

Tomasz Jasiński

## Informatyka w służbie filologii łacińskiej i historii

### Wstęp

Kultura i cywilizacja europejska opierają się na dziedzictwie antyku, przekazanym nam i wzbogaconym przez dorobek pokoleń okresu średniowiecza i nowożytności. Szczególna rolę w przekazywaniu i utrwaleniu tego dziedzictwa odegrał język łaciński, który nieprzerwanie przez ponad 2000 lat pełnił rolę krwioobiegu w funkcjonowaniu kultury europejskiej. Jeszcze w II połowie XIX w., na wielu uniwersytetach europejskich, pisano prace doktorskie i habilitacyjne po łacinie. Spuścizna łacińskich tekstów począwszy od czasów przedklasycznych aż niemal po dzień dzisiejszy jest najważniejszą częścią dziedzictwa europejskiego. Dorobek cywilizacji łacińskiej nieustannie kształtował narodowe kultury, literaturę, naukę, prawo, *etc.* Europejczyków, a za ich pośrednictwem całego świata. Dziś – chcąc siebie zrozumieć – zarówno Europejczycy, jak i mieszkańcy świata muszą ciągle na nowo badać i poznawać tę łacińską spuściznę.

Na pierwszy rzut oka wydawałoby się, że już wszystko zostało powiedziane o prawie i literaturze rzymskiej, że rozpoznano teksty Ojców Kościoła i żywoty Świętych, że poznano kroniki i dokumenty średniowiecza. Chociaż trudno zaprzeczyć osiągnięciom w tym zakresie poprzednich pokoleń uczonych, to jednak informatyka przyniosła latynistom, literaturoznawcom i historykom zupełnie nowe możliwości. Powstały liczne korpusy różnego rodzaju tekstów łacińskich, które pozwalają uczonym z niezwykłą precyzją rozwiązywać wiele kwestii badawczych, które wcześniej wymagały niekończących się kwerend. Często wiele kwestii było poza zasięgiem badających. Różnorodność tych korpusów i baz tekstowych jest tak wielka, że nawet ich wyliczenie przekroczyłoby możliwości tego artykułu. Może jeszcze większą pomocą dla uczonych niż najwspanialsze korpusy elektroniczne są wyszukiwarki internetowe, dzięki którym natychmiast można dotrzeć do szukanych fraz językowych, a także do literatury danego zagadnienia. Znalezienie dziś przysłowiowej igły w stogu siana to żadna trudność; dzięki Internetowi staje się to nie tylko możliwe, ale staje się to zwykłym standardem naukowym. Dzięki wyszukiwarkom internetowym, bazom elektronicznym i korpusom słownikowym w jedną chwilę docieramy do sedna

problemu, otrzymujemy całą bazę źródłową i pełny wykaz literatury. Oczywiście ze względu na ochronę praw autorskich *etc.* konieczne jest jeszcze odwiedzenie tradycyjnej biblioteki, ale zanim ją odwiedzimy, wiemy dokładnie, jakich książek poszukujemy i na jakiej stronie tych książek znajdziemy poszukiwane przez nas zagadnienie. Dziś latynista, historyk czy w ogóle reprezentant nauk humanistycznych dzięki informatyce ma jakby indeks do całego dziedzictwa kulturowego świata!

Czy zatem jeszcze czegoś potrzebujemy; może poza nieograniczonym dostępem do wszelkiej literatury, bez względu na prawa autorskie. Wiadomo, że gigantyczna część naszego łacińskiego europejskiego dziedzictwa spoczywa na półkach archiwów i bibliotek. Materiały te dzięki licznym programom digitalizacyjnym stają się w coraz szybszym tempie powszechnie dostępne. Wiele z nich nie jest dotychczas odczytanych i z tego powodu ich treść nie jest dostępna wyszukiwarkom internetowym. Minie zapewne wiele lat zanim cała ta spuścizna zostanie w przeważającej części odczytana za pomocą specjalnych programów Optical Character Recognition (OCR). Gdy w 1993 r. w czasie jubileuszu profesora Zbigniewa Perzanowskiego mówiłem, że w przyszłości kursywę gotycką i inne pisma będą czytały specjalne programy OCR, życzliwie śmiano się ze mnie. W zeszłym roku Biblioteka Kórnicka PAN wraz z jedną z komercyjnych firm informatycznych złożyła wniosek o fundusze w jednym z programów europejskich na sfinansowanie programu OCR wspomagającego odczytywanie średniowiecznych ksiąg grodzkich i ziemskich oraz dokumentów łacińskich. „Na szczęście” eksperci oceniający współpracę e-biznesu z nauką nie zakwalifikowali naszego projektu do realizacji, ale nie ma już żadnej bariery technologicznej, aby taki program powstał i funkcjonował. Piszę „na szczęście”, gdyż z powodu niżej opisanego projektu baliśmy się, że mamy zbyt małe rezerwy „ludzkie” (wszyscy są zaangażowani w inne granty), aby wykonać projekt OCR dla kursywy gotyckiej.

Jak widać, zarówno dokonania informatyki w dziedzinie latynistyki i powiązanych z nią nauk są ogromne, jak również perspektywy zastosowania informatyki w tym zakresie są niezwykle obiecujące. Czy czegoś jednak nie przeoczyliśmy, co mogłoby przynieść postęp w badaniach. Otóż łatwo zauważyć, iż brak programów, które byłyby wsparciem dla uczonych przy analizie tekstów łacińskich. Brak takich programów stał się oczywisty, gdy prowadziłem badania nad *Kroniką polską* Galla Anonima. Próba znalezienia analogicznej prozy, a później też pokrewnych utworów poetyckich napotyka na poważne przeszkody. Gall Anonim był niezwykle sprawnym pisarzem i poetą, nie potrzebował – jak to robiło wielu pisarzy średniowiecza – trzymać się kurczowo tekstu, z którego korzystał; nie musiał powtarzać własnych fraz. Te cechy jego stylu utrudniają odszukanie jego wzorców. Poszukiwanie analogicznego słownictwa w korpusach na niewiele się

zdało. Zresztą metodę tę należy uznać za niezwykle prymitywną. Wpisujemy jedno charakterystyczne słowo i szukamy w różnych korpusach. Wprawdzie znajdujemy to słowo w innych tekstach, ale musimy mieć naprawdę szczęście, aby trafić nie tylko na podobne pojedyncze słowo, ale na zbliżony zwrot językowy. Ta metoda, którą posługuje się wielu badaczy, jest niezwykle pracochłonna i mało owocna, a w przypadku Galla Anonima niemal bezużyteczna. Specyfika prozy Galla Anonima z konieczności nadała poszukiwaniom komputerowym zupełnie nowy, nieoczekiwany kierunek. W ramach tych poszukiwań powstało najpierw kilka programów komputerowych wyłącznie na użytek indywidualny, później jednak na podstawie tych doświadczeń zrodziła się większa inicjatywa, dzięki której zarówno latyniści, neolatyniści, literaturoznawcy i historycy otrzymają informatyczne narzędzie badawcze zupełnie nowego typu. Przyjrzyjmy się zatem najpierw owym „pierwotnym narzędziom”, które stały się pomostem do budowy potężnego „kombajnu” informatycznego.

#### **1. Początkowe badawcze doświadczenia informatyczne, które wpłynęły na kształt projektu badawczego, opracowywanego przez Bibliotekę Kórnicką PAN**

Główne doświadczenia z zakresu informatyki powstały w czasie moich studiów nad *Kroniką polską* Galla Anonima. Wprowadziłem wówczas do użycia kilka programów, które dosłownie powywracały nie tylko nasze wyobrażenia o prozie i poezji Galla, ale które przede wszystkim rzuciły całkowicie nowe światło na rozwój piśmiennictwa łacińskiego. Zapoznanie czytelnika z tymi doświadczeniami, lepiej pozwoli zrozumieć przydatność tego rodzaju programów w badaniach zarówno filologicznych, jak i historycznych. Kilkakrotnie pisałem o programie zbudowanym do poszukiwania komputerowego prozy Galla. Nigdy jednak nie zdradzałem, w jaki sposób ten program jest zbudowany. Zaznaczałem tylko, że to program „nowej generacji”, a także dodawałem, że nie jest to program opierający się na poszukiwaniu wspólnego słownictwa czy podobnych formuł gramatycznych. Był to program poszukujący tekstów prozatorskich z taką samą rytmiką, jak rytmika w *Kronice polskiej* Galla Anonima. Ponieważ był to pierwszy na świecie program badający rytmikę łacińską (nie wiem, czy przypadkiem w ogóle nie pierwszy badający rytmikę), to uważałem, że w pełni zasługuje na dumną nazwę: program „nowej generacji”. Jest to niezwykle prosty program, jego niezwykłość opiera się na kilku moich spostrzeżeniach i obserwacjach, które stały się dla mnie oczywiste w czasie analizy prozy *Kroniki* Galla Anonima. Nigdy o nich nie pisałem, gdyż do czasu zakończenia moich badań prowadzonych za pomocą tego programu, postanowiłem trzymać w tajemnicy zasadę jego funkcjonowania. Obecnie gdy te badania są zakończone i gdy zdobyte w czasie badań doświadczenia stały się podstawą budowy wielkiego narzędzia

badawczego, czyli wspomnianego „kombajnu” informatycznego, konieczne jest ujawnienie budowy tego programu do poszukiwania prozy rytmicznej analogicznej do *Kroniki polskiej* Galla Anonima. Jestem winien wyjaśnienia zasad działania tego programu już chociażby z tego względu, aby uzyskane wyniki nie budziły wątpliwości badaczy i żeby każdy mógł się przekonać, czy program ten jest wiarygodny.

„Wynalezienie” owego programu było całkowicie przypadkowe. Gdy – idąc śladem Danuty Borawskiej<sup>1</sup> – tradycyjnymi metodami badawczymi starałem się dowieść, iż *Kronika polska* Galla Anonima i *Translacja św. Mikołaja z Myry do Wenecji* (*Historia de translatione sanctorum Magni Nicolai*) anonimowego Mnicha z Lido są dziełem jednego, tego samego pisarza, część badaczy sprzeciwiła się temu pogładowi, stwierdzając, iż podobne słownictwo, zbliżona rytmika i pokrewne toposy literackie niekoniecznie potwierdzają moje ustalenia. Podkreślano (jednak bez przeprowadzenia głębszych badań), że wielu pisarzy wówczas posługiwało się takim słownictwem, że rytmika jest typowa dla epoki *etc.* Zrozumiałem, że powtarzanie przeze mnie dotychczasowych argumentów, ewentualnie dodanie jeszcze dwóch czy trzech podobnych zwrotów, nie przekona moich adwersarzy. Konieczne były zupełnie nowe przekonywujące argumenty. Prędko zdałem sobie sprawę, że dostarczyć ich może tylko analiza rytmiki Galla Anonima i Mnicha z Lido, a następnie odpowiedzenie na pytanie, czy jest to rytmika typowa dla epoki, czy też jakaś niezwykła i nieznaną. Wiele przemawiało za tym, iż była to wyjątkowa rytmika, ale należało tego dowieść. Za niezwykłością rytmiki Galla i Mnicha z Lido przemawiały ustalenia Feliksa Pohoreckiego, który w latach 1929-1930 opublikował fenomenalny artykuł o rytmice prozy, a także o rytmice poezji Galla Anonima<sup>2</sup>. W swojej rozprawie, moim zdaniem dotychczas najwybitniejszej pracy poświęconej Gallowi, pisał m.in.:

Jak długo zatem nie odkryjemy współczesnego Gallowi dzieła o tych samych właściwościach stylistycznych, a przede wszystkim o tym samym rozkładzie form kursusowych, z charakterystyczną przewagą *veloxu*, a przy równoczesnym wydatnym posługiwaniu się formą zwaną *c. trispondaicus*, tak długo nie rozwiążemy zagadki Galla-Anonima<sup>3</sup>.

Z tych słów wynikało, że proza rytmiczna ze zdecydowaną przewagą *veloxu*, przy znacznym udziale *trispondaicusa*, była czymś nadzwyczajnym, czymś,

1 D. Borawska, *Gallus Anonim czy Italus Anonim*, „Przegląd Historyczny” 1965, t. 56, nr 1, s. 111-119.

2 F. Pohorecki, *Rytmika kroniki Galla-Anonima*, „Roczniki Historyczne” 1929, t. 5, s. 105-169 (cz. I) – [www.wbc.poznan.pl/publication/413076](http://www.wbc.poznan.pl/publication/413076) [dostęp: 01.03.2015]; „Roczniki Historyczne” 1930, t. 6, z. 1, s. 12-75 (cz. II).

3 F. Pohorecki, *Rytmika...*, cz. II, s. 71.

czego pełen energii i pomysłów Pohorecki nie mógł znaleźć, poza oczywiście „naszym” Gallem. Wprawdzie nie potrafiłem wówczas (2008) ocenić podobieństwa rytmiki Galla i Mnicha z Lido, to jednak na podstawie ustaleń Mariana Plezi było mi wiadome, że również w *Translacji św. Mikołaja* przeważał *velox*, a *trispondaicus* był na drugim miejscu<sup>4</sup>. A więc jakby spełnił się postulat badawczy Feliksa Pohoreckiego: oto mamy dzieło z podobną rytmiką. Moim adwersarzom jednak to nie wystarczało. Trzeba było ustalić, czy taka proza była na przełomie XI / XII w. czymś typowym czy też nadzwyczajnym. Ponadto gnębiła mnie wówczas jeszcze jedna wątpliwość: czy wystarczy odnotować określony procentowo udział poszczególnych rytmów w danej prozie, czy też nie, bo może są różne rodzaje *veloxu*, *trispondaicusa*, *etc.*

Chociaż miałem za sobą lekturę wielu prac poświęconych rytmice, nie potrafiłem samodzielnie zidentyfikować kadencji rytmicznych prozy. Po nauczaniu się tego, co okazało się bardzo proste, rozpocząłem identyfikację każdego rytmu w *Kronice polskiej* i w *Translacji św. Mikołaja*. Z literatury wiedziałem, że rytm *velox* powstaje, gdy czterosylabowy wyraz akcentowany na przedostatnią sylabę (paroksyton), poprzedzony jest jakimkolwiek wyrazem akcentowanym na trzecią od końca sylabę (proparoksyton), natomiast *trispondaicus*, gdy czterosylabowy paroksyton jest poprzedzony też jakimkolwiek paroksytonem. Akcent w języku łacińskim – pomijam wszelkie skomplikowane wątpliwości co do charakteru akcentu (toniczny, dynamiczny, *etc.*) – podobnie jak w języku polskim pada na przedostatnią sylabę. Jedynie w wypadku, gdy przedostatnia sylaba jest krótka, to wówczas akcent pada na trzecią od końca sylabę. To, czy sylaba druga od końca jest krótka czy długa można ustalić na podstawie słownika, znajomości gramatyki, a w tym dość prostych reguł<sup>5</sup>. Już po krótkiej chwili pracy nad tekstem Galla zorientowałem się, że m.in. szereg *veloxów* stworzonych przez tego kronikarza kończy się wyrazem zakończonym na *-avit*. Dzięki temu spostrzeżeniu postanowiłem ułatwić sobie pracę. Zamiast systematycznie czytać Galla i zaznaczać każdy kolejny rytm, wpisałem do funkcji *Ctrl+f* (znajdowanie) końcówkę *-avit*, a następnie naciskałem *Enter*. Komputer szukał wyrazu na *-avit*, a ja liczyłem jego sylaby. Jeżeli wyraz ten miał cztery sylaby, a tak najczęściej było, spoglądałem na wyraz go poprzedzający. Gdy był

4 M. Plezia, *Nowe studia nad Gallem-Anonimem*, w: *Mente et litteris. O kulturze i społeczeństwie wieków średnich*, kom. red. H. Chłopocka [et al.], Poznań 1984, s. 111-120.

5 Oczywiście wiedziałem, że są też rytmy tzw. *cum consillabicatione*, które tworzą sekwencje akcentów więcej niż dwóch wyrazów, a więc np. proklityki i wyrazu trzysylabowego zamiast wyrazu czterosylabowego, ale tego rodzaju akcenty jako nieliczne mogły być na etapie wstępnym pominięte.

to proparoksyton, mogłem obok tych dwóch wyrazów wpisać – robiłem to za pomocą funkcji *Ctrl+v* – słowo: *velox* (np. *convívium preparávit*). Gdy ów czterosylabowiec poprzedzał paroksyton, wówczas wpisywałem: *trispondaicus* (np. *amicórum invitávit*). Gdy „rozprawiłem się” z wszystkimi czterosylabowcami zakończonymi na *-avit*, poszukałem kolejnych końcówek charakterystycznych dla „drugiego wyrazu” z kadencji tworzącej *velox* lub *trispondaicus*. Ostatecznie znalazłem – przeglądając tekst *Kroniki polskiej* – następujące najbardziej charakterystyczne końcówki czterosylabowców, na podstawie których Gall budował albo *veloxy* albo *trispondaicusy*:

*-abant -abat -amur -amus -antes -antur -are -aret -ari -asse -assem -assent -asses -asset -atur -avit -ebant -ebat -emus -entes -erunt -etur -isse -issem -issent -isset -ivit.*

Oczywiście rytm *velox* i *trispondaicus* konstruował Gall też na podstawie wyrazów z innymi końcówkami, ale te wyżej wymienione – jak się okazało – były najczęstsze. Dysponowałem też spostrzeżeniami dotyczącymi „pierwszych wyrazów” kadencji, czyli wyrazów poprzedzających czterosylabowy paroksyton. Najczęściej stosowanym przez kronikarza zabiegiem, aby utworzyć *velox*, było posłużenie się w przypadku „pierwszego wyrazu” takim słowem, w którego dwóch ostatnich sylabach występowały obok siebie dwie samogłoski. W takich wypadkach, poza kilkoma wyjątkami, przedostatnia sylaba jest zawsze krótka, co skutkuje przesunięciem akcentu na trzecią sylabę od końca, np. *studio speculari*; *faciem contemplari*; *naufragium evitare*. Tego rodzaju *veloxów* w księdze I spotykamy aż 227, co stanowi ponad 46% wszystkich. Ponadto Gall posługiwał się dość ściśle określonymi słowami czy raczej końcówkami, które generowały proparoksyton. Nie było ich wiele, ale większość z nich odmieniała się przez przypadki, co trzeba było uwzględnić w programie:

*-bile -bilem -biles -bili -bilis -cula -culae -culam -cule -culi -culis -culo -culos -culum -culus*

*-erit -errima -errimae -errimam -errimas -errimi -errimis -errimo -errimos -errimum -errimus*

*-ibus -ica -icae -icam -icas -ice -ici -icis -ico -icos -icum -icus -issima -issimae -issimam*

*-issimas -issimi -issimis -issimo -issimos -issimum -issimus -iter -libet -mine -minem -mines -mini -minis -ola -olae -olam -olas -ole -oli -olis -olo -olum -olus -titit -udine -udinem -udines -udini -udinis -ula -ulae -ulam -ulas -ule -uli -ulis -ulo -ulum -ulus.*

Zwróciłem też uwagę, iż obok tych końcówek w budowie *veloxów* w prozie Galla ważną rolę odgrywały dwa proparoksytony, a mianowicie: *dóminus* oraz *epíscopus*.

Gdy oznaczałem *veloxy* w *Kronice polskiej* Galla, a także w *Translacji* Mnicha z Lido za pomocą funkcji *Ctrl+f* uświadomiłem sobie nagle, że mógłbym w ten sam sposób przeszukiwać wielkie pliki tekstowe ze średniowiecznymi tekstami łacińskimi. Od wielu lat zbierałem takie pliki w Internecie, gdzie pełno było (i jest coraz więcej) przeszukiwalnych tekstów kronik, żywotów Świętych, dokumentów; były one w różnych ogólnodostępnych bazach, a ich adresy można było znaleźć na licznych forach historyków-amatorów (!). Wyobrażałem sobie, że podobnie jak w przypadku *Kroniki polskiej* na początek zacznę szukać końcówki *-avit*, a następnie – jeżeli za pomocą tej metody znalazłbym *velox* – to wówczas rzucę okiem na kolejne zdania, aby przekonać się, czy mam do czynienia z rytmiką zbliżoną do Galla. Okazało się jednak, że takich *veloxów* są – bez przesady – setki tysięcy (!), zarówno przypadkowych, jak też występujących w mniej czy bardziej licznych grupach. Szybko uzmysłowiłem sobie, że przeszukiwanie wszystkich moich zasobów tekstów łacińskich za pomocą jednej końcówki byłoby nie lada wyzwaniem. Analizowanie setek tysięcy przypadków przerastało ludzkie możliwości, a przecież trzeba było jeszcze te same teksty przejrzeć za pomocą dwudziestu kilku kolejnych końcówek! W tym momencie doszedłem do wniosku, że mógłbym napisać bardzo prosty program komputerowy do poszukiwania *veloxów* „typu Gall”. Program ten nie byłby bardziej skomplikowany niż program, który napisałem w 1992 r., a opublikowałem w 1995 r., do poszukiwania identycznych zwrotów w plikach zebranych na potrzeby edycji *Kodeksu dyplomatycznego Wielkopolski*<sup>6</sup>. Rozkazy i pętle dla programu do poszukiwania rytmiki „typu Gall” były niezwykle proste. Pierwsza pętla była następująca: 1. szukaj końcówki *-abant*; 2. gdy ją znalazłeś policz, czy znaleziony wyraz z tą końcówką ma cztery sylaby, czyli cztery samogłoski (tu musiałem zdefiniować szereg wyjątków)<sup>7</sup>. Gdy wyraz liczyłby cztery sylaby,

6 T. Jasiński, Komputer a perspektywy dyplomatyki polskiej, w: *Tradycje i perspektywy nauk pomocniczych historii w Polsce: materiały z sympozjum w Uniwersytecie Jagiellońskim dnia 21-22 października 1993 roku profesorowi Zbigniewowi Perzanowskiemu przypisane*, pod red. M. Rokosza, Kraków 1995, s. 273-282.

7 Wyraz liczył cztery samogłoski (samogłoski to: *a, e, i, o, u, y*), przy czym zgodnie z zasadami języka łacińskiego *ae* i *oe* to były pojedyncze samogłoski: *e*, jednak z jednym wyjątkiem: na początku wyrazu, kiedy głoskę *oe* poprzedza spółgłoska *c*, to wówczas *oe* program musiał liczyć jako dwie oddzielne samogłoski; przy liczeniu samogłosek występujące obok siebie *au* zgodnie z zasadami języka łacińskiego należało uwzględnić jako jedną samogłoskę *a*; podobnie w przypadku *iu* też należało pominąć *i*, a tylko na początku wyrazu oraz w zestawieniach *abiu adiu coniu subiu* (też na początku wyrazu) liczyć tylko jako jedną samogłoskę, a mianowicie jako *u*, np. *abiuravit* ma tylko 4 samogłoski (*a u a i*), bo nie liczymy pierwszego *i*; we wszystkich innych przypadkach

program miał otrzymać rozkaz: wstaw po końcówce *-abant* znak \*<sup>8</sup>. Po zakończeniu tej procedury program miał przejść do analizy wyrazu poprzedzającego, czyli do „pierwszego wyrazu” kadencji. Najpierw miał stwierdzić, czy dany wyraz ma przynajmniej trzy sylaby, czyli trzy samogłoski z wymienionymi wyżej wyjątkami, gdyż tylko wyrazy liczące trzy i więcej sylab mogą być proparoksytonami. Jeżeli wyraz liczyłby przynajmniej trzy sylaby, to wówczas program miał sprawdzić, czy wyraz ten miał którąś z końcówek proparoksytonicznych „pierwszego wyrazu”. Program tak długo miał podstawiać kolejne końcówki, aż znalazłby ją lub stwierdziłby, że nie ma tej końcówki. W pierwszym wypadku obok znaku \* miał dodać znak #. Gdyby nie było takiej końcówki, to wówczas program miał za zadanie sprawdzić, czy czasem w pierwszym wyrazie ostatnia samogłoska nie była bezpośrednio poprzedzona drugą samogłoską. Wówczas też – z małymi wyjątkami – mielibyśmy do czynienia z proparoksytonem. Gdyby tak było, program też miał dostawić do znaku \*, znak #. Jeżeli jednak „pierwszy wyraz” nie liczyłby trzech sylab i nie miał żadnej z wymienionych końcówek, czy też ostatnia samogłoska wyrazu nie byłaby poprzedzona drugą samogłoską, to wówczas program nie miałby dodawać żadnego znaku. Założyłem, że program po wykonaniu pierwszej pętli z końcówką *-abant*, wykona kolejne 26 pętli z kolejnymi końcówkami „drugiego wyrazu”. Jako wynik spodziewałem się otrzymać analizowany tekst w prawie niezmienionej formie, tylko z dodatkowymi znakami \* oraz \*#. Wiadomo było, że samodzielnie występujący znak \* oznaczał, iż mamy do czynienia z trispondaicusem, a dwa obok siebie znaki \*# – iż z veloxem. Oczywiście istniało dwu-, trzyprocentowe prawdopodobieństwo, że któryś z trispondaicusów był w rzeczywistości veloxem i odwrotnie niezwykle małe prawdopodobieństwo, że któryś z veloxów był trispondaicusem. Było to spo-

---

program miał liczyć *iu* jako dwie samogłoski (np. *martyrium*). Identycznie miał postępować z *ia* na początku wyrazu oraz w zestawieniach *abia adia conia subia* (też na początku wyrazu), wówczas takie dwie litery miały być traktowane jako jedna samogłoska (*a*), np. *abiacet* ma tylko 3 samogłoski (*a a e*), bo nie liczymy *i*; wyraz *iacet* ma tylko dwie samogłoski, bo nie liczymy *i*. We wszystkich innych przypadkach należało liczyć *ia* jako dwie samogłoski, np. wyraz *pecunia* ma 4 samogłoski *e u i a*, i jak widać w tym wypadku liczymy samogłoskę *i*. W zestawieniach *qui que qua quo* (obojętnie czy na początku czy w środku wyrazu) założono, że program nie będzie uwzględniał samogłoski *u*, gdyż jest ona w tym wypadku równa spółgłosce, czyli np. wyraz *inequale* ma 4 samogłoski (*i e a e*), bo nie liczymy *u*.

8 Był to znak umowny, który mógł być zastąpiony każdym innym nietypowym znakiem, nie występującym w tekście. Nietypowość pozwalała wykorzystać ten znak w przyszłości do natychmiastowego znajdowania veloxów i trispondaicusów (zob. niżej w głównym tekście).



wodowane tym, że nie opłacało się uwzględniać w programie szczególnych wyjątków (np. Nicoláus, María). Ta drobna nieprecyzyjność programu nie miała znaczenia, gdyż wszystkie teksty znalezione za pomocą programu miały być skontrolowane za pomocą tradycyjnych metod.

Gdy wszystkie założenia logiczne i poszczególne pętle programu zostały zdefiniowane, poradzono mi, abym zrezygnował z samodzielnego przełożenia tych rozkazów na konkretny już program komputerowy i przekazał to zadanie profesjonalnemu programiście. Zwróciłem się wówczas do Piotra Jasińskiego, studenta trzeciego roku informatyki Politechniki Poznańskiej, który przypadkowo nosił to samo nazwisko co ja i którego znałem jako niezwykle uzdolnionego informatyka. Po kilku dniach program był gotowy i działał nienagannie. Błyskawicznie „połykał” wielkie, liczące kilka, a nawet kilkanaście tysięcy znaków pliki tekstowe. Okazało się jednak, że wystąpiło to samo zjawisko, co przy „ręcznym” przeszukiwaniu dużych plików. Veloxów i trispondaicusów było tyle, że niezwykle trudno było wyłowić w tej masie teksty, które bardziej były bardziej intensywnie „nasycone” veloxami. Zwróciłem się wówczas do Piotra Jasińskiego, aby program wyposażył w „nakładkę”, za pomocą której mógłbym poszukiwać takich tekstów. Ustaliliśmy, iż program otrzyma dodatkową możliwość, a mianowicie będzie dodatkowo zaznaczał teksty o wysokim „nasyconiu” veloxami. Otrzymałem możliwość definiowania tych tekstów za pomocą dwóch okienek, w pierwszym wpisywałem liczbę veloxów, a w drugim liczbę znaków. Gdy wpisałem w okienku veloxów np. liczbę 16, a w okienku ze znakami liczbę 4000, to wówczas program wyróżniał wszystkie teksty w pliku, w których na obszarze 4 tys. znaków znalazł przynajmniej 16 veloxów. Dodatkowo na marginesie program oznaczał te fragmenty całego pliku, gdzie znajdowały się teksty spełniające ten warunek (nieco podobnie jak w Google books). Otrzymałem całkowitą swobodę definiowania liczb w obydwu okienkach, mogłem wpisywać każdą liczbę. Wiedziałem jednak, że chcąc uzyskać sensowne wyniki, musiałem wpisywać sensowne liczby, a aby je poznać konieczne było „wykalibrowanie” programu za pomocą tekstu Galla. „Wpuściłem” do komputera olbrzymi plik tekstowy, w którym *Kronika polska* stanowiła ok. 5% tekstu i wpisałem w okienkach właśnie wymienione liczby: 16 i 4000. Była chwila emocji, czy program do poszukiwania veloxów „typu Gall” znajdzie je w *Kronice polskiej* Galla. Po chwili (program przeszukuje z prędkością 1000 stron na sekundę) ukazał się analizowany plik z zaznaczonym każdym veloxem i trispondaicusem, przy czym na marginesie otrzymałem informację, iż spośród całego pliku, tylko jego niewielki fragment – *Kronika polska* – spełnia zdefiniowane wcześniej wymagania, czyli ma 16 veloxów na 4000 znaków. Dalej analizowałem ten plik za pomocą programu, stopniowo podnosząc liczbę w okienku veloxów

i dopiero przy liczbie 37 veloxów na 4000 znaków program poinformował mnie, że już żaden tekst w tym pliku nie spełnia tego warunku. A więc liczbą maksymalną dla *Kroniki polskiej* była liczba 36 veloxów na 4000 znaków. To znaczy, że program przy tym współczynniku wykrywa jedno miejsce w *Kronice polskiej*, w którym występuje aż 36 typowych dla Galla veloxów na 4000 znaków. Później poddałem podobnemu badaniu *Translację św. Mikołaja*; tu liczba maksymalna wynosiła 26. W pierwszej chwili byłem rozczarowany „niskim” wynikiem osiągniętym przez *Translację św. Mikołaja*, ale później się okazało, że był to i tak drugi – po Gallu – najwyższy wynik spośród wszystkich tekstów, a cały zbiór był przecież ogromny, liczył 900 tys. stron. Ponadto należy pamiętać, że liczba kadencji veloxowych w *Kronice polskiej* stanowi bez ułamków procenta 60%, a w *Translacji św. Mikołaja* niemal równe 50%. Przy „kalibrowaniu” programu zauważyłem, że program ten był – takie było wstępne założenie – przewidziany do poszukiwania prozy rytmicznej charakterystycznej dla Galla, a dokładnie veloxów, którymi się posługiwał. Okazało się, że skutkiem tego założenia było, iż dzieła, które miały od *Kroniki polskiej* Galla wyższą procentowo liczbę kadencji rytmicznych osiągały znacznie niższy wynik w analizach przeprowadzonych za pomocą wspomnianego programu. Tak np. dzieła Filipa z Harveng (zm. 1183), w których liczba rytmicznych kadencji wynosiła aż 96%, uzyskiwał znacznie niższy rezultat (liczba veloxów na 4000 znaków) niż *Kronika polska*. Wyniki dotyczące Filipa z Harveng były porównywalne z rezultatami dotyczącymi *Translacji św. Mikołaja*. Jak było to możliwe? Otóż Gall posługiwał się veloxem, którego „drugi wyraz” był najczęściej czasownikiem i to przeważnie w formie osobowej, natomiast „pierwszy wyraz” był przeważnie przymiotnikiem lub rzeczownikiem. Z tego względu, iż rytm ten opiera się na wspomnianych czasownikach, niekiedy nazywam go veloxem „czasownikowym”. Jak się później okazało, był to styl o bardzo długiej tradycji. Filip z Harveng układał veloxy byle jak, nie oglądając się na tę starą tradycję; jego proza miała przez to jakby grafomański charakter i mimo iż było w niej aż tyle veloxów, to jednak tylko jakaś część z nich była spokrewniona strukturalnie z Gallem.

Zanim przystąpiłem do systematycznych analiz wszystkich plików, postanowiłem zweryfikować poprawność działania tego programu w jeszcze jeden sposób, co jak się okazało doprowadziło – już na wstępie – do obalenia dotychczasowych sądów o narodzinach veloxu „czasownikowego”. Wiadomo było, że podobnym cursusem jak Gall (tj. głównie veloxem „czasownikowym”) posługiwała się przede wszystkim kancelaria papieska. Wszyscy uczeni przyjmowali jako pewnik czy dogmat, iż cursus został wprowadzony do kancelarii papieskiej w 1088 r. na skutek jej reformy za czasów pontyfikatu Urbana II, przeprowadzonej przez mnicha z Monte Cassino, Jana z Gaety,

późniejszego papieża Gelazego II<sup>9</sup>. Przekonanie o słuszności tego poglądu zostało tym bardziej umocnione, że w 1963 r. ponownie opowiedziała się za nim wybitna szwedzka znawczyni cursusu Gudrun Lindholm<sup>10</sup>. Wprawdzie w 1975 r. inny szwedzki neolatynista Tore Janson dowodził, że rodzaj cursusu (czyli z przewagą veloxu „czasownikowego”) zrodził się nieco wcześniej, bo już w I połowie XI w szkołach, gdzie nauczano retoryki w północnej Italii, w Dolinie Padańskiej. Pierwszymi pisarzami, którzy mieli się posługiwać tym rodzajem cursusu byli według Tore Jansona: Gwido z Arezzo (ok. 992-ok. 1050) oraz Piotr Damiani (1007-1072). Janson wykazał też, że cursus ten wszedł do użytku w kancelarii papieskiej, nieco wcześniej niż dotychczas przypuszczano, a mianowicie już za Aleksandra II (1061-1073)<sup>11</sup>. Ustalenia Tore Jansona w tym zakresie traktowałem jako pewnik, jako rezultat oparty na badaniach, a nie – jak to było dotychczas – na podstawie jednej zapiski źródłowej.

Wyposażony w tę wiedzę, postanowiłem użyć programu do analizy dużego pliku z tekstami z IV-VI w. Założenie było oczywiste – jeżeli program wykryje w tych tekstach cursus z przewagą veloxu „czasownikowego”, to znaczy że coś z tym programem nie tak. Było oczywiste, że w tym pliku nie mogło być tego rodzaju rytmiki. Ku mojemu zdumieniu program znalazł w tym pliku szereg tekstów z tego rodzaju cursusem; szczególnie dużo tego rodzaju tekstów było w Kodeksie Teodozjusza, a co za tym idzie też i Justyniana, w tekstach napisanych przez św. Ambrożego, św. Augustyna i w wielu tekstach tego okresu. Po przeprowadzeniu badań tradycyjnymi me-

9 *Rerum Gallicarum et Francicarum Scriptoresi. Recueil des historiens des Gaules et de la France. Contenant la suite des monumens des trois Règnes de Philippe I<sup>er</sup>, de Louis VI dit le Gros, et de Louis VII surnommé le Jeune, depuis l'an MLX, jusqu'en MCLXXX*, nouv. éd. publ. sous la dir. de M. L. Delisle, Paris 1877, t. 14, s. 688: *Papa [Urban II] litteratissimus et facundus fratrem Johannem [późniejszego papieża Gelazego II] virum utique sapientem ac providum sentiens ordinavit, admovit suumque cancellarium ex intima deliberatione constituit, ut per eloquentiam sibi a Domino traditam antiqui leporis et elegantiae stilum in sede apostolica iam pene omnem deperditum sancto dictante Spiritu Johannes Dei gratia reformaret ac Leoninum cursum lucida velocitate reduceret* – <http://gallica.bnf.fr/ark:/12148/bpt6k50> [01.03.2015].

10 G. Lindholm, *Studien zum mittellateinischen Prosarhythmus. Seine Entwicklung und sein Abklingen in der Briefliteratur Italiens*, Stockholm 1963, s. 11 i n. (Acta Universitatis Stockholmiensis, Studia Latina Stockholmiensia, t. 10).

11 T. Janson, *Prose Rhythm in Medieval Latin from the 9th to the 13th century*, Lund 1975, s. 34-56 (Acta Universitatis Stockholmiensis, Studia Latina Stockholmiensia, t. 20) – [www.academia.edu/7049867/Prose\\_Rhythm\\_in\\_Medieval\\_Latin\\_from\\_the\\_9th\\_to\\_the\\_13th\\_Century](http://www.academia.edu/7049867/Prose_Rhythm_in_Medieval_Latin_from_the_9th_to_the_13th_Century) [01.03.2015].

toдами nie miałem wątpliwości, iż program działa bez zarzutu i że nie ma mowy o jakiegokolwiek pomyłce programu. Gdy uświadomiłem sobie ten fakt, to wówczas zacząłem się zastanawiać, jak mógł pojawić się w tych tekstach *cursus*, który miał powstać dopiero w pierwszej połowie XI w. Wiedziałem, że w starożytności, od przełomu III / IV w. znany był *cursus mixtus*, czyli mieszanka rytmów iloczynowych, które jednocześnie były wyposażone w rytmy akcentowe, takie same, jakie były później znane w średniowieczu. O istnieniu *cursusu mixtus* było wiadomo od schyłku XIX w., był on przedmiotem licznych badań, wymarł gdzieś w VI wieku, a trwał dalej nawet do VII w. już jako *cursus akcentowy*. Następnie tenże *cursus akcentowy* znikł na ponad sto lat i odrodził się na przełomie VIII / IX w. Wiadomo jednak było, że zarówno *cursus mixtus*, jak i ten wczesnośredniowieczny *cursus akcentowy* w niczym nie przypominał *cursusu veloxowego „czasownikowego”*, znanego z kancelarii papieskiej, a który – jak wyżej była mowa – w świetle najnowszych badań miał się narodzić dopiero w pierwszej połowie XI w. Świadom tych wszystkich ustaleń, w pierwszej chwili – po odkryciu takiego *veloxu* w tekstach IV-VI w. – pomyślałem, że może wszyscy badacze późnoantycznych tekstów padli ofiarą podobnego złudzenia, jak niegdyś u schyłku średniowiecza ulegli takiemu złudzeniu humaniści – poszukiwacze antycznych tekstów w bibliotekach najstarszych europejskich klasztorów. Jak wiadomo, ci pionierzy humanizmu znaleźli tam teksty wielu antycznych pisarzy, spisane minuskułą karolińską. Na tej podstawie doszli do wniosku, iż jest to pismo antyczne i jako takie zostało przez nich rozpropagowane tak skutecznie, iż zostało ono sztucznie przywrócone do życia i nawet my dzisiaj nim piszemy. Pomyślałem, że podobnie mogło być z zabytkami antycznymi; zastanawiałem się, czy mogły zostać one w czasie przepisywania w okresie późnokarolińskim i ottońskim, ozdobione najnowszym i najmodniejszym stylem, a mianowicie *cursusem z veloxem „czasownikowym”*. Wydało mi się to jednak niemożliwe, gdyż w tych czasach (późnokarolińskich i ottońskich) nie mogło być – jak przypuszczałem – *veloxu „czasownikowego”*. To, że takie przypuszczenie o rzekomym ozdobieniu okazało się całkowicie fałszywe, przekonały mnie najstarsze rękopisy Kodeksu Teodozjusza. Pochodziły one z okresu wczesnego średniowiecza i nie mogły w tym czasie być „zainfekowane” *veloxem* w ramach rewolucji *veloxowej*, która wedle przypuszczeń Tore Jansona miała miejsce w pierwszej połowie XI w. w północnej Italii. Nie ulegało wątpliwości, że *velox „czasownikowy”* był znany najpóźniej od pierwszych dekad IV w. Chr. Ze zdumieniem jednak stwierdziłem, że nikt z historyków prawa rzymskiego, począwszy od Theodora Mommsena, który przygotował nowoczesne wydanie Kodeksu Teodozjusza (1817-1903), aż po współczesnego wybitnego znawcę prawa rzymskiego Detlefa Liebsa, nie miał pojęcia, że Ko-

deks ten napisany jest prozą rytmiczną, w wielu miejscach z przewagą veloxu „czasownikowego”. Zdziwiająco, że nie wiedział tego zwłaszcza ostatni uczo-ny, który badał szczegółowo teksty prawne w tym Kodeksie, a w tym szczegó-ólnie opublikowane przez Konstantyna Wielkiego<sup>12</sup>. Teksty te pochodzące z drugiego, trzeciego i czwartego dziesięciolecia IV w. okazały się najstarszy-mi tekstami napisanymi prozą rytmiczną z przewagą veloxu „czasownikowe-ego”. Detlef Liebs napisał nawet specjalną rozprawę poświęconą rytmom tekstów prawa rzymskiego<sup>13</sup>. W artykule tym nie znajdziemy jednak ani wzmianki o istnieniu kadencji rytmicznych. Wiedza o posłużeniu się w Ko-deksach Teodozjusza i Justyniana byłaby niezwykle pomocna historykom prawa, gdyż znajomość rytmiki ułatwiłaby studia nad rekonstrukcją pierwot-nego brzmienia rytmicznego tekstu. Jak wiadomo, zachowały się dziesiątki rękopiśmiennych wersji wspomnianych kodeksów, a wyboru odpowiednich lekcji i emendacji tekstów przeprowadzano bez świadomości, że są to teksty rytmiczne.

Ostatecznie użycie programu komputerowego dowiodło, że proza rytmicz-na z przewagą veloxu „czasownikowego” rozwijała się od początku IV w. aż po wiek VI włącznie, a następnie odrodziła się już na początku wieku IX, a nie, jak dowodził Tore Janson w pierwszej połowie XI w. Trudno zresztą było uwierzyć, iż proza taka wymarła na dwa wieki (VII i VIII w.), aby następnie sama z siebie się odrodzić. Dalsze przeszukiwania tekstów dowiodły, iż proza rytmiczna z ve-loxem „czasownikowym” była nieprzerwanie w użyciu w kancelarii papieskiej od VI w. do końca średniowiecza (!)<sup>14</sup>. Nie było żadnej przerwy w użyciu tego stylu, a słynna reforma Jana z Gaety to nic innego, jak zalecenie dla urzędników, jak należy pisać dokumenty w związku z wycofaniem w tym czasie z użycia w kancelarii papieskiej księgi formularzowej *Liber diurnus*, pochodzącej z około 700 r., a która niemal cała była spisana z użyciem prozy rytmicznej z przewagą veloxu „czasownikowego”. Brak tej wiedzy spowodował, iż zarówno słynna edycja Haralda Zimmermanna, jak i monografia Hansa-Henninga Kortüma,

12 D. Liebs, *Konstantin der Grosse. Geschichte – Archäologie – Rezeption; internationales Kolloquium vom 10. – 15. Oktober 2005 an der Universität Trier zur Landesausstellung Rheinland-Pfalz 2007 „Konstantin der Grosse”*, hrsg. A. Demandt, Trier 2006, s. 97-107.

13 *Idem*, *Rhythmische Rechtssätze. Zur Geschichte einiger lateinischer Rechtsregeln*, „Juristenzeitung” 1981, nr 5-6, s. 160-164 – [www.jstor.org/stable/20815070](http://www.jstor.org/stable/20815070) [01.03.2015].

14 T. Jasiński, *Początki cursusu w kancelarii papieskiej w świetle „Liber diurnus” oraz tzw. „Zacharias-Bonifatius-Privileg”. Przyczynek do badań nad prozą późnoantyczną i wczesnośredniowieczną*, w: *Między tekstem a znakiem. Prace ofiarowane Profesor Barbarze Trelińskiej w siedemdziesiątą rocznicę urodzin*, pod red. A. Jaworskiej i S. Górzyń-skiego, Warszawa 2013, s. 361-372.

poświęcona kancelarii papieskiej od końca IX w. do połowy XI w., pominęły wiele istotnych elementów badawczych<sup>15</sup>. Kortüm stara się często odgadnąć, jak wyglądały pierwotne teksty w księgach ówczesnej kancelarii papieskiej, popełniając przy tym wiele błędów, które mógłby uniknąć, gdyby wiedział, iż ma do czynienia z tekstami rytmicznymi.

Znaczenie tego rodzaju programu trudno przecenić; za jego pomocą można śledzić interpolacje w kronikach i żywotach Świętych, późniejsze wtręty w dokumentach papieskich, w tym też w formułach *Libri diurni*. Program ten ma również ogromne zasługi w poszukiwaniu tekstów pokrewnych z *Kroniką polską* Galla Anonima. W sumie udało się wykryć aż 40 tekstów z wysokim udziałem veloxu „czasownikowego”, które powstały przed Gallem. Gdy prace nad analizą „ręczną” tych tekstów się przeciągały, doszedłem do wniosku, iż należy jeszcze raz zmodyfikować program. Do tego momentu miał – jak pamiętamy dwa okienka – które służyły do podania liczby veloxów i liczby znaków. To pozwalało poszukiwać teksty z prozą rytmiczną z wysokim udziałem veloxu „czasownikowego” typu Gall. Pamiętałem jednak, iż cechą charakterystyczną prozy Galla jest również wysoki udział trispondaicusa. Na moją prośbę Piotr Jasiński dodał do programu jeszcze dwa okienka, w których mogłem wpisać dane dotyczące trispondaicusa oraz liczbę znaków. Trzecia wersja programu, po ponownej kalibracji, była gotowa do wykrycia prozy rytmicznej podobnej do Galla, a więc z dużym udziałem cursusu veloxowego „czasownikowego” oraz trispondaicusa też „czasownikowego”. Obawiając się, żeby program nie przeoczył żadnego podobnego do Galla tekstu, bardzo szeroko „ustawiłem” owe cztery okienka. Po ponownym przeanalizowaniu wszystkich tekstów (900 tys. stron) program „złowił” 11 zabytków historiograficznych. Dziewięć z nich okazało się mało zbliżonych do Galla, „złowienie” ich było skutkiem bardzo ostrożnego zdefiniowania danych w czterech okienkach. Jednym z dwóch dzieł podobnych do Galla była wspomniana *Translacja św. Mikołaja*, co było wiadome na podstawie moich wcześniejszych analiz. Drugim zabytkiem był czterostronicowy żywot pewnego merowińskiego świętego, spisany na początku XII w. w Walonii. Jednym słowem program odnalazł 4 strony wśród 900 tys. stron. Badania nad wymienionym zabytkiem zostaną przedstawione w oddzielnej pracy.

15 *Papsturkunden 896-1046*, bearb. von H. Zimmermann, t. 1: 896-996, Wien 1984 (Veröffentlichungen der Historischen Kommission. Österreichische Akademie der Wissenschaften, Philosophisch-Historische Klasse; 3. – Denkschriften. Österreichische Akademie der Wissenschaften, Philosophisch-Historische Klasse; 174); H.-H. Kortüm, *Zur päpstlichen Urkundensprache im frühen Mittelalter. Die päpstlichen Privilegien 986-1046*, Sigmaringen 1995, *passim*.

## 2. Założenia teoretyczne „kombajnu” informatycznego do badań łacińskich tekstów prozatorskich i poetyckich zarówno z okresu starożytnego, jak i średniowiecza, a nawet czasów nowożytnych

Pożytek płynący ze zdefiniowania i napisania wymienionego programu zachęcił mnie do stworzenia jeszcze kilku innych programów, a więc programu do porównywania tekstów, programu do poszukiwania 16-zgłoskowców trocheicznych, które dotychczas udało mi się odnaleźć tylko w *Kronice polskiej* i *Translacji św. Mikołaja*. W końcu doszedłem do wniosku, że najlepsze byłoby powołanie większego zespołu, gdzie informatyk wraz z filologami klasycznymi oraz historykami mógłby stworzyć zupełnie nowe dzieło informatyczne, które mogłoby służyć wszystkim badaczom, również uczonym z zagranicznych ośrodków. Program taki otrzymał w końcu 2012 r. wsparcie Narodowego Programu Rozwoju Humanistyki. Przewiduje się uruchomienie owego „kombajnu” na początku 2017 r.

Wielofunkcyjny program komputerowy realizowany w ramach tego projektu nie jest powtórzeniem programów do badania Galla ani nawet jego odmianą. Nowy program będzie opierał się na wpisaniu do bazy wszystkich łacińskich słów wraz ze wszelkimi odmianami gramatycznymi. W trakcie wpisywania, które wykonują wyłącznie filologowie łacińscy, zdefiniowana została długość każdej sylaby każdego słowa, w każdej formie gramatycznej (czasowniki mają niekiedy ponad 200 form). Przy wpisywaniu odmian gramatycznych, wraz z długością sylab, posłużyliśmy się specjalnymi „podprogramami” komputerowymi, które znacznie skracają czas wpisywania, a równocześnie ograniczą do całkowitego minimum ludzkie błędy. Dobór słowników, na podstawie których będzie określona długość samogłosek, a także główne pryncypia merytoryczne „podprogramów” powierzone zostały młodemu, doskonałemu znawcy greckiego i łacińskiego iloczasu – filologowi klasycznemu, doktorowi Piotrowi Stępniewi. On również objął nadzór nad korektą wszystkich wpisów, tak aby w bazie słownikowej nie znalazł się żaden błąd. Następnie do tak zbudowanej bazy wstępnej zostaną dodane jako warianty dodatkowe zasady pisowni średniowiecznej z licznymi wariantami średniowiecznymi – tak gramatycznymi, jak i ortograficznymi. Po dokonaniu wszystkich wpisów grupa filologów łacińskich oznaczy dodatkowo powiązania między wyrazami na podstawie rdzeniowej (odwołujemy się w tym wypadku do doświadczeń prof. Zenona Kałuży) oraz powiązania pomiędzy synonimami. Te ostatnie są wprowadzane w dwóch grupach: oddzielnie dla tekstów klasycznych i oddzielnie dla średniowiecznych; będzie jednak istniała opcja programu, która pozwoli na uwzględnienie jednocześnie dwóch grup synonimów.

Tak zbudowana baza leksykalna stanie się podstawą dla funkcjonowania programu, który będzie posiadał szereg różnych funkcji. Wszystkie te funkcje będą zestawione dla użytkownika na jednej stronie internetowej.

Pierwsza funkcja programu umożliwi określanie pokrewieństwa dwóch tekstów, z uwzględnieniem zasobu słownictwa i jego częstotliwości; mogłyby to być teksty małe, jak i olbrzymie pliki tekstowe. Zasada programu będzie następująca: wszystkie słowa obu tekstów zostaną sprowadzone do podstawowej formy gramatycznej, następnie zestawione według częstotliwości, przy czym program obliczy, jaki promil (a niekiedy jaką częśćkę promila) tekstu stanowi dane słowo oddzielnie w każdym z tekstów. Po dokonaniu tej procedury w obu tekstach program zestawia wspólnie występujące słowa wraz z ich „udziałami” w całym tekście mierzonymi w promilach. Następnie program zsumuje promile, wybierając zawsze niższą wartość; np., jeśli w tekście A udział słowa *habeo* liczy 0,45 promila, a w tekście B tylko 0,37 promila, to do sumowania zostanie wybrana mniejsza wartość, czyli 0,37 promila, gdyż tylko do tej wartości oba teksty są całkowicie paralelne zarówno w zakresie słownictwa, jak i częstotliwości. Gdyby dwa teksty były identyczne, suma owa stanowiłaby 1000 promili, a im bardziej słownictwo będzie różne i im bardziej różna będzie częstotliwość wspólnego słownictwa, to pokrewieństwo to będzie bardziej odległe, co wyrazi się niższą wielkością (sumą) w promilach. Korzystając z tego programu, będzie można ustalić, że np. słownictwo Thietmara jest bliskie kronikarzowi X czy też Ojcowi Kościoła Y. Pozwoli on również na badanie wpływu wielkich pisarzy antycznych (Cycerona, św. Augustyna, *etc.*) na poszczególnych kronikarzy średniowiecza. To, oczywiście, tylko przykładowe możliwości badawcze, płynące z wykorzystania programu. W istocie jednak jego zastosowanie w badaniach naukowych stwarza niemal nieograniczone perspektywy poznawcze.

W ramach powyżej wspomnianej funkcji każdy korzystający uzyska możliwość zestawienia wyrazów według częstotliwości w obu tekstach i oddzielnie, w kilku różnych wariantach, które ułatwią dalszą analizę tekstową. Wszystkie warianty zostały już dokładnie przemyślane, pomijam je ze względu na konieczność ograniczenia szczegółowości<sup>16</sup>.

Funkcja programu ustalająca pokrewieństwo badanych tekstów w zakresie słownictwa i jego częstotliwości pozwalałaby jedynie na zorientowanie

16 Wiadomo, że niektóre formy gramatyczne w języku łacińskim mogą wywodzić się od dwóch różnobrzmiących i mających inne znaczenie słów, np. słowo *esto* może wywodzić się od *edo* lub od *sum*. Program będzie oddzielnie zestawiał takie słowa, tak aby badający miał możliwość rozstrzygnięcia, o jaką podstawę i znaczenie w danym wypadku chodzi.



się, czy mamy do czynienia z bliskimi czy też odległymi od siebie tekstami. Historyk czy latynista potrzebuje jednak dla swoich badań wiedzy nie tylko o tego rodzaju pokrewieństwie. Badacz chciałby poszukiwać zbieżnych fraz lub zdań. Umożliwi to druga funkcja tego programu pozwalająca na zestawienie najbardziej podobnych zdań lub fraz między dwoma badanymi tekstami. Funkcję tę zilustruje poniższy przykład. Załóżmy więc, że badamy Kronikę Wincentego Kadłubka i chcemy odnaleźć w licznych tekstach zbieżne z Kroniką zdania lub frazy. Do okna A „wkładamy” wtedy elektroniczny tekst Kroniki Wincentego (będzie go można pobrać z bazy będącej częścią tego programu), a do okna B wielki plik tekstowy, zawierający liczne kroniki XI-XII wieku. Te ostatnie są dość powszechnie dostępne w Internecie oraz w specjalnych bazach tekstowych oferowanych przez wydawców tekstów elektronicznych. Wpisujemy do okna zadaniowego programu, że życzymy sobie np. 30 najbardziej zbieżnych fraz czy zdań. Program zestawi wszystkie frazy, które są identyczne w obu badanych tekstach. Wiadomo jednak, że zbliżone frazy w obu tekstach różnią się często kolejnością wyrazów, przypadkami gramatycznymi, *etc.* Te trudności pokona druga opcja tej funkcji, czyli zestawienie najbardziej podobnych zdań. Opcja ta umożliwi odnalezienie podobnych wyrażen, nawet jeśli różnią się one pod względem gramatycznym. Program, po wybraniu tej funkcji, odwoła się do swojej bazy słownikowej i sprowadzi wszystkie wyrazy w obu plikach do form podstawowych, a następnie zestawi 30 najbardziej zbliżonych zdań. Oczywiście liczba zdań jest dowolna i może być zastąpiona liczbą np. 29 albo 78. Program zestawi zdania od najbardziej podobnych do mniej podobnych. Co istotne, przy poszukiwaniu podobnych zdań istnieć będą nie tylko specjalne funkcje wyłączające z poszukiwania określone słowa (np. *et*, *ad*, *etc.*), ale również opcja wyłączająca wyrazy dwu-, trzy- czy czteroliterowe. Wyeliminuje to zestawianie zdań z przypadkowymi podobieństwami.

Po przeprowadzeniu tego rodzaju poszukiwań badający będzie miał możliwość włączenia dodatkowych opcji, czyli np. uwzględnienia podobieństw rdzeniowych oraz podobieństw z uwzględnieniem synonimów. Powracając do naszego przykładu, badający wprowadzając ogromny plik, np. kilkudziesięciu kronik, *etc.*, miałby możliwość odszukania podobieństw do Kroniki Wincentego Kadłubka za jednym „kliknięciem”. Oczywiście funkcja ta umożliwi porównywanie wszelkich tekstów, np. poszczególnych mów Cycerona między sobą, albo tychże mów tego retora z wybranymi kronikami średniowiecznymi, *etc.*

Kolejną funkcją programu będzie oznaczanie rytmiki akcentowej w badanym dziele. Jako że w bazie słownikowej podana zostanie długość każdej sylaby każdego wyrazu w każdej gramatycznej formie, to natychmiast – uwzględniając

wszystkie dodatkowe szczególne przypadki – za pomocą programu będzie można określić, gdzie pada akcent w danym słowie<sup>17</sup>.

Możliwość określenia akcentu pozwala określić rytmikę akcentową, czyli cursus całego badanego tekstu. Ponieważ pisarze średniowieczni, a także antyczni (np. Ammianus) „kładli” cursus w różnych miejscach, to funkcje programu do wykrywania cursusu będą wyposażone w wiele opcji. Wprowadzenie tych opcji pozwoli też na wykrycie cursusu w tych miejscach badanego dzieła, gdzie jego obecności nikt nie spodziewa, czyli np. w innych miejscach niż zakończenia klauzul. Planuje się następujące opcje:

- a) zidentyfikowanie wszystkich rytmów akcentowych;
- b) zaznaczenie rytmów tylko na końcu zdania;
- c) zaznaczenie rytmów na końcu akapitu;
- d) zaznaczenie rytmów tylko przed przecinkiem i kropką;
- e) zaznaczenie rytmów przed „et”, „atque”, „ac” i innych znakach np. przed wyrazem zakończonym na „que”;
- f) możliwość wpisania znaków do tekstu, przed którym ma być oznaczony cursus.

Program pozwoli zidentyfikować obok rytmów podstawowych także rytmy „*per consillabicationem*”.

Rewolucyjną wręcz innowacją będzie opcja wprowadzenia do programu dowolnej sekwencji akcentowej, zdefiniowanej według powszechnie używanych przez badaczy kodów. Np. badacz będzie mógł sprawdzić, czy dość częsty u Galla rytm pp/1-3p (rodzaj *veloxu*) występuje u Hildeberta z Lavardin.

Wszystkie funkcje określające rytmikę akcentową będą wyposażone w dodatkowe opcje umożliwiające wyszukiwanie w tekście fragmentów, gdzie mamy do czynienia z odpowiednią ilością danych rytmów na określonej liczbie znaków; pozwoli to wykryć prozę z wysokim udziałem poszczególnych rytmów, co jest bezcenne dla historyka i dla latynisty.

Możliwość natychmiastowego określenia cursusu badanego tekstu trudno przecenić; wyniki uzyskane za pomocą cursusu pozwolą nakreślić nowe, nieznane linie rozwojowe piśmiennictwa łacińskiego, ułatwią poszukiwanie pokrewnych dzieł, umożliwią datowanie zabytków literatury łacińskiej, posłużą pomocą w wyszukiwaniu interpolacji lub miejsc zepsutych w badanych tekstach tak narracyjnych, jak i dokumentowych. Program ten będzie z pewnością też bardzo pożyteczny dla prawników zajmujących się prawem rzymskim, o czym

17 W języku łacińskim niekiedy długość sylaby wynika z formy gramatycznej, np. w wyrazie *relegimus* zgłoska *le* może być długa lub krótka, w zależności czy jest to, czas *praesens* czy *perfectum*. Gdy program napotka na taką ewentualność, poinformuje, że mamy tu do czynienia z dwiema możliwościami.

pisałem wyżej. Dzięki temu programowi zostanie rozwiązana zagadka częściowego zaniku cursusu w VII-VIII w. Dziś już wiemy, że przetrwał on w kancelarii papieskiej, ale czy tylko tam?

Niezwykle możliwości badawcze przed latynistami otwierałaby możliwość badania rytmiki opartej na iloczasiu, której początki przypadają na czasy Cyncerona i która staje się wszechobecna od schyłku III w. aż do początków VI w. Funkcja ta – podobnie jak w przypadku cursusu akcentowego – będzie wyposażona w różne funkcje, umożliwiające identyfikację klauzul i fraz akcentowych w różnych miejscach. Pozwoli to skorygować dotychczasowe poglądy na rozwój późno-antycznej literatury, umożliwi zidentyfikowanie kierunków rozprzestrzeniania się wzorców literackich. Przykładowo przyjmuje się, że późno-antyczne kadencje iloczasowe rozwijały się pod przemożnym wpływem dzieł Cyncerona. To jest niewątpliwe, ale np. latyniści amerykańscy Ralph G. Hall i Steven M. Oberhelman<sup>18</sup> stwierdzili, że we fragmentach mów Cyncerona na 908 klauzul znaleźli tylko 21 klauzul kretyk-trybrach, która to klauzula stała się później bardzo popularna w łacińskiej prozie Cesarstwa Rzymskiego. Badacze ci przyjmowali, że klauzula ta zawdzięcza swoją popularność temu, iż stała się podstawą dla cursusu *tardus*. Dzięki programowi będzie można np. w jednej chwili przebadać wszystkie klauzule Cyncerona (Tadeusz Wojciechowski, naliczył ich 17 902) i wszystkie inne rytmiczne elementy jego mów, a następnie stwierdzić precyzyjnie, ile było w jego mowach klauzul kretyk-trybrach. Następnie pod tym kątem będzie można przeszukać inne dzieła piśmiennictwa łacińskiego z okresu Cesarstwa i ustalić, kiedy ta klauzula się upowszechniła i czy rzeczywiście jej geneza wiąże się z możliwością wykorzystania jej jako podstawy rytmu akcentowego *tardus*. Dzięki programowi to zagadnienie będzie można rozwiązać też „od tyłu”, a mianowicie ustalić za pomocą opcji rytmiki akcentowej najstarsze dzieła z rytmem *tardus* i sprawdzić, też za pomocą programu, w jakim stopniu opierał się on na klauzuli kretyk-trybrach.

Program otworzy też możliwości badania rytmiki całych zdań czy wypowiedzi prozy łacińskiej. Wybitni latyniści, Eduard Norden i Tadeusz Zieliński, twierdzili, że kluczowy dla rytmiki prozy był nie tylko rytm kadencji, ale też rytm obecny w całym zdaniu (wypowiedzi), a zwłaszcza na początku zdania. Program pozwoli zbadać tę rytmikę, opartą na iloczasiu, a w jakimś mniejszym stopniu najpewniej też na akcencie, od czasów narodzin piśmiennictwa łacińskiego aż po czasy, gdy zanikł iloczasiu.

18 R. G. Hall, S. M. Oberhelman, *Rhythmical Clausulae in the Codex Theodosianus and the Leges Novellae ad Theodosianum pertinentes*, „Classical Quarterly” 1985, t. 35, nr 1, s. 201 – [www.jstor.org/stable/638816](http://www.jstor.org/stable/638816) [01.03.2015].

Dodatkową opcją będzie możliwość badania równoległego obu rytmów – iloczynowego i akcentowego, co będzie niezwykle pomocne przy badaniu prozy Cesarstwa Rzymskiego, gdy najbardziej popularnym był wspomniany wyżej *cursus mixtus*, stanowiący połączenie w jedną całość kadencji rytmicznych i akcentowych. *Cursus mixtus* powstawał w okresie od połowy III w. aż po pierwsze dziesięciolecie V wieku. Z kadencji iloczynowych zalecanych i używanych przez Cicerona i jego kontynuatorów zaczęto wybierać takie, które były jednocześnie rytmami akcentowymi, np. peon-trochej dla rytmu *trispondaicus*; kretyk-trochej dla rytmu *planus*, *etc.* Program miałby możliwość równoczesnego wyświetlenia iloczynu badanych rytmów akcentowych, co pozwoliłoby na ponowne zbadanie początków, rozwoju i zaniku *cursus mixtus*. Ten ostatni był „rytmem mieszanym” też z tego względu, że w prozie tego okresu występują obok siebie trzy rodzaje rytmów:

1. całkowicie oparte na iloczynie;
2. rytmy jednocześnie oparte na iloczynie i na akcencie;
3. rytmy oparte wyłącznie na akcencie.

Udział statystyczny tych trzech rytmów jest jednym z elementów pozwalających na określenie czasu powstania dzieła. Są oczywiście wyjątki, do których zaliczają się np. Ammianus oraz częściowo św. Hieronim i św. Augustyn; powstanie tych wyjątków mogłoby zostać wyjaśnione dzięki możliwościom tego programu.

Na koniec trzeba dodać, że program będzie oferował możliwość sprawdzania ortografii zarówno tekstów klasycznych, jak i średniowiecznych. Stanie się to bardzo pożyteczne – jako wstępna „obróbka” tekstu – dla wydawców tekstów łacińskich, ale jeszcze bardziej dla badaczy, którzy za pomocą komputerowych programów typu OCR zamieniali skany i zdjęcia, a także pliki PDF z tekstami łacińskimi na programy tekstowe, przeszukiwalne.

Ponieważ Biblioteka Kórnicka od wielu lat uczestniczy w tworzeniu elektronicznych narzędzi badawczych dla nauk humanistycznych, do wyżej przedstawionego programu zostaną dołączone dwie bazy tekstowe i jedna baza urzędników Rzeczypospolitej (XII-XVIII w.). Wszystkie trzy bazy pozwolą uczonym zajmującym się problematyką polską prowadzić najważniejsze badania historyczne i latynistyczne. Nie ma bowiem potrzeby umieszczania baz z tekstami starożytnego piśmiennictwa łacińskiego, gdyż są one dostępne dzięki wielu bazom elektronicznym, którymi dysponują wszystkie ważniejsze biblioteki na świecie. Wszystkie ważniejsze teksty starożytne są też dostępne w Internecie w formie przeszukiwalnej, np. *Latin Library*. Podobnie sprawa przedstawia się z tekstami średniowiecznego piśmiennictwa łacińskiego, które dostępne są badaczowi dzięki dziesiątkom edycji elektronicznych i specjalnym bazom dostępnym na stronach internetowych (np. *Compendium Auctorum Latinorum*

Medii Aevi – C.A.L.M.A.). To właśnie z tych baz, lub jakichkolwiek innych, badacze mogą pobierać teksty łacińskie, czy to starożytne, czy średniowieczne, a nawet późniejsze i poddawać je badaniom przez ogólnodostępny program zamieszczony na stronie Biblioteki Kórnickiej PAN.

Znacznie gorzej natomiast przedstawia się sytuacja z polskimi tekstami elektronicznymi – są one dość trudno dostępne. Dlatego do dyspozycji badaczy zostaną oddane podstawowe kroniki i żywoty polskie wydrukowane w *Monumenta Poloniae Historica*, (t. I-VI, Lwów-Kraków 1864-1893). To ostatnie dzieło jest wprawdzie ogólnodostępne w wersji internetowej ([www.kpbc.umk.pl/publication/12939](http://www.kpbc.umk.pl/publication/12939) [01.03.2015]), ale nie jest dostępne w wersji przeszukiwalnej. Do bazy zostaną wprowadzone w formie przeszukiwalnej następujące dzieła (pisownia oryginalna):

1. Jana Kanaparza Żywot ś. Wojciecha, MPH 1, s. 162-171;
2. Brunona Żywot ś. Wojciecha, MPH 1, s. 189-222;
3. Galla kronika, MPH 1, s. 390-484;
4. Kronika Węgierskopolska, MPH 1, s. 495-515;
5. Ebbonis vita Ottonis episcopi babenbergensis, MPH 2, s. 32-70;
6. Herbordi vita Ottonis episcopi babenbergensis, MPH 2, s. 71-127;
7. Monachi Priefligensis vita Ottonis episcopi babenbergensis, MPH 2, s. 128-144;
8. Mierzwy kronika, MPH 2, s. 163-192;
9. Mistrza Wincentego kronika i jej skrócenie przez bezimiennego dopełniacza kroniki Mierzwy zrobione, MPH 2, s. 249-447;
10. Kronika Boguchwała i Godysława Paska, MPH 2, s. 467-600;
11. Kronika Janka z Czarnkowa, MPH 2, s. 619-758;
12. De sancto Adalberto episcopo, MPH 4, s. 209-221;
13. Kronika książąt polskich (*Chronica principum Poloniae*), MPH 3, s. 423-578;
14. Kronika polska (*Chronica Polonorum*), MPH 3, s. 578-656;
15. Miracula sancti Adalberti, MPH 4, s. 226-238;
16. Vita sancti Stanislai episcopi Cracoviensis (Vita minor), MPH 4, s. 253-285;
17. Miracula sancti Stanislai, MPH 4, s. 293-318;
18. Vita sancti Stanislai episcopi Cracoviensis (Vita maior), MPH 4, s. 355-438;
19. Exordium ordinis Cruciferorum seu Chronica de Prussia, MPH 6, s. 290-309;
20. Chronica Olivensis, MPH 6, s. 310-350.

Trzy przyczyny zdecydowały o wyborze tych dzieł. Po pierwsze, są to najważniejsze zabytki historiografii polskiego średniowiecza (z wyjątkiem *Roczników* Jana Długosza, których tu brak); po drugie, są to teksty zwarte, w przeciwieństwie do roczników i luźnych dokumentów; po trzecie, przekształcenie tych dzieł w bazę cyfrową nie narusza niczyich praw autorskich.

W przyszłości Biblioteka Kórnicka rozpocznie starania o środki, aby w tej bazie zamieścić wszystkie teksty łacińskie dotyczące Polski, powstałe przed 1500 rokiem.

Dla potrzeb dyplomatyki w ramach tego projektu badawczego zostaną zamieszczone dokumenty 11 tomów Kodeksu dyplomatycznego Wielkopolski, zawierające 4834 dokumentów oraz 8 tomów Zbioru dokumentów małopolskich z 2570 dokumentami. Łącznie w bazie znajdzie się zatem około 7400 dokumentów. Najstarsza część tych edycji nie jest już chroniona prawami autorskimi, a jeżeli chodzi o tomy wydane w XX wieku, to Biblioteka Kórnicka uzyskała zgodę od wydawców lub spadkobierców ich praw do umieszczenia obu serii w bazie tego projektu. Tysiące tych dokumentów dzięki programowi będzie mogło być punktem odniesienia lub przedmiotem badań za pomocą funkcji, które oferuje niniejszy projekt. Zarówno baza tekstów narracyjnych, jak i dokumentowych byłyby ogólnodostępne. Należy zaznaczyć, że obie serie są już gotowe i nie wymagają większych nakładów. Kodeks dyplomatyczny Wielkopolski został przez Bibliotekę Kórnicką umieszczony w formie przeszukiwalnej na stronie internetowej Wielkopolskiej Biblioteki Cyfrowej ([www.wbc.poznan.pl/publication/12458](http://www.wbc.poznan.pl/publication/12458) [01.03.2015]), a Zbiór dokumentów małopolskich wymaga tylko drugiej korekty oraz opracowania informatycznego.

Do badań nad dokumentami średniowiecznymi, a zwłaszcza nad ich autentycznością najważniejsza jest analiza listy świadków, a w tym zwłaszcza urzędników. Biblioteka Kórnicka uzyskała zgodę od autorów 25 tomów serii – wydawanej najczęściej przez Bibliotekę Kórnicką i poświęconej urzędnikom dawnej Rzeczypospolitej. W ramach tego projektu powstanie zatem baza, która byłaby niezastąpionym narzędziem do badania przede wszystkim polskich dokumentów, akt i innych tekstów (też historiograficznych) od XII do XVIII w. Wielu autorów tej serii wyraziło chęć bezpłatnego uaktualnienia rezultatów badań, gdyż po publikacji dzieła w formie papierowej udało się w wielu wypadkach dokonać nowych ustaleń. Baza danych urzędników dawnej Rzeczypospolitej, będzie zawierała informacje o osobach i sprawowanych przez nich urzędach na przestrzeni wieków (XII-XVIII w.) z terenów należących do Rzeczypospolitej. W celu łatwego i ogólnodostępnego przeglądania zgromadzonych w bazie zasobów stworzony zostanie interfejs, dostępny z poziomu przeglądarki internetowej. Dzięki niemu będzie istniała możliwość wyselekcjonowania jedynie tych danych, które będą interesujące dla danego użytkownika. Do bezproblemowej obsługi bazy danych zostanie stworzona aplikacja, umożliwiająca uzupełnianie oraz poprawianie zawartych w niej zasobów przez osoby do tego uprawnione.

Mamy nadzieję, że ów „kombajn” informatyczny dobrze będzie służył latynistom i historykom z całego świata.

## Summary

### Tomasz Jasiński *Science in the service of Latin philology and history*

At the Kórnik Library of the Polish Academy of Sciences, a team of historians, classical philologists, librarians and computer scientists have undertaken (2012-2016) a grant project entitled *A new generation of computer programs to analyze Latin narrative and documentary texts together with an electronic database of texts and a database of Polish state officials of the twelfth through the eighteenth centuries*. This program will be the world's first research tool containing a database of the entire Latin vocabulary with all the grammatical and chronemic values. It will be a multifunctional and generally accessible program to examine Latin poetry according to vowel length and rhythm, as well as to detect chronemic flow, mixture (eg. mixtus) and accent.