Mirosław Gajer[1]
Zbigniew Handzel[2]

Zastosowanie generatorów struktur syntaktycznych na potrzeby rekonstrukcji i rewitalizacji wymierających języków

Application of syntactic structures generators for the aim of reconstruction and revitalization of endangered languages

Streszczenie: Obecna sytuacja językowa świata prezentuje się wręcz katastrofalnie, gdyż wiele małych języków jest poważnie zagrożonych całkowitym wymarciem i to niestety już w najbliższych latach. Taki stan rzeczy nie jest bynajmniej korzystny z punktu widzenia wysoce pożądanego zachowania różnorodności dziedzictwa kulturowego ludzkości. Ponadto pamiętać trzeba, że każdy z będących w użyciu języków ludzkich to także nieodzowny materiał badawczy dla lingwistów, w oparciu o który rozważani naukowcy mogą testować swoje hipotezy dotyczące nie poznanej do końca jeszcze natury języka ludzkiego, jego pochodzenia, tzw. uniwersaliów językowych czy też ewolucji całych grup i rodzin językowych. W związku z powyższym podejmowane są w świecie nauki liczne inicjatywy mające na celu uratowanie przed popadnięciem w niepamięć licznych języków ludzkich o małej liczbie użytkowników. W artykule autorzy wysunęli oryginalną propozycję budowy odpowiednich narzędzi informatycznych w postaci generatorów struktur syntaktycznych, które mogłyby znacznie usprawnić proces rekonstrukcji i rewitalizacji zagrożonych wymarciem języków, a nawet i języków od wielu lat całkowicie już martwych. Obecnie przedmiotem zainteresowań autorów są peryferyjne skandynawskie języki północnogermańskie, takie jak potencjalnie zagrożony wymarciem język farerski (Wyspy Owcze) i niestety od ponad stu lat martwy już język norn (Szetlandy i Orkady).

[1] Dr inż., AGH Akademia Górniczo-Hutnicza, Wydział EAIiIB, Katedra Informatyki Stosowanej.
[2] Dr inż., Wyższa Szkoła Ekonomii i Informatyki.

Słowa kluczowe: lingwistyka komputerowa, struktury syntaktyczne, języki zagrożone wymarciem, dziedzictwo kulturowe.

Summary: Nowadays the situation of the world of languages is such that many small languages are seriously endangered in the coming years. This situation is not good from the point of view of the necessity of preserving the diversity of cultural heritage. Moreover, every human language constitute the precious research material for the linguists, basing on which they can test their scientific hypothesis concerning the nature of the human language, its origin, and the evolution of groups and families of languages. The world of sciences developed many initiatives and projects the aim of which is to preserve the small languages from total perishing. In the paper we propose to build proper computer tools in the form of the generators of syntactic structures that could support the process of reconstruction and revitalization of endangered and even dead languages. The point of authors' interest are peripheral Scandinavian languages, such as potentially endangered language of the Faroe Islands and the norn language, which is unfortunately dead for over one hundred years.

Keywords: computational linguistics, syntactic structures, endangered languages, cultural heritage.

JEL classification codes: C8

Wprowadzenie

Lingwiści szacują, że obecnie na kuli ziemskiej pozostaje wciąż w użyciu ponad 7 tysięcy różnych języków ludzkich. Podanie dokładnej ich liczby nie jest możliwe, przynajmniej z trzech zasadniczych powodów.

Po pierwsze, znakomita większość materiału językowego występującego na kuli ziemskiej nie została jeszcze należycie (a niekiedy nawet i w ogóle) przebadana przez lingwistów. Ponadto znaczny odsetek języków ludzkich pozostaje w użyciu na trudno dostępnych obszarach, głównie w Afryce, Ameryce Łacińskiej i Azji Południowo-Wschodniej. Jednak, co najważniejsze, w zdecydowanej większości są to języki małe, których liczba użytkowników wynosi niekiedy jedynie kilkaset bądź nawet kilkadziesiąt czy zaledwie wręcz kilka osób w podeszłym wieku[3].

[3)] Majewicz A. F.: *Języki świata i ich klasyfikowanie*, Państwowe Wydawnictwo Naukowe, Warszawa, 1989.

Po drugie, próbując zliczyć języki, którymi porozumiewają się ludzie na ziemi, napotykamy nieuchronnie na nierozstrzygnięty jak dotąd w lingwistyce fundamentalny problem, w jaki sposób można jednoznacznie stwierdzić, czy w danym przypadku mamy do czynienia z całkowicie odrębnym językiem czy też jest to jedynie pewna odmiana dialektalna jakiegoś innego języka[4]. Co więcej, w świetle prowadzonych w językoznawstwie badań wydaje się, że ustalenie tutaj jakiś precyzyjnych kryteriów podziału nie jest możliwe z zasady, bowiem wiele języków (jak na przykład język arabski lub chociażby język chiński) wydaje się być wręcz wewnętrznie splątanych w niekończącą się sieć wszelkiego rodzaju odmian regionalnych i dialektów[5].

Wreszcie, po trzecie, spora liczba języków będących obecnie w użyciu na ziemi jest poważnie zagrożona wymarciem, podobnie jak na wyginiecie zdaje się być skazana znaczna część występujących na kuli ziemskiej gatunków zwierząt i roślin. W swej istocie oba z wymienionych procesów są do siebie niezwykle podobne, a u ich podłoża leżą przede wszystkim czynniki związane z dynamicznym rozwojem cywilizacyjnym ludzkości, postępem technicznym wiążącym się z praktycznie niczym nie ograniczoną eksploatacją zasobów środowiskowych. W ten sposób niszczone są na naszych oczach swoiste nisze ekologiczne, w których dotychczas mogły ukrywać się rzadkie gatunki fauny i flory, a także i specyficzne języki będące w użyciu przez niewielkie społeczności ludzkie[6].

Wiele spośród zagrożonych wymarciem języków jest jeszcze pamiętanych przez zaledwie pojedyncze osoby w podeszłym wieku i, co gorsze, języki te nie są już przekazywane kolejnym pokoleniom. W takim przypadku zniknięcie tych języków z powierzchni ziemi jest tylko kwestią czasu. Ponadto warto uświadomić sobie fakt, że praktycznie w każdym tygodniu na świecie umiera przynajmniej jeden język, przy czym za moment śmierci języka przyjmuje się moment śmierci ostatniego z jego użytkowników, który znał ten język biegle w sposób czynny.

Zajmujący się problemem wymierających języków lingwiści biją wręcz na alarm, ponieważ w świetle prowadzonych badań wydaje się, że omawiane niekorzystne procesy przybrały w ostatnich latach niezwykle na sile, co może skutkować tym, że już za niecałe pół wieku liczba będących w użyciu na ziemi języków skurczy się wręcz o połowę.

---

[4] Danecki J.: *Współczesny język arabski i jego dialekty*, Wydawnictwo Akademickie DIALOG, Warszawa, 2000.

[5] Künstler J. M.: *Języki chińskie*, Wydawnictwo Akademickie DIALOG, Warszawa, 2000.

[6] Matisoff J. A.: *Zagrożona różnorodność: języki i formy życia*, Świat Nauki, październik 2002, s. 66-73.

Oczywiście, ktoś mógłby zapewne powiedzieć, że proces wymierania języków o niewielkiej liczbie użytkowników nie stanowi w ogóle jakiegokolwiek problemu, a wręcz przeciwnie, może być postrzegany nawet jako zjawisko w pewny stopniu korzystne. Przecież, gdyby wszyscy ludzie na ziemi mówili tylko jednym językiem, nie istniałyby pomiędzy nimi jakiekolwiek bariery komunikacyjne i zapewne zaoszczędzono by mnóstwo pieniędzy, które obecnie są wydawane na tłumaczenia różnego rodzaju dokumentów, tekstów naukowych, technicznych, medycznych czy też aktów prawnych.

Niestety, istnieje również i druga strona medalu, bowiem każdy język ludzki to swego rodzaju unikalny środek wyrażania myśli, a zarazem także i oryginalny sposób postrzegania świata. Bogactwo i różnorodność języków świata to również wspaniałe dziedzictwo ogólnoludzkiej kultury, a śmierć każdego języka, nawet o niewielkiej liczbie użytkowników, tę ze wszech miar pożądaną różnorodność świata języków w istotnym stopniu zubaża[7].

Obserwowane obecnie przyspieszenie procesu wymierania języków to również zła wiadomość dla językoznawców. Wziąwszy pod uwagę, że znaczna liczba będących w użyciu na ziemi języków nie została jeszcze należycie przebadana, a w wielu wypadkach nawet w ogóle nie została dostatecznie rozpoznana, lingwiści obawiają się, że w perspektywie kilkudziesięciu lat znaczna część ich materiału badawczego może po prostu zniknąć bezpowrotnie z powierzchni ziemi[8].

Taki stan rzeczy napawa niepokojem, trzeba bowiem być świadomym, że językoznawstwo jest wciąż stosunkowo jeszcze młodą dyscypliną nauki, w przypadku której toczą się liczne spory natury wręcz fundamentalnej na temat tego, czym w ogóle jest język ludzki oraz gdzie leżą granice możliwości i ograniczeń języka. Przykładowo, jednym z podstawowych pojęć w lingwistyce są tzw. uniwersalia językowe, czyli te elementy języka, które są absolutnie niezbędne i w związku z tym muszą bezwzględnie pojawić się w każdym języku ludzkim. Do tego rodzaju uniwersaliów językowych zalicza się na przykład zaimki osobowe, ponieważ jak dotychczas nie jest znany żaden język ludzki, w którym nie byłoby jakiegoś systemu zaimków osobowych dla pierwszej, drugiej i trzeciej osoby gramatycznej, aczkolwiek w przypadku różnych języków systemy zaimków osobowych mogą znacznie od siebie się różnić. Ponieważ jednak znaczna część języków świata nie została jeszcze w ogóle przebadana,

[7] Kondratow A., *Zaginione cywilizacje*, Państwowy Instytut Wydawniczy, Warszawa, 1988.

[8] Corballis M. C., Dessalles J. L., Dunbar R.: *Aux origines du langage*, La Recherche, no. 341, 2001, s. 27-39.

nie możemy na chwilę obecną mieć żadnej pewności, czy gdzieś na świecie nie istnieje jakiś język pozbawiony całkowicie takiej kategorii gramatycznej jak właśnie zaimki osobowe. Jeśli w niedalekiej przyszłości większość języków ludzkich o niewielkiej liczbie użytkowników zniknie z powierzchni ziemi, możemy się już tego nigdy nie dowiedzieć.

Osobnym zagadnieniem są kwestie związane z ewolucją języka ludzkiego. Wciąż nie wiemy dokładnie, w jaki sposób, kiedy i dlaczego nasi odlegli przodkowie zaczęli używać języka. Nie wiemy także, czy języki ludzkie, które występują obecnie, powstały w wyniku czysto biologicznej ewolucji, czy też są jedynie wytworem ogólnoludzkiej kultury i zostały przez ludzi po prostu wymyślone w celu skuteczniejszego rozwiązywania problemów związanych z wzajemną komunikacją w archaicznych społecznościach ludzkich. Wielu badaczy uważa, że języki ludzkie, które występują obecnie, musiały być poprzedzone w zamierzchłej przeszłości czymś znacznie prostszym, czyli swego rodzaju protojęzykiem, zawierającym jednie warstwę leksykalną, natomiast pozbawionym jakichkolwiek struktur składniowych. Lingwiści toczą obecnie spory na temat tego, jak dokładnie słownictwo takiego protojęzyka mogło wyglądać, a ściślej rzecz ujmując, czy poszczególne wyrazy miały charakter swego rodzaju podstawowych atomów, czy też, być może, kodowały nawet nieco dłuższe i bardziej rozbudowane frazy. Wydaje się, że nieco światła na tego rodzaju spory może rzucić właśnie badanie małych języków rozsianych w trudnodostępnych zakątkach globu. Cóż jednak z tego, gdy w stosunkowo niedługim czasie lingwiści po prostu nie będą prawdopodobnie mieli już czego badać.

Wymieranie małych języków to także zła wiadomość dla językoznawców zajmujących się lingwistyką historyczną. Przedmiotem zainteresowań lingwistyki diachronicznej jest ewolucja całych rodzin bądź grup językowych, ustalanie wzajemnego powinowactwa genetycznego pomiędzy niekiedy dość odległymi językami, a także rekonstrukcja wcześniejszych stadiów wybranych języków, jak choćby języka praindoeuropejskiego czy języka prasłowiańskiego. W tym wypadku nieocenione mogą okazać się także języki o niewielkiej liczbie użytkowników, w przypadku których zachować się mogły dawne cechy leksykalne lub składniowe, które dawno już wyszły z użycia w przypadku innych języków należących do danej grupy bądź rodziny językowej. Dobrym przykładem w tym względzie są języki, takie jak islandzki i farerski, które silnie odbiegają od współczesnych północnogermańskich języków skandynawskich (szwedzki, norweski, duński) natomiast w swej formie są bardzo zbliżone do języka staronordyckiego, tym samym umożliwiając jego zrozumienie, lepsze poznanie i dalszą rekonstrukcję.

W związku z powyższym w wielu krajach podejmowane są działania mające na celu zachowanie istniejącej obecnie na świecie różnorodności językowej i uratowanie licznych języków o niewielkiej liczbie użytkowników przed popadnięciem w całkowite zapomnienie. Niestety, jak dotychczas działania takie wydają się być raczej mało skuteczne.

Ratowanie wymierających języków

Prowadzone obecnie na świecie działania, zmierzające do ratowania przed popadnięciem w niepamięć wymierających w szybkim tempie języków, polegają głównie na rejestracji ich warstwy leksykalnej oraz sporządzaniu opisów ich warstwy syntaktycznej. Tego rodzaju działania autorzy niniejszego artykułu określają mianem działań biernych. W głębokim przekonaniu autorów działania takie w żadnym wypadku nie mogą okazać się skuteczne, w tym sensie, że nigdy nie mogą stanowić solidnej podstawy, która mogłaby być wykorzystana do przeprowadzenia rekonstrukcji, a następnie rewitalizacji wymarłych już bądź zagrożonych wymarciem języków.

Zdaniem autorów ratowanie wymierających języków ze wszech miar wymaga podjęcia działań czynnych, polegających na stworzeniu odpowiedniego oprogramowania komputerowego, za pomocą którego możliwe byłoby przeprowadzenie rekonstrukcji rozważanych języków, a także wspomaganie procesów związanych z ich rewitalizacją, czyli ponownym przywróceniem do powszechnego użycia, przynajmniej w niewielkim kręgu ochotników, czyli swego rodzaju pasjonatów zainteresowanych zachowaniem dziedzictwa kulturowego ich przodków.

Tego rodzaju oprogramowanie komputerowe to przede wszystkim tzw. generatory struktur syntaktycznych, czyli programy umożliwiające ich użytkownikowi tworzenie poprawnych składniowo wypowiedzi w danym języku, przy czym w poszczególnych przypadkach znajomość przez użytkownika danego języka może być mocno ograniczona bądź nawet i żadna. Wymienione generatory struktur syntaktycznych współpracują z opracowanymi uprzednio lingwistycznymi bazami danych (rejestrującymi warstwę leksykalną języka) oraz lingwistycznymi bazami wiedzy (rejestrującymi występujące w danym języku zależności składniowe).

Omawiane przez autorów generatory struktur syntaktycznych mogą także być postrzegane jako swego rodzaju odmiana programów komputerowych wspomagających przekład, czyli programów typu MAHT (ang. Machine-Aided Human Translation). Za pomocą programów typu MAHT użytkownik

może tworzyć poprawne gramatycznie teksty w językach, o których ma w zasadzie stosunkowo niewielkie pojęcie bądź nawet i takich, których wcale nie zna. Zdaniem autorów tego rodzaju narzędzia informatyczne, jak generatory struktur syntaktycznych i programy typu MAHT nadają się w idealny sposób do wspomagania procesu rekonstrukcji i rewitalizacji wymierających czy też nawet i dawno już wymarłych języków.

Realizacja założonego przez autorów celu wymaga przede wszystkim opracowania ergonomicznego i intuicyjnego w obsłudze graficznego interfejsu użytkownika, a ponadto dla wybranych języków, które mają zostać poddane procesowi rekonstrukcji i rewitalizacji, należy opracować lingwistyczne bazy danych, rejestrujące ich warstwę leksykalną, oraz lingwistyczne bazy wiedzy, oddające ich warstwę składniowa. Reasumując, rozwijany przez autorów projekt jest przedsięwzięciem ze wszech miar interdyscyplinarnym, wymagającym ścisłej współpracy osób reprezentujących zarówno dyscyplinę naukową informatyki technicznej, jak i lingwistyki stosowanej, a także i filologie poszczególnych języków.

W porównaniu do podejmowanych dotychczas w rozważanym obszarze działań proponowane przez siebie podejście autorzy uważają za ze wszech miar nowatorskie, a nawet i wręcz rewolucyjne, gdyż w ich opinii podejmowane działania mają realną szansę ustanowić prawdziwy przełom w dotychczasowym podejściu do mozolnego procesu ratowania wymierających języków.

Postawiona powyżej teza może zostać udowodniona, jeśli porówna się proponowane przez autorów podejście z prawdopodobnie najbardziej dotychczas zakrojonym na szeroką skalę przedsięwzięciem, określanym mianem The Rosetta Project. Nazwa rozważanego przedsięwzięcia pochodzi od tzw. kamienia z Rosetty, który został odnaleziony w 1798 roku podczas wyprawy Napoleona Bonapartego do Egiptu. Kamień z Rosetty zawiera pewien przekaz zapisany egipskimi hieroglifami, egipskim pismem demotycznym oraz sporządzony jest w języku starogreckim. Znaczenie kamienia z Rosetty było przełomowe dla rozwoju dziedziny egiptologii, ponieważ stanowiło podstawę do odczytania egipskich hieroglifów w 1822 roku przez francuskiego uczonego, Franciszka Champoliona.

Analogicznie, w ramach przedsięwzięcia The Rosetta Project powstać ma wykonana z platyny okrągła płytka, na której wyryte mają zostać napisy zawierające podstawowe słownictwo zapisane w około tysiącu języków będących obecnie w użyciu na ziemi. Szacuje się, że już w stosunkowo niedługim czasie znaczna część z owego tysiąca języków wyjdzie całkowicie z użycia. W związku z tym przewiduje się, że The Rosetta Project pozwoli

na zachowanie wiedzy o wymarłych językach i przekazanie jej kolejnym pokoleniom.

Na rysunku 1 pokazano zrzut ekranu, przedstawiający fragment platynowego dysku powstałego w ramach programu The Rosetta Project, z tym, że wykonany jest on także w wersji elektronicznej – umożliwiającej interaktywne przeglądanie jego zawartości z wykorzystaniem w tym celu przeglądarki internetowej. W ten sposób dowolnie wybrany fragmenty dysku można odpowiednio powiększyć, aby odczytać zawarte na nim napisy wykonane w wielu różnych językach świata.

Rysunek 1. Fragment dysku z Rosetty w wersji interaktywnej

Źródło: https://rosettaproject.org.

Jednak autorzy niniejszego artykułu wyrażają poważne wątpliwości odnośnie skuteczności tego rodzaju działań. Bowiem leksyka to tylko zaledwie część wiedzy o danym języku (aczkolwiek bardzo ważna). Poza samym słownictwem trzeba jeszcze znać przynajmniej podstawowe reguły składniowe, aby w rozpatrywanym języku można było tworzyć jakiekolwiek sensowne wypowiedzi. Program The Rosetta Project w żadnym wypadku tego nie umożliwia. To, co oferuje The Rosetta Project, porównać można zaledwie z ludzkim protojęzykiem, w przypadku którego występowało jedynie słownictwo pozbawione jakichkolwiek reguł składniowych. Tak naprawdę, dopiero wynalezienie składni umożliwiło dalszy dynamiczny rozwój cywilizacyjny praludzkich społeczności.

Generatory struktur syntaktycznych

Rozwijane przez autorów generatory struktur syntaktycznych, stanowiące podstawę do rozwoju systemów przewidzianych do wspomagania procesu rekonstrukcji i rewitalizacji zagrożonych wymarciem języków, bazują na koncepcji tzw. języków kontrolowanych.

Pod pojęciem języka kontrolowanego będziemy rozumieli pewien skończony podzbiór dowolnego języka naturalnego (ewentualnie języka sztucznego, takiego jak choćby język esperanto), przy czym obowiązuje zależność, że każde zdanie języka kontrolowanego musi być także poprawnym pod względem składniowym zdaniem języka naturalnego, na którym dany język kontrolowany jest wzorowany. Natomiast bynajmniej nie każde zdanie danego języka naturalnego musi być możliwe do utworzenia w opartym na nim języku kontrolowanym.

W związku z powyższym język kontrolowany może być postrzegany jako swego rodzaju zubożona wersja danego języka naturalnego. W przypadku języków kontrolowanych zwykle silnemu okrojeniu podlega gramatyka, ponieważ twórcy danego języka kontrolowanego uwzględniają jedynie te konstrukcje syntaktyczne, które w danym języku są naprawdę niezbędne, pomijając bardziej egzotyczne elementy składniowe, które mogą zostać podczas tworzenia wypowiedzi zastąpione innymi, zwykle znacznie prostszymi.

Dzięki tego rodzaju zawężeniu zbioru dopuszczalnych struktur składniowych zyskujemy znaczną kontrolę nad gramatyką języka, co stanowić może podstawę do budowy wszelkiego rodzaju komputerowych systemów przetwarzania języka naturalnego, takich jak systemy automatycznego generowania tekstów danego języka, automatycznego streszczania dokumentów, przekładu komputerowego, komputerowych systemów dialogowych i wielu innych.

Języki kontrolowane są najczęściej projektowane jako systemy leksykalnie otwarte, co oznacza, że występujące w ich przypadku słownictwo może być systematycznie poszerzane o nowe jednostki leksykalne. Tym sposobem dany język kontrolowany może skutecznie nadążać za rozwojem języka naturalnego, na którym jest pierwotnie wzorowany[9].

Natomiast jeżeli chodzi o składnię języków kontrolowanych, to w praktyce spotykane są zarówno rozwiązania zamknięte, jak i otwarte składniowo. W przypadku tych ostatnich zbiór dopuszczalnych reguł składniowych może

[9] Szczepaniak l., Królikowski Z, *Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań*, Pro dialog, nr 11, 2000, s. 47-67.

zostać w każdej chwili dowolnie poszerzony, tak aby dany język kontrolowany w swej formie jeszcze bardziej się upodobnił się do swego pierwowzoru w postaci pewnego języka naturalnego.

Opracowywane przez autorów generatory struktur syntaktycznych w swym założeniu oparte są na językach kontrolowanych, które są projektowane jako systemy otwarte zarówno pod względem leksykalnym, jak i składniowym. Na obecnym etapie dopuszczalna składnia tworzonych przez autorów języków kontrolowanych pozwala na utworzenie następujących typów wypowiedzi:

– Fraza rzeczownikowa,
– Zdanie proste,
– Zdanie złożone.

Z kolei tworzona w języku kontrolowanym fraza rzeczownikowa może składać się z samego rzeczownika albo z rzeczownika wraz z określającym go przymiotnikiem. Ponadto każda fraza rzeczownikowa może zostać dodatkowo uzupełniona rodzajnikiem, zaimkiem wskazującym bądź zaimkiem dzierżawczym.

Jeśli chodzi o tworzone w języku kontrolowanym zdania proste, są one budowane na bazie fraz czasownikowych zarówno przechodnich, jak i nieprzechodnich. Ponadto użytkownik ma możliwość wyboru trybu tworzonego zdania (orzekający, przypuszczający i rozkazujący) oraz odpowiedniego czasu gramatycznego.

Na obecnym etapie rozwoju opracowanego przez autorów systemu języka kontrolowanego założono, że tworzone w nim zdania złożone składają się jedynie z dwóch zdań prostych złożonych podrzędnie lub współrzędnie.

Opisany powyżej zbiór reguł gramatycznych jest zbiorem minimalistycznym, jednakże jest zarazem zbiorem wystarczającym, tzn. takim, przy użyciu którego można w projektowanym języku kontrolowanym przekazać w zasadzie dowolne treści. Ponadto, jak już uprzednio wspomniano, rozwijane przez autorów języki kontrolowane są przewidziane jako systemy otwarte składniowo, a zatem nic nie stoi na przeszkodzie, aby dodać do nich nowe reguły gramatyczne, jeśli tylko zajdzie taka konieczność.

Przykłady zastosowania

Obecnie w kręgu zainteresowania autorów znajduje się grupa północnogermańskich języków skandynawskich. Spośród wymienionych języków powszechnie znane są języki, takie jak szwedzki, duński i norweski. W przypadku ostatniego z wymienionych mamy do czynienia z dwoma równoprawnymi standardami, określanymi odpowiednio nazwami bokmål (dosłownie język książkowy) oraz nynorsk (czyli język nowonorweski).

Drugi z wymienionych standardów literackich języka norweskiego jest poniekąd tworem sztucznym, który został utworzony przez norweskiego językoznawcę i poetę, Ivara Aasena na bazie wybranych dialektów ludowych Norwegii. Swego czasu język nynorsk zyskał w społeczeństwie norweskim sporą popularność, jednak obecnie znajduje się zdecydowanie w defensywie, ponieważ coraz większy odsetek społeczeństwa norweskiego deklaruje jako swój pierwszy język ojczysty właśnie wariant bokmål (jednak nauka obu wariantów języka norweskiego jest nadal obowiązkowa w szkołach).

Szacuje się, że na chwilę obecną zaledwie około 10% Norwegów używa jako pierwszego wariantu językowego nynorsk, a ponadto odsetek ten systematycznie spada. Sytuację językową Norwegii komplikuje dodatkowo fakt, że w powszechnym użyciu pozostaje także wiele różnorodnych lokalnych dialektów, cieszących się stosunkowo wysokim statusem społecznym (bywają często używane również i w sytuacjach oficjalnych, co traktowane jest jako coś zgoła normalnego, w żadnym wypadku nie wzbudzającego czyjegoś oburzenia czy też rodzącego swego rodzaju sensację). Ponieważ odsetek młodych osób deklarujących nynorsk jako swój pierwszy język w ostatnich latach silnie maleje, język ten może być postrzegany jako język potencjalnie zagrożony wymarciem. Jest rzeczą wysoce prawdopodobną, że za kilkadziesiąt lat w zasadzie wszyscy mieszkańcy Norwegii posługiwać się będą wariantem językowym bokmål, a lokalne dialekty wraz ze stworzonym na ich podstawie językiem nynorsk popadną ostatecznie w zapomnienie.

Można oczywiście argumentować, że takie ujednolicenie sytuacji językowej Norwegii będzie w przyszłości czynnikiem pozytywnym, w pewnym sensie integrującym i znacznie lepiej spajającym społeczeństwo tego kraju. Pozostawanie w użyciu tylko jednego wariantu językowego to także spore oszczędności, bowiem obecnie wszystkie oficjalne dokumenty państwowe muszą być sporządzane w obu wersjach językowych. Ponadto pisząc do jakiegokolwiek norweskiego urzędu w języku nynorsk, urząd ten jest zobowiązany do odpisania petentowi również w tej samej wersji językowej, co wymusza konieczność dodatkowego zatrudnienia osób biegle posługujących się tym mniej popularnym wariantem języka norweskiego.

Jednak z drugiej strony całkowity zanik opartego na ludowych dialektach Norwegii języka nynorsk jest zjawiskiem negatywnym, ponieważ prowadzi do poważnego zubożenia dziedzictwa kulturowego tego kraju z jego bogatym folklorem. W związku z powyższym sensownym wydaje się podejmowanie szeroko zakrojonych działań mających na celu utrzymanie przy życiu również i tego mniej popularnego standardu językowego.

Oprócz wymienionych na wstępie głównych skandynawskich języków północnogermańskich w użyciu pozostają jeszcze znacznie od nich mniejsze języki peryferyjne. W przeciwieństwie do języków głównych, które występują na kontynencie europejskim, peryferyjne języki północnogermańskie są używane jedynie na izolowanych wyspach. Taki stan rzeczy powoduje, że skandynawskie języki peryferyjne w stosunku do języków głównych zachowały wiele archaicznych cech w zakresie ich leksyki i składni. W istocie peryferyjne języki północnogermańskie znacznie bliższe są używanemu około tysiąca lat temu językowi staronordyckiemu niż współczesnym językom skandynawskim, takim jak szwedzki, duński czy norweski. Ponadto, w sytuacji gdy języki szwedzki, duński i norweski są w zasadzie w znacznym stopniu wzajemnie zrozumiałe dla każdego z ich użytkowników, to w żadnym wypadku nie można powiedzieć tego o językach peryferyjnych, które ze względu na ich archaiczność są bardzo trudno zrozumiałe dla przeciętnego Szweda, Duńczyka bądź Norwega.

Największym i zarazem najważniejszym, posiadającym status oficjalnego języka państwowego, spośród skandynawskich języków peryferyjnych jest północnogermański język islandzki, posiadający prawie 300 tysięcy użytkowników. Na razie język ten nie wydaje się być zagrożony wymarciem, głównie ze względu na geograficzną izolację Islandii oraz status oficjalnego języka państwowego.

Natomiast, jako język potencjalnie zagrożony wymarciem wymieniany jest często język farerski, będący w użyciu na Wyspach Owczych i posiadający obecnie jeszcze około 30 tysięcy użytkowników.

W swej formie język farerski jest mocno zbliżony do języka islandzkiego, jednak w przeciwieństwie do niego nie ma on statusu oficjalnego języka państwowego – posiada jedynie status języka regionalnego, dzięki czemu jest obowiązkowo nauczany w szkołach. Wiąże się z tym jednak pewne zagrożenie, ponieważ jeśli oficjalnym językiem państwowym jest język duński, to młodsze pokolenie mieszkańców Wysp Owczych może zacząć postrzegać własny język farerski jako nieco mniej atrakcyjny – cieszący się znacznie niższym prestiżem społecznym. Jeśli tak się stanie, to lepiej wykształcone warstwy społeczeństwa mieszkańców Wysp Owczych przestaną po prostu języka farerskiego na co dzień używać i przejdą na cieszący się o wiele większym prestiżem społecznym język duński. Ponadto w języku farerskim raczej nie powstają żadne poważniejsze teksty o charakterze naukowo-technicznym i w tym obszarze prym zdecydowanie wiedzie język duński. To wszystko niezbyt dobrze wróży, jeśli chodzi o przyszłość języka farerskiego i w związku z tym należy oczekiwać, że liczba jego użytkowników, jak i stopień czynnej jego znajomości będą w latach kolejnych systematycznie spadać.

W związku z powyższym opracowywane przez autorów generatory struktur syntaktycznych oparte na języku kontrolowanym mogą okazać się pomocne we wspieraniu procesu rewitalizacji języka farerskiego.

Ostatnim, najbardziej tajemniczym i zarazem najmniej znanym z peryferyjnych języków północnogermańskich jest język norn. Niestety, ale od ponad stu lat jest to już język całkowicie martwy, aczkolwiek istnieją pewne świadectwa, jakoby osoby posługujące się w sposób czyny językiem norn żyły jeszcze na początku XX wieku. Język norn był przez wiele wieków w powszechnym użyciu na Orkadach i Szetlandach oraz na terytorium północnej Szkocji. Niestety, omawiany język nie wytrzymał konkurencji z cieszącym się znacznie większym prestiżem społecznym i popularnością językiem angielskim, więc ostatecznie popadł w zupełne zapomnienie.

Obecnie jako jedyne żywe świadectwo po wymarłym niegdyś języku norn pozostały pewne wyrazy powiązane semantycznie ze specyfiką regionu oraz nieliczne utarte zwroty frazeologiczne, które wciąż jeszcze są używane przez mieszkańców Orkadów i Szetlandów. Ponadto, wartym odnotowania jest fakt, że mimo tego, iż obecnie na Orkadach i Szetlandach mówi się wyłącznie po angielsku, to jednak mieszkańcy tego regionu wymawiają ten język w bardzo charakterystyczny i specyficzny sposób. Polega to na tym, że słuchając mieszkańców Orkadów czy Szetlandów z nieco większej odległości, gdy nie jesteśmy już w stanie zrozumieć znaczenia poszczególnych słów, wydawać się może, że mówią oni w jakimś północnogermańskim języku skandynawskim z jego charakterystyczną prozodią opartą na akcencie tonicznym.

Ponieważ w języku norn zachowały się do naszych czasów pewne zabytki w postaci tekstów pisanych, podejmowane są obecnie inicjatywy mające na celu rekonstrukcję tego języka, a następnie jego ożywienie poprzez czynne używanie w gronie jego miłośników. Sytuację komplikuje dodatkowo fakt, że w przeszłości język norn nie był językiem w miarę jednorodnym, lecz w użyciu było mnóstwo jego dialektów, które dodatkowo były wzajemnie silnie zróżnicowane. Wersja języka norn, która była używana na północnym terytorium Szkocji, mocno odbiegała od dialektów Orkadów i Szetlandów, które to z kolei znacznie różniły się od siebie.

Z wymienionych powodów lingwiści podejmujący próbę rekonstrukcji języka norn de facto tworzą pewien język sztuczny, stanowiący swoistą syntezę lokalnych dialektów. Rozważany język określany jest mianem języka nynorn, czyli nowego języka norn. W zamierzeniu autorów opracowywany przez nich system generatora struktur syntaktycznych może stanowić nieodzowną pomoc w obszarze działań związanych z rekonstrukcją, a następnie rewitalizacją języka nynorn.

Obecnie podejmowane przez autorów działania zmierzają do utworzenia w języku JavaScript prototypu opracowanego przez nich generatora struktur syntaktycznych bazującego na językach kontrolowanych. Opracowany prototyp stanowił będzie podstawę do budowy pełnosprawnego już systemu generatorów struktur syntaktycznych dla języków farerskiego i nynorn, mających na celu rewitalizację, a w drugim przypadku także i rekonstrukcję wymienionych języków.

Na rys. 2 zamieszczono zrzut ekranu przedstawiający panel umożliwiający tworzenie poprawnych składniowo fraz składających się z rzeczownika wraz z określającym go przymiotnikiem, uzupełnionych dodatkowo rodzajnikami, zaimkami wskazującymi lub dzierżawczymi.

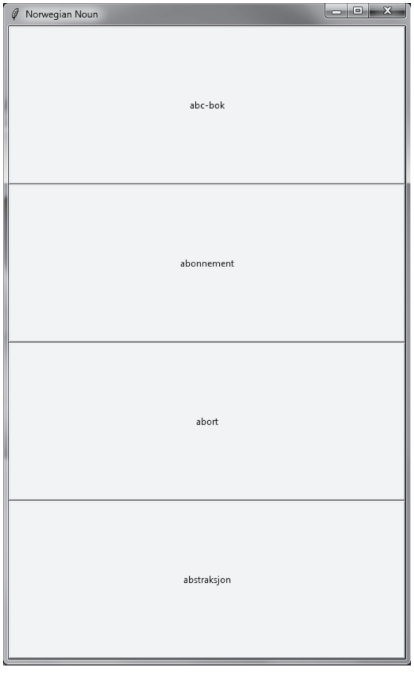
Rys. 2. Zrzut ekranu przedstawiający panel umożliwiający dokonanie wyboru elementów dodatkowych frazy rzeczownikowej

Źródło: opracowanie własne.

Jak wynika z rysunku 2, użytkownik generatora struktur syntaktycznych budując frazę rzeczownikową, jeszcze przed wyborem odpowiedniego rzeczownika i ewentualnie określającego go przymiotnika, może zadecydować, czy tworzona przez niego fraza rzeczownikowa ma być poprzedzona rodzajnikiem (określonym bądź nieokreślonym), zaimkiem wskazującym (obiekty bliższe bądź dalsze) lub zaimkiem dzierżawczym. Użytkownik może także zadecydować, że tworzona przez niego fraza rzeczownikowa ma składać się wyłącznie z rzeczownika, czy też ma to być rzeczownik wraz z określającym go przymiotnikiem.

Interesująca alternatywą, którą autorzy biorą pod uwagę, jest także zastosowanie w celu budowy generatora struktur syntaktycznych nowoczesnego języka programowania Python. W języku tym autorzy budują obecnie generator struktur syntaktycznych dla języka norweskiego. Na rys. 3 pokazano zrzut ekranu, na którym widoczne jest okno wyboru norweskiego rzeczownika.

Rys. 3. Zrzut ekranu przedstawiający panel umożliwiający dokonanie wyboru norweskiego rzeczownika



Źródło: opracowanie własne.

Z kolei na rys. 4 przedstawiono zrzut ekranu umożliwiający wybór dodatkowych elementów, o które ma być uzupełniona tworzona przez użytkownika fraza rzeczownikowa.

Rys. 4. Zrzut ekranu przedstawiający panel umożliwiający dokonanie wyboru dodatkowych elementów frazy rzeczownikowej

Źródło: opracowanie własne.

Poniżej przedstawiono fragment kodu programu w języku Python, który definiuje wszystkie cztery możliwe do utworzenia formy norweskiego rzeczownika „badeværelse" (pol. łazienka):

```
def rzecz_BA_5():
    global pojNieOkr
    global pojOkr
```

global mnoNieOkr
global mnoOkr
global rodzMas
global rodzFem
global rodzNeu
global tantum
pojNieOkr = „badeværelse”
pojOkr = „badeværelset”
mnoNieOkr = „badeværelser”
mnoOkr = „badeværelsene”
rodzMas = 0
rodzFem = 0
rodzNeu = 1
tantum = 0

Znaczenie zmiennych występujących w zamieszczonym powyżej kodzie jest następujące:

– pojNieOkr – forma nieokreślona dla liczby pojedynczej
– pojOkr – forma określona dla liczby pojedynczej
– mnoNieOkr – forma nieokreślona dla liczby mnogiej
– mnoOkr – forma określona dla liczby mnogiej
– rodzMas – wyznacznik rodzaju męskiego
– rodzFem – wyznacznik rodzaju żeńskiego
– rodzNeu – wyznacznik rodzaju neutralnego
– tantum – zmienna niosąca informację o ograniczeniach odnośnie możliwości utworzenia liczby pojedynczej lub mnogiej dla danego rzeczownika

Podsumowanie

Omawiany w artykule generator struktur syntaktycznych stanowi oryginalne przedsięwzięcie zmierzające do opracowania narzędzi informatycznych wspierających proces ratowania przed popadnięciem w niepamięć zagrożonych wymarciem języków. Opracowywany przez autorów system może istotnie usprawnić proces rekonstrukcji wymarłych już języków oraz przyczynić się do ich efektywnej rewitalizacji. Dalsza rozbudowa tworzonego przez autorów narzędzia informatycznego zmierzała będzie w kierunku budowy systemów wspomagających przekład typu MAHT (ang. Machine-Aided Human

Translation), tak aby poprawne składniowo wypowiedzi w językach zagrożonych wymarciem mogły być równie tworzone przez osoby w ogóle nie znające tych języków[10]. Kolejnym planowanym krokiem jest integracja powstałych systemów z systemami syntezy bądź rekonstrukcji mowy, tak aby tworzone w zagrożonych językach wypowiedzi mogły być również przez użytkowników odsłuchane. Tego rodzaju narzędzia informatyczne mogą okazać się niezwykle pomocne w procesie nauczania zagrożonych wymarciem języków, co jest niezwykle istotne z punktu widzenia ich dalszego przetrwania jako języków żywych.

Literatura

1. Corballis M. C., Dessalles J. L., Dunbar R.: A*ux origines du langage*, La Recherche, no. 341, 2001, s. 27-39.
2. Danecki J.: *Współczesny język arabski i jego dialekty*, Wydawnictwo Akademickie DIALOG, Warszawa, 2000.
3. Gajer M.: *Wielojęzyczne systemy automatycznego przekładu oparte na metodzie wzorców translacyjnych*, AGH Uczelniane Wydawnictwa naukowo-Dydaktyczne, Kraków, 2008.
4. Kondratow A., *Zaginione cywilizacje*, Państwowy Instytut Wydawniczy, Warszawa, 1988.
5. Künstler J. M.: *Języki chińskie*, Wydawnictwo Akademickie DIALOG, Warszawa, 2000.
6. Majewicz A. F.: *Języki świata i ich klasyfikowanie*, Państwowe Wydawnictwo Naukowe, Warszawa, 1989.
7. Matisoff J. A.: *Zagrożona różnorodność: języki i formy życia*, Świat Nauki, październik 2002, s. 66-73
8. Szczepaniak l., Królikowski Z, *Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań*, Pro dialog, nr 11, 2000, s. 47-67.

[10] Gajer M.: *Wielojęzyczne systemy automatycznego przekładu oparte na metodzie wzorców translacyjnych*, AGH Uczelniane Wydawnictwa naukowo-Dydaktyczne, Kraków, 2008.