

ANALIZA MOŻLIWOŚCI I OGRANICZEŃ SYSTEMÓW TRANSLACJI AUTOMATYCZNEJ WSPOMAGANEJ PRZEZ CZŁOWIEKA NA PRZYKŁADZIE SYSTEMU TŁUMACZĄCEGO Z JĘZYKA WŁOSKIEGO NA POLSKI

STRESZCZENIE

Translacja automatyczna jest dyscypliną nauki dostarczającą wiedzy o tym, jak programować komputery, aby były one w stanie dokonywać automatycznych przekładów pomiędzy wybranymi językami naturalnymi. Translacja automatyczna była również jedną z pierwszych aplikacji, jakie zostały zaproponowane dla komputerów. Niestety szybko okazało się, że zadanie translacji automatycznej jest znacznie trudniejsze, ale zarazem o wiele ciekawsze z naukowego punktu widzenia, niż pierwotnie sądzono. W artykule omówiono podstawowe przyczyny powodujące, że translacja automatyczna jest zadaniem tak niezwykle trudnym. Omówiono również najbardziej obiecujące kierunki rozwoju systemów translacji automatycznej. W dalszej części artykułu przedstawiono podstawowe koncepcje związane z nowym, zaproponowanym przez autora podejściem do zagadnień translacji automatycznej. Zamieszczone w artykule rozważania zilustrowano na przykładzie eksperymentalnego systemu translacji automatycznej, dokonującego przekładu zdań zapisanych w języku włoskim na polski.

Słowa kluczowe: translacja automatyczna, przetwarzanie języka naturalnego, sztuczna inteligencja

THE ANALYSIS OF POSSIBILITIES AND FRONTIERS OF THE HUMAN-AIDED MACHINE TRANSLATION SYSTEM ON THE EXAMPLE OF ITALIAN-TO-POLISH TRANSLATION SYSTEM

Machine translation is a branch of science that teaches us how to program the computers, so as they were able to translate between different human languages. Machine translation was also one of the first application that was proposed for computers. Nonetheless, it soon appeared that the task of machine translation is much more difficult, but also much more interesting from the scientific point of view, than one had ever thought before. In the paper it is thoroughly explained why machine translation is so extremely hard. The most promising directions of development of machine translation systems are also briefly described. The special attention is paid to machine translation systems that are developed for Polish language. The other part of the paper is devoted to some practical experiments of implementation of human-aided machine translation technique for the system that translates from Italian into Polish. The way in which the system operates is illustrated on numerous examples. Italian language belongs to the group of Romance languages at its main feature is a variety of flexion forms of verbs. In the paper the algorithm for Italian flexion is described and it is illustrated on some examples.

Keywords: Machine translation, Computational linguistics, Natural language processing

1. WSTĘP

Celem badaczy zajmujących się translacją automatyczną jest opracowanie programów komputerowych, które byłyby w stanie dokonywać przekładów tekstów zapisanych w pewnym języku naturalnym, np. włoskim, na wybrany inny język naturalny, np. bułgarski.

Wbrew pozorom, translacja automatyczna nie jest wcale młodą dziedziną wiedzy – znana jest od dawna, prawie tak długo jak sam wynalazek cyfrowego komputera. Za pioniera translacji automatycznej uznawany jest powszechnie Warren Weaver, który w 1949 roku wystosował do amerykańskiej fundacji wspierającej naukę (Rockefeller Foundation) memorandum, w którym postulował podjęcie na szeroką skalę badań w dziedzinie translacji [1]. Weaver zain-

spirowany został odkryciami w zakresie kryptograficznych technik szyfrowania danych, które rozwinęły niezwykle mocno w latach drugiej wojny światowej. Badacz ten uważał, że proces translacji pomiędzy dwoma językami naturalnymi polega również na pewnej swoistej deszyfracji zapisanej za pomocą obcojęzycznych słów informacji. Zatem kluczem do zautomatyzowania translacji byłaby jedynie znajomość sposobu szyfrowania.

Autor niniejszego artykułu nie dysponuje niestety żadnymi informacjami na temat tego, czy Warren Weaver znał w stopniu dostatecznym jakikolwiek język obcy, ale na podstawie przedstawionego przez niego rozumowania śmie twierdzić, iż wiedza lingwistyczna tego człowieka nie była zapewne zbyt rozległa, przy czym powody, dla których autor wyraża taką opinię, zostaną szerzej omówione w kolejnym punkcie artykułu.

* Akademia Górniczo-Hutnicza w Krakowie, Katedra Automatyki

Pierwszy publiczny pokaz systemu komputerowego, który przetłumaczył z języka rosyjskiego na język angielski 49 prostych i wybranych odpowiednio wcześniej zdań, odbył się już w 1954 roku. Możliwości tego systemu były oczywiście bardzo ograniczone, podobnie zresztą, jak ograniczone były możliwości ówczesnych komputerów. Rozważany system korzystał z rosyjsko-angielskiego słownika o pojemności zaledwie 250 haseł oraz z jedynie 16 prostych reguł gramatycznych [2].

Po pierwszych spektakularnych sukcesach szybko jednak okazało się, że zautomatyzowanie procesu translacji pomiędzy językami naturalnymi jest zadaniem o wiele ciekawszym i niestety znacznie trudniejszym, niż pierwotnie sądzono. Taki stan rzeczy spowodował pojawienie się wielu krytycznych wystąpień osób, które w ogóle kwestionowały sens prowadzenia jakichkolwiek dalszych badań nad translacją automatyczną. Z surową krytyką wystąpił w 1959 roku izraelski filozof Bar-Hillel, który przedstawił dosyć silne argumenty na poparcie głoszonej przez siebie tezy, w myśl której zautomatyzowanie translacji w taki sposób, aby programy komputerowe produkowały przekłady o jakości porównywalnej z jakością pracy tłumaczy profesjonalistów, nie jest możliwe, i to nie tylko z powodu jakichś przejściowych ograniczeń natury technicznej, ale z zasady. Wystąpienie Bar-Hillela zaowocowało ostatecznie wydaniem w 1964 roku przez Amerykańską Akademię Nauk raportu, nazwanego później raportem ALPAC (*Automatic Language Processing Advisory Committee*), który nakazywał wstrzymanie finansowania badań nad translacją automatyczną. Wskutek tego zainteresowanie badaczy tą dziedziną informatyki stosowanej zmalało praktycznie do zera na dobrych kilkanaście lat. Renesans nastąpił dopiero pod koniec lat 70. XX w. Pierwszym systemem translacji automatycznej, który odniósł duży sukces rynkowy, był opracowany w 1977 roku w Kanadzie system METEO. Zadanie stawiane przed systemem METEO ograniczało się tylko i wyłącznie do jednej wybranej dziedziny zastosowań – tłumaczenia raportów i komunikatów meteorologicznych z języka angielskiego na język francuski. System ten pracuje do chwili obecnej i tłumaczy dziennie teksty o objętości około 50 tysięcy słów, które ukazują się już bez żadnych poprawek i przeróbek we francuskojęzycznej prasie, radiu i telewizji. Począwszy od końca lat 70. zainteresowanie translacją automatyczną wzrastało systematycznie na całym świecie, zwłaszcza w Japonii, Stanach Zjednoczonych, Rosji i w Europie Zachodniej [3].

2. DLACZEGO AUTOMATYZACJA TRANSLACJI JEST TAK NIEZWYKLE TRUDNYM ZADANIEM?

Istnieje wiele różnorodnych powodów faktu sprawiającego, że proces translacji pomiędzy językami naturalnymi z tak wielkimi oporami poddaje się implementacji komputerowej. Pierwszym i najważniejszym powodem jest wieloznaczność wypowiedzi, która jest stałą i nieuniknioną cechą każdego języka naturalnego. Wieloznaczność wypowiedzi występuje na trzech poziomach:

- 1) leksykalnym (dotyczy słownictwa),
- 2) syntaktycznym (dotyczy analizy gramatycznej wypowiedzi),
- 3) semantycznym (dotyczy znaczenia wypowiedzi).

Trzeba na wstępie podkreślić, że rozważana wieloznaczność wypowiedzi nie stanowi dla ludzi jakiegoś większego problemu, który utrudniałby wzajemną komunikację. Podobnie wieloznaczność nie jest czymś, co utrudniałoby w stopniu znaczącym pracę tłumaczom profesjonalistom [5].

W związku z tym powstaje, niemalże w naturalny sposób, propozycja, aby przekładów dokonywać w ten sposób, aby owa wieloznaczność wypowiedzi została w nich zachowana. Niech zatem odbiorca przełożonych treści sam zada sobie trud rozwikłania ewentualnych wieloznaczności – po cóż więc zaprzętać tym „głowę” komputerowi zaangażowanemu w proces translacji? Niestety, jak się za chwilę okaże, sprawy nie mają się tak prosto. Wręcz przeciwnie, zachowanie wieloznaczności wypowiedzi w przełożonym tekście okazuje się bardzo często w praktyce po prostu niemożliwe. Jakże są zatem tego powody?

Najbardziej rozpowszechnionym, i sprawiającym zarazem najwięcej kłopotów podczas automatyzacji procesu translacji pomiędzy językami naturalnymi, typem wieloznaczności jest wieloznaczność występująca na poziomie leksykalnym. Istotnie, otwierając jakikolwiek duży dwujęzyczny słownik (np. francusko-polski) trudno jest doprawdy znaleźć wyrazy, które posiadałyby tylko jedno znaczenie [4]. Wręcz przeciwnie, reguła jest taka, że oprócz jednego znaczenia, które jest zwykle uznawane za podstawowe, wyrazy posiadają również całą masę pobocznych znaczeń. Językiem, który szczególnie obfituje w tego rodzaju wieloznaczności słownictwa, jest język angielski. Na przykład w *Wielkim słowniku angielsko-polskim* można znaleźć następujące polskie ekwiwalenty angielskiego słowa *bow*, występującego w zdaniu w funkcji rzeczownika:

- 1) dziób (statku),
- 2) ukłon,
- 3) łuk,
- 4) kabłąk,
- 5) smyczek,
- 6) kokarda,
- 7) węzeł,
- 8) związanie,
- 9) łęk (siodła),
- 10) pióro kreślarskie do cyrkla,
- 11) motylek (krawat).

Jak widać, tylko jedno angielskie słowo posiada aż jedenaście polskich odpowiedników. Jest rzeczą oczywistą, że rozważana wieloznaczność leksykalna musi stanowić poważną trudność dla twórców programów translacji komputerowej. Bowiem, gdy komputer w tłumaczonym przez siebie tekście napotyka angielskie słowo *bow*, wówczas musi wybrać jeden z polskich odpowiedników tego słowa. Ponieważ wymienione powyżej polskie ekwiwalenty angielskiego słowa *bow* mają różne znaczenia, czasem zupełnie nie związane ze sobą (np. kokarda i smyczek), podjęcie podczas dokonywania wyboru błędnej decyzji będzie z nie-

zwykle dużą dozą prawdopodobieństwa mieć opłakane skutki. Po prostu, odbiorca przetłumaczonego w błędny sposób tekstu najprawdopodobniej w ogóle nie będzie wiedział, o co w nim chodzi [6].

Powstaje tutaj oczywiście pytanie, skąd komputer ma wiedzieć, który z ekwiwalentów słowa *bow* należy w danej sytuacji wybrać. Człowiek tłumacz w takim wypadku kieruje się informacją wydobytą z kontekstu wypowiedzi, a pomaga mu w tym jego inteligencja, przenikliwość umysłu, inwencja twórcza, ogólna wiedza o świecie i najwyczajniejszy ludzki zdrowy rozsądek. Ale w jaki sposób można wbudować w komputer taką ogólną wiedzę o świecie oraz zdrowy rozsądek? Zautomatyzowanie procesu wnioskowania pozwalającego na wydobycie informacji z kontekstu wypowiedzi też może okazać się w praktyce niezwykle trudne [8].

Jeżeli natomiast założymy, że komputer będzie dobierał polskie odpowiedniki angielskiego słowa *bow* metodą na chybił trafił, to trzeba zauważyć, że prawdopodobieństwo popełnienia błędu wynosi aż $\frac{10}{11}$. Przy założeniu optymistycznego wariantu, że każde słowo posiada średnio tylko dwa różne znaczenia, prawdopodobieństwo to wynosi $\frac{1}{2}$. Jeżeli założymy ponadto, że typowe zdanie zbudowane jest z dziesięciu wyrazów, to prawdopodobieństwo uzyskania jego prawidłowego przekładu, przy rozwikływaniu wieloznaczności leksykalnych metodą czysto losową, wynosi zaledwie $\frac{1}{1024}$. Jeżeli tłumaczony za pomocą komputera typowy tekst składa się ze stu takich zdań, to prawdopodobieństwo trafnego doboru znaczeń wszystkich występujących w nim wyrazów wynosi zaledwie 1,0715 pomnożone przez dziesięć podniesione do potęgi minus 301. W porównaniu z uzyskaną tutaj wartością prawdopodobieństwa, prawdopodobieństwo trafienia szóstki na pojedynczym kuponie totolotka jest niewyobrażalnie duże, bo wynosi aż $\frac{1}{13983816}$, i jest o 293 rzędów wielkości większe!

Po przeprowadzonych powyżej wyliczeniach, nietrudno odgadnąć, że przekład uzyskany za pomocą takiego programu komputerowego będzie nadawał się tylko i wyłącznie do kosza na śmieci (jest to zapewne miejsce odpowiednie również dla samego rozważanego tutaj programu translacji automatycznej). Niestety, zdecydowana większość programów automatycznej translacji, dotyczy to zwłaszcza prób stworzenia takich programów dla języka polskiego (przekład z języka polskiego na język angielski bądź niemiecki i odwrotnie), działa na zasadach podobnych do wyżej opisanych – nic zatem dziwnego, że produkty te nie cieszą się na rynku zbytnią popularnością.

Innym rodzajem wieloznaczności, który sprawia programom komputerowym znaczne trudności, jest wieloznaczność syntaktyczna. Jej źródłem jest fakt, że często analizę gramatyczną zdania można przeprowadzić na kilka różnych, ale całkowicie równoprawnych sposobów. Na przykład, angielskie zdanie:

I see a boy with a telescope.

można przełożyć jako:

Widzę chłopca z teleskopem.

albo:

Widzę chłopca za pomocą teleskopu.

Jeżeli osoba dokonująca przekładu lub wyręczający ją w tym program komputerowy, nie wiedzą dokładnie o co chodzi (kto ma właściwie ten teleskop, osoba obserwowana czy obserwator), to również nie będzie w stanie dokonać prawidłowego wyboru spośród przedstawionych powyżej dwóch możliwości. Człowiek tłumacz, posługując się kontekstem wypowiedzi, potrafi zwykle rozstrzygnąć tego typu dylematy, jednakże zautomatyzowanie takiego zdroworozsądkowego wnioskowania graniczy w praktyce z cudem.

Ostatnim rodzajem wieloznaczności, z którą borykać się muszą programy automatycznej translacji, jest wieloznaczność występująca na poziomie semantyki wypowiedzi. Na przykład angielskie zdanie:

Put some toner into the printer and then switch it on.

teoretycznie można przetłumaczyć na dwa równoprawne sposoby:

Dodaj toner do drukarki, a następnie uruchom go.

albo:

Dodaj toner do drukarki, a następnie uruchom ją.

Oczywiście, w praktyce tylko drugi przekład uznany może zostać za poprawny. Trzeba po prostu wiedzieć, że w rozważanym przypadku, angielski zaimek *it* odnosi się do słowa *printer*, a nie do słowa *toner*. Człowiek tłumacz posługuje się w tym miejscu posiadaną przez siebie wiedzą dotyczącą danej dziedziny działalności ludzkiej, z której wynika, że uruchomić można na przykład drukarkę, ale nie można zrobić tego samego z tonerem. Niestety wiedza taka z niezwykle dużymi oporami poddaje się procesowi algorytmizacji i wbudowaniu w bezrozumny komputer.

Innym rodzajem trudności, występującym podczas próby zautomatyzowania translacji, są różnice systemowe zachodzące pomiędzy poszczególnymi językami. Panuje w tym względzie żelazna reguła, w myśl której im mniejsze pokrewieństwo (genetyczne, typologiczne i geograficzne) pomiędzy danymi językami, tym rozważane różnice są większe [9].

Różnice systemowe występują zwłaszcza w systemach gramatycznych różnych języków. Prawdopodobnie (bo autor nie jest tego w stu procentach pewny), w każdym języku świata występują trzy osoby gramatyczne: pierwsza odpowiadająca osobie wypowiadającej się, druga osobie, do której wypowiedź jest kierowana (jest to tzw. interlokutor), oraz trzecia dla osobnika, rzeczy, zjawiska, pojęcia bądź stanu, będącego przedmiotem komunikatu językowego. Na tym jednak zwykle podobieństwa się kończą, dalej występują już tylko prawie same różnice, a im języki są bardziej od siebie odległe, tym różnice te są większe i jest ich zarażem więcej [10].

Na przykład w języku węgierskim, należącym do ugrofińskiej grupy językowej, nie jest w ogóle znane pojęcie rodzaju gramatycznego. Z tego powodu węgierski zaimek osobowy *ő* można przetłumaczyć na polski, w zależności od kontekstu wypowiedzi, jako: *on, ona* lub *ono*. Nie trzeba chyba dodawać, że wykstrahowanie przez komputer potrzebnych do tego informacji z kontekstu wypowiedzi (trzeba wiedzieć, czy mowa jest o mężczyźnie, kobiecie czy też o dziecku) wcale nie jest rzeczą łatwą – w przeciwieństwie do człowieka, który na ogół nie ma z tym jakichś szczególnych trudności [11].

Z kolei polskiemu zaimkowi osobowemu *ty* odpowiada arabski zaimek *'anta*, jeżeli wypowiedź kierowana jest do mężczyzny (co można by wyrazić po polsku, jako „ty mężczyzno”), lub *'anti*, jeżeli wypowiedź jest kierowana do kobiety (można by powiedzieć po polsku „ty kobieto”). Jak widać, aby dokonać prawidłowego przekładu, komputerowy tłumacz musi wiedzieć, czy osoba mówiąca zwraca się do kobiety, czy też do mężczyzny.

Kolejna osobliwość dotycząca języka arabskiego polega na występowaniu w tym języku tzw. liczby podwójnej, odnoszącej się do przypadku, w którym jest mowa o dokładnie dwóch osobnikach, rzeczach, pojęciach bądź zjawiskach. Jako ciekawostkę można podać, że liczba podwójna występowała w średniowieczu również w języku polskim i jej formy wyszły z użycia dopiero w XV wieku.

Zatem aby przetłumaczyć na język arabski polski zaimek osobowy *wy*, należy wiedzieć, jaka jest liczba osób, do których mówca się zwraca. Bowiem, jeżeli są to tylko dwie osoby, należy w przekładzie użyć zaimka *'antuma* (można by powiedzieć po polsku „wy dwaj” albo „wy dwie”). Jeżeli natomiast osób tych jest więcej, to trzeba jeszcze znać ich płeć, bowiem jeśli są to sami mężczyźni bądź grupa mieszana (damsko-męska), należy posłużyć się zaimkiem *'antum*, natomiast w przypadku, gdy są to same kobiety, trzeba użyć zaimka *'antunna*.

Spośród żywych języków słowiańskich liczba podwójna występuje obecnie jeszcze tylko w języku słoweńskim. Na przykład, chcąc przetłumaczyć na język słoweński polski wyraz *bracia* trzeba wiedzieć, czy jest ich dwóch – należy wówczas użyć formy *brati* (można przełożyć na polski jako „dwaj bracia”), czy też jest ich więcej – w tym przypadku odpowiednią formą będzie *brata*.

Kolejną osobliwością zacerpniętą, tym razem, z języka hiszpańskiego jest trójstopniowy system zaimków wskazujących, podczas gdy w większości języków europejskich występuje system dwustopniowy. Na przykład w języku polskim mamy dwie formy zaimków wskazujących: *to* i *tamto*, w zależności od stopnia bliskości osoby mówiącej i przedmiotu, o którym jest mowa. Natomiast w języku hiszpańskim występują następujące formy zaimków wskazujących: *esto* (można przełożyć na polski jako „to tutaj”), *eso* (można przełożyć na polski jako „to tam”) i *aquello*.

Duże różnice systemowe można zaobserwować, analizując systemy czasów gramatycznych wybranych języków. Na przykład system czasów gramatycznych języka polskiego jest o wiele uboższy od systemu języka angielskiego.

Rozważmy angielskie zdanie warunkowe:

If I had enough money I would buy this car.

które można przełożyć na język polski jako:

*Gdybym miał wystarczającą sumę pieniędzy,
kupiłbym ten samochód.*

Należy jednak zauważyć, że identyczne tłumaczenie na język polski ma również następujące angielskie zdanie warunkowe:

*If I had had enough money
I would have bought this car.*

Natomiast znaczenia obu tych zdań są zupełnie odmienne, bowiem w pierwszym przypadku rozważany warunek jest wciąż możliwy do spełnienia, podczas gdy drugie z wymienionych zdań odnosi się do przeszłości, w której warunek ten nie został spełniony.

Należy zauważyć, że język polski nie posiada konstrukcji gramatycznych pozwalających na oddanie wymienionych różnic znaczeniowych.

Aby to uczynić, należałoby w języku polskim posłużyć się czasem zaprzeszyłym. Problem polega jednakże na tym, że gramatyka współczesnego języka polskiego użycia takiego czasu nie dopuszcza, wyszedł on bowiem z użycia na przełomie XIX i XX wieku – posługiwali się nim jeszcze w swoich utworach literackich, między innymi, Henryk Sienkiewicz i Eliza Orzeszkowa.

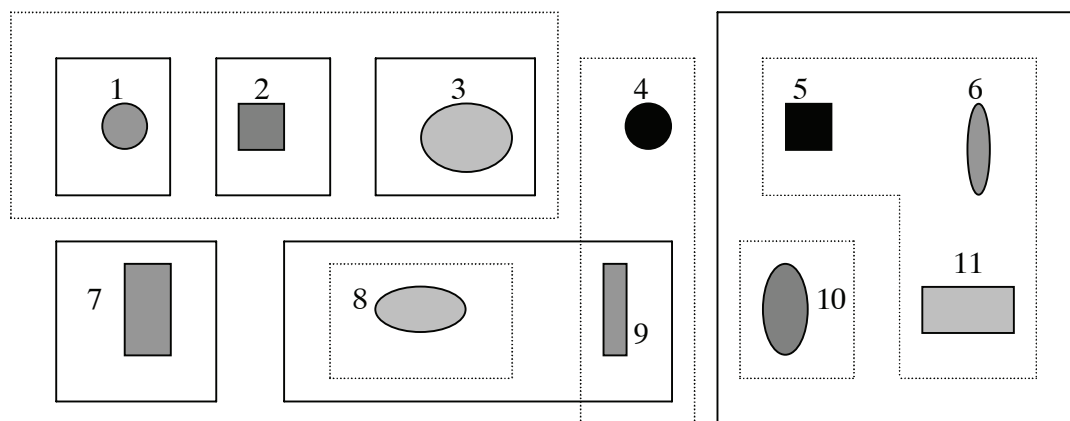
Zatem posługując się (niedopuszczalną oficjalnie) konstrukcją czasu zaprzeszyłego, rozważane angielskie zdanie można by sformułować w języku polskim jako:

*Gdybym był miał wystarczającą sumę pieniędzy,
kupiłbym był ten samochód.*

Warto zwrócić uwagę, że takie konstrukcje można jeszcze spotkać w polskim języku mówionym – zwłaszcza w przypadku starszych osób.

Różnice systemowe pomiędzy różnymi językami występują również w słownictwie, bowiem słownictwo każdego języka stanowi niezwykle skomplikowany system, o niepoznanych jeszcze do końca wzajemnych relacjach między poszczególnymi jednostkami leksykalnymi. Każdy z języków dzieli opisywaną przez niego rzeczywistość (świat osób, przedmiotów, zjawisk, pojęć itp.) w sposób właściwy tylko dla niego. Jest rzeczą zastanawiającą, że użytkownicy różnych języków postawieni w obliczu dokładnie takiej samej sytuacji, w formułowanych na jej temat wypowiedziach umieszczają zupełnie odmienne elementy tej sytuacji dotyczące. Bowiem każdy z języków dzieli otaczającą jego użytkowników rzeczywistość na odmienne fragmenty, którym następnie przypisywane są odrębne jednostki leksykalne. Proces ten został zilustrowany na rysunku 1.

Na rysunku 1 przedstawiono w sposób poglądowy jedność 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.



Rys. 1. Podział rzeczywistości na jednostki leksykalne w dwóch różnych językach
Objaśnienia w tekście

Na rysunku 1 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 bezwzględna 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.

Powyższe, nieco abstrakcyjnie brzmiące, wywody zostaną teraz zilustrowane przykładami zaczerpniętymi z konkretnych języków świata. Na przykład, chcąc przetłumaczyć na język szwedzki polski wyraz *dziadek* można użyć słowa *morfar*, jeżeli dziadek ten jest ojcem matki, lub *farfar*, jeżeli rozważany dziadek jest ojcem ojca. Jak widać, bez dodatkowej informacji wydobytej, być może, z kontekstu wypowiedzi, dokonanie poprawnego przekładu w ogóle nie jest możliwe. Nie trzeba chyba dodawać, że stworzenie programu komputerowego, który przeprowadzałby w sposób automatyczny tego typu wnioskowanie może okazać się niezmiernie trudne do realizacji.

Inny ciekawy przykład pochodzi z języka francuskiego. Otóż, gdy chcemy przetłumaczyć na język francuski polskie słowo *rzeka*, mamy do wyboru wyraz *fleuve*, który określa rzekę uchodzącą bezpośrednio do morza (rzeką taką jest np. Rodan) lub *rivière*, który określa rzekę nieposiadającą bezpośredniego połączenia z morzem. Zatem program automatycznej translacji dokonujący przekładu z języka polskiego na francuski, musiałby posiadać wbudowaną weń bazę wiedzy geograficznej, aby wiedzieć w danym przypadku, jakiego francuskiego ekwiwalentu polskiego słowa *rzeka* użyć.

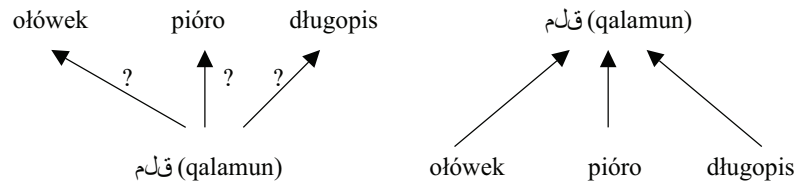
Również chcąc przełożyć na język angielski polski wyraz *skóra*, trzeba wiedzieć, czy *skóra* ta jest już w jakiś spo-

sób przetworzona (wygarbowana, zabarwiona itp.), wówczas należy użyć słowa *leather*, lub czy jest to skóra żywa bądź przetworzona w taki sposób, że przypomina żywą skórę (np. skórę węża na torebce z wężowej skóry), wówczas należy użyć słowa *skin*. Nie trzeba chyba dodawać, że wyciąganie tego typu wniosków przez program komputerowy jest mało prawdopodobne.

Podobnie w języku niemieckim, polskiemu czasownikowi *jeść* odpowiada czasownik *essen*, gdy mowa jest o jedzącym człowieku, lub *fressen*, gdy mowa jest o jedzącym zwierzęciu. Zatem komputer musiałby jakoś umieć wywnioskować z kontekstu wypowiedzi, czy mowa jest o człowieku, czy też o zwierzęciu. W jaki więc sposób zautomatyzować takie wnioskowanie?

Z jeszcze dziwniejszymi przypadkami mamy do czynienia, gdy rozważymy dwa języki bardzo do siebie odległe pod względem genetycznym i geograficznym. Na przykład, chcąc przetłumaczyć na język japoński polskie słowo *siostra*, należy użyć słowa *onesan*, gdy siostra ta jest młodsza, lub *imoto*, gdy owa siostra jest starsza. Podobnie, w indiańskim języku navaho nie ma odpowiednika polskiego czasownika *iść*, istnieje natomiast cały szereg czasowników, które określają odpowiednio poruszanie się pojedynczej osoby, dwóch osób, małej grupki osób, dużej grupy osób itp. Również walijski przymiotnik *glas* odpowiada polskim przymiotnikom *zielony*, *niebieski* i *szary* – zatem użytkownicy języka walijskiego nie rozróżniają w mowie w ogóle tych kolorów, ponieważ określają je wszystkie wspólną nazwą. W języku polskim możemy zwrócić się do drugiej osoby albo używając formy *ty*, albo też formy *pan* względnie *pani*. Natomiast w języku koreańskim istnieje aż sześć różnych stopni grzeczności, z których każdy wymaga użycia odmiennych form czasownikowych. Należy zdać sobie sprawę z faktu, że problemu tego nie można w żaden sposób ominąć, bo w końcu na którąś z alternatywnych form trzeba się przecież zdecydować.

Należy zauważyć, że problemy z automatyzacją translacji powstają nieuchronnie podczas przechodzenia z systemu gramatycznie bądź leksykalnie uboższego na bogatszy. Nie ma natomiast żadnych problemów przy przechodzeniu z systemu bogatszego na uboższy.



Rys. 2. Podczas automatyzacji translacji problemy występują jedynie w przypadku przechodzenia z systemu leksykalnie uboższego na bogatszy

Zostało to zilustrowane na rysunku 2, gdzie przedstawiono proces tłumaczenia na język polski arabskiego słowa *qalamun*, 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 chcemy przetłumaczyć na arabski jedno z polskich słów: ołówek, pióro lub długopis.

Ponadto pewne jednostki leksykalne zaczerpnięte z jednego języka mogą nie mieć swoich odpowiedników w innym języku. Dotyczy to przede wszystkim pojęć abstrakcyjnych, będących wytworem pewnej specyficznej kultury (np. pojęcia filozofii hinduskiej), jak również nazw obiektów rzeczywistych związanych z pewnymi realiami lokalnymi. Przykładowo w języku serbskim słowo *badnjak* oznacza kłodę dębową spalaną według starych serbskich zwyczajów w wigilię święta Bożego Narodzenia. Podobnie w języku arabskim słowo *halhalun* określa rodzaj specjalnej bransoletki, którą muzulmańskie kobiety zakładają na nogę. Podobnie arabskie słowo *badrun* oznacza księżyc, który akurat jest w pełni, a słowo *thukla* kobietę, która utraciła swoje jedyne dziecko. Takie jednostki leksykalne nie posiadają swoich bezpośrednich odpowiedników w języku polskim – są to tak zwane dziury leksykalne (ang. *lexical holes*) – w związku z czym, mogą zostać przetłumaczone na język polski jedynie w sposób opisowy, co w niektórych przypadkach może sprawiać nieco trudności.

Odrębnym problemem występującym podczas translacji jest tłumaczenie różnego rodzaju skostniałych związków frazeologicznych, wyrażeń idiomatycznych i przysłów. Przykładem takich skostniałych związków frazeologicznych mogą być na przykład połączenia rzeczowników ze pewnymi specyficznymi przymiotnikami [4]. Zjawisko to powoduje, że dosłowne tłumaczenie takiego przymiotnika brzmi rażąco i nienaturalnie w połączeniu z towarzyszącym mu rzeczownikiem. Na przykład dosłowne tłumaczenie włoskiego zdania:

Lui è un uomo bello.

jako:

On jest pięknym mężczyzną.

brzmi w języku polskim sztucznie. Należało by raczej powiedzieć:

On jest przystojnym mężczyzną.

Jednakże włoski przymiotnik *bello* oznacza *piękny*, ewentualnie *ładny*, natomiast w rozważanym przypadku należy mu przypisać jeszcze jedno dodatkowe znaczenie: *przystojny*.

Z kolei gdyby dosłownie przetłumaczyć na język polski szwedzkie porzekadło:

Många bäckar små gör en stor å.

otrzymano by zdanie:

Wiele małych strumyków czyni rzekę.

Należy oczekiwać, że tłumacz profesjonalista posłużyłby się zapewne polskim analogiem tego szwedzkiego porzekadła i przełożył je jako:

Kropla drąży skalę.

Podobnie niemieckie przysłowie:

Schlafende Hunde soll Mann nicht wecken.

można by dosłownie przełożyć na język polski, jako:

Nie powinno się budzić śpiących psów.

Natomiast właściwym przekładem będzie polski odpowiednik tego przysłowia:

Nie wywołuj wilka z lasu.

Aby komputer mógł prawidłowo tłumaczyć tego typu skostniałe związki frazeologiczne, porzekadła i przysłowia należałoby wyposażyć go w odpowiednią bazę danych, w której zawarte byłyby wzorce takich tłumaczeń. W zasadzie można to zrobić bez jakichś większych problemów, natomiast pewien niepokój budzić mogą rozmiary takiej bazy danych [14, 15, 16].

O wiele bardziej skomplikowane przedstawiają się sprawy w przypadku próby automatyzacji przekładu wyrażeń idiomatycznych, czyli takich związków składniowych, których znaczenia nie jesteśmy w stanie wywnioskować na podstawie analizy znaczeń ich elementów składniowych. Na przykład amerykańskie zdanie zawierające konstrukcję idiomatyczną:

He had ants in his pants.

posiada następujące znaczenie:

Wiercił się i nie mógł spokojnie usiedzieć na miejscu.

Jedyny problem z przekładem tego zdania polega na tym, że zdanie takie może zostać przetłumaczone również dosłownie, jako:

Miał mrówki w spodniach.

co też może być przekładem poprawnym w przypadku, gdy opisywana jest pewna rzeczywista sytuacja, bo przecież ktoś mógł wybrać się na przykład na wycieczkę do lasu i jakieś mrówki mogły mu istotnie powłazić do spodni (co nie było z pewnością dla tej osoby zbyt przyjemne).

Podobnie, angielskie zdanie:

Sam kicked the bucket.

znaczy w tłumaczeniu dosłownym, że:

Sam kopnął wiadro.

Natomiast w przenośni (jako dość popularny w Stanach Zjednoczonych idiom) zdanie to znaczy, że: *Sam umarł* – można by posłużyć się tutaj polskim odpowiednikiem tego idiomu i powiedzieć, że: *Sam kopnął w kalendarz*. Problem polega jedynie na tym, że w pewnych sytuacjach właśnie tłumaczenie dosłowne może być tłumaczeniem właściwym, bo przecież nasz Sam mógł rzeczywiście kopnąć w zdenerwowaniu jakieś stare wiadro, które gdzieś raczyło stanąć mu na drodze.

Oczywiście, człowiek tłumacz prawie zawsze jest w stanie wywnioskować z kontekstu wypowiedzi, czy daną frazę należy traktować dosłownie, czy też jako wyrażenie idiomatyczne. Natomiast opracowanie algorytmu, który byłby w stanie przeprowadzić tego typu wnioskowanie, pozostawia się Czytelnikowi, jako ćwiczenie, które powinien wykonać, gdy tylko znajdzie chwilę wolnego czasu...

Powyższe przykłady, zaczerpnięte z różnych języków świata pozwalają uświadomić, jak niezwykle złożonym i trudnym jest proces translacji. Bowiem czytając jakąkolwiek książkę przełożoną z języka obcego na język polski, zwykle nie zdajemy sobie nawet sprawy, jak wiele trudu musiał włożyć tłumacz, aby przedstawić czytelnikowi opisywany w tej książce świat, w sposób jak najbardziej zbliżony do sposobu, w jaki opisał go w oryginale autor [13].

3. METODY TRANSLACJI AUTOMATYCZNEJ

Do chwili obecnej w wielu różnych ośrodkach badawczych rozproszonych na całym świecie dopracowano się wielu różnorodnych podejść do zagadnienia translacji automatycznej. W punkcie niniejszym zostaną pokrótce scharakteryzowane wszystkie najważniejsze koncepcje, jakie pojawiły się do chwili obecnej w translacji automatycznej.

3.1. Translacja oparta na transferze morfologicznym

Najstarszą i chyba najpowszechniej stosowaną metodą translacji automatycznej jest metoda oparta na tzw. transferze morfologicznym (ang. *morphological transfer*). Jest to również metoda odznaczająca się najwyższym stopniem intuicyjności, przypominająca zarazem sposób, w jaki pracuje człowiek tłumacz.

Metoda transferu morfologicznego polega na dokonaniu w pierwszej kolejności rozbioru gramatycznego tłumaczonego zdania, którego celem jest ustalenie funkcji, jakie spełniają poszczególne wyrazy w zdaniu. W fazie tej powinno zostać, między innymi, ustalone, który z wyrazów jest podmiotem zdania, który jego orzeczeniem, a który dopełnieniem bliższym bądź dalszym. Po zakończeniu tych czynności należy dokonać podstawień obcojęzycznych ekwiwalentów poszczególnych wyrazów, które pobrane zostają z dwujęzycznego słownika.

Pozornie wszystko brzmi jasno i logicznie, ale niestety dość spore kłopoty zaczynają się w chwili przejścia od pięknej teorii do szarej praktyki. Okazuje się bowiem, że zautomatyzowanie procesu analizy gramatycznej zdania nie jest wcale takie łatwe. Problem jest stosunkowo łatwo rozwiązywalny dla prostych zdań pojedynczych, natomiast jest on prawie niemożliwy do rozwiązania w przypadku długich i skomplikowanych zdań wielokrotnie podrzędnie złożonych o zawilej budowie. Ponadto stopień trudności wzrasta, w miarę tego, im bardziej syntetyczny jest dany język, gdyż w miarę wzrostu stopnia syntetyczności danego języka szyk wyrazów w zdaniu staje się coraz bardziej dowolny, co niezmiernie utrudnia jego gramatyczną analizę. Dobrym przykładem ilustrującym to zjawisko będzie porównanie języka angielskiego, który jest typowym reprezentantem grupy języków analitycznych, z językiem polskim – typowym przedstawicielem języków syntetycznych. Na przykład, w języku angielskim podmiot występuje zawsze na początku zdania, natomiast w języku polskim, jego pozycja jest właściwie dowolna.

Ponadto omówione powyżej zjawisko wieloznaczności syntaktycznej wypowiedzi sprawia, że system automatyczny może dokonać rozbioru zdania na wiele równoprawnych sposobów, w związku z czym powstaje pytanie, jak ustalić na drodze automatycznego wnioskowania, który z nich jest w danej sytuacji właściwy.

Również zastępowanie poszczególnych wyrazów ich obcojęzycznymi ekwiwalentami nie jest sprawą prostą, z uwagi na wieloznaczność występującą na poziomie leksykalnym, bowiem w każdym języku istnieją takie wyrazy, które mają więcej niż jedno znaczenie. Na przykład greckie słowo $\sigma\kappa\omicron\pi\acute{o}\varsigma$ może, w zależności od kontekstu jego użycia, posiadać takie znaczenia, jak: cel, zamiar, melodia, wartownik. Niestety prawidłowy wybór polskiego ekwiwalentu słowa $\sigma\kappa\omicron\pi\acute{o}\varsigma$ dokonany na drodze całkowicie automatycznej pozostaje poza zasięgiem jakiegokolwiek istniejącego systemu automatycznej translacji i zapewne nie prędko (jeśli kiedykolwiek) system taki zostanie zbudowany.

Nic zatem dziwnego, że jakość pracy systemów automatycznej translacji opartych na metodzie morfologicznego transferu jest raczej niska. Pomimo to metoda ta jest dość powszechnie stosowana, zwłaszcza w przypadku systemów komercyjnych przeznaczonych do tłumaczenia tekstów należących do języka o tematyce ogólnej.

Jako przykład takiego systemu zostanie zaprezentowany opracowany w Polsce program automatycznej translacji, służący do tłumaczenia tekstów z języka angielskiego na polski, o nazwie *English Translator 2*. Szczegóły dotyczące

tego produktu można znaleźć na stronie internetowej firmy Techland [7].

Niestety autor niniejszego artykułu nie miał możliwości pracować z programem *English Translator 2*, a zatem wydane przez mnie wyroki mogą być niesprawiedliwe i może zbyt pochopne, ale przeglądając stronę internetową firmy Techland nasuwają się, mimo wszystko, dość poważne zastrzeżenia i wątpliwości. Chodzi o to, że firma Techland zapewnia, że opracowany przez nią program dokonuje bardzo starannej analizy gramatycznej angielskich zdań oraz korzysta z niezwykle bogatego słownika angielsko-polskiego, zawierającego aż 230 tysięcy haseł, przy czym liczba form fleksyjnych wyrazów języka polskiego wynosi aż około 3 milionów. Niestety przeglądając zamieszczone na stronie www firmy Techland tzw. „screenshoty” (zdjęcia ekranu), obrazujące pracę programu *English Translator 2*, można mieć co do podanych założeń spore wątpliwości. Generalnie trzeba przyznać, że jakość przekładów zamieszczonych na rozważanych „screenshotach” jest, mimo wszystko, dosyć dobra – przynajmniej można w większości przypadków zrozumieć sens przełożonych zdań. Niestety istnieją w tym względzie wyjątki, ale w przypadku translacji automatycznej (przynajmniej na obecnym etapie badań) potknięcia takie są i tak prawdopodobnie nieuniknione. Na przykład angielskie zdanie:

*He suddenly jumped towards them and attacked
with the sword.*

zostało przetłumaczone przez program *English Translator 2* jako:

On nagle skoczył ku nim i atakowany z szpadą.

Jak widać, moduł analizy gramatycznej tym razem zawiódł i rola słowa *attacked* w rozważanym zdaniu została nieprawidłowo rozpoznana, co odbiło się bezpośrednio na jakości przekładu.

Również inne zdanie, zamieszczone w prezentowanych na stronie internetowej „screenshotach”:

*After that time your company will be allowed to negotiate
a bigger discount.*

zostało w dość osobliwy sposób przełożone, jako:

*Po tamtym czasie twoja firma będzie pozwolona
negocjować większą zniżkę.*

Oczywiście, w tym przypadku nieprawidłowo dla języka polskiego została użyta strona bierna, bowiem czasownik *pozwalac* w ogóle nie może tworzyć w języku polskim takiej strony!

Ale to jeszcze nic, bo przecież od takich mankamentów nie jest chyba wolny żaden z programów automatycznej translacji. Natomiast, co napawać musi wręcz zdumieniem, to fakt, że w tłumaczonych (rzekomo) przez program *English Translator 2* angielskich zdaniach znalazły się błędy ortograficzne. Na przykład, w pierwszym z cytowanych zdań słowo *suddenly* zostało napisane „sudenly” (przez jed-

no „d”), a w drugim z tych zdań *allowed* napisano jako „al-
loved” przez „v”)

Powstaje zatem istotne pytanie, jak to jest możliwe, żeby program *English Translator 2* akceptował błędnie napisane wyrazy? A jeśli jednak ich nie akceptuje (bo nie powinien tego w żadnym wypadku robić), to w jaki sposób dokonał ich prawidłowego przekładu? Czy aby rozważane „screenshoty” nie zostały odpowiednio wcześniej spreparowane? A jeśli tak, to w jakim celu? Czyżby program *English Translator 2* nie mógł poradzić sobie z przekładem zamieszczonych tam tekstów? Prawdopodobnie ocena możliwości programu *English Translator 2* wymaga przeprowadzenia w tym zakresie jeszcze wielu solidnych badań i eksperymentów.

3.2. Translacja oparta na przykładach

Obecnie za jeden z najbardziej obiecujących kierunków badań uznaje się technikę translacyjną opartą na przykładach. Technika ta oznaczana jest skrótem EBMT (ang. *Example-Based Machine Translation*). Idea leżąca u podstaw tego podejścia jest bardzo prosta. Założmy, że dysponujemy pewnym tekstem napisanym, dla ustalenia uwagi, w języku polskim oraz innym tekstem, który stanowi jego przekład na np. język hiszpański. Ponadto założmy, że każdemu zdaniu z tekstu polskiego odpowiada dokładnie jedno zdanie w tekście hiszpańskim. Zatem zdania z obu tekstów możemy uporządkować w dwujęzyczne pary, a następnie umieścić je w pewnej bazie danych. Po zgromadzeniu takiej bazy danych o odpowiednio dużych rozmiarach można już przystąpić do tłumaczenia techniką EBMT.

Jeżeli tłumaczymy teraz jakiś zupełnie nowy tekst napisany w języku polskim i w tekście tym wystąpi pewne zdanie, którego tłumaczenie zostało już wcześniej umieszczone w bazie danych, to istnieje bardzo duża szansa, że umieszczając tłumaczenie tego zdania zaczerpnięte z bazy danych w tworzonym aktualnie w języku hiszpańskim tekście, dokonamy prawidłowego przekładu tego zdania. Ktoś może od razu zapytać, dlaczego nie ma całkowitej pewności, że taki przekład jest całkowicie poprawny? Czyżby do rozważanej bazy danych zakradły się jakieś błędy?

Otóż nie, wręcz przeciwnie, jedynym powodem pojawienia się nieprawidłowego przekładu nie są błędy – zakładamy, że nasza baza danych jest wolna od błędów – ale różnice systemowe pomiędzy dwoma językami. Jeżeli przypomnimy sobie teraz wcześniejsze rozważania na temat możliwości i ograniczeń dokonywania przekładów pomiędzy różnymi językami naturalnymi, sprawa stanie się jasna. Dodatkowo zagadnienie to zostanie zilustrowane jeszcze kilkoma dalszymi przykładami. Otóż, chcąc przetłumaczyć z języka węgierskiego na polski zdanie:

Ő orvos.

do wyboru są dwie możliwości:

On jest lekarzem.

albo:

Ona jest lekarką.

Jest to bezpośrednią konsekwencją faktu, że język węgierski nie zna pojęcia rodzaju gramatycznego. Ponieważ w bazie danych zdania polskie i węgierskie ułożone są parami, do bazy takiej można wpisać tylko jeden z dwóch podanych powyżej wariantów polskiego przekładu. Jeżeli założymy, że w bazie tej wpisano zdanie: *Ona jest lekarką*, a w tłumaczonym tekście chodzi akurat o mężczyznę, to oczywiście przekład taki będzie błędny.

Płynie stąd ważna wskazówka, w tekstach, które mają zostać poddane automatycznej translacji, należy wystrzegać się stosowania jakichkolwiek enigmatycznych wyrażań. W szczególności należy unikać stosowania zaimków, a w ich miejsce należy używać pełnych nazw. Na przykład, gdyby tłumaczone węgierskie zdanie brzmiało:

István orvos.

miałoby ono dokładnie jedno tłumaczenie na język polski:

Stefan jest lekarzem.

Jednak nie zawsze można się w ten sposób ustrzec przed wszelkimi pułapkami językowymi. Przykład takiej sytuacji pochodzi z języka szwedzkiego. Założmy, że chcemy przetłumaczyć na język szwedzki następujące polskie zdanie:

Babcia pytała kiedy przyjedziemy.

Mimo iż nie ma w tym zdaniu żadnych zaimków, istnieją dwie wersje jego przekładu:

- 1) *Mormor frågade när vi skulle komma.*
- 2) *Farmor frågade när vi skulle komma.*

Pierwszy przekład odnosi się do przypadku, w którym jest mowa o babci, będącej matką matki, natomiast w drugim mowa jest o babci będącej matką ojca. Ponieważ język polski nie wytworzył odrębnych kategorii leksykalnych, pozwalających na odróżnienie takich koligacji rodzinnych, w sytuacji, gdy wzięty zostanie z bazy danych przekład: *Mormor frågade när vi skulle komma*, a w tekście będzie chodziło o babcię, która jest matką ojca, to oczywiście przekład taki będzie nieprawidłowy, a ponadto u szwedzkiego czytelnika wywoła on zapewne konsternację, bo czytelnik ten nie będzie wiedział, o którą babcię w końcu chodzi.

Podobnie, gdy chcemy przetłumaczyć na język hiszpański polskie zdanie:

Rozmawialiśmy o was.

również istnieją dwie alternatywne możliwości przekładu:

- 1) *Hemos hablado de vosotros.*
- 2) *Hemos hablado de vosotras.*

Pierwszy przypadek dotyczy sytuacji, gdy rozmawiano o samych mężczyznach lub o grupie mieszanej, w której byli zarówno mężczyźni, jak i kobiety. Natomiast przypadek drugi dotyczy sytuacji, w której rozmawiano o kobietach.

Zatem gdyby w bazie danych znajdował się przekład: *Hemos hablado de vosotros*, a w rzeczywistości chodziło o mężczyzn, przekład taki byłby po prostu dla czytających go Hiszpanów zabawny.

Językiem, który potencjalnie może wyjątkowo obfitować w tego typu pułapki, jest język niderlandzki. Sytuacja taka spowodowana jest faktem, że w języku tym nie istnieją żadne specjalne formy służące do wyrażania trybu warunkowego zdania. Zatem w sytuacji wyrwania zdań z kontekstu trudno jest ustalić jednoznacznie ich znaczenie. Zostanie to pokazane na kilku następujących przykładach.

Między innymi, niderlandzkie zdanie:

Als ik in Amsterdam ben, ga ik naar het Rijksmuseum.

może zostać przetłumaczone, jako:

*Jak jestem w Amsterdamie,
to idę do Muzeum Państwowego.*

albo:

*Gdy będę w Amsterdamie,
to pójdę do Muzeum Państwowego.*

Również zdanie:

Als ik in Amsterdam was, ging ik naar het Rijksmuseum.

można przełożyć jako:

*Jak byłem w Amsterdamie,
to poszedłem do Muzeum Państwowego.*

albo:

*Gdybym był w Amsterdamie,
to poszedłbym do Muzeum Państwowego.*

Zadania takie mogą, w przypadku zastosowania rozważanej metody automatycznej translacji opartej na przykładach, przysporzyć sporo trudności.

Rzeczą, która może zastanawiać czytelnika, jest zapewne rozmiar bazy danych, która byłaby w stanie pomieścić wszystkie przykłady tłumaczeń zdań pochodzących z danego języka. Ponieważ liczba możliwych do utworzenia zdań w dowolnie wybranym języku jest zapewne astronomiczna, nikt jeszcze takiej bazy danych nie opracował. W praktyce stosuje się sklejanie potrzebnych zdań z dostępnych z bazy danych fragmentów. Rozwiązanie takie rodzi jednakże liczne problemy, które mają swe źródło głównie na styku takich fragmentów – ich wzajemne dopasowanie nie jest wcale sprawą prostą.

Jednakże trzeba uczciwie powiedzieć, że już obecnie metoda translacji automatycznej oparta na przykładach EBMT cieszy się coraz większą popularnością. Prace w tym kierunku prowadzone są, między innymi, na prestiżowej amerykańskiej uczelni – Carnegie Mellon University w Pittsburgu w stanie Pensylwania. Na uczelni tej opracowano przy wykorzystaniu techniki EBMT dwa programy komputerowe, służące do dokonywania automatycznej translacji z języka angielskiego na język francuski i hiszpański, dla tekstów o tematyce ogólnej. Uzyskane wyniki są zachęcają-

ce, ponieważ pomimo względnie niewielkich rozmiarów bazy danych, w której mieszczą się przykłady translacyjne – zgromadzono około trzech milionów takich przykładów o średniej długości nieco ponad trzech słów – uzyskano pokrycie tłumaczonych dokumentów tymi przykładami rzędu 97%. Ponadto badacze z Carnegie Mellon University zaproponowali oryginalne podejście polegające na uogólnianiu przykładów translacyjnych, poprzez zastępowanie w nich wybranych słów tzw. żetonami (ang. *tokens*). Taki uogólniony przykład zawierający pewną liczbę żetonów jest w stanie pomieścić w sobie wiele indywidualnych przykładów, podpadających pod daną kategorię, co posiada kolosalny wpływ na rozmiary potrzebnej do translacji bazy danych.

Spektakularny sukces, związany z zastosowaniem techniki translacji automatycznej opartej na przykładach EBMT, odniesiono również w Hiszpanii, gdzie opracowany został system służący do dokonywania automatycznych przekładów z języka hiszpańskiego na język kataloński, który jest językiem ojczystym dla około 10 milionów osób – Katalończyków – zamieszkujących południową Hiszpanię (między innymi rejon Barcelony i wyspy Baleary) oraz część Sardynii (okolice miasta Alghero) [13]. Rozważany system został zastosowany do automatycznego tłumaczenia, z urzędowego języka hiszpańskiego na język kataloński, ukazującego się w Barcelonie czasopisma regionalnego zatytułowanego „*Periódico de Catalunya*”. Co najciekawsze, uzyskany tą drogą przekład jest na tyle dobrej jakości, że nie są już później wykonywane żadne dalsze poprawki, a czasopismo kierowane jest bezpośrednio do druku. Nakład tego czasopisma wynosi ponad 60 tysięcy egzemplarzy oraz ukazuje się ono codziennie, a tłumaczone jest automatycznie począwszy od 1999 roku. Jest to chyba pierwszy i jedyny przykład zastosowania translacji automatycznej, do języka o tematyce ogólnej, na tak dużą skalę i z tak dobrym skutkiem. Równie intrygujący jest fakt, że w rozważanym systemie translacji automatycznej nie wykorzystano żadnych wyników badań prowadzonych w dziedzinie tzw. lingwistyki komputerowej. Wręcz przeciwnie, system ten analizuje podawane na jego wejście zdania w możliwie najprostszy sposób, podobny do pracy programu sprawdzającego pisownię. Po prostu słowa hiszpańskie, bądź zlepki takich słów, czyli tzw. związki frazeologiczne, o maksymalnej długości sześciu wyrazów, są zastępowane ich katalońskimi odpowiednikami. Rozważany system translacji automatycznej dysponuje jedynie słownikiem o bardzo dużych rozmiarach, w którym mieszczą się wszelkie potrzebne wzorce tłumaczeń poszczególnych wyrazów i całych związków frazeologicznych. Uzyskane za pomocą tego systemu teksty odznaczają się wysoką jakością i wysokim stopniem dokładności lingwistycznej przekładu. Również godna uznania jest duża szybkość pracy rozważanego systemu. Mianowicie, jest on w stanie przetłumaczyć dokument o rozmiarach 15 KB w ciągu zaledwie dwóch sekund. System ten znajduje się ciągle jeszcze w fazie rozwojowej, ponieważ zaangażowana w jego powstanie grupa lingwistów pracuje nieustannie nad uaktualnianiem jego słownika, poprzez dopisywanie do niego nowych słów i nowych form czasowników oraz sekwencji słów – związków frazeologicznych – o długości nie przekraczającej jednakże sze-

ściu wyrazów. Podsumowując, można powiedzieć, że jest to typowy przykład praktycznej implementacji systemu translacji automatycznej, bazującego jedynie na dopasowywaniu przykładów translacyjnych i nie korzystającego z żadnych wyrafinowanych technik translacyjnych. Gdzie zatem leży tajemnica odniesionego sukcesu? Jak to możliwe, że tak trudne zadanie jak translacja automatyczna może być rozwiązane w tak prosty sposób?

Otóż, tajemnica odniesionego sukcesu leży w bardzo bliskim pokrewieństwie języka hiszpańskiego i katalońskiego. Są to bowiem języki romańskiej grupy językowej, należące ponadto do tej samej podgrupy iberyjskiej tych języków. Różnice pomiędzy tymi dwoma językami występują głównie w ich fonetyce, natomiast różnice we frazeologii i w gramatyce tych języków są znikome. Aby polski czytelnik mógł sobie wyrobić pewien pogląd na tę sprawę, można powiedzieć, że języki hiszpański i kataloński są do siebie mniej więcej tak samo podobne jak języki czeski i słowacki lub duński i norweski.

Wypływa stąd ważny wniosek, odnośnie do przydatności techniki automatycznej translacji opartej na przykładach EBMT. Mianowicie, należy oczekiwać, że jakość uzyskiwanych tą drogą przekładów będzie bardzo silnie zależała od stopnia pokrewieństwa genetycznego pomiędzy językami zaangażowanymi w przekład. Prawdopodobnie technikę EBMT można by z bardzo dobrym skutkiem zastosować do tłumaczenia tekstów z języka polskiego na języki czeski i słowacki (ewentualnie na języki górno- i dolnołużyckie oraz na język kaszubski – oczywiście gdyby w ogóle istniała taka potrzeba), ponieważ wszystkie z wymienionych języków należą do jednej tzw. zachodniej podrodziny języków słowiańskich. Nieco gorszych rezultatów należało by się spodziewać w przypadku tłumaczenia techniką EBMT z języka polskiego na inne języki słowiańskie, należące np. do wschodniej grupy tych języków (rosyjski, ukraiński i białoruski) lub do ich grupy południowej (słoweński, chorwacki, serbski, macedoński i bułgarski) – zwłaszcza język bułgarski i macedoński znacznie odbiegają od innych współczesnych języków słowiańskich. Natomiast mam duże wątpliwości, czy zastosowanie metody EBMT przyniosłoby tak dobre rezultaty w przypadku tłumaczenia z języka polskiego na języki romańskiej i germańskiej grupy językowej. Jeszcze gorszych rezultatów należałoby się spodziewać w przypadku tłumaczenia systemem EBMT z języka polskiego na języki innych grup językowych indoeuropejskiej rodziny językowej, bardziej terytorialnie i kulturowo odległe, takie jak np. języki grupy indoirañskiej czy celtyckiej. Zapewne jeszcze o wiele gorzej miałyby się sprawy w przypadku przekładu na języki należące do odmiennych rodzin językowych, bardzo odległych w sensie klasyfikacji genetycznej i geograficznej, takich jak np. język arabski, suahili, hausa, japoński, chiński, tajski, indonezyjski, tybetański, wietnamski, koreański itp.

3.3. Translacja oparta na wiedzy o języku i dziedzinie tłumaczonych tekstów

W niniejszym podpunkcie scharakteryzowane zostanie pokrótce podejście oparte na wiedzy o języku oraz na wiedzy

o dziedzinie działalności ludzkiej, której formułowane w tym języku wypowiedzi dotyczą. Opracowane zgodnie z tą metodą systemy translacji automatycznej nazywane są systemami opartymi na wiedzy i oznaczane są skrótem KBMT od angielskiej nazwy *Knowledge Based Machine Translation*.

Główna idea, leżąca u podstaw działania takich systemów, polega na opracowaniu specjalnego sztucznego i abstrakcyjnego języka, który spełniał będzie rolę języka pośredniczącego w procesie translacji pomiędzy dwoma językami naturalnymi [17]. Taki język pośredniczący nosi nazwę *interlingua*. Zatem każdy z języków naturalnych, występujących w systemie translacji automatycznej, w którym zastosowano koncepcję języka *interlingua*, wyposażony musi zostać w blok syntezy oraz w blok analizy. Blok syntezy służy do generacji wypowiedzi w rozważanym języku naturalnym z wejściowego tekstu zapisanego w języku *interlingua*. W przeciwieństwie do tego, blok analizy służy do syntezy tekstu w języku *interlingua* na podstawie wypowiedzi w rozważanym języku naturalnym [18].

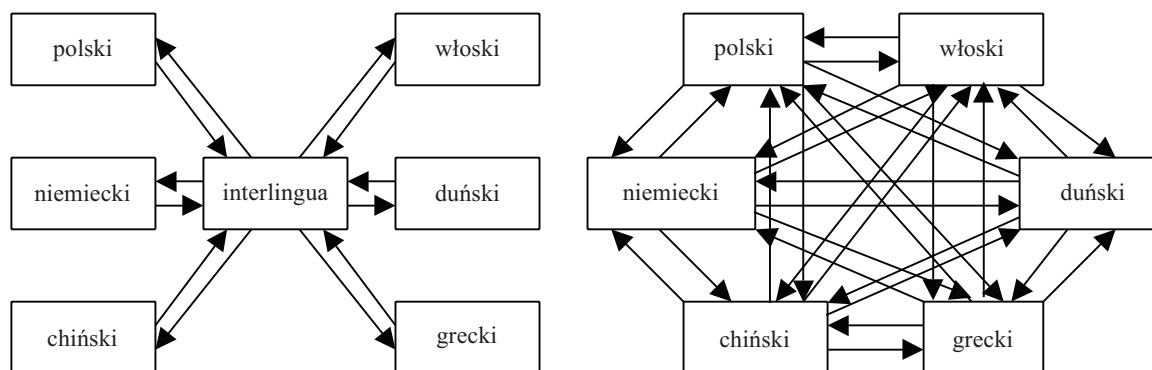
Główną zaletą koncepcji *interlingua* jest to, iż w przypadku systemu n języków naturalnych, wystarczy opracować jedynie $2n$ programów translacji automatycznej (dokładnie n programów syntezy i n programów analizy), aby móc dokonywać przekładów w dowolnym kierunku, pomiędzy dowolnie wybraną parą danych języków. Dla porównania w klasycznej metodzie translacji automatycznej, opartej na tzw. transferze struktur morfologicznych języka, aby stworzyć równoważny system tłumaczenia pomiędzy dowolnie wybraną parą języków, w obu kierunkach, należałoby utworzyć aż $n(n-1)$ programów translacji automatycznej. Jak widać, w koncepcji *interlingua* problem ma złożoność liniową, podczas gdy w metodach klasycznych jego złożoność jest kwadratowa, co zilustrowane zostało na rysunku 3.

Niestety koncepcja języka *interlingua* posiada również wady. Wady te można łatwo dostrzec, przypominając sobie omówione powyżej różnice systemowe w gramatyce oraz w słownictwie, jakie występują pomiędzy różnymi językami naturalnymi [19]. Aby język *interlingua* dobrze spełniał swoją funkcję, język taki powinien być językiem uniwer-

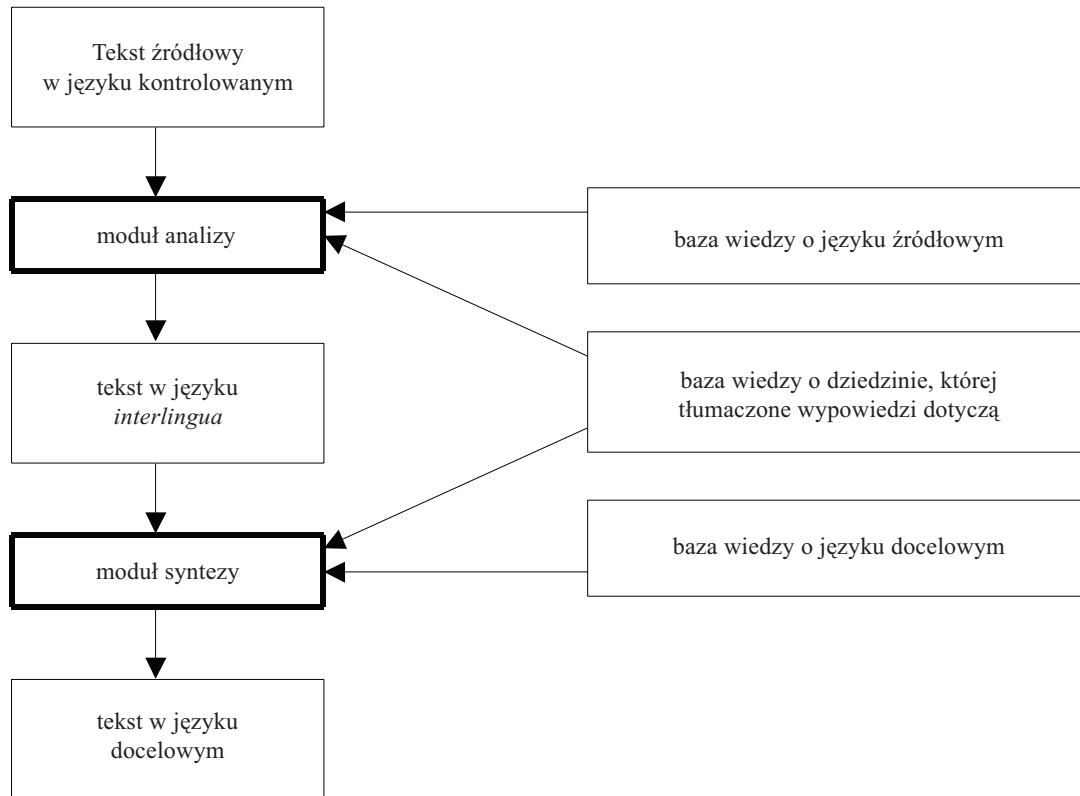
salnym, to znaczy powinien ujmować w swe ramy wszelkie zjawiska językowe, występujące w językach naturalnych zaangażowanych w proces translacji automatycznej. Rozwiązanie takie rodzi natychmiast liczne problemy, nie znane uprzednio w metodach klasycznych. Dla przykładu założmy, że dysponujemy systemem trzech języków: angielskim, niemieckim i japońskim, dla których chcemy opracować uniwersalny język *interlingua*, umożliwiający tłumaczenie pomiędzy dowolnie wybraną parą tych języków.

Założmy, dla ustalenia uwagi, że chcemy przetłumaczyć z języka angielskiego na język niemiecki rzeczownik *the brother* (brat). W klasycznym systemie translacji automatycznej metodą transferu dokonano by, po prostu, podstawienia niemieckiego ekwiwalentu tego słowa, czyli *der Bruder*. Natomiast w rozważanym systemie, ponieważ język *interlingua* musi uwzględniać specyfikę wszelkich języków wchodzących w skład systemu, sprawa nie jest już taka prosta. Dzieje się tak za sprawą języka japońskiego, w którym słowo *brat* odpowiada rzeczownikowi *oniisan*, gdy brat ten jest starszy lub *ototo*, gdy ów brat jest młodszy. Zatem stosując podejście typu *interlingua*, należy rozwiązać pewien dodatkowy problem (wnioskowanie o starszeństwie w rodzinie na podstawie kontekstu wypowiedzi), który w klasycznym podejściu w ogóle nie występuje. Ponadto stworzenie jednego uniwersalnego języka o tak bogatej strukturze, aby obejmował wszystkie zjawiska występujące w wybranych językach świata, leży poza zasięgiem i możliwościami współczesnej nauki i techniki.

Warto również wspomnieć o pewnych próbach zastosowania koncepcji *interlingua* w połączeniu z klasyczną metodą transferu, z tym, że jako język pośredniczący *interlingua* stosowany był jeden z języków naturalnych, między innymi, przeprowadzono taki eksperyment dla języka esperanto. Z powyższych względów zakres zastosowań systemów translacji automatycznej typu KBMT jest mocno ograniczony. Systemy tego typu raczej nie nadają się do tłumaczenia tekstów należących do języka ogólnego. Natomiast ich polem zastosowań jest tłumaczenie wypowiedzi należących do pewnego języka kontrolowanego, służącego do opisu pewnej wybranej dziedziny działalności człowieka.



Rys. 3. Opracowanie systemu tłumaczenia w dowolnym kierunku pomiędzy sześcioma wybranymi językami wymaga w przypadku języka *interlingua* jedynie 12 programów automatycznej translacji (6 syntezy i 6 analizy), podczas gdy w przypadku podejścia klasycznego programów tych musi być aż 30



Rys. 4. Architektura systemu KANT

Rozważany język kontrolowany stanowi ograniczony podzbiór danego języka naturalnego, co oznacza, że nie wszystkie słowa i konstrukcje składniowe, które występują normalnie w danym języku naturalny, są w kontrolowanym podzbiórze tego języka dopuszczalne. Zatem język kontrolowany stanowi język w dużym stopniu sformalizowany – podobny do języka programowania komputerów. Oczywiście, język taki raczej nie nadaje się do pisania utworów literackich, ale za to świetnie może posłużyć do opracowywania tekstów o charakterze naukowym, a zwłaszcza wszelkiej dokumentacji technicznej. Jego dużą zaletą jest to, iż wymusza na osobie w nim piszącej jasny i precyzyjny sposób wyrażania myśli.

Kierując się powyższą przesłanką w amerykańskim Carnegie Mellon University opracowano system translacji automatycznej o nazwie KANT (ang. *Knowledge-based Accurate Natural-language Translation*) (rys. 4). Pole działania systemu KANT jest bardzo ograniczone, system ten służy tylko i wyłącznie do tłumaczenia z języka angielskiego na kilka wybranych języków, między innymi na: francuski, włoski, hiszpański, portugalski, niemiecki i turecki, podręczników obsługi i użytkownika sprzętu ciężkiego, takiego jak np. koparki, dźwigi, spychacze, wywrotki itp. Sprzęt taki produkowany jest w Stanach Zjednoczonych i jest eksportowany do wielu krajów świata, niestety nie można go sprzedawać bez odpowiednich podręczników obsługi, które muszą zostać uprzednio przetłumaczone z języka angielskiego na inne języki i właśnie realizacji tego zadania ma służyć program KANT. Program KANT odniósł

już duży rynkowy sukces, a uzyskiwane dzięki niemu przekłady są już tak dobre, że tylko w sporadycznych przypadkach wymagają interwencji człowieka tłumacza. Obecnie w Carnegie Mellon University trwają intensywne prace nad dostosowaniem systemu KANT do innych dziedzin tłumaczonych tekstów, takich jak np. medycyna i farmacja.

3.4. Podejście oparte na statystyce

Na przestrzeni ostatnich kilkunastu lat zauważyć można było zwiększone zainteresowanie wielu ośrodków badawczych zastosowaniem w translacji automatycznej metod statystycznych. Mówiąc w wielkim skrócie, statystyczne metody translacji automatycznej mogą zostać scharakteryzowane, podobnie zresztą jak translacja oparta na przykładach EBMT, jako podejścia, w których następuje zerwanie z formułowaną *explicite* wiedzą o języku i regułach rządzących procesem translacji.

Większość dokonań w tej dziedzinie polegała na bezpośrednim przeniesieniu na grunt translacji automatycznej metod wypracowanych podczas badań nad systemami automatycznego rozpoznawania mowy (ang. *speech recognition*). Ze statystycznym podejściem do translacji automatycznej wiążą się takie pojęcia, jak:

- statystyczne modelowanie języka (ang. *language model*),
- statystyczne modelowanie translacji (ang. *translation model*).

Statystyczne modelowanie języka udziela odpowiedzi na pytanie o wielkość prawdopodobieństwa pojawienia się w tek-

ście danego ciągu wyrazów. Prawdopodobieństwo to oznaczane jest jako $P(S)$, gdzie S jest pewnym ciągiem wyrazów.

$$S = S_1, S_2, S_3, S_4, \dots, S_N.$$

Prawdopodobieństwo $P(S)$ może zostać policzone jako

$$P(S) = P(S_1) \cdot P(S_2|S_1) \cdot P(S_3|S_1, S_2) \cdot P(S_4|S_1, S_2, S_3) \cdot \dots$$

Prawdopodobieństwo $P(S_N|S_1, S_2, S_3, \dots, S_{N-1})$ nazywane jest N -gramem. W praktyce wartości N -gramów przybliżane są tzw. bigramami $P(S_N|S_{N-1})$ lub trigramami $P(S_N|S_{N-2}, S_{N-1})$. Bardzo rzadko stosuje się w praktyce N -gramy wyższych rzędów. Powodem tego są wielkie trudności z oszacowaniem wartości prawdopodobieństw takich N -gramów na podstawie dostępnych tekstów zapisanych w modelowanym języku. Bowiem już nawet w przypadku pewnych trigramów, które pojawiają się w danym języku bardzo rzadko, oszacowanie ich prawdopodobieństwa następuje z dużymi trudnościami, a uzyskiwane na podstawie dostępnych tekstów wartości i tak są obciążone dużym błędem.

Z kolei statystyczne modelowanie translacji dostarcza wartości prawdopodobieństwa $P(T|S)$ wystąpienia ciągu wyrazów T w języku docelowym translacji, pod warunkiem, że zadany jest ciąg wyrazów w języku źródłowym S .

Ostatecznie prawdopodobieństwo wystąpienia paralelnych ciągów wyrazów w językach źródłowym i docelowym wyznaczone jest ze wzoru

$$P(S, T) = P(T|S) \cdot P(S).$$

Następnie korzystając ze wzoru Bayesa otrzymuje się

$$P(S|T) = P(T|S) \cdot P(S) / P(T).$$

Zatem statystyczne podejście do translacji polega na znalezieniu sekwencji wyrazów S , która posiada najwyższe prawdopodobieństwo wystąpienia, pod warunkiem, że zadana jest sekwencja wyrazów T [12].

Otrzymane do chwili obecnej rezultaty, powstałe w wyniku zastosowania w translacji metod statystycznych, nie napawają zbyt dużym optymizmem. Badania takie prowadzone były między innymi przez amerykańską firmę IBM, gdzie tłumaczone były proste zdania o ubogim słownictwie. W przypadku systemu opartego na czystej statystyce udało się w miarę poprawnie przetłumaczyć zaledwie 39% wprowadzonych do systemu zdań. Z kolei dodanie do rozważanego systemu dodatkowego modułu analizy gramatycznej opartego na pewnej bazie wiedzy lingwistycznej pozwoliło podnieść rozważany wskaźnik do 60%.

Reasumując, można powiedzieć, że statystyczne podejście do translacji automatycznej odznacza się spośród wszystkich rozważanych metod najwyższym stopniem zaawansowania matematycznego, ale jednocześnie najmniejszą użytecznością praktyczną.

4. TRANSLACJA AUTOMATYCZNA WSPOMAGANA PRZEZ CZŁOWIEKA

W artykule niniejszym autor zaproponował własne, oryginalne podejście do zagadnienia translacji automatycznej.

Punktem wyjścia dla opracowanego systemu były niżej wymienione spostrzeżenia:

- Prawdopodobnie w każdym języku naturalnym można wyróżnić pierwszą, drugą i trzecią osobę gramatyczną, przy czym pierwsza osoba odpowiada osobie mówiącej, druga osobie, do której kierowana jest wypowiedź, a trzecia osobie lub przedmiotowi, będącej treścią komunikatu językowego. Inne elementy, takie jak liczba pojedyncza, podwójna lub mnoga, rodzaj gramatyczny są już opcjonalne i zależą od danego języka. Ponadto w niektórych językach pomiędzy podmiotem zdania a jego orzeczeniem zachodzi związek zgody, to znaczy podmiot musi zgadzać się z orzeczeniem pod względem osoby, liczby i rodzaju. Z kolei pomiędzy orzeczeniem zdania a jego dopełnieniem zachodzi może związek rzędu, to znaczy orzeczenie rządzi przypadkiem dopełnienia.
 - Prawdopodobnie w każdym języku naturalnym w budowie zdania można wyróżnić takie elementy jak podmiot zdania, będący sprawcą czynności, orzeczenie zdania, określające rodzaj wykonywanej przez podmiot czynności i dopełnienie zdania określające przedmiot na którym wykonywana jest dana czynność. Ponadto w zdaniu wystąpić mogą okoliczniki precyzujące dodatkowo rodzaj i sposób wykonania danej czynności. Ze względu na kolejność występowania poszczególnych członów zdania: podmiotu S (ang. *subject*), orzeczenia V (ang. *verb*) i dopełnienia O (ang. *object*) wyróżnić można sześć różnych typów składniowych. Na przykład, w języku angielskim jedynym dopuszczalnym typem składniowym jest konstrukcja SVO, wedłu jej jest zbudowane jest przykładowe zdanie:
A cat has drunk milk.
- Z kolei język polski dopuszcza wszystkie sześć możliwych typów składniowych, czyli:
- 1) SVO: *Kot wypił mleko.*
 - 2) SOV: *Kot mleko wypił.*
 - 3) VSO: *Wypił kot mleko.*
 - 4) VOS: *Wypił mleko kot.*
 - 5) OVS: *Mleko wypił kot.*
 - 6) OSV: *Mleko kot wypił.*

Na podstawie powyższego przykładu widać, jak duże są różnice pomiędzy językiem angielskim, który jest typowym językiem analitycznym o ściśle zdefiniowanej składni, a językiem polskim, który jest językiem syntetycznym, w którym składnia jest w bardzo dużym stopniu swobodna.

W każdym języku istnieje pewien typ składni, który może zostać uznany za podstawowy, i który nawet jeśli nie jest obowiązkowy, to i tak bywa używany najczęściej. Dla zdecydowanej większości języków indoeuropejskim takim typem składniowym jest typ SVO. Wyjątkiem jest tutaj na przykład język perski, w którym szykiem obowiązkowym jest SOV.

- Wieloznaczność leksykalna jest wszechobecna w każdym języku naturalnym, bowiem praktycznie każde

słowo posiada co najmniej dwa różne znaczenia, na przykład angielskie słowo: *cape* to:

- 1) peleryna,
- 2) przylądek.

Rozwikłanie wieloznaczności leksykalnych w przypadku języka o tematyce ogólnej wymaga posiadania ogólnej wiedzy o świecie i zdrowego rozsądku, czyli elementów, które prawdopodobnie z zasady nie podlegają algorytmizacji [9]. W związku z powyższym autor stawia tezę, że całkowite zautomatyzowanie procesów translacyjnych nie jest możliwe. Aby system translacji automatycznej mógł działać sprawnie, konieczne jest jego wspomoczenie przez człowieka, aktywnie zaangażowanego w procesy translacyjne.

Biorąc pod uwagę powyższe wskazówki, można uznać, że zaproponowany przez autora system translacji automatycznej działa w sposób następujący:

- System pracuje w trybie on-line, to znaczy tłumaczenie powstaje równolegle z procesem wprowadzania tekstu w języku źródłowym.
- Elementy zdania takie, jak: podmiot S, orzeczenie V, dopełnienie O oraz poszczególne okoliczniki P, wprowadzane są do systemu oddzielnie, w kolejności charakterystycznej dla danego języka.
- W przypadku gdy którykolwiek z elementów: podmiot S, orzeczenie V, dopełnienie O lub okolicznik P, jest wieloznaczny pod względem leksykalnym, system zwraca się do użytkownika z prośbą o rozwikłanie tej wieloznaczności, poprzez wybranie jednej z co najmniej dwóch zasugerowanych możliwości. Na przykład gdyby użytkownik wprowadził jako podmiot zdania polskie słowo *zamek*, wówczas system poprosiłby o dokonanie wyboru pomiędzy następującymi możliwościami:
 - 1) siedziba króla,
 - 2) zamek błyskawiczny,
 - 3) zamek do zamykania drzwi,
 - 4) zamek od karabinu.

- System tłumaczy kolejno wprowadzane elementy składniowe zdania, wypisując je w kolejności właściwej dla danego języka docelowego, uwzględniając przy tym występujące w języku docelowym związki zgody i rzędu.

Działanie tego typu systemu translacji automatycznej wspomaganą przez człowieka zostanie zilustrowane w kolejnym punkcie na przykładzie systemu tłumaczącego z języka włoskiego na polski, którego próbna wersja została opracowana przez autora.

5. AUTOMATYCZNA TRANSLACJA Z JĘZYKA WŁOSKIEGO NA JĘZYK POLSKI

Język włoski należy do grupy języków romańskich i jest typowym reprezentantem języków analitycznych, posiadających ściśle określony szkielet budowy zdania, w tym przypadku typu SVO [20]. Zatem w rozważanym systemie

translacji automatycznej wspomaganą przez człowieka użytkownik musi wprowadzać kolejne elementy składniowe zdania włoskiego właśnie w kolejności:

1. podmiot S,
2. orzeczenie V,
3. dopełnienie O,
4. ewentualne okoliczniki zdania P.

Funkcję podmiotu, orzeczenia i dopełnienia tłumaczonego zdania mogą pełnić nie tylko pojedyncze wyrazy, ale także całe związki frazeologiczne. Zatem słuszniej byłoby mówić w tym przypadku o frazie podmiotowej S, frazie orzeczeniowej V oraz o frazie dopełnieniowej O. Również w funkcji okoliczników zdania mogą wystąpić całe frazy okolicznikowe P.

Zaproponowane podejście bazuje na technice translacji automatycznej opartej na przykładach EBMT, bowiem rozważany system wyposażony został w bazę danych, w której umieszczono przykłady translacyjne dla tłumaczonych fraz podmiotowych, orzeczeniowych, dopełnieniowych i okolicznikowych.

Szczegółnej uwagi wymaga uwzględnienie faktu, że w języku polskim, który jest językiem docelowym omawianego systemu, pomiędzy podmiotem i orzeczeniem zdania występuje związek zgody, a pomiędzy orzeczeniem zadania a jego dopełnieniem zachodzi związek rzędu. Z tego powodu opracowaną bazę danych wyposażono w cztery dodatkowe atrybuty o nazwach:

- 1) <osoba>,
- 2) <liczba>,
- 3) <rodzaj>,
- 4) <przypadek>.

Atrybut <osoba> może przyjmować wartości: 1, 2 lub 3 – co odpowiada pierwszej, drugiej i trzeciej osobie gramatycznej.

Z kolei atrybut <liczba> może przyjmować wartości: 1 lub 2 – co odpowiada liczbie pojedynczej i mnogiej.

Atrybut <rodzaj> może przyjmować wartości: 1, 2 lub 3 – co odpowiada rodzajowi męskiemu, żeńskiemu i nijakiemu.

Natomiast atrybut <przypadek> może przyjmować wartości: 1, 2, 3, 4, 5 lub 6 – co odpowiada mianownikowi, dopełniaczowi, celownikowi, biernikowi, narzędnikowi i miejscownikowi.

Każdorazowo przed wprowadzeniem do systemu nowego zdania, które ma zostać przetłumaczone, atrybutowi <przypadek> przypisywana jest wartość 1, ponieważ podmiot zdania w języku polskim występuje zawsze w mianowniku.

Po wprowadzeniu do systemu podmiotu zdania z bazy danych, oprócz jego tłumaczenia odczytywane są również odpowiednie wartości atrybutów: <osoba>, <liczba> i <rodzaj>, tak aby orzeczenie zdania spełniało związek zgody.

Po wprowadzeniu do systemu orzeczenia z bazy danych pobierane jest jego tłumaczenie oraz odczytywana jest wartość atrybutu <przypadek>, aby dopełnienie zdania wystąpiło we właściwym związku rzędu.

Działanie omawianego systemu zostanie zilustrowane na następującym przykładzie tłumaczenia zdania zapisanego w języku włoskim:

*Mio fratello ha comprato dischi
di musica classica oggi.*

Na początku, przed wprowadzeniem do systemu rozważanego zdania atrybutowi <przypadek> zostaje przypisana wartość 1, ponieważ podmiot zdania zawsze występuje w mianowniku.

<przypadek> = 1;

Następnie użytkownik systemu wprowadza podmiot zdania, w rozważanym przypadku jest to wyrażenie *mio fratello*, które w zależności od wartości atrybutu przypadek może mieć różną postać przekładu na język polski (tab. 1).

Tabela 1

<i>min bror</i>	<przypadek>
mój brat	1
mojego brata	2
mojemu bratu	3
mojego brata	4
moim bratem	5
moim bracie	6

Zatem tłumaczeniem wprowadzonej frazy podmiotowej *mio fratello* będzie „mój brat”. Ponadto atrybutom <osoba>, <liczba> i <rodzaj> muszą zostać przypisane odpowiednie wartości zawarte w bazie danych i skojarzone z hasłem *mio fratello*:

<osoba> = 3,

<liczba> = 1,

<rodzaj> = 1.

Jak widać z zamieszczonego powyżej zestawienia, atrybutowi <osoba> została przypisana wartość 3, ponieważ fraza podmiotowa „mój brat” występuje w trzeciej osobie. Podobnie atrybutowi <liczba> została przypisana wartość 1, ponieważ rozważana fraza występuje w liczbie pojedynczej. Również atrybutowi <rodzaj> przypisana została wartość 1, ponieważ rodzajem gramatycznym tej frazy jest rodzaj męski.

W kroku kolejnym użytkownik wprowadza do systemu frazę pełniącą w rozważanym włoskim zdaniu funkcję orzeczenia, czyli *ha comprato*. Fraza ta w zależności od wartości atrybutów <osoba>, <liczba> i <rodzaj> może mieć różne tłumaczenia (tab. 2).

Z zamieszczonego powyżej zestawienia wynika, że tłumaczenie frazy orzeczeniowej przybiera postać „kupił”. Ponadto z bazy danych oprócz rozważanego tłumaczenia odczytywana jest wartość atrybutu <przypadek>, tak aby pomiędzy orzeczeniem zdania a jego dopełnieniem zachodził związek rzędu.

Tabela 2

<i>ha comprato</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
kupił	3	1	1
kupiła	3	1	2
kupiło	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3

W rozważanym przypadku atrybutowi <przypadek> zostanie przypisana wartość 4, ponieważ dopełnienie tego zdania musi wystąpić w bierniku

<przypadek> = 4.

W kroku kolejnym użytkownik wprowadza frazę pełniącą w zdaniu funkcję dopełnienia, przy czym może to być zarówno dopełnienie bliższe, jak i dopełnienie bliższe wraz z dopełnieniem dalszym. W rozważanym przykładzie wprowadzona zostanie fraza *dischi di musica classica*, która w zależności od wartości atrybutu <przypadek> może posiadać następujące tłumaczenia (tab. 3).

Tabela 3

<i>dischi di musica classica</i>	<przypadek>
płyty z muzyką klasyczną	1
płyt z muzyką klasyczną	2
płytom z muzyką klasyczną	3
płyty z muzyką klasyczną	4
płytami z muzyką klasyczną	5
płytach z muzyką klasyczną	6

Ponieważ atrybutowi <przypadek> przypisana została uprzednio wartość 4, zostanie wybrane tłumaczenie „płyty z muzyką klasyczną”. Z bazy danych zostaną odczytane również wartości atrybutów <osoba>, <liczba> i <rodzaj>, które jednakże nie będą już dalej wykorzystywane:

<osoba> = 3,

<liczba> = 2,

<rodzaj> = 2.

Na zakończenie użytkownik może wprowadzić, w kolejności występowania, dowolną liczbę fraz spełniających w tłumaczonym zdaniu funkcję okoliczników. W rozważanym zdaniu występuje jedna tego typu fraza *oggi*, która posiada tylko jedno tłumaczenie, niezależne od wartości jakiegokolwiek atrybutu (tab. 4).

Tabela 4

<i>oggi</i>
dzisiaj

Po wprowadzeniu do systemu ostaniej frazy użytkownik informuje system o tym, że więcej elementów nie będzie, w wyniku czego atrybutowi <przypadek> przypisana zostanie wartość 1, aby podmiot kolejnego wprowadzanego do systemu zdanie wystąpił w mianowniku

<przypadek> = 1.

Ostatecznie w wyniku pracy rozważanego systemu otrzymuje się następujące polskie tłumaczenie włoskiego zdania *Mio fratello ha comprato dischi di musica classica oggi*:

Mój brat kupił płyty z muzyką klasyczną dzisiaj.

Należy zwrócić uwagę, że jest to tłumaczenie wierne, przekazujące dokładnie treść włoskiego oryginału. Ponadto uzyskane zdanie zapisane w języku polskim jest poprawne pod względem budowy gramatycznej, ponieważ zarówno pomiędzy jego podmiotem i orzeczeniem jest spełniony związek zgody, jak i pomiędzy orzeczeniem i dopełnieniem spełniony jest związek rzędu. Jedynie miejsce występowania okolicznika czasu może wydawać się dla użytkownika języka polskiego nieco nienaturalne, ponieważ umieściłby on zapewne ten okolicznik na początku zdania, które miało by wówczas postać:

Dzisiaj mój brat kupił płyty z muzyką klasyczną.

Niestety zaproponowany system translacji automatycznej nie może zdobyć się na aż tak wysoki poziom finezji i budować zdania piękne pod względem stylistycznym. Zresztą zadanie takie pozostaje poza zasięgiem jakiegokolwiek istniejącego systemu translacji automatycznej. Poza tym tłumaczenie literatury pięknej nie jest celem proponowanego systemu translacji automatycznej, w którym chodzi jedynie o w miarę wierne przekazanie treści oryginału w sposób poprawny pod względem gramatycznej budowy zdania.

6. NA CZY POLEGA WSPOMAGANIE TRANSLACJI PRZEZ CZŁOWIEKA W ROZWAŻANYM SYSTEMIE?

Na wstępie rozważany system został sklasyfikowany jako system translacji automatycznej wspomagany przez człowieka – tzw. HAMT (ang. *Human Aided Machine Transla-*

tion). W rozważanym systemie owo wspomaganie procesów translacyjnych przez człowieka odbywa się w dwojaki sposób.

Po pierwsze to człowiek – użytkownik systemu – dokonuje podziału tłumaczonego zdania na frazy, spełniające kolejno rolę podmiotu, orzeczenia, dopełnienia i okoliczników rozważanego zdania. Podejście takie zabezpiecza system przed niebezpieczeństwem pojawienia się wieloznaczności (ang. *ambiguity*) na poziomie analizy syntaktycznej tłumaczonego zdania, ponieważ to właśnie sam użytkownik określa w sposób jednoznaczny funkcje spełniane przez poszczególne elementy składowe tłumaczonego zdania. Istotne znaczenie posiada tutaj przestrzeganie sztywno narzuconego szyku poszczególnych elementów składowych zdania (np. szyku SVO w języku włoskim). Należy zwrócić uwagę, że podejście takie wymaga posiadania przez użytkownika systemu minimum wiedzy lingwistycznej o języku, w którym formułowane są tłumaczone przez system zdania, co niestety może być pewną przeszkodą w rozpoznaniu i popularności takiego systemu.

Po drugie, w sytuacjach wątpliwych, gdy wprowadzona przez użytkownika fraza może posiadać wiele całkowicie różnych od siebie znaczeń, czyli w przypadku wystąpienia wieloznaczności na poziomie semantycznym użytkownik systemu musi sam zdecydować, które z możliwych znaczeń tej frazy chodzi. W przeciwieństwie do rozważanej uprzednio sytuacji nie jest to trudne zadanie, ponieważ oprócz dobrej znajomości leksyki danego języka (z czym dla rodowitego użytkownika danego języka nie powinno być najmniejszych problemów – jeśli tylko tłumaczone treści nie dotyczą jakiejś wąskiej, specjalistycznej dziedziny, która jest mu obca) nie jest już tutaj potrzebna żadna wiedza lingwistyczna. Po prostu, sam system przedstawia użytkownikowi kilka propozycji do wyboru, z których musi on wybrać właściwą. Proces ten zostanie zilustrowany na kilku następujących przykładach w dalszej części artykułu. Na razie jednak należy zaznajomić się z budową rekordów bazy danych, w której zawarte są przykłady translacyjne.

7. ZAWARTOŚĆ BAZY DANYCH PRZYKŁADÓW TRANSLACYJNYCH

Głównym elementem rozważanego systemu translacji automatycznej jest baza danych o odpowiednio dużych rozmiarach. W bazie tej zawarte są nie tylko przykłady translacyjne, pokazujące jak należy tłumaczyć dane frazy, ale także wartości, jakie należy przypisać odpowiednim atrybutom, aby w tłumaczonym zdaniu na język polski zachowany został pomiędzy podmiotem i orzeczeniem związek zgody, a pomiędzy orzeczeniem i dopełnieniem zdania związek rzędu. Zatem pozyskanie takiej bazy danych o odpowiednio dużych rozmiarach stanowi klucz do poprawnej i odznaczającej się dużą skutecznością pracy systemu translacji automatycznej wspomaganej przez człowieka.

Zawarte w bazie danych frazy można ogólnie podzielić na frazy rzeczownikowe, czasownikowe i okolicznikowe.

Frazy rzeczownikowe mogą spełniać w tłumaczonych zdaniach funkcję podmiotu lub dopełnienia zdania. Z kolei frazy czasownikowe spełniają funkcje orzeczenia zdania. Natomiast frazy okolicznikowe – to okoliczniki zdania, wśród których wyróżnić można okoliczniki miejsca, czasu, sposobu, celu i przyczyny.

Rekordy bazy danych dla fraz rzeczownikowych mają następującą postać (tab. 5).

Tabela 5

„FR_włoski”	<przypadek>	
FR_polski_1	1	
FR_polski_2	2	
FR_polski_3	3	
FR_polski_4	4	
FR_polski_5	5	
FR_polski_6	6	
<osoba>	<liczba>	<rodzaj>
OSOBA	LICZBA	RODZAJ

- „FR_włoski” – oznacza frazę rzeczownikową należąca do języka włoskiego, która ma zostać przetłumaczona na język polski, w sposób zależny od wartości atrybutu <przypadek>.
- FR_polski_1 – oznacza tłumaczenie włoskiej frazy rzeczownikowej na język polski, w sytuacji gdy atrybut <przypadek> przyjmuje wartość 1.
- FR_polski_2 – oznacza tłumaczenie włoskiej frazy rzeczownikowej na język polski, w sytuacji gdy atrybut <przypadek> przyjmuje wartość 2.
- FR_polski_3 – oznacza tłumaczenie włoskiej frazy rzeczownikowej na język polski, w sytuacji gdy atrybut <przypadek> przyjmuje wartość 3.
- FR_polski_4 – oznacza tłumaczenie włoskiej frazy rzeczownikowej na język polski, w sytuacji gdy atrybut <przypadek> przyjmuje wartość 4.
- FR_polski_5 – oznacza tłumaczenie włoskiej frazy rzeczownikowej na język polski, w sytuacji gdy atrybut <przypadek> przyjmuje wartość 5.
- FR_polski_6 – oznacza tłumaczenie włoskiej frazy rzeczownikowej na język polski, w sytuacji gdy atrybut <przypadek> przyjmuje wartość 6.
- OSOBA – oznacza wartość przypisywaną atrybutowi <osoba> w celu spełnienia związku zgody pomiędzy podmiotem a orzeczeniem zdania polskiego.
- LICZBA – oznacza wartość przypisywaną atrybutowi <liczba> w celu spełnienia związku zgody pomiędzy podmiotem a orzeczeniem zdania polskiego.
- RODZAJ – oznacza wartość przypisywaną atrybutowi <rodzaj> w celu spełnienia związku zgody pomiędzy podmiotem a orzeczeniem zdania polskiego.

Jako ilustrację powyższego opisu można podać przykładowy rekord z bazy danych dla włoskiego rzeczownika *albergo* (tab. 6).

Tabela 6

<i>albergo</i>	<przypadek>	
hotel	1	
hotelu	2	
hotelowi	3	
hotel	4	
hotelem	5	
hotelu	6	
<osoba>	<liczba>	<rodzaj>
3	1	1

Jak już wspomniano, w rozważanej bazie danych rekordami mogą być nie tylko pojedyncze wyrazy, jak w zamieszczonym powyżej przykładzie, ale całe frazy rzeczownikowe pełniące w zdaniu funkcję podmiotu bądź dopełnienia. Przykład rekordu danych dla takiej frazy rzeczownikowej zamieszczono w tabeli 7.

Tabela 7

<i>la casa nuova</i>	<przypadek>	
nowy dom	1	
nowego domu	2	
nowemu domowi	3	
nowy dom	4	
nowym domem	5	
nowym domie	6	
<osoba>	<liczba>	<rodzaj>
3	1	1

Inny przykład jeszcze dłuższej frazy rzeczownikowej może wyglądać następująco (tab. 8).

Tabela 8

<i>notizie urgenti da mandare</i>	<przypadek>	
pilne wiadomości do wysłania	1	
pilnych wiadomości do wysłania	2	
pilnym wiadomościom do wysłania	3	
pilne wiadomości do wysłania	4	
pilnymi wiadomościami do wysłania	5	
pilnych wiadomościach do wysłania	6	
<osoba>	<liczba>	<rodzaj>
3	2	2

Z kolei rekordy bazy danych dla fraz czasownikowych mają następującą postać (tab. 9).

Tabela 9

„FC_włoski“	<osoba>	<liczba>	<rodzaj>
FC_polski_1	1	1	1
FC_polski_2	1	1	2
FC_polski_3	1	2	1
FC_polski_4	1	2	2
FC_polski_5	2	1	1
FC_polski_6	2	1	2
FC_polski_7	2	2	1
FC_polski_8	2	2	2
FC_polski_9	3	1	1
FC_polski_10	3	1	2
FC_polski_11	3	1	3
FC_polski_12	3	2	1
FC_polski_13	3	2	2
FC_polski_14	3	2	3
<przypadek>			
PRZYPADEK			

- „FC_włoski” – oznacza frazę czasownikową należąca do języka włoskiego, która ma zostać przetłumaczona na język polski, w sposób zależny od wartości atrybutów <osoba>, <liczba> i <rodzaj>.
- FC_polski_1 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 1, 1 i 1.
- FC_polski_2 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 1, 1 i 2.
- FC_polski_3 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 1, 2 i 1.
- FC_polski_4 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 1, 2 i 2.
- FC_polski_5 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 2, 1 i 1.
- FC_polski_6 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 2, 1 i 2.
- FC_polski_7 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybu-

ty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 2, 2 i 1.

- FC_polski_8 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 2, 2 i 2.
- FC_polski_9 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 3, 1 i 1.
- FC_polski_10 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 3, 1 i 2.
- FC_polski_11 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 3, 1 i 3.
- FC_polski_12 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 3, 2 i 1.
- FC_polski_13 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 3, 2 i 2.
- FC_polski_14 – oznacza tłumaczenie włoskiej frazy czasownikowej na język polski, w sytuacji gdy atrybuty <osoba>, <liczba> i <rodzaj> przyjmują odpowiednio wartości: 3, 2 i 3.
- PRZYPADEK – wartość przypisywana atrybutowi <przypadek>, aby pomiędzy orzeczeniem zdania polskiego a jego dopełnieniem zachodził związek rządu.

Jako przykład można podać rekord z bazy danych dla włoskiej frazy czasownikowej *hai fatto* (tab. 10).

Tabela 10

<i>hai fatto</i>	<osoba>	<liczba>	<rodzaj>
	1	1	1
	1	1	2
	1	2	1
	1	2	2
zrobiłeś	2	1	1
zrobiłaś	2	1	2
	2	2	1
	2	2	2
	3	1	1
	3	1	2
	3	1	3
	3	2	1
	3	2	2
	3	2	3
<przypadek>			
4			

W powyższym przykładzie atrybutowi <przypadek> została przypisana wartość 4, ponieważ polski czasownik „zrobić” wymaga dopełnienia w bierniku.

Z kolei rekordy bazy danych dla fraz okolicznikowych mają najprostszą z możliwych postaci, ponieważ zawierają tylko i wyłącznie przykłady tłumaczenia danej frazy okolicznikowej z języka wejściowego na język wyjściowy, które nie zależą od wartości żadnego z atrybutów (tab. 11).

Tabela 11

FO_włoski
FO_polski

- FO_włoski – oznacza frazę okolicznikową należąca do języka włoskiego, która ma zostać przetłumaczona na język polski
- FO_polski – oznacza tłumaczenie włoskiej frazy okolicznikowej na język polski

Spośród rozważanych fraz okolicznikowych wyróżnić można między innymi okoliczniki:

1) czasu, np. (tab. 12–14);

Tabela 12

<i>adesso</i>
teraz

Tabela 13

<i>poi</i>
potem

Tabela 14

<i>fra una settimana</i>
za tydzień

2) miejsca, np. (tab. 15 i 16);

Tabela 15

<i>dappertutto</i>
wszędzie

Tabela 16

<i>in un negozio</i>
w sklepie

3) sposobu, np. (tab. 17 i 18);

Tabela 17

<i>male</i>
źle

Tabela 18

<i>caramente</i>
drogo

4) pierwszego stopnia miary, np. (tab. 19 i 20);

Tabela 19

<i>noco</i>
mało

Tabela 20

<i>troppo</i>
za dużo

8. PRZYKŁADY TRANSLACJI

Aby czytelnik mógł wyrobić sobie pewne pojęcie na temat ograniczeń i możliwości prezentowanego systemu, poniżej zamieszczono przykłady uzyskanych za jego pomocą tłumaczeń wybranych zdań języka włoskiego.

Rozważone zostanie następujące włoskie zdanie wypowiedziane w czasie teraźniejszym (*presente*):

*Il sole splendeva con il suo splendore
per tutto il giorno.*

Najpierw użytkownik wprowadza do systemu podmiot rozważanego zdania (tab. 21), czyli:

<przypadek> = 1.

Tabela 21

<i>il sole</i>	<przypadek>	
słońce	1	
słońca	2	
słońcu	3	
słońce	4	
słońcem	5	
słońcu	6	
<osoba>	<liczba>	<rodzaj>
3	1	3

Następnie wprowadzane jest orzeczenie zdania (tab. 22).

Tabela 22

<i>splendeva con</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
świecił	3	1	1
świeciła	3	1	2
świeciło	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
5			

W następnej kolejności wprowadzane jest dopełnienie zdania (tab. 23).

Tabela 23

<i>il suo splendore</i>	<przypadek>		
swój blask	1		
swojego blasku	2		
swojemu blaskowi	3		
swój blask	4		
swoim blaskiem	5		
swoim blasku	6		
<osoba>	<liczba>	<rodzaj>	
3	1	1	

Na koniec podawany jest jeszcze okolicznik (tab. 24).

Tabela 24

<i>per tutto il giorno</i>
przez cały dzień

W rezultacie otrzymujemy następujące tłumaczenie zdania włoskiego:

Słońce świeciło swoim blaskiem przez cały dzień.

Jako kolejny przykład może posłużyć translacja następującego włoskiego zdania:

L'italiano conserva il modo di parlare della sua città.

Użytkownik wprowadza najpierw podmiot zdania (tab. 25):

<przypadek> =1;

Tabela 25

<i>L'italiano</i>	<przypadek>	
Włoch	1	
Włocha	2	
Włochowi	3	
Włocha	4	
Włochem	5	
Włochu	6	
<osoba>	<liczba>	<rodzaj>
3	1	1

Następnie wprowadzane jest orzeczenie zdania (tab. 26).

Tabela 26

<i>conserva</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
zachowuje	3	1	1
zachowuje	3	1	2
zachowuje	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Na koniec użytkownik wprowadza dopełnienie zdania (tab. 27).

W rezultacie na wyjściu systemu otrzymuje się następujący przekład włoskiego oryginału:

Włoch zachowuje zawsze sposób wymowy charakterystyczny dla swojego miasta.

Tabela 27

<i>il modo di parlare della sua città</i>	<przypadek>	
sposób wymowy charakterystyczny dla swojego miasta	1	
sposobu wymowy charakterystycznego dla swojego miasta	2	
sposobowi wymowy charakterystycznemu dla swojego miasta	3	
sposób wymowy charakterystyczny dla swojego miasta	4	
sposobem wymowy charakterystycznym dla swojego miasta	5	
sposobie wymowy charakterystycznym dla swojego miasta	6	
<osoba>	<liczba>	<rodzaj>
3	1	1

Należy zwrócić uwagę, że otrzymane w wyniku działania systemu przekłady zdań włoskich stanowią ich wierne tłumaczenia, a ponadto są to zdania zbudowane poprawnie z punktu widzenia reguł gramatyki języka polskiego.

Język włoski posiada w porównaniu z językiem polskim o wiele bardziej rozbudowany system czasów gramatycznych. W związku z tym poniżej zamieszczono przykłady obrazujące, jak zachodzi proces tłumaczenia w przypadku zapisania tego samego zdania w wybranych czasach gramatycznych.

Podmiotami rozważanych zdań włoskich będą zaimki osobowe, dla których odpowiednie rekordy z bazy danych wyglądają następująco (tab. 28–34).

Tabela 28

<i>io</i>	<przypadek>	
ja	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
1	1	2

Tabela 29

<i>tu</i>	<przypadek>	
ty	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
2	1	2

Tabela 30

<i>lui</i>	<przypadek>	
on	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
3	1	1

Tabela 31

<i>lei</i>	<przypadek>	
ona	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
3	1	2

Tabela 32

<i>noi</i>	<przypadek>	
my	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
1	2	2

Tabela 33

<i>voi</i>	<przypadek>	
wy	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
2	2	2

Tabela 34

<i>loro</i>	<przypadek>	
one	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
3	2	2

Z kolei w roli dopełnienia zdań tłumaczonych za pomocą rozważanego systemu, występuje wyrażenie „una tazza di caffè”, dla którego rekord bazy danych ma postać następującą (tab. 35).

Tabela 35

<i>una tazza di caffè</i>	<przypadek>	
filiżanka kawy	1	
filiżanki kawy	2	
filiżance kawy	3	
filiżankę kawy	4	
filiżanką kawy	5	
filiżance kawy	6	
<osoba>	<liczba>	<rodzaj>
3	1	2

Rozważane zdania w czasie teraźniejszym (*presente*) mają postać następującą:

Io bevo una tazza di caffè.

Tu bevi una tazza di caffè.

Lui beve una tazza di caffè.

Lei beve una tazza di caffè.

Noi beviamo una tazza di caffè.

Voi bevete una tazza di caffè.

Loro bevono una tazza di caffè.

Kolejne rekordy z bazy danych dla orzeczeń zamieszczonych powyżej zdań wyglądają następująco (tab. 36–41).

W rezultacie pracy rozważanego systemu translacji automatycznej otrzymuje się następujące tłumaczenia zdań włoskich:

Ja piję filiżankę kawy.

Ty pijesz filiżankę kawy.

On pije filiżankę kawy.

Ona pije filiżankę kawy.

My pijemy filiżankę kawy.

Wy pijecie filiżankę kawy.

One piją filiżankę kawy.

Tabela 36

<i>bevo</i>	<osoba>	<liczba>	<rodzaj>
piję	1	1	1
piję	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 37

<i>bevi</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
pijesz	2	1	1
pijesz	2	1	2
–	2	2	1
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 38

<i>beve</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
pije	3	1	1
pije	3	1	2
pije	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 39

<i>beviamo</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
pijemy	1	2	1
pijemy	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 40

<i>bevete</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
pijecie	2	2	1
pijecie	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 41

<i>bevono</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
piją	3	2	1
piją	3	2	2
piją	3	2	3
<przypadek>			
4			

Z kolei rozważane zdania włoskie w czasie przeszłym dokonanym (*passato prossimo*) mają następującą postać:

Io ho bevuto una tazza di caffè.

Tu hai bevuto una tazza di caffè.

Lui ha bevuto una tazza di caffè.

Lei ha bevuto una tazza di caffè.

Noi abbiamo bevuto una tazza di caffè.

Voi avete bevuto una tazza di caffè.

Loro hanno bevuto una tazza di caffè.

Kolejne rekordy z bazy danych dla orzeczeń zamieszczonych powyżej zdań wyglądają następująco (tab. 42–47).

Tabela 42

<i>ho bevuto</i>	<osoba>	<liczba>	<rodzaj>
wypiłem	1	1	1
wypiłam	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 43

<i>hai bevuto</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
wypiteś	2	1	1
wypitaś	2	1	2
–	2	2	1
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 44

<i>ha bevuto</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
wypił	3	1	1
wypiła	3	1	2
wypiło	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 45

<i>abbiamo bevuto</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
wypiliśmy	1	2	1
wypiliśmy	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 46

<i>avete bevuto</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
wypiliście	2	2	1
wypiliście	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
–	3	2	1
–	3	2	2
–	3	2	3
<przypadek>			
4			

Tabela 47

<i>hanno bevuto</i>	<osoba>	<liczba>	<rodzaj>
–	1	1	1
–	1	1	2
–	1	2	1
–	1	2	2
–	2	1	1
–	2	1	2
–	2	2	1
–	2	2	2
–	3	1	1
–	3	1	2
–	3	1	3
wypili	3	2	1
wypili	3	2	2
wypili	3	2	3
<przypadek>			
4			

W rezultacie pracy rozważanego systemu translacji automatycznej otrzymuje się następujące tłumaczenia zdań włoskich:

Ja wypiałam filiżankę kawy.

Ty wypitaś filiżankę kawy.

On wypił filiżankę kawy.

Ona wypila filiżankę kawy.

My wypiliśmy filiżankę kawy.

Wy wypiliście filiżankę kawy.

One wypiliły filiżankę kawy.

Należy zwrócić uwagę, że we wszystkich z przedstawionych powyżej przykładów otrzymano wierne przekłady

włoskich oryginałów, przy czym uzyskane drogą automatyczną zdania polskie są zbudowane poprawnie pod względem syntaktycznym.

Zgodnie z wcześniejszą obietnicą zostanie wyjaśnione teraz, w jaki sposób system wykorzystuje wiedzę współpracującego z nim człowieka, który wspomaga proces translacji, do rozwikływania wieloznaczności występujących na poziomie leksykalnym – czyli użytego słownictwa.

Na przykład w sytuacji, gdy użytkownik wprowadzi do systemu jako frazę rzeczownikową włoski zaimek osobowy *tu*, w niektórych przypadkach translacji mogą powstać sytuacje wątpliwe, kiedy dokonanie niewłaściwego wyboru może prowadzić do błędnego przekładu. Włoski zaimek osobowy *tu* odpowiada dokładnie polskiemu zaimkowi „ty”, ale niestety nic nie mówi na temat płci osoby do której się zwracamy. Brak takiej informacji, w niektórych przypadkach może sprawić, że system translacji automatycznej znajdzie się na rozdrożach, ponieważ będzie miał do wyboru dwie różne możliwości i niestety żadnej wskazówki odnośnie do tego, na którą z nich należy się zdecydować. To właśnie w tym momencie potrzebna jest pomoc człowieka współpracującego na bieżąco z systemem i wspomagającego proces translacji.

Zatem po wprowadzeniu do systemu słowa *tu*, system skieruje do użytkownika zapytanie o płeć tej osoby. Oczywiście zapytanie takie zostanie sformułowane w języku włoskim, ponieważ właśnie tym językiem włada użytkownik systemu. System podsunie użytkownikowi dwie propozycje:

- 1) *tu – un uomo* (mężczyzna),
- 2) *tu – una donna* (kobieta).

Jeżeli użytkownik wybierze odpowiedź nr 1, wówczas z bazy danych wybrany zostanie rekord (tab. 48).

Tabela 48

<i>tu</i>	<przypadek>	
ty	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
2	1	1

W przypadku przeciwnym, czyli po wyborze odpowiedzi nr 2, system wybierze z bazy danych rekord o postaci (tab. 49).

Jak łatwo można zauważyć, oba rekordy różnią się jedynie wartością atrybutu <rodzaj>, jednakże nawet tak subtelna różnica niesie za sobą doniosłe konsekwencje. Bowiernie, jeżeli w kroku kolejnym użytkownik wprowadzi frazę *ha mangiato*, wówczas z odpowiedniego rekordu bazy danych wybrane zostaną różne możliwości, co wynika bezpośrednio z tabeli 50.

Tabela 49

<i>tu</i>	<przypadek>	
ty	1	
–	2	
–	3	
–	4	
–	5	
–	6	
<osoba>	<liczba>	<rodzaj>
2	1	2

Tabela 50

<i>ha mangiato</i>	<osoba>	<liczba>	<rodzaj>
zjadłem	1	1	1
zjadłam	1	1	2
zjedliśmy	1	2	1
zjadłyśmy	1	2	2
zjadłeś	2	1	1
zjadłaś	2	1	2
zjedliście	2	2	1
zjadłyście	2	2	2
zjadł	3	1	1
zjadła	3	1	2
zjadło	3	1	3
zjedli	3	2	1
zjadły	3	2	2
zjadły	3	2	3
<przypadek>			
5			

Jak widać, w przypadku wybrania przez użytkownika odpowiedzi nr 1 wypisane zostanie słowo „zjadłeś”, a w przypadku wybrania odpowiedzi nr 2 „zjadłaś”. Jak łatwo można zauważyć, brak współpracy człowieka z systemem automatycznej translacji mógłby na przykład spowodować, że do mężczyzny (polskiego odbiorcy tłumaczonego komunikatu językowego) skierowana zostałaby wypowiedź „ty zjadłaś”, co oczywiście nie jest poprawne pod względem językowym i ponadto brzmi dosyć zabawnie.

Ilustracją dla innego przypadku wspomaganego procesu translacji przez człowieka może być przykład sytuacji, w której na wejście systemu podany zostanie włoski rzeczownik *stufa*, który posiada dwa całkowicie różne znaczenia:

- 1) piec.
- 2) cieplarnia.

Zatem po wprowadzeniu do systemu słowa *stufa* na ekranie monitora powinny pojawić się dwie propozycje (podpowiedzi), z których użytkownik musi wybrać właściwą:

- 1) *Tu pensi a un forno?* (Masz na myśli piec?)
- 2) *Tu pensi a una serra?* (Masz na myśli cieplarnię?)

W przypadku zdecydowania się na pierwszą z zasugerowanych przez system podpowiedzi wybrany zostanie z bazy danych następujący rekord (tab. 51).

Tabela 51

<i>stufa</i>	<przypadek>	
piec	1	
pieca	2	
piecowi	3	
piec	4	
piecem	5	
piecu	6	
<osoba>	<liczba>	<rodzaj>
3	1	1

Natomiast w przypadku zdecydowania się na podpowiedź nr 2 wybrany z bazy danych rekord będzie miał postać następującą (tab. 52).

Tabela 52

<i>stufa</i>	<przypadek>	
cieplarnia	1	
cieplarni	2	
cieplarni	3	
cieplarnię	4	
cieplarnią	5	
cieplarni	6	
<osoba>	<liczba>	<rodzaj>
3	1	2

Również w tym przypadku współpraca człowieka z systemem automatycznej translacji staje się nieodzowna, jeśli tylko tłumaczenie ma być poprawne. Bowiem niewłaściwe użycie polskich ekwiwalentów włoskiego słowa *stufa* z całą pewnością doprowadzi do błędnego zrozumienia przetłumaczonego zdania.

9. PODSUMOWANIE

Artykuł wprowadza czytelnika w fascynujący świat translacji automatycznej. Jednakże pamiętać należy o tym, że pomimo swej egzotyczności translacja automatyczna nie jest jakąś nowo powstałą dziedziną wiedzy, ponieważ historia prowadzonych w tym kierunku badań liczy sobie już ponad pięćdziesiąt lat. Od czasu pierwszego wystąpienia Warrena Weavera przebyto już długą drogę i zdano sobie dobrze sprawę ze specyfiki i trudności problemu. Obecnie mądrzejsi o kilkadziesiąt lat doświadczeń wiemy, że rozumo-

wanie Weavera było bardzo (można wręcz powiedzieć dziecinnie) naiwne, ponieważ badacz ten mylnie zakładał, że proces translacji można oprzeć na pewnym sposobie szyfrowania.

Postępując w sposób proponowany przez Weavera, można jedynie dokonywać przekładów z pewnego sztucznie wykreowanego języka (np. na bazie dowolnego języka naturalnego), poprzez odwrócenie kolejności występowania liter w poszczególnych wyrazach. Zatem znając sposób szyfrowania, można (zgodnie z postulatem Weavera) przełożyć następujące zdanie, które pozornie sprawia wrażenie, że należy do jakiegoś (zapewne egzotycznego) języka obcego:

Eno fo ym sdneirf thguob a wen rac tsal raey.

Tymczasem stosując prostą deszyfrację, polegającą na ponownym odwróceniu kolejności liter w poszczególnych wyrazach, otrzymuje się:

One of my friends bought a new car last year.

Niestety w przypadku dokonywania przekładu pomiędzy dwoma dowolnymi językami naturalnymi sprawy nie mają się już tak prosto, nie istnieją bowiem żadne reguły (nawet najbardziej złożone), które pozwoliłyby przekształcić dowolnie angielskie zdanie w jego polski ekwiwalent. Na przykład, dla zamieszczonego powyżej angielskiego zdania system translacji automatycznej powinien podać na wyjściu zdanie:

*W zeszłym roku jeden z moich przyjaciół
kupił nowy samochód.*

W istocie proces translacji jest procesem, w który zaangażowany musi zostać wysoki stopień kreatywności, wyobraźni i intuicji uczestniczącej w nim osoby. Z tego powodu translacja z wielkimi oporami poddaje się wszelkim próbom algorytmizacji i automatyzacji. Wobec wysuwanych przez niektórych badaczy i filozofów [9, 10] poważnych wątpliwości, co do możliwości zautomatyzowania procesów leżących u podstaw funkcjonowania ludzkiego umysłu, należy również powątpiewać, czy całkowite zautomatyzowanie translacji i zupełne wyeliminowanie z tego procesu człowieka, jest w ogóle możliwe.

Z tego powodu autor niniejszej pracy wysunął tezę, iż na obecnym etapie rozwoju technik przetwarzania informacji, proces translacji automatycznej – jeśli tylko ma dostarczać poprawnego przekładu – wymaga aktywnej współpracy człowieka, którego zadaniem jest rozstrzygnięcie wszelkich leksykalnych i syntaktycznych wieloznaczności tłumaczonych wypowiedzi.

Celem udowodnienia tak sformułowanej tezy autor opracował eksperymentalny system, służący do translacji zdań napisanych w języku włoskim na język polski. Pierwsze, wstępne wyniki uzyskane podczas eksperymentów przeprowadzonych z opisanym systemem wyglądają dość obie-

ciągają. System dostarcza w miarę wiernych przekładów podanych na jego wejście włoskich zdań. Również uzyskane tą drogą przekłady są zbudowane poprawnie pod względem reguł gramatyki języka polskiego, tzn. pomiędzy podmiotem a orzeczeniem zdania zachodzi związek zgody, a pomiędzy orzeczeniem zdania a jego dopełnieniem związek rządu. Niestety stylistyka tworzonych w ten sposób zdań pozostawia wiele do życzenia. Być może pewną poprawę przyniosłoby tutaj zastosowanie dodatkowego programu postprocesora, który dokonywałby przegrupowania poszczególnych elementów składniowych zdań wyjściowych, np. okoliczniki czasu umieszczałyby na początku polskiego zdania itp.

Natomiast dużą zaletą zaprezentowanego systemu jest to, iż łączy on w sobie elementy pochodzące z różnorodnych podejść do translacji automatycznej, wykorzystując jednocześnie ich najlepsze cechy. Zaczerpnięte z techniki automatycznej translacji opartej na przykładach EBMT podejście oparte na bazie danych, w której zawarte są tłumaczenia całych fraz, sprawia, że otrzymane za pomocą rozważanego systemu teksty tłumaczeń brzmią dla człowieka w sposób bardziej naturalny – nie odczuwa się tak dotkliwie sztuczności języka pojawiających się na wyjściu systemu wypowiedzi jak w przypadku innych technik translacji automatycznej. Z kolei wbudowanie do rozważanego systemu minimum wiedzy lingwistycznej, dotyczącej związków zgody i rządu, umożliwiło prawidłowe łączenie pochodzących z bazy danych przykładów translacyjnych. Również wprężenie w proces translacji aktywnie uczestniczącego w nim człowieka pozwoliło na wyeliminowanie niebezpieczeństwa wpadnięcia procesu translacji w pułapkę wieloznaczności leksykalnej bądź syntaktycznej.

Jednocześnie należy zdać sobie sprawę, że poważnym ograniczeniem zaproponowanego systemu, jest to, iż działa on jedynie na pewnym języku kontrolowanym, który jest zaledwie małym podzbiorem danego języka naturalnego. Zatem nie wszystkie zdania, które są poprawnymi zdaniami w języku włoskim mogą zostać przez opisany system zaakceptowane.

Z wymienionych powyżej powodów opisany system nie nadaje się oczywiście do tłumaczenia literatury pięknej. Jednakże odpowiedź na tak sformułowany zarzut może być tylko jedna: tłumaczenie tekstów literackich nigdy nie było, nie jest i prawdopodobnie nigdy nie będzie celem żadnego systemu translacji automatycznej. Zarzut, że jakiś system translacji automatycznej nie potrafi tłumaczyć w sposób literacki dzieł Szekspira jest absurdalny – to tak, jakby ktoś zarzucał robotom wykorzystywanym w przemyśle do automatyzacji procesów produkcyjnych i montażowych, że nie potrafią zatańczyć w balecie na lodzie. Jakie są zatem przewidywane kierunki zastosowań opisanego systemu?

Po pierwsze system taki może okazać się bardzo przydatny podczas turystycznych kontaktów osób nieznających w dostatecznym stopniu żadnego wspólnego języka obcego. Pełniłyby wówczas role popularnych rozmówek kieszonkowych, ale w porównaniu z nimi odznaczałyby się zdecy-

dowanie większą elastycznością, gdyż liczba możliwych do wygenerowania za jego pomocą zdań byłaby w praktyce nieograniczona.

Ponadto system taki, po uprzednim wzbogaceniu bazy danych w odpowiednie jednostki leksykalne, mógłby posłużyć również do tłumaczenia pewnych tekstów specjalistycznych – naukowych i technicznych. Na przykład można by za jego pośrednictwem dokonywać przekładów podręczników użytkownika i instrukcji obsługi różnych urządzeń technicznych. Również za pomocą takiego systemu można by tłumaczyć ulotki umieszczane na produktach, które są eksportowane do krajów z innego obszaru językowego. Zapewne potencjalnych zastosowań zaproponowanego systemu translacji automatycznej jest bez liku, np. można poważnie pomyśleć o tłumaczeniu za jego pośrednictwem poczty elektronicznej, i tylko od pomysłowości jego użytkowników zależy jego popularność i stopień rozpowszechnienia.

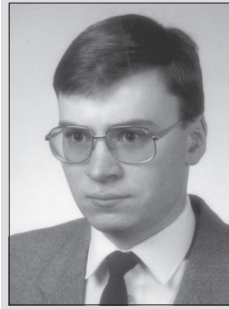
Nie należy jednakże zapominać, że o sukcesie zaproponowanego systemu zadecyduje ostatecznie rozmiar jego bazy danych. Chodzi bowiem o to, aby zgromadzone tam jednostki leksykalne można było zastosować do jak największej liczby spotykanych w praktyce językowej wyrazów i związków frazeologicznych. Trzeba dobrze zdawać sobie sprawę jak niełatwe i jak czasochłonne jest to zadanie, bowiem każdy język naturalny jest wręcz przebogata strukturą, a leksyka każdego języka to prawdziwa mozaika. W praktyce efektywna konstrukcja takiej bazy danych przekracza możliwości pojedynczego człowieka – w istocie, jest to zadanie dla dużego zespołu złożonego z informatyków, filologów i lingwistów.

Literatura

- [1] Baker K.L., Franz A.M., Jordan P.W., Mitamura T., Nyberg E.H.: *Coping with ambiguity in a large-scale machine translation system*. Pittsburgh, Center for Machine Translation, Carnegie Mellon University 1992
- [2] Arnold D., Balkan L., Meijer S., Humphreys R.L., Sadler L.: *Machine translation: An introductory guide*. London, NCC Blackwell 1994
- [3] Melby A.: *Machine translation and philosophy of language*. Machine Translation Review, No. 9, April 1999, pp. 6–17
- [4] Blekhman M., Pevzner B.: *First Steps of language engineering in the USSR: The 50s through 70s*. Machine Translation Review, Issue No. 11, December 2000, pp. 5–7
- [5] Waibel A., Geutner P., Tomokiyo L.M., Schultz T., Woszczyna M.: *Multilinguality in speech and spoken language systems*. Proceedings of the IEEE, vol. 88, No. 8, August 2000, pp. 1297–1313
- [6] Zue V.W., Glass J.R.: *Conversational interfaces: advances and challenges*. Proceedings of the IEEE, vol. 88, No. 8, August 2000, pp. 1166–1180
- [7] strona internetowa firmy Techland o adresie <http://www.techland.com.pl>
- [8] Penrose R.: *Nowy umysł cesarza*. Warszawa, Wydawnictwo Naukowe PWN 1995, ISBN 83-01-11819-9
- [9] Wiślicki W.: *Czy możliwy jest naukowy opis tego, co dzieje się w umyśle?* *Wszechświat*, t. 101, nr 1–3, 2000, pp. 8–12
- [10] Penrose R.: *Makroświat, mikroświat i ludzki umysł*. Warszawa, Wydawca Prószyński i S-ka 1997

- [11] Carbonell J.G., Mitamura T., Nyberg E.H.: *The KANT perspective: A critique of pure transfer (and pure interlingua, pure statistics, ...)*. Pittsburgh, USA, Center for Machine Translation, Carnegie Mellon University 1996
- [12] 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, No. 1, January 2000, pp. 24–36
- [13] Canals R., Esteve A., Garrido A., Guardiola M.I., Iturraspe-Bellver A., Montserrat S.: *InterNOSTRUM: A Spanish-Catalan machine translation system*. Machine Translation Review, No. 11, December 2000, pp. 21–25
- [14] Loukachevitch N.V., Dobrov B.V.: *Thesaurus-based structural thematic summary in multilingual information systems*. Machine Translation Review, No. 11, December 2000, pp. 10–20
- [15] Fukutomi O.: *Report on commercial machine translation in a manufacturing domain*. Machine Translation Review, No. 11, December 2000, pp. 16–25
- [16] Murphy D.: *Keeping translation technology under control*. Machine Translation Review, No. 11, December 2000, pp. 7–10
- [17] Nyberg E., Mitamura T., Carbonell J.: *The KANT machine translation system: From R&D to initial deployment*. Centre for Machine Translation, Carnegie Mellon University, USA, 1998
- [18] Mitamura T.: *Controlled languages for multilingual machine translation*. Centre for Machine Translation, Carnegie Mellon University, USA, 1996
- [19] Mitamura T., Nyberg E.H., Carbonell J.: *An efficient interlingua translation system for multi-lingual document production*. Centre for Machine Translation, Carnegie Mellon University, USA, 1998
- [20] Kreisberg A., de Fanti S.: *Mówimy po włosku*. Warszawa, Wiedza Powszechna 1992, ISBN 83-214-0282-8

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