

MAGDALENA KLAPPER
DOROTA KOŁODZIEJ

Instytut Języka Polskiego PAN, Kraków

Elektroniczny Tezaurus Rozproszonego Słownictwa Staropolskiego do 1500 roku. Perspektywy i problemy

Artykuł powstał w związku z tworzeniem koncepcji nowego projektu w Pracowni Języka Staropolskiego IJP PAN obejmującego dygitalizację i udostępnienie średniowiecznych źródeł językowych zawierających polskie słownictwo — Elektronicznego Repozytorium Języka Staropolskiego do 1500 r. (ERJS). Celem projektu będzie stworzenie bazy danych zabytków języka polskiego od najstarszych zachowanych poświadczeń do końca XV wieku.

Jedną z najważniejszych cech piśmiennictwa średniowiecznego na ziemiach polskich jest jego dwujęzyczność — początkowo jedynym językiem tekstów pisanych była łacina, z czasem zaczęła jej towarzyszyć polszczyzna, która w kolejnych wiekach stopniowo się usamodzielniała (Michałowska 1995: 329–335, Łoś 1922: 1–2). Polskie słownictwo pojawiało się także w otoczeniu innych języków, np. czeskiego czy niemieckiego.

Założenia projektu

Projekt ERJS został podzielony na dwa moduły pod względem proporcji elementów polskich do łacińskich. Pierwszy moduł (Elektroniczny Korpus Tekstów Staropolskich do 1500 r. — EKTS¹) będzie zawierał polskie teksty ciągłe, to znaczy utwory napisane wyłącznie po polsku (np. „Bogurodzica”) bądź z nielicznymi łacińskimi dodatkami (np. orytle, „Biblia królowej Zofii” czy „Psałterz floriański”). Natomiast w tekstach nieciągłych elementy łacińskie pod względem ilościowym przeważają nad polskimi, inaczej mówiąc w tekstach tych występują jedynie pojedyncze polskie wyrazy w obcym otoczeniu. Nie mogłyby trafić do korpusu bez naruszenia jego zwartej struktury. Jednak

¹ Wstępne założenia tej części projektu, w tym wykaz i charakterystyka tekstów ciągłych z kanonu źródeł „Słownika staropolskiego” (Sstp), zostały przedstawione w artykule: Klapper, Kołodziej 2014: 205–212.

ze względu na to, że są świadectwem szczególnego związku obu języków i bezcennym źródłem materiału leksykalnego niepoświadczonego w innych zabytkach, konieczne jest zebranie ich w osobnej bazie danych, umożliwiającej ich sprawne przeszukiwanie i porównywanie. To zadanie ma spełnić Elektroniczny Tezaurus Rozproszonego Słownictwa Staropolskiego do 1500 r. (ETRSS), w którym będą zgromadzone wyrazy polskie występujące w kontekstach obcojęzycznych. Ten moduł uzupełni niespełna milionowy korpus języka staropolskiego².

Charakterystyka rozproszonego słownictwa staropolskiego

Średniowieczne zabytki językowe zawierające rozproszone polskie słownictwo są ważnym, choć często niedocenianym źródłem wiedzy na temat leksyki dawnej. O bogactwie i różnorodności tego materiału świadczy to, że poświadczone w nim słownictwo obejmuje wiele ważnych dziedzin życia ludzi na ówczesnych ziemiach polskich (poniższy wykaz ma charakter przeglądu, a nie ścisłej klasyfikacji):

1. Społeczeństwo, prawo, polityka

Do grupy tej należy słownictwo opisujące relacje polityczne i społeczne, zasady ustroju gospodarczego i społecznego, a szczególnie słownictwo związane z procedurami sądowymi, dotyczące organizacji miast, handlu, rzemiosła i edukacji, różnego rodzaju przywilejów, obowiązujących podatków, danin, opłat, powinności.

2. Przyroda i medycyna

W tej grupie znajduje się słownictwo związane z fizyką, astronomią, geografą, geologią, biologią i medycyną, a szczególnie słownictwo opisujące zjawiska przyrodnicze, cechy przedmiotów i ich relacje, nazwy minerałów i drogich kamieni, nazwy botaniczne, zoologiczne, nazwy anatomiczne i fizjologiczne, nazwy chorób, przypadłości i lekarstw.

3. Wytwórczość

Kolejną grupę tematyczną tworzą nazwy surowców oraz efektów działalności wytwórczej człowieka, a także nazwy związanych z nią czynności i ich wykonawców, m.in. słownictwo z zakresu rolnictwa, rzemiosła i eksploatacji zasobów naturalnych.

4. Umysłowość i psychika człowieka

W piśmiennictwie średniowiecznym bogato reprezentowane jest również słownictwo związane z cechami umysłowymi i psychicznymi człowieka oraz opisujące stosunki między ludźmi, sferę emocjonalną, a także procesy poznawcze i twórczość artystyczną.

5. Duchowość i religia

Ostatnia grupa to słownictwo związane z teologią, moralnością, opisujące relacje między światem doczesnym i nadprzyrodzonym, objaśniające realia biblijne, a także

² Nawiązując do ram czasowych Sstp, ilekroć w artykule posługujemy się określeniem „staropolski”, mamy na myśli okres rozwoju języka polskiego począwszy od najstarszych zachowanych poświadczeń do końca XV wieku.

słownictwo związane z organizacją roku kościelnego, instytucjami kościelnymi, religiami niechrześcijańskimi itd.

Ilość rozproszonego polskiego słownictwa w tekstach łacińskich jest trudna do oszacowania. Tak naprawdę dopiero opracowanie tezaury da pełną odpowiedź na pytanie o to, jak wiele staropolskich wyrazów dochoowało się do naszych czasów poza tekstami ciągłymi.

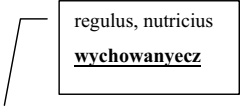
W zabytkach średniowiecznych pojedyncze polskie wyrazy w obcym otoczeniu najczęściej są utożsamiane z glosami. Tak też były traktowane w „Słowniku staropolskim” (Sstp). W bardziej precyzyjnym ujęciu „glosą nazywa się wyraz lub wyrażenie wprowadzone dodatkowo przez pisarza, przepisywacza, czytelnika, komentatora itp. w tekst oryginału” (Klemensiewicz 2009: 138). Zatem słownictwo rozproszone w zabytkach do końca XV w. to pojęcie szersze niż glosa. Ze względu na formę współistnienia z łacińskim kontekstem słownictwo rozproszone możemy podzielić na: glosy, polonizmy wkomponowane w tekst łaciński oraz bezkontekstowe ekwiwalenty translacyjne. Będzie to miało odzwierciedlenie w sposobie prezentacji materiału w bazie danych ETRSS, co postaramy się pokazać na kilku przykładach.

1. Glosy

Pojęciem *glosa* określamy polski wyraz, wyrażenie lub frazę posiadającą odpowiednik w tekście łacińskim, pełniącym funkcję tekstu podstawowego. Tym odpowiednikiem może być pojedynczy wyraz, wyrażenie, zdanie lub kilka zdań. Podstawową funkcją glos jest objaśnienie sensu poszczególnych fragmentów tekstu łacińskiej podstawy.

Jednemu wyrazowi łacińskiemu w tekście podstawowym może odpowiadać więcej niż jedna glosa polska, niekiedy zestawiana z dodatkowymi glosami łacińskimi. Elektroniczna postać tezaury umożliwi precyzyjne odtworzenie położenia poszczególnych glos względem tekstu głównego w rękopisie (glosy interlinearne, marginalne, wewnętrzne), co ma wpływ na interpretację ich znaczenia.

a) glosa interlinearna

Biblioteka Narodowa w Warszawie Rkp BAW 29 k. 89v XV med.	
	<div data-bbox="426 1350 609 1443" style="border: 1px solid black; padding: 5px;"> regulus, nutricius <u>wychowanecz</u> </div> <p data-bbox="177 1465 880 1493">Cuius in locum <u>alumpnus</u> Polonie Wladislaus nomine Boleslai sufficitur.</p> <p data-bbox="177 1548 293 1576"><i>GIWinc 147</i></p>

Między rzędkami tekstu kroniki Wincentego Kadłubka glosator dopisał dwie bliższoznaczne glosy łacińskie objaśniające semantykę wyrazu *alumpnus* (tj. *alumnus*) oraz jego polski odpowiednik *wychowanyecz* (w transkrypcji *wychowaniec*).

b) glosa marginalna

Biblioteka Narodowa w Warszawie Rkp IV 3018 k. 33v XV in.		
Nasci voluit, ut nobis <u>tribueret</u> GKazB IV 41	<table border="1"> <tr> <td data-bbox="669 458 891 550"> pacem et salutatem <u>pokoy asdrowe</u> </td> </tr> </table>	pacem et salutatem <u>pokoy asdrowe</u>
pacem et salutatem <u>pokoy asdrowe</u>		

Na marginesie tekstu kazania dopisano po łacinie uzupełnienie wywodu *pacem et salutatem*, po czym podano jego polskie tłumaczenie *pokoy asdrowe* (w transkrypcji *pokoj a zdrowie*).

c) glosa wewnętrzna

Archiwum Główne Akt Dawnych Warszawa rkp 9052 1360	
<p>Volen<tesque, ut> ... kmethones et incolas ... ab omnibus contribucionibus, exacionibus, <u>expeditionalibus</u>, alias <u>woyenne</u>, et nuptialibus, wlgariter <u>swadzebne</u> dictis, ...</p> <p>absolvimus et perpetuo liberamus</p> <p>KodMaz III 51</p>	

W przytoczonym fragmencie podano polskie odpowiedniki łacińskiej nazwy daniny: *woyenne* (w transkrypcji *wojenne*), oraz opłaty: *swadzebne* (w transkrypcji *swadziebne*).

W wypadku niektórych glos ustalenie ich związku z łacińską podstawą jest trudne bądź wręcz niemożliwe. Dzieje się tak, gdy podczas kolejnego przepisywania dokumentu dochodzi do przemieszania lub opuszczenia partii tekstu bądź włączenia glos interlinearnych lub marginalnych pierwowzoru do tekstu głównego kopii, nie zawsze we właściwych miejscach³.

³ Por. np. uwagi E. Belcarzowej na temat położenia glos wewnętrznych w rękopisach kazań ze zbiorów Biblioteki Narodowej BN III 3027, BN IV 3021 (Belcarzowa 2001: 9, 12).

Nie zawsze funkcja objaśniająca glosy sprowadza się do podania dosłownego tłumaczenia ściśle określonego miejsca podstawy. Odstępstwa mogą dotyczyć zarówno gramatyki (odpowiednik polski jest inną częścią mowy niż podstawa, występuje w innym przypadku itp.), jak i semantyki. Polska glosa może niekiedy oddawać nietypowe lub metaforyczne znaczenie wyrazu łacińskiej podstawy. Zdarza się również, że glosowanie opiera się na luźnym skojarzeniu semantycznym.

W rękopisach kazań często natrafiamy na glosy, których znaczenie odbiega dość daleko od dosłownego znaczenia wyrazu objaśnianego⁴. Należy pamiętać, że w średniowiecznej praktyce homiletycznej używane były liczne zbiory kazań zapisywanych i przepisywanych po łacinie, ale wygłaszanych w języku wernakularnym. Obecność glosy pomagała kaznodziei w szybkim znalezieniu stosownego słowa w języku słuchaczy przy zachowaniu poprawności teologicznej wywodu, co może tłumaczyć, dlaczego glosowane były nie tylko wyrazy trudne i rzadko spotykane, ale też powszechnie używane, jak *miser cordia* (miłosierdzie) czy *obedientia* (posłuszeństwo)⁵.

Przykłady glosowania tekstu w celu zastąpienia obcych realiów elementami bliższej, dobrze znanej rzeczywistości odnajdujemy np. w opisach egzotycznych zwierząt.

Biblioteka Kapitulna i Seminaryjna w Gnieźnie ink. 504 k. C1a ca 1500	<p>Tunc etiam temporis contigit ut duae tigrides ferocissimae — zamb Tigris est serpens</p> <p><i>GI Gn 40</i></p>
-----------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Na przykład łaciński wyraz *tigris* ('tygrys') w zabytkach staropolskich zwykle otrzymuje odpowiednik w postaci nazwy innego potężnego dzikiego zwierzęcia, tyle że występującego w klimacie umiarkowanym. W przytoczonym powyżej przykładzie takim odpowiednikiem, dopisanym na marginesie karty inkunabułu, jest żubr, w zapisie *zamb* (w transkrypcji *zabr*)⁶.

Skrajnym przypadkiem swobodnego tłumaczenia jest glosa *ad sensum*, której nie da się powiązać z żadnym konkretnym wyrazem łacińskim, oddaje ona bowiem sens

⁴ Na przykład w ustępie kazania przywołującym fragment Księgi Wyjścia opisujący odlanie przez Izraelitów złotego cielca glosa *obalkowasty* nie jest dosłownym odpowiednikiem łacińskiego *conflatilem* (*conflatilis* 'ulany, stopiony'). Glosator zastępuje opis sposobu wykonania cielca określeniem jego kształtu ('obły, owalny') (Belcarzowa 2001: 18).

⁵ Por. np. Belcarzowa 1997: 87.

⁶ W jednym z odpisów „Georgik” Wergiliusza tę samą funkcję spełnia tur: *tigres turowye* (w transkrypcji *turowie*) *VergGeorg* 79. Por. Sstp 11, 599.

dłuższego fragmentu utworu⁷. Zdarza się również, że objaśniając łaciński wyraz polisemiczny, glosator podaje polski odpowiednik dla innego znaczenia niż to, w jakim wyraz ten został użyty w tekście⁸. Bywa i tak, że glosa zostaje wpisana wskutek błędnego rozumienia łaciny przez tegoż glosatora.

Zasygnalizowane problemy nie wyczerpują listy trudności, na jakie można natrafić podczas interpretacji średniowiecznego tekstu glosowanego, dają jednak wyobrażenie o tym, jak skomplikowanym zadaniem jest przygotowanie edycji tego typu rękopisu.

2. Polonizmy wkomponowane w tekst łaciński

W wypadku dokumentu glosowanego tekst podstawowy jest samodzielny i może funkcjonować bez polskich wtretów. Zupełnie inaczej sytuacja wygląda w wypadku polonizmów wkomponowanych w tekst łaciński. Są to polskie wyrazy, wyrażenia lub frazy niemające odpowiednika w tekście podstawy, pełniące funkcję nie tyle objaśniającą, co dookreślającą i uzupełniającą. Mogą to być wyrazy pospolite, jak i nazwy własne. W tym wypadku tekst łaciński jest niesamodzielnym. Wyrazy polskie są jego nieodłącznym elementem.

Archiwum Narodowe w Krakowie Advocatalia nr 92 p. 229 1496
Item 3 pepla, 2 nigra et tercium grizeum; item kopa schypow byalich czeskych ; item 4 baliste, arcus cum sagittis
<i>RocznKrak XVI 64</i>

W powyższym fragmencie pochodzącym z ksiąg sądów wójtowskich krakowskich wśród łacińskich formuł wykazu mienia zmarłego pojawia się zapis *kopa schypow byalich czeskych* (w transkrypcji *kopa szypow białych czeskich*) złożony z polskich leksemów: *kopa, szyp, biały, czeski*.

W tego typu zapisach obok polonizmów wkomponowanych w tekst łaciński mogą również pojawiać się glosy:

⁷ Na przykład glosa marginalna *krzeschil* (w transkrypcji *krzesił*) w rękopisie BJ 2368 nawiązuje do fragmentu ewangelii św. Łukasza o wskrzeszeniu młodzieńca z Naim (Luc 7, 11–15), choć nie ma bezpośredniego odpowiednika w tekście kazania; por. komentarz wydawczynie (Belcarzowa 1997: 44).

⁸ Na przykład w rękopisie Biblioteki Publicznej w Petersburgu nr 223, opisanym przez A. Brücknera (1894: 299–310), w komentarzu do hymnu „Clarum decus” fragment „tekel interpretatum comprehensum” jest opatrzony dopiskiem *zwązano jest* (w transkrypcji *związano jest*). Łaciński czasownik *comprehendo*, *-ere* może znaczyć zarówno ‘ocenić, rozważyć’, jak i ‘opasać’. Por. Sstp 11, 500.

Wojewódzkie Archiwum Państwowe w Poznaniu
Księga grodzka nakielska 6 k. 60v–61
1440

A scrupulis alias **od kopczow** oppidi dicti **Wlothowa** a **przeroszl** penes lacum dictum
Pyczyno

KodWP X 197

W cytowanym fragmencie rękopisu poświadczane są następujące polskie wyrazy: *od kopczow*, *Wlothowa*, *przeroszl*, *Pyczyno* (w transkrypcji odpowiednio: *od kopcow*, *Wlotowa*, *przerośl*, *Piczyno*). W tym przykładzie polonizmy wkomponowane to apelatywum *przerośl* oraz ojkonimy: *Wlotów* (obecnie Złotów (Wielatów, Flatau) i *Piczyno* (obecnie Pieczyn), natomiast wyrażenie *od kopcow* jest głosem wewnętrzną (por. Gąsiorowski, Jasiński (wyd.) 1993: 197–198).

Wkomponowanie polonizmów w tekst łaciński mogło nastąpić na etapie powstania tekstu (np. wyliczania nazw miejscowości w dokumentach prawno-administracyjnych) bądź podczas przepisywania glosowanego pierwowzoru, kiedy glosy zostały wciągnięte do tekstu podstawy w miejsce swoich łacińskich odpowiedników.

Biblioteka Jagiellońska
rkps 1508 k. 96v
XV in.

Sed aliquando dyabolus **splōczye** registrum et librum consciencie

GIKazB III 75

W cytowanej kopii kazania pominięto łacińskie orzeczenie zdania, zastępując je pasującym do kontekstu wyrazem polskim *splōczye* (w transkrypcji *splącze*), który w pierwowzorze najprawdopodobniej był zapisany pomiędzy rządkiem. Wydawczynie opisuje zawartość rękopisu jako „tekst makaronizowany”, a polonizmy bez łacińskich odpowiedników jako glosy wewnętrzne „bez łaciny” (Belcarzowa 1997: 68–69, 72–73).

3. Bezkontekstowe ekwiwalenty translacyjne

Pod tym pojęciem rozumiemy wyrazy, wyrażenia lub frazy polskie w tekstach dwujęzycznych o charakterze komparatywnym. W słownikach wyrazy polskie są pozbawione szerszego kontekstu, również materiał łaciński jest tekstem nieciągłym.

Biblioteka Państwowa w Berlinie Theol. Lat. F 655 k. 48v ca 1471		
Vani	Marny	Sap 13, 1
In versacionem	Kwprzebywanyv	Sap 13, 11
Verticibus	Wyrzchowini	Sap 13, 13
Aliqui dicunt Vertex	søk uel garb	Sap 13, 13
Sculpat	dlubye uel rige	Sap 13, 13
[...]		
Nam quia laboraturus est	nyeprezzo essma dzalaczy a. proczovaci	Sap 13, 9
[...]		
Et magice artis appositi erant derisus	atak gym przydano naklam czarnoksyaskye ksągy	Sap 13, 7
<i>(MamLub 155–156)</i>		

Na przykładzie przytoczonych fragmentów mamotrektu lubińskiego dotyczących tekstu trzynastego rozdziału Księgi Mądrości⁹ widać, że ekwiwalent polski może mieć postać pojedynczego wyrazu, grupy wyrazów lub całej frazy. Dla jasności podajemy je w transkrypcji: *marni, ku przebywaniu, wirzchowiny, sęk, garb, dlubie, ryje, nie przeto eż ma działaci, procowaci, a tak jim przydano na kłam czarnoksięskie księgi*.

Rozproszone słownictwo staropolskie w pracach edytorskich

Wydania średniowiecznych źródeł językowych zawierających rozproszone polskie słownictwo mają długą, ponadstuletnią tradycję. Badaniom nad nim były poświęcone prace m.in. E. Belcarzowej, A. Brücknera, K. Górskiego, M. Kowalczyk, H. Kowalewicz, W. Kuraskiewicz, R. Laskowskiego, H. Łopacińskiego, J. Łosia, W.A. Maciejowskiego, L. Malinowskiego, W. Nehringa, F. Piekosińskiego, J. Reczka, J. Rostafińskiego, W.R. Rzepki, B. Ulanowskiego, S. Urbańczyka, W. Wisłockiego, J. Wojtkowskiego, J. Woronczaka, W. Wydry, F. Wysockiej, W. Żurowskiej-Góreckiej. Wraz z rozwojem badań i unowocześnianiem technik wydawniczych edycje te stawały się coraz liczniejsze, bardziej precyzyjne i coraz doskonalsze.

Wydania glos

Zdając sobie sprawę z niewielkiej liczby zachowanych staropolskich tekstów ciągłych, filologowie już w połowie XIX wieku docenili znaczenie polskich glos rozsianych po łacińskich rękopisach w rekonstrukcji systemu leksykalnego polszczyzny w średnio-

⁹ W rękopisie Biblioteki Państwowej w Berlinie wyrazy łacińskie i polskie są zapisane naprzemiennie jednym ciągiem, jednak dla łatwiejszej orientacji czytelnika w treści mamotrektu przyjmujemy trójkolumnową formę edycji, nawiązując do wydania „Mamotrektów staropolskich”. W pierwszej kolumnie podajemy fragment łaciński, w drugiej — polski odpowiednik, w trzeciej — sigła biblijne.

wieczu. Jako pierwszy glosy do trzech kanonów mszy świętej wydał W.A. Maciejowski w 1852 r. (por. Bolz, Żurowska-Górecka 1977: 61). Przeprowadzając kwerendę dostępnych wówczas zbiorów bibliotecznych, badacze wypisywali przynajmniej w wyborze napotykanego w manuskryptach wyrazy, które identyfikowali jako polskie. Prace A. Brücknera, H. Łopacińskiego, J. Łosia, J. Rostafińskiego, W. Wisłockiego i innych po raz pierwszy dały wgląd w nieznaną dotąd i niezbadany materiał, choć autorzy często z konieczności odstępowali od jego systematycznego i wyczerpującego opracowania.

A. Brückner pisał we wstępie do „Kazań średniowiecznych”:

Wartość tych glos mała; glosowaniu ulegają niemal wyłącznie pojęcia oderwane; wyraz ukuty lub przejęty z czeskiego powtarza się więc ciągle; nieraz kazania nawet z tysiącem glos polskich nie zawierają ani dziesięciu rzeczywiście wartych wyszczególnienia; drukowanie więc wszelkich glos wszystkich zbiorów byłoby srogim marnowaniem czasu i pracy; na wyczerpanie zasługują chyba tylko najdawniejsze pomniki, np. z pierwszego ćwierćwiecza [XV w.] (Brückner 1895: 40).

Doświadczenia zespołu opracowującego Sstp wykazały, że takie podejście badawcze jest niewystarczające, przynajmniej w pracach leksykograficznych. Wyraża tę opinię E. Belcarzowa:

Dawne wydania zawiodą w dwóch punktach: po pierwsze podawanie wyrazu polskiego tylko z najbliższym mu wyrazem łacińskim (a często nawet bez niego) bardzo utrudnia, niekiedy wręcz uniemożliwia ustalenie znaczenia tego wyrazu [...] drugi słaby punkt dawnych wydań: wydawcy przeważnie opuszczali glosy, ich zdaniem mało ciekawe, wybierając tylko „rodzynki”, tj. te wyrazy, które uznali za ciekawe dla fonetyki, fleksji czy leksyki. Porównanie tych wydań z rękopisami wykazało, że wśród opuszczonych, rzekomo mało ciekawych słów, zawsze się znajdują takie, które dotąd nie były poświadczone w materiałach Sstp, lub takie, które przedstawiają nieudokumentowaną dotąd formę czy znaczenie (Belcarzowa 1981: 5–6).

Wiele edycji źródłowych z drugiej połowy XX w. było opracowywanych właśnie z myślą o wykorzystaniu w Sstp, co miało wpływ na ich formę graficzną. Wydawcy zamieszczali informacje na temat badanych zabytków we wstępie, zaś główną część pracy zajmowały fragmenty tekstów łacińskich dostarczające kontekstu niezbędnego do ustalenia znaczenia kolejnych glos. Przyjęcie takiej formy wydania nierzadko utrudnia poprawną interpretację poszczególnych urywków mocno okrojonego łacińskiego tekstu podstawowego, np. identyfikację tytułu kazania (tj. dnia, na jaki było przeznaczone), przywoływanych w nim cytatów ze źródeł kościelnych itp. Ograniczenia techniczne druku uniemożliwiały precyzyjne odwzorowanie położenia glos względem tekstu głównego. Glosy interlinearne oraz marginalne, zarówno polskie, jak i łacińskie, często trzeba było wciągać do tekstu podstawowego (jak to ma miejsce w wypadku poświadczeń cytowanych w Sstp), sygnalizując ich umiejscowienie za pomocą złożonego systemu nawiasów lub odmiennego kroju czcionki. Niekiedy powoduje to problemy z przyporządkowaniem trudniejszych, mniej jasnych glos właściwemu wyrazowi łacińskiemu.

Wydania glos w tej postaci, mimo niedoskonałości, stanowią nieocenione źródło wiedzy o polskim słownictwie rozsianym po rękopiśmiennych zabytkach łacińskich. Niestety, prowadzona przez badaczy kwerenda nie była systematyczna. Ponadto różne

okoliczności sprawiły, że niektóre manuskrypty zostały opublikowane kilkakrotnie, podczas gdy inne do dziś nie doczekały się krytycznego opracowania bądź nie zostało ono wydane drukiem. Na przykład zbiór ponad dziewięćdziesięciu kazań *de tempore* i *de sanctis*, pochodzący z rękopisu Archiwum i Biblioteki Krakowskiej Kapituły Katedralnej oznaczony w Sstp skrótem *Serm* (*Sermones*), dostępny jest tylko w postaci rękopiśmiennych ekscerptów (OZSstp 178). Prace nad ekskserpcją zabytków językowych są nadal prowadzone w różnych ośrodkach, jednakże ich efekty są upowszechniane wciąż w niewystarczającym stopniu. Również podczas opracowywania suplementu do Sstp znacznie zwiększyła się dokumentacja źródłowa, która czeka na prezentację:

Tzw. ekscerpty paleografa, czyli odczytania polskich glos z ich łacińskimi kontekstami, mające charakter dokumentacyjny, po zdigitalizowaniu będą mogły służyć na równi z wydaniem źródłowymi średniowiecznych rękopisów (Deptuchowa i in. 2014: 252).

Wydania tekstów łacińskich z wkomponowanymi polonizmami

W edycjach najstarszych zabytków prawnych i sądowych polskie wyrazy w tekście łacińskim często, choć nie zawsze, wyróżniane były graficznie poprzez zastosowanie czcionki pochyłej lub rozstrzelonej, co ułatwia ich identyfikację. Edycje tego typu źródeł, opracowywane głównie przez historyków, zawierają szczegółowe dane na temat poszczególnych manuskryptów, a także osób i miejsc, które są w nich wymienione. Publikowane są pełne teksty zachowanych ksiąg, co daje gwarancję, że żaden polski wyraz nie został pominięty, jednak zdarzają się błędy w odczytaniu.

Wśród źródeł Sstp znajdują się wydania statutów i ustaw królewskich, akt grodzkich i ziemskich, zapisków sądowych, dyplomów, zapisków heraldycznych, przywilejów rodowych, rachunków, inwentarzy itd. z różnych regionów Polski, jednak tylko te, w których zapisano polskie apelatywa. Nie daje to pełnego obrazu piśmiennictwa doby staropolskiej, a zwłaszcza jego początków. Zenon Klemensiewicz w „Historii języka polskiego” pisze:

Umownie otwiera dobę staropolską rok 1136, w którym papież Innocenty II wysłał do arcybiskupa gnieźnieńskiego Jakuba bullę zabezpieczającą majątności arcybiskupstwa (Klemensiewicz 2009: 31).

Kanon źródeł Sstp nie uwzględnia wielu cennych dokumentów łacińskich zawierających wyłącznie polskie nomina propria, również najstarszych, jak „Bulla gnieźnieńska”. Baza danych aspirująca do ukazania pełni zachowanego do dziś słownictwa staropolskiego musi wykraczać poza ten kanon i odejść od sztywnego podziału na nazwy własne i pospolite.

Podobnie jak w wypadku tekstów glosowanych, nie wszystkie znane źródła zawierające polonizmy wkomponowane w tekst łaciński doczekały się edycji drukowanej, np. księgi sądowe grodzkie i ziemskie Krakowa, Biecza oraz Czchowa z Archiwum Narodowego w Krakowie. Materiał ten został włączony do bazy materiałowej Sstp w postaci luźnych ekscerptów i wciąż wymaga systematycznego opracowania.

Wydania słowników

Średniowieczne zapiski zestawiające leksykę łacińską i polską w formie słownika nie są licznie reprezentowane w źródłach, które przetrwały do naszych czasów. Obecnie dostępne są wydania krytyczne łacińsko-polskich słowników ogólnych, np. „Wokabularza trydenckiego”, uważanego za najstarszy zachowany słownik łacińsko-polski, oraz leksykonów specjalnych, w tym dwóch mamotrektów (słowniczków biblijnych podających tłumaczenie trudniejszych wyrazów występujących w kolejnych księgach Wulgaty i prologach do nich). Interesującego materiału leksykalnego dostarcza również słowniczek nazw pokrewieństwa (zob. OżSstp 108). Badacze znajdowali także pomniejsze wokabularze — przeważnie gromadzące po kilka czy kilkanaście wyrazów z zakresu prawa, liturgii, teologii, przyrody czy medycyny — zapisane na pojedynczych kartach bądź okładkach kodeksów.

Niestety, część tego typu zabytków została bezpowrotnie utracona. Przykładem może być grupa pomniejszych słowniczków z manuskryptów Cesarskiej Biblioteki Publicznej w Petersburgu, którą opisał A. Brückner w pracy: „Z rękopisów petersburskich. Część II. Średniowieczne słownictwo polskie” (Brückner 1899). Rękopisy — zawierające m.in. 3 niewielkie słowniczkiki o treści przygodnej, 2 słowniczkiki o tematyce biblijnej, 2 zbiorki nazw przyrodniczych oraz 1 zbiorek leksyki prawniczej — zwrócone Bibliotece Narodowej w Warszawie na mocy traktatu ryskiego — zostały spalone przez hitlerowców po upadku powstania warszawskiego lub zaginęły w innych okolicznościach podczas II wojny światowej (OżSstp 123–126).

Zasady opracowania i udostępnienia materiału

Przegląd materiału, choć tylko skrótowy, pokazuje, jak wielkim bogactwem są edycje źródeł zawierających rozproszone słownictwo staropolskie. Wskazuje też, jakie mają słabe strony oraz jak wiele pracy wymaga ich usystematyzowanie, uzupełnienie i przystosowanie do wykorzystania w formie edycji elektronicznej. Postępująca dygitalizacja zbiorów umożliwi bibliotekom powszechne udostępnianie coraz większej liczby cyfrowych fotografii zabytków rękopiśmiennych przechowywanych w ich zasobach. Pozwala to na weryfikację ustaleń badaczy i poszerza bazę materiałową poświadczeń dawnego słownictwa. Stąd rośnie potrzeba zebrania, uporządkowania i zaktualizowania stanu badań w postaci elektronicznej ogólnodostępnej i przeszukiwalnej bazy danych o rozproszonym słownictwie staropolskim.

Projekt ETRSS będzie polegał na bezpłatnym i powszechnym udostępnieniu transliteracji i transkrypcji rękopisów, których powstanie jest datowane na okres do 1500 r., zawierających polskie wyrazy. Będą one dostępne w ramach internetowej bazy danych zaopatrzonej w narzędzia ułatwiające kwerendę i analizę porównawczą. Główny nacisk zostanie położony na prezentację staropolskiego materiału językowego w obcym otoczeniu. Dlatego podjęto decyzję o niezamieszczaniu pełnych tekstów zabytków, a jedynie obszernych fragmentów (kontekstów) umożliwiających ustalenie znaczenia wy-

razu polskiego. Materiał źródłowy zostanie zaczerpnięty z wydań krytycznych zabytków. Te wydania będą stanowić podstawę transliteracji, identyfikacji fragmentów tekstu jako wyrazów polskich i ich interpretacji leksykalno-morfologicznej. Ustalenia wydawców będą jednak w miarę możliwości weryfikowane na podstawie zachowanych rękopisów oraz najnowszej dostępnej wiedzy historycznej i językowej. Prezentowane zabytki zostaną opatrzone metadanymi. Użytkownik będzie miał możliwość łatwego sprawdzenia, z jakiego rękopisu pochodzi interesujący go polski wyraz, na której karcie został zapisany, bez trudu dotrze również do szczegółowych danych na temat zabytku (np. czasu i miejsca powstania, autorstwa, gatunku, edycji krytycznej), o ile są one dostępne.

Ułatwieniem w korzystaniu z bazy danych będzie graficzne wyodrębnienie wyrazów polskich i ich obcojęzycznych ekwiwalentów (o ile takie są), a także dostęp do komentarzy merytorycznych i edytorskich zamieszczanych w wypadku jakichkolwiek wątpliwości co do odczytania, interpretacji, datowania itp. Komentarze zostaną przygotowane głównie na podstawie uwag wydawców, a także dostępnej literatury przedmiotu. Podczas opracowywania tekstów staropolskich badacze niejednokrotnie natrafiają na problemy związane ze złym stanem zachowania zabytków, nieznormalizowaną ortografią, koniecznością odróżnienia specyficznych zjawisk językowych staropolszczyzny od oczywistych błędów kopisty. Z tych powodów należy się spodziewać, że komentarze będą nader liczne.

Materiał udostępniony w teaurusie będzie anotowany, tzn. każdemu polskiemu segmentowi (który w większości wypadków odpowiada ciągowi znaków między spacjami) zostanie przyporządkowany możliwie najdokładniejszy opis gramatyczny, obejmujący podstawową formę słownikową (lematyzacja), określenie części mowy i formy fleksyjnej. Należy liczyć się z tym, że w wypadku niektórych wyrazów informacja gramatyczna będzie niejednoznaczna bądź niepełna. Sposób przedstawienia opisu gramatycznego polskich wyrazów będzie taki sam jak materiału w EKTS — możliwie jak najbliższy rozwiązaniom stosowanym w Narodowym Korpusie Języka Polskiego, ale z uwzględnieniem specyfiki polszczyzny najstarszej doby piśmiennej. Przewidujemy również specjalną anotację polskich nazw własnych, hybryd polsko-łacińskich (wyrazów polskich z łacińskimi końcówkami) oraz wyrazów, których odczytanie jest niepewne bądź niemożliwe.

Perspektywy współpracy i rozwoju projektu

ETRSS i EKTS będą udostępnione na wspólnej platformie w postaci dwóch niezależnych, ale sprzężonych modułów Elektronicznego Repozytorium Języka Staropolskiego (ERJS). Wspólne ich użytkowanie pozwoli na zebranie pełnej bazy poświadczeń polszczyzny zachowanych w tekstach najstarszej doby.

ETRSS będzie komplementarny wobec opracowywanego obecnie w Pracowni Łaciny Średniowiecznej IJP PAN projektu „Elektronicznego korpusu łaciny średniowiecznej na ziemiach polskich (eFontes Mediae et Infimae Latinitatis Polonorum)”.

Planowany tezaurus częściowo obejmie ten sam materiał językowy, ale nawet w tym zakresie nie będzie jego powieleniem. Należy raz jeszcze podkreślić, że podstawą doboru średniowiecznych zabytków do ETRSS jest obecność w nich wyrazów polskich — a teksty korpusu łaciny średniowiecznej na ziemiach polskich zostały wybrane na podstawie innych kryteriów. Również sposób opracowania i anotacji będzie różny w tych dwóch projektach. W ETRSS podstawa łacińska będzie potraktowana jako element obcy — tło, na którym zostanie zaprezentowane rozproszone polskie słownictwo. Także narzędzia wyszukiwawcze będą wyspecjalizowane w identyfikacji i analizie polskich wyrazów. Te dwa projekty umożliwią wieloaspektowe badania średniowiecznych tekstów łacińsko-polskich. Każdy w ramach swojej specjalności.

Inny projekt związany tematycznie z pracami nad ERJS jest opracowywany przez Mariusza Leńczuka z Pracowni Języka Staropolskiego IJP PAN: „XV-wieczne przekłady Nowego Testamentu — projekt elektronicznej konkordancji staropolskiej. Internetowa baza danych”. Podejmuje on próbę zebrania wszystkich zachowanych przykładów tłumaczeń Pisma Świętego na język polski z tego okresu (pojedynczych wyrazów i dłuższych fragmentów). W badaniach nad glosami polskimi, jak choćby tymi występującymi w tekstach homiletycznych, duże znaczenie oprócz podstawy łacińskiej kazania ma również odniesienie do cytowanego bądź parafrazowanego wersetu biblijnego. Sprzężenie projektu M. Leńczuka z przyszłym ERJS zapewni użytkownikom tej bazy danych dostęp do informacji na temat biblijnego źródła polskiego słownictwa.

Projekt ERJS stwarza również nowe możliwości dla badaczy średniowiecznych nazw osobowych i miejscowych. Podczas opracowywania elektronicznej edycji średniowiecznych zabytków językowych współpraca historyków języka zajmujących się apelatywami ze specjalistami z zakresu onomastyki przyniosłaby niewątpliwie obojętne korzyści.

Podsumowanie

Dwujęzyczność średniowiecznych zabytków językowych zawierających polskie słownictwo sprawia, że Elektroniczny Tezaurus Rozproszonego Słownictwa Staropolskiego (ETRSS) jest metodologicznym i technicznym wyzwaniem. Zdajemy sobie sprawę z trudności opracowania skomplikowanego, niejednoznacznego i niejednorodnego materiału oraz z potrzeby stworzenia narzędzia wiarygodnego i na tyle elastycznego, by mogło sprostać potrzebom współczesnych badań językoznawczych. Cenne będą dla nas wszelkie uwagi i sugestie mogące udoskonalić projekt, aby jak najlepiej służył on użytkownikom. Moduł nieciągi ERJS jeszcze przez długi czas powinien pozostać otwarty, umożliwiając weryfikację dotychczasowych ustaleń, a także uzupełnianie zbioru o nowe, nieopracowane jeszcze, źródła.

Bibliografia

Teksty źródłowe
(skrótów wg OżSstp)

- GIKazB* IV: E. Belcarzowa, Głosy polskie w łacińskich kazaniach średniowiecznych. Część IV, Kraków 2001.
- GIWinc*: H. Kowalewicz, W. Kuraszkiewicz, Polskie głosy w rękopisie Kroniki Wincentego Kadłubka, *Slavia Occidentalis* XXXIV, 1977, s. 134–153.
- KodMaz* III: I. Sułkowska-Kuraś, S. Kuraś, Nowy kodeks dyplomatyczny Mazowsza. Część III, Warszawa 2000.
- KodWP*: A. Gąsiorowski, T. Jasiński (wyd.), Kodeks dyplomatyczny Wielkopolski, t. 10, Poznań 1993.
- MamLub*: W. Żurowska-Górecka, V. Kyas (opr.), Mamotrekty staropolskie. Cz. I–III, Wrocław–Warszawa–Kraków–Gdańsk 1977–1980.
- RocznKra*k: J. Ptaśnik, Studia nad patrycjatem krakowskim wieków średnich. Część druga, *Rocznik Krakowski* XVI, 1914, s. 1–90.
- VergGeorg*: B. Bolz, W. Żurowska-Górecka, Vergili Georgicorum schedae Gnesnenses, *Symbolae Philologorum Posnaniensium Graecae at Latinae* III, 1977, s. 41–89.

Literatura

- Belcarzowa E., 1981–2001, Głosy polskie w łacińskich kazaniach średniowiecznych. Część I–IV, Wrocław–Warszawa–Kraków–Gdańsk–Łódź.
- Bolz B., Żurowska-Górecka W., 1977, Głosy staropolskie w gnieźnińskim rękopisie Wergilego, *Symbolae Philologorum Posnaniensium Graecae at Latinae* III, s. 61–72.
- Brückner A., 1894, Średniowieczna poezja łacińska w Polsce. Część trzecia, [w:] *Rozprawy Akademii Umiejętności* 23, Kraków, s. 268–319.
- 1895, Kazania średniowieczne. Część pierwsza, [w:] *Rozprawy Akademii Umiejętności* 24, Kraków, s. 38–97.
- 1899, Z rękopisów petersburskich. Część II: Średniowieczne słownictwo polskie, [w:] *Prace Filologiczne* 5, Warszawa, s. 1–52.
- Deptuchowa E., Frodyma M., Szelachowska-Winiarzowa L., Wanicowa Z., Wysocka F., 2014, Skarbnica leksyki staropolskiej, *Język Polski* XCIV, s. 250–264.
- Klapper M., Kołodziej D., 2014, Elektroniczny Korpus Tekstów Staropolskich do 1500 r. Perspektywy i problemy, *Prace Filologiczne* LXV, s. 205–212.
- Klemensiewicz Z., 2009, *Historia języka polskiego*, Warszawa.
- Łoś J., 1922, *Początki piśmiennictwa polskiego. Przegląd zabytków językowych*, Lwów–Warszawa–Kraków.
- Michałowska T., 1995, *Średniowiecze*, Warszawa.
- OżSstp: W. Twardzik, E. Deptuchowa, L. Szelachowska-Winiarzowa (red.), *Opis źródeł Słownika staropolskiego*, Kraków 2005.
- Sstp: *Słownik staropolski*, red. S. Urbańczyk, Warszawa–Wrocław–Kraków 1953–2002.

SUMMARY

Digital Thesaurus of Scattered Old Polish Lexis until 1500. Prospects and issues

Key words: medieval Polish language, Old Polish lexis, scattered vocabulary, gloss.

Słowa kluczowe: polszczyzna średniowieczna, leksyka staropolska, słownictwo rozproszone, glosa.

The aim of this paper is to present the project called Elektroniczny Tezaurus Rozproszonego Słownictwa Staropolskiego (ETRSS — Digital Thesaurus of Scattered Old Polish Lexis). The planned database will gather digital editions of linguistic sources preserved to our times, from the oldest until 1500, which contain Polish vocabulary. The authors present the assumptions of the project, describe the characteristics of scattered Old Polish lexis, outline history of its research, as well as formulate basic principles of the database structure and functionality. They provide examples showing the diversity of the material and some issues related to its presentation and interpretation. The paper also describes the prospects for the development of the project and cooperation with researchers pursuing related projects.