

## WPROWADZENIE DO METODY TRANSLACJI AUTOMATYCZNEJ OPARTEJ NA WZORCACH

### STRESZCZENIE

Artykuł niniejszy jest pierwszym artykułem z większego cyklu prac poświęconych opracowanej przez autora metodzie translacji automatycznej opartej na wzorcach PBMT (Pattern Based Machine Translation). Metoda ta stanowi połączenie istotnych cech znanych do chwili obecnej metod translacji automatycznej, takich jak tłumaczenie poprzez analogię EBMT (Example-Based Machine Translation) oraz tłumaczenie oparte na regułach gramatycznych RBMT (Rule-Based Machine Translation). Zdaniem autora zaproponowana metoda może stanowić istotny krok na drodze do zbudowania w pełni automatycznych systemów komputerowego przekładu, charakteryzujących się stosunkowo wysoką jakością uzyskiwanych tłumaczeń tekstów należących do języka ogólnego. Jednakże zbudowanie pierwszego pełnosprawnego systemu translacji automatycznej opartego na wzorcach translacyjnych wymaga zarówno dużego nakładu pracy lingwistycznej, jak i dalszych badań, w związku z czym jest sprawą raczej odległej przyszłości. Istota metody PBMT została w artykule zilustrowana na licznych przykładach zaczerpniętych z kilku języków świata. Ponadto w artykule zamieszczono krótkie wprowadzenie do dziedziny automatycznego przekładu, omówiono historię badań w tym zakresie oraz scharakteryzowano podstawowe metody i kierunki dalszego rozwoju systemów translacji automatycznej.

**Słowa kluczowe:** lingwistyka informatyczna, translacja automatyczna, wzorce translacyjne

### THE INTRODUCTION TO THE MACHINE TRANSLATION METHOD BASED ON PATTERNS

The paper is the first one of the series of papers concerning the new method of machine translation that was proposed by this author. The method was called by this author Pattern-Based Machine Translation (PBMT) and it is a combination of the most important features of Example-Based and Rule-Based machine translation techniques. In this author's opinion this method can make possible to build fully-automatic high-quality machine translation systems for unrestricted text. Building the first fully-operating pattern-based machine translation system requires a lot of time and work of the team of trained linguists. In the paper the most important aspects of the pattern-based machine translation method are described. Moreover the way in which such machine translation system operates was illustrated on many examples for different languages of the world. In the paper a short introduction to the field of machine translation was also presented. The brief history of machine translation was described and the most common machine translation techniques are discussed.

**Keywords:** computational linguistics, machine translation, pattern-based translation

### 1. WPROWADZENIE

W ostatnich latach można zaobserwować wzmożone zainteresowanie technikami informatycznymi związanymi z przetwarzaniem, analizą, rozpoznawaniem i interpretacją języka naturalnego [1, 2]. W związku z powyższym opracowywane są systemy informatyczne operujące na tzw. kontrolowanych językach naturalnych, stanowiących ograniczone podzbiory języków naturalnych, które mogą być efektywnie i precyzyjnie przetwarzane w systemie komputerowym [3]. Prowadzone są również prace nad matematycznym modelowaniem języków naturalnych [4]. W tym miejscu można wspomnieć, między innymi, o badaniach autora dotyczących systemów automatycznego rozpoznawania języka tekstów [5].

Odrębnym obszarem badawczym jest wypracowywanie technik automatycznego rozpoznawania sygnału mowy i jego konwersji do postaci tekstowej [6, 7]. Z kolei systemy automatycznego rozpoznawania mowy znajdują zastosowanie w tzw. automatycznych systemach dialogowych, któ-

rych zadaniem jest prowadzenie w czasie rzeczywistym dialogu z użytkownikiem, przy czym tematyka tego dialogu dotyczy zwykle pewnej wąskiej dziedziny działalności ludzkiej takiej, jak np. rezerwacja biletów lotniczych lub miejsc w hotelach [1, 8, 9].

Innym obszarem badawczym związanym z przetwarzaniem języka naturalnego jest translacja automatyczna, w tym również translacja mowy [10].

Translacja automatyczna, zwana również niekiedy przekładem mechanicznym lub translacją maszynową MT (Machine Translation), jest dziedziną zastosowań informatyki, dostarczającą wiedzy o tym, jak pisać programy komputerowe, które byłyby w stanie dokonywać automatycznych przekładów pomiędzy wybranymi językami naturalnymi, np. między językiem niemieckim i francuskim [10, 11].

Pierwsze rozważania nad możliwościami zautomatyzowania translacji datują się co najmniej na lata 30. XX wieku. W roku 1933 radziecki inżynier Piotr Trojanskij opatentował w Moskwie urządzenie, które miało pełnić rolę

\* Katedra Automatyki AGH

mechanicznego słownika, dokonującego automatycznej zamiany wprowadzanych słów należących do języka źródłowego na ich odpowiedniki w języku docelowym przekładu [13]. Niestety pomysł Trojanskiego spotkał się z chłodnym przyjęciem w świecie nauki radzieckiej, nie wzbudził większego zainteresowania, został zapomniany i w związku z tym nie doczekał się nigdy praktycznej realizacji, a pierwszeństwo w dziedzinie automatyzacji przekładu przypadło Amerykanom.

Otóż, już w rok po skonstruowaniu pierwszego komputera amerykański teoretyk informacji Warren Weaver, rozważając możliwości innych niż numeryczne zastosowań maszyn cyfrowych, wymienił również automatyzację przekładu pomiędzy językami naturalnymi. Dowodem tego jest korespondencja, jaką prowadził Weaver w 1947 roku z brytyjskim krystalografem Andrew Boothem. W 1949 roku Weaver wystosował do amerykańskiej pozarządowej instytucji wspierającej rozwój nauki, zwanej Rockefeller Foundation, której był zresztą wiceprezydentem, swoje słynne memorandum zatytułowane „Przekład”. W memorandum tym Weaver postulował podjęcie badań naukowych nad możliwościami zastosowań komputerów w automatyzacji translacji pomiędzy językami naturalnymi [11].

Memorandum Weavera znalazło duży oddźwięk w świecie nauki amerykańskiej, bowiem już w 1952 roku na MIT została utworzona pierwsza grupa badawcza ukierunkowana specjalnie na translację automatyczną [13]. Z kolei w 1954 roku odbył się w Nowym Jorku pierwszy publiczny pokaz działającego systemu translacji automatycznej. Podczas pokazu komputer dokonał przekładu 49 prostych rosyjskich zdań na język angielski. O prostocie rozważanego systemu translacji automatycznej świadczyć może fakt, że dysponował on bilingwicznym słownikiem o objętości zaledwie 250 haseł oraz posługiwał się tylko sześcioma regułami gramatycznymi. Pokaz ten wywołał wśród badaczy falę entuzjazmu i spowodował znaczne zintensyfikowanie badań w obszarze translacji automatycznej [12]. Od 1956 roku badania nad translacją automatyczną zaczęto z powodzeniem prowadzić również w Związku Radzieckim [14].

Niestety po kilku latach szalonej euforii nadszedł czas krytyki. Największy oddźwięk w świecie nauki wywołał opublikowany w 1959 roku raport izraelskiego filozofa Bar Hillela. W raporcie tym Bar Hillel podał szereg argumentów na poparcie głoszonej przez siebie tezy, zgodnie z którą w pełni automatyczna translacja o wysokiej jakości FAHQMT (*Fully-Automatic High-Quality Machine Translation*) nie jest możliwa, i to nie tylko w chwili głoszenia tej tezy, z powodów natury czysto technicznej, ale nie jest możliwa z zasady, i w związku z tym nie będzie mogła zostać zrealizowana również w przyszłości [11].

Bezpośrednią konsekwencją publikacji Bar Hillela było wydanie w 1964 roku przez Amerykańską Akademię Nauk słynnego raportu ALPAC (Automatic Language Processing Advisory Committee). Podsumowując raport ALPAC, można powiedzieć, że uznawał on prowadzenie prac nad zautomatyzowaniem translacji za zwykłą stratę czasu i pieniędzy,

co było spowodowane tym, że istniejące wówczas systemy translacji automatycznej charakteryzowały się niską jakością przekładu i wolnym tempem pracy. Ponadto przy ówczesnym stanie techniki nie było widać żadnych perspektyw na poprawienie istniejącego stanu rzeczy [11].

Opublikowanie raportu ALPAC spowodowało w Stanach Zjednoczonych obcięcie wszelkich funduszy przeznaczanych na badania w dziedzinie translacji automatycznej, co jednocześnie sprawiło, że szeregi badaczy gotowych poświęcić się tej dyscyplinie naukowej stopniały praktycznie do zera. Nieliczni, którzy pozostali wierni translacji automatycznej, musieli szukać sponsorów w kręgach pozarządowych. Jako ciekawostkę można podać, że w owym czasie jednym ze sponsorów badań nad translacją automatyczną był, między innymi, amerykański kościół mormonów, który był zainteresowany zautomatyzowaniem przekładu tekstów biblijnych [13].

Renesans w dziedzinie translacji automatycznej nastąpił dopiero pod koniec lat 70. XX w. Pierwszym wdrożeniem systemu translacji automatycznej, które stało się komercyjnym sukcesem, było oddanie do użytku w 1977 roku w Montrealu systemu METEO. Jako ciekawostkę można podać, że system ten działa do dnia dzisiejszego i tłumaczy codziennie z języka angielskiego na francuski biuletyny zawierające informacje meteorologiczne o średniej objętości 45 000 słów [11].

W latach 80. i 90. XX w. opracowano na świecie wiele systemów translacji automatycznej, które całkiem dobrze spełniły stawiane przed nimi oczekiwania. Wymienić w tym miejscu można systemy takie, jak: SYSTRAN, LOGOS, ALPS, ENGSPAN, SPANAM, METAL, GLOBALINK, WEIDNER, TAUM, GETA, KANT, PS i inne [15].

Obecnie obserwuje się znaczne nasilenie zainteresowań dziedziną translacji automatycznej związane, między innymi, ze światową ekspansją sieci Internet. Już dzisiaj Internet jest praktycznie niewyczerpanym źródłem informacji na niemalże dowolny temat. Niestety, jedną z głównych przeszkód utrudniających pełne korzystanie z dobrodziejstw Internetu jest bariera językowa [16]. W związku z powyższym, realizowanych jest wiele projektów mających na celu wypracowanie narzędzi umożliwiających automatyczne tłumaczenie stron internetowych na ojczysty język ich użytkownika.

Szczególnie duże zainteresowanie translacją automatyczną można zaobserwować w Japonii [17], w Stanach Zjednoczonych [18], w krajach Unii Europejskiej [19], zwłaszcza w Wielkiej Brytanii [20] i Hiszpanii [21], a także w Indiach [22]. Nic w tym zresztą dziwnego, ponieważ są to kraje, w których, ze względu na ich wielonarodową sytuację etniczną lub pozycję gospodarczą i polityczną oraz związane z tym kontakty z innymi państwami świata, zapotrzebowanie na wszelkiego rodzaju tłumaczenia jest największe [23].

Jednakże, pomimo ponad pół wieku badań prowadzonych w dziedzinie translacji automatycznej, w pełni automatyczny przekład dla dowolnych tekstów należących do

nieograniczonego żadnymi sztucznymi więzami języka, którego jakość dorównywałaby efektem pracy tłumaczy profesjonalistów, wciąż pozostaje niedoścignionym ideałem – można wręcz powiedzieć św. Graalem poszukiwań badawczych [11]. Pytanie, czy kiedykolwiek zamierzony cel uda się osiągnąć, wciąż pozostaje problemem otwartym. Niniejsza problematyka wiąże się bezpośrednio z dociekaniami ogólniejszej natury na temat możliwości i ograniczeń systemów sztucznej inteligencji oraz z pytaniem, czy można w pełni poznać tajniki działania ludzkiego mózgu oraz czy można w pełni symulować funkcjonowanie ludzkiego umysłu za pomocą komputera [24, 25, 26, 27, 28, 29, 30, 31, 32].

Istnieje wiele przyczyn, z których powodu translacja z wielkim trudem i z wielkimi oporami poddaje się automatyzacji i implementacji w postaci programu komputerowego. Podstawową przyczyną jest fakt polegający na tym, że przekład pomiędzy językami naturalnymi jest rodzajem działalności wymagającym od biorącego w nim udział człowieka dużej pomysłowości i inwencji twórczej. Bezpośrednią przyczyną tego faktu jest ścisły związek języka z cywilizacją oraz z realiami lokalnymi, w których żyje dana społeczność, przez co pewne treści nie dają się przetłumaczyć w sposób bezpośredni [33, 34]. Interesujący jest również fakt, że członkowie różnych społeczności językowych postawieni w obliczu identycznej sytuacji, częstokroć w swych wypowiedziach zwracają uwagę na zupełnie odmienne elementy dotyczące tej sytuacji [35].

Nie bez znaczenia są również różnice zachodzące pomiędzy systemami gramatycznymi różnych języków, różnice w leksyce poszczególnych języków oraz idiomatyczny sposób przekazywania niektórych treści. Wiele trudności sprawia badaczom związanym z dziedziną translacji automatycznej również wieloznaczność wypowiedzi formułowanych w języku naturalnym, która może objawić się aż na trzech różnych poziomach analizy wypowiedzi [11]:

- 1) leksykalnym,
- 2) syntaktycznym,
- 3) semantycznym.

Autor podjął próbę usystematyzowania zjawisk językowych, które mogą stanowić potencjalną przeszkodę w zautomatyzowaniu translacji, czego wyrazem są między innymi jego prace [36, 37, 38, 39, 40, 41, 42, 43, 44, 45].

Oprócz w pełni automatycznych systemów translacji FAMT (*Fully-Automatic Machine Translation*) spotyka się również systemy translacji automatycznej wspomagane przez człowieka HAMT (*Human-Aided Machine Translation*) oraz narzędzia informatyczne wspomagające pracę człowieka tłumacza MAHT (*Machine-Aided Human Translation*).

Artykuł niniejszy poświęcony jest w całości opracowanej przez autora metodzie translacji automatycznej opartej na tzw. wzorcach translacyjnych PBMT (*Pattern-Based Machine Translation*) oraz opisowi prób zastosowania tej metody w systemach dokonujących automatycznych przekładów z wybranych języków świata na język polski.

W przekonaniu autora zaproponowana przez niego metoda może stanowić istotny krok na drodze prowadzącej do skonstruowania w pełni automatycznych systemów translacji, dostarczających przekładów o względnie wysokiej jakości, dla tekstów należących do języka ogólnego.

Zdaniem autora nie pozbawiony znaczenia jest również fakt, że praca niniejsza jest jedną z nielicznych w Polsce publikacji dotyczących bezpośrednio dziedziny translacji automatycznej. Jedyna polskojęzyczna monografia z rozważanej dziedziny [12], w świetle najnowszych osiągnięć informatyki, jest już całkowicie przestarzała, a ponadto jej autor zdecydowanie zbyt wiele miejsca poświęcił rozważaniom, w gruncie rzeczy jałowym, na temat tego co będzie, a co nie będzie możliwe w przyszłości. Przy czym rozważań tych jest stanowczo zbyt wiele w stosunku do rzetelnego opisu istniejących ówczesnie metod translacji automatycznej i możliwych do osiągnięcia za ich pomocą rezultatów.

W świetle powyższych uwag widać, że na polskim rynku wydawniczym istnieje swoista luka, gdyż brak jest w miarę nowej pozycji książkowej traktującej o niezwykle pasjonującej i interdyscyplinarnej dziedzinie badań, jaką jawi się translacja automatyczna. Brak ten jest tym bardziej dokuczliwy, że w ostatnich latach badania w dziedzinie translacji automatycznej nabrały wyraźnego przyspieszenia, spowodowanego niezwykle dynamicznym rozwojem samej informatyki, jak również opracowaniem zupełnie nowych metod translacji, konkurencyjnych do uprzednio istniejących. W przekonaniu autora napisany przez niego cykl artykułów będzie w stanie tę lukę, przynajmniej częściowo, zapełnić.

## 2. PRZEGLĄD METOD TRANSLACJI AUTOMATYCZNEJ

Amerykański teoretyk informacji Warren Weaver w swoim słynnym memorandum postulował ukierunkowanie badań na zastosowanie w translacji automatycznej technik kryptograficznych. Uczony ten twierdził wręcz, że każdy tekst zapisany w języku rosyjskim jest *de facto* tekstem zapisanym w języku angielskim, z tym tylko zastrzeżeniem, że można go potraktować, jako tekst zaszyfrowany w pewien szczególny sposób. Translacja polegałaby zatem na odkryciu sposobu szyfrowania tekstu, następnie dokonaniu jego deszyfracji, po czym rozważany tekst mógłby już zostać odczytany bezpośrednio w języku angielskim [11].

Jednak już pierwsze przeprowadzone próby dowiodły, że proces automatyzacji translacji jest zadaniem o wiele trudniejszym, ale zarazem o wiele ciekawszym z naukowego punktu widzenia, niż pierwotnie sądzono.

Działanie pierwszych systemów translacji automatycznej polegało na stosowaniu tzw. bezpośredniej metody przekładu (*direct method*). W metodzie tej zdania tekstu źródłowego tłumaczone były po prostu wyraz po wyrazie. Nic zatem dziwnego, że jakość uzyskiwanych tą drogą



przekładów była rażąco niska. Później wprowadzono kolejne udoskonalenie, polegające na wprowadzeniu drugiej fazy translacji – tzw. fazy uporządkowania (*rearrangement phase*), w której, stosując odpowiednie reguły gramatyki języka docelowego przekładu, starano się uporządkować kolejność wyrazów w uzyskanym z komputera tekście oraz odpowiednio dopasować ich formy fleksyjne [15]. W ten sposób wykrystalizowały się pierwsze systemy translacji automatycznej oparte na regułach RBMT (*Rule-Based Machine Translation*).

W systemach translacji automatycznej opartych na regułach można wyróżnić dwa odrębne podejścia:

- 1) metodę transferu struktur syntaktycznych TBMT (*Transfer-Based Machine Translation*),
- 2) metodę KBMT (*Knowledge-Based Machine Translation*) wykorzystującą wiedzę o językach źródłowym i docelowym przekładu oraz wiedzę o dziedzinie, której tłumaczone teksty dotyczą.

Systemy implementujące metodę TBMT działają trójetażowo.

W etapie pierwszym dokonywana jest analiza syntaktyczna i semantyczna wypowiedzi należącej do języka źródłowego, a następnie wypowiedź ta jest przekształcana do pewnej formy reprezentacji charakterystycznej dla języka źródłowego. W etapie drugim dokonywane jest przejście, czyli transfer – stąd nazwa metody – z reprezentacji właściwej językowi źródłowemu do reprezentacji właściwej językowi docelowemu. W etapie trzecim następuje wygenerowanie wypowiedzi w języku docelowym na podstawie reprezentacji charakterystycznej dla języka docelowego [11].

W przeciwieństwie do systemów TBMT, działanie systemów KBMT sprowadza się do dwóch etapów [15].

W etapie pierwszym, zwanym etapem analizy, dokonywana jest analiza syntaktyczna wypowiedzi sformułowanych w języku źródłowym oraz bardzo szczegółowa analiza semantyczna tych wypowiedzi, przy czym system translacji automatycznej korzysta tutaj z dziedzicznych baz wiedzy. W wyniku tego tworzona jest reprezentacja wypowiedzi języka źródłowego, która jest całkowicie od tego języka niezależna. W etapie drugim z takiej niezależnej od języka reprezentacji wypowiedzi generowana jest wypowiedź w języku docelowym przekładu. Faza ta nazywana jest fazą syntezy.

Dużą zaletą systemów KBMT jest ich stosunkowo łatwa możliwość poszerzania o dalsze języki. Po prostu wystarczy dla każdego z nowo dodanych języków opracować odpowiedni moduł analizy i syntezy, aby można było tłumaczyć z rozważanego języka na dowolny inny język uwzględniony dotychczas przez system, jak również tłumaczyć z dowolnego języka ujętego w ramach systemu na nowo dodany język [11, 46].

Jeśli w systemie KBMT reprezentacja wypowiedzi ma formę linearną, przypominającą zapis w pewnym języku naturalnym, język taki nazywa się językiem *interlingua*, czyli językiem pośredniczącym, a sam system oznaczany jest skrótem IBMT (*Interlingua-Based Machine Translation*) [46].

Obecnie systemy TBMT oraz KBMT są powszechnie stosowane, przy czym podejście TBMT znajduje zastosowanie w przypadku systemów dokonujących tłumaczeń tekstów należących do języka ogólnego, natomiast systemy KBMT są stosowane do tłumaczenia tekstów dotyczących pewnej szczególnej dziedziny działalności technicznej bądź naukowej człowieka [11]. Metodę TBMT zastosowano również w dostępnym komercyjnie systemie „English Translator” dokonującym automatycznych przekładów pomiędzy językami polskim i angielskim.

Na przełomie lat 80. i 90. XX w. pojawiły się w translacji automatycznej alternatywne podejścia, zrywające całkowicie z tradycją uprzednio znanych metod. Wymienić można w tym miejscu systemy translacji automatycznej, których działanie oparto wyłącznie na metodach statystycznych SBMT (*Statistically-Based Machine Translation*) oraz systemy translacji automatycznej oparte na przykładach, które w literaturze oznaczane są skrótem CBMT (*Case-Based Machine Translation*) lub EBMT (*Example-Based Machine Translation*). Oba z wymienionych podejść charakteryzują się tym, że podczas translacji nie jest dokonywana żadna analiza syntaktyczna ani semantyczna tłumaczonych tekstów [11].

W systemach SBMT tekst zapisany w języku naturalnym traktowany jest jako ciąg pewnych symboli, których częstotliwość pojawiania się oraz relacje wzajemnego umiejscowienia w tekście podlegają prawom rozkładu statystycznego [47].

W ostatnich kilku latach dużą popularność zdobywa metoda EBMT, która staje się coraz poważniejszą alternatywą dla innych znanych metod translacji automatycznej [48]. System translacji automatycznej oparty na metodzie EBMT dysponuje bazą danych, w której znajduje się bilingwiczny tekst o dużej objętości, zwykle co najmniej kilku milionów wyrazów. W rozważanym bilingwicznym tekście poszczególne zdania są uporządkowane w ten sposób, że każdemu zdaniu w języku źródłowym odpowiada dokładnie jedno zdanie w języku docelowym. Zawartość takiej bazy danych nazywa się korpusem równoległym (*parallel corpus*). System EBMT dokonuje tłumaczeń w ten sposób, że stara się odnaleźć w części korpusu równoległego zapisanej w języku źródłowym fragmenty zdań, które mają zostać przetłumaczone. Następnie w części korpusu równoległego zapisanej w języku docelowym, system stara się odnaleźć przekłady odnalezionych uprzednio fragmentów tłumaczonych zdań. Po odnalezieniu takich fragmentów zapisanych w języku docelowym są one podstawiane w miejsce fragmentów zdań w języku źródłowym. Zatem przekład powstaje poprzez proste wyszukiwanie fragmentów zdań w bazie danych i podstawianie w ich miejsce ich ekwiwalentów w języku docelowym [11]. Często systemy EBMT wyposażane są w dodatkowe moduły, których celem jest wzajemne dopasowywanie odnalezionych fragmentów, uzgadnianie ich form fleksyjnych itp.

Pewnym minusem systemów EBMT jest konieczność posiadania korpusu równoległego o dużych rozmiarach. Niestety często akwizycja takiego bardzo licznego zbioru dopasowanych do siebie zdań języka źródłowego i docelo-

wego nie jest sprawą łatwą. Dążąc do wydatnego zmniejszenia rozmiarów korpusu równoległego opracowano modyfikację metody EBMT, którą nazwano metodą opartą na uogólnionych wzorcach GEBMT (*Generalized Example-Based Machine Translation*). System GEBMT pracuje w sposób bardzo podobny do systemu EBMT, z tą tylko różnicą, że na samym początku występuje dodatkowa faza, polegająca na oznaczaniu specjalnymi znacznikami (*tokens*) wybranych elementów tekstu (*tokenization phase*). W fazie tej wybranym wyrazom zostają przypisane znaczniki, które określają pewną szerszą kategorię, do której dane wyrazy można zaliczyć. Dzięki temu możliwe jest późniejsze uogólnienie przykładów translacyjnych, polegające na tym, że przykład uogólniony może zawierać wzorce tłumaczeń obejmujące setki lub nawet tysiące przykładów zwykłych. Jak dowodzą badania, zastosowanie metody GEBMT pozwala na zmniejszenie rozmiarów potrzebnego korpusu równoległego nawet o rząd wielkości [49, 50, 51, 52].

W tym miejscu warto również wspomnieć o podejściach do translacji automatycznej, które stanowią hybrydę wymienionych uprzednio metod. W systemach hybrydowych translacja oparta na przykładach może być na przykład połączona z metodami opartymi na transferze struktur syntaktycznych lub z metodami statystycznymi. Jak sądzą niektórzy badacze, przyszłość translacji automatycznej należy będzie właśnie do tego typu systemów, stanowiących hybrydę różnych metod [11, 46].

Do nieco innej kategorii, aczkolwiek blisko związanej z translacją automatyczną opartą na przykładach, zalicza się technikę tzw. pamięci translacyjnej TM (*Translation Memory*). Technika pamięci translacyjnej jest chętnie stosowana przez tłumaczy w przypadku dokonywania przekładów zmodyfikowanych wersji tekstów, które już kiedyś zostały przez nich przetłumaczone i zostały zapisane w bazie danych, w taki sposób, że każdemu zdaniu w języku źródłowym odpowiada dokładnie jedno zdanie w języku docelowym [11]. Podczas tłumaczenia nowej wersji dokumentu komputer stara się w nim odnaleźć te zdania, które były już w wersji poprzedniej, a następnie podstawia w ich miejsce ich odpowiedniki zapisane w języku docelowym przekładu. Jak uczy doświadczenie, podejście takie może przyczynić się nawet do kilkukrotnego skrócenia czasu poświęconego na dokonanie przekładu nowej wersji dokumentu [53].

Pomiędzy wymienionymi uprzednio technikami translacji automatycznej a techniką pamięci translacyjnej uplasować można jeszcze inną kategorię systemów, którą stanowią tzw. leksykony związków frazeologicznych PL (*Phrasal Lexicons*). Główne idee związane z leksykonami związków frazeologicznych zostały zaprezentowane już w 1975 roku przez amerykańskiego lingwistę Josepha Beckera [54].

Becker w swoim referacie poczynił słuszną uwagę, że lingwiści, podobnie zresztą jak większość badaczy związanych z innymi dyscyplinami naukowymi, chcieliby najchętniej być fizykami i posiadać wdzięczny obiekt badań, który zawsze zachowuje się w sposób deterministyczny, i w związku z tym, jest możliwe odkrycie ścisłych praw nim rządzących [54]. Niestety język naturalny takim obiektem

badania nie jest [35]. W związku z powyższym Becker zaproponował podejście całkowicie odmienne od obowiązujących do tej pory trendów w lingwistyce. Zamiast uważać proces produkcji językowej jako polegający na dokonywaniu zestawień słów bądź morfemów, które dopiero tworzą wypowiedzi, Becker zaproponował coś zgoła odmiennego. Uznał mianowicie, że dopiero związki frazeologiczne, zbudowane co najmniej z dwóch wyrazów stanowią prawdziwy budulec wypowiedzi językowych [54].

Współczesne badania zdają się potwierdzać hipotezę Beckera. Okazuje się bowiem, że ludzie budują swoje wypowiedzi głównie przez powtarzanie, modyfikowanie i łączenie uprzednio poznanych związków frazeologicznych, które są na bieżąco adaptowane do nowych sytuacji [54]. Takie związki frazeologiczne nazywane są stereotypami i występują w każdym języku naturalnym w wielkiej obfitości [34]. Jak dowodzi Becker proces działalności językowej polega co najmniej w równym stopniu na pozyskiwaniu z pamięci gotowych stereotypów, z których budowana jest wypowiedź, jak na generacji całkowicie nowych wypowiedzi. Dzieje się tak dlatego, że wiele sytuacji nie wymaga twórczej działalności językowej, ponieważ daje się zaliczyć do jednej z kategorii znanych osobie generującej wypowiedź [54]. Zamiast tworzyć coraz to nowsze warianty wypowiedzi, ludzie najchętniej operują gotowymi formułami, banałami, idiomami, aluzjami, sloganami i innymi podobnymi wzorcami szablonowymi [54].

Jako przykład występowania w języku tego typu skostniałych związków frazeologicznych zostanie rozważone zdanie zaczerpnięte z przedmowy do książki [55]. Zdanie to ma postać:

*Mając na względzie dotkliwy brak na polskim rynku wydawniczym materiałów potrzebnych do praktycznego opanowania języka rumuńskiego, uważamy, że nasze rozmówki okażą się pomocne dla wszystkich zainteresowanych.*

W rozważanym zdaniu można wyróżnić następujące szablonowe związki frazeologiczne: „mając na względzie”, „dotkliwy brak”, „polski rynek wydawniczy”, „materiały potrzebne do”, „praktyczne opanowanie języka”, „okazać się pomocnym” i „wszyscy zainteresowani”. Budowa wypowiedzi zawartej w rozważanym zdaniu polega na odpowiednim „posklejaniu” wymienionych związków frazeologicznych oraz na użyciu ich w odpowiednich formach fleksyjnych.

System translacji automatycznej, którego działanie oparte zostało na leksykonie związków frazeologicznych, zwanych dalej po prostu frazami, korzysta z dwujęzycznego słownika fraz, w którym każdej frazie języka źródłowego odpowiada dokładnie jedna fraza w języku docelowym przekładu. Sposób działania systemu translacji automatycznej opartego na leksykonie fraz, bardzo przypomina pracę modułu sprawdzania poprawności pisowni w edytorze tekstowym takim jak np. popularny „MS Word”. Po prostu system odnajduje w leksykonie kolejne frazy, z których zbudowany jest tłumaczony tekst i podstawia w ich miejsce ich odpowiedniki w języku docelowym.

Spektakularny przykład implementacji takiego systemu pochodzi z Hiszpanii, gdzie bilingwiczny leksykon fraz zo-

stał utworzony dla języków hiszpańskiego i katalońskiego [56]. Jakość pracy rozważanego systemu translacji automatycznej jest na tyle wysoka, że został on z powodzeniem zastosowany do mechanicznego tłumaczenia z języka hiszpańskiego na kataloński czasopisma „El Periódico de Catalunya”, które następnie już bez żadnych poprawek kierowane jest do druku w nakładzie 60 000 egzemplarzy. Szybkość pracy rozważanego systemu wynosi około 7 KB tekstu na sekundę. Jakość przekładu jest wysoka, głównie dzięki temu, iż system jest ciągle udoskonalany, w sposób polegający na tym, że sztab lingwistów systematycznie uzupełnia bilingwiczny leksykon fraz, wprowadzając doń nowe związki frazeologiczne o maksymalnej długości sześciu wyrazów [56]. Jest to prawdopodobnie do chwili obecnej najbardziej spektakularny przykład wykorzystania systemu translacji automatycznej do tłumaczenia tekstów należących do języka ogólnego, które zakończyło się tak dużym sukcesem. Niestety tajemnica tego sukcesu polega głównie na bardzo bliskim pokrewieństwie genetycznym zachodzącym pomiędzy językami źródłowym i docelowym przekładu. Należy oczekiwać, że w przypadku zastosowania rozważanej metody do automatycznego tłumaczenia tekstów pomiędzy językami bardzo od siebie odległymi w kategoriach klasyfikacji genetycznej, typologicznej i geograficznej, uzyskane rezultaty nie będą się odznaczały już tak wysoką jakością. Jednakże ostateczną odpowiedź może dać tutaj tylko zbudowanie działającego systemu translacji automatycznej opartej na bilingwicznym leksykonie fraz oraz wykonanie odpowiedniej liczby testów.

Zgodnie z tym, co wiadome jest autorowi, metoda translacji automatycznej wykorzystująca bilingwiczny leksykon fraz nie została nigdy zastosowana dla języka polskiego. Autor podjął próbę zastosowania rozważanej metody do dokonywania mechanicznych przekładów z wybranych języków świata na język polski. Poczynione podczas takich prób liczne przemyślenia i wyciągnięte na ich podstawie wnioski doprowadziły ostatecznie autora do sformułowania własnej modyfikacji rozważanej metody translacji automatycznej, którą autor nazwał translacją automatyczną opartą na wzorcach PBMT. Zostanie ona omówiona szczegółowo w kolejnym punkcie.

### 3. TRANSLACJA AUTOMATYCZNA OPARTA NA WZORCACH

Podstawową zaletą techniki translacji automatycznej opartej na leksykonie fraz jest to, iż zapisane w nim frazy w sposób niejako automatyczny zawierają znaczną porcję informacji kontekstowej. Zdaniem autora jest to najważniejsza i najistotniejsza cecha metody translacji opartej na bilingwicznym leksykonie fraz, która jednocześnie decyduje o jej przewadze nad pozostałymi metodami, w których tłumaczenie odbywa się wyraz po wyrazie. Okazuje się bowiem, że dysponując jedynie pojedynczym wyrazem, wyrwanym z kontekstu wypowiedzi, bardzo trudno w sposób jednoznaczny, ustalić jego odpowiednik znaczeniowy

w języku przekładu. Odpowiedzialne jest za to głównie zjawisko wieloznaczności języka naturalnego, objawiające się na poziomie leksykalnym wypowiedzi. Na przykład angielskie słowo **case** zgodnie ze słownikiem [57], rozpatrywane wyłącznie w funkcji rzeczownika, może mieć aż 28 różnych polskich odpowiedników:

- |                                      |                       |
|--------------------------------------|-----------------------|
| 1) wypadek,                          | 14) szkatułka,        |
| 2) przypadek chorobowy,              | 15) futerał,          |
| 3) sprawa sądowa,                    | 16) etui,             |
| 4) proces sądowy,                    | 17) puzdro,           |
| 5) argumenty przemawiające za czymś, | 18) kasetka,          |
| 6) przypadek gramatyczny,            | 19) kasetka,          |
| 7) stan,                             | 20) koperta zegarka,  |
| 8) sytuacja,                         | 21) okno inspektowe,  |
| 9) warunki,                          | 22) gablotka,         |
| 10) skrzynia,                        | 23) oszklona szafa,   |
| 11) skrzynka,                        | 24) torebka nasienna, |
| 12) paka,                            | 25) neseser,          |
| 13) pudło,                           | 26) pokrowiec,        |
|                                      | 27) papierośnica,     |
|                                      | 28) cygarnica)        |

Jest rzeczą całkowicie oczywistą, że dokonanie przez system translacji automatycznej niewłaściwego wyboru ekwiwalentu znaczeniowego tłumaczonego słowa **case** skutkuje najprawdopodobniej kompletnym niezrozumieniem przekładu przez jego odbiorcę. Cóż bowiem ma wspólnego „wypadek” ze „skrzynią”, „gablotką” czy też z „papierośnicą”?

Zjawisko wieloznaczności leksykalnej nie jest jakąś szczególną cechą, charakterystyczną tylko i wyłącznie dla języka angielskiego. Wręcz przeciwnie, jest ono wszechobecne w każdym języku naturalnym. W istocie, otwierając jakikolwiek dwujęzyczny słownik, np. francusko-polski, trudno znaleźć wyrazy, które posiadałyby tylko jeden odpowiednik znaczeniowy w języku docelowym.

Poniżej podano kilka przykładów zjawiska wieloznaczności leksykalnej, występującego w wybranych językach świata:

- francuski – słowo **perle**: 1. perła, 2. paciorek, 3. kapsułka;
- hiszpański – słowo **fondo**: 1. dno, 2. głębia, 3. tło;
- portugalski – słowo **filète**: 1. niteczka, 2. włókienka, 3. paseczek, 4. kreska, 5. listewka, 6. filet;
- włoski – słowo **stufa**: 1. piec, 2. cieplarnia;
- niemiecki – słowo **Absatz**: 1. ustęp, 2. obcas, 3. osad, 4. złożę, 5. osadzenie, 6. zbyt;
- angielski – słowo **butt**: 1. beczka, 2. pień, 3. pniak, 4. grubszy koniec, 5. kolba karabinu, 6. płastuga, 7. nasypanie za strzelnicą, 8. pośmiewisko, 9. uderzenie głową;
- niderlandzki – słowo **boodschap**: 1. poselstwo, 2. polecenie, 3. wiadomość, 4. zakupy;
- szwedzki – słowo **tomten**: 1. parcela, 2. plac, 3. krasnoludek;
- norweski – słowo **hytte**: 1. chata, 2. szałas, 3. buda, 4. huta, 5. kabina;
- duński – słowo **løber**: 1. biegacz, 2. dywanik;



- fiński – słowo **kanta**: 1. podstawa, 2. obcas, 3. stanowisko, 4. baza;
- węgierski – słowo **kert**: 1. ogród, 2. sad;
- grecki – słowo **σκαπός**: 1. zamiar, 2. melodia, 3. wartownik;
- perski – słowo **استاد**: 1. profesor, 2. mistrz;
- arabski – słowo **وصل**: 1. połączenie, 2. łączy, 3. kontakt, 4. związek, 5. zawias, 6. dodatek;
- hebrajski – słowo **ציאה**: 1. wyjście, 2. odjazd;
- turecki – słowo **öğür**: 1. rówieśnik, 2. klasa, 3. warstwa, 4. sfera.

Jak wynika z zamieszczonego zestawienia, prawdopodobieństwo losowego trafienia przez system translacji automatycznej we właściwy ekwiwalent znaczeniowy tłumaczonego słowa jest niewielkie. Natomiast podstawowa zaleta metody translacji opartej na bilingwicznym leksykonie fraz polega właśnie na tym, że nie są tłumaczone pojedyncze wyrazy, ale ich całe grupy wyrazów. Z kolei jak wykazały badania [13], w celu ustalenia właściwego odpowiednika tłumaczonego słowa wystarczy w zdecydowanej większości wypadków dysponować jego kontekstem, na który składa się wyłącznie jedno słowo poprzedzające tłumaczony wyraz lub jedno słowo bezpośrednio po tym wyrazie następujące. W sporadycznych przypadkach potrzebne są dwa takie słowa. Spostrzeżenie to wyjaśnia, dlaczego metoda translacji automatycznej oparta na bilingwicznym leksykonie fraz odznacza się tak dużą skutecznością.

Istotę metody translacji automatycznej opartej na bilingwicznym leksykonie fraz najlepiej będzie można zrozumieć z następującego przykładu.

Niech będzie dany zatem pewien bilingwiczny tekst, na przykład taki, jak zamieszczony poniżej:

*Upon my return to the United States a few months ago, **after the extraordinary series of adventures** in the South Seas and elsewhere, of which an account is given in the following pages, accident threw me into the society of several gentlemen in Richmond, Va. , who felt deep interest in all matters relating to the region and who were constantly urging it upon me, as a duty, to give my narrative to the public.*

*Kilka miesięcy temu, gdy wróciłem do Stanów Zjednoczonych **po wielu nadzwyczajnych przygodach** na morzach południowych i w innych okolicach, o czym będę mówił dalej, przypadkowo znalazłem się w Richmond, stan Wirginia, w towarzystwie kilku dżentelmenów, których ogromnie zainteresowały wszystkie szczegóły zwiedzanych przeze mnie okolic i którzy ustawicznie wmawiali mi, że obowiązkowo powinienem opublikować moje wspomnienia.*

Górny tekst zapisany w języku angielskim stanowi fragment powieści Edgara Allana Poe'a pod tytułem: *The Narrative of Arthur Gordon Pym of Nantucket*. Natomiast tekst w języku polskim stanowi przekład tekstu angielskiego w wykonaniu Stanisława Wyrzykowskiego zatytułowany *Opowieść Artura Gordona Pyma z Nantucket*.

Niech w tekście zapisanym w języku źródłowym wystąpi pewna fraza, na przykład taka, jak została wyróżniona

w tekście angielskim, czyli: „after the extraordinary series of adventures”. Ponadto, niech równocześnie w bazie danych zostanie zapisane polskie tłumaczenie tej frazy, czyli: „po wielu nadzwyczajnych przygodach”, wówczas istnieje bardzo wysokie prawdopodobieństwo (ale niestety nie całkowita pewność), że jeżeli w pewnym innym tłumaczonym mechanicznie tekście angielskim pojawi się jeszcze raz taka sama fraza, czyli „after the extraordinary series of adventures”, i dokonujący przekładu komputer podstawy w jej miejsce odczytaną z bazy danych polską frazę „po wielu nadzwyczajnych przygodach”, to mechaniczny przekład tekstu angielskiego otrzymany na drodze takich podstawień będzie przekładem poprawnym.

Dla przykładu zostanie rozważony następujący bilingwiczny tekst, który został zapisany w języku szwedzkim i polskim. Część szwedzką stanowi urywek pewnego krótkiego opowiadania zaczerpnięty z podręcznika [58], natomiast część polska jest tłumaczeniem tego tekstu dokonany przez autora niniejszej pracy.

<i>Jag är på semester på Gotland.</i>	<i>Jestem na wakacjach na Gotlandii.</i>
<i>I dag har jag varit och titat på den här stenen.</i>	<i>Dzisiaj poszedłem popatrzeć na ten oto kamień.</i>
<i>Den är mycket fin med många bilder, men det är inga runor på den.</i>	<i>Jest bardzo ładny z wieloma obrazkami, ale nie ma na nim napisów runicznych.</i>
<i>Här har varit fint väder hela tiden.</i>	<i>Tutaj <b>przez cały czas</b> była ładna pogoda.</i>
<i>Jag har solat och badat varje dag.</i>	<i>Opalałem i kąpałem się w każdy dzień.</i>
<i>I går var jag i Visby.</i>	<i>Wczoraj <b>byłem</b> w Visby.</i>
<i>Det är en mycket vacker stad.</i>	<i>Jest to bardzo piękne miasto.</i>
<i>Jag träffade Erik i en gammal kyrka.</i>	<i>Spotkałem Eryka w <b>starym kościele</b>.</i>
<i>Han besökte goda vänner här på Gotland.</i>	<i>On odwiedził swoich dobrych przyjaciół tutaj na Gotlandii.</i>
<i>Han åkte hem i går.</i>	<i>On wyjechał wczoraj do domu.</i>
<i>Jag ska åka härifrån på lördag.</i>	<i>Ja wyjadę stąd w sobotę.</i>
<i>Jag har beställt plats på båten för här är så många turister.</i>	<i>Zarezerwowałem miejsca na statku, ponieważ tutaj jest tak wielu turystów.</i>

Niech na podstawie rozważanego bilingwicznego tekstu, w bazie danych zostaną zapisane następujące pary fraz w języku szwedzkim i polskim (zostały one podkreślone w obydwu tekstach):

- i dag – dzisiaj,
- hela tiden – przez cały czas,
- var jag – byłem,
- i en gammal kyrka – w starym kościele.

Wówczas, na podstawie rekordów skonstruowanej w ten sposób bazy danych można dokonać przekładu następują-

cego szwedzkiego zdania, które nie wystąpiło wcześniej w rozważanym bilingwicznym tekście:

### I dag var jag i en gammal kyrka hela tiden

Jak widać, podstawiając z bazy danych w miejsce kolejnych fraz tłumaczonego mechanicznie szwedzkiego zdania ich polskie ekwiwalenty znaczeniowe, otrzymuje się zdanie będące wiernym przekładem szwedzkiego oryginału:

### Dzisiaj byłem w starym kościele przez cały czas

Poczynione przez autora próby bezpośredniego zastosowania metody translacji automatycznej opartej na bilingwicznym leksykonie fraz w systemach dokonujących przekładów z wybranych języków świata na język polski nie przyniosły oczekiwanych rezultatów. Podstawową trudnością, na jaką natrafiono, okazała się mnogość form fleksyjnych wyrazów języka polskiego. Zjawisko fleksyjności języka polskiego sprawia, że niezwykle trudnym zadaniem staje się zestawienie bilingwicznego leksykonu, w którym każdej frazie w języku źródłowym odpowiadałaby dokładnie jedna fraza w języku docelowym, w tym wypadku polskim.

Na przykład, biorąc pod uwagę frazę zaczerpniętą z języka niemieckiego „ein psychologischer Roman” można się zastanawiać, który właściwie z jej możliwych polskich odpowiedników należałoby umieścić w bilingwicznym leksykonie fraz: „powieść psychologiczna”, „powieści psychologicznej”, „powieść psychologiczną”, czy też „powieścią psychologiczną”? Dopiero zaczerpnięcie z tekstu języka źródłowego dłuższej frazy takiej, jak np. „Wir haben einen psychologischen Roman gelesen” doprowadziłoby do rozstrzygnięcia tego problemu. Niestety fraza należąca do języka źródłowego stała się znacznie dłuższa, co oznacza, że aby system translacji mógł działać sprawnie, bilingwiczny leksykon fraz musiałby zawierać znacznie większą liczbę długich fraz (co grozi oczywiście kombinatoryczną eksplozją liczby możliwych połączeń wyrazów), niż w przypadku, w którym frazy te posiadałyby umiarkowaną długość.

Próbując dostosować metodę translacji automatycznej opartej na bilingwicznym leksykonie fraz do tłumaczenia na język polski z wybranych języków świata, autor opracował jej pewną modyfikację, polegającą na wprowadzeniu pojęcia wzorca translacyjnego. W związku z powyższym opracowana przez autora metoda translacji automatycznej została nazwana metodą opartą na wzorcach translacyjnych i została oznaczona, przez analogię do oznaczeń innych podejść do translacji automatycznej, jako PBMT.

Wprowadzone przez autora wzorce translacyjne pełnią rolę uogólnionych bilingwicznych fraz, co oznacza, że zawierają one wszelkie potencjalnie możliwe tłumaczenia frazy zapisanej w języku źródłowym na język docelowy. Ponieważ w języku polskim występują takie zjawiska, jak deklinacja rzeczowników oraz koniugacja czasowników, autor wprowadził trzy typu wzorców translacyjnych:

- 1) wzorce typu deklinacyjnego,
- 2) wzorce typu koniugacyjnego,
- 3) wzorce nie podlegające fleksji.

Wprowadzone zostały ponadto trzy atrybuty takie, jak <przypadek>, <liczba> i <rodzaj>. Od aktualnej wartości tych atrybutów zależy, który z możliwych przekładów frazy w języku źródłowym zostanie wybrany spośród kilku różnych propozycji oferowanych przez wzorzec translacyjny.

## 4. WZORCE TRANSLACYJNE

Pierwszym typem wzorca translacyjnego wprowadzonym przez autora jest tzw. wzorzec translacyjny podlegający deklinacji. Za pomocą tego typu wzorca tłumaczone są przede wszystkim frazy rzeczownikowe, których polskie ekwiwalenty znaczeniowe odmieniają się przez przypadki. Wzorzec translacyjny typu deklinacyjnego, w najbardziej dogodny i zwarty sposób, może zostać przedstawiony w postaci tabelarycznej (tab. 1).

Tabela 1. Wzorzec translacyjny typu deklinacyjnego

N_SOURCE	<przypadek>
N_TARGET_1	1
N_TARGET_2	2
N_TARGET_3	3
N_TARGET_4	4
N_TARGET_5	5
N_TARGET_6	6
<liczba> = VAL_N; <rodzaj> = VAL_G; <przypadek> = VAL_C;	

Poszczególne pola wzorca translacyjnego podlegającego deklinacji mają następujące znaczenie. W polu N\_SOURCE zawarta jest fraza rzeczownikowa należąca do języka źródłowego, która ma zostać przetłumaczona przez system translacji automatycznej na język docelowy. Z kolei w polach N\_TARGET\_1, N\_TARGET\_2, N\_TARGET\_3, N\_TARGET\_4, N\_TARGET\_5 i N\_TARGET\_6 zawarte są przekłady rozważanej frazy źródłowej na język docelowy (polski), występujące odpowiednio w mianowniku, dopełniaczu, celowniku, bierniku, narzędniku i miejscowniku. Podczas dokonywania przekładu, wybór odpowiedniej polskiej frazy dokonywany jest na podstawie aktualnej wartości atrybutu <przypadek>. Jeśli atrybut <przypadek> ma wartość 1, wówczas w miejsce frazy języka źródłowego N\_SOURCE podstawiana jest polska fraza N\_TARGET\_1. Jeśli atrybut <przypadek> ma wartość 2, wówczas podstawiana jest fraza N\_TARGET\_2 itd.

Rozważany wzorzec translacyjny nie zawiera polskiej frazy odpowiadającej wołaczowi. Postąpiono tak celowo, za względu na specyfikę siódmego przypadku, który występuje wyłącznie w wyrażeniach wykrzyknikowych i tłumaczony jest, w związku z tym, za pomocą innego typu wzorca translacyjnego – tzw. wzorca nie podlegającego fleksji, który zostanie omówiony w dalszej części artykułu.



Wzorzec translacyjny podlegający deklinacji zawiera jeszcze jedno pole, w którym wyszczególnione jest, jakie wartości mają zostać przypisane atrybutom <liczba>, <rodzaj> i <przypadek>. Wartości te wynoszą odpowiednio VAL\_N, VAL\_G i VAL\_C. Symbol VAL\_N może przybierać wartości 1 lub 2, co odpowiada liczbie pojedynczej i mnogiej. Z kolei symbol VAL\_G może przyjmować wartości 1, 2 lub 3, co odpowiada rodzajowi męskiemu, żeńskiemu i nijakiemu. Natomiast symbol VAL\_C może przyjmować wartości 1, 2, 3, 4, 5 lub 6, co odpowiada mianownikowi, dopełniaczowi, celownikowi, biernikowi, narzędnikowi i miejscownikowi.

Rozważane pole wzorca translacyjnego jest opcjonalne, to znaczy w pewnych szczególnych przypadkach może w ogóle nie wystąpić, i w związku z tym, atrybutom <liczba>, <rodzaj> i <przypadek> nie zostaną przypisane żadne wartości. W polu tym mogą wystąpić również tylko niektóre atrybuty, np. <liczba> i <rodzaj>, i w takim wypadku tylko im zostaną przypisane odpowiednie wartości.

Drugim typem wzorca translacyjnego jest wzorzec translacyjny podlegający koniugacji. Za pomocą tego typu wzorca translacyjnego tłumaczone są głównie frazy czasownikowe. Wzorzec translacyjny typu koniugacyjnego może zostać przedstawiony w postaci tabeli (tab. 2).

Tabela 2. Wzorzec translacyjny typu koniugacyjnego

V_SOURCE	<liczba>	<rodzaj>
V_TARGET_1_1	1	1
V_TARGET_1_2	1	2
V_TARGET_1_3	1	3
V_TARGET_2_1	2	1
V_TARGET_2_2	2	2
V_TARGET_2_3	2	3
<liczba> = VAL_N; <rodzaj> = VAL_G; <przypadek> = VAL_C;		

Pole V\_SOURCE zawiera frazę należącą do języka źródłowego, która ma zostać przetłumaczona na język docelowy (polski). Z kolei w polach V\_TARGET\_1\_1, V\_TARGET\_1\_2, V\_TARGET\_1\_3, V\_TARGET\_2\_1, V\_TARGET\_2\_2 i V\_TARGET\_2\_3 zawarte są przekłady rozważanej frazy źródłowej na język docelowy, w zależności od aktualnych wartości atrybutów <liczba> i <rodzaj>. Na przykład fraza V\_TARGET\_1\_1 stanowi polski przekład frazy V\_SOURCE w sytuacji, gdy atrybut <liczba> przyjmuje wartość 1 i atrybut <rodzaj> również przyjmuje wartość 1, co odpowiada liczbie pojedynczej i rodzajowi męskiemu. Podobnie fraza V\_TARGET\_2\_2 stanowi polski przekład frazy V\_SOURCE w sytuacji, gdy atrybut <liczba> przyjmuje wartość 2 i atrybut <rodzaj> również przyjmuje wartość 2, co odpowiada liczbie mnogiej i rodzajowi żeńskiemu.

Ostatnie pole wzorca translacyjnego podlegającego koniugacji pełni taką samą rolę, jak ostatnie pole wzorca translacyjnego typu deklinacyjnego, to znaczy zawarta w nim jest informacja, jakie wartości mają zostać przypisane atrybutom <liczba>, <rodzaj> i <przypadek>. Podobnie jak w przypadku wzorca typu deklinacyjnego, pole to ma charakter opcjonalny, to znaczy w pewnym szczególnym przypadku może w ogóle nie wystąpić, albo też w polu tym wartości mogą zostać przypisane tylko niektórym atrybutom, np. tylko atrybutowi <przypadek>.

Ostatnim typem wzorca translacyjnego, wprowadzonym przez autora, jest wzorzec nie podlegający fleksji. Wzorzec ten został również ujęty w formie tabeli (tab. 3).

Tabela 3. Wzorzec translacyjny nie podlegający fleksji

SOURCE
TARGET
<liczba> = VAL_N; <rodzaj> = VAL_G; <przypadek> = VAL_C;

Pole SOURCE stanowi frazę należącą do języka źródłowego, która ma zostać przetłumaczona przez system translacji automatycznej na język polski. Natomiast pole TARGET stanowi polski przekład frazy zawartej w polu SOURCE.

Ostatnie z pól wzorca translacyjnego nie podlegającego fleksji ma charakter opcjonalny i zawiera informacje o tym, jakie wartości mają zostać przypisane atrybutom <liczba>, <rodzaj> i <przypadek>.

## 5. PRZEBIEG PROCESU TRANSLACJI

Sposób, w jaki system translacji automatycznej tłumaczy na język polski zadane na jego wejście zdania należące do języka źródłowego przekładu, zostanie zilustrowany na przykładzie translacji następującego zdania zapisanego w języku francuskim:

### La liste des noms géographiques et le tableau des verbes irréguliers forment un supplément au présent dictionnaire

Na początku atrybutowi <przypadek> przypisywana jest wartość 1. Jest to podyktowane faktem, polegającym na tym, że w języku polskim podmiot zdania zwykle występuje w mianowniku:

<przypadek> = 1;

Po wykonaniu tej czynności, system translacji automatycznej próbuje odnaleźć w bazie danych frazę stanowiącą całe zdanie:

### la liste des noms géographiques et le tableau des verbes irréguliers forment un supplément au présent dictionnaire

Jest to podyktowane tym, że niektóre zdania mogą od-  
dawać pewne treści w sposób idiomatyczny, w zwią-  
zku z czym, nie jest możliwe wydedukowanie ich znacze-  
nia z analizy ich części składowych. Zatem przekłady  
takich zdań muszą zostać zapamiętane w postaci odpowied-  
nych rekordów bazy danych, zawierającej wzorce trans-  
lacyjne.

W przypadku gdy system translacji automatycznej nie  
odnajdzie rozważanego zdania w bazie danych, wówczas  
odrzuca od rozważanego zdania jego ostatni wyraz i podej-  
muje kolejną próbę odnalezienia w bazie danych wzorców  
translacyjnych następującej frazy:

#### **la liste des noms géographiques et le tableau des verbes irréguliers forment un supplément au présent**

W przypadku kolejnej porażki odniesionej podczas prze-  
szukiwania bazy danych wzorców translacyjnych, system  
translacji automatycznej ponownie odrzuca ostatni wyraz  
rozważanej frazy, po czym usiłuje odnaleźć w bazie danych  
następującą frazę:

#### **la liste des noms géographiques et le tableau des verbes irréguliers forment un supplément au**

Opisane czynności powtarzane są wielokrotnie, aż do  
momentu, w którym system translacji automatycznej podej-  
muje próbę wyszukania w bazie danych następującej frazy:

#### **la liste des noms géographiques**

Rozważana operacja zostaje uwieńczona sukcesem, po-  
nieważ system translacji automatycznej odnajduje w bazie  
danych wzorzec translacyjny typu deklinacyjnego, zawie-  
rający poszukiwaną frazę:

<b>la liste des noms géographiques</b>	<przypadek>
lista nazw geograficznych	1
listy nazw geograficznych	2
liście nazw geograficznych	3
listę nazw geograficznych	4
listą nazw geograficznych	5
liście nazw geograficznych	6
<liczba> = 1; <rodzaj> = 2;	

Ponieważ na początku atrybutowi <przypadek> zosta-  
ła przypisana wartość 1, jako polski przekład francuskiej  
frazy „la liste des noms géographiques” zostanie wybrana  
faza „lista nazw geograficznych”. Ponadto atrybutom  
<liczba> i <rodzaj> zostaną przypisane odpowiednio war-  
tości 1 i 2, ponieważ fraza „lista nazw geograficznych” wy-  
stępuje w liczbie pojedynczej i jest rodzaju żeńskiego.  
Wcześniej atrybuty <liczba> i <rodzaj> posiadały wartości  
nieokreślone.

Następnie system translacji automatycznej podejmuje  
próbę odnalezienia w bazie danych frazy, stanowiącej część  
tłumaczonego zdania, która powstała po odrzuceniu frazy  
uprzednio odnalezionej:

#### **et le tableau des verbes irréguliers forment un supplé- ment au présent dictionnaire**

Niestety próba odnalezienia w bazie danych wzorców  
translacyjnych takiej frazy kończy się porażką, w związku  
z czym system translacji automatycznej odrzuca jej ostatni  
wyraz i ponawia próbę odszukania w bazie danych następu-  
jącej frazy:

#### **et le tableau des verbes irréguliers forment un supplé- ment au présent**

Również i ta próba kończy się porażką, w związku  
z czym cała procedura powtarzana jest aż do skutku. Osta-  
tecznie systemowi translacji automatycznej udaje się odna-  
leźć w bazie danych frazę:

et

Rozważaną frazę zawiera następujący wzorzec transla-  
cyjny nie podlegający fleksji:

et
i

Rozważany wzorzec nie przypisuje atrybutom <liczba>, <rodzaj> i <przypadek> żadnych wartości.

W kolejnym kroku system translacji automatycznej po-  
dejmuje próbę odnalezienia w bazie danych frazy powstałej  
z tłumaczonego zdania po odrzuceniu od niego fraz już  
uprzednio odnalezionych:

#### **le tableau des verbes irréguliers forment un supplément au présent dictionnaire**

Próba odnalezienia w bazie danych takiej frazy kończy  
się porażką, w związku z czym, system translacji automa-  
tycznej odrzuca ostatni wyraz i ponawia próbę odnalezienia  
frazy:

#### **le tableau des verbes irréguliers forment un supplément au présent**

Niestety, również tej frazy systemowi translacji auto-  
matycznej nie udaje się odnaleźć w bazie danych i rozważa-  
na procedura musi być kontynuowana aż do skutku, to zna-  
czy aż do momentu, gdy w bazie danych poszukiwana jest  
faza:

#### **le tableau des verbes irréguliers**

Rozważana fraza zostaje odnaleziona w bazie danych w postaci wzorca translacyjnego typu deklinacyjnego:

<b>le tableau des verbes irréguliers</b>	<przypadek>
tablica czasowników nieregularnych	1
tablicy czasowników nieregularnych	2
tablicy czasowników nieregularnych	3
tablicę czasowników nieregularnych	4
tablicą czasowników nieregularnych	5
tablicy czasowników nieregularnych	6
<liczba> = 1; <rodzaj> = 2;	

Ponieważ aktualna wartość atrybutu <przypadek> wynosi 1, jako przykład francuskiej frazy „le tableau des verbes irréguliers” zostaje przyjęta polska fraza „tablica czasowników nieregularnych”, po czym atrybutom <liczba> i <rodzaj> zostają odpowiednio przypisane wartości 1 i 2, ponieważ rozważana fraza występuje w liczbie pojedynczej i jest rodzaju żeńskiego.

Następnie system translacji automatycznej podejmuje próbę odnalezienia w bazie danych frazy powstałej z tłumaczonego zdania po odrzuceniu odnalezionych wcześniej fraz:

#### **forment un supplément au présent dictionnaire**

Próba odnalezienia w bazie danych takiej frazy kończy się porażką, w związku z powyższym, system translacji automatycznej odrzuca ostatni wyraz i próbuje odnaleźć w bazie danych frazę:

#### **forment un supplément au présent**

Również tej frazy systemowi translacji automatycznej nie udaje się odnaleźć w bazie danych, w związku z czym, rozważana procedura jest kontynuowana do skutku, czyli do momentu, w którym system podejmuje próbę odnalezienia w bazie danych frazy:

#### **forment**

Fraza ta zostaje odnaleziona w postaci wzorca translacyjnego typu koniugacyjnego:

<b>forment</b>	<liczba>	<rodzaj>
stanowi	1	1
stanowi	1	2
stanowi	1	3
stanowią	2	1
stanowią	2	2
stanowią	2	3
<przypadek> = 4;		

Ponieważ atrybut <liczba> ma wartość 1 i atrybut <rodzaj> ma wartość 2, jako przykład francuskiej frazy „forment” zostaje wybrana polska fraza „stanowi”. Ponadto atrybutowi <przypadek> zostaje przypisana wartość 4, ponieważ w języku polskim czasownik „stanowi” wymaga użycia po nim rzeczownika w bierniku, aby spełniony był związek rządu zachodzący pomiędzy orzeczeniem i dopełnieniem zdania.

W kroku kolejnym system translacji automatycznej podejmuje próbę odnalezienia w bazie danych następującej frazy:

#### **un supplément au présent dictionnaire**

Niestety próba odnalezienia w bazie danych takiej frazy kończy się niepowodzeniem, po czym system translacji automatycznej odrzuca ostatni wyraz i ponawia próbę odnalezienia w bazie danych frazy:

#### **un supplément au présent**

Również tej frazy systemowi translacji automatycznej nie udaje się odnaleźć w bazie danych, w związku z czym, odrzucany jest jej ostatni wyraz i ponawiana jest próba odnalezienia frazy:

#### **un supplément au**

Próba ta kończy się sukcesem, ponieważ rozważana fraza zostaje odnaleziona w postaci wzorca translacyjnego typu deklinacyjnego:

<b>un supplément au</b>	<przypadek>
dodatek do	1
dodatku do	2
dodatkowi do	3
dodatek do	4
dodatkiem do	5
dodatku do	6
<liczba> = 1; <rodzaj> = 1; <przypadek> = 2;	

Ponieważ aktualna wartość atrybutu <przypadek> wynosi 4, jako tłumaczenie frazy „un supplément au” zostaje wybrana polska fraza „dodatek do”. Ponadto atrybutom <liczba>, <rodzaj> i <przypadek> zostają przypisane odpowiednio wartości 1, 1 i 4. Dzieje się tak dlatego, że fraza „dodatek do” występuje w liczbie pojedynczej, jest rodzaju męskiego oraz wymaga użycia następującego po niej rzeczownika w dopełniaczu (dodatek do kogo? czego?).

W kroku kolejnym system podejmuje próbę odnalezienia w bazie danych frazy powstałej po odrzuceniu od tłumaczonego zdania odnalezionych uprzednio fraz:

#### **présent dictionnaire**



Próba odnalezienia takiej frazy zostaje uwieńczona sukcesem, ponieważ baza danych zawiera wzorzec translacyjny typu deklinacyjnego odpowiadający rozważanej frazie:

<b>présent dictionnaire</b>	<przypadek>
niniejszy słownik	1
niniejszego słownika	2
niniejszemu słownikowi	3
niniejszy słownik	4
niniejszym słownikiem	5
niniejszym słowniku	6
<liczba> = 1; <rodzaj> = 1;	

Ponieważ aktualna wartość atrybutu <przypadek> wynosi 2, jako odpowiednik francuskiej frazy „présent dictionnaire” zostaje wybrana polska fraza „niniejszego słownika”. Ponadto obu atrybutom <liczba> i <rodzaj> zostaje przypisana wartość 1, ponieważ polska fraza „niniejszy słownik” występuje w liczbie pojedynczej i jest rodzaju męskiego.

Zestawiając kolejno polskie tłumaczenia francuskich fraz, otrzymuje się następujący przekład rozważanego francuskiego zdania:

#### **Lista nazw geograficznych i tablica czasowników nieregularnych stanowi dodatek do niniejszego słownika**

Proces wyszukiwania w bazie danych wzorców translacyjnych poszczególnych francuskich fraz, które kolejno podlegają tłumaczeniu przez system translacji automatycznej, można prześledzić na podstawie następującego zestawienia:

- 1) la liste des noms géographiques et le tableau des verbes irréguliers forment un supplément au présent dictionnaire
- 2) la liste des noms géographiques et le tableau des verbes irréguliers forment un supplément au présent
- 3) la liste des noms géographiques et le tableau des verbes irréguliers forment un supplément au
- 4) la liste des noms géographiques et le tableau des verbes irréguliers forment un supplément
- 5) la liste des noms géographiques et le tableau des verbes irréguliers forment un
- 6) la liste des noms géographiques et le tableau des verbes irréguliers forment
- 7) la liste des noms géographiques et le tableau des verbes irréguliers
- 8) la liste des noms géographiques et le tableau des verbes
- 9) la liste des noms géographiques et le tableau des
- 10) la liste des noms géographiques et le tableau
- 11) la liste des noms géographiques et le

12) la liste des noms géographiques et

#### **13) la liste des noms géographiques**

14) et le tableau des verbes irréguliers forment un supplément au présent dictionnaire

15) et le tableau des verbes irréguliers forment un supplément au présent

16) et le tableau des verbes irréguliers forment un supplément au

17) et le tableau des verbes irréguliers forment un supplément

18) et le tableau des verbes irréguliers forment un

19) et le tableau des verbes irréguliers forment

20) et le tableau des verbes irréguliers

21) et le tableau des verbes

22) et le tableau des

23) et le tableau

24) et le

#### **25) et**

26) le tableau des verbes irréguliers forment un supplément au présent dictionnaire

27) le tableau des verbes irréguliers forment un supplément au présent

28) le tableau des verbes irréguliers forment un supplément au

29) le tableau des verbes irréguliers forment un supplément

30) le tableau des verbes irréguliers forment un

31) le tableau des verbes irréguliers forment

#### **32) le tableau des verbes irréguliers**

33) forment un supplément au présent dictionnaire

34) forment un supplément au présent

35) forment un supplément au

36) forment un supplément

37) forment un

#### **38) forment**

39) supplément au présent dictionnaire

40) supplément au présent

#### **41) supplément au**

#### **42) présent dictionnaire**

Jak można łatwo zauważyć, w celu dokonania przekładu rozważanego francuskiego zdania system translacji automatycznej podejmował aż 42 próby odnalezienia w bazie danych wzorców translacyjnych fraz, z których składało się tłumaczone zdanie, przy czym sukcesem zakończyły się próby nr 13, 25, 32, 38, 41 i 42.

Poniżej zamieszczono kolejny przykład tłumaczenia na język polski, przy użyciu zaproponowanej przez autora metody opartej na wzorcach translacyjnych, następującego zdania zapisanego w języku angielskim:

**Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time**

Na początku nowego zdania atrybutowi przypadek przypisywana jest wartość jeden.

<przypadek> = 1;

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce:

– wzorec translacyjny typu deklinacyjnego:

Weaver's memorandum	<przypadek>
wezwanie Weavera	1
wezwania Weavera	2
wezwanii Weavera	3
wezwanie Weavera	4
wezwaniami Weavera	5
wezwanii Weavera	6
<liczba> = 1;	
<rodzaj> = 3;	

– wzorec translacyjny typu koniugacyjnego:

brought to the attention of a wide circle	<liczba>	<rodzaj>
zwrócił uwagę szerszych kręgów na	1	1
zwróciła uwagę szerszych kręgów na	1	2
zwróciło uwagę szerszych kręgów na	1	3
zwrócili uwagę szerszych kręgów na	2	1
zwróciły uwagę szerszych kręgów na	2	2
zwróciły uwagę szerszych kręgów na	2	3
<przypadek> = 4;		

– wzorec translacyjny typu deklinacyjnego:

the possibilities	<przypadek>
możliwości	1
możliwości	2
możliwościom	3
możliwości	4
możliwościami	5
możliwościach	6
<liczba> = 2;	
<rodzaj> = 3;	

– wzorec translacyjny nie podlegający fleksji:

of
<przypadek> = 2;

w tym wzorcu translacyjnym angielskie słowo „of” nie jest w ogóle tłumaczone, ponieważ nie posiada ono w języku polskim jakiegoś bezpośredniego odpowiednika, zamiast tego atrybutowi <przypadek> przypisana zostaje wartość 2, aby tłumaczenie frazy rzeczownikowej występującej po słowie „of” wystąpiło w dopełniaczu;

– wzorec translacyjny typu deklinacyjnego:

a new and exciting application	<przypadek>
nowe i ekscytujące zastosowanie	1
nowego i ekscytującego zastosowania	2
nowemu i ekscytującemu zastosowaniu	3
nowe i ekscytujące zastosowanie	4
nowym i ekscytującym zastosowaniem	5
nowym i ekscytującym zastosowaniu	6
<liczba> = 1;	
<rodzaj> = 3;	

– wzorec translacyjny nie podlegający fleksji:

of
<przypadek> = 2;

– wzorec translacyjny typu deklinacyjnego:

the computers	<przypadek>
komputery	1
komputerów	2
komputerom	3
komputery	4
komputerami	5
komputerach	6
<liczba> = 2;	
<rodzaj> = 3;	

– wzorec translacyjny typu koniugacyjnego:

whose potentialities	<liczba>	<rodzaj>
którego możliwości	1	1
której możliwości	1	2
którego możliwości	1	3
których możliwości	2	1
których możliwości	2	2
których możliwości	2	3
<liczba> = 2;		
<rodzaj> = 3;		

– wzorec translacyjny typu koniugacyjnego:

were being discovered and proclaimed	<liczba>	<rodzaj>
został odkryty i ogłoszony	1	1
została odkryta i ogłoszona	1	2
zostało odkryte i ogłoszone	1	3
zostali odkryci i ogłoszeni	2	1
zostały odkryte i ogłoszone	2	2
zostały odkryte i ogłoszone	2	3

– wzorzec translacyjnie nie podlegający fleksji:

<b>with</b>
z
<przypadek> = 5;

– wzorzec translacyjny typu deklinacyjnego:

<b>enthusiasm</b>	<przypadek>
entuzjizm	1
entuzjazmu	2
entuzjizmowi	3
entuzjizm	4
entuzjazmem	5
entuzjaźmie	6
<liczba> = 1;	
<rodzaj> = 1;	

– wzorzec translacyjnie nie podlegający fleksji:

<b>and</b>
i

– wzorzec translacyjny typu deklinacyjnego:

<b>optimism</b>	<przypadek>
optymizm	1
optymizmu	2
optymizmowi	3
optymizm	4
optymizmem	5
optymizmie	6
<liczba> = 1;	
<rodzaj> = 1;	

– wzorzec translacyjnie nie podlegający fleksji:

<b>at this time</b>
w owym czasie

W rezultacie pracy systemu translacji automatycznej otrzymano następujący polski przekład angielskiego zdania:

**Wezwanie Weavera zwróciło uwagę szerszych kręgów na możliwości nowego i ekscytującego zastosowania komputerów których możliwości zostały odkryte i ogłoszone z entuzjazmem i optymizmem w owym czasie.**

Proces wyszukiwania w bazie danych wzorców translacyjnych poszczególnych angielskich fraz, które kolejno podlegają tłumaczeniu przez system translacji automatycznej, można prześledzić na podstawie następującego zestawienia:

- 1) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time
- 2) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this
- 3) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at
- 4) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism
- 5) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and
- 6) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm
- 7) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with
- 8) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed
- 9) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered and
- 10) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being discovered
- 11) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were being
- 12) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities were
- 13) Weaver's memorandum brought to the attention of a wide circle the possibilities of a new and exciting application of the computers whose potentialities





- 57) brought to the attention of a wide circle the possibilities of
- 58) brought to the attention of a wide circle the possibilities
- 59) brought to the attention of a wide circle the
- 60) brought to the attention of a wide circle**
- 61) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time
- 62) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this
- 63) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at
- 64) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism
- 65) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and
- 66) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm
- 67) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with
- 68) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and proclaimed
- 69) the possibilities of a new and exciting application of the computers whose potentialities were being discovered and
- 70) the possibilities of a new and exciting application of the computers whose potentialities were being discovered
- 71) the possibilities of a new and exciting application of the computers whose potentialities were being
- 72) the possibilities of a new and exciting application of the computers whose potentialities were
- 73) the possibilities of a new and exciting application of the computers whose potentialities
- 74) the possibilities of a new and exciting application of the computers whose
- 75) the possibilities of a new and exciting application of the computers
- 76) the possibilities of a new and exciting application of the
- 77) the possibilities of a new and exciting application of
- 78) the possibilities of a new and exciting application
- 79) the possibilities of a new and exciting
- 80) the possibilities of a new and
- 81) the possibilities of a new
- 82) the possibilities of a
- 83) the possibilities of
- 84) the possibilities**
- 85) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time
- 86) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this
- 87) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at
- 88) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism
- 89) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and
- 90) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm
- 91) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed with
- 92) of a new and exciting application of the computers whose potentialities were being discovered and proclaimed
- 93) of a new and exciting application of the computers whose potentialities were being discovered and
- 94) of a new and exciting application of the computers whose potentialities were being discovered
- 95) of a new and exciting application of the computers whose potentialities were being
- 96) of a new and exciting application of the computers whose potentialities were
- 97) of a new and exciting application of the computers whose potentialities
- 98) of a new and exciting application of the computers whose
- 99) of a new and exciting application of the computers
- 100) of a new and exciting application of the
- 101) of a new and exciting application of
- 102) of a new and exciting application
- 103) of a new and exciting
- 104) of a new and
- 105) of a new
- 106) of a
- 107) of**
- 108) a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time
- 109) a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this
- 110) a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at
- 111) a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism

- 112) a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm and
- 113) a new and exciting application of the computers whose potentialities were being discovered and proclaimed with enthusiasm
- 114) a new and exciting application of the computers whose potentialities were being discovered and proclaimed with
- 115) a new and exciting application of the computers whose potentialities were being discovered and proclaimed
- 116) a new and exciting application of the computers whose potentialities were being discovered and
- 117) a new and exciting application of the computers whose potentialities were being discovered
- 118) a new and exciting application of the computers whose potentialities were being
- 119) a new and exciting application of the computers whose potentialities were
- 120) a new and exciting application of the computers whose potentialities
- 121) a new and exciting application of the computers whose
- 122) a new and exciting application of the computers
- 123) a new and exciting application of the
- 124) a new and exciting application of
- 125) a new and exciting application**
- 126) of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time
- 127) of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this
- 128) of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at
- 129) of the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism
- 130) of the computers whose potentialities were being discovered and proclaimed with enthusiasm and
- 131) of the computers whose potentialities were being discovered and proclaimed with enthusiasm
- 132) of the computers whose potentialities were being discovered and proclaimed with
- 133) of the computers whose potentialities were being discovered and proclaimed
- 134) of the computers whose potentialities were being discovered and
- 135) of the computers whose potentialities were being discovered
- 136) of the computers whose potentialities were being
- 137) of the computers whose potentialities were
- 138) of the computers whose potentialities
- 139) of the computers whose
- 140) of the computers
- 141) of the
- 142) of**
- 143) the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time
- 144) the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this
- 145) the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism at
- 146) the computers whose potentialities were being discovered and proclaimed with enthusiasm and optimism
- 147) the computers whose potentialities were being discovered and proclaimed with enthusiasm and
- 148) the computers whose potentialities were being discovered and proclaimed with enthusiasm
- 149) the computers whose potentialities were being discovered and proclaimed with
- 150) the computers whose potentialities were being discovered and proclaimed
- 151) the computers whose potentialities were being discovered and
- 152) the computers whose potentialities were being discovered
- 153) the computers whose potentialities were being
- 154) the computers whose potentialities were
- 155) the computers whose potentialities
- 156) the computers whose
- 157) the computers**
- 158) whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this time
- 159) whose potentialities were being discovered and proclaimed with enthusiasm and optimism at this
- 160) whose potentialities were being discovered and proclaimed with enthusiasm and optimism at
- 161) whose potentialities were being discovered and proclaimed with enthusiasm and optimism
- 162) whose potentialities were being discovered and proclaimed with enthusiasm and
- 163) whose potentialities were being discovered and proclaimed with enthusiasm
- 164) whose potentialities were being discovered and proclaimed with
- 165) whose potentialities were being discovered and proclaimed
- 166) whose potentialities were being discovered and
- 167) whose potentialities were being discovered
- 168) whose potentialities were being
- 169) whose potentialities were
- 170) whose potentialities**
- 171) were being discovered and proclaimed with enthusiasm and optimism at this time
- 172) were being discovered and proclaimed with enthusiasm and optimism at this



- 173) were being discovered and proclaimed with enthusiasm and optimism at  
 174) were being discovered and proclaimed with enthusiasm and optimism  
 175) were being discovered and proclaimed with enthusiasm and  
 176) were being discovered and proclaimed with enthusiasm  
 177) were being discovered and proclaimed with  
**178) were being discovered and proclaimed**  
 179) with enthusiasm and optimism at this time  
 180) with enthusiasm and optimism at this  
 181) with enthusiasm and optimism at  
 182) with enthusiasm and optimism  
 183) with enthusiasm and  
 184) with enthusiasm  
**185) with**  
 186) enthusiasm and optimism at this time  
 187) enthusiasm and optimism at this  
 188) enthusiasm and optimism at  
 189) enthusiasm and optimism  
 190) enthusiasm and  
**191) enthusiasm**  
 192) and optimism at this time  
 193) and optimism at this  
 194) and optimism at  
 195) and optimism  
 196) and  
 197) optimism at this time  
 198) optimism at this  
 199) optimism at  
**200) optimism**  
**201) at this time**

W celu dokonania przekładu rozważanego angielskiego zdania, system translacji automatycznej podjął aż 201 prób odnalezienia w bazie danych wzorców translacyjnych fraz, z których składało się tłumaczone zdanie, przy czym sukcesem zakończyły się próby nr 34, 60, 84, 107, 125, 142, 157, 170, 178, 185, 191, 196, 200, 201.

Bliższego wyjaśnienia wymaga jeszcze sytuacja, w której system translacji automatycznej, wykonując cykl poszukiwań wzorców w bazie danych, dochodzi do momentu, gdy poszukiwany jest już tylko ostatni wyraz frazy, pozostały po kolejnym odrzuceniu pozostałych wyrazów, i wyraz taki nie zostaje odnaleziony w bazie danych. W takiej sytuacji system translacji automatycznej przepisuje po prostu taki wyraz w tłumaczonym tekście bez żadnych zmian.

Dużą zaletą zaproponowanej przez autora metody automatycznej translacji opartej na wzorcach jest fakt polegający na tym, że metoda ta umożliwiła prawidłowe funkcjonowanie mechanizmów związanych z fleksyjnością języka polskiego. Wartości przypisywane podczas translacji atrybutom <liczba>, <rodzaj> i <przypadek> pozwalają na pra-

widłowy wybór form fleksyjnych podczas tłumaczenia kolejnych fraz odnalezionych w bazie danych, dzięki czemu między podmiotem a orzeczeniem zdania zachodzi związek zgody, a pomiędzy orzeczeniem a dopełnieniem spełniony jest związek rządu. Ponadto należy zwrócić uwagę fakt polegający na tym, iż w zdecydowanej większości języków świata podstawowym szykiem budowy zdania jest szyk typu SVO (*Subject Verb Object*). Jedyne nieliczne języki wykazują odmienny szyk budowy zdania, co jednak nie stanowi jakiegos zasadniczego problemu z zastosowaniem w ich przypadku metody automatycznej translacji zaproponowanej przez autora.

Na przykład język arabski posiada szyk budowy zdania typu VSO (*Verb Subject Object*). W takim przypadku grupy orzeczenia i podmiotu zdania są ujęte w ramach jednego wzorca translacyjnego. Można to zaobserwować na przykładzie translacji następującego arabskiego zdania:

كتب ولد رسالة.

Na początku nowego zdania atrybutowi przypadek przypisywana jest wartość jeden.

<przypadek> = 1;

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce:

- wzorzec translacyjny nie podlegający fleksji:

كتب ولد
chłopiec napisał
<przypadek> = 4;

- wzorzec translacyjny typu deklinacyjnego:

رسالة	<przypadek>
list	1
listu	2
listowi	3
list	4
listem	5
liście	6
<liczba> = 1;	
<rodzaj> = 1;	

Otrzymuje się zatem polski przekład rozważanego arabskiego zdania:

### Chłopiec napisał list

Z kolei w języku perskim obowiązuje szyk zdania typu SOV (*Subject Object Verb*), w związku z czym grupy dopełnienia i orzeczenia zdania tłumaczone są w ramach jednego wzorca translacyjnego. Szyk SOV posiada na przykład zdanie:

آدم روح دارد.

Na początku nowego zdania atrybutowi przypadek przypisywana jest wartość jeden.

<przypadek> = 1;

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce:

– wzorzec translacyjny typu deklinacyjnego:

أدم	<przypadek>
człowiek	1
człowieka	2
człowiekowi	3
człowieka	4
człowiekiem	5
człowieku	6
<liczba> = 1;	
<rodzaj> = 1;	

– wzorzec translacyjny typu koniugacyjnego:

روح دار	<liczba>	<rodzaj>
ma duszę	1	1
ma duszę	1	2
ma duszę	1	3
mają duszę	2	1
mają duszę	2	2
mają duszę	2	3

Otrzymuje się zatem polski przekład rozważanego perskiego zdania:

### Człowiek ma duszę

Opracowana przez autora metoda translacji automatycznej oparta na wzorcach translacyjnych PBMT okazała się dość skutecznym narzędziem umożliwiającym dokonywanie przekładów o relatywnie wysokiej jakości z różnych języków świata na język polski. Dowodem tego są przykłady tłumaczeń tekstów należących do wybranych języków świata, które zostaną zamieszczone w kolejnych artykułach niniejszego cyklu publikacji poświęconych PBMT. Teksty te charakteryzują się różnorodną tematyką treści, a także różnym stopniem trudności, co podyktowane zostało częściowo różnym stopniem znajomości rozważanych języków przez autora, jak również różnym stopniem dostępności materiałów lingwistycznych (gramatyk, słowników, wyborów tekstów) dotyczących rozważanych języków.

## 6. PRÓBA OSZACOWANIA LICZBY WZORCÓW TRANSLACYJNYCH

Podstawowe pytanie związane z metodą translacji automatycznej zaproponowanej przez autora dotyczy rozmiarów bazy danych potrzebnej do efektywnego działania systemu, czyli odnosi się bezpośrednio do liczby wzorców transla-

cyjnych, które muszą zostać w bazie danych zamieszczone. Oszacowanie liczby wymaganych wzorców translacyjnych nie jest sprawą łatwą i ostateczną odpowiedź może przynieść tutaj tylko odwołanie się do empirii, czyli wykonanie eksperymentu, polegającego na zbudowaniu systemu translacji automatycznej i przetestowaniu skuteczności jego działania na wybranych próbkach tekstów. Jednak pomimo oczywistych trudności można podjąć próbę oszacowania zakresu, w którym mieści się liczba wzorców translacyjnych  $L$  wymaganych do sprawnego funkcjonowania systemu.

Poszukiwana liczba  $L$  jest na pewno większa niż liczba haseł  $S$  zawartych w słowniku danego języka. Jedną z bezpośrednich przyczyn jest fakt, iż każdy język odznacza się pewnym stopniem fleksyjności, to znaczy od podstawowych form wyrazów można utworzyć szereg form pochodnych. W związku z powyższym, zdefiniowany może zostać tzw. współczynnik fleksyjności języka  $f$ , jako stosunek liczby możliwych do wygenerowania w danym języku form wyrazowych  $F$  do liczby form podstawowych  $S$  (haseł zawartych w słowniku danego języka), czyli

$$f = \frac{F}{S} \quad (1)$$

Górne ograniczenie przedziału  $L_{\max}$ , w którym zawarta jest liczba wzorców translacyjnych  $L$ , może zostać oszacowane jako iloczyn

$$L_{\max} = F \cdot F \cdot F \cdot \dots \cdot F \quad (2)$$

Liczba czynników  $F$ , występujących we wzorze (2) zależy od maksymalnej długości frazy języka źródłowego, jaka może wystąpić we wzorcu translacyjnym. Widać, że nawet dla umiarkowanych długości fraz występujących we wzorcach translacyjnych wartość liczby  $L_{\max}$  jest wręcz astronomiczna. Na przykład, dla fraz dwuwyrazowych wartość liczby  $L_{\max}$  wynosi  $F^2$ , dla fraz trójwyrazowych  $F^3$  itd. Są to z całą pewnością wartości zawyżone o wiele rzędów wielkości, ponieważ zestawienie ze sobą przypadkowych form fleksyjnych kilku losowo wybranych wyrazów z bardzo wysokim prawdopodobieństwem utworzy frazę niepoprawną z syntaktycznego punktu widzenia. Nawet jeżeli składnia takiej frazy będzie dopuszczalna przez reguły gramatyki danego języka, to i tak z dużym prawdopodobieństwem fraza taka pozbawiona będzie jakiegokolwiek sensownego znaczenia. Zatem rzeczywista wartość liczby  $L$  jest o wiele rzędów wielkości mniejsza od  $L_{\max}$ .

Z kolei dolny kres przedziału  $L_{\min}$ , w którym zawiera się wartość liczby  $L$ , może zostać oszacowany ze wzoru

$$L_{\min} = F \cdot \left( 1 + \frac{l_2 + l_3 + l_4 + l_5 + l_{\geq 6}}{l_1} \right) \quad (3)$$

We wzorze (3) liczba form wyrazowych  $F$  jest przemnożona przez czynnik określający, ile razy liczba fraz dwuwyrazowych, trójwyrazowych, czterowyrazowych, pięciowyrazowych, szściowyrazowych i dłuższych jest większa od liczby fraz składających się z pojedynczego wyrazu, które to frazy wystąpiły w pewnej reprezentatywnej dla danego języka próbie tekstu. Oczywiście może się okazać, że rzeczywista liczba wymaganych wzorców translacyjnych potrzebnych do sprawnego funkcjonowania systemu będzie większa, ale liczba  $L_{\min}$  stanowi z całą pewnością jej kres dolny. Występujący we wzorze (3) czynnik  $F$  może zostać policzony ze wzoru (1), jednakże w tym celu znana musi być wartość współczynnika fleksyjności  $f$  danego języka. Różne języki mają odmienną wartość współczynnika fleksyjności  $f$ , co podyktowane jest głównie faktem, iż języki te mają różną ilość możliwych do utworzenia form deklinacyjnych rzeczowników oraz form koniugacyjnych czasowników.

Na przykład w języku arabskim występują tylko trzy przypadki: mianownik, dopełniacz i biernik, ale za to istnieją trzy liczby: pojedyncza, podwójna i mnoga. Wzór deklinacji arabskiego rzeczownika „طالب” (student) został zamieszczony w tabeli 4.

**Tabela 4.** Wzór deklinacyjny arabskiego rzeczownika „طالب”

Przypadek	Liczba pojedyncza	Liczba podwójna	Liczba mnoga
Mianownik	طالب	طالبان	طلاب
Dopełniacz	طالب	طالبين	طلاب
Biernik	طالبا	طالبين	طلابا

Z kolei w języku islandzkim istnieją cztery przypadki: mianownik, dopełniacz, celownik i biernik. Wzór odmiany przez przypadki islandzkiego rzeczownika „barn” (dziecko) został zamieszczony w tabeli 5.

**Tabela 5.** Wzór deklinacyjny islandzkiego rzeczownika „barn”

Przypadek	Liczba pojedyncza	Liczba mnoga
Mianownik	barn	börn
Dopełniacz	barns	barna
Celownik	barni	börnum
Biernik	barn	börn

Również w języku starogreckim istniało pięć przypadków: mianownik, dopełniacz, celownik, biernik i wołacz. Wzór deklinacji starogreckiego rzeczownika „ποιητής” (poeta) został zamieszczony w tabeli 6.

**Tabela 6.** Wzór deklinacyjny starogreckiego rzeczownika „ποιητής”

Przypadek	Liczba pojedyncza	Liczba mnoga
Mianownik	ποιητής	ποιηταί
Dopełniacz	ποιητου	ποιητων
Celownik	ποιητη	ποιηταις
Biernik	ποιητήν	ποιητάς
Wołacz	ποιητά	ποιηταί

Z kolei w języku łacińskim wyróżnić można sześć przypadków: mianownik, dopełniacz, celownik, biernik, ablativus i wołacz. Wzór odmiany łacińskiego rzeczownika „puer” (chłopiec) zamieszczono w tabeli 7.

**Tabela 7.** Wzór odmiany łacińskiego rzeczownika „puer”

Przypadek	Liczba pojedyncza	Liczba mnoga
Mianownik	puer	pueri
Dopełniacz	pueri	puerorum
Celownik	puero	pueris
Biernik	puerum	pueros
Ablativus	puero	pueris
Wołacz	puer	pueri

Również sześć przypadków posiada język słoweński, z tym że oprócz liczby pojedynczej i mnogiej występuje w nim jeszcze liczba podwójna. W tabeli 8 zamieszczono wzór deklinacji słoweńskiego rzeczownika „lipa” (lipa).

**Tabela 8**

Wzór odmiany słoweńskiego rzeczownika „lipa”

Przypadek	Liczba pojedyncza	Liczba podwójna	Liczba mnoga
Mianownik	lipa	lipe	lipi
Dopełniacz	lipe	lip	lip
Celownik	lipi	lipam	lipama
Biernik	lipo	lipe	lipi
Narzędnik	lipo	lipami	lipama
Miejscownik	lipi	lipah	lipah

W języku polskim przypadków jest siedem, natomiast w sanskrycie przypadków było aż osiem: mianownik, biernik, narzędnik, celownik, ablativus, dopełniacz, miejscownik i wołacz (nazwy przypadków podano w kolejności powszechnie przyjętej w gramatyce języka sanskryckiego). W tabeli 9 zamieszczono wzór deklinacyjny sanskryckiego rzeczownika „sakha” (przyjaciel).

Tabela 9

Wzór odmiany sanskryckiego rzeczownika „sakha”

Przypadek	Liczba pojedyncza	Liczba podwójna	Liczba mnoga
Mianownik	sakhā	sakhāyau	sakhāyah
Biernik	sakhāyam	sakhāyau	sakhīn
Narzędnik	sakhyā	sakhibhyām	sakhibhih
Celownik	sakhye	sakhibhyām	sakhibhyah
Ablativus	sakhyuh	sakhibhyām	sakhibhyah
Dopełniacz	sakhyuh	sakhyoh	sakhīnām
Miejscownik	sakhyau	sakhyoh	sakhisu
Wołacz	sakhe	sakhyau	sakhāyah

Różne języki świata posiadają także odmienne wzorce koniugacji czasowników. Na przykład w języku szwedzkim fleksja czasownikowa uległa maksymalnemu uproszczeniu. W tabeli 10 zamieszczono wzór koniugacyjny szwedzkiego czasownika „att aka” (jechać) w czasie teraźniejszym.

Tabela 10. Wzór odmiany szwedzkiego czasownika „att aka”

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	åker	åker
Druga	åker	åker
Trzecia	åker	åker

Nieco więcej form fleksyjnych posiada czasownik w języku niderlandzkim. W tabeli 11 zamieszczono wzór odmiany w czasie teraźniejszym niderlandzkiego czasownika „spreken” (mówić).

Tabela 11. Wzór odmiany niderlandzkiego czasownika „spreken”

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	spreek	spreken
Druga	spreekt	spreken
Trzecia	spreekt	spreken

Zdecydowanie bogatszą fleksję form czasownikowych wykazuje język węgierski. W tabeli 12 zamieszczono wzór odmiany w czasie teraźniejszym węgierskiego czasownika „tudni” (wiedzieć).

Tabela 12. Wzór odmiany węgierskiego czasownika „tudni”

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	tudok	tudunk
Druga	tudsz	tudtok
Trzecia	tud	tudnak

Równie bogata fleksja czasownikowa występuje w języku perskim. W tabeli 13 zamieszczono wzór koniugacji perskiego czasownika „داشتن” (mieć).

Tabela 13. Wzór odmiany perskiego czasownika „داشتن”

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	دارم	داریم
Druga	داری	دارید
Trzecia	دارد	دارند

Natomiast w języku arabskim fleksja czasownikowa jest już bardzo rozbudowana. W tabeli 14 zamieszczono wzór odmiany w czasie teraźniejszym arabskiego czasownika „عمل” (pracować).

Tabela 14. Wzór odmiany arabskiego czasownika „عمل”

Osoba (rodzaj)	Liczba pojedyncza	Liczba podwójna	Liczba mnoga
Osoba pierwsza	أعمل	أعمل	أعمل
Osoba druga (rodzaj męski)	تعملون	تعملان	تعمل
Osoba druga (rodzaj żeński)	تعلمن	تعلمان	تعلمين
Osoba trzecia (rodzaj męski)	يعملون	يعملان	يعمل
Osoba trzecia (rodzaj żeński)	يعلمن	يعلمان	يعلم

Jak wynika z zamieszczonych rozważań, wartość współczynnika fleksyjności  $f$  musi zostać wyznaczona indywidualnie dla każdego z języków na drodze analizy możliwych do utworzenia w tych językach form fleksyjnych wyrazów. Autor podjął próbę oszacowania wartości współczynników fleksyjności dla 10 języków należących do germańskiej



i romańskiej rodziny językowej, co zostało szczegółowo omówione w kolejnym artykule należącym do niniejszej serii publikacji.

Ponadto w kolejnych artykułach zamieszczono przykłady zastosowania opracowanej przez autora metody automatycznej translacji opartej na wzorcach (PBMT) dla wybranych języków świata.

## 7. ZAKOŃCZENIE

Artykuł niniejszy stanowi wprowadzenie do opracowanej przez autora metody automatycznej translacji opartej na wzorcach PBMT. Zdaniem autora zaproponowana metoda może stanowić istotny krok na drodze do zbudowania w pełni automatycznych systemów komputerowego przekładu, charakteryzujących się stosunkowo wysoką jakością uzyskiwanych tłumaczeń tekstów należących do języka ogólnego. Jediną przeszkodę stanowi tutaj rozmiar bazy danych, zawierającej wzorce translacyjne. Niestety opracowanie takiej bazy danych nie może zostać zautomatyzowane, ponieważ każdy z wzorców translacyjnych wymaga indywidualnego potraktowania i przełożenia na język docelowy przez tłumacza specjalistę. Zatem, aby osiągnąć konkretne rezultaty w postaci działającego systemu PBMT, konieczna jest szeroko zakrojona współpraca z zespołami lingwistów, którzy będą w stanie poświęcić wiele godzin swej mozolnej pracy na zapewnienie bazy danych występującymi najczęściej w danym języku wzorcami translacyjnymi. Niestety na realizację takiego projektu potrzeba zarówno czasu, jak i pieniędzy, gdyż w grę wchodzi przynajmniej kilka lat pracy zespołu tłumaczy złożonego, co najmniej, z kilkunastu osób. Na razie system PBMT jest rozwijany, dla kilku wybranych języków, przez autora oraz jego dyplomantów, a rezultaty dotychczasowej pracy można obejrzeć na stronie internetowej o adresie <http://machtran.go.pl>.

## Literatura

- [1] Juang B.H., Furui S.: *Automatic Recognition and Understanding of Spoken Language – A First Step Toward Natural Human-Machine Communication*. Proceedings of the IEEE, vol. 88, 2000, 1142–1165
- [2] Bole L., Cichy M., Różańska L.: *Przetwarzanie języka naturalnego*. Warszawa, Wydawnictwa Naukowo-Techniczne 1982
- [3] Szczepaniak L., Królikowski Z.: *Kontrolowane języki naturalne – przegląd rozwiązań i zastosowań*. Pro Dialog, nr 11, 2000, 47–67
- [4] Siu M., Ostendorf M.: *Variable N-grams and Extensions for Conversational Speech Language Modeling*. IEEE Transactions on Speech and Audio Processing, vol. 8, 2000, 63–74
- [5] Gajer M.: *System automatycznej identyfikacji języka tekstu*. Półrocznik AGH Elektrotechnika i Elektronika, t. 20, z. 2, 2001, 100–107
- [6] Bellegarda J.R.: *Large Vocabulary Speech Recognition with Multi-span Statistical Language Models*. IEEE Transactions on Speech and Audio Processing, vol. 8, 2000, 76–84
- [7] Bu L., Chiueh T.D.: *Perceptual Speech Processing and Phonetic Feature Mapping for Robust Vowel Recognition*. IEEE Transactions on Speech and Audio Processing, vol. 8, 2000, 105–114
- [8] Levin E., Pieraccini R., Eckert W.A.: *A Stochastic Model of Human-Machine Interaction for Learning Dialog Strategies*. IEEE Transactions on Speech and Audio Processing, vol. 8, 11–22
- [9] Riccardi G., Gorin A.L.: *Stochastic language Adaptation over Time and State in Natural Spoken Dialog Systems*. IEEE Transactions on Speech and Audio Processing, vol. 8, 3–10
- [10] Waibel A., Geatner P., Tomokiyo L.M., Schultz T., Woszczyna M.: *Multilinguality in Speech and Spoken Language Systems*. Proceedings of the IEEE, vol. 88, 2000, 1297–1313
- [11] Arnold D., Balkan L., Meijer S., Humphreys R.L., Sadler L.: *Machine Translation: An Introductory Guide*. London, NCC Blackwell 1994
- [12] Ludskanow A.: *Tłumaczy człowiek i maszyna cyfrowa*. Warszawa, Wydawnictwa Naukowo-Techniczne 1973
- [13] Hutchins W.J.: *Machine Translation – Past, Present, Future*. London, Ellis Horwood Series in Computers and Their Applications 1986
- [14] Blekhman M., Pevzner B.: *First Steps of Language Engineering in the USSR: The 50s through 70s*. Machine Translation Review, vol. 11, 2000, 5–7
- [15] Whitelock P., Kilby K.: *Linguistic and Computational Techniques in Machine Translation System Design*. London, UCL Press 1995, ISBN 1-85728-216-7
- [16] Gajer M.: *Translacja automatyczna w sieci Internet*. Studia Informatica, vol. 23, No. 2B (49), 133–140
- [17] Vasconcellos M.: *Users of systems: The current state of MT usage or: How do I use these? Let me count the ways*. Fourth Machine Translation Summit in Kobe, 1993, <http://www.eamt.org/archive/summit93.html>
- [18] Mitamura T.: *Controlled languages for Multilingual Machine Translation*. Language Technologies Institute, School of Computer Science, Carnegie Mellon University, USA, 1999, <http://www.lti.cs.cmu.edu/Research/Kant>
- [19] Fukutomi O.: *Report on Commercial Machine Translation in a Manufacturing Industry Domain*. Machine Translation Review, No. 10, 1999, 16–25
- [20] Murphy D.: *Keeping Translation Technology under Control*. Machine Translation Review, No. 11, 2000, 7–10
- [21] Canals R., Esteve A., Garrido A.: *interNOSTRUM: A Spanish-Catalan Machine Translation System*. Machine Translation Review, No. 11, 2000, 21–25
- [22] Bandyopadhyay S.: *State and Role of Machine Translation in India*. Machine Translation Review, No. 11, 2000, 1–3
- [23] Loukachevitch N.V., Dobrov B.V.: *Thesaurus-Based Structural Thematic Summary in Multilingual Information Systems*. Machine Translation Review, No. 11, 2000, 10–20
- [24] Harel D.: *Rzecz o istocie informatyki*. Warszawa, Wydawnictwa Naukowo-Techniczne 1992, ISBN 83-204-1525-X
- [25] Penrose R.: *Nowy umysł cesarza. O komputerach, umyśle i prawach fizyki*. Warszawa, Wydawnictwo Naukowe PWN 1995, ISBN 83-01-11819-9
- [26] Wiślicki W.: *Czy możliwy jest naukowy opis tego, co dzieje się w umyśle?* Wszechświat, t. 101, nr 1–3, 2000, 8–12
- [27] Kasperski M.: *Sztuczna inteligencja – Droga do myślących maszyn*. Gliwice, Wydawnictwo Helion 2003
- [28] Tadeusiewicz R.: *Problemy biocybernetyki*. Warszawa, PWN 1994
- [29] Gajer M.: *Czy komputer może mieć świadomość?* Elektronizacja, nr 3/2002, 2002, 20–23
- [30] Gajer M.: *Czy komputery zastąpią kiedyś ludzi? Pomiary Automatyka Robotyka – Miesięcznik Naukowo-Techniczny*, nr 12, 2001, 14–19
- [31] Gajer M.: *Gdzie leżą granice możliwości systemów sztucznej inteligencji? (część I)*. Elektronizacja, nr 10/2002, 2002, 2–6
- [32] Gajer M.: *Gdzie leżą granice możliwości systemów sztucznej inteligencji? (część II)*. Elektronizacja, nr 12/2002, 2002, 2–6
- [33] Rutkowski P.: *Jak wiele można przetłumaczyć?* Wiedza i Życie, luty 2001, 42–44
- [34] Lebedziński H.: *Przekładoznawstwo ogólne wobec teorii enroi*. Warszawa, Państwowe Wydawnictwo Naukowe 1989
- [35] Majewicz A.F.: *Języki świata i ich klasyfikowanie*. Warszawa, Państwowe Wydawnictwo Naukowe 1989, ISBN 83-01-08163-5

- [36] Gajer M.: *Możliwości i ograniczenia systemów translacji automatycznej*. Półrocznik AGH Elektrotechnika i Elektronika, t. 21, z. 1, 2002, 1–17
- [37] Gajer M.: *Interaktywny system translacji automatycznej oparty na podejściach typu KBMT oraz EBMT*. Półrocznik AGH Elektrotechnika i Elektronika, t. 21, z. 1, 2002, 18–37
- [38] Gajer M.: *System translacji automatycznej oparty na dialogu z użytkownikiem i przykładach translacyjnych*. Pro Dialog – Computer Programming and Applications, No. 14, 2002, 77–115
- [39] Gajer M.: *Analiza możliwości i ograniczeń współczesnych systemów translacji automatycznej*. IX Konferencja „Systemy Czasu Rzeczywistego”, Ustroń, 16–19 września 2002, 481–491
- [40] Gajer M.: *Translacja automatyczna – ograniczenia i perspektywy rozwoju*. Automation 2002, Konferencja Naukowo-Techniczna „Automatyzacja – Nowości i Perspektywy”, 20–22 marca 2002, Warszawa, 102–109
- [41] Gajer M.: *Wprowadzenie do zagadnień translacji automatycznej*. Kwartalnik Elektroniki i Telekomunikacji, t. 48 z. 1, 2002, 117–160
- [42] Gajer M.: *Translacja automatyczna – podejścia oparte na koncepcji języka interlingua i na przykładach translacyjnych*. Kwartalnik Elektroniki i Telekomunikacji, t. 48, z. 3–4, 2002, 523–551
- [43] Gajer M.: *System translacji automatycznej dla języka polskiego*. Wiadomości Elektrotechniczne, Rok LXX, nr 7–8, 2002, 275–280
- [44] Gajer M.: *Komputerowe systemy automatycznego tłumaczenia dokumentacji technicznej*. Wiadomości Elektrotechniczne, Rok LXX, nr 6, 2002, 228–232
- [45] Gajer M.: *Fakty i mity na temat systemów translacji automatycznej*. Elektronizacja, nr 1–2/2002, 2002, 33–36
- [46] Nyberg E., Mitamura T., Carbonell J.: *The KANT Machine Translation System: From R&D to Initial Deployment*. Proceedings of LISA Workshop on Integrated Advanced Translation Technology, Washington, 1997, <http://www.lti.cs.cmu.edu/Research/Kant>
- [47] Ney H., Niessen S., Och F.J., Sawaf H., Tillmann C., Vogel S.: *Algorithms for Statistical Translation of Spoken Language*. IEEE Transactions on Speech and Audio Processing, vol. 8, 200, 24–35
- [48] Brown R.D.: *Transfer-Rule Induction for Example-Based translation*. MT Summit VIII, Santiago de Compostela, 2001, <http://www.eamt.org/summitVIII/workshop-papers.html>
- [49] Brown R.D.: *Transfer-Rule Induction for Example-Based Translation*, Language Technologies Institute, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 2000, <http://www.lti.cs.cmu.edu/Research/GEBMT>
- [50] Brown R.D.: *Automated Dictionary Extraction for “Knowledge-Free” Example-Based Translation*. Language Technologies Institute, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 2000, <http://www.lti.cs.cmu.edu/Research/GEBMT>
- [51] Brown R.D.: *Example-Based Machine Translation in the Pangloss System*. Language Technologies Institute, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 2000, <http://www.lti.cs.cmu.edu/Research/GEBMT>
- [52] Brown R.D.: *Adding Linguistic Knowledge to a Lexical Example-Based Translation System*. Language Technologies Institute, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 2000, <http://www.lti.cs.cmu.edu/Research/GEBMT>
- [53] Planas E., Furuse O.: *Formalizing Translation Memories*. Proceedings of MT Summit VII, Singapore, 1999, 331–339
- [54] Schäler R.: *Beyond Translation Memories*. MT Summit VIII, Santiago de Compostela, 2001, <http://www.eamt.org/summitVIII/workshop-papers.html>
- [55] Geambașu C., Odrobińska E.: *Konwersacje polsko-rumuńskie*. Kraków, Wydawnictwo Uniwersytetu Jagiellońskiego 2001
- [56] Pérez C.R.: *From Novelty to Ubiquity: Computers and Translation at the Close of the Industrial Age*. Translation Journal, vol. 5, No. 1, 2001
- [57] Stanisławski J.: *Wielki słownik angielsko-polski z suplementem*. Warszawa, Państwowe Wydawnictwo Wiedza Powszechna 1983
- [58] Lindquist M., Stypińska M.: *På svenska – Podręcznik do nauki języka szwedzkiego dla początkujących*. Warszawa, PWN 1991

Wpłynęło: 10.02.2005

Miroslaw GAJER



Urodził się 25 kwietnia 1971 roku. Ukończył kierunek elektroniki ze specjalnością aparatura elektroniczna na Wydziale Elektrotechniki, Automatyki i Elektroniki Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. Dyplom magistra inżyniera otrzymał w 1996 roku. Następnie kontynuował naukę na studiach doktoranckich na Wydziale Elektrotechniki, Automatyki i Elektroniki Akademii Górniczo-Hutniczej, które zakończył złożeniem rozprawy doktorskiej. W dniu 26 października 2000 roku Rada Wydziału Elektrotechniki, Automatyki, Informatyki i Elektroniki Akademii Górniczo-Hutniczej nadała mu tytuł doktora nauk technicznych w dyscyplinie informatyka oraz przyznane mu zostało wyróżnienie. Obecnie jest pracownikiem naukowo-dydaktycznym w Katedrze Automatyki AGH, zatrudnionym na stanowisku adiunkta. Swoje obecne zainteresowania wiąże z dziedziną sztucznej inteligencji, a zwłaszcza z technikami przetwarzania języka naturalnego i automatycznej translacji.

e-mail: mgajer@ia.agh.edu.p