canada’s result like others. In the Canada result, instead of “converged”, we found “combined”.



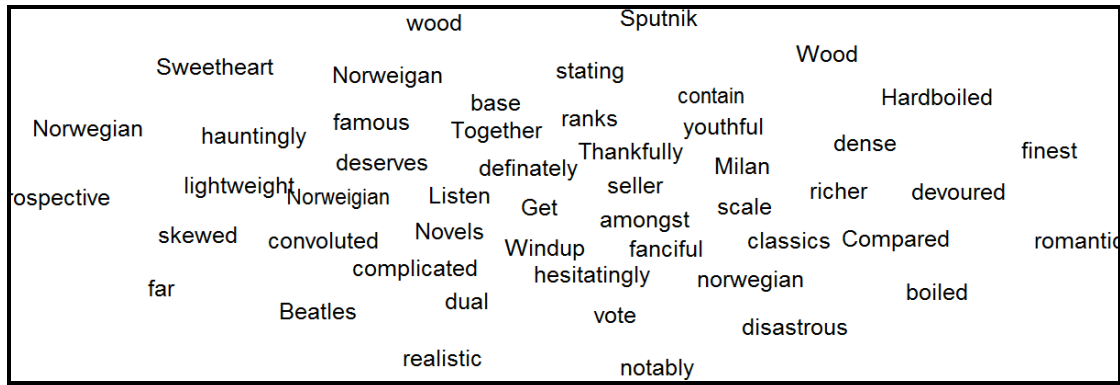


Figure 10: The vector model similar to “Norwegian (Wood)” in the Canada data.

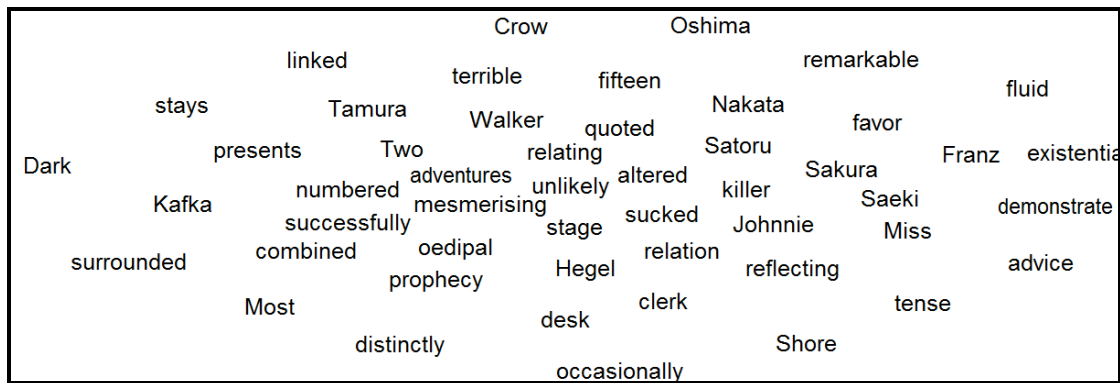


Figure 11: The vector model similar to “Kafka (On The Shore)” in the Canada data.

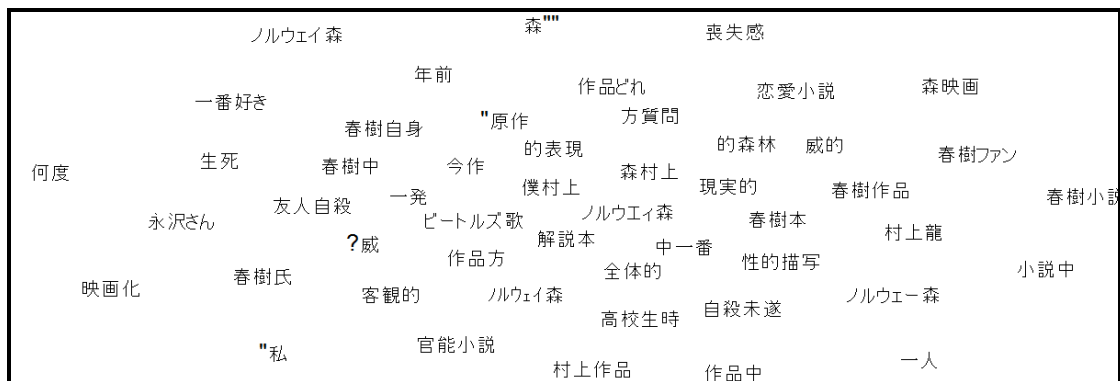


Figure 12: The vector model similar to “ノルウェイ+森” (Norwegian Wood) in the Japanese data.

Finally, we shall analyze the Japanese reviews. Figure 12 shows the word list similar to *Norwegian Wood*. The vector representation model was generated from the reviews collected with the retrieval keyword “Norwegian Wood”, not “Haruki Murakami”. Because we use a bigram as a word unit, a bigram is supposed to be the given word to the vector model.

Given the word “Norwegian Wood”, the subset vector space was extracted and there the word “Love story”, “Beatles song”, “High school age”, “Friend suicide”, “Attempted suicide”, “Sense of loss“, “Ms. Nagasawa (major character)“, “Sexual

portrayal “, and " Sensual novel “ were extracted. These are all appearing as words similar to “Norwegian Wood” also in the analysis of the vector space model for the retrieval keyword "Murakami Haruki".

Compare to the Western review results, a notable difference is that the word “love story” appears only in the Japanese reviews and the concept “love” related words did not appear in the Western reviews.

## **Conclusion**

We conducted a comparison of reviews on the two Murakami novels, *Norwegian Wood* and *Kafka On The Shore*. The reviews first were collected in various countries Amazon review web sites. The analysis methods are generation of the vector representation model by word2vec and its vector model visualization by t-SNE.

In the reviews on *Kafka On The Shore*, there is no difference among the countries. The appearing words and review concepts are, for example, “Oedipal complex”, “Convergence of two worlds” which is leitmotif of Murakami novels, and “Mr Nakata” who can converse with cats. However, on the *Norwegian Wood* reviews, a notable difference is found that the word “love story” appears only in the Japanese reviews and the concept “love” related words did not appear in the Western reviews. When *Norwegian Wood* was sold explosively in Japan at the beginning, it seems that young people read and sympathize more than aged persons. We think of the difference reason as follows; When *Norwegian Wood* was translated into English as a popular novel and read by UK readers, it might have been read from the beginning with the stance of a best-selling novel. As a result, it seems that the reader became more aged, and the spirit of deeply reading the philosophical theme of the novel, like life and death, was strong.

Fujii says that the young middle class in the Chinese-speaking world is reading Murakami as a manual of urban consumption culture which was rapidly spread along with their economic growth. We are also interested in the relationship between economic growth and Murakami boom in Asia. As a future work, we will analyze the SNS written in Thai. As Japanese literature, there has not been a best-selling one like Murakami novels. For the cultural difference survey, we will continue to analyze and compare the reviews of Murakami novels in various countries.

## **Acknowledgement**

This research was partly supported by funds from the Gakushuin University Research Institute for Oriental Cultures as the research project in 2016 to 2017.

## References

Fujii, Shozo (2007), *村上春樹のなかの中国 (China in Haruki Murakami)* (Asahi sensho 826; Tokyo: Asahi-shinbun).

--- (2009), '札幌「ドルフィン・ホテル」の思い出と「東アジアと村上春樹」シンポジウムへの展開 (Memories of Dolphin Hotel in Sapporo and Development of " East Asia and Haruki Murakami" Symposium)', *世界は村上春樹をどう読むか (How the world reads Haruki Muakami) "A Wild Haruki Chase"* (Tokyo: Bungei-Shunju), 277-84.

Maaten, Laurens Van Der (2014), 'Accelerating t-SNE using tree-based algorithms', *J. Mach. Learn. Res.*, 15 (1), 3221-45.

Mikolov, Tomas, et al. (2013), 'Efficient Estimation of Word Representations in Vector Space', *arXiv:1301.3781*.

Miller, Laura (February 6th, 2005), 'Deep Chords: Haruki Murakami's 'Colorless Tsukuru Tazaki and His Years of Pilgrimage'', *The New York Times*.

Murakami, Haruki (2000), *Norwegian Wood*, trans. Jay Rubin (London: Harvill Press).

--- (2005), *Kafka On The Shore*, trans. J. Philip Gabriel (London: Harvill Press).  
Rubin, Jay (2005), *Haruki Murakami and the music of words* (London: Vintage Books ).

Rudder, Christian (2014a), *Dataclysm: What our online lives tell us about our offline selves* (London: 4th Estate).

--- (2014b), *Dataclysm: Who we Are\* \*When We Think No One's Looking* (New York: Crown Publishers).

Smith, Patti (August 5th, 2014), 'Deep Chords: Haruki Murakami's 'Colorless Tsukuru Tazaki and His Years of Pilgrimage'', *The New York Times*.

Sugimoto, Yoshio (2009), 'Japanese culture: An overview', in Yoshio Sugimoto (ed.), *The Cambridge Companion to Modern Japanese Culture (Cambridge Companions to Culture)* (Cambridge University Press).

**Contact email:** yukari.shirotaATgakushuin.ac.jp