

Piotr Wierzchoń  
Poznań

## ODTWARZALNOŚĆ W GRANICACH PARY PRZEKŁADOWEJ

### 1. Wstęp

17 października 2005 roku w „Gazecie Wyborczej” ukazał się następujący tekst:

Skończył się wzrost notowań Donalda Tuska. To wyniki sondażu „Gazety” zrobionego w piątek i sobotę. Większość wywiadów została przeprowadzona w sobotę. Wówczas Polacy mieli za sobą pięć dni „afery wehrmachtowej” — już wiedzieli, że dziadek kandydata PO był wcielony do Wehrmachtu, choć nie jako ochotnik.

Postawmy przed sobą zadanie, by w powyższym fragmencie oznaczyć wszystkie połączenia dwuwyrazowe, ściślej — związki nominalne np. rzeczownika z przymiotnikiem (w ujęciu składniowym — z przydawką w postpozycji). W czasie lektury tekstu odnajdujemy m.in. połączenie *afery wehrmachtowej* (w postaci hasłowej: *afera wehrmachtowa*)<sup>1</sup>. W ten sposób

---

<sup>1</sup> Przymiotnik *wehrmachtowy* przed datą utworzenia kolokacji *afera wehrmachtowa* wchodził w nieliczne związki. Od rzeczownika *Wehrmacht* urabiano raczej przymiotnik *wehrmachtowski*: *wehrmachtowski „drag”* (dekstryna w tabletkach), *wehrmachtowska manierka*, *wehrmachtowska rejestracja*, *wehrmachtowski opel*, *wehrmachtowska Somua S-35*, *wehrmachtowski demobil* czy *wehrmachtowski mundur*. Dokumentacja wyrazu *wehrmachtowy* jest bardzo pouczająca i daje asumpt do kolejnego potwierdzenia stanów wypowiedzianych przez tych badaczy języka, którzy nie absolutyzują zasobów internetowych wystarczającej podstawy prowadzenia badań lingwistycznych. Otóż próżno szukać w zasobach internetowych obfitych potwierdzeń materiałowych omawianego tu przymiotnika przed dniem 17 X 2005, i to około godziny 04:00 GMT. Po 17 października 2005 roku serwisy indeksujące oferują już kilkaset kontekstów z przymiotnikiem *wehrmachtowy*. Identyczną — symptomatyczną semajologicznie — sytuację wzrostu frekwencji (notowaną z dokładnością godzinową) wyrazu obserwowaliśmy w przypadku jednostki *Bieslan*. Do 1 września 2004 roku bardzo trudno poszukiwać potwierdzenia tekstowego tego wyrazu (np. pozaencyklopedycznego). Obecnie w tekstach polskojęzycznych odnajdujemy kilkadziesiąt tysięcy kontekstów dla wyrazu *Bieslan*, w tym także derywaty: *biesłański* (wraz z istotnymi wyrażeniami idiomatycznymi: *syndrom biesłański*, *biesłański horror*, *biesłańska tragedia* itd.).

kończymy pracę z tekstem, której wynikiem jest interesująca nas jednostka wielowyrazowa.

W niniejszym artykule operujemy — za Wojciechem Chlebda — specyficznym pojęciem jednostki wielowyrazowej:

Jednostki te nazywamy frazeologicznymi (idiomatycznymi) w szczególnym sensie tego terminu, odbiegającym od rozumienia standardowego, zbliżonym natomiast do pojęcia „odtworzalnych wielowyrazowców” w rozumieniu A. Bogusławskiego. To złożone z 2–3 (i więcej) wyrazów całości nominatywne, które werbalizują rozmaite pojęcia, sądy, oceny i emocje w sposób do tego stopnia regularny, że zyskały charakter wyrazowych połączeń reprodukowanych [...] (Chlebda, red., 2007: 7).

Metodom poszukiwania wielowyrazowych jednostek językowych poświęcono w literaturze przedmiotu bardzo wiele miejsca<sup>2</sup>, czynność tę można uznać nawet za wyraz mody<sup>3</sup> naukowej. Całość omówień podzielić można na opracowania *stricte* filologiczne — akwantytatywne, oparte na manualnym oglądzie materiału tekstowego (Chlebda 2006, Bańko 2005, Wawrzyńczyk 2006), oraz na ujęcia statystyczne, odwołujące się w swych strukturach do rozważań (następnie do rozwiązań) matematycznych (Yamamoto, Church 2001; por. przypisy 2 i 3). Poza wspomnianymi ujęciami istnieją także prace szczególne, podejmujące tematykę w sposób ponadprzeciętnie wszechstronny, wieloaspektowy, kontaminujące problematykę lingwistyczną, filologiczną, historyczną, kulturową, filozoficzną, obyczajową, realizmową itp. Syntetyzującym osiągnięciem tych badań jest monografia Chlebdy (2005). Jak każdy zwerbalizowany podział w przyrodzie i ten jest wyrazem pewnej propozycji: w jednej z interpretacji wskazanie dowolnej wielowyrazowej jednostki językowej w ramach tradycyjnych ujęć filologicznych już jest czynnością noszącą znamiona świadomości kwantytatywnej. Przemawia za tą perspektywą chociażby parametr zaproponowany przez Chlebda w pionierskim na gruncie frazematyki polskiej opracowaniu (Chlebda 1991), a mianowicie odtwarzalność. Skoro zatem frazematyk wyróżnia spośród innych ciągów jednostkę *jak za króla Ćwieczka*, to zapewne dlatego, że napotkał ten fragment w swych kwerendach więcej niż jeden raz. Lub: napotkał jeden raz, ale już z gotową informacją metajęzykową, iż jest to właśnie jednostka godna wyróżnienia (a więc jej odtwarzalność została stwierdzona już uprzednio). Systematyczne ujęcie frazematyki

<sup>2</sup> Por. m.in.: Barnbrook 2007, Cantos-Gomez, Sánchez 2001, Čermák, Klégr 2004, Claveau, L'Homme 2006, Daudaravičius, Marcinkevičienė 2004, Dayrell 2007, Nesselhauf 2005, Stefanowitsch, Gries 2003, Williams 2002, 2005, Yamasaki 2008; por. omówienie syntetyzujące: Siepmann 2005.

<sup>3</sup> Por. m.in.: Dernowicz 2007, Piskorski, Sydow 2007, Przepiórkowski et al. 2007, Rodríguez et al. 2007, Sagot 2007, Suzuki, Fukumoto 2007, Todirascu et al. 2007.

przez Chlebdę wypływa w pewnej mierze, co pokazuje powyższy cytat, z myśli Andrzeja Bogusławskiego, który wskazał na szczególne miejsce „odtworzalnych wielowyrazowców” w systemie językowym (w tym także w leksykografii<sup>4</sup>; por. Bogusławski 1988)<sup>5</sup>.

W niniejszej publikacji na wstępie omówimy wykorzystywane w praktyce typy metod ekscerpcji wielowyrazowych jednostek języka (celowo w tym miejscu unikamy obciążonych ładunkiem terminologicznym określeń: *wielowyrzowiec*, *syntagma*, *idiom*, *frazemologizm*, *frazem*, *kolokacja* itp.), by w drugiej części tekstu zaprezentować dodatkowy sposób kolekcjonowania nielitycznych w obrębie słownika jednostek wielowyrzowych, oparty na analizie par przekładowych w układzie *a tergo*. Naszym celem jest wstępnie przybliżenie możliwie nieskomplikowanych metod wydobywania z tekstów pisanych tego typu jednostek językowych, by wskazać, w jakim zakresie możliwe są do wykorzystania — poza dotychczas znanymi — dodatkowe praktyczne metody ekscerpcji.

Tradycyjnie edukowany filolog, dysponując wzorem typu:

$$\text{Dice}_{(x,y)} = 2 P_{(x,y)} / (P_{(x)} + P_{(y)}),$$

tak sformułowanej informacji nie użyje w żaden praktyczny sposób, gdyż ścieżka filologa raczej nie przewiduje realizacji celów badawczych z wykorzystaniem powyższej aparatury obliczeniowej; filolog potraktuje tego typu formułę statystyczną indyferentnie. Wobec tego staramy się ukazać, w jakim zakresie możliwa jest droga dotarcia do wielowyrzowych (wstępnie „podejrzanych” o idiomatyczność, *resp.* frazematyczność) jednostek językowych przez badacza nieposługującego się instrumentarium matematycznym. Otóż tylko o pozajęzykowa wiedza dotycząca matur przeprowadzanych w roku 2006 pozwala zinterpretować połączenie wyrazowe *matura Giertycha* jako nazwę rodzaju systemowej abolicji maturalnej (w przeciwieństwie do określenia *matura Giertycha* jako nazwy indywidualnego, osobisto-edukacyjnego doświadczenia ministra czy też obywatela Romana Giertycha). W procesie idiomatycznego (frazematycznego) dekodowania tekstu pomocne będzie użycie cudzysłowu<sup>6</sup>:

<sup>4</sup> Por. Bogusławski, Garnysz-Kozłowska 1979.

<sup>5</sup> Por. także w kwestii określania jednostek języka: Bogusławski 1976a, 1976b, 1987.

<sup>6</sup> Jednym z możliwych sposobów usprawnienia ekscerpcji jest obserwacja jednostek wielowyrzowych ujętych w cudzysłowach. We wcześniejszych pracach (np. Wierzchoń 2005) ukazał się, jak efektywna w wymiarze rejestracyjnym, a także dla potrzeb gramatyczno--opisowych może się okazać skrótowa jednostek cudzysłowowych. Wystarczy w tym miejscu nadmienić, że jednostki w cudzysłowach stanowią nieprzeciętnie atrakcyjny pod względem filologicznym element językowy, zawierający w sobie własność neologiczności, w tym neofrazematyczności. Stąd ekscerpcja jednostki *afera wehrmachtowa* wydaje się naturalną konsekwencją prac rozwi-

Państwowe szkoły dały bowiem do zrozumienia, że absolwenci z „maturą Giertycha” nie mają czego u nich szukać. Giertych nagoni więc klientów prywatnym, nastawionym głównie na robienie kasy uczelniom („Trybuna”, 2006.08.03).

Powracając w tym miejscu do zadań niniejszej publikacji, należy stwierdzić, że jej celem jest przedstawienie serii konkretnych kroków operacyjnych zmierzających do osiągnięcia pewnego zbioru wyników. Pracę tę przedstawimy na przykładzie szczególnego tekstu — *Wielkiego słownika rosyjsko-polskiego z kluczem polsko-rosyjskim* pod red. J. Wawrzyńczyka (dalej: WSRP<sup>7</sup>).

## 2. Strategie poszukiwania połączeń wyrazowych

Istnieją generalnie dwa podejścia badawcze ujmujące jednostki dyskontynuacyjne w centrum analiz<sup>8</sup>. Badacze zorientowani niekomputerowo każdorazowo z założenia wnikają w ontologię sytuacji, analizują konteksty kulturowe, historyczne, relacje metaforyczne, układy nadawczo-odbiorcze, interakcje itp. Druga grupa badaczy — stosująca narzędzia komputerowe — odchodzi od wspomnianych płaszczyzn komunikacyjnych, uznając je za nieuchwytnie w procesie badawczym i koncentruje się na własnościach kwantytatywnych (czy też innych własnościach parametryzowanych — np. gramatycznych lub graficzno-tekstowych: użycie cudzysłowu, zastosowanie markerów tekstowych<sup>9</sup> typu: *zw.*, *so-called*, *так называемый* itp.)<sup>10</sup>. Dla

niętych we wspomnianych publikacjach. Cudzysłów wskazywany był już przez Chlebę (1991) jako jeden z możliwych strukturalnych formalnych wskaźników frazematyczności. Kolejnym krokiem badawczej penetracji stało się przedstawienie owej immanentnie tkwiącej w jednostce językowej frazematyczności w postaci wyrażenia regularnego [a-z]+ [a-z]+ (*notabene* wykorzystywanego regularnie przez informatyków w ich pracy programistycznej). Zdajemy sobie sprawę, że dla grupy programistów formuła „[a-z]+ [a-z]+” jest zapisem trywialnym, z drugiej strony jednak stwierdzenie, że jednym z wyznaczników frazematyczności jest ujęcie cudzysłowowe, dla pewnej grupy językoznawców także nie przedstawi żadnej istotniej wartości.

<sup>7</sup> Skrótom WSPR oznaczamy z kolei *Wielki słownik polsko-rosyjski* pod red. J. Wawrzyńczyka.

<sup>8</sup> Wydaje się, że badacz, który staje przed zadaniem przeglądu literatury dotyczącej ogólnej pojętej frazeologii, frazematyki, idiomatyczności, kolokacyjności etc., chcąc zachować świeżość, dystynktywność własnego ujęcia, winien roztropnie pominąć jeden krok badawczy: zrezygnować z omawiania kilkuset (dane dotyczące liczby opracowań wyszukanych w *BL/LB*, *Linguistic Abstracts*, *SBJO* itp. zbiorach) opracowań dotyczących powyższych zagadnień i skoncentrować się jedynie na krytycznych globalnych ujęciach problemu (w rodzaju Čermak 2001), w których wskazuje się na nieogarnialność masy literatury przedmiotu.

<sup>9</sup> „Ta metoda odznacza się wysoką precyzją (ang. *precision*), ale niską kompletnością (ang. *recall*), to znaczy że wśród otrzymanych za pomocą wyrażen regularnych wyników jest stosunkowo dużo kolokacji idiomatycznych, ale nie wszystkie kolokacje idiomatyczne znajdują się wśród wyników (w wynikach znajdują się tylko kolokacje jawnie definiowane w tekście)” (Buczyński 2004: 31).

<sup>10</sup> Intuicja podpowiada, że podział ten najczęściej ma podłoże pokoleniowe, co oznacza zwrot młodszych badaczy w stronę ujęć komputerowych.

rozwiązań np. anglosaskich<sup>11</sup> możliwe jest dodatkowo poczynienie kolejnego kroku, a mianowicie odwołanie się w opracowaniach do porządku leksykalnego wyrażonego np. poprzez ontologię FrameNet. Na przykład Atkins et al. w oparciu o korpus 14 000 wystąpień słowa *argue* w BNC wskazują, że najbardziej efektywną drogą do uzyskania rzetelnej informacji leksyko-graficznej dotyczącej danego słowa (w ujęciu Atkins et al. 2003 — głównie tzw. *frame evoking words*) jest obserwacja jego otoczenia gramatycznego (Atkins et al. 2003: 252). Autorzy przeprowadzają rozumowanie w oparciu o strukturę FrameNet<sup>12</sup>.

Na przypadku języka polskiego wyróżniającym się ujęciem w zakresie ekscerpcji jednostek wielowyrazowych jest pionierska praca A. Buczyńskiego:

Program, który opisuje niniejsza praca, stosuje do wykrywania kolokacji miary statystyczne (określane też jako probabilistyczne), badając parametry takie jak częstość wystąpienia kolokacji, rozrzut występowania kolokacji pomiędzy teksty wchodzące w skład korpusu lub odchylenie częstości występowania kolokacji od wartości oczekiwanej przy założeniu niezależności słów wchodzących w jej skład (Buczyński 2004: 30).

Poszukując połączeń terminologicznych (operacje dotyczące poszukiwania połączeń terminologicznych znacznie różnią się od czynności, których celem jest określenie zbioru np. frazemów nieterminologicznych typu: *Jaś i Małgosia, przepraszamy za usterki, nie chcem ale muszem*<sup>13</sup> itp.), zazwyczaj operuje się technikami kluczowości (K-score), wykluczenia słów funkcyjnych (stop-list — por. także rozdział 4.3.5. w pracy Buczyński 2004<sup>14</sup>) oraz lematyzacji (Petrovic et al. 2006).

Z powodu fleksyjności języka polskiego i braku homografii dla niektórych form fleksyjnych (*biała niedziela, białej niedzieli, białe niedziele* itd.) metody kwantytatywne typu *T-score*, *Gravity Counts*, *Dice*, *Mutual information* itp. (por. Buczyński 2004, Daudaravičius, Marcinkevičienė 2004) traci na wartości, bowiem ich aplikacja uzależniona jest w pierwszej kolejności od

<sup>11</sup> W przypadku dysponowania odpowiednio opracowaną ontologią języka polskiego badania takie także są możliwe do przeprowadzenia.

<sup>12</sup> W kwestii *Frame semantics* por. m.in.: Atkins et al. 2003, Baker et al. 2003, Fillmore et al. 2003a, b, c, Fontenelle 2000, Hanks 2004, Zaenen 2002; w sprawie nowszych wypowiedzi por.: Boas 2005: 453, Piasecki, Koczan 2007, Vetulani et al. 2007, Bajkowski 2007, Fišer 2007.

<sup>13</sup> Por. z pojęciem „tematu” i „wariacji” w pracy: Chlebda 2005: 27.

<sup>14</sup> „Przy próbach zastosowania programu do celów praktycznych, dość szybko okazało się, że — zwłaszcza przy pytaniach o kolokacje konkretnych słów — znaczącą część wyników stanowią kolokacje z krótkimi, często występującymi słowami — przyimkami, spójnikami i innymi małymi słówkami (określenie jednej z użytkowniczek)” (Buczyński 2004: 64).

zastosowania odpowiednio kompetentnego analizatora morfologicznego lematyzującego formy fleksyjne do kanonicznych postaci słownikowych. W przypadku słownictwa współczesnego nazywającego zjawiska związane z realiami taka preanaliza morfologiczna zostanie znacząco zahamowana. Wydaje się zatem, że niemożliwe jest zaproponowanie jedynej optymalnej metody kwantytatywnej, np. ekscerpcji bigramowej (tj. takiej, w której wykorzystuje się obserwację liczby wystąpień danych 2-połączeń wyrazowych; por. Stubbs 2002, Yamamoto, Church 2001) dla automatycznej ekscerpcji połączeń wyrazowych w językach fleksyjnych. Dlatego, pomiędzy innymi, konieczne jest określenie także takiej metody, w której główny akcent kładzie się na analizę pozakwantytatywnych determinantów kolokacyjności (frazematyczności; Chlebeda 1991: 115). Taką metodą może być, o czym wspomnieliśmy wyżej, np. obserwacja połączeń dwu-, trój- i czterowyrazowych poprzedzonych skrótem *tw.* oraz ograniczonych z prawej strony znakiem niefonogramicznym w terminologii np. Tarajło (1988), tj. kropką, przecinkiem, średnikiem itp. Wielkość korpusu w tego typu badaniach nie ma znaczenia, co jest największą zaletą tej metody.

Na tle wspomnianych ujęć badawczych, jakkolwiek byśmy formułowali ich warunki techniczne (np. w postaci: idiomem jest takie połączenie, które w tekście o mocy 100 000 wyrazów występuje 5 razy), nie istnieje możliwość takiego sformułowania definicji, aby w czasie jej stosowania nie odwoływać się do czynności nieprzewidzianych w sformalizowanej postaci definicji; innymi słowy, w pracy ekscerpta należy się odwołać do intuicji językoznawczej, kompetencji językowej, źródeł encyklopedycznych<sup>15</sup> etc. Rodzi się zatem problem fundamentalny: co faktycznie badamy? Czy w ogóle jesteśmy w stanie generalnie (ogólnie<sup>16</sup>) zdefiniować jednostki, których poszukujemy? Nie są to bowiem najczęściej występujące połączenia wyrazowe.

Degand i Bestgen wskazują:

To test the retrieval technique, we will concentrate first on opaque idioms only, *i.e.* stable, holistic combinations of two or *more* words whose meaning cannot be derived from its component parts [...] (Degand, Bestgen 2003: 25).

<sup>15</sup> Większość użytkowników języka polskiego spotkała się w szkole podstawowej (np. na lekcjach historii) z połączeniem wyrazowym np.: *plan Marshalla*. Bez odwoływania się do źródeł encyklopedycznych nie każdy językoznawca będzie w stanie rozpoznać cechę idiomatyczności takich jednostek, jak np.: *kostka Napiera*, *kryterium Rayleigha*, *liczba Avogadra*, *liczba Fermata*, *liczba Mersenne'a*, *liczba Reynoldsa*, *liczba Wolfa*, *mechanizm Higgsa*, *metal Dore'a*, *most Einsteina-Rosena*, *notacja Backusa-Naura*, *odczyn Herxheimera*, *odczynnik Grignarda*, *owale Cassiniego*, *paradoks Fermiego*, *paradoks Seeliger*, *pas Kuipera*, *pierścienie Einsteina*, *plan Weinera*, *powierzchnia Horgana* czy *prawo Evina*.

<sup>16</sup> W sensie: w ramach językoznawstwa ogólnego.

Wydaje się, że i w naszym przypadku poszukujemy właśnie tych obiektów, o których mówią cytowani tu autorzy. Być może adekwatnym opisem tych jednostek byłoby nazwanie ich tak, jak zrobił to autor *Słownika peryfraz czyli wyrażen omownych* (Bańko 2005). Wówczas jednak ponownie mielibyśmy do czynienia bardziej z jednostkami wyróżnianymi na podstawie intuicji (wiedzy) leksykograficznej, wiedzy pozajęzykowej, a nie na podstawie metod typu: Altenberg 1998, Buczyński 2004, Choueka et al. 1983, Daudaravičius, Marcinkevičienė 2004, De Cock 1998, Degand, Bestgen 2003, Jones, Sinclair 1974, Mason 1997, Smadja 1989, 1993, Stubbs 2002, Yamamoto, Church 2001 (to jedynie wybór z dziesiątek podejść badawczych).

W zakresie opisu terminologii celne ujęcie problemu odczytujemy w pracy Pedersena<sup>17</sup>: „[...] terminological word combinations are characterised by their ability be defined as units” (Pedersen 1995: 62). Faktycznie, połączenia typu: *abolycja podatkowa, abstrakcja ekspresyjna, adopcja anonimowa, adopcja prenatalna, aktywator plazminogenu, aktywna eutanazja, aktywne wakacje, aktywny wypoczynek* itd. zapewne mogłyby stanowić w słowniku języka ogólnego oddzielne (pod)hasła leksykograficzne. Problem upatrujemy jednak w tym, że bezsporne określenie tego, czy *aktywne wakacje* to połączenie odtwarzalne, czy nie, jest zadaniem ze wszech miar skomplikowanym. A jeżeli nie jest to połączenie odtwarzalne, to jakie jest to połączenie? Czy ma ono taki sam status onomazjologiczny, jak połączenia *miłe wakacje* albo *nudne wakacje*? Pytanie to ma też zdecydowany wymiar translatorski.

Na tle tak naszkicowanych komplikacji dotyczących żmudnej kwalifikacji danych połączeń do zbioru „odtworzalnych wielowyrazowców” w niniejszej pracy zaproponować chcielibyśmy osobną metodę prowadzącą do poszerzenia inwentarza idiomów. W tym celu ponownie odwołamy się do pojęcia odtwarzalności. Nie chcąc wprowadzać dodatkowych terminów lingwistycznych, lecz próbując się oprzeć na dotychczasowym dorobku badawczym, reinterpretujemy — na gruncie prezentowanego w niniejszym artykule ujęcia — złożoną przez Chlebę propozycję rozumienia odtwarzalności, którą będziemy tu traktować jako odtwarzalność składnika połączenia wielowyrazowego. W opisywanej metodzie będziemy się posługiwać już opracowanym materiałem słownikowym, a więc nie eksplorujemy zbioru tekstów, lecz ograniczamy się w ekscerpcji jedynie do słownika. Innymi słowy: poszu-

<sup>17</sup> W tym miejscu jednak natychmiast zwraca naszą uwagę kwestia ograniczenia badanego materiału do jednostek terminologicznych. To ujęcie wydaje się z kolei zbyt rygorystyczne, pomija bowiem w rejestracji istotną grupę połączeń nieterminologicznych.

kiwać będziemy połączeń wielowyrazowych w słowniku przekładowym. Zadanie to w pierwszym oglądzie może się wydać czynnością redundantną, oto bowiem dysponujemy już opracowanym leksykograficznie artykułem hasłowym, w którym przedstawia się w ramach np. podhasel zbiory połączeń wyrazowych. Na przykład z artykułu hasła **МОГИЛЬНЫЙ**:

**МОГИЛЬНЫЙ** (напр. тишина) grobowy ♦ **МОГИЛЬНАЯ ПЛИТА** płyta nagrobkowa;  
**МОГИЛЬНЫЙ КРЕСТ** krzyż nagrobny<sup>18</sup>

otrzymujemy dwa połączenia:

- ♦ **МОГИЛЬНАЯ ПЛИТА** płyta nagrobkowa
- ♦ **МОГИЛЬНЫЙ КРЕСТ** krzyż nagrobny

W związku z tak przygotowanym materiałem słownikowym można zadać pytanie: co należy rozumieć przez ekscerpcję par przekładowych? Jaki układ typograficzny tekstu będzie pomocny w ekscerpcji par przekładowych powiązanych tym samym członem — i co za tym idzie: czym jest dwujęzyczny indeks *a tergo*? Komu służy i czemu? W jaki sposób układ ten będzie pomocny w odnajdywaniu połączeń wielowyrazowych z członami o cesze odtwarzalności? Jaki krok naprzód w badaniach nad inwentarzem idiomatykonu czyni dwujęzyczny układ *a tergo*? Jaki rodzaj informacji — niedostępny w bezpośrednim kontakcie ze słownikiem — materiał *a tergo* (i tylko ten materiał) udostępnia?

### 3. Jednostki przekładowe w układzie *a tergo*

W 2004 roku ukazał się *Wielki słownik rosyjsko-polski z kluczem polsko-rosyjskim* pod red. J. Wawrzyńczyka (WSRP). W roku 2006 nakładem wydawnictwa Takt ukazał się *Indeks a tergo jednostek przekładowych Wielkiego słownika rosyjsko-polskiego z kluczem polsko-rosyjskim* (Wierchoń 2006)<sup>19</sup>, który na ponad dwóch tysiącach stron prezentuje ponad 120 tysięcy jednostek przekładowych<sup>20</sup>. Poniżej postaramy się wyliczyć okoliczności,

<sup>18</sup> Cytaty pobrane z WSRP wprowadzamy do niniejszego artykułu czcionką o jednym kroju.

<sup>19</sup> Dalej: *Indeks a tergo*...

<sup>20</sup> Korzystanie z tej pozycji w warunkach przekładowych uzależnione jest od perfekcyjnej, celującej (bliskiej tzw. bilingwalnej) znajomości języków polskiego oraz rosyjskiego. Krąg osób, które z pożytkiem sięgną po *Indeks a tergo*..., zawęży stan znajomości języka rosyjskiego oraz języka polskiego jednocześnie. Ze źródła tego nie skorzysta osoba, która doraźnie poszukuje odpowiedniego ekwiwalentu w celu dokonania tekstowego podstawienia właściwych jednostek. Brak dyrektyw wyboru, informacji



w jakich narzędzie to może się stać pomocą w procesie tworzenia, opracowywania (kolekcjonowania) list wielowrazowych połączeń, m.in. w celu opracowywania idiomatykonów<sup>21</sup> (zob. Chlebda, red., 2006, 2007, 2008).

Przed wszystkim należy ustalić pojęcie pary przekładowej. Pary przekładowe w obrębie naszych badań budujemy na podstawie WSRP. Para zatem to zbiór dwuelementowy (inaczej: relacja) składający się z hasła lub podhasła WSRP (pisane w WSRP czcionką półgrubą) języka rosyjskiego oraz właściwego mu odpowiednika polskiego. Powtórzmy: w skład takiej jednostki wchodzi hasło lub podhasło języka rosyjskiego wraz z jego ekwiwalentem pobranym z języka polskiego. Zwróćmy uwagę, że mówimy tu jedynie o hasle lub podhasle (w sensie: napisu hasłowego), natomiast nie wspominamy o żadnych dodatkowych informacjach leksykograficznych (fonetycznych<sup>22</sup>, morfologicznych czy składniowych)<sup>23</sup>. Utworzenie pary przekładowej o takiej konstrukcji pozwoli w układzie *a fronte* oraz *a tergo* kontrolować przebiegi morfologiczne (grupy wyrazów mające np. to samo zakończenie w sensie morfologicznym<sup>24</sup> lub leksykalnym<sup>25</sup>) zbiorów wyrazów. Dla dalszego operowania materiałem wygodne terminologicznie będzie przypomnienie pojęć *transland* i *translat* (Bogusławski 1988): *translandem* w obrębie tego artykułu nazywać będziemy jednostkę języka rosyjskiego, natomiast *translatem* — jednostkę języka polskiego.

metaleksykograficznych, pragmatycznych, gramatycznych itp. wyklucza komfortowe, tj. niesystematyczne, doraźne korzystanie ze wspomnianego indeksu. Dwujęzyczny układ *a tergo* jednakże powoduje, że pewne systematyczne, zwłaszcza morfologiczne oraz leksykalne obserwacje serii par przekładowych stają się — poprzez opracowany układ — w ogóle możliwe.

<sup>21</sup> W przypadku wykorzystania *Indeksu a tergo* w kontekście opracowania szczególnego dzieła (Chlebda, red., 2006, 2007, 2008) można mówić o wyjątkowo szczęśliwych okolicznościach: *Indeks a tergo...* wprost podpowiada relację przekładową ustaloną dla języka rosyjskiego oraz polskiego.

<sup>22</sup> Miejsce oraz symbol akcentu wyrazów rosyjskich przejmujemy się z WSRP i pozostawiamy bez zmian.

<sup>23</sup> Warto w tym miejscu zaznaczyć, że praktyka ta nie jest własnością jedynie tu prezentowanej metody uzyskiwania par przekładowych. Wyborny *Podręczny idiomatykon polsko-rosyjski* (Chlebda, red., 2006, 2007, 2008) jednostki takie nazywa intuicyjnie i jasno „czystym” ekwiwalentem: „*Idiomatykon* jest **słownikiem przekładowym**, co oznacza, że jego priorytetem informacyjnym jest podawanie drugojęzycznych ekwiwalentów jednostek hasłowych, informacje gramatyczne zaś o tych jednostkach, ich parametry stylistyczne, pragmatyczne i im podobne przesunięte zostały na plan drugi lub są w ogóle nieobecne. Równocześnie jednak staramy się nie tylko pokazać «czysty» ekwiwalent (*translat*) danej jednostki polskiej w języku drugim (rosyjskim), ale też skierować uwagę użytkownika na sposób korzystania z niego w trakcie tworzenia wypowiedzi w języku rosyjskim” (Chlebda, red., 2006: 8; por. także: Chlebda, red., 2007: 7–8).

<sup>24</sup> Dla jednostek jednowyrazowych.

<sup>25</sup> Dla jednostek wielowrazowych.

### 3a. Przygotowanie materiału dla uzyskania par przekładowych

Poniżej zaprezentujemy proces „obliczania” jednostek przekładowych<sup>26</sup>. W praktyce otrzymanie tych jednostek jest możliwe dzięki temu, że WSRP przejawia wysoką kulturę leksykograficzną: na podstawie tego tekstu (słownika) możliwe jest wygenerowanie listy jednostek przekładowych, co nie jest stanem oczywistym, bowiem generowanie list jednostek przekładowych nie było przecież ujęte w programie tego słownika (w czasie, gdy był on tworzony dla celów publikacji drukowanej). Najistotniejsze z naszej perspektywy w WSRP jest użycie ukośnika (ze spacjami) oraz średnika. Znaki te, jako (nigdy) niewchodzące do translatu, doskonale pełnią funkcję metaleksykograficzną, oddzielając od siebie kolejne potencjalne translaty. WSRP ukazuje strukturę haseł o różnej złożoności akapitów, od artykułów hasłowych najprostszyc

**Ааре** Aare

przez:

**Аббас** [-ca] Abbas  
**Аббасиды** [-дов] Abbasydzi  
**аббат** [-та] 1. opat 2. (*настоятель*) przeor  
 [...]

do typu:

**а** *грамматическое слово* 1. (*соединяет предложения или их части, выражая противопоставление, сопоставление, добавление*) а 2. (*после отрицания*) *лечз* ∇ не я, а вы не я, *лечз* вы 3. (*вопросительное слово*) со ∇ не *хотите*, а? не *хচ্ছেе*, со? ♦ **а не то / а то** (*в противном случае*) *во / во* *иначеж*; **а то и** (*выражение, присоединяющее сообщение о чём-либо нежелательном, неожиданном*) а *може* i / *албо* i / *албо* *навет*

WSRP wprowadza w siatce haseł również hasła wielowyrazowe<sup>27</sup>:

**авгиевы конюшни** [авгиевых конюшен] stajnia Augiasza

Tworzenie indeksu rozpoczyna się od usunięcia artykułów odsyłaczowych, które nie zawierają żadnej informacji przekładowej (strona lewa oraz strona

<sup>26</sup> Informacje zawarte w tym podrozdziale stanowią przeredagowaną postać rozdziału *Metoda tworzenia indeksu w układzie WSRP* pracy Wierchoń 2006.

<sup>27</sup> „Dość często w nagłówkach artykułów hasłowych umieszczamy — jako tzw. hasła — nie pojedyncze wyrazy, lecz całe związki wyrazowe (por. np. **анютинны глазки**, **Вербное воскресенье**, **женщина-врач**)” (WSRP: 7-8).

prawa ujmuje jednostki tylko jednego języka — języka rosyjskiego). Są to jednostki o funkcji pomocniczej, których jednym z zadań jest albo lematyzacja (hasłowanie, sprowadzanie do formy tzw. podstawowej, kanonicznej, słownikowej<sup>28</sup>) nieregularnych form gramatycznych (np. **их** → **он**), albo oznaczenie wariantywności sylabicznych pewnych przyimków (np. **ко** → **к**). Dlatego usuwamy artykuły hasłowe typu: **ешь** → **есть**, **дальше** → **далёкий**, **далеко**, **ж частица** → **же** itp.

W przygotowującym materiale znika ujęta w WSRP po niektórych wyrazach hasłowych informacja wymawianiowa. Usuwamy informację dotyczącą konkretnego sposobu wymowy (**авиамоделизм** /дэ/, **АФП** /а-эф-пэ/) czy też wymowy wariantywnej (np.: **АН** /а-эн; ан/).

W celu uzyskania wyłącznie tekstowej informacji przekładowej (bez informacji uzupełniających) zrezygnowano z podawania wskazówek uściślających:

**атомарный** 1. (относящийся к атому) atomowy 2. (дробный, нецелостный) zatomizowany

Zrezygnowano też z podawania wskazówek dotyczących akomodacji składniowych, łączliwości, rekcji itp. WSRP formułuje te informacje w sposób dwojaki, a zatem kursywnie w nawiasach okrągłych, np.:

**артистическая** (комната) garderoba (dla aktorów)

a także w ilustracyjnych frazach oznaczonych znakiem  $\Delta$ <sup>29</sup>, np:

**адриатический** adriatycki  $\Delta$  Адриатическое море Morze Adriatyckie  
**недержание** [-ния] nietrzymanie  $\nabla$  недержание мочи nietrzymanie moczu

Kolejny typ informacji składniowej w WSRP realizowany jest za pomocą zaimkowych oraz niezaimkowych wskaźników<sup>30</sup> typu: *кого*, *чего*,

<sup>28</sup> Por. Szafran 1997.

<sup>29</sup> Ponieważ uwagi formułowane po nabli nie mają charakteru systemowego (są to przykłady wprowadzone z inicjatywy autorów WSRP, stanowią raczej luźne związki wyrazowe), w myśl przyjętej metody poza zainteresowaniem pozostaną pary: *Адриатическое море* Morze Adriatyckie, *недержание мочи* nietrzymanie moczu, gdyż nie stanowią one par przekładowych. Kwestia ta może budzić pewne kontrowersje, bowiem np. para *Адриатическое море* Morze Adriatyckie poprzedzona jest nablą (a nie rombem), w przeciwieństwie np. do uwzględnionej przez nas pary przekładowej *Баренцево море* Morze Barentsa. Decyzja tego typu (konkretnie: umieszczanie przykładów połączeń wyrazowych po nabli) nie należała do piszącego te słowa, gdyż pracowaliśmy na gotowym materiale WSRP.

<sup>30</sup> Informacje te nazywane są także *sygnałami zaimkowymi*, por. np. konstrukcję pary przekładowej w pracy Chlebeda, red., 2006: „Zwroty czasownikowe wyposażone są w syg-

*кого-чего, что, во что, с кем-чем, сделать что, кого, чего, кого-чего, со, в со, з kim-czym, že itp.*, zatem w artykule hasłowym:

**агитировать** [–рую, –руешь] *за кого-что* agitować *за kim-czym, на rzecz кого-чего*

pomija się informację: *за кого-что* oraz: *за kim-czym, на rzecz кого-чего*.

Tzw. dyrektywy wyboru i wskazówki interpretacyjne<sup>31</sup>, zarówno formułowane w języku rosyjskim:

**Ауди** (марка автомобиля, фирма) Audi

jak i w języku polskim:

**аура** [–ры] aura (*настрój; zespół objawów przed atakiem padaczki*)

także nie są częścią hasła lub podhasła, lecz jedynie artykułu hasłowego, dlatego i one podlegają usunięciu. Informacja morfologiczna jako informacja — w myśl przyjętej tu zasady — pozaprzekładowa, nienależąca ani do translatu, ani do translandu, zawarta w nawiasach kwadratowych po hasle, również zostaje wykasowana, np. [–цы] w:

**читательница** [–цы] czytelniczka

Pełne formy fleksyjne ujęte w nawiasy kwadratowe także zostają usunięte:

**Абель** [Абеля] Abel

Jeżeli w polu informacji morfologicznej dodatkowo poza konkretnymi zakończeniami lub całym wyrazami w danej postaci fleksyjnej występują inne informacje (np.: *род. мн.*), również były one usuwane, por. w przykładzie:

**авантюристка** [–тки; *род. мн. –ток*] awanturka

Pomija się także kwalifikatory rodzaju (*м, ж, ср, т, ж, н*) i liczby (*ед., мн., лр, лт*) oraz informację dotyczącą odmienności lub nieodmienności (*одт.*), (*неизм., неодт.*) danego hasła, np.:

nały zaimkowe markujące ich typową łączliwość leksykalną w tekstach, co również ułatwia przekształcanie tych zwrotów w zdania” (s. 9).

<sup>31</sup> „Wskazówki interpretacyjne formułowane są po polsku, występują po polskich odpowiednikach w nawiasach okrągłych i wskazują, o jakie użycie — z dwu czy większej liczby użyć możliwych — chodzi w danym wypadku, tzn. przy danym wyrażeniu rosyjskim” (WSRP: 14).

**адажно** *неизм.* adagio *odm., nieodm.* / adadžio *odm., nieodm.*

Zlikwidowano oznaczenie homonimiczności<sup>32</sup>.

### 3b. Multiplikacja hasel oraz podhasel

Dla dwu rodzajów artykułów hasłowych przewiduje się multiplikację par przekładowych. Sens tej operacji wyjaśnimy na przykładzie konkretnego artykułu WSRP:

**навязывать** [–ваю, –ваешь] 1. narzucać 2. (*на что: привязывать*) wiązać na czym

Jak wskazaliśmy wyżej, po redukcji informacji nieprzekładowych (dodatkowych, np. gramatycznych), uzyskujemy:

**навязывать** 1. narzucać 2. wiązać

Dla tak ustalonego materiału tworzymy poprzez multiplikację dwie pary przekładowe<sup>33</sup>:

**навязывать** narzucać  
**навязывать** wiązać

WSRP przewiduje oprócz wyszczególnionych liczbowo (na zasadzie opozycji semantycznej) odpowiedników (translatów) wprowadzenie także kilku odpowiedników synonimicznych i w operacji tej stosuje się specjalny symbol: ukośnik (/). Symbol ten w zamyśle redakcyjnym winien hierarchizować odpowiedniki, regulując kolejność ich pojawienia się<sup>34</sup>:

**невозвращенец** [–нца] azylant / emigrant (*który odmówił powrotu do kraju*)  
**невозвращенка** [–нки; *род. мн. –нок*] azylantka / emigrantka (*która odmówiła powrotu do kraju*)

<sup>32</sup> „*имя существительное* informacja na szarym tle — tu przykładowa — służy do różnicowania hasel homonimicznych (o identycznej wymowie i pisowni, ale o różnej etymologii i znaczeniu)” (WSRP: 14).

<sup>33</sup> W przypadku gdy w WSRP kolejne odpowiedniki oznaczane są numerami (1., 2., 3. itd.), likwiduje się numerację, traktując każdy odpowiednik jako oddzielny składnik pary przekładowej.

<sup>34</sup> „[...] od adekwatniejszego, w naszym odczuciu, do mniej adekwatnego, a także od częstszego do rzadszego” (WSRP: 13).

Tworzymy zatem osobno (jako osobne pary) cztery pary przekładowe<sup>35</sup>, a zatem:

невозвращенец azylant  
 невозвращенка azylantka  
 невозвращенец emigrant  
 невозвращенка emigrantka

Szczególnie traktujemy informacje ujęte w podhasłach, np. w artykule hasłowym:

читательский czytelniczy ♦ читательский билет karta biblioteczna

wprowadzono podhasłową informację: **читательский билет** karta biblioteczna. Jeżeli po znaku rombu występuje znak średnika, oznacza to, że znak rombu zostaje powtórzony, w związku z czym po średniku należy wprowadzić nową parę przekładową, np.:

♦ **частное** 1. szczegół / szczegóły 2. (результат деления) iloraz; **частным образом** prywatnie

♦ **частное** 1. szczegół / szczegóły 2. iloraz; **частным образом** prywatnie

♦ **частное** 1. szczegół / szczegóły 2. ♦ **частным образом** prywatnie

♦ **частное** 1. szczegół / szczegóły

♦ **частным образом** prywatnie

**частное** szczegół

**частное** szczegóły

**частным образом** prywatnie

Podsumowując, w celu wygenerowania indeksu par przekładowych (transland — translat) likwiduje się wszystkie ciągi w nawiasach, a także same nawiasy: kwadratowe, okrągłe lub ukośne (usuwa się przecinki i inne znaki interpunkcyjne w nawiasach). Likwiduje się ciągi kursywne beznawiasowe. Usunięte zostają ilustracje umieszczane po nabli<sup>36</sup> wraz z nabłą oraz dodatkowo jedna spacja przy niej. Na tym etapie pracy pozostają numery punktów z przynależnymi im kropkami, ukośniki i romby. Romby zamienia się na znaki początku paragrafu, czyli na nowe akapity. Numery punktów (zarówno w części głównej artykułu, jak i w podhasłowej) z kropkami zamienia się na ukośniki — z wyjątkiem wszakże punktów pierwszych (1.),

<sup>35</sup> Oczywiście w takim afrekwencyjnym i aonomazjologicznym układzie informacja dotycząca frekwencji odpowiednika dla konkretnego hasła zostaje całkowicie zamazana.

<sup>36</sup> A to dlatego, że „ilustracje podawane po znaku Δ **uzupełniają** [podkr. — P.W.] opis przekładowy i uwypuklają różne jego aspekty (zwłaszcza składnię)” (WSRP: 13). Usunięcie tych fragmentów WSPR jest najbardziej spornym punktem opracowanej metody.

gdyż te są likwidowane. Tak ujednoliconą strukturę można wykorzystać do zautomatyzowanej operacji dopisywania translandów występujących w nagłówkach artykułów hasłowych — w części głównej artykułów. Indeks nie wskazuje punktów (translatów), haseł homonimicznych (szarego tła), skrótów: *сокр.*, instrukcji wymawianiowych. Powyższe rozważania teoretyczne przedstawimy obecnie w zastosowaniu praktycznym z odpowiednimi przykładami, usuwając z materiału WSRP informację oznaczoną poniżej zaciemnionym polem:

**абиссинец** [–нца; абиссинцы, –нцев] Abisyńczyk

**Абиссиния** [–ний] Abisynia

**абиссинка** [–нки; *род. мн. –нок*] Abisynka

**частный** /сн/ 1. (личный) prywatny 2. (касающийся подробностей) szczegółowy 3. (специфический) szczególnie 4. (отдельный, отдельно взятый) poszczególny / pojedynczy ♦ **частное** 1. szczegół / szczegóły 2. (результат деления) iloraz; **частным образом** prywatnie

Z tak przygotowanego materiału w procesie redukcji pozostawia się jedynie części niezakreślone. Dla przykładu, z artykułu hasłowego **КОЛОТЬ**:

**КОЛОТЬ** [колю, колешь; колотый] 1. kłuć 2. (рассекать) rąbać 3. (убивать) zarzynać 4. (кого-либо; упрекать) dogryzać komu 5. (разбивать) tłuc 6. (орехи) łupać / łuskać ♦ **КОЛОТЫЙ** 1. kłuty 2. (напр. дрова) porąbany; **КОЛОТЬ ГЛАЗА** (попрекать) wykłuwać oczy

do inwentarza-indeksu par przekładowych trafiają po przeprowadzonej operacji następujące jednostki:

**КОЛОТЬ** kłuć

**КОЛОТЬ** rąbać

**КОЛОТЬ** zarzynać

**КОЛОТЬ** dogryzać

**КОЛОТЬ** tłuc

**КОЛОТЬ** łupać

**КОЛОТЬ** łuskać

**КОЛОТЫЙ** kłuty

**КОЛОТЫЙ** porąbany

**КОЛОТЬ ГЛАЗА** wykłuwać oczy

Zreasumujmy: jednostka słownikowa pozostająca w centrum naszego zainteresowania to segment składający się z dwu części: części rosyjskojęzycznej (translandu) oraz części polskojęzycznej (translatu).

**абиджанский** abidżański

**негодование** oburzenie

**в негодование** oburzyć się

transland

translat

Część rosyjskojęzyczną stanowią hasła oraz podhasła (i tylko te segmenty, tj. bez informacji dodatkowych, np. gramatycznych) WSRP, natomiast w części drugiej znajdują się polskojęzyczne odpowiedniki, przy czym zachowana jest bezwyjątkowo zasada, zgodnie z którą jednemu hasłu lub podhasłu języka rosyjskiego odpowiada dokładnie jeden odpowiednik języka polskiego. W przypadku, gdy w oryginalnym materiale WSRP dla jednego hasła ujmowano np. odpowiednik składający się z dwu synonimicznych (oddzielonych ukośnikiem) wyrazów języka polskiego (**негабаритный** niewymiarowy / niestandardowy), hasło języka rosyjskiego zostało powielone (tj. uzyskano: **негабаритный** niewymiarowy oraz **негабаритный** niestandardowy). Sytuacja taka pozwala obserwować w jednej kolumnie informacje grafemowe w postaci maksymalnych (wyznaczanych materiałem WSRP) serii. W seriach tych ujawnia się materiał, który w normalnym trybie użytkowania słownika nie byłby nigdy zauważony i wychwycony; aby bowiem obserwować przykładową serię:

**kanański** ханаанский  
**kananejski** ханаанский  
**albański** албанский  
**Urbanski** Урбанский  
**Urbański** Урбанский  
**durbański** дурбанский  
**kubański** кубанский  
**hawański** гаванский

...

należałoby przeczytać cały słownik WSPR i wypisać z odpowiednich miejsc hasła: *albański, czarencawański, durbański, erewański, hawański, kanański, kananejski, kubański, moldawiański...* lub przeczytać cały słownik WSRP i wypisać z odpowiednich miejsc hasła: *албанский, чаренцаванский, дурбанский, ереванский, гаванский, ханаанский, ханаанский, кубанский, молдаванский...*

W kręgu zainteresowania pozostaje zatem osobliwa (na gruncie tradycyjnej leksykografii) forma danych: forma, która odzwierciedla jedynie minimalne jednostki (tekstowe pary) przekładowe (translandowo-translatowe lub translatowo-translandowe). Zekwiwalentyzowanie rosyjskiego napisu (pod)hasłowego z każdym z możliwych odpowiedników języka polskiego oraz uszeregowanie wyniku tej operacji — czyli serii par — pod względem jedynie alfabetycznym skutkuje uzyskaniem (w masie 120 000 jednostek) obrazu współczesnej ruszczyzny w perspektywie kontrastywnej, z akcentem położonym na grafemowe ujęcie danych w układzie *a tergo*. Przygotowane pary przekładowe pozbawione są jakichkolwiek informacji uzupełniających, dyrektyw wyboru, informacji o znaczeniach, wymowie, wariantowości itp. Przyjrzyjmy się np. parom:



Urbanski Урбанский  
Urbański Урба́нский

Użytkownik dwujęzycznego indeksu *a tergo* musi wiedzieć, że spolonizowana forma rosyjskiego wyrazu *Урба́нский* może (wariantywnie) mieć postać: *Urbański*, w przeciwnym wypadku nie jest bowiem w stanie poprawnie zinterpretować tak przyjętego rozwiązania translacyjnego. Indeks w tej specyficznej postaci może mieć też walor edukacyjny: użytkownik, przygodnie napotykając dwie wyżej wymienione pary, sięgnie do tekstu WSRP, by dostrzec, że:

**Урба́нский** 1. Urbanski 2. (в полонизир. форме) Urbański

Indeks *a tergo* celowo „przemieszał” wszystkie rodzaje par przekładowych (stylistycznie przecież nierzadko zróżnicowane). Szczegółowa obserwacja serii grafemowych pozwoli bez większych problemów wyprowadzić z danego materiału odpowiednie (wedle potrzeb badawczych) wnioski (nas w dalszym ciągu interesować będą jedynie połączenia wielowyrazowe). Reasumując, całość automatycznych działań można w praktyce sprowadzić do operacji na informacjach wyłożonych w poniższym akapicie:

Artykuły zawierające informację przekładową zaczynają się od nagłówka, czyli wyrażenia hasłowego (krócej: hasła), które zawsze jest drukowane czcionką prostą pogrubioną; w ten sam sposób podawane są wyrażenia podhasłowe (po znaku ◊). Odpowiedniki polskie są drukowane czcionką prostą jasną. Kursywa (zawsze czcionką jasną) jest zarezerwowana wyłącznie dla komentarzy odnoszących się czy to do wyrażen rosyjskich, czy to polskich (WSRP: 9).

W indeksie pozostawia się jedynie hasła oraz podhasła wraz z polskimi odpowiednikami. Całość sprowadzono do układu *a tergo* wedle kolejności wyznaczonej alfabetem nierosyjskim.

#### 4. Odtwarzalność w granicach pary przekładowej

Odtwarzalność pojmujemy jako ponadjednokrotne występowanie składnika (np. wyrazu) translandu lub translatu na tej samej, inicjalnej bądź terminalnej, pozycji w parze przekładowej. Aby przekonać się, w jaki sposób rozumiemy odtwarzalność w granicach jednostek przekładowych, wystarczy skierować swoją uwagę na artykuł hasłowy hasła **глаз**, by odnaleźć tam potencjalnie interesujące połączenia wyrazowe. Oto jaki materiał oferuje WSRP w hasle **глаз**:

**глаз** [-за (-зу), o **глазе**, в **глазу**; **глаза**, **глаз**, **глазам**] **1.** oko **2.** (*зрение*) wzrok / **oczy** / oko ♦ **в глаза** (*прямо*) prosto w **oczy**; **в глаза не видеть** nie widzieć na **oczy**; **в глазах мелькает** w oczach się mieni; **в глазах темнеет** robi się ciemno w oczach; **глаз на глаз 1.** (*наедине*) sam na sam **2.** (*без свидетелей*) w cztery **oczy**; **глаз отдыхает на ком—чём** *co* raduje oko / *co* cieszy oko; **глаза** (*зрение*) wzrok / **oczy**; **глаза навькат** / **глаза навькате** wylupiate **oczy**; **глаза не отрываються у кого кто** nie może **oczy** oderwać; **глаза чьи, у кого слипаются** **oczy** się kleją *komi*; **за глаза** *za* **oczy**ta; **куда глаза глядят** dokąd **oczy** poniosa; **на глазах** (*на виду*) w oczach; **с глазу на глаз** (*напр. разговор*) w cztery **oczy**

Oto fragment listy par przekładowych w układzie *a tergo*:

**zapłakane oczy** *наплаканные* **глаза**  
**wytrzeszczone oczy** *выпяченные* **глаза**  
**ma zeza** *косые* **глаза**  
**wywrócić oczy** *завести* **глаза**  
**wylupiate oczy** *рачьи* **глаза**  
**kątem oka** *краем* **глаза**  
**kątem oka** *краешком* **глаза**  
**zaćmiło się w oczach** *затмило* **глаза**  
**oczy zapłonęły** *загорелись* **глаза**  
**zerkać** *скашивать* **глаза**  
**zezować** *скашивать* **глаза**  
**zamydlić oczy** *замазать* **глаза**  
**razić** *резать* **глаза**  
**szczypać w oczy** *щипать* **глаза**  
**zrobić okrągłe oczy** *округлить* **глаза**  
**zaprószyć oczy** *засорить* **глаза**  
**zezować** *косить* **глаза**  
**spojrzeć z ukosa** *покосить* **глаза**  
**spojrzeć zezem** *скосить* **глаза**  
**zerknąć** *скосить* **глаза**  
**wywrócić oczy** *закатить* **глаза**  
**wytrzeszczyć oczy** *выкатить* **глаза**  
**wykluwać oczy** *колоть* **глаза**  
**robić okrągłe oczy** *округлять* **глаза**  
**zaprószać oczy** *засорять* **глаза**

Gdyby zrezygnować z porównywania inwentarza jednostek przekładowych fundowanych podhasłami hasła **глаз**, to poniższy przykład hasła **состояние** (WSRP) powinien lepiej unaocznić postać rzeczy:

**состояние** [-ния] 1. stan 2. (*имущество*) majątek

Układ *a tergo* daje:

**należyty stan** *исправное* **состояние**  
**stan zdenerwowania** *нервное* **состояние**  
**śniączka** *коматозное* **состояние**

**stan podniecenia** возбуждённое состояние  
**stan nieprzytomności** бессознательное состояние  
**stan omdlenia** бессознательное состояние  
**stan półprzytomności** полусознательное состояние

Można sądzić, że przykład hasła **состояние** nie jest reprezentatywne dla ilustracji problemu, gdyż sam artykuł hasłowy nie był rozbudowany o odpowiednie podhasła. Mogłoby to fałszywie podsuwać trop, iż, przykładowo, połączenie *нервозное состояние* (*stan zdenerwowania*) nie jest jednostką na tyle istotną, by wyróżniać ją w ramach osobnego podhasła. Zanalizujemy zatem jeszcze dwa przykłady artykułów hasłowych (**больной** i **лист**), w których utworzono podhasła.

**больной** нездоровый [бoлен, больна] 1. (страдающий, болезненный; ненормальный) chory / niezdrowy 2. (чем: нездоров) chory na co ♦ **болен** (болею) choruję / jestem chory; **больна** (болею) choruję / jestem chora; **больное место** czułe miejsce / słaba strona; **больной вопрос** paląca kwestia; у кого **больной** (желудок, сердце, почки и т. п.) jest chory na co ∇ у отца **больной** желудок ojciec jest chory na żołądek; у неё **больное** сердце jest chora na serce

Indeks *a tergo*... (z całego WSRP) notuje:

**chronicznie chory** хронический больной  
**chory na epilepsję** эпилептический больной  
**psychicznie chory** психический больной  
**obłożnie chory** лежачий больной  
**chory zakaźnie** заразный больной  
**chory na białaczkę** лейкозный больной  
**chory na grypę** гриппозный больной  
**chory na tyfus** тифозный больной  
**chory na dur brzuszny** брюшнотифозный больной  
**chory na ospę** оспенный больной  
**chory na cholere** холерный больной  
**niebezpiecznie chory** опасный больной  
**chory na płuca** лёгочный больной  
**chory na wątrobę** печёночный больной  
**wątrobiarz** печёночный больной  
**chory na świerzb** чесоточный больной

**лист** *напр. бумаги* [-та; -ты, -тов] 1. arkusz 2. (*напр. титульный*) karta z ♦ **в ... листа** (два, три, четыре: *objęmom*) / **в ... листов** (пять, шесть i т. д.: *objęmom*) o objętości ... arkuszy (dwóch, trzech, czterech, pięciu, sześciu itd.); **с листа** (*напр. неть, переводить*) z kartki

Fundowany c a ł y m (a-я) materiałem WSRP indeks jednostek przekładowych notuje pary:

**karta obiegowa** обходной лист  
**lista skladek** подписной лист  
**lista składowa** подписной лист  
**listek bobkowy** лавровый лист  
**liść bobkowy** лавровый лист  
**glejt** охранный лист  
**list żelazny** охранный лист  
**ankieta** опросный лист  
**kwestionariusz** опросный лист  
**zwolnienie lekarskie** больничный лист  
**tytuł wykonawczy** исполнительный лист  
**strona tytułowa** титульный лист

Istnieją takie konstrukcje „wielowyrazowców”, które składają się z wyrazów występujących we fleksyjnej formie zależnej (*casus obliqui*), nienotowanej w artykule hasłowym, por. np. połączenia z formą fleksyjną *взглядом*, a więc formą zapewne przyporządkowaną hasłu słownikowemu **ВЗГЛЯД**:

**взгляд** [-да] 1. spojrzenie 2. (*мнение*) pogląd ♦ **на ... взгляд** (*мой, твой, её* i t.p.: *по чьему-либо мнению*) ... zdaniem (*моим, твоим, jej* itp.)

*Indeks a tergo...* dostarcza pary:

**ogarniać spojrzeniem** окидывать взглядом  
**ogarniać wzrokiem** окидывать взглядом  
**unicestwić spojrzeniem** испепелить взглядом  
**ogarnąć spojrzeniem** окинуть взглядом  
**ogarnąć wzrokiem** окинуть взглядом  
**kierować wzrok na** прицеливаться взглядом  
**skierować wzrok na** прицелиться взглядом  
**spojrzeć z ukosa** покоситься взглядом

**рука** [-ки, вин. руку; руки, рук, рукам] 1. ręka 2. (*кисть*) dłoń 3. (*почерк*) charakter pisma / pismo 4. (*разг.: протекция*) protekcja / plecy ♦ **на руках** 1. (*в наличии*) przy sobie 2. (*на попечении*) na opiece 3. (*в ведении*) na głowie; **обеими руками** oburącz; **от руки** odręcznie / odręczny; **пожать руку** uściśnąć dłoń; **по обе руки** po obu stronach; **рука об руку** ręka w rękę; **руки-крюки** (*неумелые руки*) dwie lewe ręce; **руки по швам!** baczność!; **с руки** (*подходит*) na rękę

W Indeksie a tergo natomiast:

**рęce do pracy** рабочие руки  
**spracowane ręce** натруженные руки  
**machać rękami** размахивать руки  
**wykręcać ręce** закручивать руки  
**wykręcać ręce do tyłu** заламывать руки

**zalamywać ręce** ломать руки  
**zacierać ręce** потирать руки  
**wyciągać ręce** простирать руки  
**zatrzeć ręce** потереть руки  
**wyciągnąć ręce** простереть руки  
**urobić sobie ręce** натрудить руки  
**urobić sobie ręce po łokcie** натрудить руки  
**wykręcić ręce do tyłu** заломить руки  
**wykręcać ręce** крутить руки  
**wykręcić ręce** закрутить руки

Zastanówmy się w tym miejscu nad jeszcze jednym problemem. Otóż można powiedzieć, że badacz nie musi w swej ekscerpji opierać się jedynie na materiale WSRP, połączeń wielowyrazowych można bowiem poszukiwać także (komplementarnie) w WSPR. Skoro więc WSRP nie oferuje pełnego (jakkolwiek ową pełność byśmy dookreślali) zestawu połączeń wielowyrazowych danego hasła, to zbiór ten prawdopodobnie odnajdziemy w słowniku o kierunku polsko-rosyjskim. Zanalizujmy dla przykładu hasło **mleko** w WSPR oraz **молоко** w WSRP.

**молоко** [-ка] mleko

Hasło WSRP **молоко** w ogóle nie zawiera połączeń wyrazowych z wyrazem *молоко*. Wobec tego badacz zainteresowany postaciami *mleka* ma przed sobą wnikliwą, detaliczną lekturę tysiąca kilkuset (czyli wszystkich) stron WSRP. Naturalnie, uprzednio dokonujemy sprawdzenia hasła WSPR, tj. **mleko**:

**mleko** молоко ♦ **mleka pełnotłustego** (*np. białko*) цельномолочный; **mleko acydofilne** ацидофилин; **mleko odtłuszczone** обрат; **mleko pełnotłuste / mleko nieodciągane** цельное молоко; **mleko prosto od krowy** парное молоко; **mleko skondensowane** сгущёнка *разг.*; **mleko w proszku** сухое молоко; **na mleku** (przyrządzony) молочный

Rzeczywiście, kwerenda ta okazuje się nadzwyczaj efektywna. Wypisujemy z WSPR: **mleko acydofilne** ацидофилин; **mleko odtłuszczone** обрат; **mleko pełnotłuste** цельное молоко; **mleko nieodciągane** цельное молоко; **mleko prosto od krowy** парное молоко; **mleko skondensowane** сгущёнка; **mleko w proszku** сухое молоко. Ta metoda nie pozwoli badaczowi jednak odnaleźć *zsiadłego mleka* (*квашеное молоко*), częściej w przyrodzie (i w tekstach!) występującego niż *mleko acydofilne*<sup>37</sup>. Dodatkowo, w WSPR

<sup>37</sup> Por. natomiast listę: *mleko kwaśne, mleko pasteryzowane, mleko pełne, mleko skondensowane, mleko UHT, mleko w proszku* ujętą w pracy Chlebda, red., 2006: 24. Warto

notujemy parę: **mleko odtłuszczone** обрат, lecz w WSRP odnajdujemy bardzo interesującą wielowyrazową rosyjską parę:

**снятой** odciągany ♦ **снятое молоко** odtłuszczone mleko

*Indeks a tergo...* tymczasem łączy zestawienie „mlek” w jednym miejscu:

**kwaśne mleko** квашеное молоко  
**zsiadłe mleko** квашеное молоко  
**mleko prosto od krowy** парное молоко  
**mleko nieodciągane** цельное молоко  
**mleko pełnotłuste** цельное молоко  
**odtłuszczone mleko** снятое молоко  
**mleko w proszku** сухое молоко

Naturalnie, serie mające to samo (trzykrotnie odtwarzalne) zakończenie *a tergo* mogą być mniej liczne. Oto przykłady trójparowe:

**program nauczania** учебная программа  
**program koncertu** концертная программа  
**program edukacyjny** обучающая программа

**naga prawda** голая истина  
**banal** азбучная истина  
**truizm** избитая истина

**wojna domowa** гражданская война  
**wojna ojczyźniana** отечественная война  
**wojna w obronie ojczyzny** отечественная война

**kasa wymiany pieniędzy** разменная касса  
**kasa oszczędności** сберегательная касса  
**kasa oszczędnościowa** сберегательная касса

**rozdziierać serce** резать сердце  
**z ciężkim sercem** скрепя сердце  
**zacinawszy zęby** скрепя сердце

**jak z igły** с иголки  
**prosto spod igły** с иголки  
**prosto z igły** с иголки

**no właśnie** вот именно  
**otóż to** вот именно  
**otóż właśnie** вот именно

**cichaczem** под сурдинку  
**cichcem** под сурдинку  
**z cicha** под сурдинку

itd.

---

tu zaznaczyć różny w stosunku do WSRP przekład translandu *mleko kwaśne* (*nota bene* z prepozycyjną przydawką w WSRP).

Nawiążmy jeszcze raz do postaci układu *a fronte* par przekładowych. Dwujęzyczny indeks *a tergo* jest przedsięwzięciem nietypowym pod względem konstrukcji, zakłada bowiem dwujęzyczność jego użytkownika; obcowanie z nim wymaga pewnej świadomości lingwistycznej. Powyżej prezentowaliśmy jedynie układ *a tergo*. Konstrukcja jednak samej pary przekładowej jako pozbawionej np. gramatycznych informacji o relacji pomiędzy translandem i translatem umożliwia także masowy, przez to niepowtarzalny, wgląd w materiał danego słownika w układzie *a fronte*, ujawniając treści niedostępne w procesie klasycznej lektury samego WSRP; por. hasło **żołnierz**:

**żołnierz** 1. солдат 2. (*uczestnik bitwy*) боец 3. (*wojskowy: określenie uroczyste*) воин ♦ **żołnierz artylerii przeciwlotniczej** зенитчик; **żołnierz batalionu medyczno-sanitarnego** медсанбатовец; **żołnierz NATO** НАТОвец *пазг.*; **żołnierz oddziałów nieliniowych** нестроевик *пазг.*; **żołnierz piechoty zmotoryzowanej** мотопехотинец; **żołnierz straży granicznej** пограничник

Oto pary przekładowe utworzone na podstawie materiału WSRP:

**żołnierz artylerii przeciwlotniczej** зенитчик  
**żołnierz batalionu medyczno-sanitarnego** медсанбатовец  
**żołnierz jednostki karnej** штрафник  
**żołnierz jednostki karnej** штрафной  
**żołnierz kontraktowy** контрактник  
**żołnierz NATO** НАТОвец  
**żołnierz ochrony pogranicza** пограничник  
**żołnierz oddziałów nieliniowych** нестроевик  
**żołnierz oddziału karnego** штрафник  
**żołnierz oddziału karnego** штрафной  
**żołnierz piechoty zmotoryzowanej** мотопехотинец  
**żołnierz pierwszego rocznika** первоодок  
**żołnierz sił pokojowych** миротворец  
**żołnierz straży granicznej** пограничник  
**żołnierz wojsk rakietowych** ракетчик

Stworzenie indeksu par pozwoliło wydobyć z WSRP dodatkowo 9 jednostek wielowyrazowych.

## 5. Podsumowanie

Pośród wielu typologii, klasyfikacji itp. wyróżnić można dwa nurty badań językoznawczych: teoretyczny (przykłady ilustrujące określone tezy dobierane są okazyjnie, „z głowy”, „by hand” — Meurers 2005: 1621 — lub stanowią nikły fragment tekstów już wyprodukowanych) i materiałowy

(każdy przykład ilustrujący dany etap opisu lingwistycznego pobierany jest z uprzednio przygotowanego materiału, zazwyczaj określanego jako korpus).

Przedkładane w tym miejscu opracowanie winno było ukazać, w jakim zakresie możliwe jest połączenie komputerowych rozwiązań automatyzujących proces badawczy z klasyczną, detaliczno-filologiczną analizą uzyskanych wyników. W gruncie rzeczy wychodzimy z założenia, że metody komputerowe winny w pierwszym kroku operacyjnym radykalnie optymalizować (także poprzez odpowiednią wizualizację wyników) postać materiału poddawanego w dalszej kolejności oglądowi manualnemu. Istotą naszych rozważań było zaprezentowanie:

- a) praktycznej metody konstrukcji pary przekładowej — jednostki słownika dwujęzycznego składającej się z *translandu* oraz *translatu*,
- b) uszeregowanie otrzymanych par w układzie *a tergo*,
- c) zastosowanie cechy odtwarzalności w odniesieniu do składników par przekładowych,
- d) wyznaczenie zbiorów jednostek wielowyrazowych.

Wykrywanie jednostek wielowyrazowych polega głównie na śledzeniu ich odtwarzalności w zbiorze tekstów. Jak jednak należy postępować, gdy obserwowany materiał jest tekstem szczególnym, tj. słownikiem przekładowym, w którym wspomniane jednostki notowane są jednokrotnie? W niniejszym tekście wskazaliśmy na odmienne od przedstawionego w rozdziałach 2. i 3. rozumienie pojęcia odtwarzalności, relatywizując tę własność do składników (np. wyrazów) par przekładowych. Jednym z narzędzi ułatwiających z kolei obserwację takich połączeń, zwłaszcza dla indeksu par przekładowych, jest właśnie układ *a tergo*.

## LITERATURA

- J. Aarts, W. Meijs, red.: *Theory and Practice in Corpus Linguistics*. Amsterdam: Rodopi 1990.
- S. Allén: *On phraseology in lexicology*. „Cahiers de Lexicologie“ 1976, nr 29 (2), s. 83–90.
- B. Altenberg: *On the phraseology of spoken English: the evidence of recurrent word-combinations*. W: A.P. Cowie, red., 1998, s. 101–122.
- S. Atkins, M. Rundell, H. Sato: *The Contribution of Framenet to Practical Lexicography*. „International Journal of Lexicography“ 2003, nr 16, s. 333–357.
- C.F. Baker, Ch.J. Fillmore, B. Cronin: *The Structure of the Framenet Database*. „International Journal of Lexicography“ 2003, nr 16, s. 281–296.
- M. Bańko: *Słownik peryfraz czyli wyrażen omownych*. Warszawa: Wydawnictwo Naukowe PWN 2005.



- G. Barnbrook: *Sinclair on collocation*. „International Journal of Corpus Linguistics” 2007, nr 12 (2), s. 183–199.
- M. Benson: *The structure of the collocational dictionary*. „International Journal of Lexicography” 1989, nr 2 (1), s. 1–14.
- M. Benson: *Collocations and general-purpose dictionaries*. „International Journal of Lexicography” 1990, nr 3 (1), s. 23–34.
- H.C. Boas: *Semantic Frames as Interlingual Representations for Multilingual Lexical Databases*. „International Journal of Lexicography” 2005, nr 18, s. 445–478.
- M. Bobran, red.: *Problemy językoznawstwa porównawczego*. Rzeszów: Wyd. Uczelniane WSP 1976.
- A. Bogusławski: *O zasadach rejestracji jednostek języka*. „Poradnik Językowy” 1976 (a), nr 8, s. 356–364.
- A. Bogusławski: *Zagadnienia jednostek przekładowych*. W: M. Bobran, red., 1976 (b), s. 41–62.
- A. Bogusławski: *Obiekty leksykograficzne a jednostki języka*. W: Z. Saloni, red., 1987, s. 13–34.
- A. Bogusławski: *Dwujęzyczny słownik ogólny. Projekt instrukcji z komentarzami*. W: Z. Saloni, red., 1988, s. 19–65.
- A. Bogusławski: *Uwagi o pracy nad frazeologią*. W: Z. Saloni, red., 1989, s. 13–30.
- A. Bogusławski, T. Garnysz-Kozłowska: *Addendum to Polish phraseology. Introductory issue. Addenda do polskiej frazeologii. Zeszyt wstępny*. Edmonton 1979.
- K. Bolton, G. Nelson, J. Hung: *A corpus-based study of connectors in student writing*. „International Journal of Corpus Linguistics” 2002, nr 7 (2), s. 165–182.
- E. Breidt, F. Segond, G. Valetto: *Local grammars for the description of multi-word lexemes and their automatic recognition in texts*. W: F. Kiefer, G. Kiss, J. Pajzs, red., 1996, s. 19–28.
- A. Buczyński: *Pozyskiwanie z Internetu tekstów do badań lingwistycznych*. Warszawa: Instytut Informatyki UW [praca magisterska napisana pod kierunkiem dra hab. Janusza S. Bienia] 2004.
- Y. Choueka, S.T. Klein, E. Neuwitz: *Automatic retrieval of frequent idiomatic and collocational expressions in a large corpus*. „Journal of the Association for Literary and Linguistic Computing” 1983, nr 4 (1), s. 34–38.
- P. Cantos-Gomez, A. Sánchez: *Lexical Constellations: What Collocates Fail to Tell*. „International Journal of Corpus Linguistics” 2001, nr 6 (2), s. 199–228.
- F. Čermák: *Substance of idioms: perennial problems, lack of data or theory?* „International Journal of Lexicography” 2001, nr 14 (1), s. 1–20.
- F. Čermák, A. Klégr: *Extending collostructional analysis: A corpus-based perspective on ‘alternations’*. „International Journal of Corpus Linguistics” 2004, nr 9 (1), s. 97–129.
- W. Chlebda: *Elementy frazematyki. Wprowadzenie do frazeologii nadawcy*. Opole: WSP 1991.
- W. Chlebda: *Szkice o skrzydlatych słowach. Interpretacje lingwistyczne*. Opole: Wyd. Uniwersytetu Opolskiego 2005.
- W. Chlebda, red.: *Podręczny idiomatykon polsko-rosyjski, z. 1. próbny*. Opole: Wyd. Uniwersytetu Opolskiego 2006.
- W. Chlebda, red.: *Podręczny idiomatykon polsko-rosyjski, z. 2*. Opole: Wyd. Uniwersytetu Opolskiego 2007.
- W. Chlebda, red.: *Podręczny idiomatykon polsko-rosyjski, z. 3*. Opole: Wyd. Uniwersytetu Opolskiego 2008.
- K.W. Church, P. Hanks: *Word association norms, mutual information, and lexicography*. „Computational Linguistics” 1990, nr 16 (1), s. 22–29.

- V. Claveau, M.-C. L'Homme: *Discovering and organizing noun-verb collocations in specialized corpora using inductive logic programming*. „International Journal of Corpus Linguistics” 2006, nr 11 (2), s. 209–243.
- J. Clear: *From Firth principles: Computational tools for the study of collocation*. W: M. Baker, F. Gill, E. Tognini-Bonelli, red., 1993, s. 271–292.
- A.P. Cowie, red.: *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press 1998.
- A.P. Cowie: *Phraseology and corpora: some implications for dictionary-making*. „International Journal of Lexicography” 1999, nr 12 (4), s. 307–323.
- V. Daudaravičius, R. Marcinkevičienė: *Gravity counts for the boundaries of collocations*. „International Journal of Corpus Linguistics” 2004, nr 9 (2), s. 321–348.
- C. Dayrell: *A quantitative approach to compare collocational patterns in translated and non-translated texts*. „International Journal of Corpus Linguistics” 2007, nr 12 (3), s. 375–414.
- S. De Cock: *A recurrent word combination approach to the study of formulae in the speech of native and non-native speakers of English*. „International Journal of Corpus Linguistics” 1998, nr 3 (1), s. 59–80.
- L. Degand, Y. Bestgen: *Towards Automatic Retrieval of Idioms in French Newspaper Corpora*. „Literary and Linguistic Computing” 2003, nr 18 (3), s. 249–259.
- W. Dernowicz: *Automatic Acquiring of Semantic Relations From Text Collection*. W: Z. Vetulani, red., 2007, s. 310–312.
- Ch.J. Fillmore, Ch.R. Johnson, M.R.L. Petruck: *Background to Framenet*. „International Journal of Lexicography” 2003, nr 16, s. 235–250.
- Ch.J. Fillmore, M.R.L. Petruck: *Framenet Glossary*. „International Journal of Lexicography” 2003, nr 16, s. 359–361.
- Ch.J. Fillmore, M.R.L. Petruck, J. Ruppenhofer, A. Wright: *Framenet in Action: The Case of Attaching*. „International Journal of Lexicography” 2003, nr 16, s. 297–332.
- T. Fontenelle: *A Bilingual Lexical Database for Frame Semantics*. „International Journal of Lexicography” 2000, nr 13, s. 232–248.
- S. Glucksberg, M.S. McGlone, M.E. Brown: *Conceptual metaphors are not automatically accessed during idiom comprehension*. „Memory and Cognition”, 1993, nr 21, s. 711–719.
- St. Th. Gries, B. Hampe, D. Schönefeld: *Converging Evidence on the Association of Verbs and Constructions*. „Cognitive Linguistics” 2005, nr 16 (4), s. 635–676.
- P. Hanks: *The Syntagmatics of Metaphor and Idiom*. „International Journal of Lexicography” 2004, nr 17, s. 245–274.
- R.R.K. Hartmann, red.: *Dictionaries and their Users*. Exeter: University of Exeter 1979.
- S. Jones, J.M. Sinclair: *English lexical collocations*. „Cahiers de Lexicologie” 1974, nr 24, s. 15–61.
- G. Kjellmer: *Patterns of collocability*. W: J. Aarts, W. Meijs, red., 1990, s. 163 — 178.
- B. Lewandowska-Tomaszczyk, P. Melia, red.: *PALC'97: International Conference on Practical Applications in Language Corpora, Łódź, Poland, 10–14 April 1997*. Łódź: Łódź University Press 1997.
- O. Mason: *The Weight of Words: An Investigation of Lexical Gravity*. W: B. Lewandowska-Tomaszczyk, P. Melia, red., 1997, s. 361–375.
- D. Meurers: *On the use of electronic corpora for theoretical linguistics. Case studies from the syntax of German*. „Lingua” 2005, nr 115 (11), s. 1619–1639.
- N. Nesselhauf: *Structural and Functional Properties of Collocations in English. A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence*. „International Journal of Corpus Linguistics” 2005, nr 10 (2), s. 266–270.

- J. Pedersen: *The identification and selection of collocations in technical dictionaries*. „Lexicographica” 1995, nr 11, s. 60–73.
- S. Petrovic, J. Snajder, B. Dalbelo-Basic, M. Kolar: *Comparison of collocation extraction measures for document indexing*. „Journal of Computing and Information Technology” 2006, nr 14:4, s. 321–327.
- J. Piskorski, M. Sydow: *Usability of String Distance Metrics for Name Matching Tasks in Polish*. W: Z. Vetulani, red., 2007, s. 403–407.
- A. Przepiórkowski, Ł. Degórski, B. Wójtowicz: *On the evaluation of Polish definition extraction grammars*. W: Z. Vetulani, red., 2007, s. 473–477.
- M.B. Rodríguez, E.D. Noya, P.G. Otero, M.L. Martínez, E.M.M. Mato, G.Rojo, M.P. Santalla del Río, S.S. Docío: *A Corpus and Lexical Resources for Multi-word Terminology Extraction in the Field of Economy in a Minority Language*. W: Z. Vetulani, red., 2007, s. 359–363.
- B. Sagot: *Building a morphosyntactic lexicon and a pre-syntactic processing chain for Polish*. W: Z. Vetulani, red., 2007, s. 423–427.
- Z. Saloni, red.: 1987. *Studia z polskiej leksykografii współczesnej II*. Białystok: Dział Wydawnictw Filii UW 1987.
- Z. Saloni, red.: 1988. *Studia z polskiej leksykografii współczesnej*. Wrocław: ZniO 1988.
- Z. Saloni, red.: *Studia z polskiej leksykografii współczesnej III*. Białystok: Dział Wydawnictw Filii UW 1989.
- D. Siepmann: *Collocation, Colligation and Encoding Dictionaries*. Part I: *Lexicological Aspects*. „International Journal of Lexicography” 2005, nr 18, s. 409–443.
- F. Smadja: *Co-occurrence: the missing link*. „Literary and Linguistic Computing” 1993, nr 4 (3), s. 163–168.
- F. Smadja: *Retrieving collocations from text: Xtract*. „Computational Linguistics” 1993, nr 19, s. 143–177.
- A. Stefanowitsch, S. Gries: *Collostructions: Investigating the interaction of words and constructions*. „International Journal of Corpus Linguistics” 2003, nr 8 (2), s. 209–243.
- M. Stubbs: *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers 2001.
- M. Stubbs: *Two quantitative methods of studying phraseology in English*. „International Journal of Corpus Linguistics” 2002, nr 7 (2), s. 215–244.
- Y. Suzuki, F. Fukumoto: *Extracting coordinate terms from newspaper articles and patent documents*. W: Z. Vetulani, red., 2007, s. 275–279.
- K. Szafran: *Automatyczne hasłowanie tekstu polskiego*. „Polonica” 1997, XVIII, s. 51–64.
- Z. Tarajło: *Znaki niefonogramiczne we współczesnym piśmie polskim*. Wrocław: Wyd. Politechniki Wrocławskiej 1998.
- A. Todirascu, Ch. Gledhill, D. Stefanescu: *Extracting Collocations in Contexts*. W: Z. Vetulani, red., 2007, s. 255–259.
- Z. Vetulani, red.: *Human Language Technologies as a Challenge for Computer Science and Linguistics*. W: *Proceedings of the 3rd Language and Technology Conference, Poznań 5–7 October*. Poznań: Wyd. Poznańskie 2007.
- M. Wallace: *What is an idiom? An applied linguistic approach*. W: R.R.K. Hartmann, red., 1979, s. 63–70.
- J. Wawrzyńczyk: *Skrzydlate słowa, konstrukcje idiomatyczne, hapaks legomena języka rosyjskiego w polskim przekładzie literackim*. Warszawa: Semiosis Lexicographica 2006.
- J. Wawrzyńczyk, red.: *Wielki słownik rosyjsko-polski z kluczem polsko-rosyjskim*. Warszawa: Wydawnictwo Naukowe PWN 2004.

- J. Wawrzyńczyk, red.: *Wielki słownik polsko-rosyjski*. Warszawa: Wydawnictwo Naukowe PWN 2005.
- P. Wierchoń: *Z cudzysłowów do poczekalni leksykograficznej II*. Warszawa: Takt 2006.
- P. Wierchoń: *Indeks a tergo jednostek przekładowych Wielkiego słownika rosyjsko-polskiego z kluczem polsko-rosyjskim. Porządek translandowy*. Warszawa: Takt 2006.
- P. Wierchoń: *Indeks a tergo jednostek przekładowych Wielkiego słownika rosyjsko-polskiego z kluczem polsko-rosyjskim. Porządek translatowy*. Warszawa: Takt 2007.
- G. Williams: *In search of representativity in specialised corpora: Categorisation through collocation*. „International Journal of Corpus Linguistics” 2002, nr 7:1, s. 43–64.
- G. Williams: *English Collocation Studies: The OSTI report*. „International Journal of Corpus Linguistics” 2005, nr 10 (2), s. 257–266.
- G. Williams: *Collocational networks: interlocking patterns of lexis in a corpus of plant biology research articles*. „International Journal of Corpus Linguistics” 1998, nr 3 (1), s. 151–171.
- M. Yamamoto, K.W. Church: *Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus*. „Computational Linguistics” 2001, nr 27 (1), s. 1–30.
- N. Yamasaki: *Collocations and colligations associated with discourse functions of unspecific anaphoric nouns*. „International Journal of Corpus Linguistics” 2008, nr 13 (1), s. 75–98.
- A. Zaenen: *Polysemy. Theoretical and Computational Approaches*. „International Journal of Lexicography” 2002, nr 15, s. 238–242.

*Piotr Wierchoń*

## REPRODUCIBILITY WITHIN THE CONFINES OF TRANSLATION UNITS

### Summary

The article presents a multi-word unit retrieval procedure set against the background of other computer science-based methods for the retrieval of multi-word idiomatic units (methods: *T-score*, *Mutual Information*, *n-gramming* etc. and, on the other hand, observation of structural idiomaticity determinants). In Polish language and contrastive Polish-Russian linguistics major inspirations concerning the description of idiomatic combinations (multi-word units, such as *przepraszamy za usterki*, *czwórka z Liverpoolu*, *lista Wildsteina* etc.) were proposed by A. Bogusławski and their extensive development was carried out in W. Chlebda's group (retrieval, collection, position of multi-word units in cultural contexts, reconstruction of authorship, semantic interpretation and even translation suggestions). Automatic methods can be divided into advanced statistical methods (n-gramming, see Yamamoto, Church 2001, Buczyński 2004) and “infantile” approaches (e.g. so-called quotation retrieval, cf., among others, Wierchoń 2005). The present article puts forward a new method for the retrieval of idiomatic combinations applied to a specific text, that is, a reverse order index of translation units. The reverse index lists translation pairs made up of Russian and Polish language segments (words, phrases etc.) ordered according to the sequence of the Russian and Polish alphabet, respectively. Reproducibility of their components is searched for within the confines of those pairs. The retrieval method in question to some extent uses ready (lexicographically compiled) multi-word units, the improvement being, however, the development of their presentation in order to enable relatively uncomplicated observation of the reproducibility of translation pair components.

Петр Вежхонь

## ВОСПРОИЗВОДИМОСТЬ В ПРЕДЕЛАХ ПЕРЕВОДНЫХ ПАР

### Резюме

Статья содержит краткий обзор методов эксцерпции устойчивых сверхлексемных образований, используемых в информатических кругах (напр., *T-score*, *Mutual Information*, *n-gramming*, анализ структурных показателей идиоматичности и т.п.). Предлагая еще одну процедуру выделения устойчивых сверхлексемных образований, автор исходит из идей, предложенных на почве польского и сопоставительного польско-русского языкознания Анджеем Богуславским (концепция единиц языка типа *przepraszamy za usterki, czwórka z Liverpoolu, lista Wildsteina*) и развитых Войцехом Хлебдой (вопросы эксцерпции, связи с культурным контекстом, восстановления авторства, семантической интерпретации, возможности перевода и т.п.). Автоматические методы можно разделить на продвинутые статистические (*n-gramming* — Yamamoto, Church 2001, Buczyński 2004) и «инфантильные» подходы (эксцерпция «по кавычкам»; см. Wierzchoń 2005). Дополнительный метод эксцерпции устойчивых сверхлексемных образований основан на анализе специфического текста, которым является обратный словарь переводных пар, разработанный автором. В пределах этих пар и ведется поиск воспроизводимости их компонентов. Эксцерпция сводится, таким образом, к использованию готовых, т.е. лексикографически уже разработанных, устойчивых сверхлексемных образований, новой же является форма их демонстрации, позволяющая вести несложные наблюдения за воспроизводимостью компонентов переводных пар.