



ALEKSANDRA JAGODA ZIĘBA

Uniwersytet Śląski w Katowicach



<https://orcid.org/0000-0001-6897-5777>

Aplikacja Treq w procesie ustalania rosyjsko-hebrajskich par przekładowych (na przykładzie potocznych nazw medycznych)

THE TREQ APPLICATION IN THE PROCESS OF ESTABLISHING RUSSIAN-HEBREW TRANSLATION PAIRS
(BASED ON THE EXAMPLE OF COLLOQUIAL MEDICAL NAMES)

Summary: This article describes the possibilities that the Treq application provides in the process of translating Russian colloquial medical names into Hebrew. The author presents the tool's available functions that make it easier to find the adequate equivalent and also provides information about the Russian and Hebrew collections included in the InterCorp parallel corpus. The article analyzes specific examples of the translation of Russian colloquial names of diseases into Hebrew using the English language as an intermediary language. Preliminary reference to printed and electronic resources made it possible to establish a potentially correct equivalent, whose adequacy should be confirmed by the application. The results of the analysis made it possible to determine the advantages and disadvantages of the application, and the found translations can serve as the basis for more thorough research in the field of identifying translation pairs of the everyday medical vocabulary of the analyzed languages.

Keywords: corpus linguistics, Treq, Russian, Hebrew, colloquial medical names

ПРИЛОЖЕНИЕ TREQ В ПРОЦЕССЕ УСТАНОВЛЕНИЯ РУССКО-ИВРИТСКИХ ПЕРЕВОДНЫХ ПАР
(НА ПРИМЕРЕ РАЗГОВОРНЫХ ОПРЕДЕЛЕНИЙ БОЛЕЗНЕЙ)

Резюме: В настоящей статье описываются возможности, какие предоставляет приложение Treq в процессе перевода русских нетерминологических названий болезней на иврит. Автор представляет доступные функции инструмента, облегчающие поиск нужного переводного эквивалента, а также приводится информация о русских и ивритских коллекциях, содержащихся в параллельном корпусе InterCorp. В статье были проанализированы конкретные примеры перевода русских разговорных определений болезней на иврит при помощи английского языка, как языка-посредника. Предварительное обращение к печатным и электронным ресурсам позволило установить потенциально правильный эквивалент, который программа должны была удостоверить. Результаты анализа позволили определить преимущества и недостатки приложения, а найденные трансляты могут послужить основой для проведения более тщательных исследований в области выявления переводных пар обиходной медицинской лексики анализируемых языков.

Ключевые слова: корпусная лингвистика, Treq, русский язык, иврит, нетерминологические названия болезней

Narzędzia aplikacji Treq

Aplikacja Treq jest narzędziem służącym do wyszukiwania odpowiedników tłumaczeniowych, z którego można korzystać od września 2014 roku. Przedstawiona baza tekstowa jest powiązana z zasobami Czeskiego Narodowego Korpusu InterCorp, przez który należy rozumieć obszerny i stale ewoluujący synchroniczny korpus równoległy, obejmujący teksty w języku czeskim i trzydziestu ośmiu innych językach. Pomysł projektu zrodził się w Pradze na Uniwersytecie Karola w 1994 roku i jest rokrocznie aktualizowany, czyli wzbogacany o nowe zbiory tekstowe. Główny trzon korpusu, jakim są zasoby literatury pięknej, uzupełniają teksty prawne z korpusu Acquis Communautaire, teksty publicystyczne ze strony Project Syndicate, sprawozdania z obrad Parlamentu Europejskiego Euro-parl, teksty publicystyczne i wiadomości ze stron internetowych VoxEurop, napisy filmowe z bazy OpenSubtitles oraz przekłady Biblii¹. Obecnie aplikacja korzysta już z czternastej wersji korpusu InterCorp, co znacznie zwiększa prawdopodobieństwo ekscerpji prawidłowego ekwiwalentu ze względu nie tylko na wzrost liczby dostępnych zasobów tekstowych, ale także na coraz większe zróżnicowanie gatunkowe dodawanych źródeł. Należy podkreślić, że wymienione kolekcje nie rejestrują tekstów we wszystkich językach, dlatego warunki dla każdej oddzielnej leksykografii przekładowej i pracy tłumacza reprezentują inny stopień przydatności.

Podstawowym celem analizowanej aplikacji jest wyszukiwanie odpowiedników tłumaczeniowych dla języka czeskiego, a przed aktualizacją użytkownik był ograniczony wyłącznie do pracy nad jednowyrazowymi jednostkami leksykalnymi (lemmy lub formy). Znaczące zmiany przyniosła wersja 2.1 aplikacji Treq, w której możliwe stało się dodatkowe wyszukiwanie dostępnych translatów² i synonimów dla języka angielskiego i hiszpańskiego, porównywanie jednostek wielowyzrazowych (*Multiword*) oraz zastosowanie wyrażeń regularnych (*Regex*). Ponadto narzędzie zostało wzbogacone o instrument, który pozwala ustalić wszystkie istniejące formy wyrazowe danego leksemu. Przedstawiony opis byłby niekompletny, gdyby nie uzupełnić go informacją, że aplikacja posiada również takie funkcje, jak parametr ($A=a$), dzięki któremu można wprowadzić zapytanie bez rozróżniania wielkich i małych liter, czy opcję *Restrict*

¹ E. Kaczmarek, A. Rosen, *Praktyczny przewodnik po korpusie równoległym InterCorp*, w: *Praktyczny przewodnik po korpusach języków słowiańskich*, red. M. Hebal-Jeziarska, Wydawnictwo Wydziału Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014, s. 208.

² Przez translat należy rozumieć jednostkę języka docelowego (odpowiednik przekładowy albo ekwiwalent przekładowy).

to, zawężając wyszukiwanie do określonych kolekcji. Poszczególne części korpusu różnią się nie tylko zasobami i rodzajami tekstów, ale także sposobami wiązania, spośród których wyróżniamy ręczne i automatyczne. Ograniczenie wyboru tekstów korpusowych będzie skutkowało mniejszą wiarygodnością uzyskanych danych w przypadku analizy identycznych zbiorów tekstowych.

Twórcy aplikacji założyli, że użytkownikami mogą być nie tylko specjaliści (tłumacze, leksykografowie), ale też osoby, które posiadają przeciętną znajomość dowolnego języka interfejsu aplikacji. Dzięki swojej prostocie aplikacja szybko zyskuje popularność wśród odbiorców, a sam proces wyszukiwania odpowiedników jest możliwy bez wcześniejszej rejestracji i dostępny przez przeglądarkę internetową (<https://treq.korpus.cz/>)³. Po wybraniu języka wyjściowego i docelowego należy w odpowiedniej kolumnie wpisać jednostkę leksykalną, a następnie kliknąć przycisk *Search*.



Wynikiem zapytania będą wszystkie dostępne ekwiwalenty tłumaczeniowe wybranego słowa na język docelowy, które zostaną automatycznie posortowane według liczby wystąpień z uwzględnieniem udziału procentowego w całkowitej objętości korpusu InterCorp. Autorzy i niektórzy lingwiści, badający użyteczność narzędzia w procesach tłumaczeniowych, zgodnie doszli do wniosku, że aplikacja działa skuteczniej w przypadku wyszukiwania przekładów jednowyrazowych jednostek języka wyjściowego. W zdecydowanej większości Treq oferuje wiarygodne tłumaczenie, a badanie kontekstów ich użycia tylko pomaga upewnić się w poprawności znalezionej odpowiedzi. Ekwiwalenty oferowane przez aplikację można uznać za potencjalne odpowiedniki słownikowe poprzez wygenerowanie przez narzędzie listy jednostek „pretendujących” do bycia prawidłowymi tłumaczeniami⁴. Treq

³ M. Škrabal, M. Vavřín, *Databáze překladových ekvivalentů Treq*, „Časopis pro moderní filologii” 2017, 2 (99), s. 245–260.

⁴ M. Škrabal, M. Vavřín, *The Translation Equivalents Database (Treq) as a Lexicographer's Aid*, in: *Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 Conference*, Lexical Computing Ltd, Brno 2017, s. 130.

nie podaje gotowych rozwiązań translatorskich, a jedynie generuje listę słownikową, której dokładna analiza pozwala wyzyskać potencjalnie prawidłowe ekwiwalenty przekładowe. Sytuacja staje się znacznie bardziej skomplikowana w przypadku przekładu jednostek wielowyrzowych, w których przedstawiony materiał korpusowy zawiera sporą liczbę błędów. W takim wypadku lepiej wykorzystać wspomnianą już funkcję *RegEx*, dzięki której możemy tworzyć bardziej złożone zapytania i tym samym uzyskać dokładniejszy przekład wybranego przez nas leksemu.

Autorzy narzędzia przekonują, że brak weryfikacji nadesłanych tłumaczeń nie przeszkadza w uznaniu odpowiedników o największej frekwencji występowania za możliwie wiarygodne. Zawarte w aplikacji wielojęzyczne bazy danych zostały sporządzone na podstawie analizy kwantytatywno-kwalitatywnej, a teksty oryginalne i ich przekłady zostały powiązane automatycznie na poziomie wyrazów za pomocą programu GIZA++⁵. Podczas przygotowania zbioru źródeł dla aplikacji wybierane są tylko zdania związane metodą 1:1, a w starszych wersjach programu wykorzystywano metodę przecięcia, co stwarzało znacznie większe prawdopodobieństwo wystąpienia potencjalnych błędów w wyniku automatycznego dopasowywania niewłaściwych ekwiwalentów jednostki wyjściowej. Ponadto tłumaczenie maszynowe GIZA++ niesie z sobą ryzyko uzyskania przypadkowych wyników analizy oraz uznania otrzymanych ekwiwalentów za potencjalnie poprawne (i wymagające dodatkowej weryfikacji w innych źródłach), a nie w pełni wiarygodne. Prostota i dostępność narzędzia również stwarzają istotne przeszkody — nie ma możliwości przeprowadzenia zaawansowanej analizy wieloczynnikowej, mimo że użytkownik może sprawdzić konteksty tłumaczeniowe wybranego słowa i sprawdzić liczbę jego wystąpień w *InterCorp*.

Zasoby rosyjskiego i hebrajskiego korpusu Przedmiot analizy

Pracując z narzędziem *Treq*, należy zawsze brać pod uwagę fakt, że moduły poszczególnych języków, zawierające się w równoległym korpusie, znacznie różnią się od siebie liczbą tekstów i określonymi kolekcjami. Zróżnicowane warunki wyszukiwania translatów są mocno widoczne w przypadku języków rosyjskiego

⁵ E. Kaczmarek, *Metody ustalania ekwiwalentów czasowników wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*, Wydawnictwo Wydziału Polonistyki Uniwersytetu Warszawskiego, Warszawa 2019, s. 48.

i hebrajskiego. Czternasta wersja rosyjskiego korpusu InterCorp zawiera ponad 21 mln słów, a hebrajska — ok. 16 mln jednostek leksykalnych i wyrażen. Jak można zauważyć, różnica w zasobach wymienionych korpusów nie jest olbrzymia. Jednak korpus hebrajski zawiera wyłącznie teksty z napisów filmowych, w których najczęściej rejestrowane są błędy tłumaczeniowe, korpus rosyjski składa się zaś z czterech kolekcji (OpenSubtitles, Project Syndicate, literatura piękna i przekłady Biblii).

Na podstawie przytoczonych danych użytkownik może wysnuć błędny wniosek, że aplikacja ta nie jest odpowiednim narzędziem do tłumaczenia rosyjsko-hebrajskiego lub hebrajsko-rosyjskiego ze względu na brak zróżnicowania gatunkowego w korpusie hebrajskim. Zupełnie różne zbiory korpusów poszczególnych języków utrudniają znalezienie prawidłowych odpowiedników. Dodatkowe wątpliwości może budzić obecność tylko trzech języków zapytań i brak możliwości bezpośredniego tłumaczenia rosyjsko-hebrajskiego i hebrajsko-rosyjskiego. Należy również pamiętać, że platforma OpenSubtitles dostarcza spory zasób materiału językowego, jednak praktycznie niezawierającego przez zawodowych tłumaczy. Dostępne w tej kolekcji napisy filmowe nie powstawały zgodnie z określonymi kryteriami listy dialogowej, z którą mamy do czynienia w przypadku napisów profesjonalnych. Jednak przy dużej ostrożności badacza możliwe jest odnalezienie jedno- i wielowrazowych jednostek mowy potocznej, które często są pomijane we współczesnych słownikach papierowych⁶. Ze względu na spore różnice w zasobach kolekcji wymienionych języków i dostępność wyłącznie platformy nieprofesjonalnych napisów filmowych dla języka hebrajskiego nieuniknione wydaje się traktowanie uzyskanych poświadczeń jako potencjalnie akceptowalnych, a nie trafnych. Celem krótkiej analizy będzie próba wykorzystania aplikacji do wyszukania potencjalnych i rzadkich ekwiwalentów przekładowych dla tłumaczenia rosyjsko-hebrajskiego. Obiektem badania będą potoczne określenia terminów medycznych (np. nazwy jednostek chorobowych, towarzyszących im symptomów czy stanu fizyczno-psychicznego człowieka), które rzadko odnotowują dostępne słowniki rosyjsko-hebrajskie bądź podają nieprawidłowy przekład. Podjęta próba uwiarygodnienia odnalezionych ekwiwalentów ze źródeł internetowych za pomocą aplikacji może posłużyć za punkt wyjścia do dalszych poszukiwań wiarygodnych translatów w tym obszarze i umieszczenia ich w odpowiednich pracach leksykograficznych.

⁶ A. Charciarek, *Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równoległym InterCorp*, „Studia Rossica Posnaniensia” 2021, vol. 46/1, s. 42.

Potoczna leksyka medyczna

Zagadnienia związane z szeroko pojętym zdrowiem, diagnostyką chorób i ich leczeniem są nieodłącznym elementem ludzkiej codzienności, a współczesne osiągnięcia w dziedzinie nauki pozwalają rozwijać się medycynie w bardzo szybkim tempie i tym samym tworzyć okazały repertuar słownictwa medycznego. Z fachową terminologią medyczną mamy do czynienia głównie w tekstach specjalistycznych i publicystycznych. Z kolei słownictwo niefachowe tej kategorii zajmuje istotne miejsce w języku mówionym w każdym z języków, cechując się wysoką frekwencją użycia w sytuacjach codziennych. Mając do dyspozycji ogromną i zróżnicowaną bazę leksykalną, językoznawcy nie pozostali obojętni w obliczu możliwości badań i analiz, co zaowocowało licznymi pracami naukowymi z zakresu leksykografii przekładowej, etymologii czy onomastyki (A.B. Юдин, Е.К. Акимова, В.А. Меркулова). We współczesnej lingwistyce obserwuje się próby wyodrębnienia leksyki medycznej jako zróżnicowanej gałęzi terminologii typologicznej (ros. *типологическое терминоведение*) opartej na specjalistycznych materiałach i koncepcjach wymienionej dziedziny⁷. Naturalne jest, że opis najbardziej rozpowszechnionych chorób i towarzyszących im symptomów, a także metody ich leczenia były od zawsze obecne w medycynie ludowej, skąd do języka ogólnego przeniknęło wiele potocznych określeń. Z reguły nazwy te są synonimami terminologii fachowej i są najczęściej używane w mowie potocznej, publicystyce lub Internecie, np. *анемия* — *малокровие, гепатит* — *желтуха*⁸.

Гусиная кожа

W dalszej części artykułu zaprezentujemy sposób wyzyskania za pomocą aplikacji Treq potencjalnie wiarygodnych hebrajskich tłumaczeń wybranych potocznych rosyjskich nazw dla określonych terminów medycznych. Dla ilustracji wybraliśmy jednostki, dla których w analizie wstępnej udało się uzyskać hebrajski ekwiwalent przekładowy przy użyciu narzędzia i których przekład został zweryfikowany w innych źródłach systemowych. Ze względu na brak możliwości bezpośredniego tłumaczenia rosyjsko-hebrajskiego językiem pośrednim będzie jeden z trzech głównych języków interfejsu, tj. język angielski. W pierwszym przykładzie

⁷ С.Г. Казарина, *Типологическое терминоведение как дифференцированная лингвистическая дисциплина*, «Вестник АГУ» 2012, № 297, s. 192–195.

⁸ Д.С. Иванова, *Нетерминологические названия болезней как объект исследования: к постановке проблемы*, «Вестник АГУ» 2017, № 202, s. 32–36.

posłużymy się potoczną rosyjską nazwą reakcji pilomotorycznej, czyli leksemem *гусиная кожа*. W celu uniknięcia możliwych dylematów znaczeniowych przy tłumaczeniu przytoczmy dokładną definicję analizowanego określenia reakcji skórnej: „гусиная кожа — небольшие пупырышки у основания мелкого волосяного покрова на коже человека, непроизвольно возникающие в случаях, когда организму холодно или же человек испытывает сильное эмоциональное возбуждение, например страх”⁹. Warto zaznaczyć, że fachowa nazwa tego zjawiska, tj. *полиэрекция*, nie jest odnotowana w Narodowym Korpusie Języka Rosyjskiego, a w przypadku przedstawionego translatu korpus rejestruje 24 poświadczenia. Zanim przystąpimy do analizy, odwołamy się do drukowanych słowników przekładowych i zasobów elektronicznych, czyli podstawowych źródeł systemowych, w celu ustalenia potencjalnych tłumaczeń angielskich i hebrajskich. Termin „źródła systemowe”, sformułowany przez Wojciecha Chlebę, należy rozumieć jako bezkontekstowe zbiory jednostek języka¹⁰. *Полный англо-русский и русско-английский словарь* V.R. Mullera dla rosyjskiego wyrażenia *гусиная кожа* podaje ekwiwalent *goose flesh*. Dodatkowa weryfikacja przekładu na stronie Wolny, wielojęzyczny Wikisłownik potwierdza użycie tego translatu, jednak wyższą frekwencją cechuje się równoznaczne wyrażenie *goose bumps* (dosł. ‘gęsie guzki’). Z kolei elektroniczny słownik GLOSBE podaje następujące tłumaczenie hebrajskie: עור ברזז (*or barwaz*). W danym przypadku mamy do czynienia z zachowaniem rzeczownika wyjściowego (hebr. *or* — ‘skóra’) i jednoczesną zmianą przymiotnika dzierżawczego (hebr. *barwaz* — ‘kaczka’). Po ustaleniu możliwych ekwiwalentów spróbujmy teraz znaleźć odpowiednie tłumaczenie przy użyciu aplikacji. Treq dla zapytania *гусиная кожа* z zaznaczeniem funkcji *Multiword* i parametru ($A=a$) przynosi następujące rezultaty:

Tabela 1. Treq dla zapytania *гусиная кожа*

Liczba poświadczeń (Frequency)	Udział procentowy (Proportion)	Rosyjski transland ¹¹ (Russian)	Angielski ekwiwalent (English)
1	100,0	гусиная кожа	bumps

⁹ A. Спектор, *Большой иллюстрированный атлас анатомии человека*, Издательство АСТ, Москва 2016, s. 31.

¹⁰ W. Chlebda, *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2011, s. 26.

¹¹ Transland — jednostka języka wyjściowego.

W danym przypadku na pierwszy plan wysuwają się problemy związane z pojedynczym wystąpieniem wyrażenia w korpusie i dodatkową koniecznością zweryfikowania kontekstu użycia wyszukanego ekwiwalentu. Jak się okazuje, poświadczenie translatu pochodzi z platformy OpenSubtitles i przedstawia się następująco:

1. *Найдите нам сирены Анны Франк. Это пугает. Это хорошо. Гусиня кожа.*

2. *Find us the Anne Frank sirens. That's chilling. That's good. That's giving me goose bumps.*

Przekład angielski, składniowo nieco bardziej rozbudowany, pozwala potwierdzić użycie znalezionej uprzednio ekwiwalentu pomimo otrzymania niepełnego przekładu przez aplikację. Korzystając z proponowanego tłumaczenia, spróbujmy sprawdzić, czy narzędzie przytoczy potencjalnie wiarygodne tłumaczenie translantu na język hebrajski:

Tabela 2. Treq dla zapytania *goose bumps*¹²

Liczba poświadczeń (Frequency)	Udział procentowy (Proportion)	Angielski translant (English)	Hebrajski ekwiwalent (Hebrew)
2	22,2	goose bumps	עור ברווז
2	22,2	goose bumps	ברווז
1	11,1	goose bumps	צמרמורת
1	11,1	goose bumps	פרפרים

Zaproponowany sposób tłumaczenia rosyjskiego wyrażenia *гусиня кожа* w danym przypadku okazał się prawidłowy. Jednak przeprowadzenie dokładnej i wiarygodnej analizy przy tak niewielkiej liczbie poświadczeń jest naturalnie niemożliwe, choć wyniki wyszukiwania mogą nakierowywać tłumacza na potencjalnie właściwy ekwiwalent przy podjęciu dalszych badań. Pierwszy translant reprezentuje sobą pełny przekład jednostki wyjściowej, a drugi, o identycznej liczbie poświadczeń, zaledwie część ekwiwalentu. Aplikacja zaproponowała również nazwę innej reakcji skórnej (צמרמורת) *'cmarmoret'* — pol. 'dreszcze', co może wynikać z błędnego przeświadczenia, że nazwy te można stosować synonimicznie. Z kolei weryfikacja ostatniego poświadczenia i kontekstu jego użycia w InterCorp pokazała, że wyrażenie *goose bumps* zostało przetłumaczone w platformie OpenSubtitles jako *motylki*

¹² Ze względu na objętość artykułu reprezentowana jest skrócona lista poświadczeń korpusowych.

w brzuchu (*parparim be-beten*). W tym przypadku aplikacja podała odpowiednik o zupełnie odmiennym znaczeniu, którego użyjemy w innej sytuacji komunikacyjnej (określenie ekscytacji przy zakochaniu). Jednakże tego rodzaju błędy umożliwiają niekiedy wyzyskanie takich form wyrazowych w języku obcym, których próżno szukać w dostępnych słownikach przekładowych. Oferowany przez platformę SketchEngine paralelny korpus rosyjsko-hebrajski potwierdził użycie wyrażenia עור ברזוו (or barwaz) jako potencjalnego ekwiwalentu dla idiomu *зусиная кожа*. Należy jednak nadmienić, że zmiana kierunku tłumaczenia wykazała, iż hebrajski wielowyrazowiec częściej jest tłumaczony na język rosyjski jako synonim omawianej reakcji pilomotorycznej, tj. leksem *мурашки*.

Лихорадка

Kolejnym przykładem potocznego określenia, szeroko używanym zarówno w mowie potocznej Rosjan, jak i literaturze pięknej, jest jednowyrazowiec *лихорадка*. W języku rosyjskim pod danym wieloznacznym leksemem należy rozumieć objaw chorobowy, jakim jest podwyższona temperatura ciała, a także stan niespokojnego pobudzenia emocjonalnego. W analizie nasza uwaga będzie skupiona wyłącznie na poświadczeniach, które rejestrują znaczenie medyczne danego leksemu. Samo słowo pochodzi od czasownika *лихорадить* (złożenie *лихо* 'zło' i *радить* 'życzyć', dosł. 'życzyć zła'), a nierzadko służy również określeniu chorób, których głównym towarzyszącym symptomem jest gorączka (np. *геморрагическая лихорадка*, *геморрагическая лихорадка Марбург*, *жёлтая лихорадка* itd.). W tym miejscu analizy postaramy się uzyskać prawidłowy ekwiwalent hebrajski bez wcześniejszego odniesienia się do zasobów papierowych i elektronicznych. Treq dla zapytania *лихорадка* w przekładzie rosyjsko-angielskim z zaznaczeniem funkcji *lemmaty* (umożliwiającej wyszukanie wszystkich form fleksyjnych danego leksemu) przynosi następujące informacje:

Tabela 3. Treq dla zapytania *лихорадка*

Liczba poświadczeń (Frequency)	Udział procentowy (Proportion)	Rosyjski transland (Russian)	Angielski ekwiwalent (English)
83	66,9	лихорадка	fever
13	10,5	лихорадка	Ebola
7	5,6	лихорадка	dengue

Uzyskane poświadczenia pozwalają wyodrębnić pierwszy ekwiwalent i uznać go za potencjalnie prawidłowy ze względu na najwyższą liczbę wystąpień w korpusie InterCorp. Należy zaznaczyć, że w procesie weryfikacji angielskiego przekładu okazało się, że w 41 poświadczeniach wyraz *fever* występuje w zespole z innymi komponentami (np. *hay fever*, *yellow fever*, *dengue fever*), tworząc tym samym nazwy innych jednostek chorobowych. Drugi translac można traktować jako rezultat tłumaczenia metonimicznego, gdyż danej chorobie zakaźnej (wywołanej wirusem o tej samej nazwie) zawsze towarzyszy gorączka. Zaproponowany translac z najwyższą frekwencją użycia w tekstach umożliwia kontynuowanie analizy i wpisanie zapytania *fever* dla tłumaczenia angielsko-hebrajskiego:

Tabela 4. Treq dla zapytania *fever*

Liczba poświadczeń (Frequency)	Udział procentowy (Proportion)	Angielski translac (English)	Hebrajski ekwiwalent (Hebrew)
36	41,4	fever	חום
12	13,8	fever	החום
11	12,6	fever	קדחת

Analiza danych korpusowych wskazuje, że w przypadku hebrajskich odpowiedników angielskiego *fever* najczęściej używany jest rzeczownik חום (*chom*), a zaraz po nim rekomendowany jest ten sam wyraz, któremu towarzyszy przedimek określony (*ha-chom*). Wiarygodność przytoczonych poświadczeń należy teraz potwierdzić innymi źródłami systemowymi. *Англо-русский и русско-английский медицинский словарь* A. Ju. Bołotina i Je. O. A. Jakuszewa wskazuje jako ewentualne ekwiwalenty słowa *лихорадка* leksemy *fever* i *febris*, co potwierdza odnotowanie przez aplikację prawidłowego przekładu, przy czym druga propozycja określa termin łaciński. Z kolei oksfordzki angielsko-hebrajski i hebrajsko-angielski słownik autorstwa Ja'akowa Lewi'ego proponuje rzeczowniki חום (*chom*, przytoczony przez narzędzie) i קדחת (*kadachat*), z których ostatni winien być używany w odniesieniu do malarii lub febry. Taki sam translac (bez innych ekwiwalentów) jest odnotowany w słowniku *Русско-ивритский иврит-русский словарь*, co potwierdza wcześniejszą obserwację, że dostępne rosyjsko-hebrajskie prace leksykograficzne nie do końca trafnie rejestrują przekład potocznej leksyki medycznej. Uzyskane z platformy SketchEngine poświadczenia oraz analiza kontekstu użycia wy-

zyskanych leksemów potwierdziły ich zastosowanie przy nazewnictwie gorączki (174 poświadczenia) i choroby tropikalnej (155 poświadczeń). Uwiarygodnienie znaczeń ekscerpowanych poświadczeń na podstawie danych słownikowych jest niezbędnym elementem pracy z aplikacją Treq, jednak poprzedzająca analizę weryfikacja słownikowa nie zawsze jest konieczna. Może ona nastąpić dopiero po uzyskaniu danych korpusowych, a poprawność pozyskanej pary przekładowej powinna być sprawdzona bez względu na poziom kompetencji językowej tłumacza.

Ветрянка

Ostatnia część analizy będzie poświęcona wyzyskaniu hebrajskiego translatu dla potocznego określenia ospy wietrznej w języku rosyjskim, w tym przypadku z wcześniejszym uwzględnieniem źródeł systemowych. Fachowa nazwa *ветряная оспа* została zastąpiona w mowie potocznej jednowyrazowcem *ветрянка* i zgodnie z danymi statystycznymi Narodowego Korpusu Języka Rosyjskiego charakteryzuje się znacznie wyższą frekwencją niż termin profesjonalny (*ветрянка* — 170 poświadczeń, *ветряная оспа* — 41 poświadczeń). Analiza kontekstów użycia pozwoliła dostrzec fakt wieloznaczności leksemu, gdyż słowo *ветрянка* służy także jako określenie wiatraka, co zostało uwzględnione przy selekcji konkordancji. Dla danej choroby zakaźnej charakterystyczne jest występowanie wysypki, początkowo przybierającej formę grudek, które ulegają przemianom w pęcherzyki i bolesne krosty. Elektroniczny słownik GLOSBE podaje cztery angielskie odpowiedniki: 1) *chicken pox*, 2) *chickenpox*, 3) *chicken-pox*, 4) *varicella*. W przypadku czwartego terminu mamy do czynienia z łańską nazwą jednostki chorobowej, co poświadczają słowniki medyczne. Z kolei pozostałe wyniki reprezentują różną pisownię tego samego wyrażenia. Weryfikacja przykładów użycia danej formy wyrazowej potwierdziła, że najwyższą frekwencją występowania w tekstach cechuje się drugi wyzyskany translant, tj. *chickenpox*. Wspomniany już *Полный англо-русский и русско-английский словарь V.R. Mullera* również rekomenduje użycie danego ekwiwalentu przekładowego. Etymologia tej jednostki chorobowej nie jest do końca wyjaśniona, chociaż istnieją hipotezy mówiące o podobieństwie pęcherzyków do ciecierzycy (ang. *chickpea*) lub do ugryzień kurczaka (ang. *chicken*)¹³. Przytoczony oksfordzki słownik angielsko-hebrajski proponuje użycie terminu *אבעבועות רוח* (*awabu'ot ru'ach*), co dostarcza pełnej informacji słownikowej i pozwala dokonać analizy, której wyniki przedstawiają się następująco:

¹³ T. Shors, *Understanding Viruses*, Jones & Bartlett Learning, Burlington 2017, s. 459.

Tabela 5. Treq dla zapytania *ветрянка*

Liczba poświadczeń (Frequency)	Udział procentowy (Proportion)	Rosyjski transland (Russian)	Angielski ekwiwalent (English)
7	87,5	ветрянка	pox
1	12,5	ветрянка	chickenpox

Weryfikacja kontekstów użycia w korpusie InterCorp pozwala wyjaśnić, dlaczego znaleziony uprzednio odpowiednik posiada tylko pojedyncze poświadczenie, w przeciwieństwie do ekwiwalentu *pox*, który reprezentuje jedynie drugą część odnotowanego wcześniej tłumaczenia. Wszystkie przykłady zastosowania danej jednostki pochodzą z platformy OpenSubtitles, w której tłumacze amatorzy konsekwentnie stosują zasadę oddzielnej pisowni analizowanej potocznej nazwy jednostki chorobowej. W takim przypadku aplikacja ze względu na automatyczną analizę korpusową pomija komponent *chicken*, co może doprowadzić do późniejszych nieścisłości znaczeniowych, gdyż część *pox* w języku angielskim jest używana także w nazwie ospy prawdziwej (ang. *smallpox*). Posługując się ekwiwalentem *chickenpox*, przystępujemy do tłumaczenia angielsko-hebrajskiego:

Tabela 6. Treq dla zapytania *chickenpox*

Liczba poświadczeń (Frequency)	Udział procentowy (Proportion)	Angielski transland (English)	Hebrajski ekwiwalent (Hebrew)
2	40,0	chickenpox	אבעבועות
1	20,0	chickenpox	באבעבועות
1	20,0	chickenpox	נמרה
1	20,0	chickenpox	אבעבועות

Tłumaczenie ponownie zawiera błąd, związany z pominięciem przez aplikację drugiego członu *רוח*, który jednak widnieje w kontekstach użycia i pozwala prawidłowo rozróżnić rodzaj ospy w tłumaczeniu. Ekwiwalent drugi (*be-awabu'ot*) należy rozumieć jako użycie rzeczownika wraz z wieloznacznym przyimkiem *ב*, który w tej sytuacji wskazuje na zastosowanie formy narzędnika z czasownikiem. Z kolei przykład trzeci nie ma nic wspólnego z poszukiwanym przez nas odpowiednikiem wyrazu *ветрянка* i pochodzi od bezokolicznika *להימרה* (*lechimara* — 'być pomazanym'). Wspomniany już paralelny korpus rosyjsko-hebrajski platformy Sketch-

Engine zarejestrował w tekstach 94 poświadczenia użycia słowa *ветрянка*, a w 78 przekładach znalazł się ekwiwalent zaproponowany przez analizowaną aplikację (tj. אַבערױת — *awabu'ot*). W obu przypadkach udało się za pomocą narzędzia Treq potwierdzić użycie potencjalnie prawidłowego ekwiwalentu przekładowego. Należy pamiętać, że każdorazowa praca z aplikacją wymaga uprzedniego zwrócenia się do sprawdzonych słowników przekładowych, w szczególności w wypadku tłumaczenia nieprofesjonalnej terminologii. Jeśli słowniki jedno- i wielojęzyczne nie odnotowują wybranych przez nas leksemów, pozostaje korzystać z poświadczeń korpusowych jako źródeł potencjalnie prawidłowych translatów.

Wnioski i dalsze perspektywy badawcze

Analiza wybranych jednostek leksykalnych z zakresu medycyny udowodniła, że mimo dysproporcji dotyczących dostępnych kolekcji badanych przez nas języków docelowo udało się wyodrębnić znalezione uprzednio pary przekładowe, które mogą posłużyć za punkt odniesienia do dalszych rozważań. Przesadą byłoby stwierdzenie, że w przypadku tłumaczenia potocznych nazw jednostek chorobowych aplikacja zawsze wykazuje się wysokim stopniem przydatności dla przekładu rosyjsko-hebrajskiego. Poza przedstawionymi badaniami dokonano próby odnalezienia hebrajskich ekwiwalentów dla 50 innych potocznych nazw medycznych, pozostając przy języku angielskim jako pośrednim w procesie przekładu. Potencjalnie prawidłowy translata udało się odnaleźć w przypadku następujących jednostek: *близорукость, ожирение, ожог, понос, рахит, проказа, рак, столбняк, чума*, co stanowi niecałe 5% analizowanych jedno- i wielowyrazowców. Platforma OpenSubtitles, która rejestruje przekłady napisów filmowych, tylko w niewielkim stopniu zakłada spełnienie kryterium sytuacyjności i dlatego nie należy oczekiwać, że odnotuje ona ekwiwalent docelowy o jednakowym stopniu potoczności. Automatyczna analiza korpusowa zmusza również użytkownika do skrupulatnego zbadania odnalezionych przez narzędzie jednostek i sprawdzenia kontekstów użycia, gdyż wyniki nierzadko uwzględniają tylko część poszukiwanego translata, co może wpływać na pochopną ocenę i przekonanie o nieprzydatności aplikacji.

Niewątpliwą zaletą narzędzia Treq jest jego nieskomplikowany interfejs użytkownika, a cały proces generowania listy słownikowej przebiega szybko i sprawnie, przy czym należy zaznaczyć, że znajomość jednego z trzech podstawowych języków aplikacji jest

konieczna w celu zidentyfikowania występujących błędów. Niekiedy obszerna lista odpowiedników tekstowych pozwala odnaleźć nie tylko synonimy dla badanej jednostki, ale także (zupełnie przypadkowo) inne ekwiwalenty używane w analogicznej sytuacji komunikacyjnej, których daremnie szukać w dostępnych słownikach przekładowych. Hebrajskojęzyczne zasoby korpusowe niewątpliwie wymagają powiększenia liczby dostępnych tekstów i zróżnicowania swoich kolekcji, co znacznie ułatwi przyszłe badania w obszarze przekładu rosyjsko-hebrajskiego i wpłynie na jego reprezentatywność. Niewielka liczba poświadczeń z jednej strony zawęży możliwości wyboru translatu, co nie zmusza użytkownika do szczegółowego przeanalizowania proponowanych tłumaczeń, a z drugiej uniemożliwia pozyskanie synonimicznych odpowiedników dla wybranego translądu.

Na uwzględnienie zasługuje również fakt, że skuteczna praca z zaprezentowanym narzędziem wymaga w pierwszej kolejności dokładnego zapoznania się z instrumentami aplikacji i pozyskania umiejętności budowy zapytań¹⁴. Niekiedy pominięcie jednej funkcji może pozbawić użytkownika dostępu do szerokiego repertuaru potencjalnych odpowiedników, a nawet doprowadzić do nieodnotowania przez aplikację żadnego ekwiwalentu. Praca na materiale języka wyjściowego i ustalenie postaci kształtowej¹⁵ wybranej przez nas jednostki jest elementem niezbędnym przy pracy z narzędziami korpusowymi i dla prawidłowego procesu tłumaczenia. W tym przypadku najbardziej przydatne okażą się słowniki jednojęzyczne, w których granice weryfikatora są już odnotowane. Bardziej precyzyjny dobór jednostek wyjściowych niewątpliwie umożliwi wprowadzenie prawidłowego zapytania i przeprowadzenie analizy bez początkowych błędów użytkownika, które w konsekwencji mogą doprowadzić do niewłaściwych wniosków.

Bibliografia

- Charciarek A., *Korpus równoległy InterCorp w leksykografii przekładowej — możliwości i ograniczenia*, „Roczniki Humanistyczne” 2019, t. 67, z. 7, s. 79–92.
- Charciarek A., *Napisy filmowe jako źródło ekwiwalentów przekładowych w polsko-czeskim korpusie równoległym InterCorp*, „Studia Rossica Posnaniensia” 2021, vol. 46/1, s. 39–49.

¹⁴ M. Hebal-Jeziarska, *Podstawowe zasady korzystania z korpusów przy badaniu języka*, w: *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, red. W. Chlebda, Wydawnictwo Uniwersytetu Opolskiego, Opole 2013, s. 19.

¹⁵ Termin używany za Wojciechem Chlebda. Przez postać kształtową należy rozumieć połączenie komponentów o określonym znaczeniu, składających się na jednostkę leksykalną.

- Charciarek A., *Narzędzie Treq w procesie ustalania polsko-czeskich par przekładowych*, „Bohemistyka” 2020, nr 4, s. 531–550.
- Charciarek A., *Programma dlya ustanovleniya perevodnykh ekvivalentov Treq: preimushchestva i nedostatki*, „Studia Rossica Posnaniensia” 2020, vol. 45/1, s. 121–132 [Charciarek A., *Программа для установления переводных эквивалентов Treq: преимущества и недостатки*, „Studia Rossica Posnaniensia” 2020, vol. 45/1, s. 121–132].
- Charciarek A., *Treq — narzędzie do ustalania ekwiwalentów przekładowych w korpusie równoległym InterCorp*, w: *Słowiński krąg. Słowo — myśl — obraz w tradycji i współczesności*, red. A. Banaszek-Szapotałowa, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2019, s. 293–304.
- Chlebda W., *Na tropach translatów. W poszukiwaniu odpowiedników przekładowych*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2011.
- Hebal-Jeziarska M., *Podstawowe zasady korzystania z korpusów przy badaniu języka*, w: *Na tropach korpusów. W poszukiwaniu optymalnych zbiorów tekstów*, red. W. Chlebda, Wydawnictwo Uniwersytetu Opolskiego, Opole 2013, s. 17–30.
- Ivanova D.S., *Neterminologicheskiye nazvaniya bolezney kak obyekt issledovaniya: k postanovke problemy*, „Vestnik AGU” 2017, no. 202, s. 32–36 [Иванова Д.С., *Нетерминологические названия болезней как объект исследования: к постановке проблемы*, «Вестник АГУ» 2017, № 202, s. 32–36].
- Kaczmarek A., *Metody ustalania ekwiwalentów czasowników wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*, Wydawnictwo Wydziału Polonistyki Uniwersytetu Warszawskiego, Warszawa 2019.
- Kaczmarek E., Rosen A., *Praktyczny przewodnik po korpusie równoległym InterCorp*, w: *Praktyczny przewodnik po korpusach języków słowińskich*, red. M. Hebal-Jeziarska, Wydawnictwo Wydziału Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014, s. 207–231.
- Kazarina S.G., *Tipologicheskoye terminovedeniye kak differentsirovannaya lingvisticheskaya distsiplina*, „Vestnik AGU” 2012, no. 297, s. 192–195 [Казарина С.Г., *Типологическое терминоведение как дифференцированная лингвистическая дисциплина*, «Вестник АГУ» 2012, № 297, s. 192–195].
- Shors T., *Understanding Viruses*, Jones & Bartlett Learning, Burlington 2017.
- Škrabal M., Vavřin M., *Databáze překládových ekvivalentů Treq*, „Časopis pro moderní filologii” 2017, 2 (99), s. 245–260.
- Škrabal M., Vavřin M., *The Translation Equivalents Database (Treq) as a Lexicographer's Aid*, in: *Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 Conference*, Lexical Computing Ltd, Brno 2017, s. 124–137.
- Spektor A., *Bol'shoy illyustrirovannyy atlas anatomii cheloveka*, Izdatel'stvo AST, Moskva 2016 [Спектор А., *Большой иллюстрированный атлас анатомии человека*, Издательство АСТ, Москва 2016].