ZARZĄDZANIE W KULTURZE
2017, 18, z. 2, s. 113–134
doi:10.4467/20843976ZK.17.008.7100
www.ejournals.eu/Zarzadzanie-w-Kulturze

Zbigniew Handzel, Mirosław Gajer

JĘZYKI KONTROLOWANE JAKO NARZĘDZIE INFORMATYCZNE WYKORZYSTYWANE W SŁUŻBIE RATOWANIA DZIEDZICTWA KULTUROWEGO WYMIERAJĄCYCH JĘZYKÓW

Abstract

CONTROLLED LANGUAGES AS A COMPUTER TOOL USED FOR SAVING THE CULTURAL HERITAGE OF ENDANGERED LANGUAGES

At present, still over 7000 languages are in use all over the world, however, this number is systematically declining from year to year. With the death of languages, an important part of human heritage is being lost. For this reason, saving endangered languages from a complete oblivion is a an urgent task that linguists try to cope with. The attempts undertaken so far have mainly concentrated on a passive recording of lexical resources of endangered languages, which, however, fails to be a successful method of keeping such languages alive. The authors of the article have proposed an innovative approach based on the application of an IT related concept of a controlled language, which allows for recording not only vocabulary of an endangered language, but its syntactic structures as well. The controlled languages along with electronic tools in form of syntactic structures generators and e-translators that the authors work on will equip future learners of endangered languages with a helpful and innovative instrument to enhance the learning process.

SŁOWA KLUCZE: języki zagrożone wymarciem, ratowanie języków, przekład komputerowy, lingwistyka komputerowa, języki kontrolowane

KEY WORDS: endangered languages, saving languages, machine translation, computational linguistics, controlled languages

1. Wprowadzenie

Według różnych szacunków obecnie na świecie jest wciąż jeszcze w użyciu około 7 tysięcy różnych języków.[1] Oczywiście, w rozważanym wypadku precyzyjne podanie dokładnej ich liczby nie jest w ogóle możliwe, choćby z tego względu, że nie są znane, przynajmniej jak dotychczas, żadne precyzyjne kryteria, które pozwalałyby na definitywne rozstrzygniecie kwestii, czy w danym przypadku mamy do czynienia z już całkowicie odrębnym językiem, czy też może w grę wchodzi jedynie pewien dialekt bądź jakaś regionalna lub środowiskowa odmiana danego języka[2].

Z tych powodów podczas dokonywania klasyfikacji języków świata częstokroć zdecydowanie górę biorą względy natury politycznej nad argumentami o charakterze czysto merytorycznym. Przykładowo, jeszcze trzydzieści lat temu mówiono powszechnie o wspólnym dla Serbów i Chorwatów języku serbsko-chorwackim, natomiast obecnie, po rozpadzie dawnej Jugosławii, terminem tym się już w zasadzie nikt nie posługuje, gdyż powszechnie wyróżnia się dwa całkowicie odrębne języki: chorwacki (zapisywany wyłącznie alfabetem łacińskim) oraz serbski (zapisywany zarówno alfabetem łacińskim, jak i cyrylicą). Tymczasem różnice pomiędzy językiem chorwackim a serbskim są naprawdę niewielkie i sprowadzają się głównie do lokalnych odmian leksykalnych oraz pewnych różnic fonetycznych w wymowie niektórych wyrazów. Istnieją także drobne różnice o charakterze składniowym, ale nie mają oczywiście jakiegoś większego znaczenia i w żadnym wypadku nie stanowią jakiejkolwiek przeszkody we wzajemnej komunikacji użytkowników obu wymienionych języków.[3]

Ponadto można śmiało zaryzykować twierdzenie, że różnice występujące między językami serbskim a chorwackim nie są wcale większe niż różnice zachodzące pomiędzy standardową polszczyzną a używanymi na południu naszego kraju gwarami góralskimi, podczas gdy tym ostatnim nie przyznaje się w żadnym wypadku statusu odrębnego języka bądź języków (wziąwszy pod uwagę również ich wzajemne zróżnicowanie).

Podobnie w grupie języków słowiańskich wyróżnia się język bułgarski i język macedoński, podczas gdy równie dobrze języki te mogłyby być uznane za regionalne odmiany jednego języka, ponieważ ich użytkownicy nie mają bynajmniej najmniejszych problemów ze wzajemną komunikacją[4].

Jednocześnie w językoznawstwie powszechnie mówi się, na przykład, o licznych dialektach języka chińskiego, i to w sytuacji, gdy zachodzące pomiędzy nimi różnice są naprawdę poważnej natury. Szczególnie duże różnice dostrzegane są pomiędzy dialektami języka chińskiego używanymi na północy i na południu Chin.

[1] J.A. Matisoff, *Zagrożona różnorodność: języki i formy życia*, „Świat Nauki", październik 2002, s. 66–73.

[2] A.F. Majewicz, *Języki świata i ich klasyfikowanie*, Warszawa 1989.

[3] H. Dalewska-Greń, *Języki słowiańskie*, Warszawa 2002.

[4] T. Lehr-Spławiński, W. Kuraszkiewicz, F. Sławski, *Przegląd i charakterystyka języków słowiańskich*, Warszawa 1954.

Przykładowo, używany na obszarze Hong-Kongu chiński dialekt kantoński różni się od standardowego chińskiego mandaryńskiego mniej więcej tak, jak język niemiecki odróżnia się od języka islandzkiego (a może nawet bardziej), w związku z czym wzajemna komunikacja ich użytkowników jest w zasadzie niemożliwa. Warto zwrócić uwagę, że według powszechnie panujących na świecie opinii wszyscy Chińczycy mówią po prostu po chińsku, gdy tymczasem w rzeczywistości mamy niewątpliwie do czynienia z wieloma bardzo różnymi językami, splątanymi w niekończącą się sieć wzajemnych powiązań i dialektów[5].

Gdy analizujemy poszczególne języki świata, tym, co przykuwa szczególną uwagę, są gigantyczne wręcz dysproporcje zachodzące pomiędzy liczbami użytkowników poszczególnych języków. Z jednej strony mamy zaledwie kilka wielkich języków, które mają aż po kilkaset tysięcy użytkowników (chiński mandaryński – 845 mln, hiszpański – 329 mln, angielski – 328 mln, arabski – 221 mln, hindi – 182 mln, bengalski – 181 mln, portugalski – 178 mln, rosyjski – 176 mln, japoński – 122 mln). Podane powyżej liczby dotyczące użytkowników danych języków mają oczywiście charakter szacunkowy i w przypadku różnych opracowań mogą nawet dość znacznie odbiegać od siebie. Zamieszczone wartości autorzy podali na podstawie serwisu internetowego dostępnego na stronie o adresie http://www.ethnologue.com.

W pobliżu setnego miejsca na liście będących obecnie w użyciu języków świata są języki mające około siedmiu milionów użytkowników, gdy tymczasem na drugim biegunie występuje kilka tysięcy języków mających co najwyżej po kilkadziesiąt, kilkanaście bądź zaledwie po kilka tysięcy, czy nawet znacznie mniej użytkowników.

W rozważanym kontekście warto jest także wspomnieć o podejmowanych próbach budowania odpowiednich modeli matematycznych i symulacyjnych, których zadaniem jest wyjaśnianie, w jaki sposób mogły powstać tego rodzaju dysproporcje odnośnie do liczby użytkowników poszczególnych języków, jednak wciąż są to jeszcze stosunkowo słabo poznane i niedostatecznie zrozumiane przez badaczy procesy o bardzo złożonej naturze[6].

Jednak tym, co niezwykle istotne z punktu widzenia rozważań prowadzonych w dalszej części niniejszego artykułu, jest dość powszechnie akceptowana przez lingwistów prognoza, według której przynajmniej połowa z będących obecnie w użyciu języków świata do końca bieżącego wieku całkowicie zniknie z powierzchni kuli ziemskiej[7]. Taki stan rzeczy jest wysoce niepokojący, ponieważ każdy z istniejących języków naturalnych wnosi bezsporny wkład do szeroko pojmowanego dziedzictwa kulturowego ludzkości. Ponadto każdy język naturalny stanowi swoisty i jedyny w swym rodzaju sposób postrzegania świata oraz unikatowy środek, służący do formułowania myśli o obiektywnej rzeczywistości otaczającej jego użytkowników, jednak zawsze silnie osadzonej w pewnym oryginalnym kontekście kulturowym. Z tego powodu śmierć każdego języka naturalnego, następująca zawsze w momencie

[5] J.M. Künstler, *Języki chińskie*, Warszawa 2000.

[6] C. Schulze, D. Staffer, S. Wichmann, S. Birth, *Survivaland Death of languages by Monte Carlo Simulation*, „Communications in Computational Physics" 2008, vol. 3, nr 2.

[7] J.A. Matisoff, *Zagrożona różnorodność: języki...*, s. 66–73.

zejścia ze świata ostatniego z jego natywnych użytkowników, stanowi bez jakiegokolwiek cienia wątpliwości niepowetowaną stratę dla całej ludzkości, gdyż w istotnym stopniu zubaża bogactwo różnorodności kulturowej świata[8].

Dodatkowo proces wymierania języków naturalnych stanowi także niepowetowaną stratę dla światowej nauki. Językoznawstwo jest mimo wszystko dyscypliną naukową, która może być wciąż jeszcze uznawana za stosunkowo młodą, w związku z czym w dziedzinie tej pozostało zapewne jeszcze bardzo wiele do zrobienia. W rozważanym kontekście wystarczy tylko wspomnieć fakt, że przeważająca liczba języków naszej planety nie została jak dotychczas w dostateczny sposób opisana (brak odpowiednich słowników, leksykonów, gramatyk opisowych itp.) i należycie sklasyfikowana, gdyż przynależność wielu ważnych języków do jednostek wyższego rzędu, takich jak grupa czy rodzina językowa, wciąż w licznych przypadkach wywołuje kontrowersje i budzi poważne wątpliwości[9].

W tym miejscu wystarczy tylko wspomnieć o głoszonej powszechnie jeszcze kilkadziesiąt lat temu koncepcji istnienia wielkiej rodziny języków chińsko-tybetańskich, podczas gdy obecnie mówi się o przynajmniej czterech odrębnych rodzinach językowych: chińskiej, tybetańskiej, birmańskiej i tajskiej, przy czym ich wspólne pochodzenie od jakiegoś, będącego w użyciu w zamierzchłych czasach, prajęzyka jest raczej definitywnie wykluczone[10]. Istniejące pomiędzy wymienionymi rodzinami językowymi podobieństwa – sprowadzające się do tego, że zawierają one języki, które są językami wybitnie analitycznymi (pozycyjnymi), natomiast od strony fonologicznej są to języki monosylabiczne i wykazujące silnie rozbudowany system toniczny, zawierający w niektórych wypadkach kilka odrębnych rejestrów tonów – tłumaczy się wyłącznie wpływem wzajemnych kontaktów ich użytkowników. W związku z powyższym wymienione podobieństwa mają charakter wyłącznie typologiczny, a nie genetyczny[11].

Analogiczne kontrowersje występują w wypadku rodziny języków ałtajskich, gdyż obecnie coraz większa liczba badaczy skłania się ku poglądowi, że w rozważanym przypadku mamy do czynienia aż z trzema odrębnymi rodzinami językowymi: turecką, mongolską i mandżurską, a dostrzegane pomiędzy nimi pewne podobieństwa są li tylko typologicznej natury i nie stanowią w żadnym wypadku dowodu na ich pochodzenie od wspólnego przodka[12,13]. Pewni badacze próbują jednak włączyć do rodziny ałtajskiej również takie języki, jak koreański i japoński, które według nich miałyby być najdalej na wschód wysuniętą grupą rozważanej rodziny językowej[14].

Proces wymierania licznych języków świata o małej liczbie użytkowników jest również niepokojący z punktu widzenia wielu ważnych badań językoznawczych

[8] M.C. Corballis, J.L. Dessalles, R. Dunbar, *Aux origines du langage*, „La Recherche" 2001, nr 341, s. 27–39.

[9] A.F. Majewicz, *Języki świata...*

[10] A. Bareja-Starzyńska, M. Mejor, *Klasyczny język tybetański*, Warszawa 2002.

[11] J.M. Künstler, *Języki chińskie*, Warszawa 2000.

[12] S. Godziński, *Współczesny język mongolski*, Warszawa 1998.

[13] S. Kałużyński, *Klasyczny język mongolski*, Warszawa 1998.

[14] A. Kondratow, *Zaginione cywilizacje*, Warszawa 1988.

prowadzonych nad tzw. uniwersaliami językowymi. Lingwiści od dawna zadają sobie pytanie, czy istnieje zbiór pewnych niezbędnych kategorii gramatycznych, które muszą być obecne w każdym języku ludzkim i bez których proces międzyludzkiej komunikacji językowej byłby zupełnie niemożliwy. Obecnie do tego rodzaju kategorii gramatycznych zalicza się, między innymi, zaimki osobowe, ponieważ jak dotychczas nie jest bynajmniej znany żaden ludzki język, który byłby całkowicie pozbawiony jakiegoś (nawet bardzo prostego) systemu zaimków osobowych. Aczkolwiek pomiędzy różnymi językami świata istnieją w tym względzie spore rozbieżności, gdyż równie licznie występują na świecie zarówno języki mające bardzo proste systemy zaimków osobowych (na przykład języki chińskie), jak i języki posiadające bardzo rozbudowane tego rodzaju systemy gramatyczne (na przykład języki arabskie)[15,16]. Jak już wspomniano, na świecie nie jest znany żaden język naturalny pozbawiony całkowicie systemu zaimków osobowych, gdy jednak weźmie się pod uwagę fakt, że zdecydowana większość będących obecnie w użyciu języków świata nie została jeszcze należycie zbadana, opisana i skatalogowana, formułowanie w rozważanym zakresie jakichkolwiek kategorycznych twierdzeń może być mimo wszystko dość ryzykowne. Z tego powodu intensywne obecnie wymieranie licznych niewielkich języków świata stanowi poważne ryzyko, że potencjalnie interesujący materiał badawczy może już wkrótce zniknąć całkowicie – i to bez jakiegokolwiek śladu – z powierzchni ziemi, w związku z czym pewne ważne z punktu widzenia lingwistyki problemy badawcze pozostaną na zawsze już nierozstrzygnięte, a na potwierdzenie odpowiednich hipotez badawczych próżno będzie szukać dowodów w istniejącym materiale językowym.

2. Języki zagrożone wymarciem

Proces wymierania języków zawsze w przeszłości następował i nie jest bynajmniej czymś nowym, jednak w ostatnich latach przybrał na sile z niespotykaną dotychczas intensywnością. Jak już wspomniano, języki naturalne wciąż wymierały i losu tego nie były w stanie uniknąć nawet języki wielkich cywilizacji, czego przykładem może być całkowite wymarcie takich języków, jak sumeryjski, akadyjski, babiloński, fenicki, aramejski, hetycki, egipski czy minojski[17].

Podobnie za martwe należy uznać języki dwóch wielkich cywilizacji europejskich, czyli łacinę i starożytną grekę. Aczkolwiek łacina dała początek całej gałęzi współczesnych języków romańskich, to jednak różnice, które zaszły podczas rozwoju tych języków w ciągu kilkunastu wieków, są tak duże, że raczej wykluczają swobodne zrozumienie tekstów zapisanych w języku łacińskim przez współczesnych użytkowników języków, takich jak włoski, francuski, hiszpański, kataloński, portugalski bądź

[15] J. Danecki, *Klasyczny język arabski*, Warszawa 1998.
[16] J. Danecki, *Współczesny język arabski i jego dialekty*, Warszawa 2000.
[17] A. Kondratow, *Zaginione...*

rumuński. Zwłaszcza ostatni z wymienionych języków zdecydowanie odbiega od pozostałych języków romańskich, ponieważ jego leksyka i w pewnej mierze również niektóre konstrukcje składniowe zostały w dużej mierze ukształtowane pod wpływem sąsiadujących z nim języków słowiańskich. W wypadku współczesnego języka greckiego różnice zachodzące w stosunku do jego starożytnego poprzednika są nawet jeszcze większe, niż w wypadku łaciny i współczesnych języków romańskich.

W ramach indoeuropejskiej rodziny językowej za całkowicie wymarłą uważana jest grupa języków anatolijskich, która nie była w przeszłości w stanie przetrwać niezwykle silnej ekspansji języka tureckiego[18]. Z kolei w czasach nam współczesnych za zagrożoną wymarciem uważana jest w pierwszym rzędzie grupa języków celtyckich[19].

Językiem, który z wymienionej grupy językowej najwcześniej odszedł w zapomnienie, jest język kornicki, będący jeszcze w XVII wieku w powszechnym użyciu na terenie Kornwalii. Prawdopodobnie śmierć ostatniej osoby, która biegle posługiwała się językiem kornickim, jako językiem pierwszym, nastąpiła w roku 1777. Analogicznie w 1974 roku zmarł prawdopodobnie ostatni z użytkowników języka manx, będącego ongiś w użyciu na brytyjskiej wyspie Man. Sytuacja pozostałych języków grupy celtyckiej wcale nie wygląda lepiej. Przykładowo, szkockim językiem gaelickim posługuje się obecnie prawdopodobnie nie więcej niż około 20 tysięcy osób. Podobnie, językiem irlandzkim biegle włada zaledwie niewiele ponad 1% populacji Irlandii, pomimo uznania tego języka za język urzędowy Republiki Irlandii i łożenia znacznych środków finansowych na jego szeroko zakrojone propagowanie, mające na celu ratowanie go przed całkowitym wyginięciem. Nieco lepiej sytuacja wygląda w wypadku języka walijskiego (około 550 tysięcy użytkowników) i języka bretońskiego (około 500 tysięcy użytkowników), chociaż w perspektywie najbliższych kilkuset lat potencjalnego zagrożenia tych języków również nie można wykluczyć.

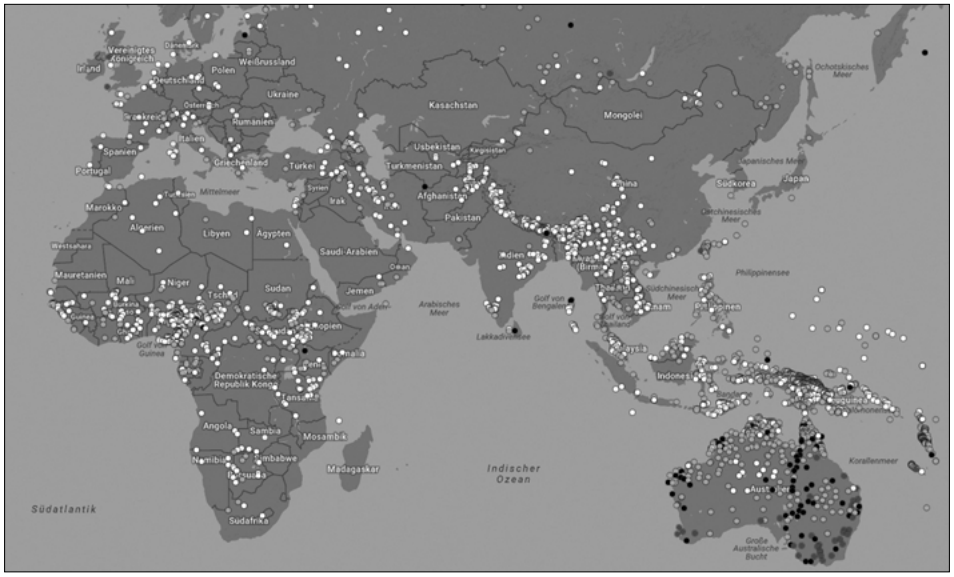
W językoznawstwie uważa się powszechnie, że jeżeli tylko jakiś język zaczyna być postrzegany przez używającą go społeczność jako w pewnym sensie język gorszy, zwłaszcza odznaczający się niższym statusem społecznym w stosunku do języka sąsiedniej społeczności, wówczas jego los jest już w zasadzie z góry przypieczętowany, a sytuacja całkowitego wyjścia z powszechnego użycia i w konsekwencji definitywne wymarcie jest li tylko kwestią upływającego czasu[20].

Z wymienionego powodu w roku 1880 odszedł, wraz ze śmiercią ostatniego z jego użytkowników, używany niegdyś na Szetlandach i Orkadach, północnogermański język norn, który został całkowicie wyparty przez dokonujący spektakularnej ekspansji język angielski. Warto nadmienić, że po wymienionym języku zostało w zasadzie bardzo niewiele jakichkolwiek zabytków o charakterze piśmiennym, w związku z czym obecnie jego pełna rekonstrukcja w ramach odpowiednich projektów badawczych nie wydaje się w ogóle możliwa.

[18] M. Popko, *Ludy i języki starożytnej Anatolii*, Warszawa 1999.
[19] A.F. Majewicz, *Języki świata*...
[20] J.A. Matisoff, *Zagrożona różnorodność: języki*..., s. 66–73.

Na il. 1 zaprezentowano fragment mapy świata z zaznaczonymi lokalizacjami, w których występują społeczności ludzkie posługujące się językami zagrożonymi w najbliższym czasie całkowitym wymarciem. Jak wynika z mapy przedstawionej na il. 1, statystycznie najwięcej języków zagrożonych wymarciem występuje w centralnej części Afryki oraz w Azji Południowo-Wschodniej, a także w Australii. Jednak wielu zagrożonych wymarciem języków naturalnych można doszukać się również w Europie Zachodniej. To przede wszystkim liczne małe języki należące w ramach rodziny indoeuropejskiej przede wszystkim do grup językowych romańskiej i germańskiej. Ponadto potencjalnie zagrożone wymarciem są wszystkie języki należące do grupy języków celtyckich.

Z kolei w ramach grupy języków słowiańskich za poważnie zagrożone wymarciem należy uznać języki dolnołużycki i górnołużycki. Oba wypada postrzegać jako ostatnie z żywych reliktów dawnej słowiańskiej warstwy językowej, występującej niegdyś powszechnie na terenie wschodnich Niemiec.



Il. 1. Lokalizacje geograficzne języków zagrożonych wymarciem
(źródło: http://www.endangeredlanguages.com [odczyt: 2.02.2017])

Język dolnołużycki używany jest jeszcze przez nie więcej niż 10 tysięcy osób w rejonie niemieckiego miasta Cottbus. Nieco więcej użytkowników, gdyż około 20 tysięcy, posiada będący w użyciu w okolicach miasta Bautzen język górnołużycki. Jednak liczba użytkowników obu wymienionych języków z roku na rok systematycznie spada, co stwarza realne zagrożenie, że po upływie kilku pokoleń podzielą one nieuchronnie los pozostałych języków słowiańskich, będących ongiś w użyciu

na terenie obecnych Niemiec[21]. Autorzy mają w tym miejscu na myśli takie języki, jak między innymi połabski, który wymarł całkowicie około roku 1750, i język słowiński, który wyszedł z użycia mniej więcej w roku 1900.

Niestety, sytuacja obu języków łużyckich nie wygląda obecnie dobrze, nawet pomimo podejmowania zintensyfikowanych działań, mających na celu ich ratowanie. Język niemiecki posiada w stosunku do rozważanych języków słowiańskich zdecydowanie wyższy prestiż społeczny, ponadto jest to język publikacji tekstów o charakterze naukowym i technicznym, w związku z czym młodsze pokolenia Łużyczan nie są specjalnie zainteresowane kultywowaniem tradycji ich przodków i zamiast zgłębiać tajniki nader skomplikowanej gramatyki języków łużyckich – między innymi w językach tych występują formy liczby podwójnej, tzw. dualisu, które tworzone są w przypadku rzeczowników, przymiotników i czasowników, gdy mowa jest o dokładnie dwóch osobach, rzeczach bądź zjawiskach – niejako w sposób naturalny przechodzą na język niemiecki, co sprawia, że w perspektywie najbliższych kilkudziesięciu lat los języków łużyckich wydaje się w zasadzie już przesądzony, a wszelkie działania mające na celu ich ratowanie okazują się w praktyce mało skuteczne. Powodem wymienionego stanu rzeczy jest oczywiście miażdżąca przewaga języka niemieckiego, który przez młodsze pokolenia Łużyczan jest postrzegany jako o wiele bardziej atrakcyjny i charakteryzujący się o wiele wyższym statusem społecznym, dlatego jego biegła znajomość jest bezwzględnie konieczna, aby móc się kształcić i później efektywnie funkcjonować w społeczeństwie. Niestety, tego samego nie można w żadnym wypadku powiedzieć o jakimkolwiek z języków łużyckich[22].

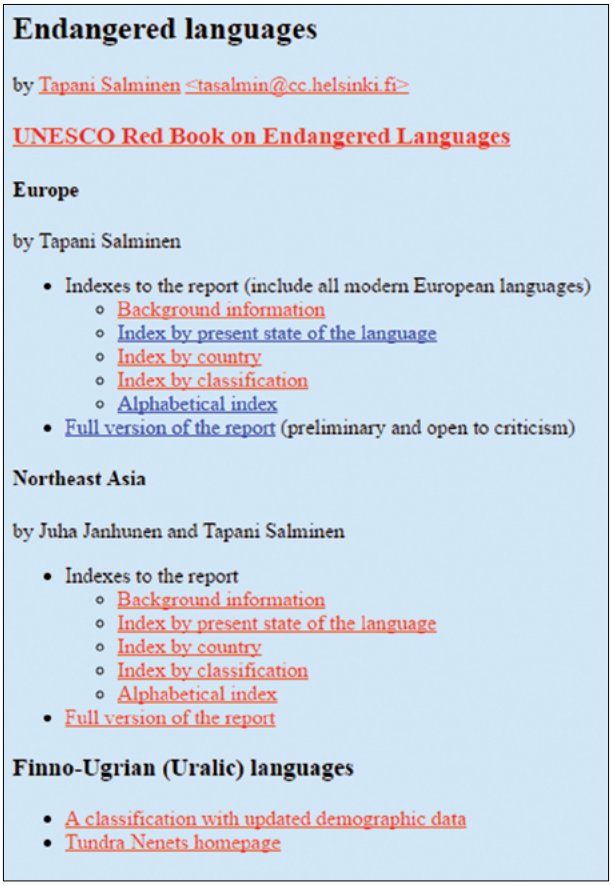
3. Ratowanie wymierających języków

Obecnie na świecie podejmowane są różnorodne inicjatywy mające na celu ratowanie ginących w zastraszającym wręcz tempie języków przed całkowitą zagładą i popadnięciem w definitywne zapomnienie. W tym miejscu wypada przede wszystkim wspomnieć o tzw. czerwonej księdze, zawierającej spis zagrożonych języków świata, która została sporządzona przez UNESCO.

Wymieniony dokument jest dostępny w postaci strony internetowej o adresie http://www.helsinki.fi/~tasalmin/endangered.html [odczyt: 4.02.2017]. To zarazem jedno z najbardziej obszernych repozytoriów wiedzy na temat obecnie zagrożonych wymarciem języków świata, w którym można znaleźć liczne cenne informacje dotyczące bieżącego stanu poszczególnych języków, takie jak między innymi: aktualna liczba ich użytkowników, stopień kompetencji językowych wśród młodszego pokolenia oraz ocena stopnia zagrożenia wymarciem. Natomiast w przypadku języków, które uważane są już za całkowicie martwe, podawana jest przybliżona data śmierci ostatnich z ich natywnych użytkowników. Strona internetowa została przedstawiona na il. 2.

[21] J. Danecki, *Współczesny język arabski...*

[22] A.F. Majewicz, *Języki świata...*



Endangered languages

by Tapani Salminen <tasalmin@cc.helsinki.fi>

UNESCO Red Book on Endangered Languages

Europe

by Tapani Salminen

- Indexes to the report (include all modern European languages)
  - Background information
  - Index by present state of the language
  - Index by country
  - Index by classification
  - Alphabetical index
- Full version of the report (preliminary and open to criticism)

Northeast Asia

by Juha Janhunen and Tapani Salminen

- Indexes to the report
  - Background information
  - Index by present state of the language
  - Index by country
  - Index by classification
  - Alphabetical index
- Full version of the report

Finno-Ugrian (Uralic) languages

- A classification with updated demographic data
- Tundra Nenets homepage

Il. 2. Czerwona księga UNESCO dotycząca języków świata zagrożonych wymarciem (źródło: http://www.helsinki.fi/~tasalmin/endangered.html [odczyt: 4.02.2017])

W rozważanym kontekście warto również wspomnieć o projekcie będącym cenną inicjatywą związaną z ratowaniem wymierających języków. Na stronie internetowej o adresie http://www.endangeredlanguages.com [odczyt: 4.02.2017] zgromadzono repozytorium, składające się z różnorodnych zasobów związanych tematycznie z wymierającymi językami świata. Widok fragmentu strony głównej rozważanego projektu został przedstawiony na il. 3.

Na zakończenie warto jest jeszcze wspomnieć o niezwykle interesującej inicjatywie, jaką jest projekt Rosetta. Na il. 4 przedstawiono widok fragmentu strony internetowej poświęconej rozważanemu projektowi, dostępnej pod adresem https://rosettaproject.org. [odczyt: 4.02.2017].

Il. 3. Strona internetowa projektu związanego z ratowaniem wymierających języków świata (źródło: http://www.endangeredlanguages.com [odczyt: 4.02.2017])

Il. 4. Strona internetowa projektu Rosetta
(źródło: https://rosettaproject.org [odczyt: 4.02.2017])

Celem rozważanego projektu jest stworzenie swoistego banku wiedzy o ludzkich językach, który będzie w stanie bezpiecznie przetrwać nawet i wiele tysięcy lat, a w związku z tym stanowić będzie dla potomnych bezcenne źródło informacji dotyczącej historii ludzkich języków. Na potrzeby projektu Rosetta zebrano informacje o ponad 1500 ludzkich języków, które łącznie zajęłyby ponad 13 000 stron maszynopisu w formacie A4. Wszelkie zebrane informacje dotyczące badanych języków zostały następnie zapisane na wykonanym z niklu metalowym krążku o średnicy kilkunastu centymetrów. Wyryty za pomocą promieni lasera tekst widoczny jest dopiero pod mikroskopem po aż około sześciusetkrotnym powiększeniu. Dla każdego z uwzględnionych w projekcie Rosetta języków ludzkich na metalicznym dysku wyryto słownik zawierający ponad tysiąc haseł, co ma w założeniu chronić zagrożone wymarciem języki przed całkowitym zapomnieniem (przynajmniej ich podstawową warstwę leksykalną).

W opinii autorów niniejszego artykułu tego rodzaju inicjatywy należy oczywiście uznać za niezwykle cenne i jak najbardziej pożądane, jednak – ich zdaniem – nie są one w stanie zachować danego języka w świadomości kolejnych pokoleń. Cóż bowiem z tego, że będziemy mieli zapis poszczególnych jednostek pewnego podstawowego zbioru ich słownictwa, gdy bez znajomości gramatyki i konkretnych reguł składniowych nie zdołamy z poszczególnych zachowanych słów utworzyć jakichkolwiek dłuższych wypowiedzi. Również w sytuacji, gdy w danym języku odnalezione zostaną jakieś próbki tekstów, będą one dla badaczy całkowicie niezrozumiałe, a odtworzenie ich pierwotnego sensu będzie zapewne niezwykle trudnym zadaniem.

Z powyższego stwierdzenia wynika ważny wniosek. Mianowicie, chcąc ratować jakikolwiek zagrożony język przed popadnięciem w całkowite zapomnienie, należy oprócz odpowiednio obszernego zbioru słownictwa utrwalić również spory zasób jego gramatyki, a zwłaszcza niezbędne do tworzenia wypowiedzi reguły składniowe. W tym celu niezwykle pomocna może okazać się koncepcja tzw. języków kontrolowanych oraz powiązane z nimi odpowiednie narzędzia informatyczne.

4. Języki kontrolowane

Z definicji przez pojęcie języka kontrolowanego będziemy rozumieli pewien podzbiór języka naturalnego, na którym dany język kontrolowany jest pierwotnie wzorowany. W związku z powyższym każdy język kontrolowany jest w zasadzie zawsze językiem w pewnej mierze sztucznym (tego typu języki sztuczne określane są mianem języków naturalistycznych), przy czym powszechnie przyjmowana jest reguła, w myśl której każde zdanie utworzone w dowolnym języku kontrolowanym powinno być zarazem poprawnym pod względem syntaktycznym i semantycznym zdaniem danego języka naturalnego, który stanowił pierwotnie podstawę do utworzenia rozważanego języka kontrolowanego[23].

[23] M. Gajer, *Wielojęzyczne systemy automatycznego przekładu oparte na metodzie wzorców translacyjnych*, Kraków 2008.

Niestety, w kierunku przeciwnym tego rodzaju prawidłowość już zdecydowanie nie obowiązuje, gdyż bynajmniej nie każde zdanie, które możemy utworzyć w rozpatrywanym języku naturalnym, będzie mogło zostać uznane za poprawne, to znaczy takie, które możemy przeanalizować, wykorzystując w tym celu zbiór dostępnych reguł syntaktycznych danego języka kontrolowanego. Z tego względu każdy język kontrolowany jest bez wątpienia systemem zdecydowanie uboższym w porównaniu z jakimkolwiek językiem naturalnym, na podstawie którego został niegdyś utworzony.

Ważną cechą języków kontrolowanych jest ich otwartość zarówno semantyczna, jak i niekiedy również składniowa. Przez otwartość semantyczną będziemy rozumieli możliwość dodawania do zasobów słownictwa danego języka kontrolowanego nowych jednostek leksykalnych. Z kolei przez otwartość składniową języka kontrolowanego będziemy rozumieli możliwość dalszej rozbudowy jego reguł gramatycznych, aczkolwiek z założenia gramatyka języków kontrolowanych jest zwykle mocno okrojona w porównaniu z gramatyką charakterystyczną dla języków naturalnych[24].

Ponadto uważa się powszechnie, że prostota reguł składniowych języka kontrolowanego jest jego główną zaletą i w związku z tym w gramatyce tego języka powinno znajdować się zdecydowanie tylko to, co jest absolutnie niezbędne do efektywnego budowania w nim różnego typu wypowiedzi. Z tego powodu zasadniczą cechą języków kontrolowanych jest ich parataktyczność, czyli sposób budowania wypowiedzi przy użyciu prostych składniowo zdań. Natomiast stosowanie różnego typu zawiłych w swej budowie i rzadko używanych w praktyce językowej konstrukcji składniowych, aczkolwiek poprawnych w danym języku naturalnym, jest zdecydowanie odradzane w przypadku języków kontrolowanych.

W dobie powszechnego rozwoju i spektakularnej ekspansji technik informatycznych języki kontrolowane są wykorzystywane przede wszystkim w dwóch obszarach, do których można zaliczyć komunikację typu człowiek–komputer (ang. *man–machine interaction*) oraz przekład komputerowy (ang. *machine translation*). Zaletą języków kontrolowanych jest fakt, że w porównaniu z językami naturalnymi odznaczają się one wyższym stopniem precyzji formułowania wypowiedzi oraz zdecydowanie niższym stopniem wieloznaczności.

To właśnie wieloznaczność każdego języka naturalnego jest tym czynnikiem, który w głównej mierze utrudnia sprawną realizację przetwarzania oraz interpretacji języka ludzkiego przez komputer. Wieloznaczność języka naturalnego objawia się w zasadzie na każdym poziomie analizy językowej i dotyczy jego leksyki, morfologii, składni, semantyki, a także pragmatyki.

W komputerowym przetwarzaniu języka naturalnego kłopotliwa jest zwłaszcza wieloznaczność leksykalna, sprowadzająca się do tego, że poszczególne wyrazy mogą mieć nieraz wiele całkowicie od siebie odmiennych znaczeń. Człowiek na podstawie szerszego kontekstu wypowiedzi nie ma zwykle większych problemów z wyborem odpowiedniego znaczenia danego wyrazu, jednak zautomatyzowanie tego rodzaju wnioskowania, opracowanie odpowiedniego algorytmu i zakodowanie go

[24] Tamże.

w postaci programu komputerowego w ogólnym przypadku wciąż jeszcze pozostaje dużym wyzwaniem dla informatyków działających w obszarze badawczym dziedziny sztucznej inteligencji[25].

Nie mniej uciążliwa w komputerowej interpretacji języka naturalnego jest jego wieloznaczność morfologiczna i składniowa, polegająca na tym, że analizę gramatyczną danego fragmentu wypowiedzi można przeprowadzić na co najmniej dwa alternatywne sposoby, skutkujące zwykle także dwiema jej różnymi interpretacjami semantycznymi.

Jak już wspomniano, języki kontrolowane w zasadzie powinny być pozbawione jakichkolwiek wieloznaczności, jednak w praktyce ideał ten okazuje się niezwykle trudny do zrealizowania. W każdym razie poziom wieloznaczności dowolnego języka kontrolowanego jest bez wątpienia o wiele niższy niż poziom wieloznaczności dowolnego języka naturalnego. Z kolei wyeliminowanie w istotnym stopniu wieloznaczności z języka kontrolowanego pozwala na jego o wiele bardziej sprawne i bardziej efektywne przetwarzanie przez komputer, co w wypadku komunikacji człowieka z maszyną w języku naturalnym (co prawda kontrolowanym) bądź w wypadku tłumaczenia tekstów z jednego języka naturalnego (aczkolwiek kontrolowanego) na inny język naturalny skutkuje zasadniczo o wiele mniejszą liczbą błędów i ewentualnych pomyłek popełnianych przez programy komputerowe. To sprawia, że rozwijanie programów komputerowych umożliwiających komunikację człowieka z maszyną w języku naturalnym oraz systemów tłumaczenia komputerowego ma jak najbardziej sens i przedstawia także dobrze rokujące perspektywy sukcesu komercyjnego.

W tym miejscu warto także wspomnieć o całkowicie sztucznych językach, tzw. językach apriorycznych. Przykładem tego rodzaju języków jest język *lojban*, a także jego wcześniejszy pierwowzór, określany nazwą *loglan*. Tego rodzaju języki nie są w ogóle zaliczane do języków sztucznych typu naturalistycznego, gdyż pierwotnie nie są wzorowane na żadnym ze znanych języków naturalnych, lecz tworzone są całkowicie niezależnie w oparciu o mechanizmy logiki formalnej, zwane rachunkiem predykatów. Z tego powodu wypowiedzi tworzone w rozważanych sztucznych językach typu apriorycznego jedynie w bardzo niewielkim stopniu przypominają zdania tworzone w innych językach ludzkich. W wypadku tego rodzaju języków udało się w zasadzie w ostateczny sposób wyeliminować jakąkolwiek wieloznaczność z tworzonych w nich wypowiedzi, jednak ceną, którą przyszło za to zapłacić, jest ich niezwykle wysoki stopień formalnej komplikacji i związana z tym trudność ich praktycznego opanowania i biegłego posługiwania się nimi w mowie czy w piśmie[26].

Z wymienionych powodów wydaje się obecnie, że to właśnie języki kontrolowane, stanowiące pewien podzbiór, który nie jest bynajmniej podzbiorem zamkniętym – zwłaszcza pod kątem warstwy leksykalnej języka, są słusznym kierunkiem rozwoju technologii informatycznych ukierunkowanych na rozwijanie komunikacji człowieka z maszyną w języku naturalnym oraz realizację przekładu komputerowego.

[25] M. Gajer, *Wielojęzyczne systemy*...

[26] Tamże.

Wyczerpujący przegląd zagadnień związanych z budową i zastosowaniami różnych typów języków kontrolowanych czytelnik może znaleźć, między innymi, w pracy L. Szczepaniaka i Z. Królikowskiego – *Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań*[27].

5. Języki kontrolowane tworzone na podstawie języków zagrożonych wymarciem

Oczywiście, języki kontrolowane mogą być w zasadzie tworzone na podstawie dowolnego języka naturalnego. W związku z tym językiem, na którym jest wzorowany dany język kontrolowany, może być także dowolny język zagrożony wymarciem, aczkolwiek dobrze byłoby, aby język taki posiadał przynajmniej jakąś swoją powszechnie akceptowaną wersję pisaną o ustalonej formalnie ortografii, o skodyfikowanych regułach gramatycznych nawet nie wspominając. Wiele spośród języków zachodnioeuropejskich, należących do grupy romańskiej, germańskiej bądź celtyckiej, z powodzeniem spełnia przedstawione powyżej warunki. Warunki te spełniają także będące w użyciu na terenie Niemiec języki słowiańskie: górnołużycki i dolnołużycki[28].

Zgodnie z najlepszą wiedzą autorów, do chwili obecnej nie była jeszcze podejmowana próba utworzenia języka kontrolowanego na podstawie jakiegokolwiek języka o niewielkiej liczbie użytkowników, i do tego jeszcze zagrożonego w najbliższej perspektywie całkowitym wymarciem. Jako pierwowzór do tworzenia języków kontrolowanych służy przede wszystkim język angielski, którego dominacja na rozpatrywanym polu jest wręcz przytłaczająca, ewentualnie pewne znaczenie mają także inne duże języki europejskie (niemiecki, hiszpański czy francuski). Natomiast pomysł tworzenia języków kontrolowanych na podstawie języków zagrożonych wymarciem wydaje się w kontekście przedstawionych rozważań całkowicie nowy.

Jak już uprzednio wspomniano, podejmowanie różnorodnych inicjatyw na rzecz ratowania zagrożonych wymarciem języków świata, w postaci gromadzenia w różnego typu elektronicznych repozytoriach zbiorów słownictwa takich języków czy też dłuższych próbek tekstu, wydaje się oczywiście bardzo cenną inicjatywą, jednak w opinii autorów nagromadzenie tego rodzaju materiału o pasywnej w swej istocie naturze nie jest w stanie przyczynić się do uratowania jakiegokolwiek języka w taki sposób, aby kiedyś w przyszłości możliwe było jego wskrzeszenie, tak jak się to stało na przykład w przypadku współczesnego języka hebrajskiego (to zarazem

[27] L. Szczepaniak, Z. Królikowski, *Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań*, „Pro Dialog" 2000, nr 11, s. 47–67.

[28] T. Lehr-Spławiński, W. Kuraszkiewicz, F. Sławski, *Przegląd i charakterystyka języków słowiańskich*, Warszawa 1954.

prawdopodobnie jedyny tak spektakularny przykład ożywienia języka od dawna uznawanego za martwy, obecny jedynie w tekstach o charakterze liturgicznym)[29].

W wypadku języków kontrolowanych sprawa wygląda zgoła odmiennie, ponieważ oprócz zasobów leksyki danego języka kodowane są również podstawowe struktury składniowe, co łącznie umożliwi budowanie poprawnych gramatycznie zdań w danym języku, a zatem jego ponowne ożywienie staje się z zasady jak najbardziej możliwe, jeśli tylko znalazłaby się nawet stosunkowo niewielka grupa osób chętnych do podjęcia jego nauki[30].

Oprócz zdefiniowania samej postaci języka kontrolowanego, czyli określenia zasobów jego leksyki oraz dokonania wyboru niezbędnych reguł składniowych, należy także utworzyć odpowiednie narzędzia informatyczne, umożliwiające użytkownikowi efektywną pracę z danym językiem kontrolowanym. Autorzy niniejszego artykułu pracują obecnie nad budową dwóch rodzajów tego typu narzędzi informatycznych.

Pierwszym z nich są tzw. generatory struktur syntaktycznych, których zadaniem jest umożliwienie użytkownikowi budowania poprawnych wyrażeń lub całych zdań (prostych bądź złożonych) w danym języku kontrolowanym. Co istotne, w tym celu nie jest wymagana od użytkownika znajomość jakichkolwiek reguł składniowych danego języka kontrolowanego, ponieważ nad poprawną budową zdań języka kontrolowanego (zgodnie z odpowiednimi regułami gramatyki) czuwa odpowiedni program generatora struktur syntaktycznych.

Do drugiego typu narzędzi informatycznych zalicza się natomiast automatyczne translatory, których zadaniem jest dokonywanie komputerowego przekładu tworzonych przez użytkownika zdań w języku kontrolowanym na wybrane języki naturalne. Obecnie autorzy planują uwzględnienie przekładu komputerowego na następujące języki europejskie, do których zalicza się języki romańskie (takie jak: francuski, włoski, hiszpański, kataloński, portugalski i rumuński), języki germańskie (angielski, niemiecki, niderlandzki, szwedzki, duński i norweski w wersji bokmål), a także języki słowiańskie (polski, czeski, słowacki, słoweński, chorwacki i rosyjski) oraz sztuczny międzynarodowy język esperanto. W celu efektywnej realizacji przekładu komputerowego na tak wiele języków pracę komputerowego translatora postanowiono oprzeć na idei wykorzystania języka pośredniczącego przekładu. W związku z powyższym przekład z danego języka kontrolowanego na wybrane języki naturalne dokonywany będzie dwuetapowo. W pierwszej kolejności zdania utworzone przez użytkownika w języku kontrolowanym będą tłumaczone na język pośredniczący przekładu (tzw. *interlingua*), a następnie przekładane na wybrane docelowe języki naturalne.

[29] A.F. Majewicz, *Języki świata*...

[30] J.A. Matisoff, *Zagrożona różnorodność: języki*..., s. 66–73.

6. Opracowywany przez autorów system języka kontrolowanego

System języka kontrolowanego, nad którym obecnie pracują autorzy niniejszego artykułu, przewidziany jest przede wszystkim dla języków zagrożonych w najbliższej przyszłości całkowitym wymarciem, a należących do wielkiej rodziny języków indoeuropejskich. Według różnych szacunków przeprowadzonych przez lingwistów rozdzielenie się wielkiej rodziny języków indoeuropejskich na poszczególne grupy językowe nastąpiło około pięciu tysięcy lat temu. Obecnie w ramach indoeuropejskiej rodziny językowej wyróżnia się następujące grupy językowe: indyjską, irańską, bałtycką, słowiańską, celtycką, romańską i germańską. Do rodziny języków indoeuropejskich zalicza się także języki takie, jak grecki, ormiański i albański. W przeszłości w ramach rozważanej rodziny językowej istniały również takie grupy językowe, jak anatolijska i tocharska, ale do czasów nam współczesnych nie zachował się żaden z należących do nich języków.

Wszystkie języki zaliczane do rodziny indoeuropejskiej posiadają wiele cech wspólnych w zakresie leksyki i składni, co umożliwia objęcie ich wspólną koncepcją uniwersalnego języka kontrolowanego. Nie jest także wykluczone, że w przyszłości rozwijane przez autorów idee związane z systemem języków kontrolowanych będzie można przenieść także na grunt innych rodzin językowych. Autorzy myślą tu przede wszystkim o rodzinie języków afroazjatyckich, a także o rodzinach językowych: uralskiej, ałtajskiej, drawidyjskiej i południowokaukaskiej, które objęte są hipotezą istnienia w zamierzchłej przeszłości wspólnego dla nich prajęzyka i określane mianem tzw. języków nostratycznych.

Jak już uprzednio wspomniano, w przypadku języków kontrolowanych zbiór reguł składniowych ograniczony jest do niezbędnego minimum, natomiast warstwę leksykalną języka można w zasadzie dowolnie rozbudowywać, systematycznie uzupełniając ją o nowe jednostki wyrazowe.

W wypadku rozwijanego przez autorów systemu języka kontrolowanego możliwe jest tworzenie trzech rodzajów wypowiedzi, do których zalicza się frazy rzeczownikowe, zdania proste oraz zdania złożone. We frazach rzeczownikowych użytkownik ma do wyboru albo sam rzeczownik, albo rzeczownik z przymiotnikiem, przy czym przymiotnik może wystąpić zarówno w funkcji przydawkowej, jak i orzecznikowej. Po dokonaniu przez użytkownika wyboru z odpowiedniego menu rzeczownika i ewentualnie przymiotnika, można zdecydować także, czy tego rodzaju fraza rzeczownikowa ma być poprzedzona rodzajnikiem, ewentualnie zaimkiem wskazującym lub zaimkiem dzierżawczym. Ponadto należy dokonać wyboru, czy dana fraza rzeczownikowa ma występować w liczbie pojedynczej, liczbie podwójnej lub w liczbie mnogiej.

Użytkownik ma do wyboru poprzedzenie wybranej przez siebie frazy rzeczownikowej rodzajnikiem nieokreślonym, określonym bądź rodzajnikiem cząstkowym – oczywiście, jeżeli konkretny typ rodzajnika w danym języku naturalnym istnieje. Przykładowo, w języku polskim pojęcie rodzajnika nie jest w ogóle znane, podczas gdy w języku francuskim występują wszystkie typy wymienionych uprzednio rodzajników.

Analogicznie, w wypadku zaimków wskazujących mamy w języku kontrolowanym do czynienia z systemem trójstopniowym, ponieważ istnieją odrębne formy zaimków wskazujących dla obiektów położonych bliżej, obiektów położonych dalej oraz dla obiektów bardzo odległych. Tego typu trójstopniowe systemy zaimków wskazujących spotykane są w wielu językach europejskich, między innymi w językach hiszpańskim i portugalskim.

Gdyby użytkownik wybrał opcję utworzenia zdania prostego, możliwe będzie budowanie zdań zarówno typu SVO (ang. *Subject – Verb – Object*), jak i typu SV (ang. *Subject – Verb*). W pierwszym wypadku mamy do czynienia z czasownikami bądź z frazami czasownikowymi, które są przechodnie, czyli wymagają użycia dopełnienia bliższego. Natomiast w drugim przypadku zdania są tworzone na podstawie fraz czasownikowych nieprzechodnich, czyli takich, w których dopełnienie bliższe zdania w ogóle nie występuje.

W roli podmiotu i dopełnienia bliższego budowanych przez użytkownika zdań mogą wystąpić zarówno frazy rzeczownikowe utworzone z rzeczowników i przymiotników, jak i z samych rzeczowników. Takie frazy mogą być oczywiście również poprzedzone rodzajnikami, zaimkami wskazującymi bądź zaimkami dzierżawczymi. Ponadto w roli podmiotu zdania występować mogą zaimki osobowe. Ponieważ w wybranych językach indoeuropejskich systemy zaimków osobowych mogą przyjmować różnorodne i całkowicie od siebie odmienne formy, dlatego w celu dokonywania wyboru odpowiedniej formy zaimka osobowego, a także zaimka dzierżawczego, autorzy zdecydowali się na opracowanie odpowiedniego interfejsu graficznego, umożliwiającego w prosty sposób realizację wyboru odpowiedniej formy danego zaimka. Na il. 5 przedstawiono projekt graficzny interfejsu użytkownika, służącego do dokonywania wyborów odpowiednich form zaimków osobowych oraz zaimków dzierżawczych.

Il. 5. Widok graficznego interfejsu użytkownika służącego do wyboru odpowiednich form zaimków osobowych i dzierżawczych

(źródło: opracowanie własne)

W niektórych językach należących do rodziny języków indoeuropejskich mamy do czynienia z maksymalnie uproszczonymi systemami zaimków osobowych. Przykładem tego rodzaju języków jest między innymi język perski, w którym istnieje jedynie sześć form zaimka osobowego – odpowiednio dla pierwszej, drugiej i trzeciej osoby liczby pojedynczej i mnogiej[31,32]. Nieco bardziej rozbudowane systemy zaimków osobowych występują w językach germańskich, gdzie zaimki dla trzeciej osoby liczby pojedynczej, i niekiedy również liczby mnogiej, odmieniają się jeszcze przez rodzaje gramatyczne. Analogicznie sytuacja przedstawia się także w wypadku języków należących do grupy romańskiej. Natomiast najbardziej złożone systemy zaimków osobowych występują w językach słowiańskich. Przykładowo, w językach takich, jak dolnołużycki, górnołużycki i słoweński w wypadku niektórych zaimków osobowych występują również odmienne ich formy dla liczby podwójnej[33].

Trzecim typem wypowiedzi, który może być utworzony w języku kontrolowanym przez użytkownika systemu, są zdania złożone. Obecnie autorzy postanowili ograniczyć zbiór dostępnych reguł umożliwiających tworzenie takich zdań jedynie do zdań podrzędnie złożonych zbudowanych z dwóch zdań prostych, połączonych za pomocą odpowiedniego wyrazu bądź frazy. Do szczególnego typu tego rodzaju zdań zaliczają się zdania warunkowe, w których system musi przestrzegać odpowiednich reguł związanych z doborem właściwych czasów gramatycznych dla obu zdań składowych.

Wiedza dotycząca leksyki i składni danego języka zapisana jest w odpowiednich lingwistycznych bazach danych. W opracowywanym przez autorów systemie języka kontrolowanego istnieją trzy rodzaje tego typu lingwistycznych baz danych. Pierwsza z nich służy do przechowywania rzeczowników danego języka, z kolei druga przeznaczona jest do przechowywania przymiotników, natomiast w trzeciej zgromadzone są frazy czasownikowe, które pełnią funkcję orzeczenia w budowanych przez użytkownika zdaniach należących do języka kontrolowanego. Jak już uprzednio wspomniano, każde z utworzonych przez użytkownika zdań języka kontrolowanego musi być zarazem poprawnym składniowo zdaniem języka naturalnego, na którym rozważany język kontrolowany jest wzorowany. Natomiast, oczywiście, nie każde zdanie danego języka naturalnego będzie można w języku kontrolowanym utworzyć, co mimo wszystko nie powinno stanowić jakiegoś istotnego mankamentu opracowywanego przez autorów systemu, bowiem dowolny z języków ludzkich jest zapewne tworem w takim stopniu elastycznym, że wszelkie potrzebne treści zawsze można wyrazić w nim w jakiś alternatywny sposób, korzystając z dostępnych w danym momencie zasobów leksyki i struktur składniowych języka kontrolowanego.

Budowa lingwistycznych baz danych zostanie omówiona na przykładzie bazy danych przechowującej rzeczowniki. Rozważana baza danych została odpowiednio wcześniej utworzona przez autorów dla języka norweskiego. Dane lingwistyczne dotyczące rzeczowników języka norweskiego zapisano w formie pliku, który

[31] M. Chwilczyńska-Wawrzyniak, *Język perski*, Warszawa 1998.

[32] K.P. Rahnama, *Język perski*, Warszawa 1999.

[33] T. Lehr-Spławiński, W. Kuraszkiewicz, F. Sławski, *Przegląd i charakterystyka...*

przechowuje kolejne rekordy, dotyczące poszczególnych rzeczowników. Przykładowy rekord z rozważanej lingwistycznej bazy danych ma następującą postać:

> *bilmekaniker&bilmekanikeren&bilmekanikerer&bilmekanikerene@0^0$1%!_machanik_samochodowy#

Każdy nowy rekord z rozważanej bazy danych rozpoczyna się zawsze znakiem „*". Po tym znaku występuje forma podstawowa danego norweskiego rzeczownika, czyli w naszym wypadku jest to rzeczownik „bilmekaniker". Kolejne znaki „&" oddzielają formy pochodne danego rzeczownika. W podanym przykładzie są to „bilmekanikeren" – forma określona danego rzeczownika w liczbie pojedynczej, „bilmekanikerer" – forma nieokreślona liczby mnogiej, oraz „bilmekanikerene" – forma określona liczby mnogiej.

Z kolei po znaku „@" zamieszczona jest informacja, czy dany rzeczownik może zostać poprzedzony rodzajnikiem nieokreślonym. Jeżeli tak, wówczas bezpośrednio po znaku „@" pojawia się „0". W wypadku przeciwnym, czyli wtedy, gdy danego rzeczownika w ogóle nie można poprzedzić rodzajnikiem nieokreślonym, pojawi się tam „1". Ponadto wszelkie inne wartości rozważanego pola w opracowywanym przez autorów systemie kontrolowanego języka norweskiego są zabronione.

Analogicznie, po znaku „^" podawana jest informacja, czy dany rzeczownik może występować zarówno w liczbie pojedynczej, jak i mnogiej – w takim wypadku bezpośrednio po rozważanym znaku występuje wartość numeryczna równa „0". Z kolei jeżeli dany rzeczownik należy do kategorii *singularia tantum*, czyli występuje jedynie w liczbie pojedynczej, wówczas po znaku „^" pojawi się „1". Podobnie, jeżeli dany rzeczownik należy do kategorii *pluralia tantum*, czyli występuje jedynie w liczbie mnogiej, wówczas po znaku „^" pojawi się „2". Pozostałe wartości dla wymienionego pola są zabronione.

Po znaku „$" podawana jest natomiast informacja o rodzaju gramatycznym danego rzeczownika, przy czym wartość „1" występuje dla rodzaju męskiego, wartość „2" dla rodzaju żeńskiego i wartość „3" dla rodzaju nijakiego, a pozostałe wartości są w rozważanym polu zabronione. Z kolei po znaku „%" może zostać zamieszczony dowolny, opcjonalny komentarz, który jednak w rozpatrywanym przypadku w ogóle nie występuje. Analogicznie, po znaku „!" podawany jest odpowiednik semantyczny danego norweskiego rzeczownika w języku pośredniczącym przekładu. W omawianym przypadku jest to fraza „_mechanik_samochodowy", przy czym język pośredniczący przekładu stanowi jedynie wewnętrzny mechanizm systemu, umożliwiający automatyczny przekład na inne języki naturalne, i nie jest w żaden sposób bezpośrednio dostępny dla użytkownika systemu. Ponadto każdy nowy rekord danych zakończony jest zawsze znakiem „#".

7. Podsumowanie

Ratowanie dziedzictwa kulturowego reprezentowanego przez liczne języki świata zagrożone wymarciem wydaje się obecnie palącym problemem. Podejmowania tego rodzaju działań nie można bynajmniej odkładać na jakąś bliżej nieokreśloną przyszłość, ponieważ już za kilkadziesiąt lat może się okazać, że nie ma w zasadzie czego ratować, a wiedza o licznych językach o niewielkiej liczbie użytkowników może całkowicie odejść w niepamięć.

Zaproponowana przez autorów inicjatywa polegająca na wykorzystaniu w procesie ratowania zagrożonych wymarciem języków znanej z technik informatycznych (związanych z obszarem sztucznej inteligencji, lingwistyki komputerowej, przetwarzania języka naturalnego oraz inżynierii lingwistycznej) koncepcji języków kontrolowanych wydaje się pomysłem ze wszech miar nowatorskim. W opinii autorów podejmowane dotychczas inicjatywy, takie jak na przykład projekt Rosetta, nie są w stanie uratować zagrożonych języków przed całkowitym zapomnieniem. Cóż bowiem z tego, że na zabezpieczonym przed korozją metalicznym dysku zapiszemy mikrodrukiem ponad tysiąc najczęściej używanych słów jakiegoś małego języka zagrożonego obecnie wymarciem, gdy po iluś tam latach nie będziemy mieli już najmniejszego pojęcia o regułach składniowych rządzących gramatyką tego języka. W związku z powyższym od form podstawowych zanotowanego na dysku słownictwa nie będziemy w stanie utworzyć żadnych form pochodnych i, co ważniejsze, nie będziemy w stanie łączyć utrwalonych przed zapomnieniem wyrazów w jakiekolwiek poprawne składniowo dłuższe frazy, o budowie całych, nawet najprostszych, zdań nawet nie wspominając.

W przeciwieństwie do tego, co napisano powyżej, opracowywane przez autorów języki kontrolowane, stanowiące swego rodzaju dokumentację nie tylko podstawowego zasobu leksyki danego języka naturalnego, lecz także zbioru jego podstawowych reguł składniowych, wraz z odpowiednimi narzędziami informatycznymi w postaci automatycznych generatorów struktur syntaktycznych, są w stanie niejako zachować wymierający język przy życiu. Opracowywany przez autorów system pozwalał będzie na tworzenie w zagrożonym wymarciem języku poprawnych składniowo zdań, nawet przez osoby niemające żadnego pojęcia o jego strukturach składniowych. Ponadto utworzone w zagrożonym wymarciem języku zdania będą automatycznie tłumaczone na kilkanaście języków europejskich, co umożliwi, po pierwsze – ich zrozumienie, a po wtóre – będzie swego rodzaju pomocą dydaktyczną wspomagającą wydatnie proces nauki takiego języka.

Kolejnym krokiem będzie połączenie opracowywanego przez autorów systemu generacji struktur syntaktycznych z systemem syntezy mowy, co umożliwi także odsłuchanie tworzonych w zagrożonym wymarciem języku wypowiedzi. Takie narzędzie informatyczne będzie wręcz nieocenione dla osób pragnących uczyć się wybranych języków indoeuropejskich, które obecnie uważane są za języki zagrożone wymarciem.

W najbliższej przyszłości autorzy planują budowę prototypowego systemu generatora struktur syntaktycznych oraz komputerowych translatorów na wybrane

główne języki europejskie dla północnogermańskiego (skandynawskiego) języka fererskiego.

Językiem fererskim posługuje się obecnie około 48 tysięcy mieszkańców Wysp Owczych, gdzie obok języka duńskiego jest on językiem urzędowym. Jak wszystkie języki o niewielkiej liczbie użytkowników, język ten w perspektywie najbliższych stuleci jest z całą pewnością poważnie zagrożony wymarciem, dlatego wszelkie inicjatywy mające na celu jego zachowanie i utrwalenie dla przyszłych pokoleń są z pewnością bardzo pożądane.

Język fererski należy do archaicznej warstwy języków skandynawskich i znacznie bliżej mu do języka staronorweskiego niż do współczesnych języków będących współcześnie w użyciu na terenie Norwegii, Danii czy Szwecji. Spośród będących obecnie w użyciu języków północnogermańskich językowi fererskiemu najbliżej do języka islandzkiego, wraz z którym tworzą swego rodzaju „żywą skamielinę" dawnych języków skandynawskich. Z tych powodów język fererski jest bardzo interesujący dla lingwistów, ponieważ wspomaga on proces rekonstrukcji starszych wersji innych współczesnych języków używanych na terenie Skandynawii.

Język fererski posiada o wiele bardziej skomplikowaną gramatykę w porównaniu ze współczesnymi językami: norweskim, szwedzkim i duńskim, gdyż zachowało się w nim wiele archaicznych kategorii gramatycznych i związanych z nimi konstrukcji składniowych. Te osobliwości składni języka fererskiego muszą być oczywiście uwzględnione w opracowywanym przez autorów generatorze struktur syntaktycznych języka fererskiego.

Bibliografia

Bareja-Starzyńska A., Mejor M., *Klasyczny język tybetański*, Warszawa 2002.

Chwilczyńska-Wawrzyniak M., *Język perski*, Warszawa 1998.

Corballis M.C., Dessalles J.L., Dunbar R., *Aux origines du langage*, „La Recherche" 2001, nr 341, s. 27–39.

Dalewska-Greń H., *Języki słowiańskie*, Warszawa 2002.

Danecki J., *Klasyczny język arabski*, Warszawa 1998.

Danecki J., *Współczesny język arabski i jego dialekty*, Warszawa 2000.

Gajer M., *Wielojęzyczne systemy automatycznego przekładu oparte na metodzie wzorców translacyjnych*, Kraków 2008.

Godziński S., *Współczesny język mongolski*, Warszawa 1998.

Kałużyński S., *Klasyczny język mongolski*, Warszawa 1998.

Kondratow A., *Zaginione cywilizacje*, Warszawa 1988.

Künstler J.M., *Języki chińskie*, Warszawa 2000.

Lehr-Spławiński T., Kuraszkiewicz W., Sławski F., *Przegląd i charakterystyka języków słowiańskich*, Warszawa 1954.

Majewicz A.F., *Języki świata i ich klasyfikowanie*, Warszawa 1989.

Matisoff J.A., *Zagrożona różnorodność: języki i formy życia*, „Świat Nauki", październik 2002, s. 66–73.

Popko M., *Ludy i języki starożytnej Anatolii*, Warszawa 1999.

Rahnama K.P., *Język perski*, Warszawa 1999.

Schulze C., Staffer D., Wichmann S., *Birth, Survivaland Death of languages by Monte Carlo Simulation*, „Communications in Computational Physics" 2008, vol. 3, nr 2.

Szczepaniak L., Królikowski Z., *Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań*, „Pro Dialog" 2000, nr 11, s. 47–67.