

ANDRZEJ CHARCIAREK

Программа для установления переводных эквивалентов Treq: преимущества и недостатки

Treq – a tool for searching translation equivalents: benefits and shortcomings

Abstract. The article presents how the Treq corpus tool can be used in translation and lexicography. It also provides a description of Polish, Czech and Russian text resources found in the Interpcorp parallel corpus as well as Treq features. Furthermore, the article provides specific examples illustrating how the Treq application can be used in Polish-Czech and Russian-Czech translation lexicography. The author uses samples of translation including single and multi-word units in order to show benefits and shortcomings of the Treq application.

Keywords: Treq, InterCorp, translation equivalence, Czech, Polish, Russian

Andrzej Charciarek, Uniwersytet Śląski w Katowicach, Katowice – Polska,
andrzej.charciarek@us.edu.pl, ORCID ID: <https://orcid.org/0000-0001-5651-5730>

Чешский национальный корпус (Český národní korpus) был основан в 1994 году на философском факультете Карлова университета в Праге. Программа Treq, названная базой переводных эквивалентов (databáze překladových ekvivalentů), относительно молода – ею можно пользоваться только с 2014 года. Это двусторонние чешско-иноязычные и англо-иноязычные словари, автоматически составленные из текстов параллельного корпуса InterCorp. Язык интерфейса Treq по выбору – чешский или английский.

Так, Treq адресован в первую очередь богемистам и англистам, поскольку одним из сопоставляемых друг с другом языков обязательно должен быть либо чешский, либо английский. Этот факт имеет существенное значение для потенциальных пользователей Treq. Для польского

русиста или русского полониста эта программа не представляет собой большой ценности, но русский или польский богемист/англист находит в Treq вспомогательный инструмент в процессе перевода (польско-чешского, русско-чешского и т. д.). Правда, в зависимости от сопоставляемых друг с другом языков Treq создает переводчику дифференцированные условия для поиска переводных эквивалентов. Это следует из того, что модули отдельных языков, которые содержит параллельный корпус InterCorp, различны по количеству текстов и их жанрам. Примечательно, что приложение Treq пользуется не последней (одиннадцатой), но девятой версией InterCorp 2016 года. Это обусловлено тем, что каждая последующая версия InterCorp отличается от предыдущей количеством текстов; в последних их значительно больше. К примеру, девятая версия русского корпуса объемом в 15 млн. слов, одиннадцатая – свыше 18 млн. слов. Аналогичные данные можно привести относительно и других языков. По сравнению с русским польский InterCorp v9 в несколько раз больше: в нем почти 84 млн. слов.

Следует отметить, что польские, чешские и русские тексты, так называемые коллекции, содержащиеся в InterCorp v9, полностью не совпадают. Русский InterCorp не охватывает текстов из корпусов EuroParl, Acquis Communautaire и Presseurop; он содержит художественные тексты (так называемое ядро), публицистические тексты на сайте Project Syndicate, а также субтитры на сайте OpenSubtitles. Польский InterCorp v9 – один из самых больших по величине иноязычных корпусов и содержит все коллекции InterCorp за исключением публицистических текстов на сайте Project Syndicate.

Вышеприведенные данные однозначно показывают, что для польско-чешского и русско-чешского перевода Treq создает совершенно различные условия. Во-первых, они касаются разного объема польского и русского корпусов (польский почти в шесть раз больше русского), во-вторых, русский InterCorp по сравнению с польским жанрово ограничен. В нем явно не хватает устных текстов, их отсутствие в какой-то мере компенсируют любительские субтитры со всеми их несовершенствами. Преимуществом польского InterCorp являются содержащиеся в нем юридические тексты Евросоюза (в корпусе Acquis Communautaire) и записи заседаний парламента Евросоюза за период 2007–2011 гг. (в корпусе EuroParl). Очевидно, успешность поиска переводных эквивалентов в первую очередь зависит от объема и качества корпусных ресурсов.

Важным фактором, который серьезным образом может повлиять на результативность Treq, является выравнивание текстов по словам, осуществляемое автоматически и практически не пересматриваемое. Этот процесс авторы программы описывают следующим образом:

Originální a překladové texty jsou nejprve na základě statistických výpočtů zarovnány po slovech pomocí programu GIZA++ (Och–Ney 2003). Zarovnané dvojice slov jsou pak seříděny a sumarizovány. Výsledek automatické excerptce není nijak revidován, jako ukazatel relevance překladového ekvivalentu však může posloužit relativní frekvence příslušné dvojice slov. Čím častěji se ekvivalent zadaného slova vyskytl ve srovnání s ostatními ekvivalenty, tím větší je pravděpodobnost, že je funkční (*Treq, Nápo věda*, электронный ресурс).

Из вышесказанного следует, что программа Treq предоставляет пользователям перечень возможных эквивалентов, т. е. приводят статистические данные о возможных переводных эквивалентах. Форма представления результирующих данных – табличная с учетом числа вхождений и их процентных долей в общем объеме, начиная с наиболее частотных единиц и кончая наименее частотными. Поиск переводных эквивалентов может проводиться в пределах всех доступных коллекций (польско-чешских – 5, русско-чешских – 3) или выбранных. Разумеется, следует иметь в виду, что ограничение обзора корпусных ресурсов (функция *omezit na...*) ведет как к снижению числа вхождений, так и к снижению достоверности полученных статистических данных. Иногда недостаточное число вхождений значительно осложняет или даже делает невозможным проведение достоверного анализа. Любое вхождение можно проверить с помощью гипертекстовой ссылки, которая позволяет получить контексты употребления анализируемой единицы.

Вторая версия программы Treq дает возможность искать не только переводные эквиваленты однолексемных, но и многолексемных единиц. Однако в случае поиска эквивалентов многолексемных единиц следует считаться с большей ошибочностью полученных результатов.

Что касается однолексемных единиц, достаточно ввести в соответствующую колонку (чешск. *dotaz*) языковую единицу, для которой ищем переводные эквиваленты. Для получения большего числа вхождений следует использовать дополнительную функцию леммы (чешск. *lemmata*), которая обнаружит все словоформы одной и той же лексемы. К примеру, приведем вхождения русской лексемы *собака* в следующем порядке: число вхождений, процентная доля, русский трансланд, чешский транслат:

без функции *lemmaty*

Lemmata [?]
 Včeslovné [?]
 Regulární [?]
 A = a [?]

▲ Frekvence ▼	▲ Procenta ▼	▲ Ruština ▼	▲ Čeština ▼
256	79.3	собака	pes
40	12.4	собака	psa
6	1.9	собака	čokl
4	1.2	собака	Pes
4	1.2	собака	pejsek
3	0.9	собака	pse
1	0.3	собака	psik
1	0.3	собака	fenka
1	0.3	собака	umfel
1	0.3	собака	Pejsek
1	0.3	собака	hahaha
1	0.3	собака	fenou
1	0.3	собака	zbavuje
1	0.3	собака	pejsku
1	0.3	собака	hroch
1	0.3	собака	Barzoj
323			

с функцией *lemmaty*

Lemmata [?]
 Včeslovné [?]
 Regulární [?]
 A = a [?]

▲ Frekvence ▼	▲ Procenta ▼	▲ Ruština ▼	▲ Čeština ▼
1479	92.5	собака	pes
24	1.5	собака	pejsek
13	0.8	собака	čokl
9	0.6	собака	psí
7	0.4	собака	on
5	0.3	собака	zvíře
5	0.3	собака	fena
4	0.3	собака	psisko
3	0.2	собака	fenka
3	0.2	собака	psik
3	0.2	собака	hajzl
2	0.1	собака	vlčák
2	0.1	собака	fuj
2	0.1	собака	pejsánek
2	0.1	собака	čubka
1	0.1	собака	potvora
1	0.1	собака	svěfenec

Как можно заметить, несмотря на явное сходство между двумя способами поиска, обнаруживаются также существенные различия. Они касаются в первую очередь числа вхождений: *pes* без функции *lemmaty* – 256, с функцией *lemmaty* – 1479. Кроме того, поиск с использованием функции *lemmaty* предоставляет значительно больше возможных переводных эквивалентов, которые в определенном контексте оказываются адекватными. К сожалению, в перечне возможных эквивалентов обнаруживается также программная слабость Treq – фиксация ошибок. Они возникают как результат либо автоматического выравнивания текстов, либо неудачного перевода. Иногда это может быть и результат обоснованных решений переводчика. Следует осознавать, что в принципе любая трансформация исходного текста может привести к тому, что составленная пара трансланд/транслят будет вызывать сомнения.

Полезность Treq в процессе перевода доказывает поиск многозначных лексем, напр. *pociąg* на чешский язык:

▲ Frekvence ▼	▲ Procenta ▼	▲ Polština ▼	▲ Čeština ▼
3833	90.5	pociąg	vlak
27	0.6	pociąg	vláček
19	0.4	pociąg	vagón
18	0.4	pociąg	četa
17	0.4	pociąg	souprava
16	0.4	pociąg	vlakový
15	0.4	pociąg	železnice
15	0.4	pociąg	přítahlivost
13	0.3	pociąg	metro
12	0.3	pociąg	železniční
12	0.3	pociąg	jet
9	0.2	pociąg	sklon
8	0.2	pociąg	vůz
8	0.2	pociąg	přítahovat
8	0.2	pociąg	vlak
7	0.2	pociąg	rychlík
6	0.1	pociąg	lokomotiva
5	0.1	pociąg	nádraží
5	0.1	pociąg	přívěsový
4	0.1	pociąg	vystoupit
4	0.1	pociąg	cesta
4	0.1	pociąg	spojení
4	0.1	pociąg	náklonnost

Прежде чем перейти к анализу отдельных вхождений в Treq, следует обратиться к содержанию словарных статей в переводных польско-чешских словарях.

Электронный польско-чешский словарь *LINGEA* приводит лишь один эквивалент *vlak*, вместе с тем фиксируя словосочетание *pociąg pośpieszny*, его эквивалент *rychlík*, а также предложения *Mój pociąg odjeżdża o szóstej – Vlak mi jede v šest* (VSPČ).

Бумажный польско-чешский словарь *AKADEMIA* под редакцией Карла Оливы фиксирует два основных значения слова *pociąg*: 1. *vlak* и 2. (*do czego*) *náklonnost* (*k čemu*), *sklon* (*k čemu*). Кроме того, словарь приводит несколько словосочетаний (PČS 56).

Следовательно, пользователь вышеназванного электронного словаря получает неполную информацию о возможных переводных эквивалентах. Ситуация с бумажным словарем выглядит по-другому; учитывается значение *‘upodobanie, które trudno jest opranować’* и приводятся его два эквивалента (*náklonnost, sklon*).

Сопоставление словарной информации с информацией, полученной из Treq, демонстрирует большие возможности корпусного анализа. Эта программа выявляет не только эквиваленты *vlak, náklonnost* и *sklon*, но обнаруживает и другие, которые могут быть использованы в переводе. Обращает на себя внимание тот факт, что Treq приводит в качестве возможного эквивалента *príťažlivosť* (15 вхождений). Подробный анализ сочетаемости этой лексемы, а точнее контекстов ее употребления, позволяет также выделить словосочетания: *sexuálni príťažlivosť, fyzická príťažlivosť*. Treq подвергает сомнению слово *náklonnost* в качестве первого эквивалента лексемы *pociąg*, которая скорее всего является эквивалентом польских *skłonność* или *upodobanie*. Корпусные данные, которыми располагает Treq, чаще указывают на пару эквивалентов *pociąg – sklon*, о чем свидетельствуют словосочетания *pociąg do alkoholu – sklon k alkoholu*. Справедливости ради следует обратить внимание также на многочисленные ошибки, поэтому отдельные вхождения надо обязательно верифицировать.

Еще больше ошибочных вхождений фиксирует Treq в случае поиска многолексемных единиц с помощью функции *многолексемные* (*víceslovné*). Значительно помогает в установлении пар эквивалентов многолексемных единиц функция *регулярные* (*regulárni*), позволяющая формулировать более сложные запросы. Так называемые регулярные слова содержат специальные символы, с помощью которых можно выявить определенный набор слов. Регулярные слова состоят:

- 1) из специальных символов,
- 2) специальных и альфанумерических символов или
- 3) альфанумерических символов.

Эта функция не столько способствует фиксации большего количества, сколько дает возможность выявить возможные варианты как трансландов, так и транслятов.

Наиболее универсальный символ представляет собой точка (.), которая может заменить любую букву. Так, последовательность трех точек позволяет обнаружить трехбуквенные слова, четырех точек – четырехбуквенные и т. д.

Звездочка (*) заменяет любую последовательность символов (ноль и больше). Аналогичную функцию выполняет плюс (+), но он заменяет как минимум один или больше символов (один и больше). Использование первого или второго символа имеет значение и дает другие результаты.

мужчин.*

▲ Frekvence ▼	▲ Procenta ▼	▲ Ruština ▼	▲ Čeština ▼
377	23.5	мужчина	muž
204	12.7	мужчины	muži
169	10.5	мужчин	mužů
94	5.8	мужчину	muže
82	5.1	мужчин	muže
82	5.1	мужчины	muže
76	4.7	мужчин	muži
76	4.7	мужчина	chlap
55	3.4	мужчиной	mužem
46	2.9	мужчинами	muži
21	1.3	мужчинам	mužům
21	1.3	мужчине	muži
19	1.2	мужчиной	muž
17	1.1	мужчины	chlapi
13	0.8	мужчина	muže
12	0.7	мужчину	chlapa
10	0.6	мужчинах	mužích
10	0.6	мужчиной	chlapem
9	0.6	мужчиной	chlap

мужчин.+

▲ Frekvence ▼	▲ Procenta ▼	▲ Ruština ▼	▲ Čeština ▼
377	30.2	мужчина	muž
204	16.4	мужчины	muži
94	7.5	мужчину	muže
82	6.6	мужчины	muže
76	6.1	мужчина	chlap
55	4.4	мужчиной	mužem
46	3.7	мужчинами	muži
21	1.7	мужчине	muži
21	1.7	мужчинам	mužům
19	1.5	мужчиной	muž
17	1.4	мужчины	chlapi
13	1.0	мужчина	muže
12	1.0	мужчину	chlapa
10	0.8	мужчинах	mužích
10	0.8	мужчиной	chlapem

Пользователь Treq может применять и другие символы с целью выявления вариантов многолексемных единиц: круглые скобки, квадратные скобки, обратную косую черту, вертикальную черту, каре. Четвертая функция (A=a) предоставляет возможность не различать прописные и строчные буквы, т. е. будет обнаружена любая форма слова, например: Женщина, ЖЕНЩИНА, женщина.

Представляется, что приложение Treq успешно применяется при поиске переводных эквивалентов однолексемных единиц исходного языка. Процесс поиска осуществляется быстро и эффективно, от пользователя не требуется особой квалификации или специальной подготовки, например, усвоения регулярных слов. Найденные при помощи Treq переводные эквиваленты в подавляющем большинстве случаев правильны. Конечно, относительно многозначных лексем необходим дополнительный анализ, включая исследование контекстов их употребления.

Тем не менее Treq позволяет решать и более сложные переводческие задачи, которые в большой степени напоминают режим работы с основной программой НКЧЯ – KonText. На практике попытаемся с помощью Treq установить польские и русские эквиваленты чешских фразем с компонентом *sranda*.

Чешско-польский словарь не фиксирует ни одну фразему с этой лексемой, приводя только польские эквиваленты лексемы *sranda*: *heca*, *komedia*, *szopa*, *szopka*. Словарь отмечает также вульгарную окраску этой лексемы (SCzP 742).

Чешско-русский словарь приводит три ее русских эквивалента: *хохма*, *умора* и *потеха*, отмечая их стилистические характеристики, кстати, спорные (VČRS 987).

Большой чешско-польский фразеологический словарь приводит следующую словарную статью:

dělat si/mít z někoho/něčeho legraci/psinu/srandu/švandu pot.

‘robić sobie z kogoś, czegoś żarty, bawić się czyimś kosztem’

stroić sobie z kogoś/czegoś żarty // robić sobie z kogoś/czegoś jaja // robić coś dla kawału // używać sobie na kimś (WCzPSF 226–227)

Уже на первый взгляд сомнения вызывает синонимический ряд слов (*legrace/psina/sranda/švanda*), отличающихся разными стилистическими оттенками: *legrace* принадлежит к разговорной речи (*hovorová čeština*), *psina/sranda/švanda* к обиходно-разговорному койне (*obecná čeština*). В отличие от стилистически нейтрального *legrace*, остальные слова отличаются эмоционально-экспрессивной окрашенностью (*psina*, *švanda*), а даже вульгарностью (*sranda*). *Legrace* фиксируется в словаре литератур-

ного чешского языка (SSČ), остальные три в словаре нелитературного чешского языка (SNČ).

Treq дает следующие результаты отдельных фразем:

dělat si/mít z někoho/něčeho legraci – 669 вхождений
 dělat si/mít z někoho/něčeho psinu¹ – 6 вхождений
 dělat si/mít z někoho/něčeho srandu – 1211 вхождений
 dělat si/mít z někoho/něčeho švandu – 0 вхождений

Как можно заметить, проверка польских эквивалентов может эффективно верифицировать обоснованность фиксации и последовательности отдельных вариантов исходного транслянда. Оказывается, что два из четырех практически не фиксируются в ресурсах InterCorp. Учитывая стилистические характеристики *dělat si z někoho/něčeho legraci* и *dělat si z někoho/něčeho srandu*, с помощью Treq можно установить следующие пары эквивалентов:

Děláš si legraci?/To si děláš legraci? – Żartujesz?/Żartujesz sobie?
 Dělám si legraci. – Żartuję.
 Nedělám si legraci. – Nie żartuję.
 To si děláš srandu? – Jaja sobie robisz?
 Děláš si srandu? – Kpisz sobie?
 Nedělám si srandu! – Nie robię (sobie) jaj!

Эти пары эквивалентов выявляются в результате дополнительной обработки отдельных вхождений, т. е. пользователь вынужден проанализировать предоставленный Treq материал.

Примечательно, что найденные Treq фраземы представляют собой коммуникативные единицы, так редко фиксируемые в словарях. Неслучайно их подавляющее большинство содержится в субтитрах, которые с известной долей условности отражают реальную непринужденную, спонтанную речь в условиях прямого контакта коммуникантов преимущественно в неофициальной ситуации общения.

Само собой разумеется, что некоторые найденные вхождения не имеют никакого отношения к возможным переводным эквивалентам. Понятно, что устранение сомнительных эквивалентов требует от пользователя этой

¹ Примечательно, что и в случае однолексемных единиц надо считаться с возможной ошибочностью полученных результатов. По полученным результатам из Treq русским эквивалентом чешского слова *psina* является *псина* (3 вхождения в InterCorp v9). В современном чешском языке слово *psina* имеет исключительно значение ‘забава, развлечение, веселье’. В современном русском языке оно употребляется в значениях: 1. ‘собачье мясо’, 2. ‘запах собаки, собачьей шерсти’, 3. ‘пес’. Еще раз оказывается, что малое количество вхождений (всего 3) требует дополнительной тщательной проверки, поскольку пары эквивалентов могут оказаться неадекватными.

программы языковой компетенции. Однако несомненно одно: Treq успешно наводит не только на правильный эквивалент, но и нередко подсказывает другие варианты перевода.

Аналогичный анализ можно осуществить, пытаясь установить русские эквиваленты.

Treq успешно находит русские эквиваленты, а кроме того контексты их употребления. Так, выстраивается следующий ряд эквивалентов: *смешно, весело, забавно, прикольно, здорово, шутка* и др. Из трех вышеприведенных эквивалентов (*хохма, умора, потеха*) появляется только *умора* (всего одно вхождение). Тем не менее пользователь Treq получает ценный материал, который дает возможность составить, например, такие пары чешско-русских эквивалентов:

To není sranda. – Это не смешно.

To je sranda!//Je to sranda! – Это (же) смешно!/Это весело!/Это забавно!

Как правило, вышеприведенные фраземы употребляются в качестве эмоциональной реакции для выражения оценки сложившейся ситуации, которая связана с чьими-либо действиями или отсутствием таковых.

Однако вышеприведенные пары эквивалентов не следует считать окончательно установленными. Носителю русского языка сразу приходят в голову возможные эквиваленты *To je sranda!; Вот умора!, Вот потеха!, Просто потеха!, Просто умора!, Это хохма!, Вот хохма!*. Подобного рода варианты можно обнаружить даже в ресурсах параллельного корпуса InterCorp, в частности в его последней одиннадцатой версии². В очередной раз обнаруживается слабость не столько программы Treq, сколько небольшого объема корпусных ресурсов русского корпуса InterCorp. Именно по этой причине Treq не составляет полную картину возможных эквивалентов.

Следует отметить, что подобного рода речевых фраз, выражающих различные эмоциональные реакции коммуникантов, явно не хватает в переводных словарях. Проблема фиксации подобных выражений в лексикографических трудах заключается, к сожалению, в традиционном восприятии слова как основной единицы словаря.

Поиск русских эквивалентов для фраземы *dělat si z někoho/něčeho srandu* с помощью Treq дает следующие результаты (учитываем только часть из них):

² Фразема *Вот умора!* имеет 5 вхождений в InterCorp v11, ни одного в InterCorp v9. Фразема *Вот потеха!* фиксируется 5 раз в InterCorp v11, в InterCorp v9 имеет 4 вхождения. Фразема *Вот хохма!* как в InterCorp v9, так и в InterCorp v11 не фиксируется.

▲ Frekvence ▼	▲ Procenta ▼	▲ Čeština ▼	▲ Ruština ▼
42	11.9	Děláš si srandu	шутить
35	9.9	Děláš si srandu	Ты шутить
15	4.3	Dělám si srandu	шучу
14	4.0	Děláš si srandu	Шутишь
7	2.0	Děláte si srandu	шутите
5	1.4	Děláte si srandu	Вы шутите
5	1.4	Dělám si srandu	Шучу
5	1.4	Dělám si srandu	Я шучу
3	0.9	Děláš si ze mě srandu	Издеваешься
3	0.9	Děláš si srandu	Издеваешься
3	0.9	Děláš si srandu	Вы шутите
3	0.9	Dělám si srandu	пошутил
3	0.9	Děláš si ze mě srandu	шутить
3	0.9	Děláš si srandu	шутите
3	0.9	Děláš si srandu	издеваешься
3	0.9	Dělám si srandu	Я пошутил
2	0.6	Děláte si ze mě srandu	стебешься надо мной
2	0.6	Děláte si ze mě srandu	Вы шутите
2	0.6	Děláš si ze srandu	разыгрываешь
2	0.6	Děláš si ze mě srandu	издеваешься
2	0.6	Dělám si srandu	Я просто шучу

Как можно заметить, вышеприведенный материал предоставляет пользователю Treq очень ценный материал для дальнейшей обработки, наводя на возможные переводные пары, например:

Děláš si srandu? – Ты шутишь?/Ты что, шутишь?/Шутишь?/Ты издеваешься?/Ты что, издеваешься?/Издеваешься?

Děláš si ze mě srandu? – Ты издеваешься надо мной?

Dělám si srandu. – Я шучу.

Следует отметить, что среди возможных эквивалентов *dělat si z někoho/něčeho srandu* находим также глагол *стебаться*, который принадлежит к молодежному сленгу.

По результатам анализа отобранных нами лексем/фразем представляется возможным сделать несколько выводов.

Во-первых, необходимо подчеркнуть полезность программы Treq, которая может успешно помочь переводчику в решении конкретных переводческих задач. Ее главное преимущество – это текстовые эквиваленты, составляющие довольно обширный список. Можно предположить, что их поиск и установление баз Treq превосходит реальные возможности даже самого опытного переводчика.

Во-вторых, извлеченный корпусной материал требует детального исследования, поскольку содержит большое количество очевидных ошибок, которые в первую очередь касаются эквивалентов многолексемных единиц. Эти ошибки значительно усложняют картину эквивалентности, а их устранение требует от пользователя Treq достаточно высокой языковой компетенции.

В-третьих, существенной проблемой является иногда малое число вхождений, которое не позволяет признать полученные корпусные данные репрезентативными и достоверными. Такая ситуация нередко возникает в случае чешско-русского Treq, отличающегося не только относительно небольшим объемом текстов, но и ограниченным жанровым разнообразием. Наоборот, чешско-польский Treq подтверждает правило: чем больше корпус, тем больше вероятность того, что в нем найдется искомый эквивалент или эквиваленты.

В-четвертых, нельзя забывать о том, что Treq пользуется ресурсами InterCorp, т. е. в основном переводами (исходным языком чаще всего является английский). Так, на практике Treq сопоставляет друг с другом, например, два перевода с английского на русский и чешский языки.

В-пятых, множество неоднократно приведенных Treq эквивалентов показывает нам возможности языковой системы, разные переводческие стратегии, возможные переводческие трансформации.

Библиография

Škrabal, Michal, Martin Vavřín. „Databáze překladových ekvivalentů Treq“. *Časopis pro moderní filologii*, nr 99 (2), 2017, s. 245–260.

Словари

- PČS – Oliva, Karel. *Polsko-český slovník*. Praha, Academia, 1999.
SCzP – Siatkowski, Janusz, Basaj Mieczysław. *Słownik czesko-polski*. Warszawa, Wiedza Powszechna, 2002.
SNČ – *Slovník nespisovné češtiny*. Praha, Maxdorf, 2009.
SSČ – *Slovník spisovné češtiny pro školu a veřejnost*. Praha, Academia, 2001.
Treq – <https://treq.korpus.cz>.
VČRS – *Velký česko-ruský slovník*. Voznice, Leda, 2005.
VSPČ – *Velký slovník polsko-český [a] česko-polský slovník*. Lexicon 5. Brno, Lingea, 2010.
WCzPSF – *Wielki czesko-polski słownik frazeologiczny*. Red. Teresa Zofia Orłoś. Kraków, Wydawnictwo Uniwersytetu Jagiellońskiego, 2009.