

METODA WYODRĘBNIANIA ODTWARZALNYCH KONSTRUKCJI WERBO-NOMINALNYCH JĘZYKA ROSYJSKIEGO Z KORPUSÓW TEKSTOWYCH

THE EXTRACTION METHOD OF REPRODUCIBLE RUSSIAN VERBAL-NOMINATIVE STRUCTURES FROM TEXT CORPORA

YURY FEDORUSHKOV, DANIEL DZIENISIEWICZ

ABSTRACT. The article describes an extraction method of Russian verbal-nominative structures from electronic texts with the use of grammatical annotation and regular expressions syntax. The stages of the retrieval process are outlined. The list of extracted structures can be subsequently verified for the presence of reproducible units (phrasems).

Keywords: corpus, morphological analysis, regular expressions, reproducibility, collocation

Yury Fedorushkov, Uniwersytet im. Adama Mickiewicza w Poznaniu, Poznań – Polska, jerfed@amu.edu.pl

ORCID: 0000-0001-9433-0956

Daniel Dzienisiewicz, Uniwersytet im. Adama Mickiewicza w Poznaniu, Poznań – Polska, dzienis@amu.edu.pl

ORCID: 0000-0003-0400-5143

1. Uwagi wstępne

W niniejszym artykule zaprezentowana zostanie metoda ekscerpacji konstrukcji werbo-nominalnych języka rosyjskiego, tj. konstrukcji czasownikowo-rzeczownikowych, dla potrzeb leksykografii¹. W tym celu utworzona została specjalna próbka korpusowa, dalej określana jako Korpus (zob. rozdz. 2). Nie podejmuje się przy tym próby określenia statusu wyekscerpowanych konstrukcji², aczkolwiek w ramach wprowadzenia do dalszych analiz po-

¹ Należy zaznaczyć, że przedstawiana w artykule metoda znalazła zastosowanie w wyodrębnianiu innych typów konstrukcji. Warsztat, w którym dokonano ekscerpacji tzw. *dwuwyrazowców* (na przykładzie konstrukcji przymiotnikowo-rzeczownikowych) na niewielkiej próbce korpusowej (ok. 80000 wyrazów) został zaprezentowany w artykule [Fedorushkov, Szutkowski 2010].

² Por. interesujące sformułowanie nazwy kategorii: „być podejrzanym o frazematyczność” [Wierzchoń 2010: 90].

święconych opisywanym jednostkom poruszone zostaną niektóre zagadnienia teoretyczne z zakresu frazematyki. Dylemat merytoryczny – czy zwrot jest zakorzeniony w języku, czy też nie – jest bowiem często pomijany przez badaczy.

Skomputeryzowane techniki wyodrębniania konstrukcji wielowyrzowych, np. oparte o filtry ankodowe³, dają rezultat „masowy”: potencjalne kolokacje znajdują się obok połączeń okazjonalnych, niereprodukowalnych⁴, jedynie generowanych w sposób przypadkowy z określonym stopniem entropii – w zależności od potrzeb komunikacyjnych. Przykładowo, z poniższej grupy wyekscerpowanych konstrukcji werbo-nominalnych tylko jedną możemy zakwalifikować jako potencjalną (wymagającą weryfikacji) kolokację – por. poniższe kwalifikatory **S** oraz **F**:

вернуться на смену
вернуться на родину

S (połączenie okazjonalne: klasa swobodna; produkt);
F (frazem: kolokacja; reprodukt)⁵.

Należy podkreślić, że obiekty badań umieszczamy właśnie w ramach szeroko pojmowanej frazematyki. Przed opisaniem warsztatu ekscerpacji przywołałyśmy niektóre powstałe na polskim gruncie koncepcje dotyczące semantyki i składni analizyzmów werbo-nominalnych oraz ich związków z frazeologią, za Wojciechem Chlebdą nakreśliłyśmy teoretyczne ramy analitycznego i syntetycznego nurtu we frazeologii oraz wyjaśnimy, co rozumieć będziemy pod

³ Por. także metody ekscerpacji kolokacji typu *log-likelihood ratio* [Захаров, Хохлова 2010] oraz stosunkowo nowa metoda modelowania tematycznego – *topic modeling*: „Na przykład w powieści detektywistycznej można się spodziewać słów typu «poszlaka», «ślady», «trup», «policja» itp., a ich obecność (i współobecność) będzie tym bardziej znacząca, im rzadziej wystąpią one w pozostałych tekstach korpusu, dajmy na to, w powieści historycznej czy sentymentalnej. [...] Porównanie próbek rzeczywistych z losowymi ujawnia nieprzypadkowe połączenia wyrazowe w próbkach rzeczywistych: im większa różnica danych obserwowanych i danych teoretycznych (tj. losowych), tym większy jest związek badanych wyrazów. Cała procedura zostaje następnie powtórzona w bardzo dużej liczbie iteracji, dzięki czemu niedoskonały z początku model sukcesywnie się optymalizuje przez eliminację współwystąpień przypadkowych” [Eder 2016: 38].

⁴ Por. problem kwalifikacji wyekscerpowanych konstrukcji w obrębie odtwarzalności w „totalnym” wyniku ekscerpacji zautomatyzowanej: „[...] efekty ekscerpacji manualnej są dokładniejsze (ang. *precision*) niż w przypadku ekscerpacji zautomatyzowanej (ang. *recall*). [...] Znalezienie odpowiedniej równowagi pomiędzy tymi dwoma stanami (*precision* i *recall*) stanowi największy problem w zakresie automatycznego poszukiwania wielowyrzowców” [Fedorushkov 2010: 61].

⁵ Rzecz jasna, sugerujemy się tutaj także weryfikacją kontekstową. O tym, że połączenie *вернуться на родину* jest konstrukcją odtwarzalną świadczy chociażby fakt, że jest ono jedną z czołowych nazw tematycznych bądź też nazwą usługi prawniczej, por. kontekst: „Советы и юридические консультации адвокатов и юристов по теме «Вернуться на родину» – помощь и ответы в режиме онлайн на любые ваши вопросы” – źródło elektroniczne: <https://www.9111.ru/Защитаправпотребителей> (dostęp 12.06.2016).

pojęciem **konstrukcji odtwarzalnej**, tj. *reproduktu* (por. niżej), oraz konstrukcji werbo-nominalnej.

Danuta Buttler, badając innowacje semantyczno-składniowe polszczyzny, wskazywała na proces analizacji składni, który tłumaczyła większą wyrazistością konstrukcji analitycznych w porównaniu do ich odpowiedników syntetycznych (por. *stać na przeszkodzie* <czemu> i *przeszkadzać* <czemu>). Badaczka twierdziła, iż w werbo-nominalnych analizmach składniowych element werbalny stanowi wykładnik kategorii aspektu, zaś komponent nominalny jest nosicielem znaczenia peryfrazy (np. *skończyć czytać, przeprowadzić spis*). Autorka zauważała, że w pewnych kontekstach, podyktowanych względami stylistycznymi, kulturowymi, znaczeniowymi i pragmatycznymi, powyższe analizmy mogą być substytuowane przez jednowyrazowe formy czasownikowe (*przeczytać, spisać*). Ponadto jako przejaw analizacji polskiej składni ukazywała częstsze używanie wyrażen przyimkowych niż form bezprzyimkowych w konstrukcjach werbo-nominalnych, np. *dowiedzieć się* <o czym> w miejsce *dowiedzieć się* <czego> [Loewe 2000: 17–18].

W swej klasycznej pracy poświęconej konstrukcjom analitycznym Janusz Anusiewicz dzielił analizmy przy pomocy kryterium morfologicznego i wyodrębniał m.in. konstrukcje złożone z czasownika i rzeczownika, którym odpowiadają syntetyczne czasowniki (np. *wejść w kontakt* ← *skontaktować się*). Za konstrukcje analityczne uznawał wyłącznie związki wyrazowe, których ekwiwalentami są czasowniki proste [Loewe 2000: 19 za Anusiewicz 1978: 22]. Jest to pogląd kontrowersyjny. Przykładowo, Piotr Żmigrodzki uważa, iż istnienie równoznacznego czasownika syntetycznego nie jest warunkiem koniecznym, jaki musi spełnić związek wyrazowy, aby został zaliczony w poczet analizmów. Badacz podaje przykłady analizmów bezekwiwalentnych (zob. np. *pałać miłością, wyrzucić wrażenie*) [Żmigrodzki 2000: 15–16].

Iwona Loewe twierdzi, iż konstrukcje werbo-nominalne pełnią funkcję orzeczenia peryfrastycznego (tj. opisowego), a jako synonimiczne wobec nich traktuje proste formy czasowników, związane semantycznie i morfologicznie z rzeczownikiem abstrakcyjnym stanowiącym komponent analizmu (np. *otoczyć opieką i zaopiekować się*). Autorka uważa, że w analizmie werbo-nominalnym nośnikiem semów definicyjnych jest komponent nominalny, zaś element werbalny odpowiada za tworzenie różnorodnych treści nie poddających się jednoznacznej klasyfikacji [Loewe 2000: 20]. Ewa Jędrzejko sądzi, iż nie tylko element nominalny, ale i czasownik może często być znaczeniorodny i nie ma on charakteru wyłącznie wspomagającego [Loewe 2000: 20 za Jędrzejko 1992].

Natomiast Andrzej Bogusławski wskazuje, że istnieją grupy orzeczeń peryfrastycznych, w których określony czasownik implikuje wystąpienie rzeczowników należących do określonego pola semantycznego, np. *doznać* + rze-

czownik nazywający uczucie (*radości, rozkoszy*). Lingwista zauważa ponadto, że abstrakcyjny rzeczownik może także warunkować wystąpienie czasownika *doznać*. Badacz twierdzi, iż orzeczenia peryfrastyczne charakteryzują się specyficznymi właściwościami semantycznymi, przez co nie zawsze mogą być stosowane wymiennie z prostymi orzeczeniami [Loewe 2000: 22 za Bogusławski 1978: 29].

Analityczne konstrukcje werbo-nominalne były także rozpatrywane przez frazeologów. Andrzej Maria Lewicki umiejscowił konstrukcje werbo-nominalne w grupie orzeczeń peryfrastycznych stanowiących związki czasownika i grupy nominalnej z wyrażeniem przyimkowym (np. *wpaść w rozpacz*) lub bez przyimka (np. *dostać szału*), których ekwiwalentem jest czasownik prosty. W celu rozróżnienia grup syntaktycznych i frazeologizmów Lewicki stosował kryterium semantyczne. Badacz uważał, że umożliwia ono zdefiniowanie frazeologizmów jako takich połączeń wyrazów, których całościowe znaczenie nie odpowiada znaczeniom ogółu jego komponentów [Loewe 2000: 21–22 za Lewicki 1976].

Wspomniane kryterium wyróżniania frazeologizmów stosowane przez Lewickiego wywodzi się z tradycyjnego pojmowania związków frazeologicznych, zgodnie z którym są to jednostki zmetaforyzowane, obrazowe, których globalna semantyka nie jest sumą znaczeń ich wziętych z osobna komponentów (np. *втирать очки, бить баклуши*). Owo klasyczne podejście na gruncie radzieckiego językoznawstwa było reprezentowane przez tzw. nurt winogradowski (w terminologii Chlebdy). W teorii Wiktora Władimirowicz Winogradowa znalazło się wiele kryteriów określających jednostkę frazeologiczną, np. odtwarzalność w mowie, stałość składu leksykalnego, metaforyczność, całościowość semantyczna, idiomatyczność, ekspresyjność i inne⁶. Funkcjonują jednak teorie zakładające, iż przedmiotem zainteresowania frazeologii powinny być również dosłowne struktury wielowyrazowe odznaczające się wyłącznie jedną cechą – odtwarzalnością⁷. Jedną z najspójniejszych koncepcji w danym nurcie została wypracowana przez Chlebdę, który włącza do frazeologii również jednostki odtwarzane w konkretnych sytuacjach komunikacyjnych. Chlebda rezygnuje z klasycznej terminologii i określa je mianem *frazemów* (inaczej *reproduktów*), będących jednostkami *frazematyki* i mających na celu nazywanie pojedynczych pojęć, opinii, treści

⁶ Winogradow wśród typów związków frazeologicznych wyodrębnił tzw. *фразеологические сращения, фразеологические единства і фразеологические сочетания*, tj. związki z odpowiednio malejącym stopniem metaforyczności – [por. Виноградов 1977]. Zob. także opis w pracy [Chlebda 2003: 18–20].

⁷ Mowa o tzw. podejściu syntetycznym – w opozycji do analitycznego, reprezentowanego przez szkołę winogradowską – opisywanym już w pracach Jewgienija Dmitrijewicza Poliwanowa w latach 20. XX w. [zob. Chlebda 2003: 23].

[Chlebda 2003: 198]. W terminologii Chlebdy *frazem* pełni rolę hiperonimiczną wobec hiponimów, takich jak frazeologizm, idiom, skrzydlate słowo, przysłowie, wyrażenie funkcyjne, termin złożony, formuła etykietalna bądź slogan [Chlebda 2003: 12]. W poczet wskaźników frazematyczności opolski badacz zalicza czynniki frekwencyjne (rozumiane nie jako wysoką frekwencję grup wyrazowych spowodowaną ich mechanicznym powtarzaniem, lecz jako występowanie relatywnie utrwalonych struktur niosących pewien potencjał treści typowy dla określonych sytuacji w danej wspólnocie komunikacyjnej), formalne (zarówno foniczne, jak i graficzne, np. intonacja, rytmika, tempo, cudzysłów, spacja, kursywa, które zapowiadają odtworzenie pewnej grupy wyrazowej), metatekstowe (fragmenty tekstu wskazujące, że inne fragmenty tekstu są odtwarzane, np. *tak zwany, jak to się mówi, jak to mawiają*), strukturalne (modyfikacje struktury związków wyrazowych umożliwiające dotarcie do ich postaci wyjściowej) i semantyczne (np. niedokładne odtworzenia z modyfikacjami znaczeniowymi) [Chlebda 2003: 228–254].

Mimo iż naszym celem nie jest sporządzenie wyczerpującego opisu dotychczasowych założeń teoretycznych dotyczących omawianego zagadnienia, a jedynie przedstawienie niektórych wybranych problemów ilustrujących jego zakres (jednakowoż nie podejmujemy się ich rozstrzygnięcia w tej pracy), należy zaznaczyć, że podejście prezentowane przez autorów niniejszego artykułu jest w wielu punktach zbieżne z koncepcją Chlebdy, która posiada zarówno wiele zalet, jak i miejsc spornych. Niewątpliwą zaletą teorii Chlebdy jest znaczne rozszerzenie obiektu badań o jednostki do tej pory pomijane, takie jak kolokacje czy związki luźne. Mogły one dzięki temu zostać włączone w zakres zainteresowań glottodydaktyki czy leksykografii przekładowej. W poczet wspomnianych miejsc spornych zaliczyć należy z pewnością wysoką rolę czynnika subiektywnego w badaniach prowadzonych w obrębie danej metody. Bez oparcia się na funkcjonujących klasyfikacjach związków wyrazowych [zob. np. Иорданаская, Мельчук 2007] badacze zdani są wyłącznie na własną kompetencję, której nie można uznać za obiektywne narzędzie weryfikacyjne. Pewne odtwarzalne struktury mogą bowiem być typowe dla idiolektów bądź socjolektów, nie są jednak odtwarzane masowo przez ogół użytkowników języka (por. zagadnienia idiofrazematyki, frazematyki rodzinnej i frazematyki gatunków wypowiedzi w pracy [Chlebda 2003: 208–227]). W takich sytuacjach zwykle niemożliwe jest jednoznaczne określenie statusu danej jednostki.

W celu zachowania klarowności wywodu decydujemy się na konsekwentne operowanie pojęciem odtwarzalności w ślad za Chlebdą, zdając sobie sprawę z istnienia określeń synonimicznych, jak np. *uzus*. W niniejszym artykule konstrukcja odtwarzalna rozumiana będzie za Chlebdą jako jednostka typowa, wyrażająca pewien potencjał treściowy, która jest odtwarzana

z dużym prawdopodobieństwem w określonej sytuacji komunikacyjnej. Dla potrzeb artykułu termin *konstrukcje werbo-nominalne* rozumiemy wyłącznie jako konstrukcje złożone z czasownika i rzeczownika (pomiędzy którymi może znajdować się również przyimek) występujące w tekstach języka naturalnego wyłącznie w następującym szyku:

V + (prep.) + N

Zabieg ten ma jedynie charakter ilustracyjny i jest spowodowany ograniczeniami dotyczącymi objętości artykułu. Opisane w tekście operacje można zastosować także np. do wyodrębnienia konstrukcji (prep.) + N + V, por. np. *баклуши бить, на волю выйти, на родину вернуться*.

2. Korpus

Korpus stanowiący źródło materiału badawczego składa się z artykułów zamieszczonych w rosyjskim piśmie społeczno-politycznym „Новая газета”, które zostały ogłoszone drukiem w 2013 roku. W skład Korpusu włączone zostały artykuły w formie elektronicznej, dostępne na zasadzie *open access* w internetowym archiwum ww. gazety⁸. Archiwum obejmuje 134 numery pisma z 2013 roku⁹ zawierające łącznie 3362 artykuły prasowe. Decyzja o utworzeniu Korpusu składającego się z artykułów zamieszczonych w piśmie „Новая газета” została podyktowana zróżnicowaną tematyką tekstów, dużą liczbą autorów oraz aktualnością poruszanej przez nich problematyki w obrębie dyskursu społeczno-politycznego w okresie, w którym konstruowany był Korpus. Ponadto na wybór źródła miał wpływ fakt powszechnej dostępności gazety.

3. Etapy pracy¹⁰

Narzędziami wykorzystywanymi w procesie ekscerpcji są analizator morfologiczny *LEMMATIZER*¹¹ oraz edytor tekstowy *EditPad Lite* ⁷¹². Proces wyodrębniania interesujących nas konstrukcji możemy podzielić na pięć etapów:

⁸ Źródło elektroniczne: <http://old.novayagazeta.ru/data/2013/> (dostęp 24.11.2014).

⁹ Nie uwzględniono dwunastu numerów, a mianowicie: 1–6, 64, 81, 88, 98, 101, 129, które nie zostały udostępnione online.

¹⁰ Niektóre fragmenty oraz aspekty metodologiczne opisywane są także przez jednego ze współautorów niniejszego tekstu (Yury Fedorushkov) w odrębnej, przygotowywanej obecnie monografii.

¹¹ Program dostępny jest na stronie: <http://aot.ru/download.php> (dostęp 25.11.2014).

¹² Por. stronę: <http://www.editpadlite.com/download.html> (dostęp 25.11.2014). Program obsługuje składnię wyrażen regularnych.

- 1) Automatyczna analiza morfologiczna danych korpusowych przy pomocy programu *Lemmatizer* (zob. opis programu w pracy [Fedorushkov 2005: 34]);
- 2) Eliminacja nieprzydatnych informacji¹³ z uzyskanych wyników;
- 3) Dezambiguacja (ujednoznaczenie) otrzymanych rezultatów. W przypadku lematyzatora (aot.ru) wykorzystanego w niniejszym warsztacie dezambiguacja polega na wyeliminowaniu niewłaściwie przypisanej charakterystyki morfolemmowej i pozostawieniu właściwej – najbardziej odpowiadającej analizie dla wyrazów zawartych w słowniku gramatycznym [Зализняк 1987] – [por. Fedorushkov 2010: 133];
- 4) Automatyczna ekscerpja ankodowa¹⁴ (objaśnienie niżej) wybranych typów konstrukcji werbo-nominalnych przy pomocy edytora tekstowego;
- 5) Manualna weryfikacja wyników ekscerpji.

3.1. Analiza morfologiczna

Proces analizy morfologicznej w programie *LEMMATIZER*¹⁵ polega na przypisaniu każdemu leksemowi Korpusu tzw. *ankodów* (opis – por. niżej), tj. znaczników niosących informację gramatyczną o wyrazach, do których są one przypisywane.

Ankod (ros. *анкод* – skrót od „аношкинский код”) – „уникальный двухбуквенный идентификатор, который соответствует некоторой комбинации значений селективных признаков и граммем. Конечное множество аношкинских кодов исчисляет все встречающиеся в данном языке комбинации морфологических характеристик” [Ножов 2003: 76]. Jest to zatem dwuliterowy (tj. np. **aa**, **аб**, **ав**, **аг** (...); **ба**, **бб**, **бв**, **бг** (...) itd.) identyfikator przypisywany przez program analizowanym wyrazom, kodujący iloczyn kategorii gramatycznych słowa, do którego się odnosi. Ankod jako kod zawierający informację o całym zestawie kategorii gramatycznych wyrazów może służyć jako bardzo przydatny tag w korpusie tekstów, pozwalający wyszukiwać listy wyrazów o dokładnej kategorii gramatycznej.

Wprowadzanie ankodów jest metodą anotowania korpusu, dzięki której szybko można uzyskać informacje o właściwościach gramatycznych znajdu-

¹³ W sprawie dyskwalifikacji oraz eliminacji poszczególnych wyników ekscerpji (np. błędów, okazjonalnych bytów graficznych itp.) – [por. Fedorushkov 2009].

¹⁴ W sprawie ekscerpji bigramowych połączeń atrybutywnych za pomocą ankodów – [por. Fedorushkov, Szutkowski 2010].

¹⁵ Program *LEMMATIZER* został pobrany ze strony www.aot.ru (dostęp 25.02.2004).

jących się w nim wyrazów. Poniżej przedstawiony został fragment wyniku analizy morfologicznej¹⁶:

Я	196 1 RLE AA +?? Яча 151179 0
не	198 2 RLE aa EXPR1 EXPR2 EXPR_NO324 +?? НЕяж 44229 0
знаю	201 4 RLEaa +УнЗНАТЬкб131267 0
,	205 1 PUN
только	207 6 RLE aa EXPR1 EXPR2 EXPR_NO498 +?? ТОЛЬКО яд 44007 0
только	207 6 RLE aa EXPR1 EXPR2 EXPR_NO498 +?? ТОЛЬКО яж 44268 0
догадываюсь	214 11 RLE aa +Уо ДОГАДЫВАТЬСЯ кб 122739 0
,	225 1 PUN
чем	227 3 RLE aa EXPR1 EXPR2 EXPR_NO507 +?? ЧЕМ яд 44012 0
чем	227 3 RLE aa EXPR1 EXPR2 EXPR_NO507 +?? ЧТО щрщс 151138 0
объяснить	231 9 RLE aa +Ул ОБЪЯСНИТЬ ка 150363 0
молчание	241 8 RLE aa +Фа МОЛЧАНИЕ eaer 100941 0
федеральных	250 11 RLE aa +?? ФЕДЕРАЛЬНЫЙ йуййхйч 171219 0
СМИ	262 3 RLE AA +Фа СМИ им 134036 0

3.2. Eliminacja zbędnych informacji i dezambiguacja

Następnym krokiem jest oczyszczenie pliku wyjściowego ze zbędnych danych (np. ciągów 231 9 RLE aa +УлОБЪЯСНИТЬ, tj. lematów oraz logu analizatora). Wszystkie operacje wykonujemy automatycznie przy pomocy filtrów wyrażeń regularnych, wykorzystując mechanizm *search and replace*. Ponadto przeprowadzamy proces dezambiguacji, czyli usuwamy niejednoznaczne interpretacje analizatora morfologicznego. Np. w przypadku kilku rekordów¹⁷ w wyniku dezambiguacji usuwamy interpretację z **rekordu 2**:

rekord 1 только207 6 RLE aa EXPR1 EXPR2 EXPR_NO498 +?? ТОЛЬКОяд 44007 0
rekord 2 только207 6 RLE aa EXPR1 EXPR2 EXPR_NO498 +?? ТОЛЬКОяж 44268 0

¹⁶ Ankody zostały wyróżnione pogrubioną czcionką. Więcej na temat kategorii gramatycznych określanych przez poszczególne ankody zob. w pracy [Ножов 2003].

¹⁷ Por. definicję rekordu – „**Rekord**. Linia (wiersz, ros. *строка*) zakończona znakiem enter (czyli przeniesieniem do nowego wersu). W rekordzie umieszczony jest tekst, który może być pojmowany zarówno jako jeden znak (np. litera lub cyfra), jak i ciąg znaków (np. słowo) lub ciągi znaków (np. frazy), łącznie ze spacjami” [Fedorushkov 2010: 139–140].

Biorąc pod uwagę automatyczny charakter powyższej metody, za „właściwą” uznajemy tylko pierwszą interpretację¹⁸.

Ponadto należy wyeliminować z pliku wyjściowego wszystkie ankody składające się z więcej niż dwóch znaków, pozostawiając jedynie pierwszy z rzędu dwuliterowy ankod (tj. według układu *a fronte*): np. nie **аааг**, lecz **аа**; nie **йуїхйч**, lecz **йу**. Jest to kolejna czynność dezambiguacyjna narażająca nas na utratę części materiału badawczego. Po wykonaniu powyższych operacji fragment listy uzyskuje następującą postać – wyraz + ankod:

rekord 1	Я	ча
rekord 2	не	яж
rekord 3	знаю	кб
rekord 4	,	PUN
rekord 5	только	яд
rekord 6	догадываюсь	кб
rekord 7	,	PUN
rekord 8	чем	яд
rekord 8	объяснить	ка
rekord 9	молчание	еа
rekord 10	федеральных	йу
rekord 11	СМИ	им

¹⁸ Niestety zdarza się, że jest ona błędna, np. wyszukując konstrukcje o strukturze *V + prep.* + *N₄* znajdujemy następującą konstrukcję:

Rekord	Wyraz	Ankod	Poprawność ankodu	Powinno być	Komentarz
1	нашли	кк	tak		
2	на	яв	tak		
3	полу	гг	nie (rodzaj żeński)	ар (rodzaj męski)	Błędna analiza: kategoria rodzaju

Jej obecność jest rezultatem przypisania przez analizator błędnych ankodów do formy *полу*, będącej formą rodzaju męskiego, a nie żeńskiego. Automatyzm dezambiguacji może narażać nas na utratę części materiału leksykalnego. Należy jednak pamiętać, iż zastosowanie metod automatycznych ma na celu przyspieszenie procesu ekstrakcji danych językowych. Przypadki niepoprawnego zinterpretowania danych przez analizator są marginalne i dotyczą jedynie niewielkiej części ekscerpowanego materiału. Proces manualnej eliminacji tych form przeprowadzany jest dopiero podczas weryfikacji wyekscerpowanego materiału badawczego.

Materiał otrzymany w takiej formie poddawany jest ekscerpacji z użyciem składni wyrażeń regularnych¹⁹.

3.3. Ekscerpacja oparta na ankodach i wyrażeniach regularnych

Następnie z tak otagowanej listy możemy przy pomocy filtrów wyrażeń regularnych wyodrębnić interesujące nas jednostki. Dla potrzeb badania będą to konstrukcje składające się z trzech wyrazów (trigramy) o następującej strukturze składniowej:

V + prep. + N₄

Należy podkreślić, iż interesują nas absolutnie wszystkie konstrukcje tego typu znajdujące się w Korpusie. W tym celu należy wyekscerpować konstrukcje ze wszystkimi formami zależnymi czasownika oraz wszystkimi kategoriami gramatycznymi rzeczownika. Używamy następujących filtrów, łączących ankody (przypomnijmy, iż interesują nas jedynie znaczniki składające się z dwóch liter, a nie np. 4-, 6-, 8-literowe, por. wyżej) oraz wyrażenia regularne:

1 rekord	^.*к[а-ь]\$	1 rekord	^.*л[а-ы]\$	1 rekord	^.*м[а-ы]\$	1 rekord	^.*н[а-о]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-з]г\$	3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-з]м\$	3 rekord	^.*[а-ж]н\$
1 rekord	^.*к[а-ь]\$	1 rekord	^.*л[а-ы]\$	1 rekord	^.*м[а-ы]\$	1 rekord	^.*о[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-з]м\$	3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-з]г\$
1 rekord	^.*к[а-ь]\$	1 rekord	^.*л[а-ы]\$	1 rekord	^.*н[а-о]\$	1 rekord	^.*о[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-з]м\$	3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-з]г\$	3 rekord	^.*[а-и]й\$
1 rekord	^.*к[а-ь]\$	1 rekord	^.*м[а-ы]\$	1 rekord	^.*н[а-о]\$	1 rekord	^.*о[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-з]г\$	3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-з]м\$
1 rekord	^.*л[а-ы]\$	1 rekord	^.*м[а-ы]\$	1 rekord	^.*н[а-о]\$	1 rekord	^.*о[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-з]г\$	3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-з]м\$	3 rekord	^.*[а-ж]н\$

¹⁹ Por. opis: „Full regular expressions are composed of two types of characters. The special characters (like the * from the filename analogy) are called *metacharacters*, while everything else are called *literal*, or normal text characters. What sets regular expressions apart from filename patterns is the scope of power their metacharacters provide. Filename patterns provide limited metacharacters for limited needs, but a regular expression «language» provides rich and expressive metacharacters for advanced uses” [Friedl 1997: 5].

1 rekord	^.*п[а-э]\$	1 rekord	^.*с[а-э]\$	1 rekord	^.*у[а-ы]\$	1 rekord	^.*х[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-э]г\$	3 rekord	^.*[а-э]г\$	3 rekord	^.*[а-э]г\$	3 rekord	^.*[а-э]г\$
1 rekord	^.*п[а-э]\$	1 rekord	^.*с[а-э]\$	1 rekord	^.*у[а-ы]\$	1 rekord	^.*х[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-и]й\$
1 rekord	^.*п[а-э]\$	1 rekord	^.*с[а-э]\$	1 rekord	^.*у[а-ы]\$	1 rekord	^.*х[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-э]м\$	3 rekord	^.*[а-э]м\$	3 rekord	^.*[а-э]м\$	3 rekord	^.*[а-э]м\$
1 rekord	^.*п[а-э]\$	1 rekord	^.*с[а-э]\$	1 rekord	^.*у[а-ы]\$	1 rekord	^.*х[а-ы]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-ж]н\$
1 rekord	^.*р[а-ы]\$	1 rekord	^.*т[а-ь]\$	1 rekord	^.*ф[а-ь]\$	1 rekord	^.*ц[а-э]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-э]г\$	3 rekord	^.*[а-э]г\$	3 rekord	^.*[а-э]г\$	3 rekord	^.*[а-э]г\$
1 rekord	^.*р[а-ы]\$	1 rekord	^.*т[а-ь]\$	1 rekord	^.*ф[а-ь]\$	1 rekord	^.*ц[а-э]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-и]й\$	3 rekord	^.*[а-и]й\$
1 rekord	^.*р[а-ы]\$	1 rekord	^.*т[а-ь]\$	1 rekord	^.*ф[а-ь]\$	1 rekord	^.*ц[а-э]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-э]м\$	3 rekord	^.*[а-э]м\$	3 rekord	^.*[а-э]м\$	3 rekord	^.*[а-э]м\$
1 rekord	^.*р[а-ы]\$	1 rekord	^.*т[а-ь]\$	1 rekord	^.*ф[а-ь]\$	1 rekord	^.*ц[а-э]\$
2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$	2 rekord	^.*яв\$
3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-ж]н\$	3 rekord	^.*[а-ж]н\$

Z uwagi na obszerność materiału obejmującego przeszło kilka tysięcy jednostek decydujemy się na wyodrębnienie grupy konstrukcji z przyimkiem *на*, tzn. konstrukcji typu:

$$V + на + N_4$$

Operację tę wykonujemy przy użyciu następującego wyrażenia regularnego:

1 rekord	^.*\$
2 rekord	^на\$
3 rekord	^.*\$

Ogółem wyekscerpowano 1604 konstrukcje trigramowe. Po przeprowadzeniu serii operacji, tj.:

- 1) lematyzacji członu werbalnego,
- 2) wyeliminowaniu powtarzających się form czasownika w części werbalnej,
- 3) odrzuceniu form błędnie zinterpretowanych przez analizator, zawężono liczbę takich konstrukcji do 1089. Otrzymana lista trigramów – to lista potencjalnych reproduktów.

4. Lista potencjalnych reproductów

Z powodu dużej liczby wyekscerpowanych jednostek poniżej zaprezentowano fragment listy obejmujący konstrukcje, w których czasownik rozpoczyna się na literę *в*:

валиться на землю
вдохновить на работу
везти на переработку
вернуть на базу
вернуть на родину
вернуть на службу
вернуть на сцену
вернуться на АЭС
вернуться на вершину
вернуться на Землю
вернуться на лыжню
вернуться на работу
вернуться на родину
вести на песню
вести на экскурсию
вешать на границу
вешаться на шею
взбежать на башню
взвешиваться на долю
взглянуть на Испанию
взглянуть на историю
взглянуть на карту
взглянуть на проблему
взглянуть на экономику
взлетать на пару
взлететь на ветку
взобраться на вершину
взобраться на платформу
взойти на гору
взять на заметку
взять на землянку
взять на неделю
взять на оплату
взять на подготовку
взять на работу
влезть на гору
влиять на культуру
влиять на психику
влиять на ситуацию
влиять на судьбу
влиять на экономику
водить на могилу
водить на прогулку
возвращать на землю
возвращаться на сцену
воздействовать на голову
воздействовать на клетку

воздействовать на науку
воздействовать на психику
возить на учебу
возлагать на Академию
возлагать на Милашину
возлагаться на администрацию
возложить на прокуратуру
войти на кухню
ворваться на базу
ворчать на Генри
вручать на дорожку
вскарabаться на вершину
вспорхнуть на сцену
вставать на сторону
встать на защиту
встать на ниву
встать на позицию
встать на сторону
вступать на территорию
вступить на территорию
входить на территорию
въехать на территорию
выбегать на палубу
выбегать на улицу
выбежать на лыжню
выбежать на улицу
выбрасывать на помойку
выбрасывать на улицу
выбраться на родину
выбросить на помойку
выбросить на сушу
выбросить на улицу
вывезти на родину
вывести на акцию
вывести на орбиту
вывести на сцену
вывести на улицу
выводить на сцену
выводить на улицу
выводиться на периферию
выглядеть на голову
выгонять на улицу
выдавать на покупку
выдвинуть на премию
выделить на Москву
выделить на разработку

выделять на медицину
выделять на систему
выделяться на рекламу
выезжать на работу
выезжать на улицу
выехать на охрану
выехать на полосу
выехать на середину
вызвать на беседу
вызвать на подмогу
выйти на улицу
выйти на волю
выйти на голодовку
выйти на демонстрацию
выйти на замену
выйти на защиту
выйти на лестницу
выйти на Лубянку
выйти на лыжню
выйти на Манежку
выйти на Московскую
выйти на набережную
выйти на палубу
выйти на пенсию
выйти на площадку
выйти на прогулку
выйти на публику
выйти на работу
выйти на свободу
выйти на сцену
выйти на Тверскую
выйти на убийцу
выйти на улицу
выйти на Швецию
выйти на эстраду
выкатываться на улицу
выкинуть на дорогу
выкинуть на помойку
выкинуть на улицу
выпивать на землю
вынести на защиту
вынести на льдину
вынести на обложку
вынести на свободу
вынести на улицу
выносить на защиту
выноситься на публику

выпадать на долю	выскочить на улицу	выходить на линию
выпасть на голову	выслать на столицу	выходить на лыжню
выпасть на долю	выставить на продажу	выходить на палубу
выплеснуть на сцену	выставить на улицу	выходить на пенсию
выплывать на сцену	выставляя на витрину	выходить на площадку
выплыть на Неглинку	выстроить на сдачу	выходить на пробежку
выползть на улицу	высыпать на горку	выходить на промзону
выползти на волю	высыпать на палубу	выходить на работу
выпускать на волю	высыпать на улицу	выходить на раздачу
выпускать на свободу	вытаскивать на улицу	выходить на реку
выпустить на волю	вытянуться на дорогу	выходить на Россию
выпустить на площадку	выходить на авансцену	выходить на свободу
выпустить на свободу	выходить на арену	выходить на службу
вырасти на сумму	выходить на борьбу	выходить на сцену
вырасти на халяву	выходить на дорожку	выходить на Тверскую
вырваться на свободу	выходить на Канаду	выходить на улицу
высадить на льдину	выходить на кухню	вышвыривать на улицу
выскочить на рандеву	выходить на лестницу	вышвырнуть на улицу

Do zlematyzowanych²⁰ konstrukcji werbo-nominalnych warto dodać kwalifikator (**S** lub **F**²¹) określający, czy dana konstrukcja MOŻE mieć status reproduktu: na przykład, czy jest frazemem (odtworzonym w gotowej formie słabo zidiomatyzowanym połączeniem wyrazowym²²), czy jednak pozostaje ona w obrębie „tworzonych zawsze od nowa” okazjonalnych połączeń, które są określane mianem luźnych, tj. należących do **klasy swobodnej**. Ta „kłoda pod nogami” w postaci dylematu merytorycznego – czy zwrot jest zakotwiczony w języku, czy też nie – bywa często omijana przez badaczy. Teoretycznie rzecz ujmując – poza kompetencją językową oraz naukową, tj. indywidualnym doświadczeniem badacza – nie istnieją bardziej subtelne oraz „trafne” techniki wyodrębniania odtwarzalnych wielowyrazowców.

5. Podsumowanie

Opracowaną w ten sposób listę należy następnie poddać weryfikacji. Warto w tym miejscu przytoczyć zdanie Piotra Wierzchonia: „W drugim kroku (tj. po przeprowadzeniu odpowiednich operacji ekscerpcyjnych – Y. F., D. D.)

²⁰ Tzn. sprowadzonych do formy słownikowej, tj. kanonicznej, np. forma czasownika sprowadza się do bezokolicznika: *забил гол* → *забить гол*. O potrzebie (lub jej braku) sprowadzania członów reproduktu do formy słownikowej – por. [Fedorushkov 2010: 66–67].

²¹ Por. wyżej – rozdział 1.

²² Por. definicja kolokacji jako frazeologizmu słabo idiomatycznego: „Коллокации – это слабоидиоматичные [pogr. – Y. F., D. D.] фразеологизмы преимущественно со структурой словосочетания, в которых семантически главный компонент (база) употреблен в своем прямом значении, а сочетаемость со вспомогательным компонентом (коллокатором) может быть задана в терминах семантического класса, но выбор конкретного слова предопределен узусом” [Баранов, Добровольский 2014: 73].

należy – oczywiście manualnie – wybrać te połączenia, które chcemy uznać za reprodukty” [Wierzchoń 2010: 112]. Przez sformułowanie „chcemy uznać” rozumiemy czynności leksykograficzno-weryfikacyjne, takie jak klasyfikacja, sortowanie, weryfikacja pojedynczych obiektów, ich weryfikacja kontekstowa, wielokontekstowa oraz szeroko rozumiana weryfikacja leksykograficzna. Według przyjętej przez nas metodologii dalsza weryfikacja kontekstowa powinna odbywać się w duchu przywoływanej wcześniej, a przyjętej za Chlebą, definicji konstrukcji odtwarzalnej, a więc z uwzględnieniem prawdopodobieństwa wyrażenia jej potencjału treściowego w określonej sytuacji komunikacyjnej.

Zaprezentowana metoda wyodrębniania konstrukcji werbo-nominalnych pozwala uzyskać n-gramowe (o dowolnej liczbie członów) wielowrazowce języka rosyjskiego. Posiada ona jednak swoje ograniczenia: za pomocą filtru dla czasownika, przyimka oraz rzeczownika nie odnajdziemy bowiem frazy *выйти на шумную улицу*. W danej sytuacji filtr wyrażenia regularnego należy poszerzyć o ankody obejmujące kategorię Przymiotnik. Zaprezentowane w artykule rozumienie konstrukcji werbo-nominalnej jest powierzchowne, informatyczne, statyczne²³, ponieważ odnosi się głównie do ścisłego szyku, w jakim występują komponenty danych konstrukcji, a nie do ich właściwości wtórnych (składniowych), tj. przewidujących możliwości alternatywnych dystrybucji w zdaniu/tekście.

Jednostki wyekscerpowane przy wykorzystaniu opisanej metody, po przeprowadzeniu weryfikacji ich odtwarzalności, mogą posłużyć jako materiał do konstruowania siatki haseł słowników, zarówno jednojęzycznych, jak i przekładowych. Dalsze procedury weryfikacyjne oraz szczegółowe rozwiązania znajdujące zastosowanie przy określaniu statusu uzyskanych konstrukcji zostaną opisane w monografii przygotowywanej przez Yurya Fedorushkova, poświęconej zarysowanej tu problematyce.

Bibliografia

- Баранов А. Н., Добровольский Д. О. 2007. *Словарь-тезаурус современной русской идиоматики: около 8 000 идиом современного русского языка*, Москва: Мир энциклопедий Аванта+.
- Виноградов В. В. 1977. *Об основных типах фразеологических единиц в русском языке*, [w:] В. В. Виноградов (red.). *Избранные труды. Лексикология и лексикография*, Москва: Наука, s. 140–161.
- Зализняк А. А. 1987. *Грамматический словарь русского языка: словоизменение: около 100,000 слов*, Москва: Русский язык.

²³ Por. idea definiowania frazemów w sposób statyczny – [Иорданская, Мельчук 2007].

- Захаров В., Хохлова М. 2014. Выделение терминологических словосочетаний из специальных текстов на основе различных мер ассоциации, [w:] *Технологии информационного общества в науке, образовании и культуре: сборник научных статей. Труды XVII Всероссийской объединенной конференции „Интернет и современное общество” (IMS-2014), Санкт-Петербург, 19–20 ноября 2014 г.*, Санкт-Петербург: НИУ ИТМО, s. 290–293.
- Иорданаская Л. Н., Мельчук И. А. 2007. *Смысл и сочетаемость в словаре*, Москва: Языки славянских культур.
- Ножов И. М. 2003. *Морфологическая и синтаксическая обработка текста (модели и программы)*, źródło elektroniczne: <http://www.aot.ru/docs/Nozhov/msot.pdf> (dostęp 12.02.2019).
- Ожегов С. И. 1974. *Лексикология. Лексикография. Культура речи*, Москва: Высшая школа.
- Anusiewicz J. 1978. *Konstrukcje analityczne we współczesnym języku polskim*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Bogusławski A. 1978. *Jednostki języka a produkty językowe. Problemy tzw. orzeczeń peryfrastycznych*, [w:] M. Szymczak (red.), *Z zagadnień współczesnego języka polskiego*, Wrocław: Zakład Narodowy im. Ossolińskich.
- Chlebda W. 2003. *Elementy frazematyki. Wprowadzenie do frazeologii nadawcy*, Łask: Leksem.
- Chlebda W. (red.) 2010. *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*, Opole: Wydawnictwo UO.
- Chlebda W. (red.) 2011. *Na tropach translatów – w poszukiwaniu odpowiedników przekładowych*, Opole: Wydawnictwo UO.
- Eder M. 2016. *Słowa znaczące, słowa kluczowe, słowozbiory – o statystycznych metodach wyszukiwania wyrazów istotnych*, „Przegląd Humanistyczny”, t. 3 (454), s. 31–44.
- Fedorushkov Y. 2005. *Automatyczna analiza morfologiczna jako narzędzie oceny poprawności wyrazów języka rosyjskiego*, „Investigationes Linguisticae”, t. XII, s. 33–43.
- Fedorushkov Y. 2007. *Methods for electronic excerption of new words in Russian*, [w:] B. Lewandowska-Tomaszczyk (red.), *PALC 2007: Practical Applications in Language and Computers. Papers from the International Conference at the University of Łódź, 19–22 April 2007*, Frankfurt am Main: Peter Lang Verlag, s. 163–186.
- Fedorushkov Y. 2008. *Metody ekscerpji neologizmów języka rosyjskiego (na materiale współczesnej prasy rosyjskiej)*, praca doktorska, źródło elektroniczne: <http://hdl.handle.net/10593/16715> (dostęp 12.02.2019).
- Fedorushkov Y. 2009. *Ekscerpja wyrazów cudzojęzycznych z tekstów prasy rosyjskiej*, Warszawa: Wydawnictwo Takt.
- Fedorushkov Y. 2010. *Metody automatyzacji ekscerpji konstrukcji atrybutywnych języka rosyjskiego*, [w:] W. Chlebda (red.), *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*, Opole: Wydawnictwo UO.
- Fedorushkov Y., Szutkowski T. 2010. *Лексико-грамматическая сочетаемость атрибутивных словосочетаний русского языка в контексте методов компьютерной экскерпции*, [w:] R. Guzmán Titado, L. Sokolova, I. Votyakova (red.), *II Международная конференция „Русский язык и литература в международном образовательном пространстве: современное состояние и перспективы”*, Гранада, 8–10 сентября 2010 г., t. II, Granada: Rubiños – 1860, S.A., s. 1565–1569.
- Friedl J. E. F. 1997. *Mastering regular expressions: powerful techniques for Perl and other tools*, Cambridge – Köln – Paris – Sebastopol – Tokyo: O’Reilly & Associates Inc.

- Jędrzejko, E. 1992. *Słownictwo tzw. analityczne w opisie leksykalnym (propozycja opisu i klasyfikacji)*, [w:] A. Markowski (red.). *Opisać słowa*, Warszawa: Elipsa.
- Lewicki A. M. 1976. *Wprowadzenie do frazeologii syntaktycznej. Teoria zwrotu frazeologicznego*, Katowice: Uniwersytet Śląski.
- Loewe I. 2000. *Konstrukcje analityczne w poezji Młodej Polski*, Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Wierzchoń P. 2002. *Automatyzacja ekscerpji definiowanych połączeń wyrazowych. Filtry wyrażień regularnych*, [w:] W. Krzemińska, P. Nowak (red.), *Przestrzenie informacji*, Poznań: Sorus, s. 119–184.
- Wierzchoń P. 2005. *Automatyczne metody ekscerpji neologizmów, czyli słowotwórstwo faktograficzne*, „*Scripta Neophilologica Posnaniensia*”, nr VII, s. 221–240.
- Wierzchoń P. 2008a. *Kotuś. „Verba Polona Abscondita...” (w fotodokumentacji). Szkic lingwochronologiczny. Centuria pierwsza*, Poznań: Instytut Językoznawstwa Uniwersytetu im. Adama Mickiewicza.
- Wierzchoń P. 2008b. *Odtwarzalność w granicach pary przekładowej*, „*Przegląd Rusycystyczny*”, nr 4, s. 111–139.
- Wierzchoń P. 2010. *Pięć bardzo skutecznych (sprawdzonych) sposobów na masowe wyodrębnianie wielowyrzowych segmentów podejrzanych o frazematyczność (czyli reproduktów)*, [w:] W. Chlebda (red.), *Na tropach reproduktów. W poszukiwaniu wielowyrzowych jednostek języka*, Opole: Wydawnictwo UO, s. 87–125.
- Żmigrodzki P. 2000. *Właściwości składniowe analitycznych konstrukcji werbo-nominalnych w języku polskim*, Katowice: Wydawnictwo Uniwersytetu Śląskiego.