

KONCEPCJA AUTOMATYZACJI KLASYFIKACJI RZECZOWNIKA JĘZYKA STARO-CERKIEWNO-SŁOWIAŃSKIEGO W GRUPY DEKLINACYJNE

PIOTR WIERCHOŃ*

W niniejszym artykule przedstawiam metody algorytmizacji rozpoznania numeru grupy deklinacyjnej rzeczowników języka staro-cerkiewno-słowiańskiego (dalej: scs.). Wyniki uzyskane w ramach poniższych obserwacji mogą w pierwszym odbiorze wydać się niezbyt zajmujące – mamy tu bowiem do czynienia z formalnym operowaniem na ciągach znaków; co więcej: piszący te słowa nie proponuje bezpośredniego wykorzystania własnych analiz dla jakichkolwiek sądów uogólniających. Istnieje jednak nadzieja wyjścia w pół drogi badaczom zajmującym się dziedziną, na którą niniejszym tekstem chciano spojrzeć z nieco innej, jak sądzę – dotychczas mało eksplorowanej strony.

Ufam przeto, iż sugestie wysunięte niżej w pewnej mierze podziałają inspirująco na badaczy scs-u. Jeśli moja metoda pracy nad językiem scs. wywoła przynajmniej jeden głos polemiczny, uznaję, że cel, jaki sobie postawiłem, opracowując niniejszy *algorytm*, został spełniony.

Główne problemy prezentowanych niżej rozwiązań algorytmicznych przedstawiam w następującym porządku:

- i) pojęcia ogólne; założenia algorytmu,
- ii) wprowadzenie pojęcia *wektora*, omówienie wektorów:
 - „*ьm*” – I, II
 - „*ьm*” – I, IV
 - „*on*” – I, V
 - „*iz*” – III, V
- iii) ogólne zasady działania algorytmu,
- iv) schemat algorytmu (jądro koncepcji automatyzacji klasyfikacji),

* Stypendysta Fundacji na rzecz Nauki Polskiej w 2001 roku.
1 W sensie niehilbertowskim.

- v) omówienie schematu,
- vi) końcowe wnioski.

1. Klasyfikacji rzeczowników dokonują na podstawie dwu informacji:

- przynależności kategorialnej rzeczownika, wyznaczonej własnościami wymiaru Terminalność;
 - przynależności kategorialnej rzeczownika, wyznaczonej własnościami wymiaru Rodzaj;
- Wymiar Terminalność zawiera cechy będące dowolnymi ciągami złożonymi ze znaków alfabetu scs-u, np.: „ьm”, „ьm”, „on”, „ØI” itd.
Wymiar Rodzaj zawiera cechy: „męskość – m”, „żeńskość – ż”, „neutralność – n”.

Algorytm przypisuje odpowiednią grupę deklinacyjną rzeczownika (dalej – *hasło wejściowe*) na podstawie informacji dwójki rodzaju: formy graficznej hasła rzeczownika o znaczeniu l. poj. i rodzaju gramatycznego tegoż rzeczownika. „Droge” od startu algorytmu do wyniku nazywam *procedurą*. Wynikiem zaś jest liczba z przedziału 1 – 5. Każda zatem procedura kończy się wynikiem, czyli odpowiednio ponumerowaną grupą deklinacyjną (oprócz wyniku Ø). Numerację tę podają za Bartulą (1994). Przez *strukturę* hasła rozumiem jego budowę pojmowaną jako ciąg znaków alfabetu scs. Np. *vlbkb* ma następującą strukturę: Ø, v, l, b, k, b, Ø.

Przedstawiając dziedzinę badań, zaznaczam na wstępie, iż nie dysponowałem indeksem *a tergo* języka scs., stąd ograniczam się do ilustrowania koncepcji prezentowanego algorytmu, wykorzystując ręcznie ekscerpowany materiał leksykalny zawarty w pozycji Bartula (1994). Ścisłej rzecz biorąc: na uwadze miałem hasła z zawartego tam *Słowniczka* Bartula (1994: 185-203). Obserwacją objąłem każde z haseł rzeczownikowych, zaopatrzonych w informację: *m* [rodzaj męski], *f* [rodzaj żeński] lub *n* [rodzaj neutralny].

2. Przyjrzyjmy się poniższej tabeli:

ьm	I, II	ьm	I, IV	am	III	im	III
ьž	Ø	ьž	IV	až	III	iž	III, V
ьn	Ø	ьn	Ø	an	Ø	in	Ø
em	Ø	ym	V	om	Ø	em	Ø
ež	Ø	yž	V	ož	Ø	ež	Ø
en	I	yn	Ø	on	I, V	en	V

Wektory poziome o ogólnej charakterystyce XYZ czytamy:

- X – symbol własności z wymiaru Terminalność (ostatni znak hasła),
- Y – symbol własności z wymiaru Rodzaj,
- Z – numer deklinacji.

Przykład:

ьm I, II czytamy: ostatnim znakiem hasła jest znak ь oraz rzeczownik reprezentowany przez to hasło należy do zbioru rzeczowników męskich, a ostatecznie jest elementem grupy deklinacyjnej albo I, albo II;

ьž Ø czytamy: nie istnieje taki rzeczownik, który byłby rzeczownikiem żeńskim i reprezentujące go hasło kończyłoby się znakiem ь; tym samym hasło takie nie należy do żadnej z grup deklinacyjnych.

Jak widać z powyższej tabeli, otrzymujemy 4 wektory, dla których informacje wejściowe nie implikują w sposób jednoznaczny numeru grupy deklinacyjnej; są to wektory: ьm, ьn, on, iž. Pozostałe warianty nie wymagają objaśnień.

Zatrzymajmy się z kolei przy problemie mianownika l. poj. rzeczownika męskiego zakończonego znakiem ь. Wahania w przyporządkowaniu klasy deklinacyjnej nastąpić mogą pomiędzy grupą I a II. Proponujemy zastosować tu dwa rozwiązania; pierwsze – słownikowe, polegające na spisaniu wszystkich wyrazów należących do deklinacji II. Jest to rozwiązanie proste i czytelne, ale nie indukuje żadnych spostrzeżeń natury algorytmiczno-językoznawczej. Druga możliwość przedstawia się następująco. Wyliczamy wszystkie hasła należące do II grupy deklinacyjnej; ich struktury przedstawiają się następująco:

synb, volb, domb, medb, vrbxb, polb, činb, sanb, mirb, ledb.

Z podanych struktur wyróżniamy ciąg dwu ostatnich znaków (pozostałych po odcięciu ь). Otrzymujemy dwuwartościowe struktury 2 znaków o ogólnej charakterystyce YZ, gdzie Y oznacza znak oznaczający samogłoskę lub jer, Z oznacza znak oznaczający spółgłoskę; są to:

in, ed, om,
an, yn, ьx,
ir, ol,

gdzie ol i ed występują po dwa razy.

Kolejno analizujemy materiał słownikowy zawarty w pozycji Bartula (1994). Badaniu podlegały wszystkie rzeczowniki z zakresu od a do m. Wyniki pracy przedstawia się niżej.

Jeżeli spełnione jest YZ, gdzie YZ są ostatnimi znakami ciągu zakończonego znakiem ь, to ciąg ten klasyfikujemy do grupy deklinacyjnej X. Np. jeżeli YZ = ol (co czytamy: jeżeli końcowe znaki hasła wejściowego oznaczonego YZ = ol i ostatni znak ciągu równy jest ь), to klasyfikujemy ten ciąg do II grupy deklinacyjnej. Pobieźna nawet analiza powyższej propozycji prowadzi do pytania, co zrobić, jeżeli rzeczownik scs. kończy się ciągiem -oľ i – będąc rzeczownikiem męskim – należy do I grupy deklinacyjnej? Taką sytuację zauważamy w parze:

poľb – II grupa,
glagoľb – I grupa.

Propozycja rozwiązania tego problemu przedstawia się następująco. W analizowanym materiale słownika języka scs. zauważyć można, że badana próba leksykalna nie zawiera innych rzeczowników rodzaju męskiego, które kończyłyby się znakiem Ъ (tj. należała do kategorii wyznaczanej cechą Terminalności -Ь) i posiadały dwa ostatnie znaki swej struktury równe YZ: yn, om, ed, ьx. Pozostałe ciągi: in, ir, ol zawierają także inne wyrazy, tzn. nie należące do II grupy deklinacyjnej. W badanej próbie są to:

in – *gospodinь, arxitriklinь* (gospodarz wesela), *činь*,
ir – *mirь*,
ol – *glagolь*.

Wszystkie te wyrazy należą do I grupy deklinacyjnej. Sporne mogą być ciągi² *činь* (por. Batrula 1994: 186) i *mirь*, które będą kategoryzowały wyrazy I deklinacji lub II deklinacji. Powyższe uwagi znajdują swe odzwierciedlenie w algorytmie. Możliwe jest także podanie teoretycznej liczby ciągów nas interesujących (tzn. o symbolu YZь).

W niniejszym artykule przedstawiamy wyniki analiz 41 z tych ciągów, co oznacza, że w materiale a – m *Słowniczka* Bartula (1994) znalazło się 41 różnych charakterystyk YZь.

Problem wektora ьm – I, IV został rozwiązany następująco. Sprawdzono w całym *Słowniczku* Bartula (1994) wszystkie wyrazy należące do kategorii wyznaczanej cechą „-ь”, należące do kategorii wyznaczanej cechą „męskości” z wymiaru Rodzaj, jednocześnie należące do deklinacji IV. Otrzymana liczba wyrazów równa się 8; są to m.in. *patь, zvěрь, ognь, gostь*. Wszystkie osiem przykładów realizuje ciągi (po odcięciu -ь): ěr, ąt, řt, ьt, st, od, gn, gl. Kolejną czynnością było sprawdzenie próby wyrazów rozpoczynających się od litery a do litery m włącznie i wybranie tych rzeczowników, które kończą się znakiem Ъ, należą do kategorii rzeczowników męskich i należą do deklinacji I. Otrzymane ciągi YZ to: ol, ar, el, ko, ьn, bl, el', bc, on, ęz, ec, aź, ьl.

Jak widać, żadna z par YZ właściwa dla deklinacji I nie pokrywa się charakterystyką z analogicznym ciągiem YZ rzeczownika z IV deklinacji.

Problem wektora on – I, V przedstawia się następująco. W badanym materiale słownikowym Bartula (1994) znajdują się cztery wyrazy należące do deklinacji V: *čudo, drěvo, nebo, slovo*. Niejednoznaczny jest status ciągu *lice*. Gdyby przyjąć tok wykładu Bartuli (*Przegląd (tabela) końcówek przypadkowych deklinacji* (Bartula 1994: 37) oraz skonfrontować tenże z zestawionym w cytowanej pozycji *Słowniczkiem*, czytelnik

² Celowo stosujemy w tym miejscu pojęcie ciągu, a nie wyrazu, gdyż w rzeczywistości nie mamy na uwadze wyrazu, a jedynie kategorię graficzną, wyznaczaną cechą – „čin?”, do której należy każdy ciąg znaczący tę cechę, tj. nie mamy na uwadze intersekcji cech semantycznych (ani gramatycznych, ani leksykalnych). Fałszywe jest bowiem stwierdzenie, iż wyraz *čin?* należy do I lub II deklinacji. Wyraz ten należy albo do deklinacji I, albo do deklinacji II.

napotyka problem. Otóż forma dopełniacza z rozszerzonym tematem -es-e (a taką podaje się w *Słowniczku*) miałyby przyporządkowany mianownik zakończony na -e. Powracając natomiast do zestawienia Bartula 1994: 37, łatwo skonstatować, iż rzeczownik *lice* podany w *Słowniczku* nie istnieje (sic!). Uznając nawet silniejsze oddziaływanie tematu rozszerzonego z V deklinacji, oczekiwać by można końcówki dla kategorii neutralnej -o. Nie doszukano się jednak w zabytkach scs-u. formy zakończonej na -o (bardziej konsekwentne – ale znowu jeszcze inne – interpretowanie tego rzeczownika odnajdujemy w pozycji Moszyński (1984: 331)). Zakończenie mianownika liczby pojedynczej jest różne od zakończenia rzeczowników z V deklinacji, natomiast dopełniacz -ese jest analogiczny. 25 rzeczowników z kategorii neutralnych o ostatnim znaku hasła -o należy do I deklinacji. Charakterystyka realizacji YZ wygląda, jak następuje:

ad, ud, tv, ьn, ęd, br, ig, er, ęt, ąt, yr, ęł, ęt, al, el, il, in, st, ag, at.

Charakterystyka realizacji YZ dla V deklinacji wygląda, jak następuje:

ud, ęv, eb, ok, ęł, ov.

Problem wektora iż – III, V został rozwiązany następująco. Znalaziono jedynie dwa rzeczowniki należące do kategorii wyznaczanej cechą żeńskości oraz należące do kategorii wyznaczanej cechą terminalną -i; są to *ďbřiti* i *mati* (notabene są to – wydate się – jedne dwa rzeczowniki z poświadczonym tematem -er- (-r-)). Łatwo zauważyć, iż po odcięciu ostatniego znaku wyrazy te należą do kategorii wyznaczanej znakiem -t. Pod uwagę zatem – w konstrukcji algorytmu – wystarczy wziąć tenże znak.

3. Zasada działania algorytmu.

a. Podanie graficznej postaci mianownika liczby pojedynczej.

hasło

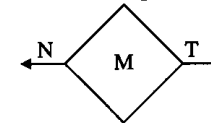


b. Podanie rodzaju gramatycznego.

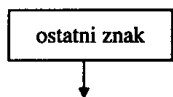
rodzaj



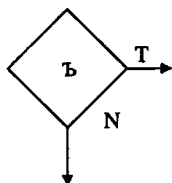
c. Analiza w polu decyzji: Jeśli M, idź w prawo, jeśli nie M, idź w lewo.



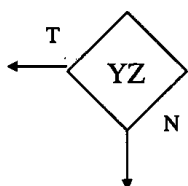
d. Podanie ostatniego znaku wejściowego.



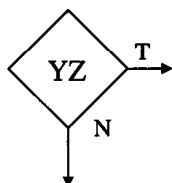
e. Analiza w polu decyzji: Jeśli ъ, analizuj dalej w prawo.



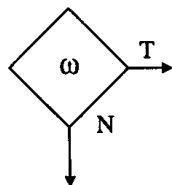
f. Analiza w polu decyzji: Np.: jeśli YZ = an, ed, yn, om, ъx, rzeczownik hasła o cesze terminalnej – jak podano w alternatywie dla YZ – należy do deklinacji II.



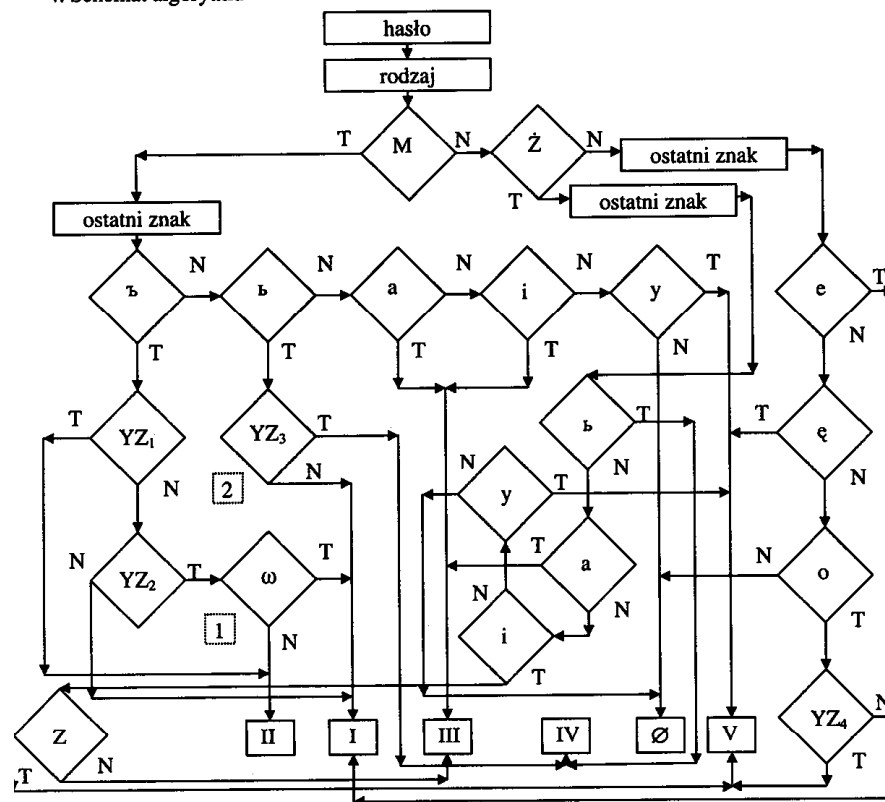
g. Analiza w polu decyzji: Jeśli YZ = in, ir, ol, idź do następnego pola decyzyjnego. Jeśli YZ ≠ in, ir, ol, rzeczownik należy do I deklinacji.



h. Analiza w polu decyzji: Jeśli ciąg równa się ω, to rzeczownik należy do I deklinacji. Jeżeli ciąg nie równa się ω, to rzeczownik należy do II deklinacji.



4. Schemat algorytmu



- YZ₁ = an, ed, yn, om,
- YZ₂ = in, ir, ol,
- YZ₃ = ěr, at, gn, gl, št, zt, od, st,
- YZ₄ = ud, eb, ěv, ov, ok, ěl,
- Z = t,
- ω = gospodinъ, arxitriklinъ, činъ, mirъ, glagolъ.

5. Omówienie schematu

Końcowych wyjaśnień wymagają dwie kwestie. Miejsca w algorytmie oznaczone cyframi 1 i 2 mogą nasuwać pewne wątpliwości. Oto próba ich rozwiązania.

a) Grupa wyrazów 1. oznaczona odpowiednimi strukturami jest niepełna z powodu badań materiałowych (tj. ich braku), a stanowi jedynie uwzględnienie wyrazów rozpoczynających się od litery a do m.

b) Uznajemy wyrazy ciągów³ *čĩnъ* oraz *mirъ* jako należące do deklinacji I. Rozwiązanie to jest rodzajem rozwiązania arbitralnego, a więc w pewnych warunkach można by te wyrazy zaklasyfikować do grupy II. Kryterium decydującym byłoby zatem uwzględnienie znaczenia leksykalnego wyrazu. Niezależnie jednak od powyższego wskazania struktura algorytmu – biorąc pod uwagę przyjętą metodę konstrukcji – pozostaje *conatus*.

Grupa wyrazów 2. oznaczona odpowiednimi strukturami – por. wyżej pkt. a.

6. Uwagi końcowe⁴

W niniejszej pracy zaprezentowano próbę algorytmizacji przyporządkowania poszczególnych rzeczowników języka scs. danym grupom deklinacyjnym. Przyporządkowanie to ma charakter funkcji, a jak powszechnie wiadomo, wynikiem działania funkcji jest przyporządkowanie jednoznaczne, tj. każdemu elementowi ze zbioru A może zostać przyporządkowany dokładnie jeden element ze zbioru B. Wynika zatem z tego to, iż każdy rzeczownik musi należeć do jednej grupy deklinacyjnej.

³ Mówimy: „wyraz ciągu *čĩn?*”, albowiem typ mówienia: „wyraz *čĩn?*” już na wstępie nakazuje jednoznaczną interpretację leksykalną wyrazu, podczas gdy pojęcie ciągu znaków takiego założenia nie może spełniać, na czym nam w tym wypadku zależy. Np. ciąg *komputer* znaczy nic innego (w ramach referowanego ujęcia!) jak własność: mienie jako formę graficzną rzędu o postaci „komputer” – wyznaczanych cechą terminalną „-komputer”, gdzie wyraz *komputer* ma własność: „urządzenie elektroniczne itd.” (tutaj dowolna definicja leksykograficzna).

⁴ Już po napisaniu niniejszego tekstu na kilka ważnych kwestii zwrócił mi uwagę pan mgr Artur Stęplewski, któremu jednocześnie chciałbym mocno podziękować za jego lekturę i krytyczne uwagi. Otóż dopowiedzieć należy w tym miejscu, iż:

a) algorytm jest adekwatny jedynie w obrębie badanego materiału; uwzględnienie innego zbioru rzeczowników powodować może zmianę jego konstrukcji; biorąc pod uwagę charakter języka scs. i uwzględniając jego wariant cyrylometodejski, można wnosić, iż klasa interesujących nas wyrazów jest już ustalona; to bardzo dobra okoliczność dla konstruowania algorytmu tego typu, co prezentowany w niniejszym artykule,

b) owa możliwość (natomiast nie: konieczność) bierze się stąd, iż (być może) ekstrapolowanie wyróżnionych ciągów na pozycji drugiej i trzeciej od końca na grunt bardziej obszernego korpusu potwierdziłoby zasadność zainteresowania się (w sensie analiz dla konkretnego algorytmu) drugim i trzecim znakiem od końca,

c) wydaje się, iż warto sfalsyfikować lub udowodnić tezę: o przynależności do grupy deklinacyjnej rzeczownika scs. decyduje postać wyrazu o znaczeniu mianownika, nie zaś formy mianownika i dopełniacza; innymi słowy: aby rozpoznać, do jakiej deklinacji należy dany rzeczownik, wystarczy znać jedynie jego postać mianownika (przynależność do kategorii z cechy wymiaru Terminalność) i informację o jego przynależności do kategorii wyznaczonej cechą wymiaru Rodzaj.

Rozważania takie wcale nie muszą być rozważaniami pozbawionymi praktycznych konsekwencji w sensie *stricte* językoznawczym i dydaktycznym. Gdyby bowiem uznać powyższą uwagę, zniknąłby problem (pozornego!) dwojakiego kwalifikowania niektórych wyrazów do niejednej grupy deklinacyjnej. Po prostu takie sytuacje nie zachodzą w świecie postarystotelesowskim.

Bibliografia

- Aho, J., E. Hopcroft, J.D. Ullman. 1983. *Projektowanie i analiza algorytmów komputerowych*. Warszawa: PWN.
- Bartula, Czesław. 1994. *Podstawowe wiadomości z gramatyki staro-cerkiewno-słowiańskiej*. Warszawa: PWN.
- Harel, D. 1992. *Rzecz o istocie informatyki. Algorytmika*. Warszawa: PWN.
- Moszyński, L. 1984. *Wstęp do filologii słowiańskiej*. Warszawa: PWN.
- Porayski-Pomsta, J. 1986. Rola przepisów algorytmicznych w kształtowaniu podstaw racjonalnego myślenia. *Poradnik Językowy* 7. 486-493.
- Syso, Maciej, M. 1998. *Piramidy, szyszki i inne konstrukcje algorytmiczne*. Warszawa. Wydawnictwa Szkolne i Pedagogiczne.
- Schmalstieg, W. R. 1983. *Introduction to Old Church Slavic (second edition, revised and expanded)*. Ohio: Columbus.