



ЧЕСТОТНОСТ И ПРОИЗВОДНОСТ НА БЪЛГАРСКАТА СПЕЦИАЛИЗИРАНА ЛЕКСИКА ОТ ОБЛАСТТА НА МЕЖДУНАРОДНИЯ ТРАНСПОРТ И СПЕДИЦИЯТА

Нина Димитрова
ndimitrova@vtu.bg

*Висше транспортно училище „Тодор Каблешков”
ул. „Гео Милев” №158, гр. София
БЪЛГАРИЯ*

Ключови думи: компютърна лексикография, корпус, честотност, производност, специализирана лексика, транспорт, спедиция

Резюме: Настоящото изследване е част от по-мощен труд по създаването на тесноотраслов Руско-български речник на термините по спедиция. Негова цел е съставянето на словник на лексикалните единици, обслужващи тази сфера на дейност в България.

За получаването на емпирични данни, които биха позволили постигането на така формулираната цел, е използван софтуерът за автоматична обработка на текстови корпуси AntConc. С негова помощ са идентифицирани лексикалните единици, които се характеризират с най-висока фреквентност. Първите позиции в честотния списък се заемат от лексемите: превоз (3,46% от обема на корпуса), товар (1,8%), транспорт (1,47%), стока (0,88%) и др. Със същия инструмент за обработка на електронни корпуси са получени данни по критерия производност на лексикалните единици в изследваната предметна област. Установено е, че най-висок деривационен потенциал притежават лексикалните единици: превоз (167 деривата), товар (158 деривата), транспорт (89 деривата) и стока (76 деривата).

В резултат на проведеното изследване бяха идентифицирани лексикалните единици, които изграждат специализираната лексика по международен транспорт и спедиция, а съпоставянето на данните от честотния списък с данните от списъка за броя на производните им показва, че в самата ѝ сърцевина са разположени лексикалните единици: превоз, товар, транспорт и стока.

1. УВОД

Съставянето на преводен тесноотраслов речник¹, какъвто е подготвяният Руско-български речник на термините по спедиция [2], включва няколко етапа: проектиране, съставяне на словник на термините в изходния език, анализ и описание на ексцерпираната терминологична лексика, фиксиране на синонимните, йерархичните и асоциативните отношения. Последният етап, на който се намира и нашата работа по

¹ Терминът *тесноотраслов речник* е на М. Попова [1, 103].

гореспоменатия речник, е подбор на еквиваленти на езика/езиците на превода и оформяне на резултатите от терминографската работа във вид на речник.

Липсата на тълковни тесноотраслови терминологични речници – както на изходния език, така и на езика на превода – затруднява в голяма степен съпоставянето на терминологичните системи и установяването на семантичната еквивалентност на подобрите речникови единици. За решаването на тази нелека задача ценна и полезна информация може да бъде извлечена от източници в сферата на функционирането на езика на превода, която да послужи за съставяне на словник и съпоставка на контекстите, в които функционират термините.

2. ИЗСЛЕДВАНЕ НА ЧЕСТОТНОСТТА И ПРОИЗВОДНОСТТА НА СПЕЦИАЛИЗИРАНАТА ЛЕКСИКА ПО МЕЖДУНАРОДЕН ТРАНСПОРТ И СПЕДИЦИЯ

Като материал за изследване на българската специализираната лексика от областта на транспорта и спедицията е използван учебникът „Международен транспорт и спедиция” с автор проф. д-р Христина Николова, Университетско издателство „Стопанство”, 2011, 268 с., който е анализиран с помощта на софтуер за автоматична обработка на текстови корпуси AntConc (Version 3.5.7, Anthony, 2018, Developed by Laurence Anthony).

В статията си „Компютърна лексикография – традиции и предизвикателства” [4, 30–31] Светла Коева привежда различни дефиниции за *корпус*. Например: „съвкупност от автентични езикови данни, които могат да се използват за лингвистични изследвания“ [5, 1]; „съвкупност от езиков материал в електронна форма, подходящ за компютърна обработка с приложение в лингвистичните изследвания или езиковите технологии“ [5, 1]; „голяма колекция от езикови примери, представени по начин, даващ възможност за компютърна обработка, и избрани по определени (лингвистични) критерии, така че да представляват адекватен езиков модел“ [6, 9] и др.

Корпусът на настоящото изследване, състоящ се от 64 119 единици (думи), след изваждането от него на служебните думи (предлози, съюзи и частици) се редуцира до 48 978 единици.

В лингвистиката *честотност* (или *фреквентност*) на думите е понятие, с което се означава тяхната повтораемост в речта. За изследването ѝ се използват различни методи, най-вече тези на компютърната лингвистика и компютърната лексикография, които позволяват чрез анализ на определен корпус (корпуси) да се получат необходимите данни за наблюдение, анализ и изводи.

Честотността на думите в даден корпус се изчислява по формулата

$$Frec_x = \frac{Q_x}{Q_{all}}$$
, където $Frec_x$ е честотността на думата «*x*», Q_x е броят на употребите на думата «*x*», а Q_{all} е общият брой думи на корпуса. В повечето случаи показателят *честотност* се изразява в проценти.

По този начин са получени данни за единиците, които се характеризират с най-висока честотност в изследвания корпус. По-долу в низходящ ред са изброени тези от тях, които попадат в диапазона от 3,46% до 0,1% от общия обем на корпуса.

Най-често срещаната единица е лексемата **превоз** и нейните производни – 1 697 случая, чиито дял е 3,46% от обема на корпуса.

На второ място се нарежда лексемата **товар** и нейните производни с регистрираните 884 случая, което формира 1,8% от обема на корпуса.

На трето и четвърто място по употреба са лексемите **транспорт** (720 случая) и **стока** (431 случая), които заедно с производните си представляват, съответно, 1,47% и 0,88% от обема на корпуса.

На пето и шесто място са лексемите *железница + железен път/железопътен* (в пълна и съкратена форма *жп*) – 346 случая или 0,75% от обема на корпуса, и *море* и производните ѝ – 222 случая или 0,45% от обема на изследвания корпус.

Седмото място се заема от лексемата *спедиция + експедиция/експедирам* и производните им, за които са регистрирани 218 случая на употреба, представляващи 0,45% от обема на корпуса.

По-надолу в честотния списък следват лексемите: *договор* (206 случая – 0,42%), *въздух* (190 случая – 0,39%), *отговорност* (186 случая – 0,38%), *документ* (174 случая – 0,36%), *автомобил* (170 случая – 0,35%), *получаване + получаващ* (164 случая – 0,33%), *мултимодален* (157 случая – 0,32%), *доставка + доставчик* (149 случая – 0,3%), *средство* (146 случая – 0,3%), *условие* (143 случая – 0,29%), *организация* (138 случая – 0,28%), *правило* (137 случая – 0,28%), *мито* (135 случая – 0,28%), *кораб* (135 случая – 0,28%), *право* (130 случая – 0,27%), *изпращане + изпращач* (127 случая – 0,26%), *конвенция* (120 случая – 0,25%), *срок* (108 случая – 0,22%), *услуга* (107 случая – 0,22%), *търговия* (107 случая – 0,22%), *комбиниран* (106 случая – 0,22%), *плащане* (100 случая – 0,2%), *цена* (92 случая – 0,19%), *контейнер* (92 случая – 0,19%), *работа + обработване* (91 случая – 0,19%), *разходи* (88 случая – 0,18%), *вреда* (79 случая – 0,16%), *вагон* (78 – 0,16%), *гара* (77 случая – 0,16%), *предаване* (75 случая – 0,15%), *липса + липсващ* (73 случая – 0,15%), *пристанище* (73 случая – 0,15%), *карнет* (73 случая – 0,15%), *оператор* (71 случая – 0,14%), *пратка* (67 случая – 0,14%), *обезщетение* (64 случая – 0,13%), *повреда* (64 случая – 0,13%), *стойност* (62 случая – 0,13%), *обем* (61 случая – 0,12%), *клиент* (60 случая – 0,12%), *застрахован* (58 случая – 0,12%), *тегло* (56 случая – 0,11%), *разрешително* (56 случая – 0,11%), *инфраструктура* (55 случая – 0,11%), *тарифа* (53 случая – 0,11%), *склад + складиране* (52 случая – 0,11%), *безопасност* (52 случая – 0,11%), *вещества* (52 случая – 0,11%), *движение* (51 случая – 0,1%) и др.

С помощта на гореспоменатия софтуер са извлечени данни и по критерия *производност* (брой на дериватите) на единиците от специализираната лексика по международен транспорт и спедиция. По-долу поради ограниченията в обема на статията са изброени само тези, които заемат челните места в списъка.

С най-голям брой деривати е лексемата *превоз (+ превозвач)* – 167 единици. Например:

- *товарни превози, сухопътни превози, железопътни превози, автомобилни превози, въздушни превози, морски превози, речни (вътрешноводни) превози, мултимодални превози, унимодални превози, смесени превози, транзитни превози* и др.;

- *превоз на тежки или обемисти товари, превози на бързоразвалящи се стоки, превоз на насипни товари, превоз на течни товари, превоз на опасни товари, превоз на дребни пратки, превоз на комплектни товари, превоз на палетизирани товари, превоз на товари с придружител, превоз на стоки под режим ТИР, превоз на товари с ниска скорост* и др.;

- *поръчка за превоз, договор за превоз, условия на превоза, тарифиране на превоза, себестойност на превоза, цена на превоза, маршрут на превоза, срок за превоз, отказ от превоз* и др.;

- *превозен документ, превозен капацитет, превозен парк, превозна операция, превозна работа, превозна цена, превозни средства* и др.;

- *превозвач, подпревозвач, лицензиран превозвач, договорен превозвач, физически превозвач, железопътен превозвач, морски превозвач* и др.

На второ място в изследвания корпус по брой деривати е лексемата *товар*, която има 158 производни. Например:

- опасни товари, сухи товари, течни товари, бързоразвалящи се товари, извънгабаритни товари, тежки товари, леки обемни товари, обемисти товари, палетизирани товари, контейнерни товари, комплектни товари, групажни товари, обратни товари, конвенционални товари, генерални товари и др.;

- тегло на товара, обем на товара, маса на товара, опаковане на товара, етикиране на товара, маркиране на товара, укрепване на товара, разкрепване на товара, превоз на товара, повреда или загуба на товара, доставяне на товара, предаване на товара на получателя и др.;

- товарен автомобил, товарен кораб, товарен вагон, товарен документ, товарен ордер, товарен пункт, товарна единица, товарна разписка, товарни места, товарни приспособления, товарни операции, товаро-разтоварни операции, товарни помещения и др.;

- автомобилна товарителница, товарителница CMR (CMR-товарителница), въздушна товарителница (AWB), морска товарителница, речна товарителница, дунавска товарителница, рейнска товарителница, железопътна товарителница, комбинирана товарителница, мултимодална товарителница, електронна товарителница и др.;

- товародател, товарополучател, товароизпращач, товароспособност на транспортното средство, товаровместимост на превозното средство и др.

Трето място по брой производни заема лексемата **транспорт** (89 деривата).
Например:

- сухоземен транспорт, автомобилен транспорт, железопътен транспорт, въздушен транспорт, речен транспорт, морски транспорт, комбиниран транспорт, мултимодален транспорт, тръбопроводен транспорт, международен транспорт, транспорт от врата до врата и др.;

- транспортен договор, транспортен коридор, транспортен оператор, транспортен пазар, транспортен документ, мултимодален транспортен документ, унимодален транспортен документ, единен транспортен документ, мултимодална транспортна верига, транспортна инфраструктура, транспортно право, транспортно средство, транспортни услуги, комбинирани транспортни услуги, мултимодални транспортни услуги, пазар на транспортните услуги, стойност на транспортните услуги, цени на транспортните услуги, качество на транспортните услуги, транспортни разходи, транспортни тарифи, транспортни такси, транспортни технологии, интелигентни транспортни технологии, транспортни операции, мултимодални транспортни операции, транспортни застраховки и др.

На четвърто място сред единици от изследвания корпус е лексемата **стока** с брой на дериватите – 76. Например:

- обем на стоката, тегло на стоката, маса на стоката, повреда на стоката, загуба на стоката, стойност на стоката, цена на стоката, качество на стоката, количество на стоката, приемане на стоката за превоз, предаване на стоката на получателя, доставка на стоката, недоставяне на стоката, придружител на стоката, натоварване на стоката, разтоварване на стоката, претоварване на стоката, стифиране на стоката, застраховане на стоката и др.;

- стокова опаковка, стокова партида, стокови документи и др.

На пето място е лексемата **железница (+ железен път)** с 49 деривата.
Например:

- отправна железница, получаваща железница, транзитни железници;

- габарити на железния път;

- железопътен подсектор, железопътен подвижен състав, железопътен терминал, железопътна администрация, железопътна безопасност, железопътна инфраструктура, железопътни средства, железопътни оператори и др.

Шестото място по брой на производните заема лексемата **мито** (+ **митница**), за която са открити 45 деривата. Например:

- *вносно мито, доставено без платено мито до ...*, с посочено място на доставка (*Delivered Duty Unpaid to... DDU*), *доставено до ...*, *мито платено, с посочено място на доставка (Delivered Duty Paid to... DDP)* и др.;

- *митническа територия, митническа декларация, митническа проверка, митническа пломба, митническа фактура, митнически контрол, митнически манифест, митнически транзит, митнически нарушения, митнически декларатор, митнически служител, митнически пункт, митнически вземания, митнически глоби, митнически документи, митнически съборове* и др.

По-нататък в списъка следват лексемите: **кораб** (36 деривата), **плащане** (33 деривата), **правило** (31 деривата), **договор** (30 деривата), **конвенция** (29 деривата), **право** (28 деривата), **море** (27 деривата), **спедиция** (26 деривата), **въздух** (25 деривата), **отговорност** (25 деривата), **услуга** (25 деривата), **търговия** (25 деривата), **доставка** (23 деривата), **контейнер** (22 деривата), **разходи** (22 деривата), **тарифа** (21 деривата) и др.

3. ЗАКЛЮЧЕНИЕ

В резултат на настоящото изследване беше съставен словник на лексикалните единици, обслужващи сферата на международния транспорт и спедицията в България. На базата на получените статистически данни честотата на употреба на отделните лексикални единици, включени в него, може да бъде определена като висока, средна или ниска честота, което дава допълнителна ценна информация за статуса им в специализираната лексика на изследваната предметна област.

Съпоставянето на данните от съставения честотен списък с данните от списъка за броя на дериватите дава възможност да бъдат идентифицирани лексикалните единици, които се отличават с най-висока честотност и същевременно имат най-висок деривационен потенциал. Това са: **превоз**, **товар**, **транспорт** и **стока**, които се намират в самата сърцевина на българската специализирана лексика по международен транспорт и спедиция.

ЛИТЕРАТУРА:

[1] Попова М., Приложно терминологично, изд. Авангард Прима, София, 2017, ISBN 978-619-160-755-6.

[2] Димитрова Н. К., Русско-болгарский словарь терминов транспортно-экспедиционной деятельности//Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. № 2. 2016. С. 65 – 73, ISSN 1814-2958.

[3] Николова Хр., Международен транспорт и спедиция, Университетско издателство „Стопанство”, София, 2011 г.

[4] Коева Св., Компютърна лингвистика – традиции и предизвикателства, 70 години българска академична лексикография, Сб. Доклади от Шестата национална конференция с международно участие по лексикография и лексикология, София, Институт за български език „Проф. Л. Андрейчин“ – БАН, 24–25 октомври 2012 г., Академично издателство „Проф. Марин Дринов”, София, 2013, С. 30–38, ISBN 978-954-322-578-1.

[5] Leech, G. Introducing corpus annotation. // Garside R., G. Leech, A. M. McEnery (eds.). Corpus Annotation: Linguistic Information from Computer Text Corpora. London: Longman, pp. 1–18.

[6] Коева, Св. Българският семантично аотиран корпус – теоретични постановки. // Българският семантично аотиран корпус. Редактор и съставител Св. Коева. София, 2010, С. 7–43.

FREQUENCY AND DERIVATIONAL POTENTIAL OF SPECIALIZED BULGARIAN LEXICAL UNITS IN THE FIELD OF INTERNATIONAL TRANSPORT AND FREIGHT FORWARDING

Nina Dimitrova
ndimitrova@vtu.bg

*Todor Kableshkov University of Transport
Sofia, 158 Geo Milev Str.
THE REPUBLIC OF BULGARIA*

Key words: *computational lexicography, corpus, frequency, derivational potential, specialized lexical units, transport, freight forwarding*

Abstract: *The present study is a part of our larger work on the creation of a specialized Russian-Bulgarian dictionary of freight forwarding terms. The purpose of the study is to compile a word list of lexical units that are in use in this field of activity in Bulgaria.*

To obtain the necessary empirical data the corpus analysis toolkit for concordancing and text analysis AntConc was used. With its help the lexical units with the highest frequency are identified. The first positions in the frequency list are occupied by the word tokens: carriage (3.46% of the total volume of the corpus), cargo (1.8%), transport (1.47%), goods (0.88%) and others. Using the same corpus analysis software tool data on the derivational potential of lexical units in the studied subject area is obtained. It was found that the lexical units that have the most derivatives are: carriage (167 derivatives), cargo (158 derivatives), transport (89 derivatives), goods (76 derivatives) and others.

As a result of the present research a word list of specialized lexical units in the field of international transport and freight forwarding was compiled. A comparison of the received data about the frequency with the data about the derivational potential of the lexical units shows that “carriage”, “cargo”, “transport” and “goods” are the core components in the analyzed vocabulary.