İSTANBUL UNIVERSITY PRESS

Acta Infologica

Acta Infologica,
ACIN 2025, 9 (2): 580–596

https://doi.org/10.26650/acin.1598277

Submitted: 08.12.2024
Revision Requested: 26.02.2025
Last Revision Received: 09.11.2025
Accepted: 18.11.2025
Published Online 03.12.2025

Research Article

Open Access

Evaluation of Domain-Specific Vocabulary with Machine Learning-Based Techniques: Japanese and Russian Case Studies

Ali Aycan Kolukısa [1] & Baktygul Kulamshaeva Kolukısa [2]

[1] Çanakkale Onsekiz Mart University, Faculty of Humanities and Social Sciences, Department of Eastern Languages and Literatures, Çanakkale, Türkiye
[2] Çanakkale Onsekiz Mart University, School of Foreign Languages - Dr. Lecturer (Article 31) and Tourism Faculty PhD Student, Çanakkale-Türkiye

Abstract Foreign language education is one of the prominent requirements. Undergraduate students at the Faculty of Tourism are offered the opportunity to learn a second foreign language, which will contribute to their professional lives. However, this second foreign language, which is taught from the beginner level, cannot contribute to the students' professional lives at a desired level unless it includes professional technical terms related to their profession. For this reason, foreign language education books should include field words related to the professional field to a certain extent. This study examines the suitability of foreign language education books used at the basic level in Russian and Japanese courses from the scope of their field speciality. First, the frequently used words in the fields of "Tourism and Hotel Management" and "Tourism Guidance" were determined and set as the keywords. Then, depending on these keywords, other frequently used words were obtained using machine learning and natural language processing techniques. For this purpose, we used Python's Gensim library, and we established corpuses of word vectors consisting of both the keywords and the near-distanced words to these keywords in each field with the help of pre-trained word vector models. This study revealed statistically to what extent the textbooks currently used contain the domain-specific vocabulary in the field.

Keywords NLP · Japanese · Russian · Foreign Language · Tourism.

Citation: Kolukısa, A. A. & Kolukısa, B. K. (2025). Evaluation of domain-specific vocabulary with machine learning-based techniques: Japanese and Russian case studies. *Acta Infologica*, *9*(2), 580-596. https://doi.org/10.26650/acin.1598277

Corresponding author: Ali Aycan Kolukısa aliaycan.kolukisa@comu.edu.tr

Introduction

Tourism is a rapidly growing sector worldwide and in Turkey. The increase in the number of tourists coming to Turkey and the growth in the income obtained from the sector at the same rate are also related to the employees' foreign language skills to a certain extent. Inadequate language education can negatively affect tourist satisfaction and country promotion. In this regard, Balcı (1998) mentioned that foreign language education is not provided at a sufficient level in higher education institutions in our country. Therefore, sector employees cannot sufficiently satisfy tourists coming to the country, and thus, our country cannot be fully promoted (see İşigüzel, 2013; Balcı, 1998).

Foreign language education is one of the most important elements in business life. The efficiency of foreign language education depends on many factors, such as the methods used, teaching staff, physical conditions, and the use of the right materials, as well as on the acquisition of vocabulary appropriate for the targeted sector. Textbooks are at the center of the education process and provide content for students. In this context, foreign language textbooks prepared for tourism undergraduate students are also required to include relevant knowledge, skills, and relevant vocabulary as well.

Related Studies

This study examines the suitability of basic-level Japanese and Russian textbooks taught as elective second foreign languages in Tourism Faculties for the field of tourism. Many studies have been conducted on the suitability analysis of textbooks (see Demirel, 2013; İşci, 2012; Çelik, 2011; Beyazit, 2013). However, it is quite difficult to come across a similar field-focused study that specifically addresses the field of tourism. Although this issue has been addressed in a similar way in Balcı and Metin (2019), only English as a foreign language has been examined.

For example, in Balcı and Metin (2019)'s study investigating the suitability of English textbooks for tourism undergraduate students, various sections of the textbook were examined using the scanning method and the relevant data (visual, exercise, topic, word preference) were collected and analyzed using the content analysis method. The words and concepts in the texts were examined and evaluated according to criteria such as:

- Cultural and educational appropriateness
- Target age group of the topics
- Gender equality
- Listening, speaking, writing, and reading skills
- Vocabulary
- Cultural comparison
- Professional knowledge and weighted professions
- etc.

(Balcı and Metin, 2019)

The textbook examined in Balcı and Metin (2019) is suitable for the field in terms of vocabulary; it includes sufficient words related to the most frequently used concepts in the tourism sector, such as hotel and hotel management, touristic activities, travel, cultural characteristics, food and beverage, self-promotion, communication, shopping, trip, and restaurant. In this context, a student who completely learns the textbook in question will flawlessly learn the vocabulary groups he/she will need regarding tourism. This situation is also directly proportional to the fact that English education has a very long and deep-rooted history. However, the situation is quite different for Russian and Japanese, which are taught as second foreign languages in Tourism Faculties. Compared to English, Japanese and Russian, which do not have a deep history like English language education, are still in their infancy in terms of foreign language education in Turkey.

From this point of view, it is normal not to come across any preliminary studies investigating the suitability of Russian and Japanese textbooks for the field of tourism, and this study aims to fill this gap to a certain extent.

Method

Natural language processing and programming techniques were used in the study. The textbooks «Minna no Nihongo-1 and 2» for Japanese and «Doroga v Rossiyu-1» for Russian were used as examples, and the degree to which the vocabulary in these books matches the words used in the professional field was evaluated. On the other hand, the study differs from previous studies in that it addresses both Japanese and Russian, and it utilizes NLP and programming techniques, as well.

Creating A Domain-Specific Word2vec Corpus

In the study, first 2 fields were determined as «Tourism and Hotel Management» and «Tourism Guidance». Subsequently, 186 field keywords that are frequently used and provided separately for each field were determined. In creating these keyword lists, based on the experience and knowledge of the authors, both the word lists published on the web by the trending big tourism companies on their own web pages in this field and the word lists prepared by the non-profit organizations in the field of tourism were used[1]. These words were accepted as keywords for our study.

These words were then input to openly accessible pre-trained word vector models using the Gensim library, an open-source library of Python. Thus, three different corpora consisting of pre-trained word vector models based on ML techniques with a total of approximately 6 million words were used.

The “corpora” mentioned here is an assembly of 3 different “pre-trained word vector corpus” formed by word vectors consisting of 200–300 dimensions that contain the vectorial calculations of the words for each language. The pre-trained models used in this study separately for Japanese and Russian for creating a large word vector corpora are shown below with the reference sources and the detailed information of each model (release date/size/dimension). Further information and the download links of the models are given below for each of the pre-trained models.

[1] https://www.tourism.jp/tourism-database/glossary/
https://www.cogito-kobo.net/OmoshiroTabiErabi/gyokaiyogo/ryokoyogoshu.html
http://insertum.com/category/107/

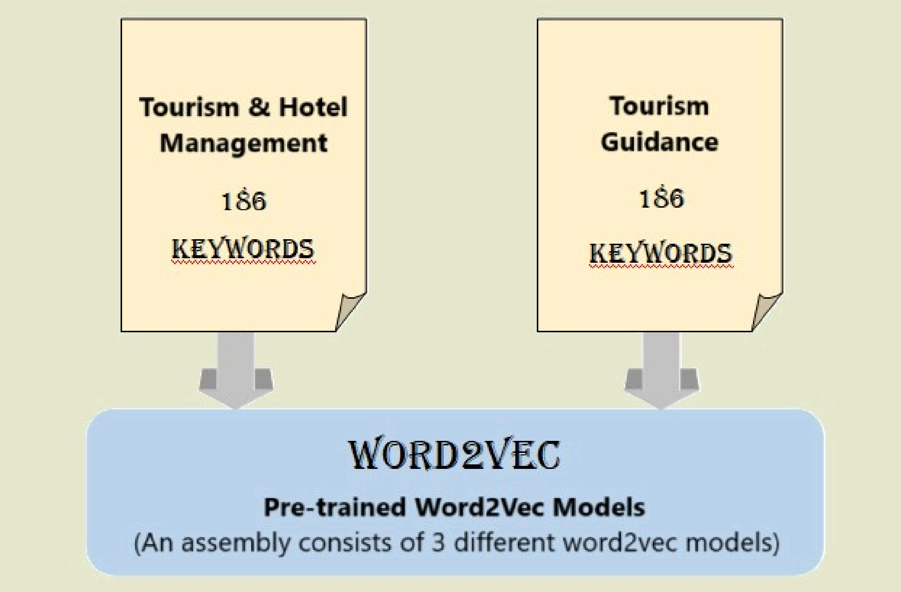
Table 1

Japanese pre-trained models	Russian pre-trained models
1) jawiki.all_vectors.200d.txt (Jun 2019/3423MB/200d) (https://github.com/singletongue/WikiEntVec/releases/)	1) model.txt (Jan 2019/631MB/300d) (http://vectors.nlpl.eu/repository/20/180.zip)
2) entity_vector.model.txt (Feb 2017/1906MB/200d) (https://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/)	2) ruwikiruscorpora_upos_skipgram_300_2_2018.vec (https://rusvectores.org/en/models/) (Dec 2017/1077MB/300d)
3) cc.ja.300.vec (Sep 2018/4429MB/300d) (https://fasttext.cc/docs/en/crawl-vectors.html)	3) cc.ru.300.vec (Feb 2018/4430MB/300d) (https://fasttext.cc/docs/en/crawl-vectors.html)

In a normal corpus, the situations of being in the same sentence or context or collocation are considered when determining the connections of words with each other. In pre-trained word vector models, this situation is expressed only numerically. Therefore, the proximity or distance of words to each other is expressed with numbers in pre-trained vector models. In such pre-trained models, the distance or proximity of any word to another word is obtained using various mathematical calculations. In the study, cosine similarity was used as a standard, and the model requested 10 words closest to each given keyword. Figure 1 shows the input of the field-related keywords into the pre-trained word vector corpora. And, Figure 2 in the next page, shows what kinds of domain-specific word2vec corpuses were created as a result.

Figure 1

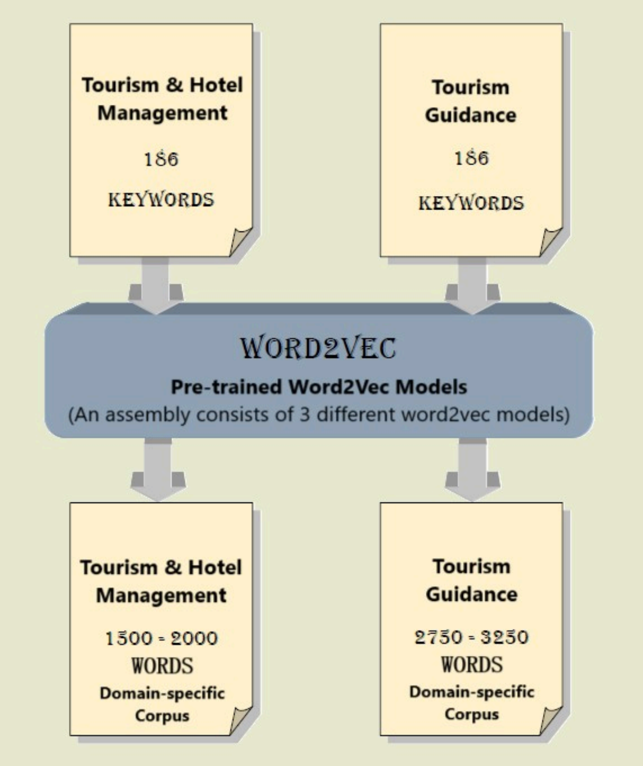
Field Keywords


Tourism & Hotel Management
186
KEYWORDS
Tourism Guidance
186
KEYWORDS
WORD2VEC
Pre-trained Word2Vec Models
(An assembly consists of 3 different word2vec models)

Thus, 186 words in Japanese and Russian, which were available separately for each of the fields, were entered as input into the word vector corpora, and withinside it was requested to return 10 closest words to each keyword from every pre-trained word2vec model. However, because the word vector corpora are consisted of 3 different pre-trained word2vec models, there were normally some cases where the model could not return the same number of words or the inputted keyword was not present in one or more of the consisted models. Therefore, the number of words returned by each keyword varied accordingly.

Figure 2

Word Corpuses

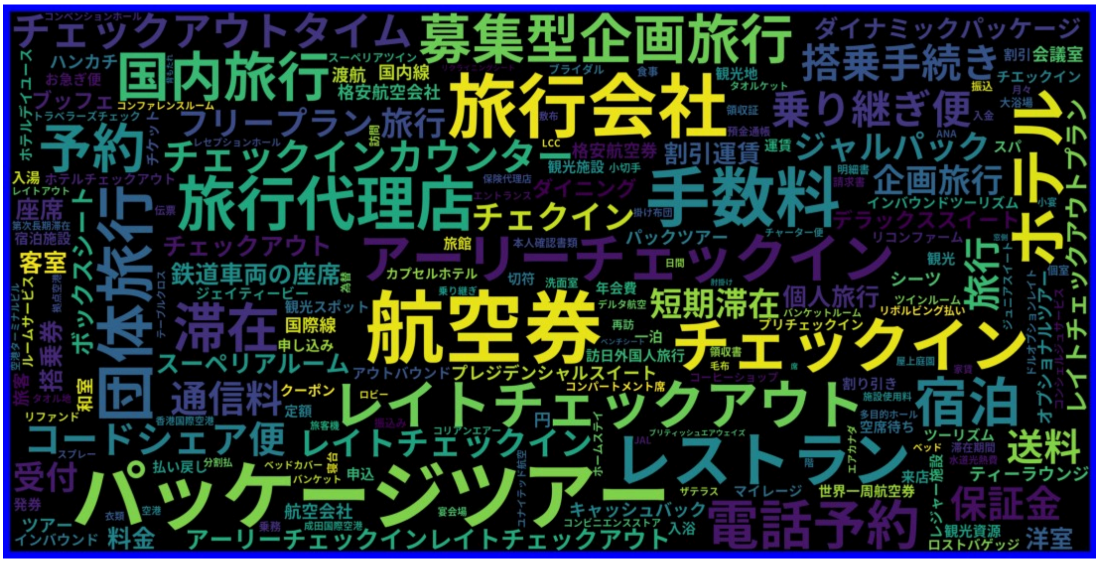


However, it was also possible to reach the same word more than once even though different keywords were used. This situation was useful for calculating the frequency of the field words reached through keywords.

To track which field words obtained from the corpus through keywords are encountered more frequently, all field words obtained by repetition were visualized and converted into a word cloud.

Figure 3

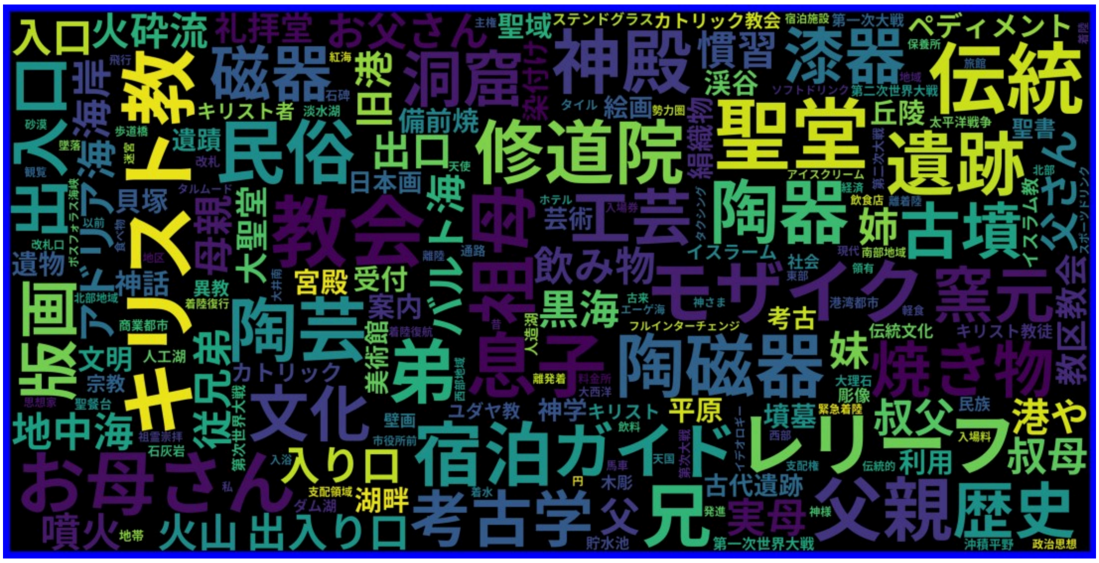
Japanese Corpus Word Cloud for Tourism and Hotel Management



In order to see the frequency of the words reached through each keyword, Figure 3 shows the Japanese word cloud for the “Tourism and Hotel Management” field, which was created before the uniquification process used afterwards to be able to obtain statistically accurate results. In the word cloud, the word that was reached with the highest frequency through various keywords was “航空券 (Flight Ticket)” and when the number of times this word appeared in the model was requested, it was found to be reached 9 times in total. For this reason, this word is written in the largest font possible. The next word was “パッケージツアー (Package Tour)”, which appeared exactly 8 times. The other words were “旅行会社 (Travel Agency)” 7 times and “手数料 (Commission)” 6 times, and their sizes are also from largest to smallest accordingly. Conversely, words that appeared only twice, such as “ベンチシート (Bench seat)” or “ベッドカバー (Bed cover)” are written in the smallest visible font and included in the word cloud, whereas words that appeared only once are not included in the word cloud because they are both numerous and have a font size below the visible size. In the next stage, the frequency of Japanese words obtained through keywords is visualized, but this time related to the “Tourism Guidance” field. Similarly, the words in the “Tourism Guidance” field, obtained through different keywords in the “Tourism Guidance” field, are considered before the uniquification process and converted into the word cloud in Figure 4.

Figure 4

Japanese Corpus Word Cloud for Tourism Guidance



The words reached with the highest frequency through the keywords in the field of tourism guidance are “聖堂 (Cathedral)”, “キリスト教 (Christianity)”, “教会 (church)” and “伝統 (tradition)”, and each one of these words was reached 5 times. For this reason, it is seen with the highest font size in the word cloud. Afterwards, the words “神殿 (temple)”, “遺跡 (ruins)”, “陶器 (pottery)”, “陶磁器 (ceramic)”, “陶芸 (ceramic art)”, “民俗 (folklore)”, “モザイク(mosaic)”, “レリーフ (relief)”, “宿泊 (accommodation)”, “お母さん (mother)”, “父親 (father)”, “祖母 (grandmother)”, “息子 (son)”, “兄(elder brother)”, “弟(younger brother)” and “修道院 (monastery)” followed, and the occurrence frequency of these words was 4; therefore, they are in the second place in terms of font size. Then, the words “文化 (culture)”, “ガイド (guide)”, “洞窟 (cave)”, “焼き物 (pottery)”, “磁器 (porcelain)”, “漆器 (varnished wooden product)”, “窯元 (pottery kiln)”, “古墳 (ancient tomb)”, “考古学 (archeology)”, “工芸 (craft)”, “歴史 (history)”, “版画 (printmaking)”, “出入口(entrance&exit)”, “出口(exit)”, “入口(introduction)”, “慣習(tradition)”, “飲み物 (beverage)”, “地中海(Mediterranean Sea)”, “アドリア海 (Adriatic

Sea)”, “バルト海 (Baltic Sea)”, “黒海(Black Sea)”, “お父さん(father)”, “母親(mother)”, “姉(older sister)”, “実母 (real mother)”, “叔母 (aunt)”, “妹(younger sister)”, “父(father)”, “叔父 (uncle)”, “従兄弟 (cousin)”, “海岸 (beach)”, “火山 (volcano)”, “噴火 (volcano eruption)”, “火砕流 (pyroclastic flow)”, “旧港 (old harbor)”, “礼拝堂 (chapel)”, “教区教会 (local church)” and “大聖堂 (large cathedral)” were encountered 3 times each, and thus their sizes are shown smaller than the ones that were reached 5 and 4 times. The words that only accessed 2 times were visualized with the smallest font that could be seen.

The high frequency observed in this way indicates that these words are likely to appear in any context related to that professional environment, and therefore, such words with high frequency should be taught most particularly.

The same operations were applied in the same way, but this time Russian words were in target, and the word cloud was established before the uniquification process, allowing word frequency to be observed.

Figure 5

Russian Corpus Word Cloud for Tourism and Hotel Management

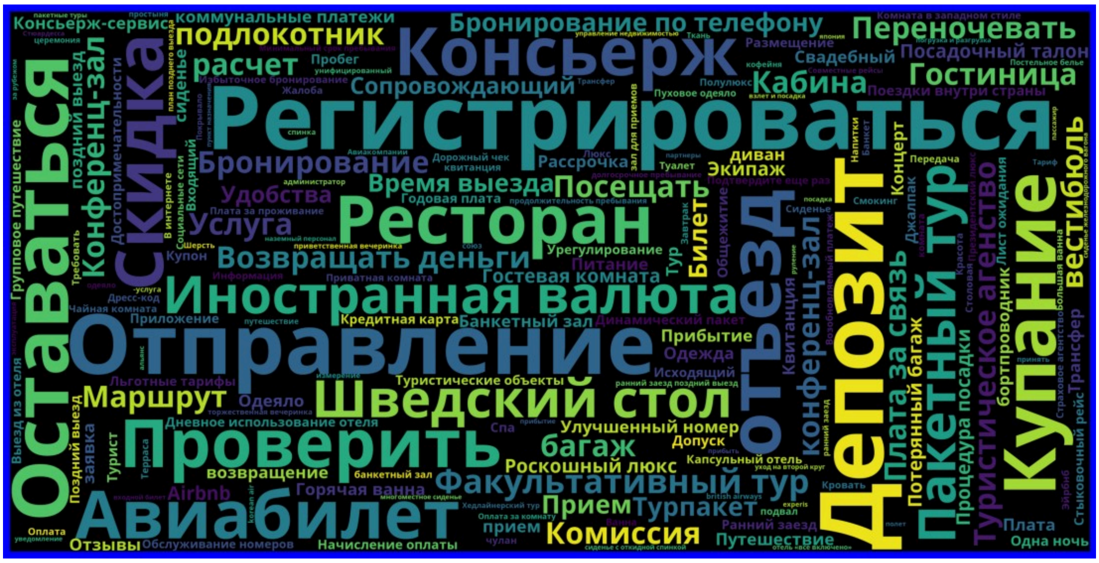
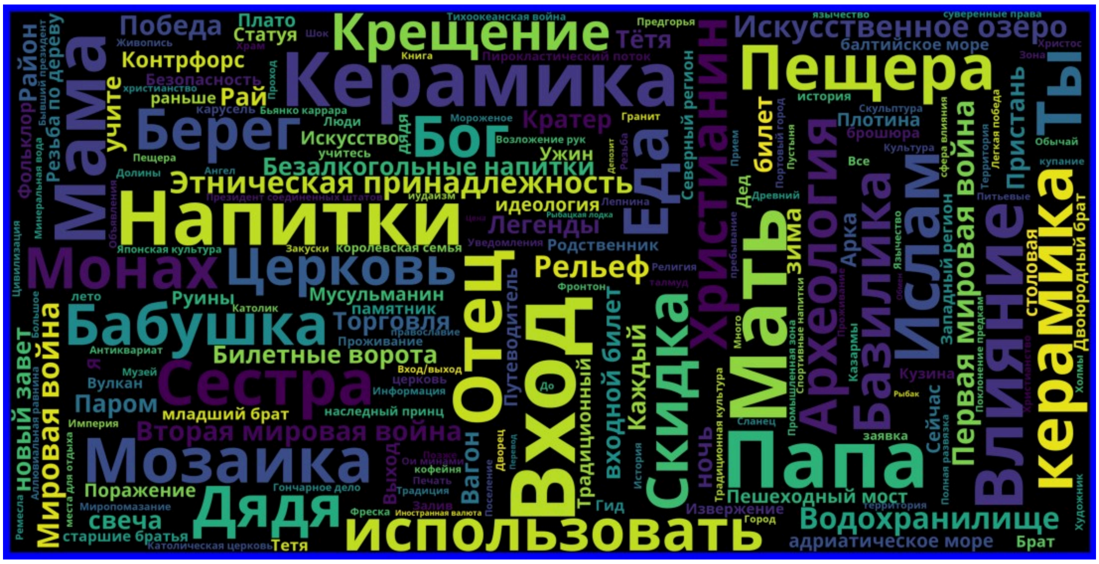


Figure 6 shows the Russian word cloud in the field of “Tourism and Hotel Management”, which was reached by Russian keywords before the uniquification process to be able to visualize the word frequency. The words with the highest frequency reached by keywords are “Регистрироваться (hotel registration)” and “Депозит (deposit).” These two words were reached 9 times and were therefore written with the largest font in the word cloud. The word “Отправление (departure)” was reached 8 times, and it is also visualized with a large font. Next, the word “Оставаться (stay/accommodation)” was reached 7 times; the words “Купание (bathing)” and “отъезд (checking out)” were found 6 times; and the words “Авиабилет (flight ticket)”, “Ресторан (restaurant)”, “Скидка (discount)”, “Проверить (check)”, “Консьерж (concierge)”, “Иностранная валюта (foreign currency)”, “Шведский стол (buffet)”, “Пакетный тур (paket tour)”, “Факультативный тур (optional tour)” were found 5 times. The font sizes of these words are also directly proportional to their frequency, from largest to smallest in the cloud. The frequency of the words “Гостиница (hotel)”, “Переночевать (overnight)”, “Посещать (to visit)”, “Туристическое агентство (travel agency)”, “Бронирование (reservation)”, “Комиссия (commission)”, “Возвращать деньги (refund)”, “Кабина (cabin)”, “Плата за связь (communication fee)”, “Конференц-зал (conference hall)”, “вестибюль (lobby)”, “Услуга (service/service)”, “расчет (payment/account)”, “подлокотник (armrest)”, “багаж (luggage)”, “Маршрут (route)" were all 4, and these words are visualized with a smaller font. The words "Turpaket (Travel Package)", "Бронирование по телефону (Phone Reservation)", "Билет (ticket)", "Прием" (reception)", "Время выезда (check-out time)", "Сопровождающий (companion)", "Удобства (convenience/comfort)", "Посадочный талон (boarding pass)", "Процедура посадки (boarding)", "Airbnb", "Потерянный багаж (lost luggage)", "Гостевая комната (guest room)", "Экипаж (crew)", "Роскошный люкс (luxury suite)", "Улучшенный номер (superior room)", "Отзывы (review/comment)", "Свадебный (wedding)", "Питание (food and beverage)", "Консьерж-сервис (Concierge service)", "Путешествие (travel)", "Горячая ванна (hot tub)", "Квитанция (receipt)", "Банкетный зал (banquet hall)", "комунальные платежи (common service fees-invoices)", "Одежда (clothing)", "Одеяло (blanket)", "поздний выезд (late check-out)", "заявка (application form)", "прием (admission)", "Тур (tour)", "Концерт (concert)", "сиденье (seat)", "диван (sofa)", "Прибытие (arrival)", "возвращение (return)", "бортпроводник (cabin crew)", "Рассрочка (installment)", "Трансфер (transfer)" and "Плата (fee)" were reached 3 times, so their font size is smaller than from the words with a frequency of 4. Other words with a frequency of 2 are visualized using the smallest readable font.

Finally, the word frequency of Russian words related to the "Tourism Guidance" field was obtained with the help of the keywords in that field. The output is again visualized as a word cloud, as shown in Figure 6.

Figure 6

Russian Corpus Word Cloud for Tourism Guidance



The Russian word with the highest frequency in the tourism guidance field through keywords was "Вход (introduction)", which appeared 13 times and has the largest font size in the word cloud. Subsequently, "'Папа (father)" and "Мать (mother)" appeared 10 times, "Керамика (ceramics)" and "Напитки (beverages)" appeared 9 times, "Мама (mother)" appeared 8 times and "Влияние (effect)" and "Отец (father)" appeared 7 times, and their sizes are from largest to smallest in the same proportion. The words "Ислам (Islam)", "Мозаика (mosaic)", "Пещера (cave)", "Еда (food)", "Ты (you)", "Сестра (sister)" were found 6 times each, and their sizes were slightly smaller than the others. Following these words, the words "Бог (God)", "Скидка (discount)", "использовать (to use)", "Дядя (uncle)", "Бабушка (grandmother)", "Берег (shore)" and "Монах (monk)" were found 5 times each; therefore, their sizes were slightly smaller than the others. Words with frequency of 4 are "Базилика (basilica)", "Крещение (christening)", "Археология (archeology)", "Христианин (Christian)", "Церковь (church)", "Этническая принадлежность (ethnicity)", "Рельеф (relief)", “Водохранилище (cistern)”, “Искусственное озеро (artificial lake)”, “Первая мировая война (World War 1)”, “Вторая мировая война (World War 2)”, “Победа (victory)”, “Билетные ворота (ticket turnstile)”, “Легенды (legends)”, “Безалкогольные напитки (non-alcoholic beverages)”, “входной билет (entrance ticket)”, “билет (ticket)”, “Рай (heaven)”, “ночь (night)”, “Вагон (wagon)”, “Торговля (commerce)”, “Район (district)”, “Тётя (aunt)”, “Кратер (crater)”, “Пристань (marina)”, “Паром (ferry)”, “свеча (candle)”, “зима (winter)”, “новый завет (Bible)”, “учите (teach)” and these words are reflected in the word cloud in a readable size. Followingly the smaller sized words which were reached 3 times, are “Арка (belt/mount)”, “Контрфорс (buttress)”, “Пешеходный мост (pedestrian bridge)”, “Путеводитель (guide)”, “Плато (plain/plateau)”, “Руины (ruins)”, “Искусство (art)”, “Мусульманин (muslim)”, “Фольклор (folklore)”, “Резьба по дереву (wood carving)”, “Статуя (sculpture)”, “Плотина (dam)”, “Поражение (defeat)”, “Выход (exit)”, “Традиционный (traditional)”, “Ужин (dinner)”, “столовая (dining room)”, “идеология (idology)”, “адриатическое море (Adriatic Sea)”, “балтийское море (Baltic Sea)”, “Северный регион (northern region)”, “Западный регион (western region)”, “Дед (grandfather)”, “Брат (brother)”, “памятник (monument)”, “Безопасность (security)”, “раньше (formerly)”, “младший брат (younger brother)”, “Двоюродный брат (male cousin)”, “Родственник (relative)”, “старшие братья (brothers)”, “Гид (tour guide)”, “Кузина (female cousin)”, “брошюра (brochure)”, “Вулкан (volcano)”, “Извержение (explosion)”, “Пирокластический поток (pyroclastic flow)”, “Залив (gulf)”, “Проживание (accommodation)”, “Казармы (barracks)”, “Все (all)”, “заявка (application form)”, “карусель (carousel)”, “лето (summer)”, “история (history)”, “Люди (people)”, “наследный принц (crown prince)” and “Королевская семья (royal family)”. The other words with the smallest font are the words that were found 2 times, and the words that reached only 1 time are not included in the cloud.

Among the words whose frequency is seen through the word cloud above, the words with a rate of 2 and higher should be included in the basic level vocabulary, as it will increase the students’ familiarity with those words in their professional lives, and this will directly affect their success in that profession. On the other hand, when we look at both Japanese and Russian field words, in the “Tourism Guidance” field, in addition to field-related terms, words encountered in daily life are also frequently encountered. The reason for this is that during the act of the guidance profession, there are more sections from daily life compared to the “Tourism and Hotel Management” field.

Results

At this stage, the repetitive words in the word corpuses of the “Tourism and Hotel Management” and “Tourism Guidance” fields, which were previously prepared and reached through multiple keywords for the calculation of word frequency, were uniquified. Thus, nonrepetitive word corpora in Japanese and Russian were obtained separately. The current numbers of the word counts of the corpuses before and after the uniquification process are given below.

Table 2

Uniquification process and word count

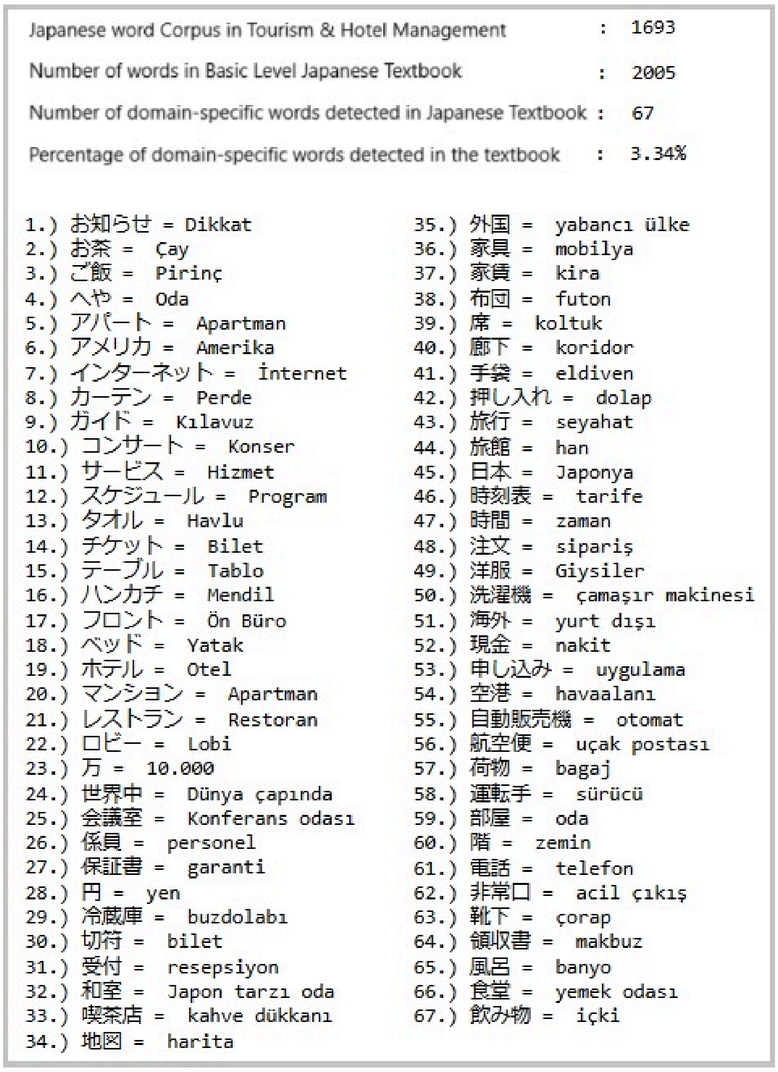
	Before the Uniquification Process		After the Uniquification Process	
Japanese Word Corpus (JWC)	Tourism and Hotel Management	2033	Tourism and Hotel Management	1693
	Tourism Guidance	3251	Tourism Guidance	2860
Russian Word Corpus	Tourism and Hotel Management	2033	Tourism and Hotel Management	1681
	Tourism Guidance	3251	Tourism Guidance	2685

In the next stage, a list of Japanese and Russian words was created by scanning the words in textbooks used to teach these languages at the basic level. A total of 2005 nonrepetitive words were identified in the "Minna no Nihongo 1-2" series, which is used for basic Japanese education. A total of 3030 nonrepetitive words were identified in the book "Doroga v Rossiyu 1", which is used for basic Russian education.

In the last stage, the percentage of words in corpora consisting of field words of "Tourism and Hotel Management" and "Tourism Guidance" that are also present in the textbooks for teaching basic levels of Japanese and Russian was calculated. The intersection set of the words was determined using the Numpy library in Python programming, and the results were given as percentages. In addition, Turkish translations were automatically added to the matching words using Google Translate for easier understanding. As a result, only 67 words out of a total of 2005 words scanned in basic level of Japanese textbooks intersected with the words of "Tourism and Hotel Management" corpora, and this corresponds to only 3.34% of the total words in the field.

Figure 7

Findings of Tourism and Hotel Management in Basic Level Japanese Textbook[2]



```
Japanese word Corpus in Tourism & Hotel Management                 :  1693

Number of words in Basic Level Japanese Textbook                   :  2005

Number of domain-specific words detected in Japanese Textbook      :  67

Percentage of domain-specific words detected in the textbook       :  3.34%


1.) お知らせ = Dikkat                   35.) 外国 =  yabancı ülke
2.) お茶 =  Çay                         36.) 家具 =  mobilya
3.) ご飯 =  Pirinç                      37.) 家賃 =  kira
4.) へや =  Oda                         38.) 布団 =  futon
5.) アパート =  Apartman                39.) 席 =  koltuk
6.) アメリカ =  Amerika                 40.) 廊下 =  koridor
7.) インターネット =  İnternet          41.) 手袋 =  eldiven
8.) カーテン =  Perde                   42.) 押し入れ =  dolap
9.) ガイド =  Kılavuz                   43.) 旅行 =  seyahat
10.) コンサート =  Konser               44.) 旅館 =  han
11.) サービス =  Hizmet                 45.) 日本 =  Japonya
12.) スケジュール =  Program            46.) 時刻表 =  tarife
13.) タオル =  Havlu                    47.) 時間 =  zaman
14.) チケット =  Bilet                  48.) 注文 =  sipariş
15.) テーブル =  Tablo                  49.) 洋服 =  Giysiler
16.) ハンカチ =  Mendil                 50.) 洗濯機 =  çamaşır makinesi
17.) フロント =  Ön Büro                51.) 海外 =  yurt dışı
18.) ベッド =  Yatak                    52.) 現金 =  nakit
19.) ホテル =  Otel                     53.) 申し込み =  uygulama
20.) マンション =  Apartman             54.) 空港 =  havaalanı
21.) レストラン =  Restoran             55.) 自動販売機 =  otomat
22.) ロビー =  Lobi                     56.) 航空便 =  uçak postası
23.) 万 =  10.000                       57.) 荷物 =  bagaj
24.) 世界中 =  Dünya çapında            58.) 運転手 =  sürücü
25.) 会議室 =  Konferans odası          59.) 部屋 =  oda
26.) 係員 =  personel                   60.) 階 =  zemin
27.) 保証書 =  garanti                  61.) 電話 =  telefon
28.) 円 =  yen                          62.) 非常口 =  acil çıkış
29.) 冷蔵庫 =  buzdolabı                63.) 靴下 =  çorap
30.) 切符 =  bilet                      64.) 領収書 =  makbuz
31.) 受付 =  resepsiyon                 65.) 風呂 =  banyo
32.) 和室 =  Japon tarzı oda            66.) 食堂 =  yemek odası
33.) 喫茶店 =  kahve dükkanı            67.) 飲み物 =  içki
34.) 地図 =  harita
```

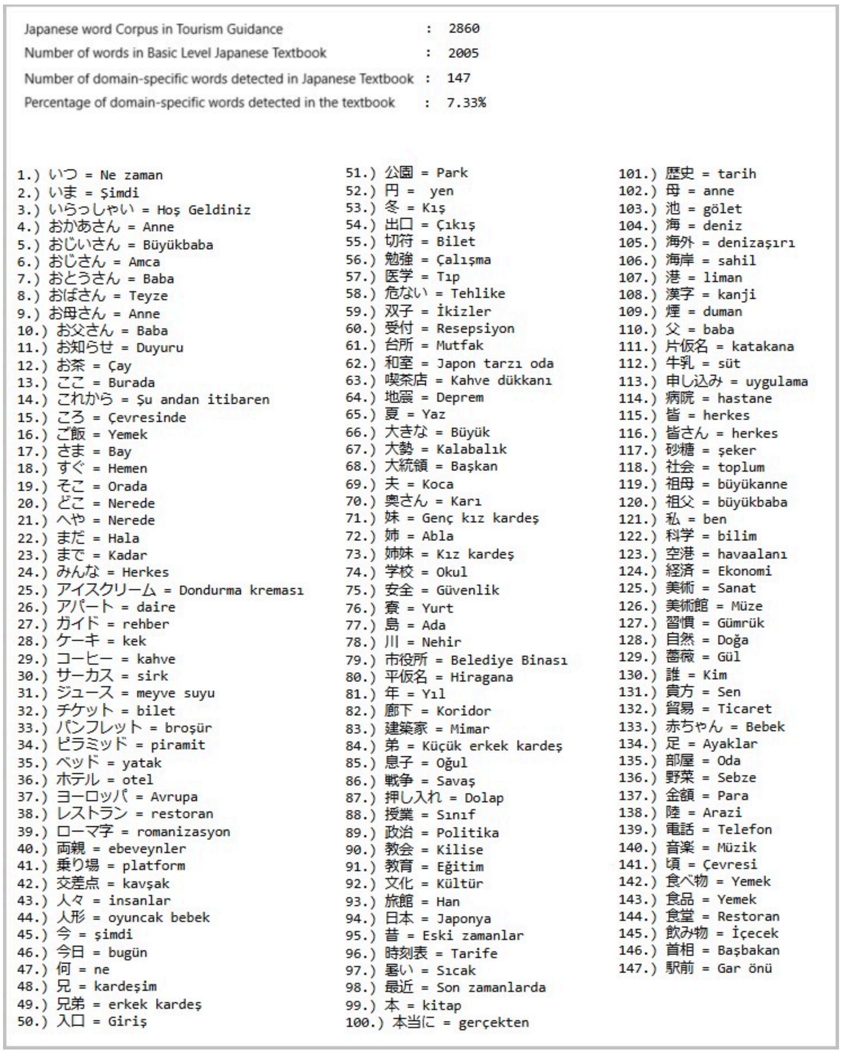
[2] Due to the automated translation embedded in the system codes, not all the meanings of the words are returned correctly, and deficiencies in the translation of some words such as "お知らせ", "ご飯", "テーブル", "旅館", "時刻表", "申し込み", "階" are encountered.

While paying attention to the words, Japanese field words with a high frequency are rarely encountered. For example, the words “ホテル (hotel)” and “レストラン (restaurant)” appear 5 times, “旅行 (travel)” appear 4 times, and “ハンカチ (handkerchief), “会議室 (Conference room)”, “円 (yen)”, “受付(reception)” and “和室 (Japanese-style room)” appear 3 times, while “チケット (ticket)”, “ベッド (bed)”, “ロビー (lobby)”, “切符 (ticket)”, “家賃 (rent)”, “席 (seat)”, “旅館 (Japanese-style accommodation facility)”, “申し込み (application)”, “空港 (airport)”, “階 (floor)” and “領収書 (receipt)” have a lower frequency and occur only twice. The frequency of the other words is 1. Therefore, only 19 out of 67 words have a frequency greater than 1, while 48 words occur only once, which shows that only 28.35% of the words that match the field corpus of the textbook consist of words with a frequency greater than 1 (at least 2 or more times). This situation is a very negative factor in terms of the quality of vocabulary education in the target language.

It was also observed that 147 words out of 2005 words scanned in basic level Japanese textbook intersected with the words of “Tourism Guidance” corpora, and this corresponds to 7.33% of the total words in the field. The intersecting words in the corpora and textbook are shown in Figure 8.

Figure 8

Findings of Tourism Guidance in Basic Level Japanese Textbook



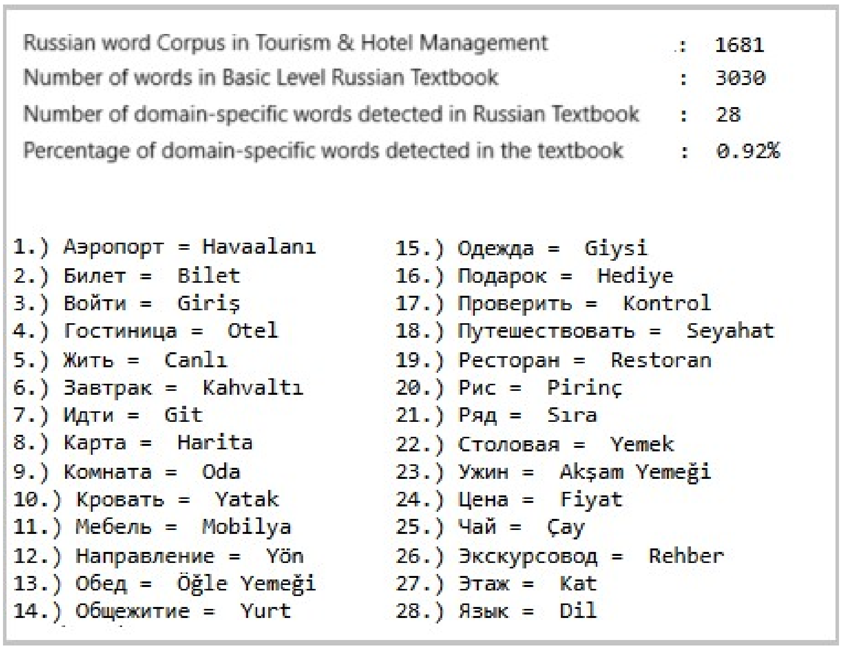
```
Japanese word Corpus in Tourism Guidance                    :  2860
Number of words in Basic Level Japanese Textbook            :  2005
Number of domain-specific words detected in Japanese Textbook :  147
Percentage of domain-specific words detected in the textbook :  7.33%

1.) いつ = Ne zaman
2.) いま = Şimdi
3.) いらっしゃい = Hoş Geldiniz
4.) おかあさん = Anne
5.) おじいさん = Büyükbaba
6.) おじさん = Amca
7.) おとうさん = Baba
8.) おばさん = Teyze
9.) お母さん = Anne
10.) お父さん = Baba
11.) お知らせ = Duyuru
12.) お茶 = Çay
13.) ここ = Burada
14.) これから = Şu andan itibaren
15.) ころ = Çevresinde
16.) ご飯 = Yemek
17.) さま = Bay
18.) すぐ = Hemen
19.) そこ = Orada
20.) どこ = Nerede
21.) へや = Nerede
22.) まだ = Hala
23.) まで = Kadar
24.) みんな = Herkes
25.) アイスクリーム = Dondurma kreması
26.) アパート = daire
27.) ガイド = rehber
28.) ケーキ = kek
29.) コーヒー = kahve
30.) サーカス = sirk
31.) ジュース = meyve suyu
32.) チケット = bilet
33.) パンフレット = broşür
34.) ピラミッド = piramit
35.) ベッド = yatak
36.) ホテル = otel
37.) ヨーロッパ = Avrupa
38.) レストラン = restoran
39.) ローマ字 = romanizasyon
40.) 両親 = ebeveynler
41.) 乗り場 = platform
42.) 交差点 = kavşak
43.) 人々 = insanlar
44.) 人形 = oyuncak bebek
45.) 今 = şimdi
46.) 今日 = bugün
47.) 何 = ne
48.) 兄 = kardeşim
49.) 兄弟 = erkek kardeş
50.) 入口 = Giriş

51.) 公園 = Park
52.) 円 =  yen
53.) 冬 = Kış
54.) 出口 = Çıkış
55.) 切符 = Bilet
56.) 勉強 = Çalışma
57.) 医学 = Tıp
58.) 危ない = Tehlike
59.) 双子 = İkizler
60.) 受付 = Resepsiyon
61.) 台所 = Mutfak
62.) 和室 = Japon tarzı oda
63.) 喫茶店 = Kahve dükkanı
64.) 地震 = Deprem
65.) 夏 = Yaz
66.) 大きな = Büyük
67.) 大勢 = Kalabalık
68.) 大統領 = Başkan
69.) 夫 = Koca
70.) 奥さん = Karı
71.) 妹 = Genç kız kardeş
72.) 姉 = Abla
73.) 姉妹 = Kız kardeş
74.) 学校 = Okul
75.) 安全 = Güvenlik
76.) 寮 = Yurt
77.) 島 = Ada
78.) 川 = Nehir
79.) 市役所 = Belediye Binası
80.) 平仮名 = Hiragana
81.) 年 = Yıl
82.) 廊下 = Koridor
83.) 建築家 = Mimar
84.) 弟 = Küçük erkek kardeş
85.) 息子 = Oğul
86.) 戦争 = Savaş
87.) 押し入れ = Dolap
88.) 授業 = Sınıf
89.) 政治 = Politika
90.) 教会 = Kilise
91.) 教育 = Eğitim
92.) 文化 = Kültür
93.) 旅館 = Han
94.) 日本 = Japonya
95.) 昔 = Eski zamanlar
96.) 時刻表 = Tarife
97.) 暑い = Sıcak
98.) 最近 = Son zamanlarda
99.) 本 = kitap
100.) 本当に = gerçekten

101.) 歴史 = tarih
102.) 母 = anne
103.) 池 = gölet
104.) 海 = deniz
105.) 海外 = denizaşırı
106.) 海岸 = sahil
107.) 港 = liman
108.) 漢字 = kanji
109.) 煙 = duman
110.) 父 = baba
111.) 片仮名 = katakana
112.) 牛乳 = süt
113.) 申し込み = uygulama
114.) 病院 = hastane
115.) 皆 = herkes
116.) 皆さん = herkes
117.) 砂糖 = şeker
118.) 社会 = toplum
119.) 祖母 = büyükanne
120.) 祖父 = büyükbaba
121.) 私 = ben
122.) 科学 = bilim
123.) 空港 = havaalanı
124.) 経済 = Ekonomi
125.) 美術 = Sanat
126.) 美術館 = Müze
127.) 習慣 = Gümrük
128.) 自然 = Doğa
129.) 薔薇 = Gül
130.) 誰 = Kim
131.) 貴方 = Sen
132.) 貿易 = Ticaret
133.) 赤ちゃん = Bebek
134.) 足 = Ayaklar
135.) 部屋 = Oda
136.) 野菜 = Sebze
137.) 金額 = Para
138.) 陸 = Arazi
139.) 電話 = Telefon
140.) 音楽 = Müzik
141.) 頃 = Çevresi
142.) 食べ物 = Yemek
143.) 食品 = Yemek
144.) 食堂 = Restoran
145.) 飲み物 = İçecek
146.) 首相 = Başbakan
147.) 駅前 = Gar önü
```

Again, the situation is not much different when attention is paid to the words, and there are only 39 words with a frequency of occurrence greater than 1, which corresponds to only 26.53% of all matching words. Out of these 39 words with a frequency greater than 1 in the corpora of Tourism Guidance, “教会 (church)” appears 5 times; “お母さん (mother)”, “兄 (brother)”, “弟 (younger brother)”, “息子 (son)”, “祖母 (grandmother)” appear 4 times; “お父さん (father)”, “ガイド (guide)”, “入口 (intro)”, “出口 (exit)”, “妹 (younger sister)”, “姉 (older sister)”, “文化 (culture)”, “歴史 (history)”, “海岸(beach)”, “父 (father)” and “飲み物 (drink)” appear 3 times, and “おかあさん (mother)”, “おじさん (uncle)”, “おばさん (aunt/aunt)”, “アイスクリーム (ice cream)”, “ホテル (hotel)”, “両親 (parents)”, “円 (yen)”, “受付 (reception)”, “夫 (husband)”, “奥さん (wife)”, “年 (year)”, “旅館 (Japanese style accommodation facility)”, “昔 (long ago/old times)”, “母 (mother)”, “港 (port)”, “社会 (society)”, “祖父 (grandfather)”, “私 (I/me)”, “経済 (economy)”, “美術館 (art museum)”, “食べ物 (food)” and “駅前 (front of the station)” appear 2 times. Therefore, although the words in the field of “Tourism Guidance” are encountered approximately twice as often as the words in the field of “Tourism and Hotel Management” in the books used for basic level Japanese education, since the percentage of the words that emerged being encountered more than once in the field corpus is 26.53, it shows that 73.47% of the matched textbook words have a low importance in terms of word frequency in the field, which can be considered as a phenomenon that can be attributed to the importance of the word, and hence to the low quality of education. For example, if it is considered that the words with high frequency are the indicators of a concept that is encountered more frequently in that profession, it can be concluded that the inability to understand such words will constitute a greater obstacle than other words in analyzing the events in professional life and therefore will have greater importance compared to the words that appear only once. Based on this, it can be said that the education of the words with high frequency will have a greater importance and play a role in the quality of vocabulary education in that language.

These rates are even lower for the Russian section. Only 28 of the 3030 words in the basic Russian textbook intersect with the “Tourism and Hotel Management” field corpus.

Figure 9

Findings of Tourism and Hotel Management in Basic Level Russian Textbook

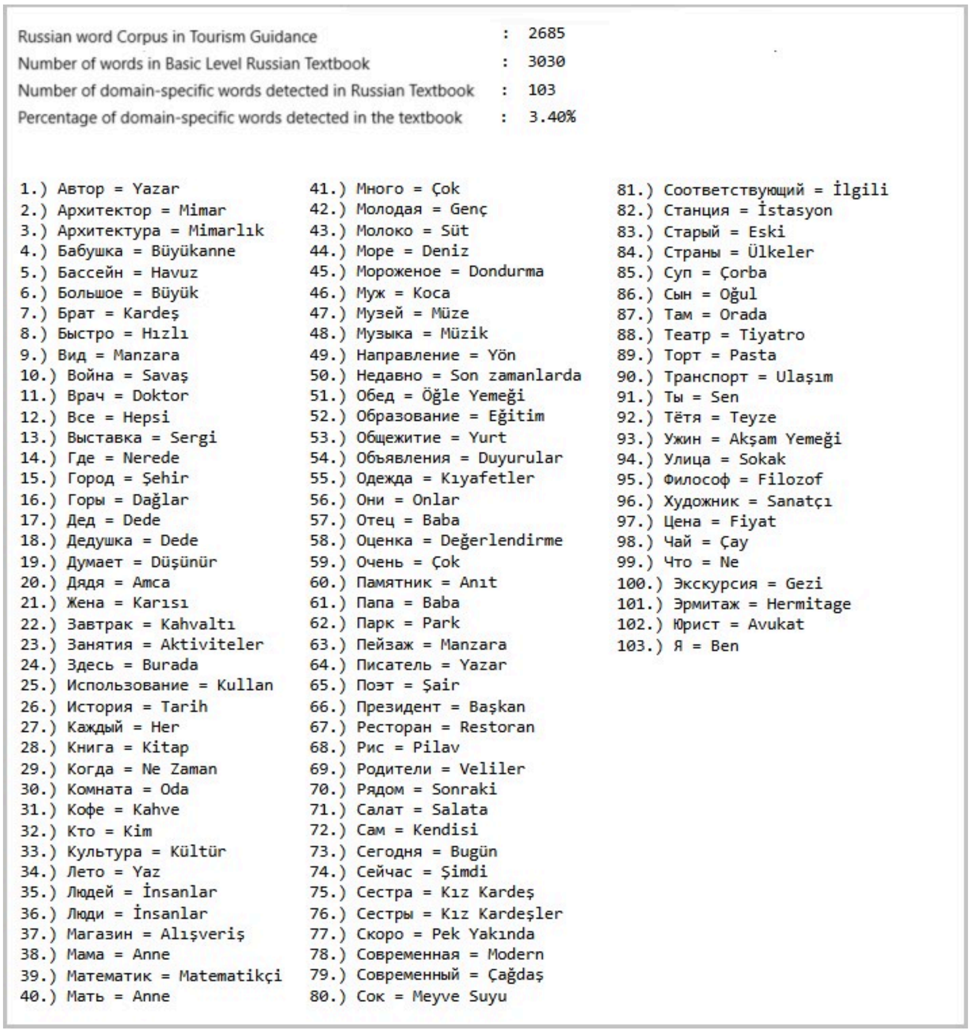


```
Russian word Corpus in Tourism & Hotel Management                .:  1681
Number of words in Basic Level Russian Textbook                   :  3030
Number of domain-specific words detected in Russian Textbook      :  28
Percentage of domain-specific words detected in the textbook      :  0.92%


1.) Аэропорт = Havaalanı         15.) Одежда =  Giysi
2.) Билет =  Bilet               16.) Подарок =  Hediye
3.) Войти =  Giriş               17.) Проверить =  Kontrol
4.) Гостиница =  Otel            18.) Путешествовать =  Seyahat
5.) Жить =  Canlı                19.) Ресторан =  Restoran
6.) Завтрак =  Kahvaltı          20.) Рис =  Pirinç
7.) Идти =  Git                  21.) Ряд =  Sıra
8.) Карта =  Harita              22.) Столовая =  Yemek
9.) Комната =  Oda               23.) Ужин =  Akşam Yemeği
10.) Кровать =  Yatak            24.) Цена =  Fiyat
11.) Мебель =  Mobilya           25.) Чай =  Çay
12.) Направление =  Yön          26.) Экскурсовод =  Rehber
13.) Обед =  Öğle Yemeği         27.) Этаж =  Kat
14.) Общежитие =  Yurt           28.) Язык =  Dil
```

These words in Figure 9 correspond to only 0.92% of the words in the Russian "Tourism and Hotel Management" field corpus, which is very low. In addition, "Проверить (to check)" and "Ресторан (restaurant)" appear 5 times; "Гостиница (hotel)" appear 4 times; "Билет (ticket)" and "Одежда (clothes)" appear 3 times, and "Завтрак (breakfast)", "Кровать (bed)", "Общежитие (dormitory)", "Столовая (dining room)" and "Экскурсовод (guide)" appear only 2 times. Words with a frequency of 2 and above correspond to only 35.71% of the total words. Figure 10 shows the intersection of the words in the Russian "Tourism Guidance" field and the words in the basic level Russian textbook.

Figure 10

Findings of Tourism Guidance in Basic Level Russian Textbook



```
Russian word Corpus in Tourism Guidance                     :  2685
Number of words in Basic Level Russian Textbook             :  3030
Number of domain-specific words detected in Russian Textbook :  103
Percentage of domain-specific words detected in the textbook :  3.40%

1.) Автор = Yazar
2.) Архитектор = Mimar
3.) Архитектура = Mimarlık
4.) Бабушка = Büyükanne
5.) Бассейн = Havuz
6.) Большое = Büyük
7.) Брат = Kardeş
8.) Быстро = Hızlı
9.) Вид = Manzara
10.) Война = Savaş
11.) Врач = Doktor
12.) Все = Hepsi
13.) Выставка = Sergi
14.) Где = Nerede
15.) Город = Şehir
16.) Горы = Dağlar
17.) Дед = Dede
18.) Дедушка = Dede
19.) Думает = Düşünür
20.) Дядя = Amca
21.) Жена = Karısı
22.) Завтрак = Kahvaltı
23.) Занятия = Aktiviteler
24.) Здесь = Burada
25.) Использование = Kullan
26.) История = Tarih
27.) Каждый = Her
28.) Книга = Kitap
29.) Когда = Ne Zaman
30.) Комната = Oda
31.) Кофе = Kahve
32.) Кто = Kim
33.) Культура = Kültür
34.) Лето = Yaz
35.) Людей = İnsanlar
36.) Люди = İnsanlar
37.) Магазин = Alışveriş
38.) Мама = Anne
39.) Математик = Matematikçi
40.) Мать = Anne
41.) Много = Çok
42.) Молодая = Genç
43.) Молоко = Süt
44.) Море = Deniz
45.) Мороженое = Dondurma
46.) Муж = Koca
47.) Музей = Müze
48.) Музыка = Müzik
49.) Направление = Yön
50.) Недавно = Son zamanlarda
51.) Обед = Öğle Yemeği
52.) Образование = Eğitim
53.) Общежитие = Yurt
54.) Объявления = Duyurular
55.) Одежда = Kıyafetler
56.) Они = Onlar
57.) Отец = Baba
58.) Оценка = Değerlendirme
59.) Очень = Çok
60.) Памятник = Anıt
61.) Папа = Baba
62.) Парк = Park
63.) Пейзаж = Manzara
64.) Писатель = Yazar
65.) Поэт = Şair
66.) Президент = Başkan
67.) Ресторан = Restoran
68.) Рис = Pilav
69.) Родители = Veliler
70.) Рядом = Sonraki
71.) Салат = Salata
72.) Сам = Kendisi
73.) Сегодня = Bugün
74.) Сейчас = Şimdi
75.) Сестра = Kız Kardeş
76.) Сестры = Kız Kardeşler
77.) Скоро = Pek Yakında
78.) Современная = Modern
79.) Современный = Çağdaş
80.) Сок = Meyve Suyu
81.) Соответствующий = İlgili
82.) Станция = İstasyon
83.) Старый = Eski
84.) Страны = Ülkeler
85.) Суп = Çorba
86.) Сын = Oğul
87.) Там = Orada
88.) Театр = Tiyatro
89.) Торт = Pasta
90.) Транспорт = Ulaşım
91.) Ты = Sen
92.) Тётя = Teyze
93.) Ужин = Akşam Yemeği
94.) Улица = Sokak
95.) Философ = Filozof
96.) Художник = Sanatçı
97.) Цена = Fiyat
98.) Чай = Çay
99.) Что = Ne
100.) Экскурсия = Gezi
101.) Эрмитаж = Hermitage
102.) Юрист = Avukat
103.) Я = Ben
```

"Tourism Guidance" field words of Russian, just as seen in the field of "Tourism Guidance" for Japanese, have more intersections than "Tourism and Hotel Management" field. Moreover, there are 103 intersected words, which corresponds to a percentage of 3.4% for the fact that the number of words in the Russian textbook is 3030. As mentioned before, this is because the "Tourism Guidance" field words include more sections from daily life than the "Tourism and Hotel Management" field words. The number of words with a frequency of at least 2 and above in the "Tourism Guidance" field in the textbook is 36. And these words are; "Мать (mother)" and "Папа (father)" which appear 10 times; "Мама (mother)" which appears 8 times; "Отец (father/father)" which appears 7 times; "Сестра (sister)" and "Ты (you)" which appear 6 times; "Бабушка (grandmother)” and “Дядя (uncle)” which appear 5 times; “Каждый (each)” and “Тётя (aunt)” which appear 4 times; “Брат (brother)”, “Все (all)”, “Дед (grandfather)”, “Сейчас (now)”, “Ужин (dinner)” and “Я (me)” which appear 3 times; “Большое (elder)”, “Война (battle)”, “Город (city)”, “Жена (wife)”, “История (past)”, “Книга (book)”, “Комната (room)”, “Культура (culture)”, “Люди (people)”, “Много (lot)”, “Мороженое (ice cream)”, “Музей (museum)”, “Общежитие (dormitory)”, “Объявления (announcements)”, “Они (them)”, “Сестры (sister)”, “Станция (station)”, “Художник (artist)”, “Цена (price)” and “Экскурсия (trip)” which appear 2 times. Words with a frequency of at least 2 and above correspond to 34.95% of the total words.

Discussion and Conclusion

When the above results are examined from a general perspective, it is seen that both the basic level Japanese and Russian textbooks are quite inadequate in terms of words related to the professional fields of “Tourism and Hotel Management” and “Tourism Guidance”. On the other hand, the percentages of words in the field of “Tourism and Hotel Management” in the textbooks are lower than those in the field of “Tourism Guidance” for both basic-level foreign language textbooks. One of the biggest reasons for this is that since sections from daily life are encountered more frequently in the practice of the profession of “Tourism Guidance”, such everyday words are also in the majority of textbooks not prepared specifically for the field. However, since such daily words are found less frequently in the field of “Tourism and Hotel Management”, the number of matches is lower at the same rate. On the other hand, Rinaldi and Yuste (2004) claim that words that can be described as technical field words only account for around 30% of all words in a corpus related to a specific technical field. From this point of view, even when this ratio could be considered a minimum requirement for a written text in any field, the percentage of field words obtained from textbooks remains well behind the figures reported by Rinaldi and Yuste (2004). For this reason, it has been concluded that the foreign language books used for teaching basic levels of Japanese and Russian at the elective foreign languages courses in the Faculty of Tourism are not sufficiently suitable for the field of tourism. Both for the field of “Tourism and Hotel Management” and “Tourism Guidance”, there is a need for basic-level Japanese and Russian textbooks that cover the field words. In addition, these books should include the most frequently used field words used by people who are actually practicing in that field in their daily business lives.

Indeed, for Russian, there are some textbooks that could be used in the field of “Tourism and Hotel Management" and “Tourism Guidance”, as shown in Table 2. However, in spite of the existence of these books, only a few of these textbooks are suitable for the basic level of Russian courses, and it could be said that there is still a need for basic Russian textbooks in the field, especially for beginners.

Table 3

Uniquification process and word count

Title of the textbook	Author	Publication Year	Field	Level
Russkiy yazyk dlya gostinits i restoranov (nachal’nyy kurs)	**Golubeva, A. V., & Zadorina, A. I.**	**2023**	**Tourism and Hotel Management, Gastronomy**	**A1-A2**
Turizmde Mesleki Rusça 1.	**Sütcü, G., & Gogunokova, E.**	**2020**	**Tourism**	**A1-A2**
Turizm'de Rusça: BENİ TAKİP EDİNİZ!	Stoyanova, M.	2023	Tourism, Travel Management and Guidance	A2

Title of the textbook	Author	Publication Year	Field	Level
Turizm'de Rusça: İYİ SEYAHATLER!	Stoyanova, M.	2023	Tourism, Travel Management and Guidance	A2
Turizm'de Rusça: SEFA GETİRDİNİZ!	Stoyanova, M.	2023	Tourism and Hotel Management, Travel Management	A2
Turizm'de Rusça: HOŞ GELDİNİZ!	Stoyanova, M.	2021	Tourism and Hotel Management	A2
Turizm'de Rusça: İYİ TATİLLER DİLERİM!	Stoyanova, M.	2021	Tourism and Hotel Management	A2
Russkiy yazyk kak inostrannyy v professional’noy podgotovke spetsialistov po turizmu	Rakova, I. V.	2022	Tourism and Hotel Management	A2-C1
Servis: Prakticheskiy kurs russkogo yazyka dlya rabotnikov servisa	S. A. Khavronina, L. A. Kharlamova, I. V. Kaznyshkina	2017	Tourism and Hotel Management, Gastronomy	A2-C1
Pyat’ zvyozd: Ekspress-kurs po russkomu yazyku dlya rabotnikov servisa.	I. V. Kaznyshkina, S. A. Khavronina	2018	Tourism and Hotel Management, Gastronomy	A2-C1
Russkiy yazyk v industrii turizma: Uchebnoe posobie (uroven’ B1–B2).	Graudynya, Z. A.	2020	Tourism and Hotel Management, Travel Management and Guidance	B1-B2
Russkiy yazyk v sfere turizma (dlya inostrannykh uchashchikhsya)	Prokudina, I. S., & Kuks, A. V.	2022	Tourism (General)	C1-C2
Russkiy yazyk kak inostrannyy v industrii turizma i gostepriimstva	Satina, T. V., Gilovaya, E. A., and Zaytseva, I. A.	2023	Tourism and Hotel Management, Gastronomy	C1-C2

However, the need for basic level Japanese textbooks for the fields of “Tourism and Hotel Management" and “Tourism Guidance” is much more serious. As for the basic level of Japanese textbooks still on the market today, only a few newly published ones eventually started to have few chapters about tourism. For example, we can see some chapters related to tourism in the new textbook prepared for the beginner-intermediate level of Japanese learners, titled “初・中級日本語教科書 日本語が広げる世界（CERF A2 レベル程度）(SHO-CHŪKYŪ NIHONGO KYŌKASHO – NIHONGO GA HIROGERU SEKAI (A2)”, which was written by KONDOH Atsuko and MARUYAMA Chika, and released in September, 2025. Therefore, it is possible to conclude from the current situation that the need for basic level of Japanese textbooks to be able to give a qualified Japanese language education at the beginner levels, in the field of "Tourism and Hotel Management" and “Tourism Guidance”, is much urgent especially compared to Russian language.

Peer Review Externally peer-reviewed.

Author Contributions Conception/Design of Study- B.K.K., A.A.K.; Data Acquisition- B.K.K., A.A.K.; Data Analysis/Interpretation- B.K.K., A.A.K.; Drafting Manuscript- B.K.K., A.A.K.; Critical Revision of Manuscript- B.K.K., A.A.K.; Final Approval and Accountability- B.K.K., A.A.K.

Conflict of Interest The authors have no conflict of interest to declare.

Grant Support The authors have no conflict of interest to declare.

Author Details

Ali Aycan Kolukısa

[1] Çanakkale Onsekiz Mart University, Faculty of Humanities and Social Sciences, Department of Eastern Languages and Literatures, Çanakkale, Türkiye

0000-0003-1315-8678 ✉ aliaycan.kolukisa@comu.edu.tr

Baktygul Kulamshaeva Kolukısa

[2] Çanakkale Onsekiz Mart University, School of Foreign Languages - Dr. Lecturer (Article 31) and Tourism Faculty PhD Student, Çanakkale-Türkiye

0000-0002-1325-9324

References

Antonova V. Ye., Nakhabina M. M., et al. (2006). Doroga v Rossiyu. Uchebnik russkogo yazyka. (elementarnyy uroven') 1, Sankt-Peterburg: Zlatoust.

Balcı, T. (1998). Türkiye'de Germanistik ve Turizm Eğitimi. Sorunlar ve somut çözüm önerileri. ÇÜ Eğitim Fakültesi Yayınları. No: 15. Adana.

Balcı, U., & Metin, F. (2019). Turizm Lisans Öğrencilerine Yönelik Hazırlanan Yabancı Dil İngilizce Ders Kitaplarının Hedef Kitle Açısından Uygunluk Analizi. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, (47), 57-76.

Beyazit, H. (2013). *Yabancı Dil Olarak Türkçe ve İngilizce Ders Kitaplarındaki Öğrenme Stratejilerinin Kullanım* [Master Thesis, Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü, Yabancı Dil Olarak Türkçe Öğretimi Anabilim Dalı]. YÖK Tez Merkezi [No: 330214]. https://tez.yok.gov.tr/UlusalTezMerkezi/

Çelik, Ş.N. (2011). *Orta Öğretim İngilizce Ders Kitabı Breeze 9 Hakkında Öğrenci, Öğretmen ve Müfettiş Görüşleri* [Master Thesis, Hacettepe Üniversitesi, Sosyal Bilimleri Enstitüsü, Eğitim Bilimleri Ana Bilim Dalı]. YÖK Tez Merkezi [No: 308429]. https://tez.yok.gov.tr/UlusalTezMerkezi/

Demirel, M.B. (2013). *Yabancı Diller için Avrupa Ortak Başvuru Metni Kapsamında Delfin Ders Kitabının İncelenmesi* [Master Thesis, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Yabancı Diller Ana Bilim Dalı]. YÖK Tez Merkezi [No: 333419]. https://tez.yok.gov.tr/UlusalTezMerkezi/

İşci, C. (2012). *Türkçenin Yabancı Dil Olarak Öğretiminde Kullanılan 'Yeni Hitit' Ders Kitabının Dört Temel Dil Becerisi ve Kültür Açısından İncelenmesi* [Master Thesis, Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü, Yabancı Dil Olarak Türkçe Öğretimi Anabilim Dalı]. YÖK Tez Merkezi [No: 330205]. https://tez.yok.gov.tr/UlusalTezMerkezi/

İşigüzel, B. (2013). Turizm işletmeciliği ve otelcilik programlarındaki mesleki Almanca dersleri üzerine bir araştırma. *NWSA-Humanities*, 8(4), 363-371.

Kara, A. & Demirel, Ş. (2022). Yabancılar için Türkçe / Rusça Öğretim Kitaplarında Kültür Unsurları. *Çukurova Üniversitesi Türkoloji Araştırmaları Dergisi*, 7(1), 40-66.

Rinaldi, F., & Yuste, E. (2004). Exploiting Technical Terminology for Knowledge Management. Proceedings of the EKAW 2004 Workshop on Application of Language and Semantic Technologies to support Knowledge Management Processes, Application of Language and Semantic Technologies to support Knowledge Management Processes (LSTKM 2004), Whittlebury Hall, October 8, 2004. https://ceur-ws.org/Vol-121/02.pdf

Surīē nettowāku hen cho. (2004). Minna no Nihongo Shokyū I Honsatsu. Tōkyō :Surīē Nettowāku Shuppan.

Reference for pre-trained models

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Joulin, A., Grave, E., Bojanowski, P. Douze, M., & Jégou, H. Mikolov, T. (2016). FastText.zip: Compressing Text Classification Models. *ICLR*, 2017.

T. Mikolov, K. Chen, G., & Corrado Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ICLR*, 2013.

Mikolov, T., Sutskever, I., Chen, K., & Corrado Dean, J. G. (2013). Distributed Representations of Words and Phrases and their Compositionality (Publik et al., 2019). *NIPS*, 2013.

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, Masaryk University - Malta. pp. 45-50. Doi: 10.13140/2.1.2393.1847.

Suzuki, M., Matsuda, K., Sekine, S., Okazaki, N. & Inui, K. (2018). A Joint Neural Model for Fine-Grained Named Entity Classification of Wikipedia Articles. IEICE Transactions on Information and Systems, Special Section on Semantic Web and Linked Data, E101-D(1), 73-81.

Suzuki, M., Matsuda, K., Sekine, S., Okazaki, N. & Inui, K. (2016). Multiple Labeling of Extended Named Entities for Wikipedia Articles. 22nd Annual Conference of the Association for Natural Language Processing (NLP2016), March 2016.

Savchuk, S. O., Arkhangelsky, T. A., Bonch-Osmolovskaya, A. A., Donina, O. V., Kuznetsova, Yu. N., Lyashevskaya, O. N., Orekhov, B. V., & Podryadchikova, M. V. (2024). National Corpus of the Russian Language 2.0: New Possibilities and Development Prospects. *Voprosy yazykoznaniya*, 2, 7–34.