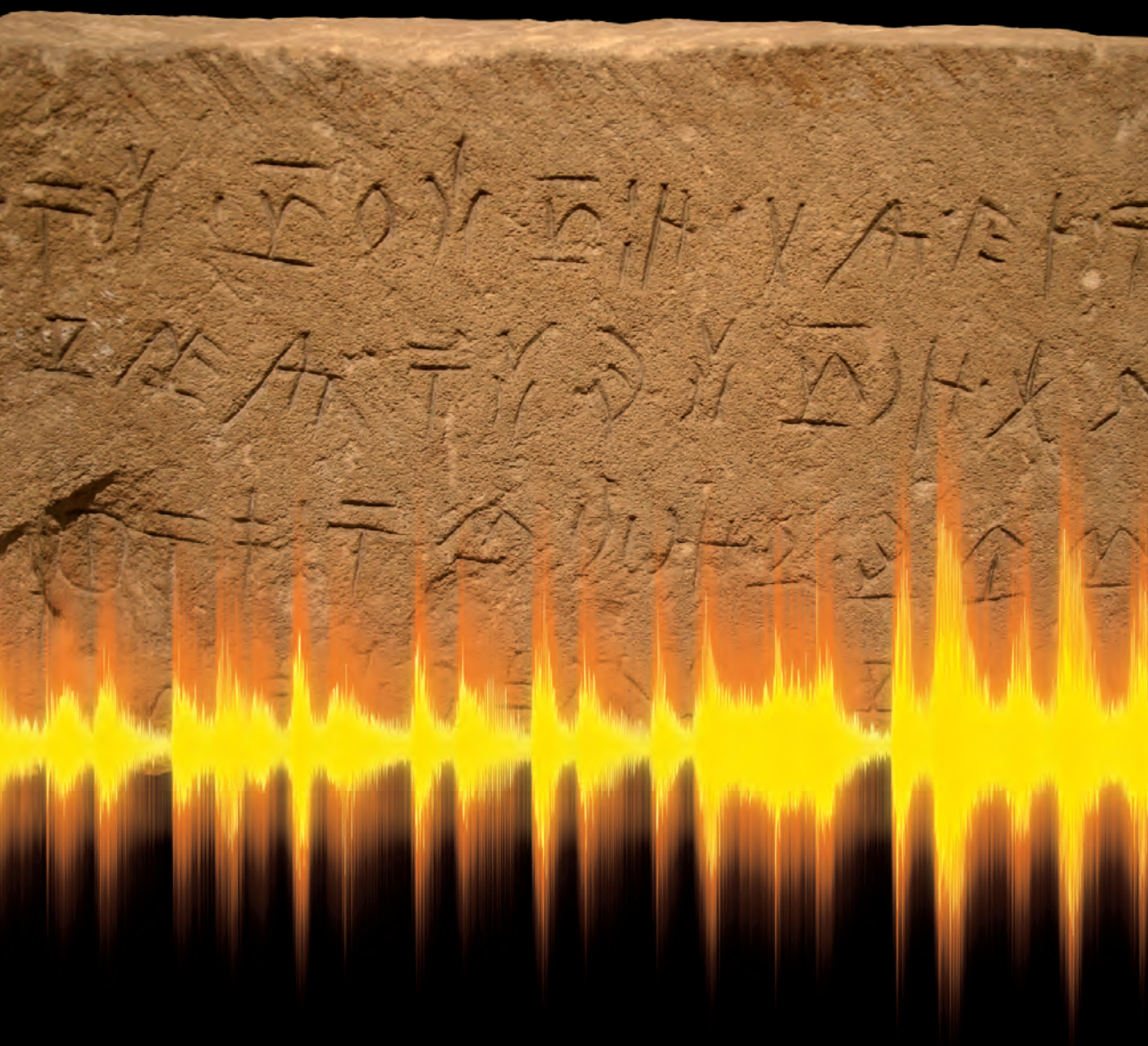
MIROSŁAW GAJER | ZBIGNIEW HANDZEL

REKONSTRUKCJA I REWITALIZACJA ZAGROŻONYCH WYMARCIEM JĘZYKÓW Z WYKORZYSTANIEM NARZĘDZI LINGWISTYKI KOMPUTEROWEJ



REKONSTRUKCJA I REWITALIZACJA ZAGROŻONYCH WYMARCIEM JĘZYKÓW Z WYKORZYSTANIEM NARZĘDZI LINGWISTYKI KOMPUTEROWEJ

MIROSŁAW GAJER | ZBIGNIEW HANDZEL

REKONSTRUKCJA I REWITALIZACJA ZAGROŻONYCH WYMARCIEM JĘZYKÓW Z WYKORZYSTANIEM NARZĘDZI LINGWISTYKI KOMPUTEROWEJ

WYDAWNICTWO KSIĘGARNIA AKADEMICKA

KRAKÓW 2021

dr inż. Mirosław Gajer
AGH Akademia Górniczo-Hutnicza w Krakowie
https://orcid.org/0000-0003-3532-9482
gajer@agh.edu.pl

dr inż. Zbigniew Handzel
Wyższa Szkoła Ekonomii i Informatyki w Krakowie
https://orcid.org/0000-0003-1470-6592
zhandzel@wsei.edu.pl

Recenzenci
dr hab. inż. Piotr Szymczyk, prof. AGH
dr hab. inż. Tadeusz Szuba, prof. UPJPII

Opracowanie redakcyjne
Ewa Dąbrowska

Projekt okładki
Lesław Sławiński

ISBN 978-83-8138-568-8 (druk)
ISBN 978-83-8138-569-5 (PDF)
https://doi.org/10.12797/9788381385695

Na okładce wykorzystano zdjęcie inskrypcji w języku eteocypryjskim ze zbiorów Ashmolean Musuem w Oxfordzie, V-IV w. p.n.e., źródło: Wikimedia Commons

Publikacja dofinansowana przez Wyższą Szkołę Ekonomii i Informatyki w Krakowie

WYDAWNICTWO KSIĘGARNIA AKADEMICKA
ul. św. Anny 6, 31-008 Kraków
tel.: 12 421-13-87; 12 431-27-43
e-mail: publishing@akademicka.pl

Księgarnia internetowa: https://akademicka.com.pl

Spis treści

Wstęp

Według nawet najbardziej optymistycznych scenariuszy ponad połowa języków będących obecnie w użyciu na kuli ziemskiej przestanie definitywnie istnieć do końca bieżącego stulecia. Śmierć języka następuje nieuchronnie wraz ze śmiercią ostatniej znającej go czynnie osoby, jednakże dochodzi do niej być może nawet nieco wcześniej: wydaje się, że aby dany język można było z pełnym przekonaniem uznać za jeszcze w pełni żywy, muszą istnieć przynajmniej dwie osoby, które za jego pośrednictwem wzajemnie się ze sobą komunikują.

Wymieranie języków to zjawisko bezdyskusyjnie niekorzystne, gdyż przyczynia się do katastrofalnego zubożenia kulturowego dziedzictwa ludzkości. Pamiętać trzeba, że każdy język stanowi oryginalny sposób postrzegania świata i opisu otaczającej jego użytkowników, nader bogatej rzeczywistości. Językoznawcy kognitywni mówią wręcz o istnieniu językowego obrazu świata, który bezpowrotnie ginie wraz ze śmiercią powiązanego z nim języka[1]. Ponadto proces wymierania języków przyczynia się do istotnego zubożenia materiału badawczego, który potencjalnie mógłby zostać wykorzystany przez lingwistów w prowadzonych przez nich dociekaniach nad ewolucją i pochodzeniem języków ludzkich[2]. Obszerniejszy materiał badawczy umożliwia również bardziej wiarygodne testowanie hipotez lingwistycznych, dotyczących na przykład kwestii istnienia tak zwanych uniwersaliów językowych czy też pytań o granice możliwości rozwoju struktur morfologicznych i syntaktycznych języków, a także ich warstwy leksykalnej, systemów fleksyjnych, fonetycznych i fonologicznych[3].

Oczywiście prawdziwe jest niewątpliwie stwierdzenie, że języki ludzkie wymierały zawsze. Jednak w przeszłości proces ten był zapewne z naddatkiem równoważony systematycznym pojawianiem się nowych języków poprzez ich nieustanne różnicowanie się dialektalne, co w konsekwencji dawało zaczątek nowym grupom, a nawet rodzinom językowym. Niemniej obecnie, wskutek niezwykle dynamicznych procesów globalizacyjnych, zjawisko wymierania języków o niewielkiej liczbie użytkowników zdecydowanie przybrało na sile, a proces powstawania nowych języków został praktycznie całkowicie zahamowany, do czego przyczyniają się głównie standaryzacja języków już istniejących, upowszechnienie systemu szkolnictwa i związana z tym presja na używanie poprawnej wersji danego języka, piętnująca zarazem jakiekolwiek odstępstwa od powszechnie obowiązującej normy. Te dwa procesy, postępujące w wyraźnie przeciwnych kierunkach,

---

[1] E. Tabakowska, *O przekładzie na przykładzie. Rozprawa tłumacza z „Europą" Normana Daviesa*, wstęp N. Davies, Społeczny Instytut Wydawniczy Znak, Kraków 1999.

[2] M.C. Corballis, J.L. Dessalles, R. Dunbar, *Aux origines du langage*, „La Recherche" 2011, nr 341, s. 27-39.

[3] J.A. Matisoff, *Zagrożona różnorodność: języki i formy życia*, „Świat Nauki" 2002, nr 10, s. 66-73.

muszą w efekcie prowadzić do drastycznego zmniejszenia się liczby języków używanych na świecie, i to w niezbyt odległej przyszłości.

W związku z powyższym w wielu krajach podejmowane są liczne inicjatywy zmierzające – tam, gdzie jest to jeszcze w ogóle możliwe – do zachowania zagrożonych wymarciem języków przy życiu. Natomiast w przypadku języków wymarłych w grę wchodzi jedynie możliwie jak najwierniejsza ich rekonstrukcja w celu przekazania wiedzy o nich kolejnym pokoleniom badaczy. W ocenie autorów podejmowane dotychczas w tym kierunku działania należy uznać za raczej mało skuteczne, ponieważ mają one charakter bierny: odnotowuje się najczęściej wyłącznie ich warstwę leksykalną. Trzeba być świadomym, że sporządzenie li tylko listy używanych w danym języku wyrazów nie sprawi, że w przyszłości język ten stanie się ponownie żywy. Do przeprowadzenia rewitalizacji języka potrzebne są bowiem jeszcze liczne dodatkowe informacje co do tego, jak poszczególne jego wyrazy można łączyć ze sobą, aby otrzymać poprawne syntaktycznie wypowiedzenia (frazy, równoważniki zdań, zdania proste i zdania złożone).

W ocenie autorów proces rewitalizacji zagrożonych wymarciem języków, a także rekonstrukcji języków od pewnego czasu martwych, wymaga podjęcia zdecydowanych działań o charakterze czynnym. Obecnie realizację tego typu inicjatyw mogą wspomagać dotychczasowe osiągnięcia lingwistyki komputerowej. W prezentowanej czytelnikowi monografii przedstawiono propozycję wykorzystania koncepcji języków kontrolowanych oraz opartych na nich generatorów struktur syntaktycznych w celu umożliwienia ich użytkownikom budowania poprawnych składniowo zdań w wybranym języku naturalnym. Dalekosiężnym celem jest budowa systemów komputerowego wspomagania przekładu na wybrane języki zagrożone wymarciem, co powinno umożliwić komunikowanie się w tych językach również osobom nieznającym ich wcale bądź znającym je jedynie w ograniczonym zakresie. W ten sposób zagrożone wymarciem języki będzie można w przyszłości skutecznie rewitalizować, a języki wymarłe – ponownie przywrócić do życia.

Niniejsza monografia składa się z sześciu rozdziałów. W rozdziale pierwszym przedyskutowana została szczegółowo obecna sytuacja językowa świata. Główną uwagę poświęcono licznym językom posiadającym stosunkowo niewielką liczbę użytkowników, które obecnie są poważnie zagrożone wymarciem. Dokonano wnikliwej analizy przyczyn takiego stanu rzeczy oraz zaproponowano pewne działania zaradcze, zmierzające do rewitalizacji języków zagrożonych wymarciem oraz rekonstrukcji języków wymarłych. Rozważane działania polegają na wykorzystaniu osiągnięć lingwistyki komputerowej, a zwłaszcza tworzonych w jej ramach narzędzi informatycznych bazujących na koncepcji tak zwanych języków kontrolowanych oraz generatorów struktur syntaktycznych.

W kolejnych rozdziałach – drugim, trzecim i czwartym – omówiono pokrótce wybrane języki należące do indoeuropejskiej rodziny językowej, zagrożone wymarciem na terytorium Europy (niektóre z nich jakiś czas temu przestały już istnieć). W tym miejscu należy koniecznie podkreślić, że dokonany przez autorów wybór rozważanych języków zagrożonych wymarciem jest w zasadniczej mierze arbitralny, a niniejsza praca w żadnym

wypadku nie rości sobie pretensji do bycia wyczerpującym przeglądem tego rodzaju języków. Autorzy monografii skupili uwagę na językach, które od pewnego czasu pozostają w obszarze ich zainteresowań i w których przypadku planują rozwijanie różnego typu narzędzi informatycznych, bazujących na koncepcji języków kontrolowanych, wspomagających proces rekonstrukcji i późniejszej rewitalizacji wybranych języków.

Rozdział drugi poświęcony został zagrożonym wymarciem językom indoeuropejskim zaliczanym do germańskiej grupy językowej, rozdział trzeci – wybranym językom romańskim, które są w różnym stopniu zagrożone wymarciem bądź wyszły już całkowicie z użycia. Z kolei w rozdziale czwartym omówiono wybrane zagrożone wymarciem języki zaliczane w ramach indoeuropejskiej rodziny językowej do mniej licznych grup, takich jak celtycka, bałtycka i helleńska. W przypadku grupy celtyckiej wszystkie należące do niej języki, które są jeszcze obecnie używane, są poważnie zagrożone wymarciem. W związku z powyższym w przeszłości mogą wymrzeć nie tyle pojedyncze języki, ile cała ta grupa językowa, co z pewnością będzie stanowiło niepowetowaną stratę dla nauki i ogólnoludzkiej kultury.

W niniejszej pracy nie zostały omówione zagrożone wymarciem języki należące w ramach indoeuropejskiej rodziny językowej do grupy słowiańskiej. Nie jest to bynajmniej skutkiem jakiegoś niedopatrzenia czy też jakichkolwiek uprzedzeń autorów – wręcz przeciwnie, decyzja została podjęta w sposób całkowicie świadomy. Autorzy uznali, że wymierające języki słowiańskie oraz narzędzia informatyczne przeznaczone do ich rekonstrukcji i rewitalizacji stanowią temat na tyle obszerny i interesujący, że warto mu z pewnością poświęcić w przyszłości odrębną monografię.

W rozdziale piątym omówiono przykładowy generator struktur syntaktycznych opracowany przez autorów dla języka norweskiego, oparty na koncepcji języka kontrolowanego. Rozważany generator struktur syntaktycznych umożliwia użytkownikowi tworzenie poprawnych składniowo zdań w kontrolowanym języku norweskim, stanowiącym pewien podzbiór bardziej rozpowszechnionej w Norwegii wersji językowej – bokmål. W przyszłości planowana jest budowa analogicznego systemu również dla mniej popularnej wersji języka norweskiego – tak zwanego języka nowonorweskiego (nynorsk), który obecnie zaczyna być powoli postrzegany jako język potencjalnie zagrożony wymarciem, ze względu na systematycznie zmniejszającą się liczbę jego użytkowników i postępującą utratę popularności wśród młodszego pokolenia Norwegów.

Ostatni rozdział stanowi podsumowanie całości zagadnień zaprezentowanych w monografii. Przedstawiono w nim także perspektywy dalszego rozwoju opracowywanych przez autorów systemów opartych na koncepcji języków kontrolowanych. Planowana rozbudowa omówionego w poprzednim rozdziale systemu generatora struktur syntaktycznych przebiegała będzie głównie w kierunku budowy narzędzi informatycznych wspomagających proces przekładu. Tego rodzaju systemy, określane w literaturze przedmiotu skrótem MAHT (ang. *Machine-Aided Human Translation*), mają za zadanie umożliwić użytkownikowi symultaniczne tłumaczenie tworzonych przezeń na bieżąco

w języku ojczystym zdań na wybrany inny język, którego w zasadzie w ogóle nie zna bądź którym posługuje się jedynie w mocno ograniczonym zakresie.

W głębokim przekonaniu autorów omówione w niniejszej monografii generatory struktur syntaktycznych oparte na koncepcji wykorzystania języków kontrolowanych mogą stanowić użyteczne narzędzia informatyczne, które powinny istotnie wspomóc niełatwe ze swej natury zadanie rekonstrukcji i rewitalizacji ginących wręcz na naszych oczach języków świata.

Przynajmniej połowa z około siedmiu tysięcy używanych obecnie na świecie języków jest skrajnie zagrożona wymarciem i większości z nich najprawdopodobniej nie uda się już uratować. Autorzy monografii próbują dociec przyczyn takiego stanu rzeczy, skupiając uwagę na europejskiej sytuacji w zakresie tej problematyki. Niestety, wciąż brakuje usystematyzowanych prac zmierzających do efektywnego gromadzenia wiedzy o językach zagrożonych wymarciem. To ostatnia szansa na podjęcie skutecznych działań w celu zachowania cennego materiału badawczego, będącego jednocześnie ginącym na naszych oczach dziedzictwem kulturowym ludzkości.

Autorzy monografii proponują zatem nowatorskie podejście polegające na opracowywaniu stosownych narzędzi informatycznych w postaci tzw. generatorów struktur syntaktycznych. Tego rodzaju programy komputerowe pozwalają na przechowanie wiedzy dotyczącej zarówno leksyki, jak i reguł składniowych obowiązujących w ginących językach. Tym samym takie oprogramowanie umożliwia użytkownikom skuteczne przeprowadzenie rekonstrukcji zdań zapisanych w językach zagrożonych wymarciem. Dalekosiężnym celem jest budowa systemów komputerowego wspomagania przekładu na wybrane języki, co powinno umożliwić komunikowanie się w nich także osobom nieznającym ich wcale bądź znającym je jedynie w ograniczonym zakresie. To kolejny, do niedawna futurologiczny etap na drodze budowania uniwersalnych sposobów komunikacji, dlatego niniejsza interdyscyplinarna monografia jest godną uwagi lekturą nie tylko dla adeptów lingwistyki, etnografii i kulturoznawstwa oraz informatyków zajmujących się problematyką lingwistyki komputerowej i zagadnieniami przetwarzania języka naturalnego.

Dr inż. Mirosław Gajer – adiunkt w Katedrze Informatyki Stosowanej AGH. Zainteresowania naukowe wiąże ze sztuczną inteligencją, zwłaszcza inteligencją obliczeniową, a także ze sztucznymi systemami ewolucyjnymi, lingwistyką komputerową, inżynierią lingwistyczną, przetwarzaniem języka naturalnego i przekładem komputerowym. Jest autorem dwóch monografii poświęconych zagadnieniom tłumaczenia komputerowego oraz ponad 200 artykułów opublikowanych w recenzowanych czasopismach naukowych i materiałach konferencyjnych. W swojej praktycznej działalności skupia się głównie na opracowywaniu w języku Python narzędzi informatycznych przeznaczonych dla lingwistów, filologów i tłumaczy.

Dr inż. Zbigniew Handzel – profesor WSEI. Do jego zainteresowań badawczych należą takie zagadnienia jak systemy czasu rzeczywistego oraz sztuczna inteligencja, a także lingwistyka komputerowa, inżynieria lingwistyczna, przetwarzanie języka naturalnego i przekład komputerowy. Jest autorem ponad 90 artykułów opublikowanych w recenzowanych czasopismach naukowych oraz w materiałach konferencyjnych. Zajmuje się też m.in. technologiami webowymi, sieciami komputerowymi i architekturą komputerową oraz problematyką bezpieczeństwa systemów informatycznych.

ISBN 978-83-8138-568-8

9 788381 385688

WYDAWNICTWO KSIĘGARNIA AKADEMICKA

https://akademicka.pl