10-2017

Using Data Visualization to Examine Translated Korean Literature

Kyung Eun (Alex) Hur

Hyokyoung Yi

Follow this and additional works at: https://scholarsarchive.byu.edu/jeal

Using Data Visualization to Examine Translated Korean Literature

Kyung Eun (Alex) Hur and Hyokyoung Yi

Introduction

In May 2016, the English translation of a South Korean fiction book called *The Vegetarian* won the Man Booker International Prize, one of the world's most prestigious literary awards. *The Vegetarian* was Han Kang's first novel to be translated into English by Deborah Smith, a British translator. The work was also chosen as one of the NY Times Book Review's "The 10 Best Books of 2016." It is considered translated Korean literature's biggest win since Kyung-Sook Shin's *Please Look After Mom* won the closing Man Asian Literary Prize in 2012.

Korean literature is becoming one of the major Korean cultural icons for global citizens, joining the Korean waves of drama and K-POP. The translation of literary works into other languages is a key factor in connecting people with different languages. In light of the increasing attention to the translation of Korean literary works, the idea emerged to gather data on the Korean literary works that have been translated into other languages to better examine the historical and current trends of Korean literary translations.

Many print publications of the compilation of translated works exist (Han'guk Munhak Pŏnyŏgwŏn, 2002; Kim, 1998; Kim, 2012; Park, 1993), and the Literature Translation Institute of Korea (LTI Korea) Library offers an online bibliography of the Korean literature translated into various languages on its website (http://library.klti.or.kr/bibliography). However, these thick reference volumes do not provide a quick and easy snapshot for understanding the whole picture of the translation trends in Korean literature. Although the online bibliography does provide some easier access to translated titles by language, publication year, country, and genre, neither resource provides visual access that can be grasped quickly. They also do not have the interactive functions that users can use to view, sort, and rank the information in a visually presentable way.

As data analysis and visualization have been rapidly introduced in many research areas, including humanities, this project's goal was set to visualize the translation data of Korean literature using Tableau, an interactive visualization tool, to better examine the topic. The study also aimed to discover any new findings or hidden information through data analysis with the help of new technology. Finally, the study hopes to bring a new perspective to the topic so that users can dig in for more information on their own.

Methodology

Data Source

In order to create an accurate and meaningful visualization, it was very important for us to carefully select accurate and reliable data. We began the data-selection process by looking through credible online open sources that hold catalogued data of translated Korean literature, such as OCLC WorldCat, National Library of Korea (국립중앙도서관), and LTI Korea (한국문학번역원). Manually collecting all bibliographic data from these open sources was not possible. For example, OCLC WorldCat offered no freely available option to search for all the books translated from Korean. Thus, we contacted the other open source organizations to get data on the translated books. However, obtaining such data from OCLC WorldCat and National Library of Korea was not successful. OCLC WorldCat's data service was very costly, and National Library of Korea's data were lacking in terms of the amount provided. In the end, we decided to go with the data provided by LTI Korea, which did not involve costs but had a comprehensive number of bibliographic records for the use of our study.

Data Description

LTI Korea's catalogue had a total of 4,526 book records. Each book's record included metadata on the author, publication date, original title, translated title, translated language, genre, and translator. In order to improve the dataset, we also manually added data on the original year of translation to determine the translation time for the books. Our target literary period was from the Korean pre-modern classics to modern literature. We filtered out non-literary materials, such as biographies, self-improvement books, children's literature, and books with multiple or unknown authors. In terms of genres, we included in the dataset fiction, poetry, soap opera scripts, and diaries/journals. The exclusion was made to make the data more consistent and relevant to each other for comparison and analysis. After filtering out the unnecessary data, we were left with a total of 2,172 books. These data were used for this project.

Data Processing

We pre-processed the finalized data to ensure data quality by making sure that the information was complete, consistent, and accurate (Lemieux et al., 2014). Much of the data, especially the titles and years of translation, was missing, inaccurate, or inconsistent. Thus, it took some time for thorough data cleaning and editing. In this step, we made sure that the data did not have any missing values or inconsistencies. We also conducted discrepancy

detection to ensure that there were no outliers or errors in the domain of each attribute. Finally, we excluded the inconsistent use of numbers, names, and data representations.

Data Visualization

After pre-processing the data, we sought the best possible way to visualize the data. We decided to use Tableau as our tool for various reasons. First, Tableau offers a tabular nature for viewing data. Second, it is the most recent and up-to-date interactive visualization tool that is flexible and fast. Third, Tableau was freely available to test and had all the necessary visual presentation features that we wanted to use for our study. In addition to the public nature of this program, which made it easier to access by anyone who wished to, we were able to get the Tableau desktop version for free so we could freely test the data for visual presentation. We also wanted to create a way for users to easily view and understand the data. Tableau fulfills this goal with its fast interactivity and good user interface. Tableau enables the quick transformation of data into a picture, allowing for various types of visualizations. It also has the ability to make the visualizations pleasing and interactive for the user to understand different parts of the data in different views.

Bar graphs were chosen to represent the ranks of translated books and authors as they are best for comparing different groups. A line graph was chosen to visualize the number of books translated each year as it is best for showing trends. Packed bubble charts were chosen to easily compare the popularity of genres and languages to the rest based on differences in sizes. To give a variety of information in one view, we applied various filters in each section.
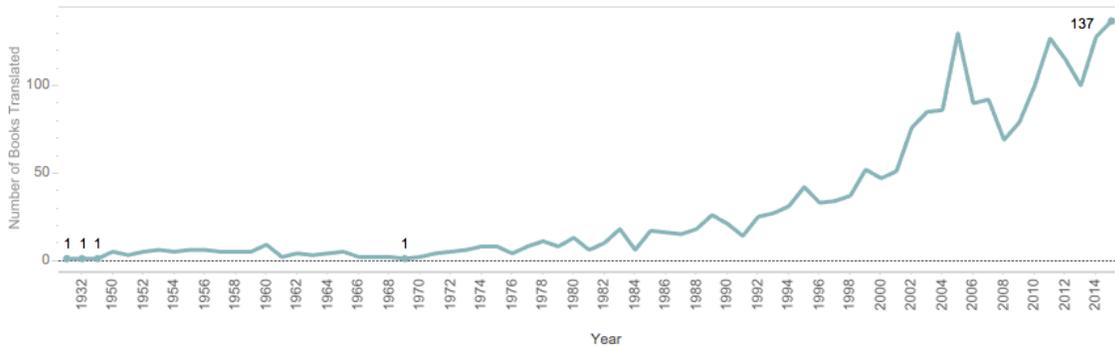
Tableau Presentation and Analysis

An interactive visualization using the LTI Korea data was created and designed using Tableau Desktop (see captured image below). This work is published on Tableau Public at https://public.tableau.com/profile/publish/TransKoreanLit/Master#!/publish-confirm. Interactivity of the visualization allows the user to view selected information in detail. When the user hovers over a section, such as a bar in a bar chart, a point on a trend line, or a section in a pie chart, a tooltip with detailed information appears and the associated graphs may change.

Translated Korean Literature

Literature Rank

Based on Literature Translation Instiute of Korea's bibliographic catalog, a total of 2,171 Korean literature has been translated from 1922 to 2015. We defined literature as books that fall into the category of fiction, poem, drama scripts, poem, and diary/journal. We excluded non literary materials such as biographies, self-improvement books, children's literature, and books with multiple or unknown authors. Below is the visulization of the data. You can filter by genre, language, or rank.

Genre	Languages	Year Translated	Rank		
(All)	(All)	(All)	1		15

Rank	Title
1	엄마를 부탁해
2	구운몽
2	우리들의 일그러진 영웅
4	나는 나를 파괴할 권리가 있다
4	시인
6	새
6	채식주의자
8	난장이가 쏘아올린 작은 공
8	오래된 정원
8	우리들의 행복한 시간
8	하늘과 바람과 별과 시
8	홍길동전
13	광장
13	님의 침묵
13	만인보
13	아름다운 영가
13	해를 품은 달
13	혀

French
German
Urdu
Chinese
Czech
English
Polish
Portuguese

137

1 1 1   1

Number of Books Translated

100

50

0

1932 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014

Year

Publication History

Average years between the literature's date of publication and translation is **16.87**.
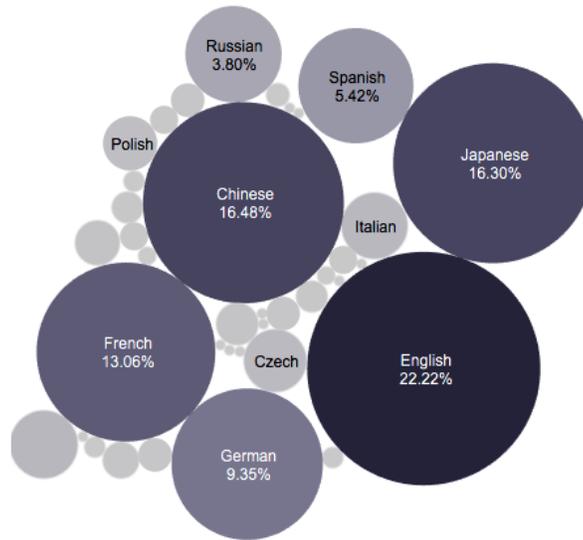
Genre

Play

Fiction
67.44%

Poem
23.61%
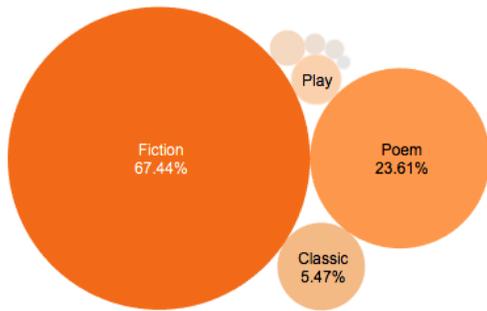
Classic
5.47%

There is a total of 8 genres.
* Classic refers to Korean literatures that have
been written before mid-19th century.

Language

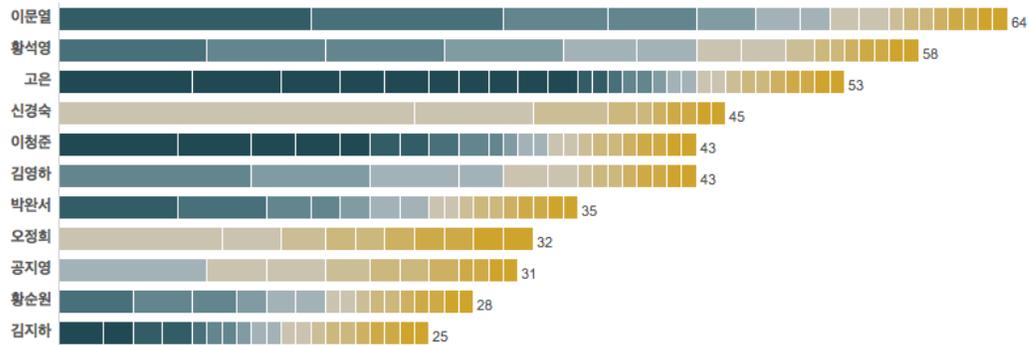A total of X languages that has been translated into.

Search

(All)

Russian
3.80%

Spanish
5.42%

Polish

Chinese
16.48%

Italian

Japanese
16.30%

French
13.06%

Czech

English
22.22%

German
9.35%

Authors Rank

There is a total of 565 authors that had their book(s) translated until 2015.

Language	Genre	Year Translated
(All)	(All)	(All)

Author	Count
이문열	64
황석영	58
고은	53
신경숙	45
이청준	43
김영하	43
박완서	35
오정희	32
공지영	31
황순원	28
김지하	25

Literature Rank

In the Literature Rank section, users can see the rank of translated Korean literature. As the bar graph shows, *Please Take Care of My Mother* (엄마를 부탁해) by Kyung-Sook Shin is the most translated piece of literature. If the user hovers over a bar, a tooltip appears with information on the author, publication year, number of languages it has been translated into, and number of times it has been published. In addition, a book's cover image, year line graph, and translated languages section change to reflect the book's information. For example, when the user hovers over the *Please Take Care of Mom* bar, the book image on the right updates to its book cover and, under it, the translated languages section updates according to the relevant data; the year line graph also changes accordingly. When *Please Take Care of Mom* is selected, it shows 15 different years and the number of published books, representing the publication trend of the translated book. When users change the slide, they can view all the ranks. The ranks can also be re-sorted by language and genre at the user's discretion.

Please Take Care of My Mother was translated into 20 different languages and published 24 times. This book is also the most recently published book among the top 20 translated Korean books. Despite having the most translated versions of a book, the author Shin ranked fourth on the list of the most translated authors. One can hypothesize that the topic of the literature is the foremost important factor of its success across different cultures. The main theme of Shin's book, motherhood, is understood and empathized universally. It would be interesting to determine if this is evident across all top books. It would also be important to see the use of the literature outside of classrooms since the primary market for translated books are for course materials (Galasso and Park, 2017).

Locally renowned literature published after 2010 is not on the top 35 list, which comprises of 45 books. A few examples are *28* by Jeong Yu-jeong (published in 2013) and *Eungyo* (은교) by Park Bum Shin. One possible reason for the lack of post 2010 literature in the top list is the long length of time needed for translation. According to LTI Korea's data, the average time for a book to be translated was 16.86 years (excluding the Korean pre-modern classics published before the New Literature era in 1894).

Year

The line graph shows an upward trend for translated books published between 1922 and 2015. The graph represents the fact that Korean literature was largely unknown to the world until the 1980s (Cerralbo, 2011). After the 1980s, the outside world showed an increased curiosity, and the government strongly pushed for the translation of Korean literature (Cerralbo, 2011). Thus, the path to translation was not always the norm. There was a major drop in the publication of translated books from 2005 to 2008, dropping from

130 published books in 2005 to 69 in 2008. Other factors affected the publication or translation market during that time, such as the 2007–2008 global financial crisis. The crisis is assumed to have affected the entire world, causing the book publishing market to diminish and thereby affecting the translation and publication of Korean literature. In 2009, the number of translated books began to increase again, and the total number of annual publications increased by 58 between 2008 and 2011. The rise could be linked to the new Korean Wave, *Hallyu 2.0,* that began in 2008 (Jin, 2012). Another point to note is the difference between the number of translated version of the top-rated books and all remaining books. The average number of translated versions for the top 20 books (0.07% of the total) is 15, compared to one for the rest of the books (98.3%). This shows that that the market for translated Korean books is not evenly distributed. It is heavily concentrated on a few books published between 1979 and 2008.

Genre and Language

The pie charts in the middle show the distribution of genres and languages. English is the most translated language (22%), and fiction is the most translated genre (67%). It is not surprising that English is the most translated language since it is a global language that is being used in more than 80 countries (CIA, 2017). Chinese (16.48%) and Japanese (16.30%) followed, which was also predicted as they are neighboring countries to Korea with cultural similarities. An examination of the translated languages rank could reveal undiscovered relationships between Korea and other countries. Fiction (67.64%) as the most popular genre is not surprising since most of Korean literature is fiction. As for poetry translations, it is the second largest genre at 23.67%. Korean poetry has a long history since the Chosŏn dynasty, which could be the reason for its large proportion. However, it would be interesting to get more insights by comparing other cultures' literature.

Authors Rank

The last section of the visualization is the rank of the authors. The top 30 authors who published the most translated versions of their work are, unsurprisingly, all authors who are renowned in Korea. The authors are ranked by the number of times their books have been translated and published. Yi Mun-yol is at the top; his books were published a total of 64 times. The bars next to the author's name give information on the books published. The segmented sections within a bar indicate one piece of literature. For example, Yi Mun-yŏl has 17 pieces of literature (segmented sections) that were translated and published into different languages. *Our Twisted Hero* (우리들의 일그러진 영웅) was published 17 times and translated 13 times. The shades of a bar represent the number of pieces of literature. The more literature, the more teal; the less literature, the more yellow. For example, Ko Un has

28 pieces of literature, so the bar shows more teal, but Oh Jung hee only has 10 pieces of literature, so her bar lacks teal and instead is more yellow.

Conclusion

As we are in the field of librarianship, we sought to use bibliographic metadata for a visualization using the most current application (i.e., Tableau) so that anyone could understand the data in easier and more meaningful ways. At the same time, the goal of this mini-study was to take full advantage of Tableau, which is increasingly popular and highly used in data visualization. We hope our study offers some thoughts for future research topics for libraries and librarians to explore and utilize bibliographic metadata, which we have long invested in, maintained, and owned across libraries.

We not only learned the importance of the data source, but also realized that—even for a smaller scale study like ours—we have to rely heavily on good quality data. Without such quality, the research could be full of fundamental flaws. In addition, the data cleaning process could be an overwhelming job, further putting at risk the subsequent step of data analysis or visualization.

The most challenging issue for our study was securing the data set using open sources so that we could minimize the expense that might be incurred for any project like this. For example, OCLC World Cat records are freely available public records, but crawling through the bibliographic metadata from the current platform is limited in terms of enabling any public to obtain and use the data freely and comprehensively. We spent quite a lot of time figuring out possible approaches through OCLC, but ultimately none were feasible without going through their fee-based services in order to obtain the information we sought. Furthermore, as librarians, we lacked very specific skills and knowledge dealing with technology involving data crawling. Even with help from experts in those areas, we spent some time concluding that the data were only available through OCLC's fee-based services due to the database's complexity and large scale. It would be our wish for OCLC to provide the bibliographic data with a pre-developed tool enabling any researcher or member of the public to use the data set more easily.

Still much remains to be discovered if more comprehensive bibliographic data from OCLC could be easily accessible for future research. For example, the current study could be expanded to world literature and translated works. It would be fascinating to understand the relationship between world literature and translated works through languages, publication dates and countries, genres, and locations of holding libraries, as such an approach would provide a much bigger scale for visualizing the whole picture. Such efforts would highlight some interesting trends or phenomena in translation studies.

Using Tableau's technology for data uploading and its convenience, it would be more meaningful and useful to update the translation data regularly if LTI Korea could provide them on a regular basis. Our study ends with the most current data (as of early 2016), but

the translated works keep growing every day, with more and more books being published in world languages. Capturing the most recent picture of translated literature today would be highly desirable.

This study has many limitations in terms of data set, analysis, visual presentation, and comparison with other Tableau presentations. Nonetheless, we hope our efforts have provided the opportunity to open up the discussion on potential use of bibliographic data for future related research involving cataloging metadata.

References

Cerralbo, Yoav. "Continuing Literary Society Under New Chief." *The Korean Herald*, November 11, 2011.
http://www.koreaherald.com/common_prog/newsprint.php?ud=20111113000287&dt=2

CIA. "The World Factbook—Field Listing: Languages." Accessed April 15, 2017.
https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html

Galasso, Regina and Chang Young Park. "An Interview with Ju-Chan and Bruce Fulton" *The Massachusetts Review,* https://www.massreview.org/node/437.

Han'guk Munhak Pŏnyŏgwŏn. *An Annotated Bibliography of Korean Literature in Translation, 1980-2002.* Seoul, Korea: Korea Literature Translation Institute, 2002.

Jin, Dal Yong. "Hallyu 2.0: The New Korean Wave in the Creative Industry." *Journal of the International Institute* 2 (2012)

Kim, Hŭng-gyu, Han'guk Munhak Pŏnyŏk Kŭmgo, and Koryŏ Taehakkyo. Minjok Munhwa Yŏn'guso. *Han'guk Munhak Pŏnyŏk Sŏji Mongnok*. Sŏul T'ŭkpyŏlsi; 서울 특별시 : Han'guk Munhak Pŏnyŏk Kŭmgo : Koryŏdae Minjok Munhwa Yŏn'guwŏn; 한국 문학 번역 금고 : 고려대 민족 문화 연구원, 1998.

Kim, Pyŏng-ch'ŏl. *Segye Munhak Pŏnyŏk Sŏji Mongnok Ch'ongnam 1895-1987*. 초판.; Ch'op'an. ed. Segye Munhak Iipsa Yŏn'gu ; Che 7-kwŏn. Sŏul-si; 서울시 : Kukhak Charyowŏn; 國學 資料院, 2012.

Lemieux, Victoria, Brianna Gormly, and Lyse Rowledge. "Meeting Big Data challenges with visual analytics: The role of records management." *Records Management Journal* 24 (2014): 122-141.

Park, On-Za. *A Bibliography of Korean Literature in English or Translated into English*. Seoul: Hanshin Pub., 1993.