

Mirosław GAJER

Akademia Górniczo-Hutnicza w Krakowie, Katedra Automatyki

PROPOZYCJA ZASTOSOWANIA TECHNIKI TRANSLACJI AUTOMATYCZNEJ OPARTEJ NA WZORCACH W SYSTEMIE TŁUMACZĄCYM Z JĘZYKA HISZPAŃSKIEGO NA POLSKI

Streszczenie. Translacja automatyczna jest dziedziną zastosowań informatyki dostarczającą wiedzy o tym, jak programować komputery, aby były w stanie dokonywać automatycznych przekładów pomiędzy językami naturalnymi. Pomimo prawie pół wieku badań nad translacją automatyczną zadanie to jest jeszcze dalekie od w pełni satysfakcjonującego rozwiązania. Obecnie duże nadzieje wiąże się ze stosunkowo niedawno opracowanymi nowymi koncepcjami, takimi jak translacja oparta na przykładach, leksykony fraz oraz tzw. pamięci translacyjne. W artykule autor zamieścił opis opracowanej przez niego metody automatycznej translacji, nazwanej translacją opartą na wzorcach oraz przedstawił propozycje zastosowania tej metody w systemie tłumaczącym z języka hiszpańskiego na polski.

Słowa kluczowe: lingwistyka komputerowa, translacja automatyczna.

THE PROPOSITION OF USING PATTERN-BASED MACHINE TRANSLATION TECHNIQUE FOR A SPANISH-TO-POLISH AUTOMATIC TRANSLATION SYSTEM

Summary. Machine translation is a branch of computer-science the aim of which is to develop computer programs that would be able to translate between human languages. High-quality fully-automatic machine translation has for a long time been a purpose of an intensive scientific effort. After over 50 years of research the problem of automation of translation is still far away from its fully satisfactory solution. Quite recently totally new approaches to machine translation have been proposed. These are: Example-Based Machine Translation, Phrasal Lexicons and Translation Memories. In the paper this author's original machine translation method is proposed. The method was called by this author Pattern-Based Machine Translation and it is a modification of Phrasal Lexicons methods.

Keywords: computational linguistics, machine translation.

1. Wprowadzenie

Translacja automatyczna jest dziedziną zastosowań informatyki, która dostarcza wiedzy o tym, w jaki sposób pisać programy komputerowe służące do dokonywania automatycznych przekładów pomiędzy wybranymi językami naturalnymi [1]. Pomimo ponad 50 lat intensywnych badań w tej dziedzinie w pełni automatyczny system komputerowego tłumacza, dostarczający przekładów o wysokiej jakości, wciąż pozostaje niedoścignionym ideałem. W ostatnich latach zaproponowano nowe metody translacji automatycznej, które pozwalają rokować pewne nadzieje na postęp w rozważanej dziedzinie zastosowań komputerów [2]. Jedną z tych metod jest tzw. translacja oparta na leksykonach związków frazeologicznych (ang. Phrasal Lexicons).

Główne idee związane z leksykonami związków frazeologicznych zostały przedstawione już w 1975 roku przez amerykańskiego lingwistę Josepha Beckera [6]. Becker w swoim referacie [7] poczynił słuszną uwagę, że lingwiści, podobnie jak większość badaczy związanych z innymi dyscyplinami, chcieliby najchętniej być fizykami i posiadać wdzięczny obiekt badań, który zawsze zachowuje się w sposób deterministyczny, i w związku z tym, jest możliwe odkrycie ściślejszych praw nim rządzących [6]. Niestety język naturalny takim obiektem badań nie jest [8]. W związku z tym Becker zaproponował podejście całkowicie odmienne od obowiązujących do tej pory trendów w lingwistyce. Zamiast uważać proces produkcji językowej jako polegający na dokonywaniu zestawień słów bądź morfemów, które dopiero tworzą wypowiedzi, Becker zaproponował coś całkowicie odmiennego. Uznał mianowicie, że dopiero związki frazeologiczne, zbudowane co najmniej z dwóch wyrazów, stanowią prawdziwy budulec wypowiedzi językowych [6]. Współczesne badania zdają się potwierdzać hipotezę Beckera. Okazuje się bowiem, że ludzie budują swoje wypowiedzi głównie przez powtarzanie, modyfikowanie i łączenie uprzednio poznanych związków frazeologicznych, które są na bieżąco adaptowane do nowych sytuacji [6]. Takie związki frazeologiczne nazywane są stereotypami i występują w każdym języku naturalnym w wielkiej obfitości [9]. Jak dowodzi Becker, proces działalności językowej polega co najmniej w równym stopniu na pozyskiwaniu z pamięci gotowych stereotypów, z których budowana jest wypowiedź, co na generacji całkowicie nowych wypowiedzi. Dzieje się tak dlatego, że wiele sytuacji nie wymaga twórczej działalności językowej, ponieważ daje się zaliczyć do jednej ze znanych osobie generującej wypowiedź kategorii [6]. Zamiast tworzyć coraz to nowsze warianty wypowiedzi, ludzie najchętniej operują gotowymi formułami, banałami, idiomami, aluzjami, sloganami i innymi podobnymi wzorcami szablonowymi [7].

System translacji automatycznej, którego działanie oparte zostało na leksykonie związków frazeologicznych, zwanych dalej po prostu frazami, korzysta z dwujęzycznego

słownika fraz, w którym każdej frazie języka źródłowego odpowiada dokładnie jedna fraza w języku docelowym przekładu. Sposób działania systemu translacji automatycznej opartego na leksykonie fraz bardzo przypomina pracę modułu sprawdzania poprawności pisowni w edytorze tekstowym takim, jak np. popularny „Word”. Po prostu system odnajduje w leksykonie kolejne frazy, z których zbudowany jest tłumaczony tekst i podstawia w ich miejsce ich odpowiedniki w języku docelowym. Spektakularny przykład implementacji takiego systemu pochodzi z Hiszpanii, gdzie bilingwiczny leksykon fraz został utworzony dla języków hiszpańskiego i katalońskiego [4]. Jakość pracy rozważanego systemu translacji automatycznej jest na tyle wysoka, że został on z powodzeniem zastosowany do mechanicznego tłumaczenia z języka hiszpańskiego na kataloński czasopisma „El Periódico de Catalunya”, które następnie już bez żadnych poprawek kierowane jest do druku w nakładzie 60.000 egzemplarzy. Szybkość pracy rozważanego systemu wynosi około 7KB tekstu na sekundę. Jakość przekładu jest wysoka, głównie dzięki temu, iż system jest ciągle udoskonalany, w sposób polegający na tym, że sztab lingwistów systematycznie uzupełnia bilingwiczny leksykon fraz, wprowadzając doń nowe związki frazeologiczne o maksymalnej długości sześciu wyrazów [4]. Jest to prawdopodobnie do chwili obecnej najbardziej spektakularny przykład wykorzystania systemu translacji automatycznej do tłumaczenia tekstów należących do języka ogólnego, które zakończyło się tak dużym sukcesem. Niestety tajemnica tego sukcesu polega głównie na bardzo bliskim pokrewieństwie genetycznym pomiędzy językami źródłowym i docelowym przekładu. Należy oczekiwać, że w przypadku zastosowania rozważanej metody do automatycznego tłumaczenia tekstów pomiędzy językami bardzo od siebie odległymi w kategoriach klasyfikacji genetycznej, typologicznej i geograficznej uzyskane rezultaty nie będą się odznaczały już tak wysoką jakością. Jednakże ostateczną odpowiedź może dać tutaj tylko zbudowanie działającego systemu translacji automatycznej, opartego na bilingwicznym leksykonie fraz oraz wykonanie odpowiedniej ilości testów. Autor niniejszego artykułu zaproponował pewną modyfikację techniki translacji automatycznej opartej na leksykonie fraz, którą nazwał translacją opartą na wzorcach (ang. Pattern-Based Machine Translation). Modyfikacja ta pozwala na dodatkowe uwzględnienie zjawisk związanych z fleksją języka polskiego.

2. Translacja automatyczna oparta na wzorcach

Podstawową zaletą techniki translacji automatycznej opartej na leksykonie fraz jest to, że zapisane w nim frazy w sposób automatyczny zawierają znaczną dawkę informacji kontekstowej. Zdaniem autora jest to najważniejsza cecha metody translacji opartej na bilingwicznym leksykonie fraz, która jednocześnie decyduje o jej przewadze nad pozostałymi

metodami, w których tłumaczenie odbywa się wyraz po wyrazie. Okazuje się bowiem, że dysponując jedynie pojedynczym wyrazem, wyrwanym z kontekstu wypowiedzi, bardzo trudno jest w sposób jednoznaczny ustalić jego odpowiednik znaczeniowy w języku przekładu. Odpowiedzialne jest za to zjawisko wieloznaczności języka naturalnego, objawiające się na poziomie leksykalnym wypowiedzi. Na przykład angielskie słowo „case” zgodnie ze słownikiem [10], rozpatrywane wyłącznie w funkcji rzeczownika, może mieć aż 28 różnych polskich odpowiedników:

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

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

Zjawisko wieloznaczności leksykalnej nie jest jakąś szczególną cechą, charakterystyczną tylko i wyłącznie dla języka angielskiego. Wręcz przeciwnie, jest ono wszechobecne w każdym języku naturalnym. W istocie, otwierając jakikolwiek dwujęzyczny słownik, np. francusko-polski, trudno jest doprawdy znaleźć wyrazy, które posiadałyby tylko jeden odpowiednik znaczeniowy w języku docelowym. Prawdopodobieństwo losowego trafienia przez system translacji automatycznej we właściwy ekwiwalent znaczeniowy tłumaczonego słowa jest niewielkie. Natomiast podstawowa zaleta metody translacji opartej na bilingwicznym leksykonie fraz polega właśnie na tym, że nie są tłumaczone pojedyncze wyrazy, ale ich zbitki. Z kolei, jak wykazały badania [11], w celu ustalenia właściwego odpowiednika tłumaczonego słowa wystarczy dysponować jego kontekstem, na który składa się jedno słowo poprzedzające tłumaczony wyraz lub jedno słowo bezpośrednio po tym wyrazie następujące. W sporadycznych przypadkach potrzebne są dwa takie słowa. Spostrzeżenie to wyjaśnia, dlaczego metoda translacji automatycznej oparta na bilingwicznym leksykonie fraz odznacza się tak dużą skutecznością.

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

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

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

Wprowadzone przez autora wzorce translacyjne pełnią rolę uogólnionych bilingwicznych fraz, co oznacza, że zawierają one wszelkie potencjalnie możliwe tłumaczenia frazy zapisanej w języku źródłowym na język docelowy. Ponieważ w języku polskim występują takie zjawiska, jak deklinacja rzeczowników oraz koniugacja czasowników, autor wprowadził trzy typu wzorców translacyjnych: wzorce typu deklinacyjnego, wzorce typu koniugacyjnego oraz wzorce nie podlegające fleksji. Wprowadzone zostały ponadto trzy atrybuty takie, jak <przypadek>, <liczba> i <rodzaj>. Od aktualnej wartości tych atrybutów zależy, który z możliwych przekładów frazy w języku źródłowym zostanie wybrany spośród kilku różnych propozycji oferowanych przez wzorzec.

Pierwszym typem wzorca translacyjnego wprowadzonym przez autora jest tzw. wzorzec translacyjny podlegający deklinacji. Za pomocą tego typu wzorca tłumaczone są przede wszystkim frazy rzeczownikowe, których polskie ekwiwalenty odmieniają się przez przypadki. Od aktualnej wartości atrybutu <przypadek> zależy, które z możliwych tłumaczeń rozważanej frazy rzeczownikowej zostanie wybrane. Wartość atrybutu <przypadek> jest zawsze ustawiana na początku nowego zdania na 1, ponieważ w języku polskim podmiot zdania występuje zawsze w mianowniku. Następnie w trakcie tłumaczenia kolejnych fraz wartość atrybutu <przypadek> jest odpowiednio modyfikowana, tak aby pomiędzy

orzeczeniem zdania a jego podmiotem zachodził związek zgody. Ponadto pierwszy typ wzorca translacyjnego zawiera jeszcze informację na temat wartości, jakie należy przypisać atrybutom <liczba> i <rodzaj>, które są następnie wykorzystywane przez drugi typ wzorca translacyjnego.

Drugim typem wzorca translacyjnego jest wzorzec translacyjny podlegający koniugacji. Za pomocą tego typu wzorca translacyjnego tłumaczone są głównie frazy czasownikowe. Wzorzec translacyjny typu koniugacyjnego zawiera wszelkie możliwe polskie przekłady frazy czasownikowej. Prawidłowy przykład jest wybierany na podstawie aktualnych wartości atrybutów <liczba> i <rodzaj>. Ponadto wzorzec koniugacyjny zawiera jeszcze informację o wartości, jaką należy przypisać atrybutowi <przypadek>, aby pomiędzy orzeczeniem zdania a jego dopełnieniem zachodził związek rządu.

Ostatnim typem wzorca translacyjnego jest wzorzec nie podlegający fleksji. Wzorzec ten zawiera frazę źródłową i jej odpowiednik w języku docelowym. Zatem jest to bezpośrednie wykorzystanie metody typu Phrasal Lexicon, z tym wyjątkiem, że wzorzec taki może opcjonalnie zawierać informację na temat tego, jakie wartości należy przypisać atrybutom <przypadek>, <liczba> i <rodzaj>.

Zaproponowana przez autora technika translacji automatycznej oparta na wzorcach została zastosowana w eksperymentalnym systemie dokonującym automatycznych przekładów z języka hiszpańskiego na polski. O wysokiej jakości możliwych do uzyskania za pomocą proponowanej metody przekładów świadczy następujący eksperyment, w którym zadaniem systemu było dokonanie przekładu następującego hiszpańskiego tekstu:

Quizás hace un siglo no habría parecido probable: hoy es realidad: los novelistas iberoamericanos se han hecho famosos. Incluso más: han dado un nuevo impulso, un nuevo vigor y un nuevo sello a la novelística. Tanto es así que los editores se disputan la edición de las obras de algunos autores, y algunos escritores y críticos españoles han notar, no sin nostalgia, que la novela sudamericana está de moda y que los sudamericanos se llevan todos los premios literarios; lo cual equivale a decir que para ser aceptado es preciso que los escritores sigan las formas en boga. Pero lo cierto es que los novelistas latinoamericanos han dado un verdadero ejemplo de cómo se pueden dar valores nuevos a las formas literarias, acercándose a los problemas del presente y utilizando, si es preciso, formas nuevas para realidades nuevas.

Borges es la figura relevante de la novela actual hispanoamericana. Jorge Luis Borges en estos momentos está trabajando en un libro de cuentos y un atlas de todos sus viajes, del que dijo que no es un libro para nadie; es un deseo exclusivamente mío. También, y en la colaboración con su compañera María Kodama, hace poco Jorge Luis Borges ha traducido un libro del islandés. Jorge Luis Borges dijo que escribe obrando como si pudiera elegir y

tratando de elegir bien. Además Jorge Luis Borges afirmó que no espera ganar el Nobel porque no lo mereze. Jorge Luis Borges dijo que aun quiere vivir a pesar de todos sus males: ceguera, úlcera, hernia y senectud; pero el ya no siente la necesidad de la acción.

W wyniku pracy systemu translacji automatycznej opartego na wzorcach otrzymano następujący polski przekład hiszpańskiego oryginału:

Być może sto lat temu nie wydawałoby się to możliwe obecnie jest rzeczywistością pisarze południowoamerykańscy stali się sławni. Co więcej dali nowy bodziec nową energię i nowe piętno powieściopisarstwu. Dochodzi nawet do tego że wydawcy walczą o wydanie dzieł pewnych autorów i niektórzy pisarze i krytycy hiszpańscy odnotowali nie bez żalu że powieść południowoamerykańska jest modna oraz że mieszkańcy Ameryki Południowej zabierają ze sobą wszystkie nagrody literackie co w sposób równoważny można powiedzieć że aby zostać zaakceptowanym trzeba żeby pisarze naśladowali formy będące w modzie. Ale prawda polega na tym że pisarze z Ameryki Łacińskiej dali prawdziwy przykład w jaki sposób można dać nowe wartości formom literackim zbliżając się do problemów teraźniejszości i używając jeśli tak można powiedzieć nowe formy dla nowych realiów.

Jerzy Luiz Borges jest znaczącą postacią współczesnej powieści latynoamerykańskiej. Jerzy Luiz Borges obecnie pracuje nad zbiorem opowiadań i atlasem wszystkich jego podróży o którym powiedział że nie jest książką dla nikogo jest wyłącznie moim pragnieniem. Również i przy współpracy z jego przyjaciółką Marią Kodamą całkiem niedawno Jerzy Luiz Borges przetłumaczył książkę z języka islandzkiego. Jerzy Luiz Borges powiedział że pisze dzieło w jaki sposób należy dokonywać wyborów i traktat o dobrym wyborze. Ponadto Jerzy Luiz Borges zapewnił że nie oczekuje zdobycia Nagrody Nobla ponieważ nie zasługuje na to. Jerzy Luiz Borges powiedział że jeszcze pragnie żyć pomimo wszystkich jego chorób ślepoty wrzodu przepukliny i starości ale on już nie odczuwa potrzeby działania.

Do dokonania tego przekładu system translacji automatycznej wykorzystał 105 wzorców translacyjnych, spośród których 33 miały długość jednego wyrazu, 31 dwóch wyrazów, 26 trzech wyrazów, 11 czterech wyrazów, 2 pięciu wyrazów oraz 2 sześciu wyrazów.

3. Uwagi końcowe

Zamieszczony w poprzednim punkcie przykład wykorzystania techniki translacji automatycznej opartej na wzorcach do przekładu na język polski hiszpańskiego tekstu dowodzi dużej skuteczności zaproponowanej przez autora metody. Pamiętać należy jednakże

o tym, że opracowanie bazy danych zawierającej okazałą liczbę wzorców translacyjnych jest bardzo czasochłonnym procesem wymagającym dużego nakładu pracy zespołu lingwistów.

Podsumowując uzyskane rezultaty badań należy jeszcze raz podkreślić, że translacja automatyczna nie należy do tej klasy zagadnień, które można rozwiązać jakimś jednym śmiałym pociągnięciem – błyskotliwym pomysłem [12], [13]. W rzeczywistości opracowanie w miarę sprawnie działającego systemu translacji automatycznej wymaga wielkiego nakładu pracy – zwykle zespół złożony z kilkunastu a nawet kilkudziesięciu specjalistów (lingwistów i informatyków) musi pracować przez okres kilku lat [15].

System translacji automatycznej o jakości działania równej efektywności pracy tłumacza przysięgłego pozostanie (przynajmniej przy obecnym stanie nauki – prawdziwa rewolucja w fizyce zdarza się rzadziej niż raz na sto lat) prawdopodobnie na zawsze niedoścignionym ideałem – św. Gralem poszukiwań badawczych [3]. Dlatego też każdy system automatycznej translacji będzie popełniał błędy – chodzi tylko o to, aby ich liczba była jak najmniejsza. Zatem prace zmierzające w kierunku opracowania doskonałego systemu translacji automatycznej przypominają nieco zmagania Don Kichota.

Każdy system translacji automatycznej wymaga nieprzerwanej pracy nad jego udoskonalaniem. Powodem takiego stanu rzeczy jest chociażby fakt, polegający na tym, że każdy język naturalny zmienia się – wciąż do języka wprowadzane są nowe wyrazy (najczęściej zapożyczenia z innych języków) oraz tworzone są nowe związki frazeologiczne. Zatem praca nad systemem translacji automatycznej tak naprawdę nigdy się nie kończy, gdyż wciąż trzeba sprawować nad nim kontrolę, aby jego jakość działania nie uległa pogorszeniu wraz z upływem czasu. Zatem utrzymanie sprawnego systemu translacji automatycznej przypomina nieco pracę Syzyfa, ponieważ bazę danych należy systematycznie uzupełniać nowymi jednostkami leksykalnymi [4].

W przyszłości należy oczekiwać, że wraz z powstawaniem i rozrastaniem się wielonarodowych tworów politycznych takich, jak Unia Europejska, generowane będzie drastyczne zwiększenie zapotrzebowania na translację wszelkiego rodzaju dokumentów. Jest to bardzo poważny problem wymagający zaangażowania olbrzymiej armii tłumaczy i okazałych środków finansowych [14]. Niestety, jak dotąd wszelkie próby uwolnienia ludzkości od przekleństwa Wieży Babel, poprzez wprowadzenie języków międzynarodowych takich, jak na przykład: esperanto, ido, adamitik, interlingua Peano, interlingua Godego, novial, occidental, solresol, volapük i wielu innych (ich liczba jest porównywalna z liczbą języków naturalnych) nie doprowadziło do niczego rozsądnego, ponieważ żaden z tych języków nie zdobył sobie powszechnego uznania. Zatem, jak widać, próbując zniszczyć prawdziwą Wieżę Babel, de facto obok wybudowano drugą o wcale nie mniejszych rozmiarach. Wszystko wskazuje na to, że ludzkość (przynajmniej w dającej się przewidzieć

przyszłości) jest skazana na wielojęzyczność, która będzie kreowała coraz większe potrzeby translacyjne [5].

LITERATURA

1. Gajer M.: System translacji automatycznej oparty na dialogu z użytkownikiem i przykładach translacyjnych. *Pro Dialog* 14(2002), Wydawnictwo NAKOM, Poznań, ss. 77-115.
2. Gajer M.: Fakty i mity na temat systemów translacji automatycznej. *Elektronizacja* 1-2/2002, Wydawnictwo SIGMA-NOT, Warszawa, ss. 33-36.
3. 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.
4. Pérez R. C.: From Novelty to Ubiquity: Computers and Translation at the Close of Industrial Age. <http://www.accurapid.com/journal/15mt2.htm>.
5. Nyberg E., Mitamura T., Carbonell J.: *The KANT Machine Translation System: From R&D to Initial Deployment*. Center for Machine Translation, Carnegie Mellon University, Pittsburgh, USA, 1997.
6. Schäler R.: Beyond Translation Memories. *MT Summit VIII*, Santiago de Compostela, 2001, <http://www.eamt.org/summitVIII/workshop-papers.html>.
7. Becker J. D.: *The Phrasal Lexicon*, *Proceedings of Theoretical Issues in Natural Language Processing*, Massachusetts, 1975, pp. 70-73.
8. Majewicz A. F.: *Języki świata i ich klasyfikowanie*. PWN, Warszawa 1989.
9. Lebedziński H.: *Przekładoznawstwo ogólne wobec teorii enroi*. PWN, Warszawa 1989.
10. Stanisławski J.: *Wielki słownik angielsko-polski z suplementem*. Państwowe Wydawnictwo Wiedza Powszechna, Warszawa 1983.
11. Hutchins W. J.: *Machine Translation - Past, Present, Future*, *Ellis Horwood Series in Computers and Their Applications*, London 1986.
12. Ludskanow A.: *Tłumaczy człowiek i maszyna cyfrowa*. Wydawnictwa Naukowo-Techniczne, Warszawa 1973.
13. Vasconcellos M.: Users of systems: The current state of MT usage or: How do I use these? Let me count the ways. *Fourth Machine Translation Summit in Kobe*, 1993, <http://www.eamt.org/archive/summit93.html>.
14. Rutkowski P.: Jak wiele można przetłumaczyć? *Wiedza i Życie*, Luty 2001, ss. 42-44.

15. Gerber L.: Working Toward Success in Machine Translation. MT Summit VIII, Santiago de Compostela, 2001, <http://www.eamt.org/summitVIII/workshop-papers.html>.

Recenzent: Dr inż. Nina Suszczańska

Wpłynęło do Redakcji 12 marca 2003 r.

Abstract

Machine translation is a science that delivers the knowledge how to program the computers, so as they were able to translate between human languages. The field of machine translation is almost as old as the invention of computer itself. In 1949 an American scientist Warren Weaver sent the memorandum to The Rockefeller Foundation, in which he demanded starting the research on the automation of translation between natural languages. Warren Weaver was inspired by cryptographic techniques, which were developed very strongly during the years of The Second World War, and he thought that there existed some fundamental similarities between these cryptographic techniques and the process of translation between human languages. After the first successes it soon appeared that the problem of machine translation is far more complicated and far more harder than Warren Weaver had ever imagined. After over 50 years of research the problem of automation of translation is still far away from its final solution. In the last years new machine translation approaches were proposed. Among these one can distinguish Example-Based machine Translation, Phrasal Lexicons, and Translation Memories. In the paper this author has proposed his original machine translation method which had been developed by himself especially for the Polish language. This method allows to take into account the peculiarities of the Polish grammar and especially the flexion nature of the Polish language. The obtained primary results are good and they allow to cherish the hope for building in the future a fully-automatic Spanish-to-Polish machine translation system of relatively high quality of output texts.

Adres

Mirosław GAJER: Akademia Górniczo-Hutnicza, Katedra Automatyki, al. Mickiewicza 30, 30-059 Kraków, Polska, mgajer@ia.agh.edu.pl.