

WNIOSKI WYNIKAJĄCE Z BADAŃ NAD METODĄ TRANSLACJI AUTOMATYCZNEJ OPARTĄ NA WZORCACH

STRESZCZENIE

Translacja automatyczna jest dziedziną lingwistyki informatycznej, której zadaniem jest opracowanie metod komputerowego przekładu tekstów zapisanych w danym języku naturalnym na inny język naturalny. Pomimo ponad 50 lat intensywnych badań automatyczny przekład jest wciąż daleki od zamierzonego ideału. Autor artykułu zaproponował nowe podejście do translacji automatycznej, które zostało przez niego nazwane translacją opartą na wzorcach PBMT (Pattern-Based Machine Translation). W artykule zamieszczono wyniki badań nad translacją opartą na wzorcach oraz wypływające z nich wnioski. Opisano również proponowane dalsze kierunki badawcze związane z automatyzacją przekładu.

Słowa kluczowe: *lingwistyka informatyczna, translacja automatyczna, translacja oparta na wzorcach*

THE CONCLUSIONS DRAWN UPON PATTERN-BASED MACHINE TRANSLATION

Machine translation is a discipline of computational linguistics the purpose of which is to develop the methods of computerized translation of text written in one natural language into another. Despite over 50 years of intensive research effort the automatic translation is still far away from what we would like to achieve. This author has proposed an original method of machine translation which was called Pattern-Based Machine Translation (PBMT). In the paper the results of implementation of PBMT for different languages are presented. Some conclusions drawn upon pattern-based machine translation method and directions of further research are also discussed.

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

1. WPROWADZENIE

Artykuł niniejszy jest trzecim artykułem z serii publikacji poświęconych opracowanej przez autora metodzie automatycznego przekładu opartej na wzorcach PBMT (Pattern-Based Machine Translation). Metoda ta stanowi hybrydę znanych uprzednio metod automatycznego przekładu, takich jak *Example-Based Machine Translation* i *Rule-Based Machine Translation*. Zdaniem autora metoda PBMT stwarza realne szanse na znaczne podniesienie jakości przekładów uzyskiwanych na drodze automatycznej. Próby zastosowania metody PBMT dla wybranych języków dały bardzo zachęcające rezultaty, ponieważ zastosowanie wzorców translacyjnych o umiarkowanej długości daje przekład w znacznym stopniu przypominający rezultat pracy człowieka tłumacza.

Uzyskane przez autora wstępne wyniki badań nad translacją opartą na wzorcach dowodzą dużej skuteczności metody PBMT. Również jakość uzyskanych na tej drodze przekładów musi być oceniona jako stosunkowo wysoka, o czym świadczyć może zamieszczony poniżej przykład translacji dokonanej za pomocą odpowiednio przygotowanych wcześniej i wyselekcjonowanych wzorców translacyjnych.

2. TRANSLACJA PRZYKŁADOWEGO TEKSTU METODĄ PBMT

Zadaniem systemu automatycznej translacji było przetłumaczenie następującego tekstu zapisanego w języku angielskim. Tekst ten został zaczerpnięty z książki [1]:

For the most part, non-English colonists adapted themselves to the culture of the original settlers. But this did not mean that all settlers transformed themselves into Englishmen. True, they adopted the English language and law and many English customs, but only as these had been modified by conditions in America. The result was a unique culture – a blend of English and continental European conditioned by the environment of the New World.

Although a man and his family could move from Massachusetts to Virginia, or from South Carolina to Pennsylvania, without making many basic readjustments. Distinctions between individual colonies were marked. They were even more marked between regional groups of colonies.

The settlements fell into fairly well-defined sections determined by geography. In the south, with its warm climate and fertile soil, a predominantly agrarian society developed. New England in the northeast, a glaciated area strewn with boulders, was inferior farm country, with generally thin, stony soil, relatively little level land, short summers, and long winters. Turning to other pursuits, the New Englanders harnessed water power and established gristmills and sawmills. Good stands of timber encouraged shipbuilding. Excellent harbors prompted trade, and the sea became a source of great wealth. In Massachusetts, the cod industry alone quickly furnished a basis for prosperity.

Settling in villages and towns around the harbors, New Englanders quickly adopted an urban existence, many of them carrying on some trade or business. Common pastureland and common wood-lots served the needs of townspeople, who worked small farms nearby. Compactness made

* Akademia Górniczo-Hutnicza w Krakowie, Katedra Automatyki

possible the village school, the village church, and the village or town hall, where citizens met to discuss matters of common interest. Sharing hardship, cultivating the same rocky soil, pursuing simple trades and crafts, New Englanders rapidly acquired characteristics that marked them as a self-reliant, independent people.

W wyniku pracy systemu translacji automatycznej otrzymano następujący polski przekład angielskiego oryginału:

W większości przypadków koloniści nie będący Anglikami zaadoptowali się do kultury pierwotnych osiedleńców. Ale to wcale nie oznacza że wszyscy osiedleńcy przekształcili się w Anglików. Wprawdzie oni przyjęli język angielski i prawo i wiele obyczajów angielskich ale tylko jak zostały zmodyfikowane przez warunki w Ameryce. Dało to w efekcie unikalną kulturę mieszankę angielskości i kontynentalnej europejskości co było uwarunkowane przez otoczenie Nowego Świata. Chociaż mężczyzna i jego rodzina mogła wyprowadzić się z Massachusetts do Wirginii lub z Południowej Karoliny do Pensylwanii nie czyniąc wiele zasadniczych nowych przystosowań. Różnice pomiędzy poszczególnymi koloniami były znaczące. One były nawet bardziej odznaczające się pomiędzy regionalnymi grupami kolonii. Miejsca osiedlania się podpadały pod dość dobrze określone obszary uwarunkowane przez geografę. Na południu z jego ciepłym klimatem i żyzną glebą rozwinęło się społeczeństwo w przeważającej mierze rolnicze. Nowa Anglia na północnym wschodzie polodowcowa kraina zarzucona otoczkami była krainą nędznych gospodarstw z ogólnie niskiej klasy kamienistą glebą względnie małym płaskim obszarem krótkimi latami i długimi zimami. Kierując uwagę na inne rodzaje działalności mieszkańcy Nowej Anglii wykorzystywali siłę wód i założyli młyny zbożowe i tartaki. Obfitość surowca drzewnego zachęcała do budowy statków. Wspaniałe porty skłaniały do aktywności handlowej i morze stało się źródłem wielkiego bogactwa. W Massachusetts sam tylko przemysł dorszowy szybko stworzył podstawy dla dobrobytu. Osiedlając się w wioskach i miastach wokół portów mieszkańcy Nowej Anglii szybko przyjęli miejski sposób życia wielu z nich zajmowało się pewną formą handlu lub działalnością gospodarczą. Wspólne pastwiska i wspólne miejsca pozyskiwania drewna zaspokajały potrzeby mieszczan którzy prowadzili małe gospodarstwa w pobliżu. Zwartość sprawiła możliwym utworzenie wiejskiej szkoły wiejskiego kościoła w pobliżu wiejskiej bądź miejskiej sali zebrań gdzie obywatele spotykali się aby omawiać wspólne sprawy. Walcząc o przetrwanie uprawiając tą samą skalistą glebę zajmując się prostym handlem i rzemiosłem mieszkańcy Nowej Anglii bardzo szybko przybrali cechy charakteru które odznaczały ich jako polegających na sobie niezależnych ludzi.

W celu dokonania porównania zaproponowanej przez autora metody translacji automatycznej opartej na wzorcach PBMT z innymi systemami translacji automatycznej, autor wykonał następujący eksperyment. W eksperymencie tym rozważany angielski tekst został zadany na wejście systemu translacji automatycznej „English Translator 2”. Pro-

gram „English Translator 2” jest dostępnym na polskim rynku systemem translacji automatycznej dokonującym przekładów w obu kierunkach pomiędzy językiem polskim i językiem angielskim [2]. Program ten jest produktem polskiej firmy „Techland”. W wyniku pracy programu „English Translator 2” otrzymano następujący polski przekład angielskiego tekstu:

Przeważnie, nie – angielski koloniści dostosowałem się do kultura oryginalny osadnicy. Ale ten nie znaczył że wszyscy osadnicy przekształcili się do Anglików. Prawdziwy, oni adoptowali angielski język i prawo i dużo angielskie zwyczaje, ale tylko jako tego został zmodyfikowany przez warunków w Ameryce. Wynik był unikalną kulturą – mieszanka języka angielskiego i kontynentalny europejski warunkowany przez środowisko Nowego Świata.

Chociaż człowiek i jego rodzina mogliby ruszać z Massachusetts do Wirginii, lub z Południowej Karoliny do Pensylwanii, robiący dużo podstawowe dopasowania. Różnice między indywidualny kolonie zostały oznaczone. Oni byli nawet bardziej oznaczeni między regionalnymi grupami kolonii.

Osady spadły do uczciwie dobrze – zdefiniowany sekcje określone przez geografę. W południu, z jego ciepłym klimatem i urodzajną ziemią, przeważnie rolnicze społeczeństwo rozwinięte. Nowy Anglia w północny wschód, pokryty lodem obszar posypany z otoczkami, był niższym gospodarstwem rolnym krajem, z ogólnie cienki, kamienista ziemia, względnie mała poziom ziemia, krótkie lata, i długie zimy. Skręt do innych ścigań, Nowe Englanders skontrolowały woda moc i założone gristmills i tartaki. Dobre stojaki drewna zachęciły budownictwo okrętowe. Doskonale przystanie umotywowwały handel, i morze stanął się źródłem wielkiego bogactwa. W Massachusetts, dorsz przemysł sam szybko zaopatrzone podstawę dla pomyślności.

Aklimatyzujący się wsi i miast dookoła przystani, Nowe Englanders szybko adoptowały miejskie istnienie, dużo nich transportowych na jakimś handlu lub biznesie. Wspólny pasturand i wspólny las – wielkie ilości służyły potrzeby mieszczan, kto pracowały małe gospodarstwa rolne w pobliżu. Compactness zrobiły możliwą wieś szkołę, wieś kościół, i wieś lub ratusz, gdzie obywatele spotkali się omówić sprawy wspólnego zainteresowania. System wielodostepowy męki, hodująca tak samo skalistą ziemię, ścigający proste handle i rzemiosła, Nowe Englanders szybko nabyte właściwości że oznaczony nich jak niezależny, niezależni ludzie.

Porównując przekład angielskiego tekstu otrzymany za pomocą programu „English Translator 2” z przekładem tego samego tekstu dokonany metodą zaproponowaną przez autora, widać, że metoda translacji automatycznej oparta na wzorcach PBMT daje w wyniku tekst o istotnie wyższej jakości. Fakt ten świadczy o słuszności obranej przez autora metodyki automatycznej translacji i stanowi wytyczenie kierunku budowy w pełni automatycznych systemów translacji automatycznej dokonujących przekładów o wysokiej jakości z dowolnych języków obcych na język polski [3].

3. ANALIZA JAKOŚCI PRACY WSPÓŁCZESNYCH SYSTEMÓW TRANSLACJI AUTOMATYCZNEJ

Potencjalnych obszarów zastosowań systemów translacji automatycznej jest dosłownie bez liku. Systemy takie mogą okazać się niezwykle przydatne i częstokroć wręcz nieodzowne w wielu międzynarodowych korporacjach, które posiadają swoje oddziały rozproszone na całym świecie i muszą w związku z tym tłumaczyć olbrzymie ilości dokumentów prawnych, handlowych i technicznych na wiele języków – co w warunkach powszechnej globalizacji gospodarki staje się codziennością. Również przedsiębiorstwa, które eksportują swoje produkty za granicę do wielu różnych krajów i obszarów językowych, stają często przed koniecznością tłumaczenia na mnóstwo języków dołączanych do tych produktów ulotek informacyjnych, instrukcji obsługi, podręczników użytkownika, katalogów, dokumentacji technicznej itp. Nie trzeba chyba w tym miejscu dodawać, że jest to ogromne pole do popisu dla badaczy pracujących w dziedzinie translacji automatycznej, a ich sukcesy łatwo mogą znaleźć tutaj przełożenie na wymierne korzyści finansowe.

Warto również nadmienić, że systemy translacji automatycznej mogą okazać się także niezwykle pomocne dla osób uczących się języków obcych. Natomiast systemy translacji automatycznej nigdy nie służyły, i prawdopodobnie nigdy nie będą służyć, do tłumaczenia utworów literackich. Nie oznacza to oczywiście, że na wejście systemu translacji automatycznej nie może zostać podany utwór literacki zapisany w języku źródłowym. Owszem może, z tym tylko zastrzeżeniem, że nie należy oczekiwać, iż na wyjściu systemu translacji automatycznej otrzyma się utwór literacki zapisany w języku docelowym. Jak możliwy do otrzymania za pomocą komputera przekład może być daleki od dzieła literackiego będącego owocem pracy wybitnego tłumacza, świadczyć może następujący przykład. Autor zadał na wejście programu translacji automatycznej „English Translator 2”, dokonującego przekładów z języka angielskiego na polski, następujący fragment utworu literackiego autorstwa Edgara Allana Poe'a zatytułowanego *The Premature Burial*:

There are certain themes of which the interest is all-absorbing, but which are too entirely horrible for the purposes of legitimate fiction. These the mere romanticist must eschew, if he do not wish to offend, or to disgust. They are with propriety handled only when the severity and majesty of truth sanctify and sustain them.

Poniżej zamieszczono efekt pracy komputerowego tłumacza – polski tekst przekładu:

Tam pewne tematy są o które zainteresowanie jest wszystkim – pochłanianiem, ale który są zbyt całkowicie straszny dla celów prawowitej fikcji. Ten zwykły romantyk musi unikać, jeśli nie życzy obrazić, lub zniechęcić. Oni są z stosownością ujęta tylko kiedy surowość i majestat prawdy święcą i podtrzymują nich.

Niestety efekt pracy programu „English Translator 2” przypomina coś, co w środowisku osób zajmujących się

translacją automatyczną określane jest angielskim terminem *word salad*.

Natomiast poniżej pokazano, jak wygląda tłumaczenie rozważanego angielskiego tekstu w wykonaniu wybitnego tłumacza literatury angielskiej Stanisława Wyrzykowskiego:

Istnieją pewne tematy przejmujące do głębi, wszelako nazbyt okropne, iżby mogły stać się przedmiotem rzetelnej literatury powieściowej. Rzetelni pisarze muszą ich unikać, jeśli nie chcą wywołać zgorznienia lub niesmaku. Godzi się je podejmować tylko w takich razach, gdy uświęca je i usprawiedliwia surowe dostojęstwo prawdy.

Czy zatem programy translacji automatycznej będą kiedykolwiek w stanie dorównać jakością swej pracy człowiekowi tłumaczowi? Jest to bardzo trudne pytanie, będące raczej domeną filozofii niż techniki. Autor wolałby raczej nie podejmować próby udzielenia na nie odpowiedzi. Jednakże kierując się pewnymi przesłankami, uważa, że komputery (przynajmniej takie, jakie znamy w chwili obecnej, czyli deterministyczne maszyny Turinga) nigdy nie pozwolą na całkowite wyeliminowanie człowieka i tym samym zwolnienie go z wszelkich wysiłków intelektualnych. Aby stało się to możliwe, musiałby nastąpić wielki przełom w fizyce, który dopiero utorowałby drogę do wypracowania nowych metod przetwarzania informacji. Wciąż jeszcze wiemy zbyt mało o zasadach budowy i funkcjonowania ludzkiego mózgu oraz o związkach zachodzących w nim procesów ze świadomością i informacją [4]. Czy jednak taki przełom kiedykolwiek nastąpi, i jaki ewentualnie będzie miał charakter, w chwili obecnej na te pytania nie sposób odpowiedzieć.

Czy zatem komputerowy tłumacz równie sprawny w swej działalności jak człowiek tłumacz profesjonalista pozostanie na zawsze świętym Graalem poszukiwań badawczych? Niestety wiele wskazuje na to, że tak, co oczywiście nie przekreśla sensu prowadzenia badań nad translacją automatyczną. Nawet jeśli tłumaczenie tekstów literackich pozostanie na zawsze domeną działalności człowieka, to i tak systemy translacji automatycznej mogą okazać się niezwykle użyteczne w tłumaczeniu innych, bardziej stereotypowych tekstów. Dzięki temu tłumacze będą mieli więcej czasu, aby zająć się tłumaczeniem bardziej ambitnych tekstów – co zawsze stanowi dla nich duże wyzwanie intelektualne.

Jak zatem rysuje się przyszłość przed translacją automatyczną? Zdaniem autora duży postęp badań w dziedzinie lingwistyki komputerowej uprawnia do wysunięcia tezy, iż za kilkanaście lat komputery będą tłumaczyć pewne teksty standardowe, ukierunkowane na konkretną dziedzinę działalności ludzkiej równie dobrze jak zawodowi tłumacze. Komputery staną się w równej mierze sprawne w translacji takich tekstów, jak już obecnie są sprawne na przykład w grze w szachy. Niestety nie można mieć natomiast żadnej nadziei, że komputery kiedykolwiek będą cokolwiek rozumieć z tłumaczonych przez siebie tekstów, podobnie jak obecnie nie rozumieją one niczego z rozgrywanej przez siebie partii szachów, co wcale nie przeszkadza im, aby tę partię szachów wygrać.

4. TRANSLACJA AUTOMATYCZNA NA TLE RÓŻNORODNOŚCI JĘZYKÓW ŚWIATA

Okazuje się, że udzielenie odpowiedzi na pytanie, ile jest języków na świecie, nie jest wcale sprawą prostą. Jednym z powodów takiego stanu rzeczy jest niemożność podania jakichś precyzyjnych kryteriów, które pozwoliłyby na uzyskanie jednoznacznej odpowiedzi na pytanie, czy w danym przypadku mamy do czynienia jedynie z dwoma różnymi odmianami dialektalnymi tego samego języka, czy też z dwoma odrębnymi językami. Zatem nader często o przyznaniu statusu odrębnego języka decydują czynniki pozalingwistyczne (najczęściej natury politycznej) [5]. Zjawisko to może zostać zilustrowane na kilku przykładach. Mianowicie przed rozpadem Jugosławii mówiono tylko o języku serbsko-chorwackim, pomimo oczywistego faktu znacznego jego regionalnego zróżnicowania dialektalnego, używania do zapisu dwóch odrębnych alfabetów, znacznych różnic w leksyce itp. Natomiast obecnie mówi się wyłącznie o językach serbskim i chorwackim, a czasami nawet bośniackim. Z drugiej strony, jeszcze do niedawna w językoznawstwie mówiono o językach holenderskim i flamandzkim, podczas gdy obecnie używa się prawie wyłącznie terminu język niderlandzki, który obejmuje swym zasięgiem oba wymienione języki.

Ponadto w Europie w wielu krajach (zwłaszcza romańskojęzycznych), oprócz języka urzędowego (włoskiego, francuskiego, hiszpańskiego, czy portugalskiego) w codziennym użyciu funkcjonuje wiele języków regionalnych, które znacznie różnią się pomiędzy sobą, a dla osób znających jedynie język urzędowy są praktycznie niezrozumiałe. Szczególnie skomplikowana sytuacja panuje we Włoszech, gdzie około 30% populacji nie zna w ogóle języka urzędowego i posługuje się wyłącznie lokalnym dialektem. Podobnie skomplikowana sytuacja językowa istnieje w południowej Francji, która od strony językowej prezentuje prawdziwą mozaikę różnorodnych dialektów.

Innym powodem trudności z precyzyjnym określeniem liczby języków świata jest fakt polegający na tym, że zdecydowana większość (ponad 80%) języków świata pozostaje językami niepisanymi. Poza tym bardzo wiele języków świata pozostaje w ogóle niezbadane, gdyż nie opracowano do tej pory ich słowników ani gramatyk [6]. Zapewne pokazanej liczby języków świata jeszcze do chwili obecnej nie odkryto. Niestety bardzo wiele z tych języków pozostanie niezbadane i nieznanie już na zawsze, ponieważ proces wymierania małych języków w ostatnich dziesiątkach lat przybrał na sile z niespotykaną dotąd intensywnością [7].

W związku z powyższymi trudnościami w językoznawstwie wprowadzono bardzo użyteczne pojęcie tzw. etnolektu, czyli środka komunikacji językowej będącego w użyciu przez daną grupę etniczną [5]. Wprowadzenie terminu etnolektu pozwala, w naturalny sposób, na ominięcie wszelkich kwestii spornych, polegających na niemożności ustalenia, czy mamy do czynienia z odrębnym językiem, czy tylko z jego dialektem bądź gwarą. Na przykład w myśl wprowadzonej terminologii, zarówno język polski, jak i kaszub-

ski mają status odrębnych etnolektów, przy czym ewentualny wynik sporów prowadzonych na temat tego czy kaszubski jest dialektem języka polskiego, czy też całkowicie odrębnym językiem (aczkolwiek blisko z językiem polskim spokrewnionym), jest zupełnie bez znaczenia.

W Centrum Lingwistyki Stosowanej (Center for Applied Linguistics) w Waszyngtonie opracowano w 1970 roku zbiór nazw etnolektów, który zawierał około 18 000 pozycji [5]. Niestety trudno jest powiedzieć, ile jest rzeczywiście istniejących etnolektów świata, bowiem na rozważanej liście poszczególne etnolekty mogą występować pod kilkoma alternatywnymi nazwami, pewne etnolekty, które są jeszcze nieznanie nauce (wciąż w wielu zakątkach świata odkrywa się nowe języki), z oczywistego powodu się tam nie znalazły, a poza tym pewne nazwy mogły zostać umieszczone na tej liście przez przypadek, czy też niekompetencję, która w konfrontacji z tak ogromnym i tak różnorodnym materiałem jest sprawą wręcz nieuniknioną [5].

Po uporządkowaniu etnolektów według liczby ich użytkowników, zawierająca je lista prezentuje się następująco (dla każdego z etnolektów podano nazwę kraju, z którego pochodzi i liczbę jego użytkowników) [8].

- 1) chiński (dialekt mandaryński) – Chiny – 885 mln
- 2) hiszpański – Hiszpania – 332 mln
- 3) angielski – Wielka Brytania – 322 mln
- 4) bengalski – Bangladesz – 189 mln
- 5) hindi – Indie – 182 mln
- 6) portugalski – Portugalia – 170 mln
- 7) rosyjski – Rosja – 170 mln
- 8) japoński – Japonia – 125 mln
- 9) niemiecki – Niemcy – 98 mln
- 10) chiński (dialekt Wu) – Chiny 77 mln
- 11) jawański – Indonezja – 76 mln
- 12) koreański – Korea – 75 mln
- 13) francuski – Francja – 72 mln
- 14) wietnamski – Wietnam – 68 mln
- 15) telugu – Indie – 67 mln
- 16) chiński (dialekt Yue) – 66 mln
- 17) marati – Indie – 65 mln
- 18) tamilski – Indie – 63 mln
- 19) turecki – Turcja – 59 mln
- 20) urdu – Pakistan – 58 mln
- 21) chiński (dialekt Min Nan) – Chiny – 49 mln
- 22) chiński (dialekt Jinyu) – Chiny – 45 mln
- 23) gudżarati – Indie – 44 mln
- 24) polski – Polska – 44 mln
- 25) arabski (dialekt egipski) – Egipt – 43 mln
- 26) ukraiński – Ukraina – 41 mln
- 27) włoski – Włochy – 37 mln
- 28) chiński (dialekt Xiang) – Chiny – 36 mln
- 29) malajalam – Indie – 34 mln
- 30) chiński (dialekt Hakka) – Chiny – 34 mln
- 31) kannada – Indie – 34 mln
- 32) orija – Indie – 31 mln
- 33) pendżabi (dialekt zachodni) – Pakistan – 30 mln
- 34) sunda – Indonezja – 27 mln

- 35) pendźabi (dialekt wschodni) – Pakistan – 26 mln
- 36) rumuński – Rumunia – 26 mln
- 37) bodźpuri – Indie – 25 mln
- 38) azerski (dialekt południowy) – Iran – 25 mln
- 39) farsi (dialekt zachodni) – Iran – 24 mln
- 40) maitili – Indie – 24 mln
- 41) hausa – Nigeria – 24 mln
- 42) arabski (dialekt algierski) – Algieria – 23 mln
- 43) birmański – Birma – 22 mln
- 44) serbochorwacki – Jugosławia – 21 mln
- 45) chiński (dialekt Gan) – Chiny – 21 mln
- 46) awadi – Indie – 21 mln
- 47) tajski – Tajlandia – 20 mln
- 48) niderlandzki – Holandia – 20 mln
- 49) joruba – Nigeria – 20 mln
- 50) sindi – Pakistan – 20 mln
- 51) arabski (dialekt marokański) – Maroko – 19 mln
- 52) arabski (dialekt saidzki) – Egipt – 19 mln
- 53) uzbecki (dialekt północny) – Uzbekistan – 18 mln
- 54) malajski – Malezja – 18 mln
- 55) amharski – Etiopia – 17 mln
- 56) indonezyjski – Indonezja – 17 mln
- 57) igbo – Nigeria – 17 mln
- 58) tagalog – Filipiny – 17 mln
- 59) nepalski – Nepal – 16 mln
- 60) arabski (dialekt sudański) – Sudan – 16 mln
- 61) saraiki – Pakistan – 15 mln
- 62) cebuano – Filipiny – 15 mln
- 63) arabski (dialekt północnolewentyński) – Syria – 15 mln
- 64) tajski (dialekt północno-wschodni) – Tajlandia – 15 mln
- 65) assamski – Indie – 15 mln
- 66) węgierski – Węgry – 14 mln
- 67) chittagonian – Bangladesz – 14 mln
- 68) arabski (dialekt iracki) – Irak – 14 mln
- 69) madura – Indonezja – 14 mln
- 70) syngaleski – Sri Lanka – 13 mln
- 71) harYanvi – Indie – 13 mln
- 72) marwari – Indie – 12 mln
- 73) czeski – Czechy – 12 mln
- 74) grecki – Grecja – 12 mln
- 75) magahi – Indie – 12 mln
- 76) chhattisgarhi – Indie – 11 mln
- 77) deccan – Indie – 11 mln
- 78) chiński (dialekt Min Bei) – Chiny – 11 mln
- 79) białoruski – Białoruś – 10 mln
- 80) zhuang (dialekt północny) – Chiny – 10 mln
- 81) arabski (dialekt saudyjski) – Arabia Saudyjska – 10 mln
- 82) paszto (dialekt północny) – Pakistan – 10 mln
- 83) somali – Somalia – 9 mln
- 84) malgaski – Madagaskar – 9 mln
- 85) arabski (dialekt tunezyjski) – Tunezja – 9 mln
- 86) rwanda – Ruanda – 9 mln
- 87) zulu – Republika Południowej Afryki – 9 mln
- 88) bułgarski – Bułgaria – 9 mln
- 89) szwedzki – Szwecja – 9 mln
- 90) lombardzki – Włochy – 9 mln
- 91) oromo (dialekt zachodniocentralny) – Etiopia – 9 mln
- 92) paszto (dialekt południowy) – Afganistan – 8 mln
- 93) kazachski – Kazachstan – 8 mln
- 94) ilocano – Filipiny – 8 mln
- 95) tatarski – Rosja – 8 mln
- 96) fulfulde – Nigeria – 8 mln
- 97) arabski (dialekt jemeński) – Jemen – 8 mln
- 98) ujgurski – Chiny – 8 mln
- 99) haitański (kreolsko-francuski) – 7 mln
- 100) azerbejdżański (dialekt północny) – Azerbejdżan – 7 mln
- 101) neapolitański-Kalabryjski – Włochy – 7 mln
- 102) kmerski – Kambodża – 7 mln
- 103) farsi (dialekt wschodni) – Afganistan – 7 mln
- 104) akan – Ghana – 7 mln
- 105) hiligaynon – Filipiny – 7 mln
- 106) kurdyjski – Turcja – 7 mln
- 107) shona – Zimbabwe – 7 mln

Zamieszczona lista języków świata, ułożona w kolejności ze względu na liczbę ich użytkowników, może być dla czytelnika pewnym zaskoczeniem. Na pierwszym miejscu znajduje się na niej dialekt mandaryński języka chińskiego, na drugim język hiszpański, a język angielski uplasował się dopiero na trzeciej pozycji. Ponadto warto jeszcze zwrócić uwagę, że na rozważanej liście znalazło się również kilka innych dialektów języka chińskiego o całkiem pokaźnej liczbie użytkowników. Na liście tej nie ma również języka arabskiego, są natomiast liczne jego dialekty, bardzo zresztą zróżnicowane i często trudno wzajemnie zrozumiałe. Język polski z 44 milionami jego użytkowników (jak widać, znaczny wkład do zamieszczonej statystyki mają również nasi rodacy przebywający na emigracji) uplasował się na 24 miejscu.

Ocenia się, że liczba etnolektów świata, które mają więcej niż 100 000 użytkowników wynosi ponad 400. Z drugiej strony na całym świecie istnieje mnóstwo bardzo małych etnolektów, które mają zaledwie po kilka tysięcy, bądź jeszcze mniej użytkowników. Etnolekty takie skazane są prawdopodobnie w niezbyt odległej przyszłości na całkowite wymarcie, a los ich w świetle zachodzących w świecie procesów globalizacyjnych wydaje się już przesądzony, mimo faktu, że w wielu krajach przeznaczają się znaczne środki na ratowanie ginących języków. Między innymi, akcja reanimowania wymierających języków wspierana jest przez UNESCO. Organizacja ta opublikowała tzw. Czerwoną Księgę Języków Zagrożonych Wymarciem (*UNESCO Red Book on Endangered Languages*), w której znalazło się również około 100 języków będących w użyciu w Europie. Na liście tej znalazły się zarówno języki, które są już całkowicie wymarłe i nie znajdują się w związku z tym w użyciu w codziennej komunikacji językowej (znane są jedynie wąskiemu gronu językoznawców bądź entuzjastów, ewentualnie funkcjonują jako języki martwe – np. język mozarabicki, który jest językiem liturgii dla bardzo wąskiej wspólnoty wyznaniowej), jak również języki potencjalnie zagrożone wymarciem [9].

Spośród języków słowiańskich zamieszczonych na liście UNESCO można wymienić:

- połabski – całkowicie wymarły około roku 1750
- słowiński – wyszedł całkowicie z użycia około roku 1900
- kaszubski – około 10 000 użytkowników
- dolnołużycki – około 10 000 użytkowników
- górnołużycki – około 20 000 użytkowników
- rusiński – około 100 000 użytkowników
- chorwacki z Burgenland – około 28 000 użytkowników
- chorwacki z Molise – około 3500 użytkowników

Z kolei spośród języków germańskich lista UNESCO wymienia:

- zachodniofryzyjski – około 350 000 użytkowników
- wschodniofryzyjski – około 1000 użytkowników
- północnofryzyjski – około 10 000 użytkowników
- cymbryjski – około 5000 użytkowników
- norn – wymarły około roku 1880
- gocki krymski – wymarły w XVIII wieku

Najwięcej na wymienionej liście znalazło się języków z romańskiej grupy językowej:

- asturiański – około 100 000 użytkowników
- leoński – około 10 000 użytkowników
- aragoński – około 30 000 użytkowników
- mozarabicki – całkowicie wymarły w XVI wieku
- gaskoński – około 250 000 użytkowników
- langwedocki – około 100 000 użytkowników
- owerniacki – około 10 000 użytkowników
- limuzyński – około 10 000 użytkowników
- prowansalski – około 250 000 użytkowników
- walloński – około 320 000 użytkowników
- frankoprowansalski – około 70 000 użytkowników
- retycki – około 30 000 użytkowników
- ladyński – około 10 000 użytkowników
- friulski – około 350 000 użytkowników
- sardyński galureski – około 100 000 użytkowników
- sardyński logurdyjski – około 500 000 użytkowników
- sardyński kapidaneski – około 500 000 użytkowników

- sardyński sassareski – około 100 000 użytkowników
- istryjski – około 1000 użytkowników
- dalmatyński – ostatni użytkownik zmarł w 1898 roku
- istrorumuński – około 1000 użytkowników
- arumuński – około 100 000 użytkowników
- maglenorumuński – około 5000 użytkowników
- kataloński z Alghero – około 21 000 użytkowników
- francuski z Wysp Normandzkich – około 6000 użytkowników

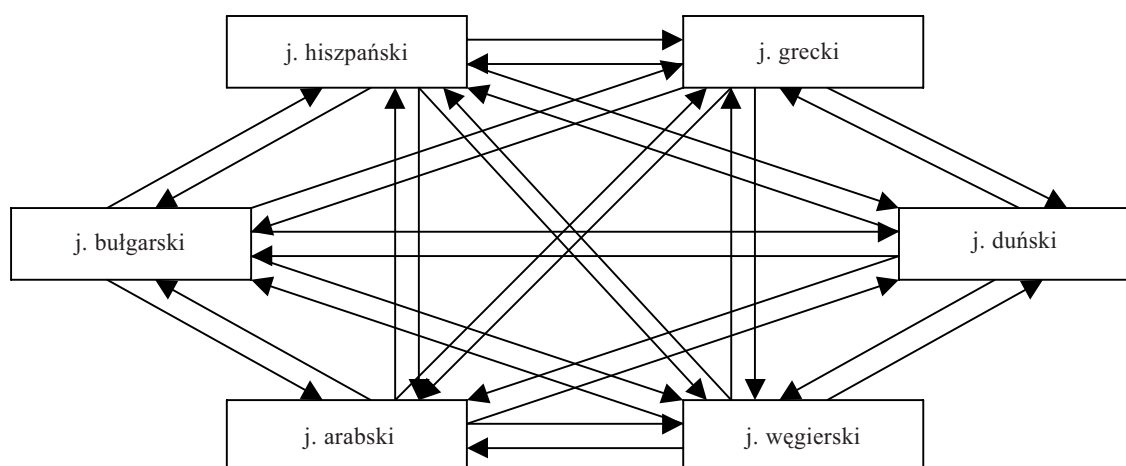
Ponadto na rozważanej liście znalazły się wszystkie języki celtyckie:

- irlandzki – około 20 000 użytkowników
- manx – ostatni użytkownik zmarł w 1974 roku
- gelicki – około 50 000 użytkowników
- walijski – około 550 000 użytkowników
- kornicki – ostatni użytkownik zmarł w 1777 roku
- bretoński – około 500 000 użytkowników

Wspomniana lista UNESCO zawiera także wszystkie języki europejskiej diaspory żydowskiej takie, jak:

- jidysz – około 300 000 użytkowników
- ladino (judeohiszpański) – około 8000 użytkowników
- shuadit (judeoprowansalski) – prawdopodobnie całkowicie wymarły
- zarphatic (judeofrancuski) – prawdopodobnie całkowicie wymarły
- italkian (judeowłoski) – prawdopodobnie całkowicie wymarły
- yevanic (judeogrecki) – około 50 użytkowników
- krimchak (judeo-krymsko-tatarski) – około 50 użytkowników

Jak wynika z zamieszczonych rozważań, liczba i różnorodność języków będących w użyciu na świecie jest wprost zdumiewająca i nawet pomimo postępującego procesu wymierania licznych małych języków (po niektórych z nich zostały tylko nieliczne świadectwa pisane – np. północno-germański język norn), w dającej się przewidzieć przyszłości z całą pewnością nie zabraknie zajęcia dla tłumaczy.



Rys. 1. Liczba programów translacji automatycznej potrzebnych do tłumaczenia w dowolnym kierunku pomiędzy parą dowolnie wybranych języków ze zbioru N języków wynosi $N \cdot (N-1)$

Można łatwo obliczyć, że liczba programów translacji automatycznej potrzebnych do tłumaczenia w dowolnym kierunku pomiędzy parą dowolnie wybranych języków ze zbioru N języków wynosi $N \cdot (N-1)$, czyli różnicę z kwadratem liczby języków. Nawet dla stosunkowo niewielkiej liczby języków liczba potrzebnych programów translacji automatycznej staje się ogromna (rys. 1).

Ze względu na tak wielką liczbę potrzebnych programów translacji automatycznej, trudno jest oczekiwać, że wszystkie te programy powstaną w niedalekiej przyszłości. Jest całkiem możliwe, że pomiędzy niektórymi parami języków (np. walijskim i maltańskim) takie programy nie powstaną nigdy, ponieważ nikt nie będzie skłonny przeznaczyć na ten cel wymaganych funduszy. Zatem jedynym środkiem zaradczym pozostaje w tym wypadku translacja wieloetapowa. Podejście takie zilustrowano na rysunku 2.

Przyglądając się uważnie rysunkowi 2, można zauważyć, że tłumaczenia tekstu zapisanego w języku bretońskim na język czeski można dokonać tylko w trzech etapach. W etapie pierwszym należy skorzystać z programu translacji automatycznej tłumaczącego z języka bretońskiego na angielski, w etapie drugim można posłużyć się programem translacji automatycznej tłumaczącym z języka angielskiego na japoński i wreszcie, w etapie trzecim, należy wykorzystać program tłumaczący z języka japońskiego na czeski. Alternatywnie w etapie drugim można było skorzystać z programu tłumaczącego z języka angielskiego na norweski, a w etapie trzecim wykorzystać program tłumaczący z języka norweskiego na czeski. Z kolei dokonanie przekładu tekstu zapisanego w języku czeskim na język angielski wymaga dwóch etapów, tzn. w etapie pierwszym dokonuje się przekładu rozważanego tekstu na język fiński, a w etapie drugim korzysta się z programu tłumaczącego z fińskiego na angielski.

Podobnie, aby przetłumaczyć tekst zapisany w języku estońskim należałoby najpierw skorzystać z programu translacji automatycznej z języka estońskiego na fiński, a następnie z programu tłumaczącego z języka fińskiego na angielski. Niestety w rozważanym na rysunku 2 systemie programów translacji automatycznej nie jest możliwe do-

konanie przekładu w odwrotnym kierunku, tzn. tekstu zapisanego w języku angielskim na język estoński.

Niewątpliwie ujemną stroną translacji wieloetapowej, oprócz być może dłuższego czasu pracy komputera, jest pogorszenie się jakości przekładu wraz ze wzrostem liczby etapów translacji.

Analizę wpływu liczby etapów translacji na jakość przekładu można zasymulować w następujący sposób. Na wejście programu translacji automatycznej „English Translator 2” można zadać pewien angielski tekst. Program „English Translator 2” dokona tłumaczenia tego tekstu na język polski. Następnie otrzymany polski tekst można ponownie podać na wejście programu „English Translator 2”, który tym razem pracuje w trybie tłumaczenia z języka polskiego na angielski. Otrzymany na tej drodze angielski tekst stanowi parafrazę oryginalnego tekstu angielskiego. Dokonując porównania tekstu oryginalnego i jego parafrazy, można ocenić jakość dwuetapowej translacji automatycznej.

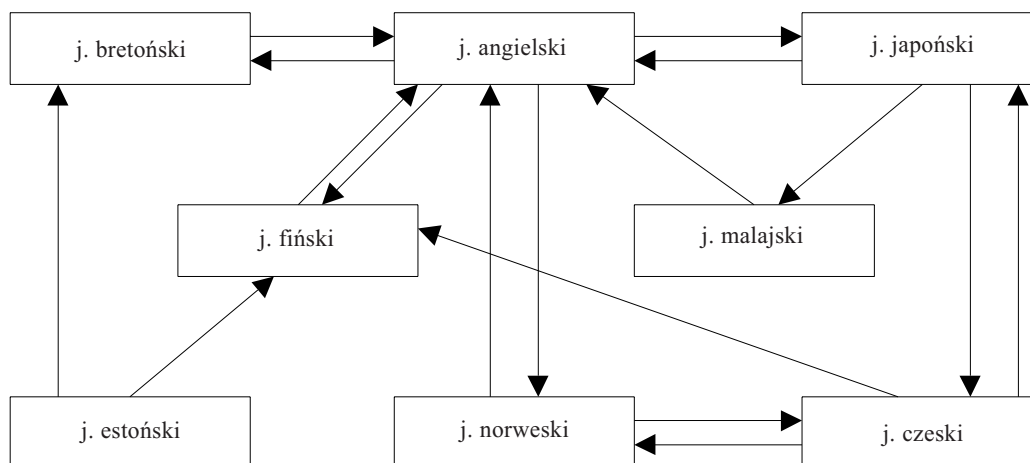
Rozważany proces można następnie powtórzyć wielokrotnie, otrzymując w ten sposób drugą, trzecią, czwartą, piątą itd. parafrazę tekstu oryginalnego. Takie postępowanie pozwala na dokonanie oceny wpływu liczby etapów translacji automatycznej na jakość tekstu wynikowego, będącego produktem takiego wieloetapowego tłumaczenia.

Autor wykonał następujący eksperyment, w którym na wejście programu translacji automatycznej „English Translator 2” został podany następujący urywek tekstu opowiadania Edgara Allana Poe'a zatytułowanego *Hop-Frog*:

When the two little friends obeyed the summons of the king, they found him sitting at his wine with the seven members of his cabinet council; but the monarch appeared to be in a very ill humor.

Rozważany tekst został zadany na wejście programu „English Translator 2” w wyniku czego otrzymano jego następujący przekład:

Kiedy dwaj małych przyjaciół słuchał wezwania do sądu króla, oni znaleźli go siadającego w jego winie z siedmiu członkami jego zgromadzenia; ale monarcha ukazała się być w bardzo chorym humorze.



Rys. 2. Translacja wieloetapowa pomiędzy parami języków, dla których nie istnieją bezpośrednie programy translacji automatycznej

Niestety, jak widać, jakość tego przekładu nie jest najlepszej jakości. Jednak mimo wszystko tekst ten został zadany na wejście modułu programu „English Translator 2” tłumaczącego z języka polskiego na angielski. Otrzymano w ten sposób następującą parafrazę angielskiego oryginału:

When two babes of friends he listened calls to court of king, they found sitting him into him fault with seven members of his meeting; but the monarch showed to be herself in ill humor very.

Jak można się było spodziewać, parafraza tekstu angielskiego dość mocno odbiega od swego pierwowzoru. Jednakże nie zważając na to, autor zadał rozważany angielski tekst na wejście programu „English Translator 2”, który przetłumaczył go ponownie na język polski w następujący sposób:

Kiedy dwa dzieciniek przyjaciół niego słuchała rozmowy zalecić się króla, oni znaleźli siadanie go do niego zwarzia z siedmioma członkami jego spotkania; ale monarcha pokazała być się w chory humor bardzo.

Powyższy polski tekst został ponownie podany na wejście programu „English Translator 2”, który przetłumaczył go na język angielski. Otrzymano w ten sposób drugą parafrazę angielskiego oryginału:

When two babies of friends of him she listened conversations to recommend herself king, they found sitting of him to him of short-circuit with seven members of his meeting; but monarch showed to be herself into ill humor very.

Niestety, druga parafraza tekstu angielskiego już w niewielkim stopniu przypomina swój oryginał, a degradacja jakości tekstu jest znaczna. Pomimo to, autor wykonał jeszcze jeden eksperyment i zadał powyższy tekst na wejście programu „English Translator 2”, który przetłumaczył go na język polski w następujący sposób:

Kiedy dwa dzieci przyjaciół niego niej słuchało konwersacje polecieć się króla, oni znaleźli siadanie niego do niego spięcia z siedmioma członkami jego spotkania, ale monarcha pokazała być się do chory humor bardzo.

Otrzymany w wyniku pracy programu „English Translator 2” polski tekst jest już praktycznie całkowicie niezrozumiały.

Po zadaniu go na wejście programu „English Translator 2” otrzymano trzecią parafrazę angielskiego oryginału:

When two children of friends of him of her it listened conversations to recommend itself king, they found sitting of him to him of short-circuit with seven members of his meeting; but monarch showed to be herself to ill humor very.

Jak widać, związek trzeciej parafrazy tekstu angielskiego z jego oryginałem jest bardzo luźny. Ponadto obserwując sposób zniekształcania znaczenia tekstu w kolejnych etapach translacji, można się zastanowić, czy aby w odniesieniu do programu „English Translator 2” nie są prawdziwe dwa następujące twierdzenia.

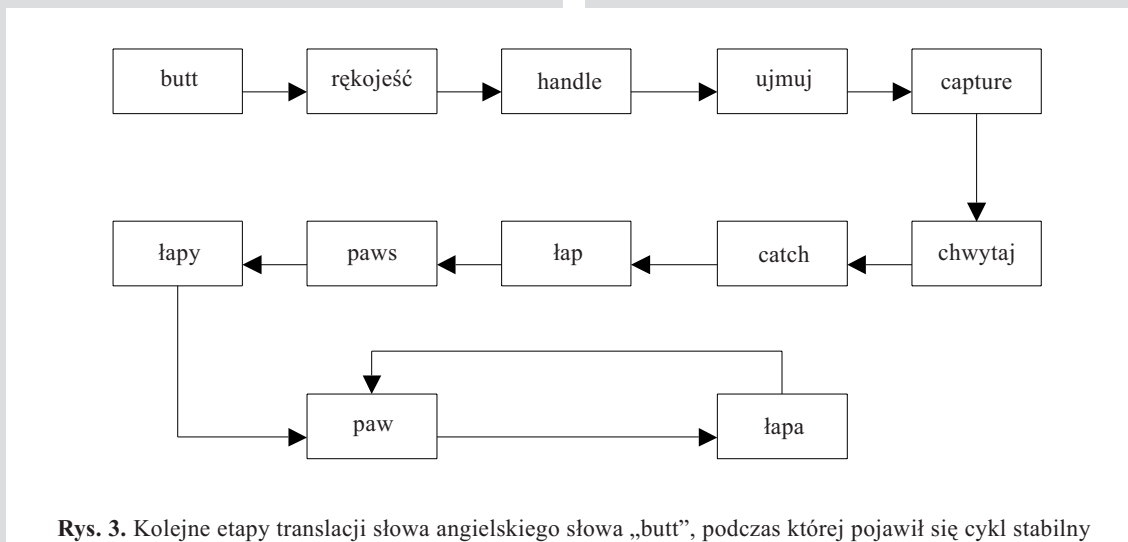
Twierdzenie 1

Dla dowolnego ciągu słów C liczącego N słów istnieje taka liczba K etapów parafrazowania ciągu C przy użyciu programu „English Translator 2”, że w wyniku otrzymuje się ponownie ciąg C .

Twierdzenie 2

Dla dwóch dowolnych ciągów słów C_1 i C_2 , z których każdy liczy N słów, istnieje taka liczba K etapów parafrazowania ciągu C_1 , że w wyniku otrzymuje się ciąg C_2 .

Dowód powyższych twierdzeń pozostaje jak dotąd problemem otwartym. Natomiast autor wykonał jeszcze jeden interesujący eksperyment związany z programem „English Translator 2” – mianowicie zadał na jego wejście wyrwane z kontekstu angielskie słowo „butt”, które zostało przetłumaczone na język polski przez rozważany program, jako „rekojeść”. Następnie słowo „rekojeść” zostało ponownie zadane na wejście programu „English Translator 2”, który tym razem przetłumaczył je jako „handle”. W kolejnym etapie słowo „handle” zostało przetłumaczone przez program „English Translator 2” jako „ujmuj”. Powyższy proces powtórzono wiele razy, aż pojawił się cykl stabilny, w którym angielskie słowo „paw” jest tłumaczone na język polski jako „łapa”, a słowo „łapa” jest tłumaczone na język angielski jako „paw”. Kolejne etapy translacji angielskiego słowa „butt” zostały zilustrowane na rysunku 3.



Rys. 3. Kolejne etapy translacji słowa angielskiego słowa „butt”, podczas której pojawił się cykl stabilny

Wniosek, jaki wypływa z zamieszczonych obserwacji, jest taki, że należy być bardzo ostrożnym w odbiorze tekstów otrzymanych za pośrednictwem wieloetapowej translacji automatycznej. Jeśli w przypadku dwuetapowej translacji automatycznej uzyskane teksty zachowują jeszcze znaczną porcję semantyki tekstu oryginału, to w przypadku dodania kolejnych etapów translacji jest rzeczą wysoce wątpliwą, czy uzyskany na tej drodze przekład będzie miał jeszcze jakąkolwiek praktyczną wartość.

5. PROPOZYCJE DALSZYCH KIERUNKÓW BADAŃ

Uzyskane wstępne wyniki badań polegających na zastosowaniu techniki PBMT do tłumaczenia na język polski tekstów zapisanych w różnych językach świata rodzą nadzieję na opracowanie programów translacji automatycznej dla języka ogólnego, charakteryzujących się stosunkowo wysoką jakością przekładu. Nie oznacza to jednak wcale, że opracowana przez autora technika translacji automatycznej oparta na wzorcach PBMT jest swego rodzaju panaceum na wszelkie trudności, jakie mogą wystąpić podczas prób zautomatyzowania przekładu. Wręcz przeciwnie wiele problemów pozostaje nadal otwartych, a ich główne źródło autor upatruje w różnicach systemów gramatycznych oraz różnicach systemów słownictwa, jakie występują pomiędzy różnymi językami świata.

W celu bliższego wyjaśnienia tych zagadnień zostaną porównane systemy zaimków osobowych w wybranych językach, głównie dlatego, że różnice pomiędzy poszczególnymi językami są tutaj znaczne, a ich wpływ na jakość translacji automatycznej jest szczególnie widoczny.

Jeden z najprostszych systemów zaimków osobowych występuje w języku węgierskim i ma postać, jak przedstawiono w tabeli 1.

Tabela 1. System zaimków osobowych języka węgierskiego

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	én	mi
Druga	te	ti
Trzecia	ő	ők

Swą prostotę system zaimków osobowych języka węgierskiego zawdzięcza głównie temu, że język ten nie zna w ogóle pojęcia rodzaju gramatycznego. Równie prosty system zaimków osobowych występuje w języku perskim, w którym również nie występuje kategoria rodzaju gramatycznego (tab. 2).

Tabela 2. System zaimków osobowych języka perskiego

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	من	ما
Druga	تو	شما
Trzecia	او	آنها

Nieco bardziej skomplikowany system zaimków osobowych występuje w języku szwedzkim, gdzie w trzeciej osobie liczby pojedynczej występują osobne formy dla rodzaju męskiego i żeńskiego (tab. 3).

Tabela 3. System zaimków osobowych języka szwedzkiego

Osoba	Liczba pojedyncza	Liczba mnoga
Oierwsza	jag	vi
Druga	du	ni
Trzecia (rodzaj męski)	han	de
Trzecia (rodzaj żeński)	hon	de

Z kolei w języku norweskim występuje jeszcze dodatkowa forma zaimka osobowego dla trzeciej osoby liczby pojedynczej odpowiadająca rodzajowi nijakiemu.

System zaimków osobowych języka norweskiego został przedstawiony w tabeli 4.

Tabela 4. System zaimków osobowych języka norweskiego

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	jæg	vi
Druga	du	dere
Trzecia (rodzaj męski)	han	de
Trzecia (rodzaj żeński)	hun	de
Trzecia (rodzaj nijaki)	det	de

Jeszcze bardziej rozbudowany system zaimków osobowych można znaleźć w języku islandzkim, gdzie odrębne formy dla rodzaju męskiego, żeńskiego i nijakiego występują zarówno w liczbie pojedynczej, jak i mnogiej (tab. 5).

Tabela 5. System zaimków osobowych języka islandzkiego

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza	ég	við
Druga	Þú	Þið
Trzecia (rodzaj męski)	hann	Þeir
Trzecia (rodzaj żeński)	hún	Þau
Trzecia (rodzaj nijaki)	Það	Þær

Niemniej osobliwy jest system zaimków osobowych języka hiszpańskiego, gdzie występują osobne formy dla rodzaju męskiego i żeńskiego również w pierwszej i drugiej osobie liczby mnogiej (tab. 6).

Tabela 6. System zaimków osobowych języka hiszpańskiego

Osoba	Liczba pojedyncza	Liczba mnoga
Pierwsza (rodzaj męski)	yo	nosotros
Pierwsza (rodzaj żeński)	yo	nosotras
Druga (rodzaj męski)	tú	vosotros
Druga (rodzaj żeński)	tú	vosotras
Trzecia (rodzaj męski)	él	ellos
Trzecia (rodzaj żeński)	ella	ellas

O wiele bardziej skomplikowany system zaimków osobowych występuje w języku arabskim, gdzie oprócz form właściwych dla liczby pojedynczej i mnogiej występują jeszcze dodatkowe formy dla liczby podwójnej (tab. 7).

Tabela 7. System zaimków osobowych języka arabskiego

Osoba	Liczba pojedyncza	Liczba podwójna	Liczba mnoga
Pierwsza	أنا	نحن	نحن
Druga (rodzaj męski)	أنت	أنتما	أنتم
Druga (rodzaj żeński)	أنت	أنتما	أنتن
Trzecia (rodzaj męski)	هو	هما	هم
Trzecia (rodzaj żeński)	هي	هما	هن

Już pobieżny rzut oka na zawartość tabel 1–7 pozwala się zorientować, że przejście z systemu zaimków osobowych jednego języka na drugi nie jest wcale sprawą łatwą i często wymaga dysponowania dodatkową wiedzą o naturze pozalingwistycznej. Ponadto wybór takiej czy innej formy zaimka osobowego przenosi się automatycznie na wybór odpowiedniej formy fleksyjnej innych części mowy występujących w tłumaczonym zdaniu: czasowników, przymiotników, liczebników itp.

Na przykład węgierski zaimek osobowy „ő” można przełożyć na język polski jako „on”, „ona” bądź „ono”. Aby wybrać odpowiedni wariant tłumaczenia, trzeba z kontekstu wypowiedzi wyłowić informację na temat tego, czego bądź kogo ów tłumaczony zaimek „ő” dotyczy. Wnioskowanie takie jest zwykle dziecinnie proste dla człowieka tłumacza, niestety jego próby ujęcia w sztywne reguły, zalgorytmizowania i zapisania w postaci programu komputerowego wciąż jeszcze pozostają obiektem badań. Różnice pomiędzy systemami zaimków osobowych języka węgierskiego i polskiego powodują, że następujące węgierskie zdanie:

Ő magas.

może zostać przetłumaczone, w zależności od kontekstu, w którym występuje, aż na trzy różne sposoby:

On jest wysoki.

Ona jest wysoka.

Ono jest wysokie.

Jak widać, wszystko zależy od tego, czy w tłumaczonym zdaniu mowa jest o mężczyźnie, kobiecie bądź dziecku. Dla człowieka tłumacza wybór właściwego ekwiwalentu znaczeniowego będzie zapewne sprawą oczywistą. Niestety nic nie jest oczywiste dla programu realizowanego przez komputer.

Podobna sytuacja występuje podczas próby przetłumaczenia angielskiego zaimka osobowego „they” na język arabski. Aby tego dokonać w sposób prawidłowy, należy dysponować informacją na temat tego, ilu osobników się pod rozważanym zaimkiem kryje. Jeżeli tylko dwóch, wówczas należy posłużyć się arabskim zaimkiem osobowym „هما”. Jeżeli jest ich więcej, wówczas trzeba dodatkowo wiedzieć, czy są to mężczyźni lub ewentualnie grupa mieszana złożona z mężczyzn i kobiet – wtedy należy użyć arabskiego zaimka „هم”, czy też są to same kobiety – wtedy należy użyć arabskiego zaimka „هن”. Z wymienionych powodów angielskie zdanie:

They are in Egypt.

może zostać przetłumaczone na język arabski na trzy sposoby:

هما في مصر .

هم في مصر .

هن في مصر .

Podobnie tłumaczenie na język hebrajski polskiego zapytania:

Kim ty jesteś?

będzie brzmiało:

מי אתה

lub

מי את

w zależności od tego, czy jest ono kierowane do mężczyzny, czy też do kobiety.

Analogicznie – chcąc przełożyć na język grecki proste polskie pytanie:

Kto to jest?

tłumacz ma do wyboru aż trzy możliwości:

Ποιος είναι αυτός;

Ποια είναι αυτή;

Ποιο είναι αυτό;

W rozważanym przykładzie wszystko zależy od tego, czy pytamy o mężczyznę, kobietę czy dziecko.

Jak już wspomniano, rozważane różnice zachodzące pomiędzy systemami gramatycznymi różnych języków nie stanowią w większości wypadków jakichś szczególnych trudności dla ludzi tłumaczy, którzy kierując się informacją pozakontekstową, ogólną wiedzą o świecie, inteligencją i zdrowym rozsądkiem, potrafią zwykle całkiem niezłe sobie w takich sytuacjach poradzić. Niestety, jeżeli chodzi o systemy translacji automatycznej, problem pozostaje ciągle otwarty, a wbudowanie w program komputerowy czegoś, co przypominałoby najwykleszy w świecie ludzki zdrowy rozsądek, chyba na zawsze pozostanie domeną popularnych powieści z gatunku *science fiction*.

Innym rodzajem trudności występujących podczas translacji są różnice w słownictwie języków źródłowego i docelowego przekładu. Okazuje się, że każdy z języków naturalnych w inny sposób dzieli otaczającą człowieka rzeczywistość na jednostki leksykalne. W wyniku tego, poszczególne słowa nie mają w drugim języku swoich dokładnych odpowiedników, a ich tłumaczenie musi być z konieczności tylko pewnym przybliżeniem pierwotnych znaczeń. Jest rzeczą naprawdę intrygującą, że użytkownicy dwóch różnych języków postawieni w obliczu dokładnie takiej samej sytuacji językowej, w formułowanych przez siebie wypowiedziach często zwracają uwagę na zupełnie odmienne elementy tej sytuacji dotyczące [5]. Dzieje się tak, ponieważ każdy z języków dzieli otaczającą jego użytkowników rzeczywistość na odmienne fragmenty, którym następnie przypisywane są odrębne jednostki leksykalne. Proces ten został zilustrowany na rysunku 4.

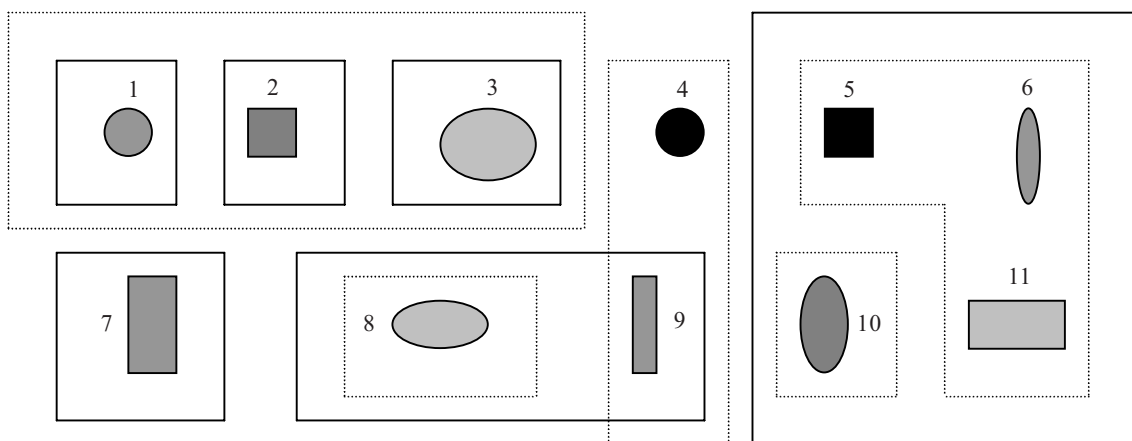
Na rysunku 4 przedstawiono w sposób poglądowy jedenaście różnych obiektów, którymi mogą być zarówno przedmioty ze świata rzeczywistego, jak również pojęcia abstrakcyjne. Poszczególne języki łączą te przedmioty w klasy, którym przypisują jednostki leksykalne, jednakże robią to na zupełnie odmienne sposoby. Dla ustalenia uwagi, podział na jednostki leksykalne w pewnym języku A został zaznaczony linią ciągłą, a w języku B linią przerywaną. Na rysunku 4 widać, że język A przypisuje każdemu

z obiektów 1, 2 i 3 odmienne jednostki leksykalne, podczas gdy język B obiekty te określa tylko jedną, wspólną jednostką leksykalną. Można powiedzieć, że w tym przypadku język A odznacza się większą szczegółowością opisu rzeczywistości niż język B. Nie jest to jednakże jakąś bezwzględną regułą, bowiem język A określa wspólną jednostką leksykalną obiekty 5, 6, 10 i 11, podczas gdy język B rozróżnia tutaj dwie odrębne jednostki leksykalne. W języku B obiekty 4 i 9 określane są wspólną jednostką leksykalną, podczas gdy w języku A obiektowi 4 w ogóle nie jest przypisana żadna jednostka leksykalna, a obiekt 9 wraz z obiektem 8 stanowi jedną jednostkę leksykalną. Należy zauważyć, że w języku B obiekt 8 to całkowicie odrębna jednostka leksykalna. Również obiekt 7 stanowi w języku A odrębną jednostkę leksykalną, podczas gdy w języku B obiekt ten nie został zaliczony do żadnej klasy.

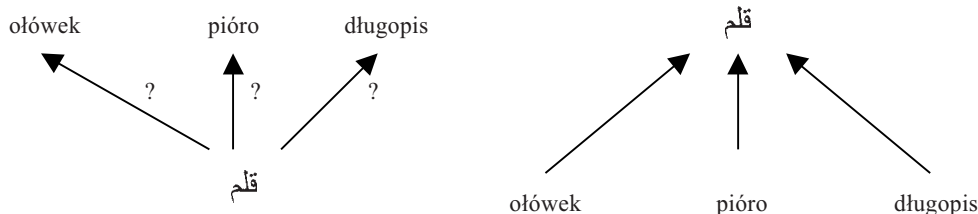
Należy zauważyć, że problemy z automatyzacją translacji powstają nieuchronnie podczas przechodzenia z systemu językowego w danej sytuacji leksykalnie uboższego na system leksykalnie bogatszy. Nie ma natomiast żadnych problemów przy przechodzeniu w kierunku przeciwnym – z systemu leksykalnie bogatszego na uboższy. Zostało to zilustrowane na rysunku 5, gdzie przedstawiono sposób tłumaczenia na język polski arabskiego słowa „قلم”, które zależnie od kontekstu może oznaczać: „pióro”, „długopis” lub „ołówek”. System komputerowy może mieć poważne problemy z wyborem odpowiedniego polskiego ekwiwalentu rozważanego arabskiego słowa. Nie ma natomiast żadnego problemu, gdy chce się przetłumaczyć na arabski jedno z polskich słów: „ołówek”, „pióro” lub „długopis”.

Człowiek tłumacz dokonując tego typu wyborów, kieruje się informacją zawartą w kontekście wypowiedzi. Jednakże zautomatyzowanie tego typu wnioskowania i jego komputerowa realizacja okazuje się w praktyce niezwykle trudna – wręcz niemożliwa.

Przy bliższym zbadaniu systemów leksykalnych różnych języków świata okazuje się, że na komputerowy program automatycznego tłumacza mogą czyhać rozmaite pułapki.



Rys. 4. Podział rzeczywistości na jednostki leksykalne w dwóch różnych językach (Objaśnienia w tekście)



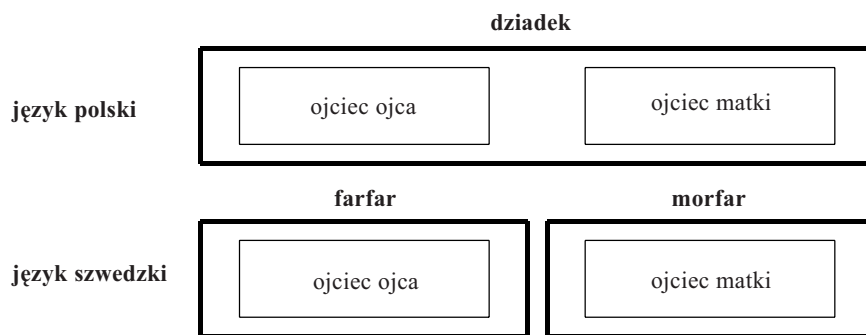
Rys. 5. Podczas automatyzacji translacji problemy występują jedynie w przypadku przechodzenia z systemu leksykalnie bogatszego na uboższy

Bardzo dobry i pouczający pod tym względem przykład pochodzi z języka szwedzkiego. Okazuje się bowiem, że przetłumaczenie na język szwedzki polskiego słowa „dziadek” nie jest wcale sprawą prostą, bowiem użytkownik języka szwedzkiego chce dodatkowo wiedzieć, czy ów „dziadek” jest z strony matki – wówczas należy użyć słowa „morfar”, czy też ze strony ojca – wówczas należy posłużyć się słowem „farfar”. Należy tutaj z całą mocą podkreślić, że problemu tego nie da się żadną miarą ominąć, ponieważ język szwedzki zna tylko słowa „morfar” i „farfar”, nie zna natomiast słowa łączącego obie wymienione kategorie znaczeniowe. Jako ciekawostkę można podać, że takie uniwersalne słowo, brzmiące „bestefar” istnieje w blisko spokrewnionych ze szwedzkim językach norweskim i duńskim.

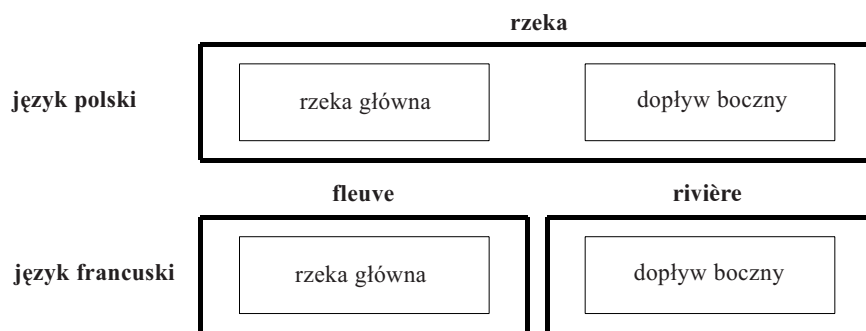
Sytuacja, jaka powstaje podczas translacji na język szwedzki polskiego słowa „dziadek”, została schematycznie przedstawiona na rysunku 6.

Podobnie nie jest sprawą łatwą dokonanie przekładu na język francuski polskiego słowa „rzeka”. Okazuje się bowiem, że użytkownik języka francuskiego będzie chciał jeszcze dodatkowo wiedzieć, czy jest to główna rzeka, mająca bezpośrednie ujście do morza, czy też jest to rzeka nie mająca bezpośredniego połączenia z morzem, która stanowi jedynie dopływ jakiejś innej większej rzeki. W pierwszym z wymienionych przypadków należy posłużyć się francuskim słowem „fleuve”, natomiast w drugim przypadku właściwym będzie użycie słowa „rivière”. Jak widać, aby komputer mógł tłumaczyć poprawnie z języka polskiego na francuski, program translacji automatycznej musiałby mieć wbudowaną dodatkowo bazę wiedzy geograficznej. Sytuacja, która powstaje podczas przekładu na język francuski polskiego rzeczownika „rzeka”, została zilustrowana na rysunku 7.

Problemy podobnej natury powstają podczas próby przetłumaczenia na język węgierski polskiego słowa „brat”.



Rys. 6. Aby przetłumaczyć na język szwedzki polskie słowo „dziadek” należy dokonać właściwego wyboru pomiędzy „morfar” i „farfar”

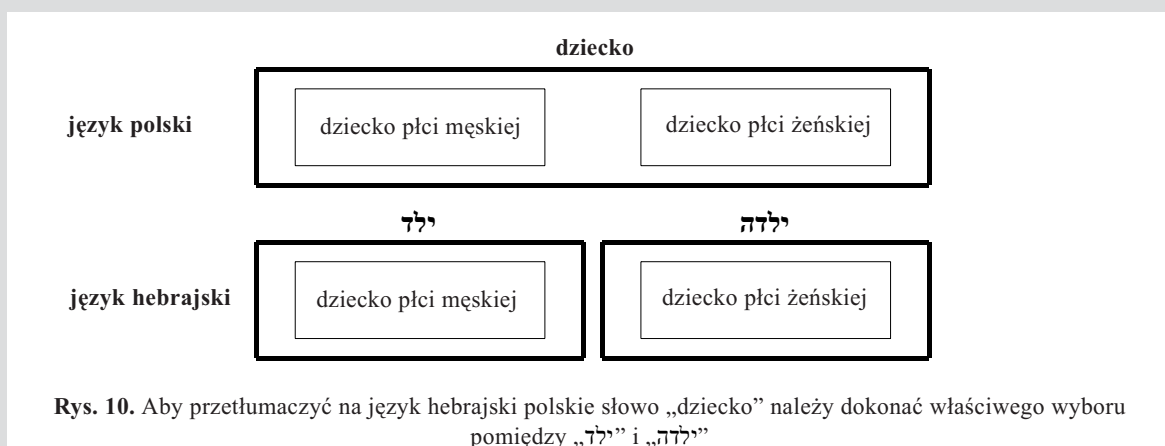
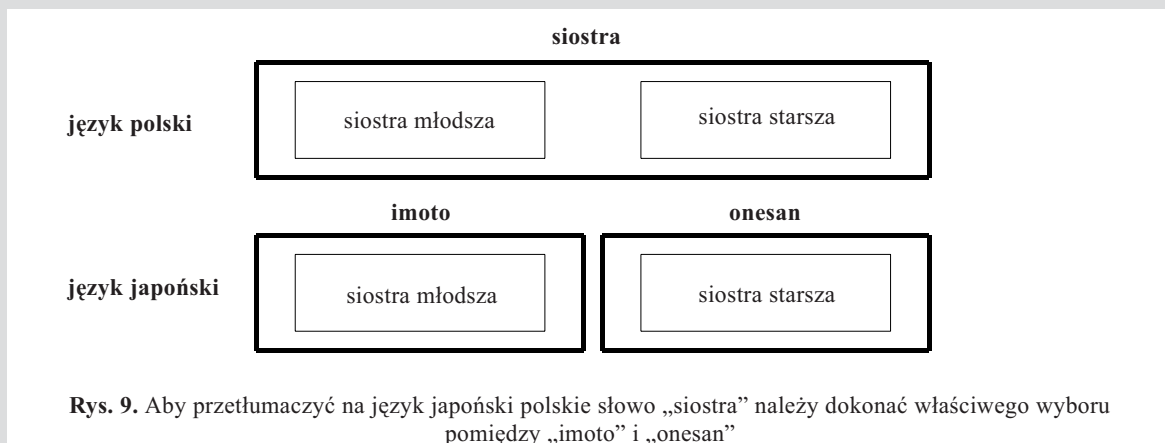
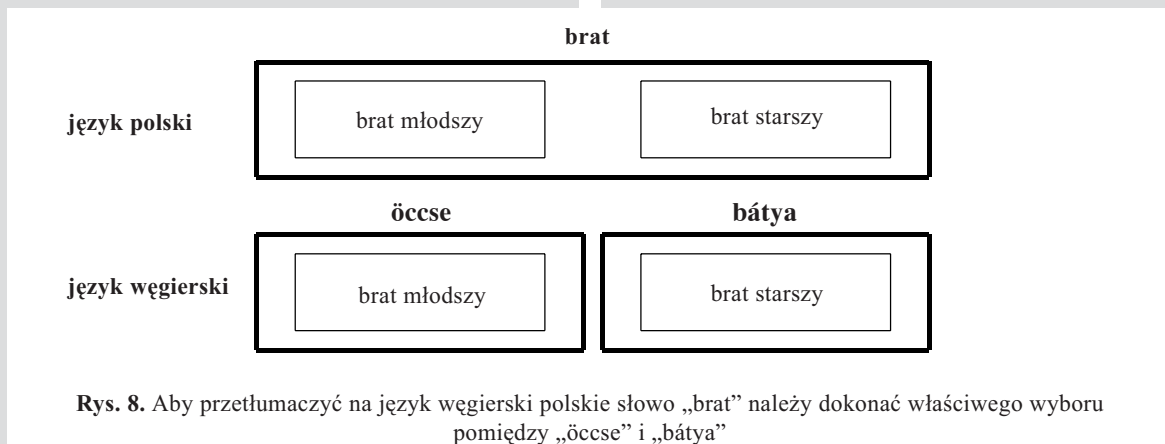


Rys. 7. Aby przetłumaczyć na język francuski polskie słowo „rzeka” należy dokonać właściwego wyboru pomiędzy „fleuve” i „rivière”

Dzieje się tak dlatego, ponieważ użytkownik języka węgierskiego z jakichś bliżej nieokreślonych powodów chce koniecznie dodatkowo wiedzieć, czy brat, o którym jest mowa, jest bratem starszym czy młodszym. Jeżeli jest to brat młodszy, wówczas należy użyć słowa „öccse”, natomiast jeżeli jest starszy, należy posłużyć się słowem „bátya”. Człowiek tłumacz dokonujący przekładu na język węgierski zapewne jakoś wyłowi z kontekstu informację dotyczącą starszeństwa w rodzinie i w związku z tym dokona prawidłowego wyboru węgierskiego ekwiwalentu polskiego słowa „brat”. Jednakże napisanie programu komputerowego, który by wykonywał rozważaną czynność automatycznie, zdaje się obecnie przerastać nawet najśmielsze wyobrażenia. Tłumaczenie na język węgierski polskiego słowa „brat” zostało zilustrowane na rysunku 8.

Z identyczną sytuacją ma się do czynienia podczas próby przekładu na język japoński polskiego rzeczownika „siostra”. W tym wypadku również trzeba dodatkowo wiedzieć, czy owa siostra jest młodsza czy starsza. Jeżeli jest młodsza, właściwym przekładem będzie japońskie słowo „imoto”, natomiast jeżeli starsza – „onesan”. Rozważana sytuacja została zilustrowana na rysunku 9.

Kolejny pouczający przykład pochodzi z języka hebrajskiego. Otóż aby przetłumaczyć na język hebrajski polskie słowo „dziecko” trzeba posiadać dodatkową pozalingwistyczną informację o płci tego dziecka. Jeżeli dziecko jest płci męskiej należy posłużyć się hebrajskim słowem „יֵלֵד”, w przeciwnym wypadku właściwym słowem będzie „יְלֵדָה”. Rozważane zawłóści języka hebrajskiego zostały zilustrowane na rysunku 10.



Również potencjalną przeszkodę na drodze do stworzenia w pełni automatycznych systemów translacji automatycznej zastępujących całkowicie człowieka tłumacza mogą stanowić niektóre wyrażenia idiomatyczne, które w zależności od kontekstu ich występowania mogą być również rozumiane dosłownie. Językiem szczególnie obfitującym w tego typu konstrukcje idiomatyczne jest między innymi afrykański język hausa należący do afroazjatyckiej rodziny językowej [10]. Na przykład następujące zdanie zapisane w języku hausa:

Gari ya yi kyau.

w dosłownym przekładzie znaczy:

Miasto jest ładne.

Jednakże zdanie to posiada jeszcze jedno, idiomatyczne znaczenie:

Jest ładna pogoda.

Człowiek tłumacz dokonujący przekładu takiego zdania zapewne będzie wiedział z jego kontekstu, czy osoba je wypowiadająca mówi dosłownie (ma na myśli rzeczywiście jakieś miasto, które jest ładne), czy też posługuje się przenośnią, aby wyrazić swoją opinie o pogodzie. Niestety napisanie programu komputerowego, który przeprowadzałby w sposób całkowicie automatyczny tego typu wnioskowanie, przy obecnym stanie nauki i techniki, przechodzi nawet najśmielsze wyobrażenia.

Podobnych przykładów zaczerpniętych z różnych języków świata można wymienić bez liku, a wszystkie dowodzą, jak niełatwym zadaniem jest zautomatyzowanie translacji. Z tych właśnie powodów komputerowy tłumacz równie sprawny jak człowiek tłumacz profesjonalista prawdopodobnie na zawsze pozostanie niedoścignionym ideałem, swoistym św. Graalem poszukiwań badawczych. Stwierdzenie powyższe wcale nie oznacza, że prowadzenie badań nad translacją automatyczną pozbawione jest sensu, czego oczywistym dowodem jest fakt dokonania się dużego postępu w tej dziedzinie w okresie ostatnich kilkunastu lat. Dostępne obecnie programy translacji automatycznej działają przecież o wiele sprawniej i skuteczniej niż ich poprzednicy z lat 80. XX w. Zatem dalsze zintensyfikowanie badań nad translacją automatyczną powinno pozwolić na jeszcze większe przybliżenie się efektów pracy programów translacji automatycznej do rozważanego niedoścignionego ideału.

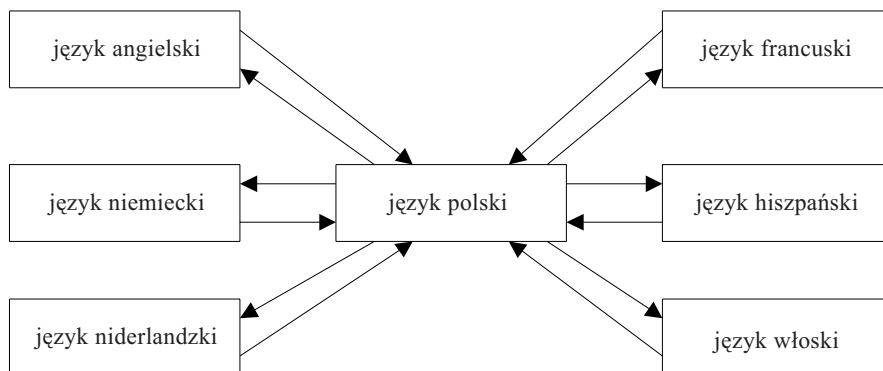
6. DALSZY PERSPEKTYWY ROZWOJU METODY PBMT

Przewidywane przez autora dalsze badania prowadzone w dziedzinie systemów translacji automatycznej koncentrować się będą w trzech różnych kierunkach.

Po pierwsze – bazy danych zawierające wzorce translacyjne dla rozważanych w niniejszej monografii systemów translacji automatycznej będą systematycznie rozbudowywane i uzupełniane o nowe jednostki leksykalne. Trzeba w tym miejscu podkreślić, że proces budowy takich baz danych nie ma nigdy końca, ponieważ możliwych do utworzenia w danym języku fraz jest praktycznie nieskończenie wiele, a ponadto każdy żywy język nieustannie się rozwija poprzez tworzenie, bądź zapożyczanie z innych języków, nowych wyrazów i związków frazeologicznych.

Drugim kierunkiem badań, który zostanie podjęty przez autora, będzie próba zastosowania opracowanej metody translacji automatycznej opartej na wzorcach PBMT do innych języków, które nie zostały do tej pory uwzględnione. Prace będą koncentrowały się zwłaszcza na nie uwzględnionych dotychczas przez autora dużych i ważnych językach indoeuropejskich oraz na językach nie należących do indoeuropejskiej rodziny językowej. Autor ma tutaj na myśli zwłaszcza duże i ważne języki Afryki i Azji, między innymi takie, jak: suahili, hausa, urdu, hindi, pendzabski, tajski, wietnamski, indonezyjski, malajski, chiński, japoński, koreański i inne.

W związku z trzecim proponowanym kierunkiem badań, autor planuje podjęcie próby zastosowania opracowanej przez siebie metody translacji automatycznej opartej na wzorcach w systemach, które dokonywałyby automatycznych przekładów z języka polskiego na inne języki. Autor podjął już wstępne próby dostosowania opracowanej przez siebie metody translacji automatycznej, tak aby nadawała się ona do tłumaczenia z języka polskiego na języki takie, jak: angielski, niemiecki, niderlandzki, francuski, hiszpański i włoski. Dzięki temu język polski stałby się językiem pośredniczącym (interlingua), za pomocą którego można byłoby dokonywać przekładów pomiędzy dowolnie wybraną parą języków i to w dowolnym kierunku, co zostało przedstawione na rysunku 11.



Rys. 11. Język polski w roli języka pośredniczącego w systemie tłumaczącym pomiędzy językami: angielskim, niemieckim, niderlandzkim, francuskim, hiszpańskim i włoskim

Jednakże każdy z rozważanych tutaj języków wymaga indywidualnego dopasowania metody translacji automatycznej opartej na wzorcach do jego struktury gramatycznej.

Na przykład dla języka angielskiego, ze względu na jego ubogą fleksję, istnieją jedynie dwa typy wzorców translacyjnych. Są to odpowiednio wzorce translacyjne nie podlegające fleksji oraz wzorce translacyjne typu koniugacyjnego. Z tym, że wzorec typu koniugacyjnego dla języka angielskiego ma nieco odmienną postać niż w przypadku języka polskiego.

Wzorec translacyjny nie podlegający fleksji w systemie tłumaczącym z języka polskiego na angielski jest zamieszczony w tabeli 8.

Tabela 8. Wzorec translacyjny nie podlegający fleksji dla języka angielskiego

SOURCE
TARGET
<liczba> = VAL_N;

Z kolei wzorec translacyjny typu koniugacyjnego w systemie tłumaczącym z języka polskiego na angielski ma postać jak w tabeli 9.

Tabela 9. Wzorec translacyjny typu koniugacyjnego dla języka angielskiego

V_SOURCE	<liczba>
V_TARGET_1	1
V_TARGET_2	2
<liczba> = VAL_N;	

Sposób działania systemu translacji automatycznej tłumaczącego z języka polskiego na angielski został zilustrowany na następującym przykładzie, w który zadaniem systemu było dokonanie przekładu na język angielski zamieszczonego poniżej polskiego zdania:

Odrobina namysłu wystarcza, aby wyjaśnić niezgodność.

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce translacyjne:

- wzorec translacyjny nie podlegający fleksji:

odrobina namysłu
a very little reflection
<liczba> = 1;

- wzorec translacyjny typu koniugacyjnego:

wystarcza aby wyjaśnić	<liczba>
suffices to explain	1
suffice to explain	2

- wzorec translacyjny nie podlegający fleksji:

niezgodność
the discrepancy
<liczba> = 1;

W wyniku pracy systemu translacji automatycznej otrzymano następujący angielski przekład polskiego zdania:

A very little reflection sufficed to explain the discrepancy.

Z kolei w systemie dokonującym automatycznych przekładów z języka polskiego na niemiecki występują wzorce translacyjne typu deklinacyjnego, typu koniugacyjnego oraz wzorce translacyjne nie podlegające fleksji.

Struktura wzorca translacyjnego nie podlegającego fleksji dla języka niemieckiego została zamieszczona w tabeli 10.

Tabela 10. Wzorec translacyjny nie podlegający fleksji dla języka niemieckiego

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

Analogicznie w tabeli 11 została przedstawiona struktura wzorca translacyjnego typu deklinacyjnego dla języka niemieckiego.

Tabela 11. Wzorec translacyjny typu deklinacyjnego dla języka niemieckiego

N_SOURCE	<przypadek>
N_TARGET_1	1
N_TARGET_2	2
N_TARGET_3	3
N_TARGET_4	4
<liczba> = VAL_N; <przypadek> = VAL_C;	

Podobnie w tabeli 12 zawarto postać wzorca typu koniugacyjnego dla języka niemieckiego.

Tabela 12. Wzorec translacyjny typu koniugacyjnego dla języka niemieckiego

V_SOURCE	<liczba>
V_TARGET_1	1
V_TARGET_2	2
<liczba> = VAL_N; <przypadek> = VAL_C;	

Praca systemu translacji automatycznej dokonującego przekładów z języka polskiego na niemiecki metodą wzorców translacyjnych została zilustrowana na następującym przykładzie. Celem było dokonanie przekładu na język niemiecki następującego polskiego zdania:

Czasopisma zawsze leżą pod nowymi gazetami.

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce translacyjne:

- na początku nowego zdania atrybutowi <przypadek> przypisywana jest wartość jeden:

<przypadek> = 1;

- wzorec translacyjny typu deklinacyjnego:

czasopisma	<przypadek>
die Zeitschriften	1
der Zeitschriften	2
den Zeitschriften	3
die Zeitschriften	4
<liczba> = 2;	

- wzorec translacyjny typu koniugacyjnego:

zawsze leżą pod	<liczba>
liegt immer unter	1
liegen immer unter	2
<przypadek> = 3;	

- wzorec translacyjny typu deklinacyjnego:

nowymi gazetami	<przypadek>
die neuen Zeitungen	1
der neuen Zeitungen	2
den neuen Zeitungen	3
die neuen Zeitungen	4
<liczba> = 2;	

W wyniku pracy systemu translacji automatycznej otrzymano następujący niemiecki przekład polskiego zdania:

Die Zeitschriften liegen immer unter den neuen Zeitungen.

Natomiast w systemie translacji automatycznej tłumaczącej z języka polskiego na język niderlandzki występują tylko wzorce translacyjne nie podlegające fleksji oraz wzorce translacyjne typu deklinacyjnego.

Dla języka niderlandzkiego wzorec translacyjny nie podlegający fleksji przedstawiono w tabeli 13.

Tabela 13. Wzorec translacyjny nie podlegający fleksji dla języka niderlandzkiego

SOURCE
TARGET
<liczba> = VAL_N;

Z kolei w tabeli 14 została zamieszczona struktura wzorca translacyjnego typu koniugacyjnego dla języka niderlandzkiego.

Tabela 14. Wzorec translacyjny typu koniugacyjnego dla języka niderlandzkiego

V_SOURCE	<liczba>
V_TARGET_1	1
V_TARGET_2	2
<liczba> = VAL_N;	

Pracę systemu translacji automatycznej tłumaczącego z języka polskiego na niderlandzki zilustrowano na przykładzie następującego zdania:

W Holandii spożycie alkoholu zawsze było o wiele niższe niż w większości krajów europejskich.

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce translacyjne:

- wzorec translacyjny nie podlegający fleksji:

w
in

- wzorec translacyjny nie podlegający fleksji:

Holandii
Nederland
<liczba> = 1;

- wzorec translacyjny nie podlegający fleksji:

spożycie alkoholu
het alcoholverbruik
<liczba> = 1;

- wzorec translacyjny typu koniugacyjnego:

zawsze było o wiele niższe	<liczba>
is altijd veel lager geweest	1
zijn altijd veel lager geweest	2

- wzorec translacyjny nie podlegający fleksji:

niż
dan

- wzorec translacyjny nie podlegający fleksji:

w
in

- wzorzec translacyjny nie podlegający fleksji:

większości krajów europejskich
de meeste Europese landen
<liczba> = 2;

W wyniku pracy systemu translacji automatycznej otrzymano następujący niderlandzki przekład polskiego zdania:

In Nederland het alcoholverbruik is altijd veel lager geweest dan in de meeste Europese landen.

Z kolei w systemie translacji automatycznej tłumaczącym z języka polskiego na francuski występują wzorce translacyjne nie podlegające fleksji oraz wzorce translacyjne typu koniugacyjnego.

Wzorzec translacyjny nie podlegający fleksji dla języka francuskiego przedstawiono w tabeli 15.

Tabela 15. Wzorzec translacyjny nie podlegający fleksji dla języka francuskiego

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

Z kolei wzorzec translacyjny typu koniugacyjnego dla języka francuskiego przedstawiono w tabeli 16.

Tabela 16. Wzorzec translacyjny typu koniugacyjnego dla języka francuskiego

V_SOURCE	<liczba>	<rodzaj>
V_TARGET_1_1	1	1
V_TARGET_1_2	1	2
V_TARGET_2_1	2	1
V_TARGET_2_2	2	2
<liczba> = VAL_N; <rodzaj> = VAL_G;		

Pracę systemu translacji automatycznej tłumaczącego z języka polskiego na francuski zilustrowano na przykładzie następującego zdania:

Polacy od dawna uczestniczyli w życiu Francji i wspomnienia o niektórych z nich pozostają wciąż żywe w pamięci Francuzów.

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce translacyjne:

- wzorzec translacyjny nie podlegający fleksji:

Polacy
les Polonais
<liczba> = 2; <rodzaj> = 1;

- wzorzec translacyjny typu koniugacyjnego:

od dawna uczestniczyli w	<liczba>	<rodzaj>
a depuis longtemps participé à	1	1
a depuis longtemps participé à	1	2
ont depuis longtemps participé à	2	1
ont depuis longtemps participé à	2	2

- wzorzec translacyjny nie podlegający fleksji:

życiu Francji
la vie française
<liczba> = 1; <rodzaj> = 2;

- wzorzec translacyjny nie podlegający fleksji:

i
et

- wzorzec translacyjny nie podlegający fleksji:

wspomnienia o niektórych z nich
le souvenir de certains d'entre eux
<liczba> = 1; <rodzaj> = 1;

- wzorzec translacyjny typu koniugacyjnego:

wciąż pozostają żywe	<liczba>	<rodzaj>
reste toujours vivant	1	1
reste toujours vivante	1	2
restent toujours vivants	2	1
restent toujours vivantes	2	2

- wzorzec translacyjny nie podlegający fleksji:

w
dans

- wzorzec translacyjny nie podlegający fleksji:

pamięci Francuzów
la mémoire des Français
<liczba> = 1; <rodzaj> = 2;

W wyniku pracy systemu translacji automatycznej otrzymano następujący francuski przekład polskiego zdania:

Le Polonais ont depuis longtemps participé a la vie française et le souvenir de certains d'entre eux reste toujours vivant dans la mémoire des Français.

Z kolei w systemie translacji automatycznej przekładającym z języka polskiego na hiszpański występują wzorce

translacyjne nie podlegające fleksji oraz wzorce translacyjne typu koniugacyjnego.

Wzorec translacyjny nie podlegający fleksji dla języka hiszpańskiego przedstawiono w tabeli 17.

Tabela 17. Wzorec translacyjny nie podlegający fleksji dla języka hiszpańskiego

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

Z kolei wzorec translacyjny typu koniugacyjnego dla języka hiszpańskiego przedstawiono w tabeli 18.

Tabela 18. Wzorec translacyjny typu koniugacyjnego dla języka hiszpańskiego

V_SOURCE	<liczba>	<rodzaj>
V_TARGET_1_1	1	1
V_TARGET_1_2	1	2
V_TARGET_2_1	2	1
V_TARGET_2_2	2	2
<liczba> = VAL_N; <rodzaj> = VAL_G;		

Pracę systemu translacji automatycznej tłumaczącego z języka polskiego na hiszpański zilustrowano na przykładzie następującego zdania:

Trudno jest spotkać miasto bardziej malownicze niż Toledo.

W bazie danych systemu automatycznej translacji zostały odnalezione następujące wzorce translacyjne:

- wzorec translacyjny nie podlegający fleksji:

trudno jest
es difícil

- wzorec translacyjny nie podlegający fleksji:

spotkać
encontrar

- wzorec translacyjny nie podlegający fleksji:

miasto
ciudad
<liczba> = 1; <rodzaj> = 2;

- wzorec translacyjny typu koniugacyjnego:

bardziej malownicze niż	<liczba>	<rodzaj>
más pintoresco que	1	1
más pintoresca que	1	2
más pintorescos que	2	1
más pintorescas que	2	2

- wzorec translacyjny nie podlegający fleksji:

Toledo
Toledo
<liczba> = 1; <rodzaj> = 1;

W wyniku pracy systemu translacji automatycznej otrzymano następujący hiszpański przekład polskiego zdania:

Es difícil encontrar ciudad más pintoresca que Toledo.

Podobnie w systemie translacji automatycznej przekładającym z języka polskiego na włoski występują wzorce translacyjne nie podlegające fleksji oraz wzorce translacyjne typu koniugacyjnego.

Wzorec translacyjny nie podlegający fleksji dla języka włoskiego przedstawiono w tabeli 19.

Tabela 19. Wzorec translacyjny nie podlegający fleksji dla języka włoskiego

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

Z kolei wzorec translacyjny typu koniugacyjnego dla języka włoskiego przedstawiono w tabeli 20.

Tabela 20. Wzorec translacyjny typu koniugacyjnego dla języka włoskiego

V_SOURCE	<liczba>	<rodzaj>
V_TARGET_1_1	1	1
V_TARGET_1_2	1	2
V_TARGET_2_1	2	1
V_TARGET_2_2	2	2
<liczba> = VAL_N; <rodzaj> = VAL_G;		

Pracę systemu translacji automatycznej tłumaczącego z języka polskiego na włoski zilustrowano na przykładzie następującego zdania:

Kościół świętego Franciszka pochodzi z czternastego wieku ale niektóre części zostały dodane w wieku osiemnastym.

W bazie danych systemu translacji automatycznej zostały odnalezione następujące wzorce translacyjne:

- wzorec translacyjny nie podlegający fleksji:

kościół świętego Franciszka
la chiesa di San Francesco
<liczba> = 1; <rodzaj> = 2;

- wzorec translacyjny typu koniugacyjnego:

pochodzi z czternastego wieku	<liczba>	<rodzaj>
è del quattordicesimo secolo	1	1
è del quattordicesimo secolo	1	2
sono del quattordicesimo secolo	2	1
sono del quattordicesimo secolo	2	2

- wzorec translacyjny nie podlegający fleksji:

ale
ma

- wzorec translacyjny nie podlegający fleksji:

niektóre części
alcune parti
<liczba> = 2; <rodzaj> = 2;

- wzorec translacyjny typu koniugacyjnego:

zostały dodane	<liczba>	<rodzaj>
è stato aggiunto	1	1
è stata aggiunta	1	2
sono stati aggiunti	2	1
sono state aggiunte	2	2

- wzorec translacyjny nie podlegający fleksji:

w wieku osiemnastym
nel diciottesimo secolo

W wyniku pracy systemu translacji automatycznej otrzymano następujący włoski przekład polskiego zdania:

*La chiesa di San Francesco e del quattordicesimo secolo
ma alcune parti sono state aggiunte
nel diciottesimo secolo.*

Jak wynika z zamieszczonych przykładów automatycznej translacji z języka polskiego na wybrane języki obce, opracowana przez autora metoda translacji automatycznej oparta na wzorcach charakteryzuje się dużą uniwersalnością, w związku z czym po stosownych modyfikacjach może zostać zastosowana do tłumaczenia pomiędzy parą dowolnie wybranych języków. Stopień modyfikacji zależy

od struktury gramatycznej języka docelowego przekładu. Im język ten posiada bardziej rozbudowaną fleksję, tym większym stopniem skomplikowania będą się charakteryzowały wzorce translacyjne. W ogólnym przypadku budowa wzorców translacyjnych może być skrajnie prosta, jak na przykład w systemie tłumaczącym na język szwedzki, bądź dość skomplikowana, jak na przykład w systemie tłumaczącym na język arabski.

Zostanie rozważony system automatycznej translacji oparty na metodzie wzorców translacyjnych, którego zadaniem jest tłumaczenie tekstów z języka polskiego na szwedzki. W systemie tym występuje tylko jeden typ wzorca translacyjnego, którym jest wzorec translacyjny nie podlegający fleksji. Wzorec translacyjny nie podlegający fleksji dla języka szwedzkiego przedstawiono w tabeli 21.

Tabela 21. Wzorec translacyjny nie podlegający fleksji dla języka szwedzkiego

SOURCE
TARGET

Fakt tak prostej budowy wzorca translacyjnego jest bezpośrednią konsekwencją zaniku we współczesnym języku szwedzkim wszelkich form deklinacyjnych rzeczowników i koniugacyjnych czasowników. Sposób działania rozważanego systemu translacji automatycznej zostanie zilustrowany na następującym przykładzie, w którym celem systemu jest dokonanie przekładu na język szwedzki następującego zdania:

Oni kupili dom po tym jak przeprowadzili się do Sztokholmu.

W bazie danych systemu translacji automatycznej zostały odnalezione następujące wzorce translacyjne:

- wzorec translacyjny nie podlegający fleksji:

oni kupili
de köpte

- wzorec translacyjny nie podlegający fleksji:

dom
ett hus

- wzorec translacyjny nie podlegający fleksji:

po tym jak przeprowadzili się
när de hade flyttat

- wzorec translacyjny nie podlegający fleksji:

do
till

- wzorec translacyjny nie podlegający fleksji:

Sztokholmu
Stockholm

W wyniku pracy systemu translacji automatycznej otrzymano następujący szwedzki przekład polskiego zdania:

De köpte ett hus när de hade flyttat till Stockholm.

Z kolei w systemie translacji automatycznej tłumaczącym z języka polskiego na arabski występują trzy typy wzorców translacyjnych, do których należą wzorce translacyjne nie podlegające fleksji, wzorce translacyjne typu deklinacyjnego oraz wzorce translacyjne typu koniugacyjnego.

Wzorec translacyjny nie podlegający fleksji dla języka arabskiego przedstawiono w tabeli 22.

Tabela 22. Wzorec translacyjny nie podlegający fleksji dla języka arabskiego

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

Z kolei w tabeli 23 przedstawiono budowę wzorca translacyjnego typu deklinacyjnego dla języka arabskiego.

Tabela 23. Wzorec translacyjny typu deklinacyjnego dla języka arabskiego

N_SOURCE	<przypadek>
N_TARGET_1	1
N_TARGET_2	2
N_TARGET_3	3
<liczba> = VAL_N; <rodzaj> = VAL_G; <przypadek> = VAL_C;	

Natomiast struktura wzorca translacyjnego typu koniugacyjnego jest dla języka arabskiego bardzo rozbudowana, co pokazuje tabela 24.

Tabela 24. Wzorec translacyjny typu koniugacyjnego dla języka arabskiego

V_SOURCE	<liczba>	<rodzaj>
V_TARGET_1_1	1	1
V_TARGET_1_2	1	2
V_TARGET_2_1	2	1
V_TARGET_2_2	2	2
V_TARGET_3_1	3	1
V_TARGET_3_2	3	2
<liczba> = VAL_N; <rodzaj> = VAL_G; <przypadek> = VAL_C;		

Należy zauważyć, że w przypadku języka arabskiego atrybut <liczba> może przybierać wartości: 1, 2 lub 3, co odpowiada liczbie pojedynczej, podwójnej i mnogiej.

Praca systemu translacji automatycznej dokonującego przekładów z języka polskiego na arabski została zilustrowana na przykładzie tłumaczenia następującego zdania.

On napisał list do przyjaciela.

W bazie danych systemu translacji automatycznej zostały odnalezione następujące wzorce translacyjne:

- wzorec translacyjny nie podlegający fleksji:

on		
		هو
<liczba> = 1; <rodzaj> = 1;		

- wzorec translacyjny typu koniugacyjnego:

napisał	<liczba>	<rodzaj>
كتب	1	1
كتبت	1	2
كتبا	2	1
كتبتا	2	2
كتبوا	3	1
كتبن	3	2
<przypadek> = 3;		

- wzorec translacyjny typu deklinacyjnego:

N_SOURCE	<przypadek>
رسالة	1
رسالة	2
رسالة	3
<liczba> = 1; <rodzaj> = 2;	

- wzorec translacyjny nie podlegający fleksji:

do przyjaciela	
	لصديقه
<liczba> = 1; <rodzaj> = 1;	

W wyniku pracy systemu translacji automatycznej otrzymano następujący arabski przekład polskiego zdania:

هو كتب رسالة لصديقه.

Zamieszczone przykłady jeszcze raz dowodzą uniwersalności zaproponowanej przez autora metody translacji automatycznej opartej na wzorcach, z tym tylko zastrzeżeniem, że metoda ta musi każdorazowo podlegać modyfikacjom adekwatnym do struktury gramatycznej języka docelowego przekładu.

7. ZAKOŃCZENIE

W tabeli 25 zamieszczono statystyczne zestawienie liczb wzorców translacyjnych o określonej długości, które wystąpiły podczas przeprowadzanych przez autora badań. W tabeli 25 zamieszczono zestawienie dostarczające informacji o tym, ile podczas tłumaczenia wystąpiło wzorców translacyjnych o długości jednego wyrazu (l_1), dwóch wyrazów (l_2), trzech wyrazów (l_3), czterech wyrazów (l_4), pięciu wyrazów (l_5) oraz dłuższych niż pięć wyrazów ($l_{\geq 6}$). W przeprowadzonych eksperymentach translacji automatycznej uwzględniono łącznie 40 języków, a do dokonania ich mechanicznego przekładu wykorzystano w sumie 2737 wzorców translacyjnych.

Tabela 25. Statystyczne zestawienie wzorców translacyjnych o różnej długości dla wybranych języków

Język	l_1	l_2	l_3	l_4	l_5	$l_{\geq 6}$
Języki słowiańskie	289	192	79	26	12	4
Angielski	64	43	16	11	3	2
Niemiecki	69	30	12	10	5	3
Niderlandzki	48	38	20	7	3	1
Szwedzki	59	32	20	10	5	0
Duński	62	39	17	11	1	1
Norweski	64	61	19	9	2	1
Francuski	51	41	18	9	12	1
Włoski	72	31	23	12	2	1
Hiszpański	33	31	26	11	2	2
Portugalski	55	20	16	8	1	1
Rumuński	40	29	22	7	4	0
Islandzki	20	23	5	4	0	0
Afrikaans	24	12	5	6	2	3
Kataloński	5	11	7	2	0	0
Litewski	42	26	6	2	0	0
Łotewski	12	19	0	1	0	0
Grecki	9	15	6	2	0	0
Albański	7	12	9	2	0	1
Perski	15	17	4	0	0	0
Węgierski	14	13	5	0	0	0
Fiński	16	13	4	0	1	0
Estoński	10	14	5	1	0	0
Arabski	86	33	6	5	0	0
Hebrajski	9	12	8	1	0	0
Turecki	17	7	4	2	0	0
Łaciński	19	21	7	2	1	1
Starogrecki	12	12	2	3	2	0
Esperanto	35	9	6	1	1	0
Suma	1258	856	377	165	59	22

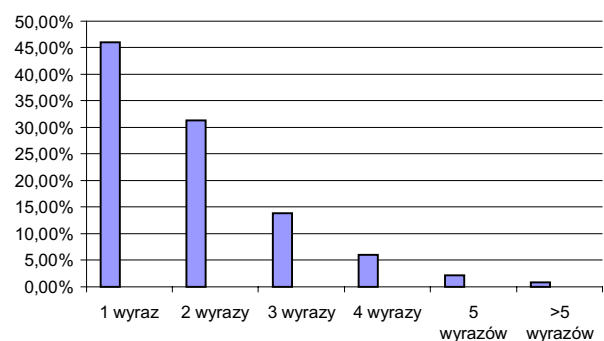
Bardziej szczegółowo wyniki uzyskane dla języków słowiańskich zamieszczono w tabeli 26.

Zamieszczone w tabelach 25 i 26 wyniki eksperymentów, polegających na zastosowaniu metody translacji automatycznej opartej na wzorcach PBMT dla wybranych języków

świata, napawają optymizmem. Okazuje się bowiem, że wzorce translacyjne o niewielkich rozmiarach, tzn. takie, których długość nie przekracza trzech wyrazów, stanowią aż 91,01% wszystkich wzorców translacyjnych, jakie zostały wykorzystane do dokonania przekładów badanych tekstów zapisanych w wybranych językach obcych. Wzorce o umiarkowanej długości, czyli takie, które składają się z czterech bądź pięciu wyrazów, stanowią tylko 8,19% wszystkich wzorców translacyjnych, z czego aż 6,03% stanowią wzorce czterowyrazowe, a tylko 2,16% wzorce pięciowyrazowe. Jednocześnie należy zauważyć, że wzorce translacyjne o długości przekraczającej pięć wyrazów stanowią zaledwie 0,80% wszystkich wzorców.

Tabela 26. Statystyczne zestawienie wzorców translacyjnych o różnej długości dla języków słowiańskich

Język	l_1	l_2	l_3	l_4	l_5	$l_{\geq 6}$
Czeski	33	21	3	0	0	0
Słowacki	18	18	9	1	0	0
Górnolужицьки	23	15	10	1	1	0
Dolnołużyцьки	22	20	10	0	1	0
Rosyjski	35	13	5	0	0	0
Ukraiński	32	12	5	2	1	0
Białoruski	27	15	2	6	1	0
Słoweński	19	14	10	4	3	4
Chorwacki	13	22	8	3	3	0
Serbowski	29	13	5	3	1	0
Bułgarski	22	19	8	4	1	0
Macedoński	16	10	4	2	0	0
Suma	289	192	79	26	12	4



Rys. 12. Graficzna ilustracja procentowej zawartości wzorców translacyjnych o różnej długości, które wystąpiły podczas automatycznej translacji na język polski tekstów zapisanych w wybranych językach obcych

Zaprezentowane wyniki liczbowe zostały dodatkowo zilustrowane na rysunku 12.

Wniosek wypływający z zamieszczonych powyżej rozważań jest taki, że stosując wzorce translacyjne o umiarkowanej długości (nieprzekraczającej pięciu bądź sześciu wyrazów), można uzyskać automatyczne przekłady tekstów charakteryzujące się stosunkowo dobrą jakością. Jest to ważne spostrzeżenie, bowiem wyrażona w liczbie wyrazów

długość wzorca translacyjnego przekłada się bezpośrednio na liczbę wzorców o takiej długości, które powinna zawierać baza danych. Jeżeli brany pod uwagę jest wzorzec translacyjny, który zawiera wiele wyrazów, to automatycznie liczba innych wzorców o takiej samej długości może być bardzo duża, co oczywiście sprawia, że ich akwizycja i umieszczenie w bazie danych nie będzie sprawą łatwą. Z drugiej strony maksymalna liczba wzorców krótkich, np. dwuwyrzawowych, jest stosunkowo niewielka, bowiem równa jest kwadratowi liczby możliwych do utworzenia w danym języku form fleksyjnych wyrazów. W związku z tym istnieje realna szansa na skompletowanie w bazie danych wszystkich potrzebnych wzorców translacyjnych o niewielkiej długości, co pozwoli na w miarę sprawne funkcjonowanie systemu automatycznej translacji. Można tutaj wykorzystać statystyczne modele języka oparte na tzw. bigramach lub trigramach (oczywiście jeżeli dla danego języka takie modele zostały stworzone) i w pierwszej kolejności do bazy danych wprowadzać te bigramy oraz trigramy, których prawdopodobieństwo pojawienia się w tekście zapisanym w danym języku jest największe.

Pewien niepokój natomiast mogą budzić rozmiary baz danych potrzebnych do pomieszczenia wzorców translacyjnych. Również ważnym problemem jest pytanie o czas dostępu do bazy danych zawierającej wzorce translacyjne, co determinuje szybkość pracy systemu translacji automatycznej. Jednakże zdaniem autora, jeżeli istnieją jeszcze obecnie jakieś trudności w tym względzie, to mają one charakter czysto techniczny, a biorąc pod uwagę tempo postępu w dziedzinie technologii półprzewodnikowej, należy oczekiwać ich rychłego przewyżczenia w przyszłości.

Podsumowując uzyskane rezultaty badań, należy jeszcze raz podkreślić, że translacja automatyczna nie należy do tej klasy zagadnień, które można rozwiązać jakimś jednym śmiałym pociągnięciem – błyskotliwym pomysłem. W rzeczywistości opracowanie w miarę sprawnie działającego systemu translacji automatycznej wymaga wielkiego nakładu pracy – zwykle zespół złożony z kilkunastu, a nawet kilkudziesięciu specjalistów (lingwistów i informatyków) musi pracować przez okres kilku lat.

System translacji automatycznej o jakości działania równoważnej efektowi pracy tłumacza przysięgłego pozostanie (przynajmniej przy obecnym stanie nauki – prawdziwa rewolucja w fizyce zdarza się rzadziej niż raz na sto lat) prawdopodobnie na zawsze niedoścignionym ideałem – św. Gralem poszukiwań badawczych. Dlatego też każdy system automatycznej translacji będzie popełniał błędy – chodzi tylko o to, aby ich liczba była jak najmniejsza. Zatem prace zmierzające w kierunku opracowania doskonałego systemu translacji automatycznej przypominają nieco zmagania Don Kichota.

Każdy system translacji automatycznej wymaga nieprzerwanej pracy nad jego udoskonalaniem. Powodem takiego stanu rzeczy jest chociażby fakt, polegający na tym, że każdy język naturalny się zmienia – wciąż do języka wprowadzane są nowe wyrazy (najczęściej zapożyczenia z innych języków) oraz tworzone są nowe związki frazeologiczne.

Zatem praca nad systemem translacji automatycznej nigdy się nie kończy, gdyż wciąż trzeba sprawować nad nim kontrolę, aby jego jakość działania nie uległa pogorszeniu wraz z upływem czasu. W związku z powyższym utrzymanie sprawnego systemu translacji automatycznej przypomina nieco pracę Syzyfa, ponieważ bazę danych należy systematycznie uzupełniać nowymi jednostkami leksykalnymi.

W przyszłości należy oczekiwać, że wraz z powstawaniem i rozrastaniem się wielonarodowych tworców politycznych takich jak Unia Europejska, generowane będzie automatycznie wręcz drastyczne zwiększenie zapotrzebowania na translację wszelkiego rodzaju dokumentów. Jest to bardzo poważny problem wymagający zaangażowania olbrzymiej armii tłumaczy i pokaźnych środków finansowych. Niestety, jak dotąd wszelkie próby uwolnienia ludzkości od przekleństwa Wieży Babel, przez wprowadzenie języków międzynarodowych takich, jak na przykład: esperanto, ido, adamik, interlingua Peano, interlingua Godego, novial, occidental, solresol, volapük i wielu innych (ich liczba jest porównywalna z liczbą języków naturalnych), nie doprowadziło do niczego rozsądnego, ponieważ żaden z tych języków nie zdobył sobie powszechnego uznania. Zatem, jak widać, próbując zniszczyć prawdziwą Wieżę Babel, *de facto* obok wybudowano drugą, o wcale nie mniejszych rozmiarach. Wszystko wskazuje na to, że ludzkość (przynajmniej w dającej się przewidzieć przyszłości) jest skazana na wielojęzyczność, która będzie kreowała coraz większe potrzeby translacyjne [11].

O tym, że bariery językowe stanowiącą mogą skuteczną zaporę przeciwstawiającą się integracji narodów Europy, świadczyć może wypadek, jaki zdarzył się na początku 2001 roku w Belgii (swego czasu było o tym bardzo głośno w środkach masowego przekazu). Warto może w tym miejscu przypomnieć, że Królestwo Belgii powstało w 1830 roku, jako państwo obejmujące terytorium zamieszkałe przez dwie narodowości: Wallonów (mówiących po francusku) i Flamandów (mówiących dialektem języka niderlandzkiego), przy czym Konstytucja Belgii w sposób szczegółowy opisuje przebieg granicy językowej między wymienionymi narodowościami (stolica Bruksela jest dwujęzyczna). Właśnie bariera językowa, jaka istnieje pomiędzy dwoma narodami zamieszkującymi Belgię, stała się przyczyną wspomnianego wypadku. Otóż, maszynista pociągu zmierzającego z obszaru francuskojęzycznego Belgii do jej obszaru niderlandzkojęzycznego wjechał na niewłaściwy tor. Fakt ten został zauważony przez dyżurnego stacji, który natychmiast zatelefonował do swojego kolegi z niderlandzkojęzycznej części Belgii i powiedział mu w swoim ojczystym języku (czyli po francusku), aby tamten nie wypuszczał ze swojej stacji pociągu jadącego w przeciwną stronę. Niestety w odpowiedzi usłyszał tylko: *Ik begrijp het niet!* (co w języku niderlandzkim znaczy: *Nie rozumiem!*), a następnie trzask odłożonej pośpiesznie słuchawki. W kilka minut później doszło do bardzo poważnej katastrofy kolejowej o tragicznych skutkach. Jakże zbawiennym mógłby się okazać w takiej sytuacji nawet najprostszy program translacji automatycznej!

Mając na uwadze powyższą dygresję, można dobrze uświadomić sobie skalę rozważanego problemu i zobaczyć zagadnienia translacji automatycznej we właściwym świetle, bowiem translacja automatyczna dawno już przestała być zagadnieniem czysto akademickim, a stała się palącą potrzebą chwili. Autor pragnie wyrazić w tym miejscu nadzieję, że niniejsza monografia przyczyni się w stopniu znaczącym do postępu w dziedzinie konstruowania systemów translacji automatycznej, a także będzie stanowiła zachętę dla innych badaczy do podjęcia dalszych prac w wytyczonym przez autora kierunku.

Literatura

- [1] Olson K.W.: *An outline of American history*. New York, United States Information Agency 1995
- [2] Gajer M.: *Interaktywny system translacji automatycznej oparty na podejściach typu KBMT oraz EBMT*. Półrocznik AGH Elektrotechnika i Elektronika, t. 21, z. 1, 2002, 17–36
- [3] Gajer M.: *Przegląd najnowszych rozwiązań w dziedzinie translacji automatycznej*. Pro Dialog – Computer Programming and Applications, No. 15, 2003, 19–40
- [4] Penrose R.: *Nowy umysł cesarza – O komputerach, umyśle i prawach fizyki*. Wydawnictwo Naukowe PWN, 1995, ISBN 183-67-11819-9
- [5] Majewicz A.F.: *Języki świata i ich klasyfikowanie*. Warszawa, PWN 1989, ISBN 83-61-08163-5
- [6] Kondratow A.: *Zaginione cywilizacje*. Warszawa, PIW 1988
- [7] Matisoff J.A.: *Zagrożona różnorodność: Języki i formy życia*. Świat Nauki, październik, 2002, 66–73
- [8] Strona internetowa organizacji „Etnologue” o adresie <http://www.cyber.vt.edu/geog1014/topics/110Religion/top100.htm>
- [9] Strona internetowa organizacji UNESCO o adresie http://www.helsinki.fi/~tasalmin/europe_report.html
- [10] Pawlak N.: *Język hausa*. Warszawa, Wydawnictwo Akademickie DIALOG 1998

- [11] Hutchins J.: *Machine translation – Past, Present, Future*. London, Ellis Horwood Series in Computers and Their Applications 1989

Wpłynęło: 10.02.2005

Mirosław GAJER



Urodził się 25 kwietnia 1971 roku. Ukończył kierunek elektronika 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.pl